

MINISTERE DE L'AGRICULTURE, DE L'AGROALIMENTAIRE ET DE LA FORET
ECOLE NATIONALE SUPERIEURE des SCIENCES AGRONOMIQUES de BORDEAUX AQUITAINE
1, cours du Général de Gaulle - CS 40201 – 33175 GRADIGNAN cedex

M E M O I R E de fin d'études
Pour l'obtention du titre

D'Ingénieur de Bordeaux Sciences Agro

Standardisation et interopérabilité dans le milieu du biologing : utilisation d'outils existants et prémice d'un système d'information

Poisot Jules

Spécialisation : AgroTIC

Étude réalisée à : Natural Solutions, 68 Rue Sainte 13001 Marseille

– 2 0 2 0 –

MINISTERE DE L'AGRICULTURE, DE L'AGROALIMENTAIRE ET DE LA FORET

ECOLE NATIONALE SUPERIEURE des SCIENCES AGRONOMIQUES de BORDEAUX AQUITAINE
1, cours du Général de Gaulle - CS 40201 – 33175 GRADIGNAN cedex

M E M O I R E de fin d'études
Pour l'obtention du titre

D'Ingénieur de Bordeaux Sciences Agro

Standardisation et interopérabilité dans le milieu du biollogging : utilisation d'outils existants et prémice d'un système d'information

Standardization and interoperability in the biollogging environment: use of existing tools and the beginnings of an information system

Poisot Jules

Maitre de stage : Aurélie Jambon

Tuteur de stage : Philippe Vismara

Remerciements

Je voudrais d'abord remercier ma maîtresse de stage, Aurélie Jambon, pour toute l'attention dont elle a fait preuve au long de ce stage.

J'aimerais ensuite remercier l'ensemble de l'équipe Natural Solutions qui m'a permis de vivre un stage agréable malgré les conditions particulières. Plus particulièrement, je remercie Jean-Vitus Albertini et Javier Blanco pour le partage de leurs connaissances et de leur expérience, Manon Frédout pour sa pédagogie, Angélique Ries pour le temps qu'elle m'a accordé, Adrien Pajot pour son aide dans la rédaction de ce mémoire ainsi que Frédéric Berton pour sa confiance et sa bienveillance.

Je souhaiterais également remercier mon tuteur de stage, Philippe Vismara, pour ses conseils précieux ainsi que l'ensemble de l'équipe pédagogique AgroTIC pour la qualité de l'enseignement dispensé au cours de la formation.

Je tiens aussi à remercier mes proches pour leur soutien qui a réellement compté en ces temps particuliers.

Enfin, je voudrais remercier Alexandra Elbakyan, sans qui la rédaction de ce mémoire et de nombreux scientifiques auraient été beaucoup plus compliqués, pour son importante contribution à la diffusion du savoir.

Table des matières

| | |
|---|----|
| Introduction | 1 |
| I – Un stage réalisé au sein d’une entreprise aux valeurs fortes | 3 |
| A) Présentation de Natural Solutions | 3 |
| 1. Une entreprise mettant l’innovation au service de la biodiversité..... | 3 |
| 2. Fonctionnement interne de l’entreprise | 4 |
| B) Déroulement et étapes clés du stage | 6 |
| II – Le monde du bio-logging, son histoire, ses promesses et ses difficultés | 7 |
| A) Genèse et démocratisation du bio-logging | 7 |
| 1. Les premières études de bio-logging | 7 |
| 2. Le début de l’utilisation des satellites | 9 |
| 3. L’avènement du bio-logging | 11 |
| B) L’absence de standardisation dans le monde du bio-logging | 13 |
| 1. Une problématique causée par le développement rapide et non structuré du milieu | 13 |
| 2. Les premières propositions de réponse à ce problème complexe..... | 15 |
| C) Les difficultés pour maximiser la valorisation des données | 16 |
| 1. La prolifération d’outils..... | 16 |
| 2. La redondance des outils | 17 |
| III – Étude d’un milieu riche en profils et en outils | 18 |
| A) Étude des profils rencontrés dans le milieu du bio-logging, leurs caractéristiques, leurs douleurs et donc leurs besoins | 19 |
| 1. Une enquête adressée à différents acteurs du milieu du bio-logging | 19 |
| 2. Les personae définis à partir des résultats de l’enquête..... | 23 |
| B) Étude d’outils existants | 24 |
| 1. Movebank, une plateforme multifonctions phare dans le milieu du bio-logging..... | 25 |
| 2. Le Parc National des Écrins, une locomotive pour la diffusion des données de biodiversité en France, et ses outils..... | 30 |
| 3. Animal Movement Cleaner Application, une application de traitement de données de mouvements d’animaux..... | 33 |
| C) Le potentiel détenu par l’association de ces outils | 35 |
| IV – Les réalisations du stage..... | 37 |
| A) Des priorités et des choix techniques | 37 |
| B) Des réalisations diverses..... | 40 |
| 1. Un module de requêtage de l’API de Movebank | 40 |
| 2. La comparaison de jeux de données sur AMCA | 42 |
| 3. Le viewer, plateforme de diffusion des données destinée au grand public | 44 |
| V – Bilan et perspectives..... | 45 |

| | | |
|----|---|----|
| A) | Prise de recul sur le travail effectué | 45 |
| 1. | Bilan sur les réalisations techniques..... | 45 |
| 2. | Bilan global sur le déroulement du stage | 46 |
| B) | Perspectives d'évolution..... | 46 |
| 1. | Amélioration des fonctionnalités existantes | 47 |
| 2. | Apport de nouvelles fonctionnalités | 47 |
| | Conclusion..... | 49 |
| | Bibliographie..... | 50 |

Table des illustrations

| | |
|---|----|
| Figure 1. Organigramme de Natural Solutions..... | 3 |
| Figure 2. Principes de fonctionnement de la méthodologie Scrum. (Subra, 2019) | 4 |
| Figure 3. Les différents cercles mis en place au sein de Natural Solutions..... | 5 |
| Figure 4. Planning du stage. | 6 |
| Figure 5. Nombre d'articles annuels portant sur la thématique de l'activité de plongée reposant sur le bio-logging (barres) ou sur de simples observations visuelles (points) en fonction du temps..... | 9 |
| Figure 6. Exactitude des émetteurs de seconde génération de Telonics..... | 10 |
| Figure 7. Evolution du poids en g de la plus petite balise de bio-logging en fonction du temps..... | 12 |
| Figure 8. Nombre d'espèces de mammifères (gauche) et d'oiseaux (droite) en fonction de leur masse corporelle. | 12 |
| Figure 9. Evolution du nombre de localisations générées par étude en fonction de l'année du début de l'étude..... | 14 |
| Figure 10. Cycle de vie des données de bio-logging..... | 15 |
| Figure 11. Type de structure représentées dans l'enquête. | 19 |
| Figure 12. Proportions de différentes classes de quantités de balises déployées..... | 20 |
| Figure 13. Outils utilisés au cours des différentes étapes du cycle de vie des données de bio-logging | 21 |
| Figure 14. Sources de douleurs au cours de l'étape de traitement des données. | 22 |
| Figure 15. Caractéristiques et fonctionnalités que les personnes enquêtées souhaiteraient retrouver dans un outil..... | 23 |
| Figure 16. Evolution du nombre d'études (A), de taxons (B) et de géolocalisations (C) que l'on peut trouver sur la plateforme Movebank. | 25 |
| Figure 17. Version simplifiée du Modèle Conceptuel de Données de Movebank. | 26 |
| Figure 18. Architecture de l'ensemble applicatif FollowDem. | 32 |
| Figure 19. Architecture de l'ensemble applicatif idéal..... | 36 |
| Figure 20. Evaluation de la valeur des différents outils présentés précédemment selon le modèle de l'échelle de qualité des données ouvertes..... | 38 |
| Figure 21. Tableaux de sélection d'études Movebank (à gauche) et d'animaux d'une étude Movebank (à droite)..... | 42 |
| Figure 22. Onglet "Local study" de AMCA..... | 43 |
| Figure 23. Page d'accueil de AMCA. | 44 |
| Figure 24. Modèle de fichier VRT permettant d'extraire des données d'une base PostgreSQL..... | 44 |
| Tableau 1. Liste non exhaustive d'outils permettant le traitement, l'analyse et/ou la visualisation de données de bio-logging..... | 17 |
| Tableau 2. Personae établis grâce aux résultats de l'enquête..... | 24 |
| Tableau 3. Fonctionnalités à développer et leur priorisation (où 1 correspond au niveau de priorité le plus élevé)..... | 38 |
| Tableau 4. Principales caractéristiques des différentes routes de l'API de Movebank..... | 41 |

Liste des abréviations

ALA : Atlas of Living Australia

AMCA : Animal movement cleaner application

BAS : British Antarctic Survey

CLUF : Contrat de Licence Utilisateur Final

DwC : DarwinCore

GBIF : Global Biodiversity Information Facility

GLS : Global Location Sensor

GPS : Global Positioning System

GSM : Global System For Mobile Communications

IBLS : Société Internationale du Bio-Logging

MCD : Modèle Conceptuel de Données

MHz: Mega Hertz

MNHN : Muséum National d'Histoire Naturelle

NASA : National Aeronautics and Space Administration

NOAA : National Oceanic and Atmospheric Administration

NSF : National Science Foundation

OBIS : Ocean Biodiversity Information System

PNE : Parc National des Écrins

PO : Product Owner

SGBD : Système de Gestion de Bases de Données

SIG : Système d'Information Géographique

TDWG : Taxonomic Databases Working Groups

USFWS : United States Fish and Wildlife Services

VHF : Very High Frequency

Glossaire

Back-end : partie d'une application web exécutée côté serveur et donc non visible aux yeux de l'utilisateur final

Balise : objet attaché à un animal qui permet d'enregistrer des positions et potentiellement d'autres mesures en cas de présence de capteurs.

Bibliothèque logicielle : collection de routines réutilisables visant à faciliter le travail des développeurs

Bio-logging : utilisation d'appareils électroniques placés sur des êtres vivants afin d'enregistrer et/ou transmettre des données portant sur la physiologie, le comportement, le mouvement et/ou l'environnement des individus

Conteneur : enveloppe virtuelle qui permet de distribuer une application avec tous les éléments dont elle a besoin pour fonctionner

Docker : logiciel libre permettant de lancer des applications dans des conteneurs logiciels

Framework : ensemble cohérent de composants logiciels structurels guidant la conception d'une architecture logicielle

Front-end : partie d'une application web exécutée côté client (i.e. le navigateur) qui constitue donc ce que voit l'utilisateur final

JavaScript : langage de programmation de scripts principalement employé dans les pages web interactives et à ce titre est une partie essentielle des applications web

Python : langage de programmation interprété

SQL (Structured Query Language) : langage informatique servant à exploiter des bases de données relationnelles (i.e. rechercher, modifier, ajouter ou supprimer des données)

Introduction

Les mouvements des animaux sont essentiels à leur survie tant à l'échelle individuelle qu'à l'échelle de l'espèce. Ces déplacements leur permettent de rechercher de la nourriture, de se reproduire, de migrer, d'échapper aux menaces ou encore d'éviter la compétition. Les mouvements ont ainsi une influence importante sur la valeur sélective individuelle (Andreassen, 1998), le flux de gènes (Banks *et al.*, 2013) et la dynamique des populations (Gebauer *et al.*, 2013). Leur impact va même au-delà de la survie ou non d'une espèce puisqu'ils jouent un rôle majeur dans de très nombreuses interactions que l'on peut observer au sein d'un écosystème. En effet les déplacements des animaux permettent également la dispersion de graines, la propagation de maladie ou encore la prédation (Lundberg *et al.*, 2003). Tant de phénomènes qui régissent le fonctionnement des écosystèmes. Toutes ces raisons mettent en évidence l'intérêt que présente l'étude des mouvements des animaux dans le cadre de la conservation de la biodiversité. De plus, cet intérêt ne cesse de grandir avec l'évolution de la conjoncture démographique et climatique.

En effet, ces mouvements si importants pour les écosystèmes sont rendus de plus en plus compliqués par les changements considérables de paysages liés à l'activité humaine, et qui impliquent perte, fragmentation et dégradation d'habitats pour les animaux (Fahrig, 2007; Tucker *et al.*, 2018). Dans un rapport publié en 2015, les Nations Unies ont estimé que d'ici 2050 la population humaine aura augmenté d'environ 2,3 milliards de personnes (United Nations, 2015). Pour que la production agricole puisse suivre cette forte croissance démographique, des chercheurs ont estimé à 1 milliard d'hectares la surface supplémentaire qui devra être utilisée pour l'agriculture (Tilman *et al.*, 2011). A cela s'ajoute le développement de la surface urbanisée qui sera sans doute tout aussi conséquente. Cette problématique de perte d'habitat est également accentuée par les dérèglements climatiques observés, responsables de changements des aires de répartition de nombreuses espèces (Hickling *et al.*, 2006; Chen *et al.*, 2011; Lister *et al.*, 2015). En effet les variations de facteurs climatiques ont une influence sur la distribution des espèces et l'étendue de leur niche écologique à travers les seuils physiologiques de tolérance à ces facteurs. Or la capacité de ces espèces à s'adapter à ces changements est souvent limitée (Davis, 2001) et cela en partie à cause de la perte, de la fragmentation et de la dégradation de l'habitat. Nous pouvons donc nous attendre à un bouleversement majeur au sein de notre biosphère.

L'étude des mouvements des animaux vise donc à nous aider à réduire autant que possible l'ampleur de ce bouleversement. Malgré les inquiétudes exprimées par certains chercheurs quant à l'évolution de cette discipline (Cooke *et al.*, 2017), notamment vis-à-vis de l'utilisation des données produites, les programmes de conservation de la biodiversité se reposent de plus en plus et autant que possible sur ce type d'étude (Fraser *et al.*, 2018). Ce phénomène peut s'expliquer notamment par les importants progrès technologiques réalisés au cours des dernières décennies. L'adoption massive d'appareils électroniques comme les smartphones et les tablettes par les consommateurs ont poussé le développement de nombreux capteurs miniatures et peu coûteux, nous permettant ainsi d'entrer dans ce que certains chercheurs considèrent comme l'âge d'or du bio-logging (Wilmers *et al.*, 2015; Kays *et al.*, 2015). Le bio-logging désigne l'utilisation d'appareils électroniques placés sur des êtres vivants afin d'enregistrer et/ou transmettre des données portant sur la physiologie, le comportement, le mouvement et/ou l'environnement des individus (Rutz and Hays, 2009). En plus des avancées matérielles viennent également s'ajouter des avancées dans le domaine de l'analyse permettant ainsi une meilleure valorisation des données récoltées (Jacoby *et al.*, 2016). Nous avons donc plus de données, qui sont de meilleure qualité et qui permettent de couvrir un plus grand spectre de problématiques. Tout cela est ainsi à l'origine d'une forte

augmentation du nombre d'études reposant sur le bio-logging ce qui entraîne également un développement du marché des outils et des services autour de ce dernier. Ce marché connaissant une croissance rapide, il devient facile de se perdre dans la multitude d'outils qui émergent, ces derniers pouvant être incomplets et redondants les uns par rapport aux autres, ce qui crée de nouveaux problèmes de standardisation et d'interopérabilité.

Ainsi, comment pourrait-on faire pour parvenir à un environnement applicatif complet qui réponde aux différentes attentes et besoins des acteurs du monde du bio-logging tout en se reposant sur la multitude d'outils déjà existants ?

Pour répondre à cette problématique, nous présenterons d'abord le contexte du stage au cours duquel s'est effectué le travail relaté par ce mémoire. Nous plongerons ensuite dans l'histoire du bio-logging afin de voir comment sont arrivés les défis auxquels ce milieu doit désormais faire face et les propositions de réponses qui ont été apportées. Nous évoquerons alors les différentes étapes de conception d'une solution réalisées au cours de ce stage puis nous présenterons les parties qui ont pu être développées.

I – Un stage réalisé au sein d’une entreprise aux valeurs fortes

A) Présentation de Natural Solutions

1. Une entreprise mettant l’innovation au service de la biodiversité

Natural Solutions est une agence digitale basée à Marseille spécialisée dans la conception et le développement d’outils numériques depuis 2008. L’entreprise s’intéresse principalement aux thématiques de collecte, de gestion, de stockage et d’exploitation des données environnementales dans le cadre de problématiques liées à la biodiversité qui est au cœur de son engagement. Ainsi, Natural Solutions édite des logiciels sur mesure, conçoit des applications web et mobiles et effectue également du conseil ainsi que de la recherche et développement. Sa double compétence dans les domaines de l’informatique et de l’environnement lui permet de dialoguer aisément avec une grande diversité d’acteurs du milieu de l’écologie. L’équipe de Natural Solutions, divisée en différents pôles (Figure 1), présente des profils variés mélangeant compétences techniques et organisationnelles. Les projets dans lesquels Natural Solutions s’engage peuvent alors avoir différentes origines : réponse à des appels d’offres, projets imaginés en interne ensuite commercialisés ou encore produits réalisés pour le client majeur qu’est Reneco – Reneco est une entreprise spécialisée dans l’élevage conservatoire de l’outarde Houbara (*Chlamydotis undulata*) qui utilise de nombreuses applications réalisées par Natural Solutions –

Les deux premiers types de projets évoqués amènent Natural Solutions à collaborer avec de nombreux partenaires qu’ils soient des organismes publics (comme des parcs nationaux, des associations de conservation de la biodiversité, des municipalités etc...) ou privés. Des exemples de solutions développées par l’entreprise peuvent être consultés sur son site web (Natural Solutions, 2020). Nous pouvons notamment citer EcoRelevé qui est un outil de saisie, gestion et partage de données de terrain intégrant un module de cartographie.

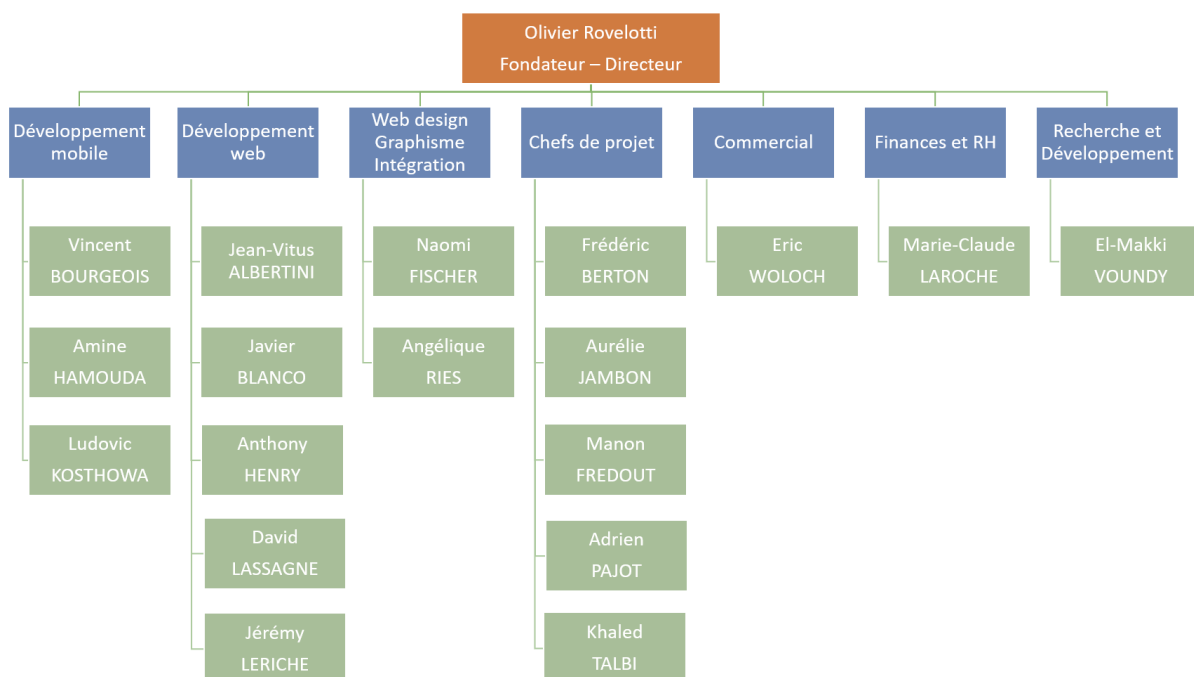


Figure 1. Organigramme de Natural Solutions.

L'écologie étant au cœur des raisons d'être de Natural Solutions, l'entreprise est très active dans ce milieu notamment en organisant et participant à des événements comme des colloques ou des hackathons.

2. Fonctionnement interne de l'entreprise

Natural Solutions n'est pas seulement innovante à travers les solutions qu'elle développe puisque c'est également une entreprise qui n'hésite pas à essayer de nouveaux modèles organisationnels. En effet, Natural Solutions fonctionne en suivant les principes des méthodes agiles, devenues très populaires au cours des dernières années notamment dans les entreprises qui font du développement informatique, et plus particulièrement du cadre de travail Scrum. Dans cette organisation qui repose sur des cycles de développement itératifs, incrémentaux et adaptatifs, les projets sont découpés en sprints qui, à Natural Solutions, durent deux semaines (Figure 2). Chaque sprint commence par l'étape de planification qui consiste à déterminer les tâches, dont la durée et/ou la difficulté de réalisation a/ont été estimée(s), à réaliser au cours du sprint. Ces tâches appartiennent à des « user stories » (récits utilisateur) qui expriment, généralement sous la forme d'une phrase simple respectant le modèle « En tant que <qui>, je veux <quoi> afin de <pourquoi> », un besoin d'un utilisateur. Chez Natural Solutions, les user stories ainsi que les tâches qu'elles contiennent sont gérées grâce à l'outil Pivotal Tracker.

Les sprints sont rythmés quotidiennement par une réunion appelée « daily scrum » (mêlée quotidienne) qui permet aux membres de l'équipe d'informer le reste du groupe sur le travail réalisé la veille, les difficultés rencontrées et le travail du jour. Une réunion de mi-sprint a lieu le lundi de la 2^{ème} semaine d'un sprint à la place du daily scrum afin de faire un point sur l'état d'avancement des objectifs fixés en début de sprint. Les sprints s'achèvent sur une rétrospective qui permet de faire des démonstrations de ce qui a été réalisé pendant le sprint, de déterminer si les objectifs du sprint ont été atteints et de recueillir des retours à propos du déroulement du sprint. L'application de cette méthodologie repose sur différents rôles dont celui de « scrum master » qui est le garant du respect du cadre Scrum ainsi que la personne qui anime les différentes réunions évoquées ci-dessus.

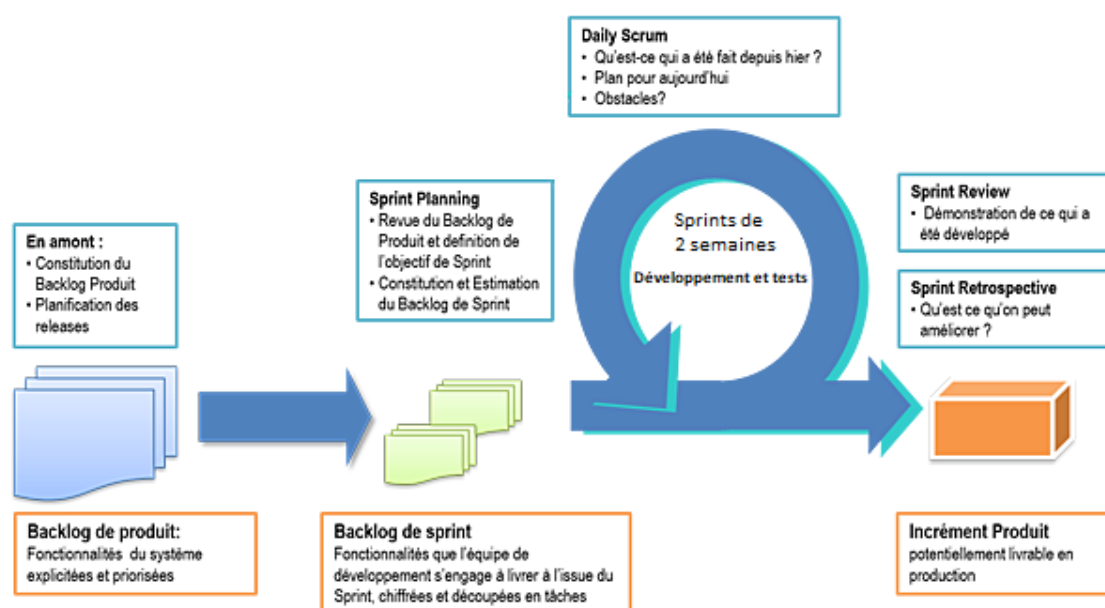


Figure 2. Principes de fonctionnement de la méthodologie Scrum. (Subra, 2019)

Chez Natural Solutions, le scrum master s'appuie sur l'outil Scrum Manager, développé en interne, qui lui permet de suivre l'avancement des sprints à travers différentes métriques et représentations comme le « burndown chart » (graphique d'avancement). Le « product owner » (PO), autre rôle fondamentale du cadre Scrum, est la personne qui oriente le projet, détermine les priorités et est responsable de l'adéquation entre les besoins des utilisateurs et les fonctionnalités développées. Il est à l'interface entre le client, le scrum master et son équipe de développeurs dont il connaît la force de travail et les compétences.

Les méthodes agiles ne sont pas les seules pratiques organisationnelles adoptées par Natural Solutions. Depuis début 2020, l'entreprise a intégré les concepts de l'holocratie à son fonctionnement. L'holocratie est un système de gouvernance qui consiste en une organisation horizontale dont l'objectif est de rendre les salariés plus autonomes et plus impliqués et intégrés dans les processus décisionnaires, visant à effacer la notion de hiérarchie. L'holocratie permet de remplacer les systèmes traditionnels en pyramide puisque l'organisation y est découpée en cercles (Figure 3). Chaque cercle possède une raison d'être, une portée ou champ d'application, qui définit le spectre des thématiques qui le concernent, ainsi que des rôles. Chaque rôle porte également une raison d'être, qui doit donc contribuer à la raison d'être du cercle auquel il appartient qui elle-même doit contribuer à la raison d'être globale de l'entreprise.

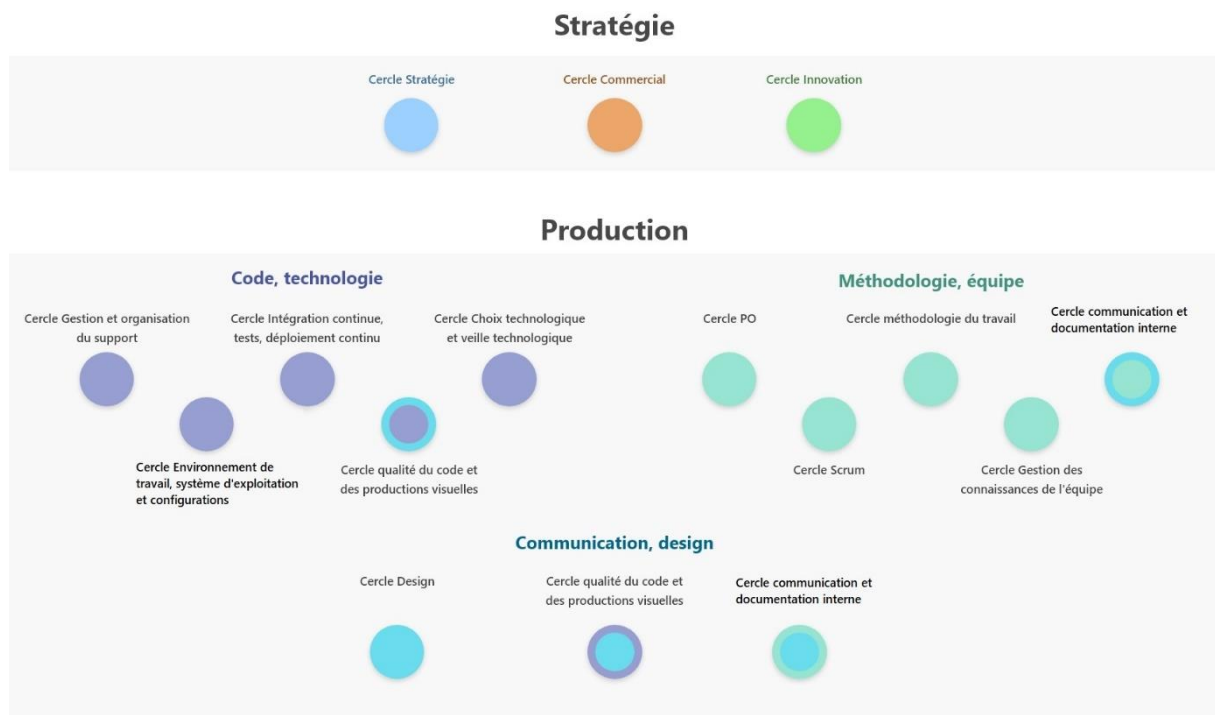


Figure 3. Les différents cercles mis en place au sein de Natural Solutions.

Ainsi, pour tout ce qui concerne le champ d'application d'un cercle, la décision revient et à ses membres qui votent au cours d'une réunion après un débat voire une période de réflexion et/ou de test. Chez Natural Solutions, la majorité des cercles est ouverte à tous les membres de l'équipe à l'exception de quelques-uns qui traitent de problématiques très spécifiques. Les réunions de cercles, qui sont organisées à une fréquence variable, sont évoquées lors des rétrospectives de fin de sprint afin de discuter d'éventuelles décisions qui concernent l'ensemble de l'entreprise. L'intégration de cette philosophie au fonctionnement de l'entreprise en est encore à sa phase expérimentale.

B) Déroulement et étapes clés du stage

Ce stage a débuté par une période d'intégration à l'entreprise (Figure 4) qui a notamment permis de découvrir sa façon de s'appropriier les méthodes agiles et l'holocratie. Cette phase a également servi à prendre connaissance des outils utilisés au sein de Natural Solutions et des produits développés par l'entreprise, qui ont fait l'objet de présentations par leurs référents techniques.

Ensuite, une phase de formation m'a permis de découvrir ou de me remettre à niveau sur certaines technologies que j'ai été amené à utiliser au cours de mon stage, comme le langage Python, le framework VueJS ou encore le Système de Gestion de Base de Données (SGBD) Microsoft SQL Server.

En parallèle, j'ai pris le temps de prendre en main certains outils, notamment Animal Movement Cleaner Application (AMCA), qui ont constitué le support du travail effectué au cours de ce stage. AMCA est l'application web développée par Aurélie Jambon au cours de son stage réalisé un an auparavant et dont ce stage constitue la suite (Jambon, 2019). Nous développerons ce sujet dans les prochaines parties du mémoire.



Figure 4. Planning du stage.

En raison des restrictions sanitaires liées à l'épidémie de COVID-19 touchant la France, l'ensemble du stage s'est déroulé en télétravail. Les activités de Natural Solutions n'ont pas été trop impactées au cours de cette période et le passage généralisé au télétravail s'est plutôt bien passé puisque l'entreprise considère actuellement la possibilité de conserver ce mode de fonctionnement. Les premiers mois ont tout de même été nécessaires pour s'adapter à cette nouvelle organisation inédite (pour moi et pour l'entreprise) qui a impacté non seulement le fonctionnement interne de Natural Solutions mais souvent également celui des différents partenaires et clients et donc les échanges entre les deux. De plus, cette période est intervenue peu de temps après le début de la mise en place de l'holocratie au sein de l'entreprise ce qui a pu générer quelques dysfonctionnements mais surtout une forte multiplication du nombre de réunions effectuées.

Lorsque les mesures de restrictions se sont assouplies autorisant un retour contrôlé aux bureaux, je me suis rendu occasionnellement aux locaux de l'entreprise pour travailler sur certaines tâches avec d'autres membres de l'équipe de Natural Solutions.

II – Le monde du bio-logging, son histoire, ses promesses et ses difficultés

Le bio-logging est un terme apparu au début des années 2000 qui désigne l'utilisation d'appareils électroniques placés sur des êtres vivants afin d'enregistrer et/ou transmettre des données portant sur la physiologie, le comportement, le mouvement et/ou l'environnement des individus (Boyd *et al.*, 2004). Dans cette partie nous verrons comment ce domaine a émergé et évolué afin d'expliquer et de mettre en évidence les problématiques rencontrées actuellement ainsi que les différentes façons d'y répondre. Les principes de fonctionnement des différentes technologies abordés ci-dessous ne seront ici que très brièvement décrits. Des explications plus détaillées peuvent être trouvées dans (Jambon, 2019) dont ce mémoire constitue une suite.

A) Genèse et démocratisation du bio-logging

1. Les premières études de bio-logging

En 1865, le médecin et physiologiste français Claude Bernard exprime l'idée suivante : « Dans l'investigation scientifique, les moindres procédés sont de la plus haute importance. Le choix heureux d'un animal, un instrument construit d'une certaine façon, l'emploi d'un réactif au lieu d'un autre, suffisent souvent pour résoudre les questions générales les plus élevées. » (Bernard, 1865). Une partie de cette idée est reprise en 1929 par le physiologiste danois August Krogh qui énonce le principe, plus tard baptisé principe de Krogh par Sir Hans Adolf Krebs, stipulant que « pour un grand nombre de problèmes, il y aura un animal de choix, ou quelques animaux de ce type, sur lequel ils seront plus facilement étudiés. » (Krogh, 1929). Ces mots mettent en évidence un obstacle souvent rencontré aujourd'hui encore en biologie et dans tous les domaines qu'elle regroupe : les sujets d'études de ces disciplines en eux-mêmes, à travers leurs caractéristiques intrinsèques, peuvent constituer une limite considérable à notre capacité à les étudier.

En effet, pour certaines espèces il était, jusqu'à très récemment, particulièrement complexe voire impossible d'obtenir des mesures directes liées aux comportements étudiés. Par exemple, la profondeur à laquelle plongeaient certains oiseaux était déterminée approximativement en capturant des individus via des filets placés à certaines profondeurs (Conroy and Twelves, 1972) ou encore en mesurant la profondeur du milieu tout en faisant l'hypothèse que ces oiseaux se nourrissaient sur le fond (Stonehouse, 1967). Cela s'explique également par le fait que les techniques à disposition des scientifiques ont longtemps présenté d'importantes limites. Nous pouvons notamment citer la taille et le poids des appareils de mesures qui non seulement empêchaient leur fixation sur des animaux trop petits et/ou trop légers mais qui en plus étaient suspectés d'altérer le comportement des individus étudiés ou de provoquer certains phénomènes comme des pertes de poids. Il est ainsi généralement considéré qu'un appareil ne doit pas peser plus de 10 voire 5 % de la masse corporelle de l'individu.

L'incapacité de transmettre les informations à distance représentait une autre limite majeure puisque cela obligeait les chercheurs à re-capturer les animaux équipés sans connaître leur position, contrainte particulièrement lourde dans le cadre d'une étude d'une espèce migratoire, nocturne ou difficilement repérable et/ou approchable de manière générale. L'arrivée de la radio-télémetrie VHF (Very High Frequency) et son gain de popularité dans le monde de l'écologie au cours des années 1960 a permis d'apporter une réponse à cette limite.

Cette technique consiste à transmettre des données grâce à un émetteur, placé donc sur un animal dans le cas des études auxquelles nous nous intéressons ici, alimenté par une pile ou une batterie et qui envoie un signal sous forme d'onde radioélectrique capté par un récepteur (une antenne radio réglée sur la fréquence de l'émetteur). Sa popularisation, tout de même limitée par le coût alors élevé de la technologie, au sein de la communauté scientifique s'explique notamment par la réduction de la taille des transistors qui ont commencé à être produits en masse en 1957 par Sony. La radio-télémetrie VHF, via différentes méthodes (e.g. la triangulation), peut notamment être utilisée afin d'obtenir des informations de localisation. L'utilisation de cette technique a ainsi permis d'élargir le spectre des espèces et des problématiques que nous étions en mesure d'étudier. En effet, cela nous a par exemple rendu capables de suivre les déplacements d'animaux évoluant dans des milieux comme les forêts denses, les déserts ou qui parcourent de longues distances. Cependant, l'utilisation de cette technique s'est surtout répandue dans le cadre de l'étude d'espèces terrestres principalement à cause de la mauvaise transmission des ondes radios sous la surface de l'eau qui représente donc un obstacle majeur dans le cas d'études portant sur des espèces aquatiques.

Revenons à l'exemple de l'étude des profondeurs de plongée. Les premières mesures directes et continues n'ont été obtenues qu'en 1963 grâce à un enregistreur de temps-profondeur, qui pesait environ 1,5 kg et dont la durée d'enregistrement était limitée à moins d'une heure, spécialement conçu pour une étude menée en Antarctique sur le phoque de Wedell (*Leptonychotes weddellii*) (Kooyman, 1965). Cette espèce a été spécifiquement sélectionnée pour le poids moyen important de ses représentants ainsi que pour leur comportement. En effet, ces phoques ont pour habitude de remonter à la surface de la banquise dans certaines zones spécifiques dans le but de respirer. Cela offre l'opportunité aux chercheurs de pouvoir attacher des appareils de mesure à des individus qui refont plus tard surface, avec une probabilité élevée, aux mêmes endroits. Les limites citées précédemment étaient ainsi contournées. Le type d'appareil utilisé dans cette étude a d'abord connu un développement relativement lent puisqu'il a fallu attendre 1975 pour parvenir à la possibilité d'enregistrer des profondeurs de plongée sur une durée de 14 jours (Kooyman *et al.*, 1976). Le rythme des progrès effectués a ensuite accéléré au cours des années 1980 pour atteindre une durée d'enregistrement de 130 jours en 1988 grâce à un appareil conçu pour une étude portant sur la tortue de mer Caouanne (*Caretta caretta*) et qui utilisait un papier sensible à la pression (Naito *et al.*, 1990). Cet enregistreur mesurait également la température de l'eau, la température de l'estomac et la distance de nage en plus de la profondeur de plongée et cela pour un poids total de 1,3 kg. Au même moment, les premières mesures directes et continues de profondeurs de plongée pour des oiseaux plongeurs volants ont été obtenues au cours d'une étude sur le Cormoran impérial (*Leucocarbo atriceps* / *Phalacrocorax atriceps*) (Croxall *et al.*, 1991). Ces travaux ont été permis par la réduction de la taille et du poids des appareils de mesure, ici des manomètres à soufflet, qui pouvaient alors être fixés avec de la résine époxy sur des plumes du centre du dos des individus. Cependant, les études citées ci-dessus constituaient alors des exceptions (Figure 5). En effet, une problématique telle que l'étude de la plongée chez les animaux était encore techniquement compliquée et les travaux se penchant dessus reposaient sur des appareils spécialement conçus pour y répondre. Il a donc fallu attendre de nouvelles avancées technologiques pour que le bio-logging puisse connaître une croissance plus importante.

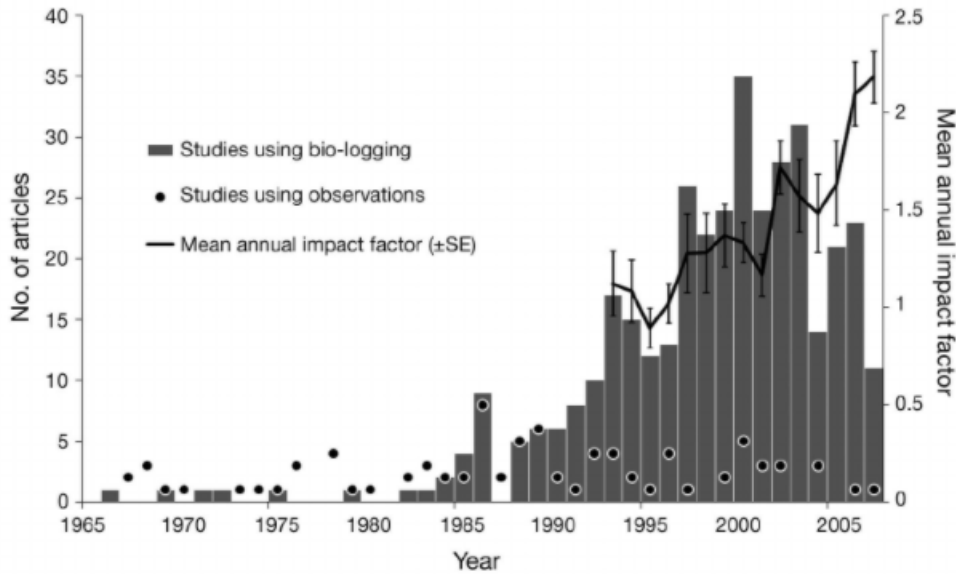


Figure 5. Nombre d'articles annuels portant sur la thématique de l'activité de plongée reposant sur le bio-logging (barres) ou sur de simples observations visuelles (points) en fonction du temps. (Ropert-Coudert et al, 2009)

2. Le début de l'utilisation des satellites

Les années 1970 ont vu les premières utilisations des satellites dans le cadre de suivi d'animaux sauvages avec notamment au départ les satellites météorologiques du programme Nimbus. En 1970, une femelle wapiti (*Cervus canadensis*) adulte du National Elk Refuge, situé dans le Wyoming aux Etats-Unis, a été équipée d'un collier de 11,3 kg (Craighead *et al.*, 1972). Ce collier transmettait la localisation de l'individu aux satellites Nimbus-3 et Nimbus-4 à raison de 2 fois par jour et sur une durée de 29 jours, la connexion ayant ensuite été perdue lorsque l'antenne s'est retrouvée dirigée vers le sol à la suite d'une rotation du collier. En 1976, l'entreprise Handar, Inc. a développé pour l'U.S. Fish and Wildlife Service (USFWS) un harnais de 5,6 kg équipé de transmetteurs satellites afin de suivre des ours polaires de plus de 180 kg en Alaska (Kolz *et al.*, 1980). Cet équipement, utilisant le satellite Nimbus-6, a affiché des erreurs de localisation allant de 1 à 14 km lors d'essais réalisés au préalable de l'étude. Ce satellite, lancé en 1975 et dont la mission s'est terminée en 1983, a fortement été mis à contribution dans le cadre d'études similaires que ce soit sur des ours polaires (Taylor, 1982; Schweinburg and Lee, 1982; Larsen *et al.*, 1983) ou sur des espèces marines comme le requin pèlerin (*Cetorhinus maximus*) (Priode, 1982), la tortue caouanne (*Caretta caretta*) (Stoneburner, 1982; Timko and Kolz, 1982), ou encore le dauphin tacheté pantropical (*Stenella attenuata*) (Woods and Kemmerer, 1982). Pour cette dernière, des essais réalisés au préalable ont établi une erreur de localisation moyenne de 23 km chez les appareils utilisés (Jennings and Gandy, 1980). Cette technologie apportait donc une réponse aux problématiques rencontrées dans l'utilisation de la radio télémétrie VHF de transmission des données dans le cadre d'études d'espèces marines, et de portée de signal dans le cadre d'études d'espèces se déplaçant sur de grandes distances. Cependant, les contraintes du poids et de la taille était toujours présente et la précision de ces systèmes devait être améliorée.

A la même période, le système Argos naît d'une collaboration entre le Centre National d'Etudes Spatiales (CNES), l'agence américaine d'étude de l'atmosphère et de l'océan (NOAA, National Oceanic and Atmospheric Administration) et l'agence spatiale américaine (NASA, National Aeronautics and Space Administration). Dans le cas de cette technologie, les balises émettent un signal à une fréquence de 401,650 MHz devant être stable car le système repose sur l'utilisation de l'effet Doppler. En effet, la localisation est calculée par des centres de traitement à partir de la date de réception du signal ainsi que de la différence entre la fréquence du signal reçu par les satellites Argos et celle du signal émis. Le système Argos a vu le jour en 1978, le principal objectif de ce projet collaboratif étant d'obtenir des données environnementales à intervalle journalier et cela n'importe où sur Terre. Cependant, les scientifiques ont rapidement compris l'intérêt que ce système présentait pour le suivi d'animaux sauvages. Entre avril 1985 et octobre 1986, 1 265 localisations ont été obtenues et analysées, l'erreur moyenne était de 829 m et 90 % des localisations estimées étaient à moins de 1 700 m de la localisation réelle (Figure 6) (Fancy *et al.*, 1988).

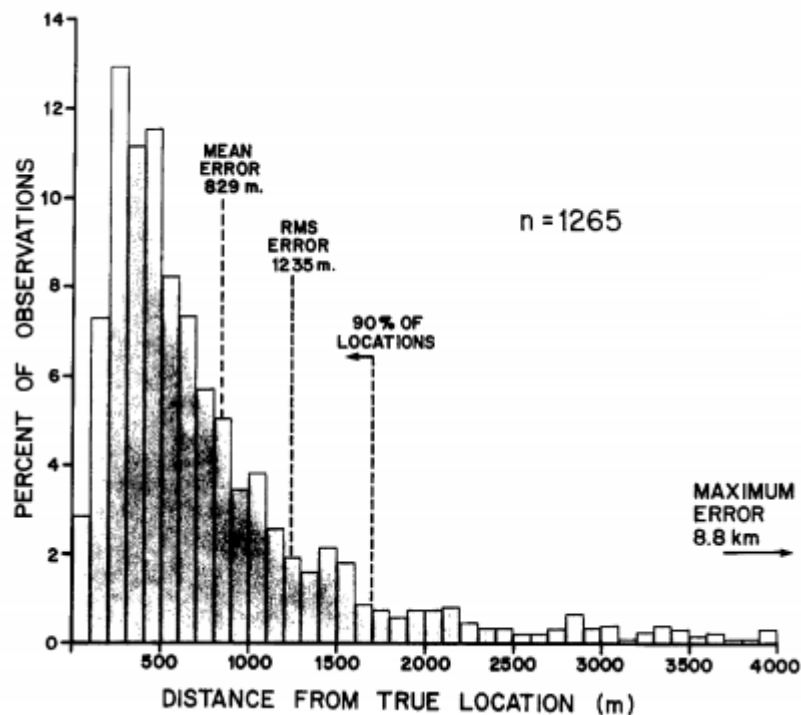


Figure 6. Exactitude des émetteurs de seconde génération de Telonics. (Fancy *et al.*, 1988)

Au cours des années 1980, de nombreuses études utilisant la technologie Argos ont été menées et en particulier sur des mammifères marins de grande taille comme le lamantin des Caraïbes (*Trichetus manatus*) (Mate *et al.*, 1986) ou sur des ruminants comme les rennes (*Rangifer tarandus granti*) (Curatolo, 1986). Les oiseaux étaient eux encore presque systématiquement exclus de ce type d'études à cause de leur poids souvent insuffisant pour porter l'appareillage nécessaire et cela malgré la réduction progressive de la taille des transmetteurs. En effet, c'est bien souvent la batterie qui constituait le facteur limitant puisque les progrès réalisés sur ces composants étaient plus lents et l'énergie nécessaire à la transmission de données sur une période de temps prolongée imposait l'utilisation de batteries d'une certaine taille. De plus, le système Argos requiert, pour son

bon fonctionnement, une grande stabilité de la fréquence des signaux émis. Ainsi en 1982 un groupe de scientifiques du Laboratoire de Physique Appliquée de l'université Johns Hopkins a étudié pour l'USFWS la faisabilité d'un transmetteur suffisamment léger pour être porté par un oiseau afin de pouvoir suivre des espèces migratoires (Strikwerda *et al.*, 1985). Un prototype combinant batteries rechargeables et cellules photovoltaïques a alors été développé pendant un an et demi à partir de janvier 1983 pour aboutir à un appareil d'environ 160 g, lui permettant ainsi d'être placé sur un pygargue à tête blanche (*Haliaeetus leucocephalus*), des pétrels géants (*Macronectes giganteus*), des cygnes siffleurs (*Cygnus columbianus*) et des cygnes trompettes (*Cygnus buccinator*) (Strikwerda *et al.*, 1986). C'est à partir de 1991 que l'on passe sous la barre des 100 g pour un appareil de ce type avec le développement d'un transmetteur de 95 g par l'entreprise Microwave Telemetry, Inc. qui permet ainsi d'étudier des individus ayant une masse corporelle moins importante.

3. L'avènement du bio-logging

La miniaturisation des composants électroniques s'est poursuivie dans les années 1990 avec le développement continu des microprocesseurs qui devenaient toujours plus puissants en même temps que plus petits. De nouvelles technologies apparaissent alors et d'autres deviennent accessibles au grand public. Le coût élevé des appareils s'appuyant sur les satellites et des frais liés à la transmission des données a poussé un groupe de chercheurs du British Antarctic Survey (BAS) à concevoir un nouveau type d'instruments dénommé Global Location Sensor (GLS) (Wilson *et al.*, 1992). Ces balises enregistrent l'intensité lumineuse et l'angle du soleil pour ensuite calculer la durée du jour qui permet de déterminer la latitude. La longitude est, elle, déterminée grâce à l'heure du zénith. Cet appareil pesant 113 g était alors capable de mesurer l'intensité lumineuse une fois toutes les 20 minutes et cela sur une durée de 220 jours. En 1989 un instrument similaire avait déjà été développé par un autre groupe de scientifiques afin d'étudier l'éléphant de mer du nord (*Mirounga angustirostris*) (DeLong *et al.*, 1992). Ces appareils avaient donc l'avantage du prix et du poids sur la plupart des systèmes reposant sur les satellites qui restaient difficilement accessibles. Cependant, deux inconvénients non négligeables viennent contrebalancer ces avantages. Le premier est que la précision des mesures est fortement dépendante du niveau de lumière parvenant à l'instrument ce qui rend son utilisation compliquée dans des milieux ombragés (sous des canopées denses ou lorsque la couverture nuageuse est importante). Le second est que ces appareils doivent être récupérés afin d'accéder aux données, ce qui implique donc de re-capturer les animaux équipés, contrainte balayée dans le cas de l'utilisation de satellites qui restait l'option préférentielle.

Les années 1990 marquent donc surtout l'accès grand public à des technologies d'importance majeure, certaines jusqu'alors réservées à des usages militaires et/ou scientifiques, comme le Global Positioning System (GPS), Internet et le Global System for Mobile Communications (GSM). Le GPS, qui est un système de positionnement par satellites appartenant au gouvernement des Etats-Unis, permet à tout utilisateur disposant d'un récepteur GPS de se géolocaliser. Les coordonnées, exprimées en se référant au système géodésique WGS 84, sont calculées par trilatération des signaux électromagnétiques synchronisés émis par les satellites. Ce système permettant d'obtenir une géolocalisation plus précise que le système Argos, des récepteurs GPS ont commencé à être utilisés sur des balises Argos. Dans le cas où l'on souhaite obtenir uniquement des données de géolocalisation sans données environnementales, les récepteurs GPS peuvent être utilisés avec un appareil communiquant sur un réseau GSM pour transmettre les coordonnées calculées. La communauté scientifique a alors disposé d'un éventail assez large de

technologies offrant ainsi la possibilité de répondre à un plus grand nombre de besoins et de problématiques, que ce soit d'un point de vue écologique ou d'un point de vue technique. Avec l'adoption massive d'appareils électroniques comme les téléphones portables ou les tablettes par les consommateurs, la miniaturisation et la production des composants se sont poursuivies à un rythme effréné. Il a fallu attendre 2005 pour passer sous la barre des 10 g (Figure 7) avec une balise de l'entreprise Microwave Telemetry, dont l'évolution des produits est représentée sur l'annexe 1.

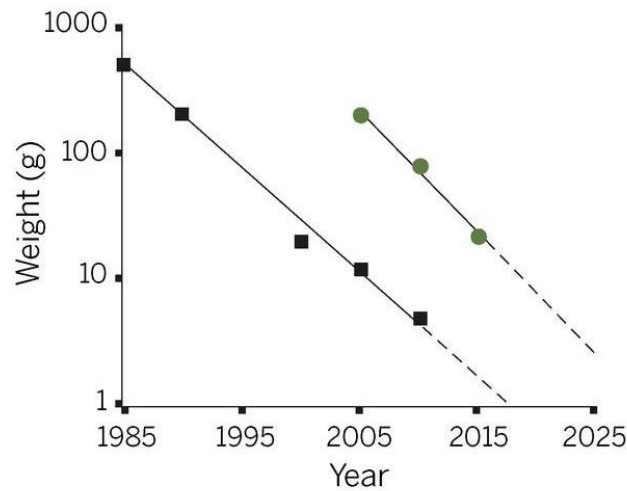


Figure 7. Evolution du poids en g de la plus petite balise de bio-logging en fonction du temps. (Kays et al, 2015)

Ces avancées étaient primordiales pour nous aider à mieux connaître notre environnement puisque la valeur médiane de la masse corporelle est d'environ 25 g pour les mammifères (Gardezi and Da Silva, 1999) et entre 30 et 40g pour les oiseaux (Blackburn and Gaston, 1994). La répartition du nombre d'espèces de ces classes en fonction de la masse corporelle moyenne peut être retrouvée sur la Figure 8.

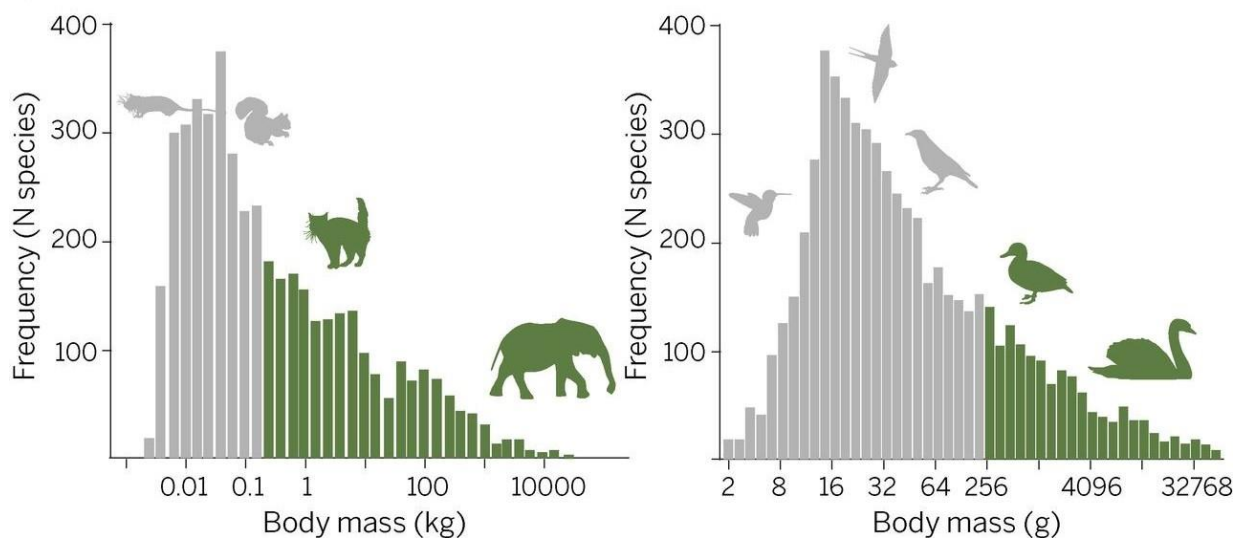


Figure 8. Nombre d'espèces de mammifères (gauche) et d'oiseaux (droite) en fonction de leur masse corporelle. Les barres vertes représentent les espèces étudiées avec une balise de 15 g. (Kays et al, 2015)

En 2014, des chercheurs ont pu suivre les déplacements et ainsi étudier le comportement de chauves-souris dont le poids moyen était inférieur à 30 g (Castle *et al.*, 2015). Au-delà de devenir plus petits, les capteurs sont également devenus plus précis, les batteries plus performantes et les émetteurs plus fiables augmentant alors la qualité des données récoltées. De plus, leur production de masse a aussi entraîné une réduction importante des coûts, tant au niveau de la récolte des données qu'au niveau de leur transmission, ce qui n'est pas négligeable dans un milieu comme l'écologie où les budgets sont généralement réduits. Cela a permis de réaliser des travaux avec un nombre d'individus suivis plus important puisque jusque-là les études étaient généralement menées sur moins d'une dizaine d'individus à cause du prix des différentes technologies. Les déploiements de masse de balises et autres capteurs sont devenus communs et la quantité de données générée par le bio-logging a donc explosé. L'écologie et le bio-logging sont ainsi rentrés dans l'ère du « big data » et les limitations techniques de l'époque ont laissé leur place à de nouvelles problématiques.

B) L'absence de standardisation dans le monde du bio-logging

Parmi les nouvelles problématiques amenées par l'entrée du bio-logging dans l'ère du « big data », certaines sont à l'origine de préoccupations d'ordre éthique et/ou environnementale. En effet, la pose d'appareils de mesures et de suivi sur des animaux a toujours été suspectée de modifier le comportement des individus mais il n'avait pas été anticipé que cela puisse aller au-delà de l'échelle individuelle. Une étude a ainsi montré que des poissons équipés de balises acoustique étaient spécifiquement ciblés par des prédateurs, ce qui est particulièrement gênant dans le cadre d'une étude portant sur la survivabilité (Cunningham *et al.*, 2014). Une autre préoccupation grandissante dans le monde du bio-logging concerne l'utilisation des données. En 2014, le gouvernement de l'État de l'Australie-Occidentale a lancé une politique permettant de tuer les grands requins repérés à proximité de plages fréquentées par des baigneurs. Au cours de la même année, des données de suivi d'un programme d'étude du requin blanc ont été utilisées par le gouvernement pour localiser et tuer un requin qui n'avait pourtant pas été identifié comme une menace potentielle (Meeuwig *et al.*, 2015). Le recours désormais presque automatique à des méthodes de bio-logging, induit par la grande accessibilité aux différentes technologies, est ainsi remis en question et il semble nécessaire que ces préoccupations soient traitées par la communauté scientifique et les organismes publics. Dans ce travail, nous nous concentrerons cependant sur la problématique de la valorisation des données issues du bio-logging.

1. Une problématique causée par le développement rapide et non structuré du milieu

Comme expliqué précédemment, les différentes technologies pouvant être utilisées dans le cadre du bio-logging sont actuellement plus accessibles que jamais, en particulier sur le plan économique. Le nombre d'études de bio-logging a donc logiquement explosé tout comme la quantité de données récoltées par étude (Figure 9).

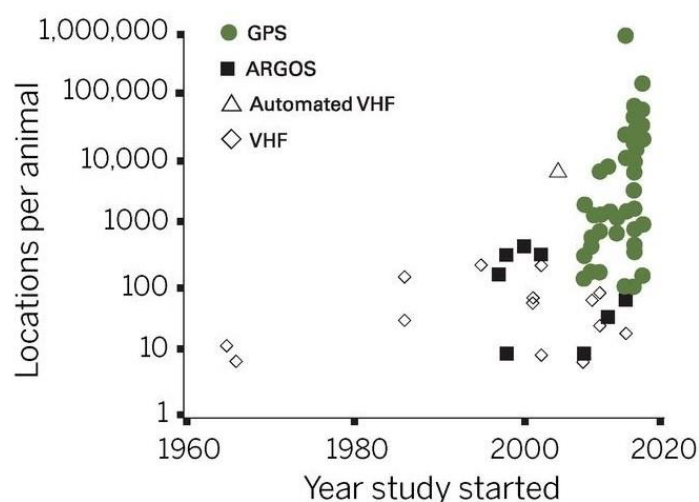


Figure 9. Evolution du nombre de localisations générées par étude en fonction de l'année du début de l'étude. (Kays et al, 2015)

Le bio-logging est donc lui aussi entré dans l'ère du « big data » et permet désormais d'apporter des réponses à des questions d'ordres assez variés, ce qui donne alors une valeur non négligeable aux données générées. Cependant, elles sont souvent sous-valorisées à cause du manque de standardisation induit par la difficulté de définir et faire appliquer un standard international.

En effet, l'augmentation du nombre d'études de bio-logging et d'individus suivis a créé une croissance de la demande en balises entraînant une multiplication de l'offre. Une liste non exhaustive de fabricants peut être trouvée en annexe 2. Nous avons donc des données qui peuvent être de natures différentes (géolocalisations, données environnementales, physiologiques...), issues de diverses technologies de collecte et de transmission elles-mêmes proposées par une grande variété de constructeurs. Le tout s'inscrivant dans un contexte de coopération scientifique internationale, le besoin de standardisation se fait de plus en plus sentir et a notamment poussé un groupe de chercheurs à publier un plaidoyer en faveur de la mise en place de standards pour les données provenant d'appareils électroniques portés par des animaux (Campbell *et al.*, 2016). Le manque de collaboration entre les différents acteurs du milieu du bio-logging y est particulièrement mis en avant. Du côté des constructeurs, on peut retrouver d'importantes divergences d'un fabricant à un autre et cela à différents niveaux. Le premier niveau d'incompatibilité est d'ordre syntaxique et concerne donc le format des données qui peuvent être transmises sous la forme de fichiers CSV, mails, fichiers XML, fichiers JSON etc... De plus, selon les fabricants les données peuvent être soit envoyées automatiquement dans l'un des formats cités précédemment soit mises à disposition sur la plateforme du constructeur. Le deuxième niveau d'incompatibilité est lui d'ordre structurel avec par exemple des différences au niveau des types des données. Enfin, nous avons un troisième niveau qui est d'ordre sémantique. En effet, d'un constructeur à un autre nous pouvons constater des divergences dans le nom des attributs, leurs définitions ou encore leurs unités et cela pour des données de même nature. Par exemple, on constate différentes façons d'indiquer le temps et l'heure ou encore le fuseau horaire qui n'est pas toujours renseigné alors même qu'il peut parfois être modifié au cours du déploiement d'une même balise. Tous ces éléments peuvent être changés par les constructeurs et les modifications ne sont pas toujours indiquées dans les métadonnées, dont la qualité est très variable, ou dans la documentation de référence. Ce niveau d'incompatibilité est peut-être le plus problématique étant donné qu'il ne concerne pas uniquement le monde du bio-logging puisqu'il comprend également les standards de description de données écologiques et géospatiales.

2. Les premières propositions de réponse à ce problème complexe

Pour s'attaquer à cette problématique, la Société Internationale du Bio-Logging (IBLS) a constitué un groupe de travail et organisé un atelier qui s'est déroulé pendant le 6^{ème} Colloque International du Bio-Logging ayant eu lieu à Constance en 2017. L'IBLS est un organisme composé de chercheurs, fondé en 2016 suite au 5^{ème} Colloque International du Bio-Logging s'étant déroulé à Strasbourg en 2014. Le principal objectif de l'IBLS est d'encourager, promouvoir et supporter le bio-logging. Ainsi, l'atelier de travail organisé a réuni différents acteurs du milieu du bio-logging dont des fabricants de balises, des chercheurs mais aussi de simples enthousiastes de ces technologies et des bases de données intéressés par la problématique de la standardisation. Une enquête à destination des constructeurs de balises a également été menée en prévision du colloque et a collecté 26 réponses pour 42 sollicitations. L'ensemble des résultats issus de l'enquête et de l'atelier de travail peut être trouvé sur le site de l'IBLS (International Bio-Logging Society, 2020). L'enquête a révélé que la majorité des personnes interrogées reconnaît que l'absence de standard constitue un problème dans le monde du bio-logging et serait prête à travailler dans ce sens. Les fabricants ont indiqué qu'ils souhaitaient être impliqués et accompagnés dans ce processus qui doit faire intervenir des utilisateurs unifiés sous une même voix.

Cependant, comme expliqué précédemment, la problématique de la standardisation ne se limite pas à la communauté du bio-logging. En effet, même si un travail avec les fabricants de balises permettrait de résoudre de nombreux problèmes intervenant sur les premières étapes du cycle de vie des données de bio-logging (Figure 10), il ne faut pas perdre de vue le contexte global.

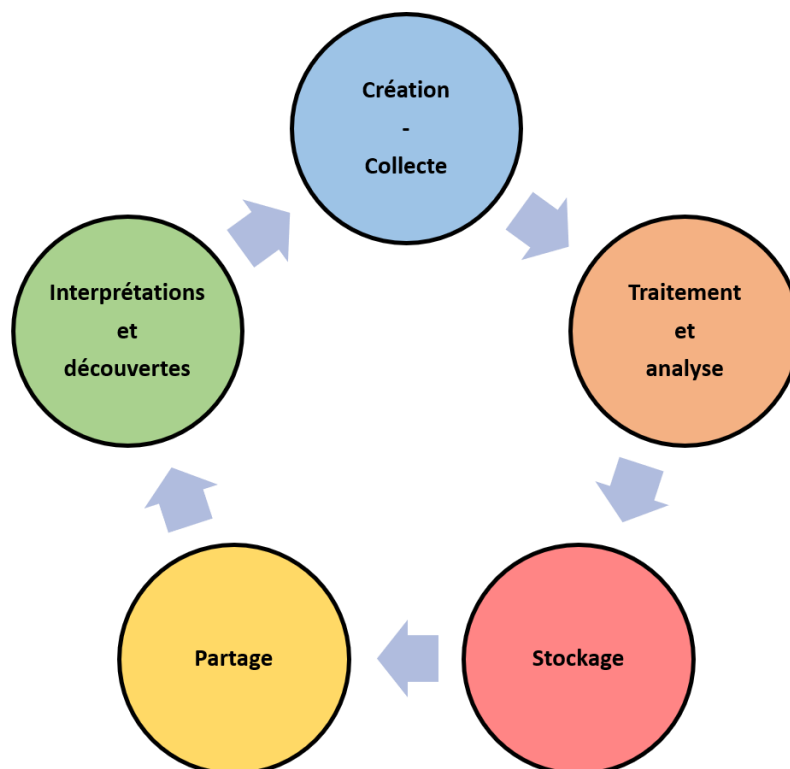


Figure 10. Cycle de vie des données de bio-logging.

Ces données sont souvent partagées sur différentes plateformes et utilisées au-delà de la seule étude au cours de laquelle elles ont été produites. (Campbell *et al.*, 2016) ont notamment dressé un catalogue d'infrastructures permettant le partage et la réutilisation des données de bio-logging. Les infrastructures de ce genre se sont également multipliées au cours des dernières années renforçant les divergences d'ordre sémantique. Bien qu'il n'existe pas une référence unique faisant l'objet d'un consensus international, le Darwin Core (DwC) est certainement le standard le plus utilisé aujourd'hui en informatique de la biodiversité. Le DwC offre un cadre stable, simple et flexible permettant la compilation de données de biodiversité venant de sources diverses et variables et joue un rôle fondamental dans le partage, l'utilisation et la réutilisation des données de biodiversité en accès libre. Conçu et maintenu par la communauté Biodiversity Information Standards (abrégée TDWG car initialement nommée Taxonomic Databases Working Group), le standard DwC est le standard de référence de plusieurs projets de grande envergure comme le Global Biodiversity Information Facility (GBIF), l'Ocean Biodiversity Information System (OBIS) ou encore l'Atlas of Living Australia (ALA). A l'issue de son atelier de travail, l'IBLS a préconisé l'utilisation de ce standard. Le milieu du bio-logging a donc des éléments de réponse aux problématiques auxquelles il est confronté, cependant la valorisation parfois insuffisante des données ne réside pas uniquement dans l'absence de standardisation.

C) Les difficultés pour maximiser la valorisation des données

1. La prolifération d'outils

Parmi les infrastructures citées dans (Campbell *et al.*, 2016), certaines ne servent pas uniquement au partage de données mais permettent également d'intervenir sur d'autres étapes du cycle de vie de ces données : leur traitement, leur analyse et leur interprétation. Prenons l'exemple de ZoaTrack, plateforme ayant pris la relève de OzTrack qui a été développée par une équipe de chercheurs de l'Université du Queensland comme une partie d'un programme partiellement financé par le gouvernement australien. Alors que d'autres infrastructures similaires sont réservées pour des groupes spécifiques d'animaux, ZoaTrack se concentre à un usage régional (au niveau de l'Australasie). Cette plateforme peut donc servir de dépôt de données mais également d'outil de traitement, d'analyse et de visualisation. Le nombre d'outils de ce type a suivi la tendance générale observée dans le monde du bio-logging connaissant ainsi une croissance qui peut s'expliquer par différents facteurs.

Comme évoqué plus tôt, les progrès technologiques ont permis le développement du bio-logging et de l'étude des mouvements des animaux. Cette discipline étant cependant récente dans le domaine de l'écologie, il a donc fallu développer des méthodes permettant d'analyser les données de déplacement afin de les valoriser au travers d'interprétations pertinentes. Les avancées techniques ont ainsi rendu nécessaires et poussé à la réalisation d'avancées analytiques. Les méthodes développées permettent notamment de s'attaquer à des problématiques techniques comme celle de la quantité de données (Nuijten *et al.*, 2020), mais également écologiques comme l'étude comportementale (Beyer *et al.*, 2013), la mise en évidence de changements dans les modes de déplacement ou du domaine vital (Patin *et al.*, 2020), ou encore pour l'étude des méthodes de sélection d'habitat (Muff *et al.*, 2020). Les travaux de ce type ont permis le développement de nombreux outils de traitement et d'analyse des données de mouvements d'animaux, notamment sous la forme de bibliothèques logicielles utilisables avec R, le langage de programmation et logiciel libre largement utilisé pour les statistiques. En 2020, près de 60 bibliothèques R, destinées à

différents usages et problématiques spécifiques à ce domaine, ont été listées et évaluées par (Joo *et al.*, 2020). Une synthèse de ces évaluations peut être trouvée dans (Jambon, 2019).

Les nombreuses avancées méthodologiques réalisées au cours de la dernière décennie ne sont cependant pas le seul facteur expliquant la prolifération de plateformes telles que ZoaTrack. La multiplication de la quantité de données générée par le bio-logging a entraîné divers effets secondaires dont celui de partiellement modifier les compétences requises pour ce genre de travaux. En effet, les études de bio-logging nécessitent désormais très souvent une approche interdisciplinaire avec des experts dans le domaine de la physiologie animale, de la modélisation, de l'écologie de terrain, de l'ingénierie ou encore de l'informatique. Cependant lorsqu'une telle association de connaissances n'est pas possible, les chercheurs, et notamment ceux issus du monde de l'écologie, sont donc forcés à devenir polyvalents et étendre leur champ de compétences. Toutefois, au-delà de cette problématique d'acquisition de nouvelles capacités, la douleur la plus citée par la communauté scientifique concerne le temps qui doit être consacré au traitement et à l'analyse des données. La quantité de données récoltées au cours des études ayant particulièrement augmenté, ces étapes sont maintenant beaucoup plus chronophages qu'auparavant et constituent alors une réelle contrainte, en particulier pour les équipes les plus réduites. Pour cette raison, les chercheurs essaient autant que possible d'automatiser ces processus afin de pouvoir consacrer plus de temps à ce qui constitue le cœur de leur métier.

2. La redondance des outils

Les plateformes permettant le traitement, l'analyse et/ou la visualisation de données de mouvements d'animaux sont donc devenues de plus en plus nombreuses au cours des dernières années. Une liste non exhaustive a été dressée dans le Tableau 1. Bien souvent développées par des membres de la communauté scientifique, elles sont souvent conçues de manière isolée avec une considération minimale pour la problématique de l'interopérabilité. De plus, beaucoup de ces outils reposent sur les bibliothèques R évoquées précédemment et peuvent donc être assez redondants.

Tableau 1. Liste non exhaustive d'outils permettant le traitement, l'analyse et/ou la visualisation de données de bio-logging.

| Nom | Résumé | Auteur(s) | Fonction(s) |
|--|--|--|--|
| ZoaTrack | Plateforme de stockage, traitement, analyse, visualisation et partage de données de suivis d'animaux | ALA | Dépôt de données, traitement, analyse, visualisation |
| Movebank | Plateforme de stockage, traitement, analyse, visualisation et partage de données de suivis d'animaux | Max Planck Institute of Animal Behavior, North Carolina Museum of Natural Sciences, University of Konstanz | Dépôt de données, traitement, analyse, visualisation |
| Animal Tracker App | Application mobile de visualisation des données présentes sur le dépôt Movebank | Max Planck Institute of Animal Behavior | Visualisation |
| Douglas Argos-Filter Algorithm | Permet de filtrer les localisations Argos selon 3 méthodes en supprimant les points inexacts sur la base de seuils définis par l'utilisateur | David Douglas, U.S. Geological Survey | Traitement |

| | | | |
|--|--|--|--|
| DYNAMO vis | Programme JAVA permettant de visualiser et d'animer des données de suivis d'animaux et qui produit des exports vidéo | Dr. Somayeh Dodge, Department of Geography, University of California Santa Barbara | Analyse, visualisation |
| Tagged Animal Movement Explorer (TAME) | Application web de visualisation et de traitement faite pour des petits jeux de données sous format csv | Jeffrey Walker & Benjamin Letcher, U.S. Geological Survey | Traitement, visualisation |
| ctmmweb | Application web de traitement, d'analyse et de visualisation de données de suivis d'animaux basée sur la bibliothèque R ctmm | X. Dong, C.H. Fleming, M.J. Noonan, J.M. Calabrese | Traitement, analyse, visualisation |

Cependant, bien que tous ces outils puissent fournir une certaine aide dans le processus de traitement et d'analyse de leurs données, les chercheurs sont parfois réticents à les utiliser. En effet, même si les chercheurs sont généralement enthousiastes à l'idée de pouvoir se passer de programmation, les outils de ce type possèdent souvent des limitations importantes telles que le manque d'interopérabilité avec d'autres services, le manque d'options de paramétrage ou encore une ergonomie peu recherchée. Ainsi, la communauté scientifique du bio-logging ne possède pas d'outil qui fasse l'unanimité à la fois pour le traitement, l'analyse et la visualisation de leurs données. Ceux qui le peuvent ont tendance à développer leur propre solution pour qu'elle colle parfaitement à leurs habitudes de travail. Les grosses structures développent parfois leurs propres outils en interne. Cependant, les écologues et chercheurs n'ayant pas de compétences en programmation voient en ces processus particulièrement chronophages une douleur qu'ils essaient de minimiser en combinant l'utilisation de multiples outils.

Ainsi, comment pourrait-on faire pour parvenir à un environnement applicatif complet qui réponde aux différentes attentes et besoins des acteurs du monde du bio-logging tout en se reposant sur la multitude d'outils déjà existant ?

III – Étude d'un milieu riche en profils et en outils

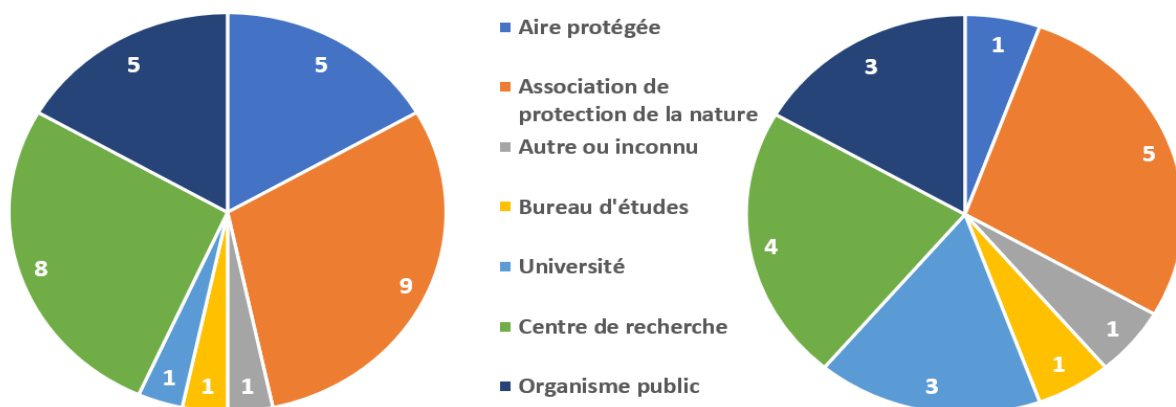
Pour nous attaquer à cette problématique, nous allons commencer par identifier et caractériser les profils des potentiels utilisateurs finaux de la solution à proposer qui seront décrits sous la forme de personae. Une persona est une personne fictive stéréotypée qui représente un groupe d'individus. Cela permettra ainsi de s'assurer de la cohérence du produit développé avec les besoins et les attentes de ses cibles.

A) Étude des profils rencontrés dans le milieu du bio-logging, leurs caractéristiques, leurs douleurs et donc leurs besoins

S'il y a bien une chose qui sera partagée par l'ensemble des utilisateurs, ce sera l'utilisation de balises de bio-logging ou au moins la possession de données de bio-logging. Nous pouvons donc déjà identifier 3 types de profils qui peuvent être concernés : les écologues (qui sont des spécialistes de l'écologie pouvant être chercheurs, ingénieurs ou biologistes), les chercheurs (pouvant être spécialisés dans différents domaines) et les naturalistes (qui sont des spécialistes des sciences naturelles pouvant être des scientifiques ou des amateurs). Ces profils peuvent se retrouver dans différentes structures, souvent publiques, telles que les centres de recherche, les universités, les parcs nationaux et autres aires protégées, ou encore les associations naturalistes et autres organismes dédiés à la conservation de la biodiversité. Notre étude reposera sur l'enquête présentée dans (Jambon, 2019) ainsi que sur différents échanges réalisés avec le Parc National des Écrins avec qui Natural Solutions collabore déjà depuis plusieurs années. Un travail d'analyse des résultats de l'enquête ayant déjà été fait dans (Jambon, 2019), nous nous pencherons ici uniquement sur les éléments que nous jugerons important pour la suite de ce mémoire.

1. Une enquête adressée à différents acteurs du milieu du bio-logging

Cette enquête, réalisée avec Google Form et diffusée par mail ainsi que sur Twitter et le groupe LinkedIn « biodiversity professionals », a été adressée à différents utilisateurs de balises de bio-logging. Une version francophone et une version anglophone ont été réalisées afin de toucher un public plus large et plus représentatif de la communauté du bio-logging. La 1^{ère} a reçu 35 réponses dont 30 de personnes utilisant des balises alors que la 2^{ème} a, elle, reçu 21 réponses dont 18 de personnes utilisant des balises. Bien que le nombre de réponses soit un peu faible pour la version anglophone, ce qui peut être déroutant d'un point de vue statistique, nous allons ici analyser les résultats des deux versions séparément. Ce choix repose principalement sur le fait qu'une partie non négligeable des réponses à la version francophone provient de contacts connus, ce qui peut introduire un biais statistique, alors que la version anglophone a été diffusée à des personnes sélectionnées aléatoirement. Ce choix peut évidemment être discuté car il a comme conséquence de diviser l'effectif qui n'est déjà pas très important au départ. Ainsi, il faudra garder cet élément à l'esprit pour l'ensemble des observations qui seront réalisées puisqu'il peut potentiellement remettre en cause la significativité des résultats.



Premièrement, nous observons que les divers types de structures cités précédemment sont bien représentés (Figure 11). On remarque également une différence de fréquence entre les types de structures selon la version du questionnaire. Notamment, les profils scientifiques (centres de recherches + universités) sont présents en plus grande proportion parmi les personnes ayant répondu à la version anglophone de l'enquête. Bien que la petite taille des échantillons ne permette pas d'affirmer la significativité de cette différence, cela permet d'illustrer la possibilité d'un biais introduit par la méthode de diffusion des différentes versions du questionnaires.

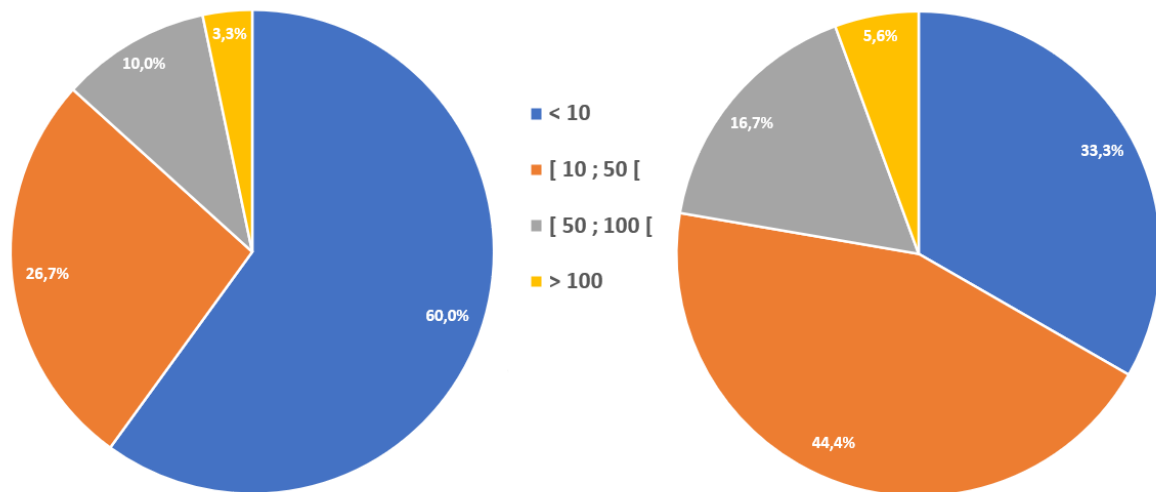


Figure 12. Proportions de différentes classes de quantités de balises déployées. L'intitulé de la question était « Combien de balises avez-vous actuellement en service ? ». Le graphique de gauche présente les réponses à la version francophone du questionnaire, celui de droite les réponses à la version anglophone. (Jambon, 2019)

Ensuite, il est intéressant de s'attarder sur le nombre de balises déployées. Nous pouvons d'abord constater que, pour les deux versions de l'enquête, la grande majorité des personnes ayant répondu ont déclaré utiliser moins de 50 balises (Figure 12). Cela peut sembler surprenant par rapport à l'accessibilité grandissante de ces dernières comme nous en avons discuté plus tôt. Cependant, ce sont des produits qui gardent un coût relativement élevé. Or, bien souvent, les acteurs du monde du bio-logging, et plus généralement de l'écologie, disposent de budgets assez restreints. Cette difficulté économique est en partie compensée par la désormais grande longévité des balises qui permet tout de même d'obtenir une quantité assez importante de données.

Nous pouvons également remarquer, encore une fois, une différence au niveau de la fréquence des différentes réponses entre les versions anglophone et francophone. En effet, pour la version francophone on observe une proportion deux fois plus importante de déploiements de moins de 10 que celle de déploiements avec un nombre de balises compris entre 10 et 50 (respectivement 60% et 26,7%). Alors que pour la version anglophone, ces deux proportions sont assez proches (respectivement 33,3% et 44,4%). Sans oublier le facteur « petite taille des effectifs », comment expliquer cette différence ? Nous pouvons formuler plusieurs hypothèses. Cette différence pourrait par exemple s'expliquer par un investissement économique pour le bio-logging moins important en France. Nous pourrions également penser que ceci est dû à la différence de proportion des différents types d'organisations constatée sur la Figure 11. Cependant, quand on s'intéresse plus en détails aux réponses, on remarque que 9 des 30 utilisateurs de balises francophones ayant répondu étudient les chiroptères. Or, pour chaque réponse concernant cet ordre de mammifères, le nombre de balises

déployées était inférieur à 10. La différence observée semble donc s'expliquer par une représentation des chiroptères inégale entre les réponses francophones et les réponses anglophones (qui elle-même peut être dû à un biais dans la sélection des contacts ou à une différence d'intérêt pour les chiroptères entre ces deux communautés). Les classes représentées sur la Figure 12 étant celles proposées en tant que réponses possibles, il aurait pu être intéressant, malgré le nombre relativement bas de réponses, de proposer un choix plus large de réponses afin d'avoir des classes plus fines.

Un autre point intéressant de l'enquête qui nous intéresse est la question des outils utilisés pour la valorisation des données de bio-logging (Figure 13). L'information étant demandée sous la forme d'une question unique et concernant donc l'ensemble des étapes du cycle de vie des données, de nombreuses réponses sont incomplètes et ne portent que sur une étape de ce cycle. Par exemple, sur 48 utilisateurs de balises ayant répondu, 5 personnes ont uniquement cité le langage de programmation R ce qui ne donne donc aucune information sur les outils permettant le stockage et la gestion des données. Il aurait alors été intéressant de séparer les questions par étape du cycle de vie des données afin d'avoir plus d'informations à ce sujet.

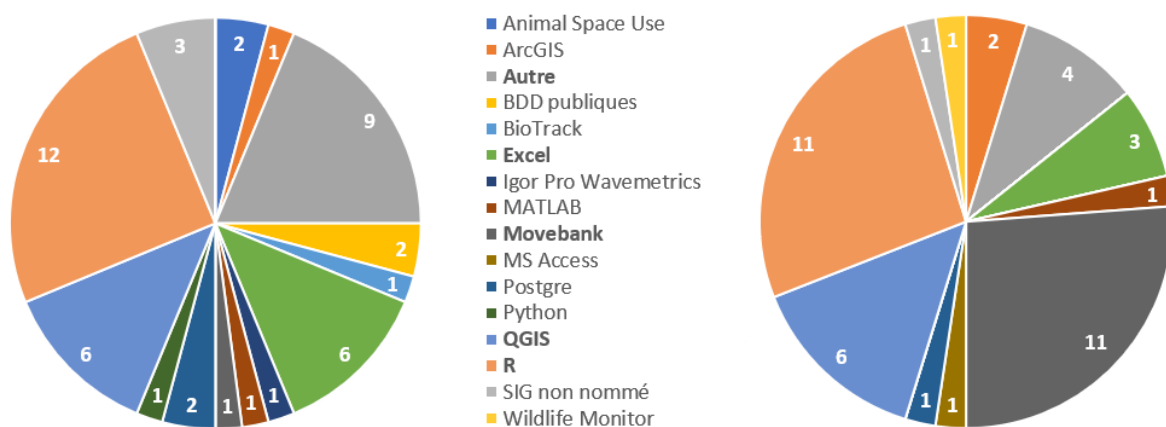


Figure 13. Outils utilisés au cours des différentes étapes du cycle de vie des données de bio-logging. La catégorie « Autre » correspond à des outils dont le nom n'a pas été donné ou des outils non disponibles au public. L'intitulé de la question était « Quel(s) outil(s) utilisez-vous pour stocker et analyser vos données ? ». Le graphique de gauche présente les réponses à la version francophone du questionnaire, celui de droite les réponses à la version anglophone. (Jambon, 2019)

Nous pouvons observer sur le graphique ci-dessus (Figure 13) une diversité relativement importante dans les outils cités pour les deux versions du questionnaire. Le langage de programmation R et l'outil de SIG QGIS ont à chaque fois une place majeure. La principale différence que l'on peut noter ici est la représentation particulièrement importante de Movebank dans les réponses anglophones, alors que la plateforme n'est citée qu'une fois dans les réponses francophones. Cela peut s'expliquer par le biais évoqué plus tôt introduit par la méthode de diffusion du questionnaire puisque Movebank a été utilisée comme source de contacts internationaux.

En regardant plus en détails les réponses, nous pouvons constater qu'un certain motif apparaît de manière récurrente dans les réponses : la combinaison d'un outil pour le stockage, d'un autre outil pour le traitement et l'analyse et d'un outil encore différent pour la visualisation. Dans la version anglophone, ce motif apparaît 10 fois sur les 18 réponses. Dans ces 10 réponses, la plateforme Movebank, citée dans le Tableau 1, est présente 9 fois. Bien que cette plateforme puisse être utilisée pour l'ensemble du cycle de vie des données, nous allons supposer qu'elle est ici utilisée

uniquement pour le stockage et le partage des données puisque d'autres outils intervenants sur d'autres étapes sont cités à ses côtés. Le langage R, utilisé principalement pour le traitement des données, est lui cité dans les 10 réponses présentant ce motif. Quant à la visualisation les outils de SIG (QGIS et/ou ARCGIS) sont cités 5 fois. Pour les 5 réponses ne citant pas de logiciel de SIG, il semble difficile de conclure si cela est dû à un oubli de la part de la personne enquêtée ou à l'utilisation d'un autre outil, comme Movebank, pour la visualisation. Pour la version francophone de ce questionnaire, ce motif est bien moins représenté puisque nous pouvons compter seulement 10 occurrences pour 29 réponses.

Le dernier point de l'enquête sur lequel il serait intéressant de s'attarder concerne les douleurs, ainsi que les besoins et attentes qui en découlent, citées par les personnes ayant répondu. Cette question ouverte portait spécifiquement sur les difficultés rencontrées avant, pendant, ou après la phase de traitement des données. Nous pouvons d'abord remarquer sur le graphique ci-dessous (Figure 14) que les limites liées à l'utilisation de balises et aux sujets des études n'ont pas encore complètement disparu puisque des difficultés sont toujours évoquées. Il est intéressant de constater la diversité des sources de douleurs. Cependant, il est compliqué d'en tirer des conclusions à cause du nombre de réponses n'ayant pas explicité la nature de leurs difficultés et qui pourraient par exemple rentrer dans la catégorie problème de temps, de compétences ou encore de volume de données.



Figure 14. Sources de douleurs au cours de l'étape de traitement des données. La catégorie « Difficulté de nature non précisée dans le traitement des données » correspond aux réponses n'ayant pas explicité s'il s'agissait d'un problème de temps, de compétences... La catégorie « Autre » correspond, elle, aux réponses qui ne semblaient pas correspondre à une douleur. L'intitulé de la question était « Actuellement, quelles difficultés rencontrez-vous avant, pendant ou après ce processus (traitement et analyses des données) ? ». Le graphique de gauche présente les réponses à la version francophone du questionnaire, celui de droite les réponses à la version anglophone. (Jambon, 2019)

Pour ce qui est des besoins et des attentes, le graphique ci-dessous (Figure 15) présente les caractéristiques et fonctionnalités que les personnes enquêtées aimeraient trouver dans un outil. Premièrement, on peut constater que l'automatisation et le gain de temps semblent particulièrement souhaités, ce qui semble cohérent avec les résultats présentés sur le graphique ci-dessus (Figure 14). Ensuite, on remarque dans les réponses francophones que l'accent est également mis sur la facilité d'utilisation et la présence d'explications claires du fonctionnement de l'outil. Cela peut s'expliquer par la plus faible représentation des profils scientifiques, plus habitués aux outils de traitement de données, dans cet échantillon de personnes, comme évoqué plus tôt. Une lecture des réponses permet de vérifier cette hypothèse.

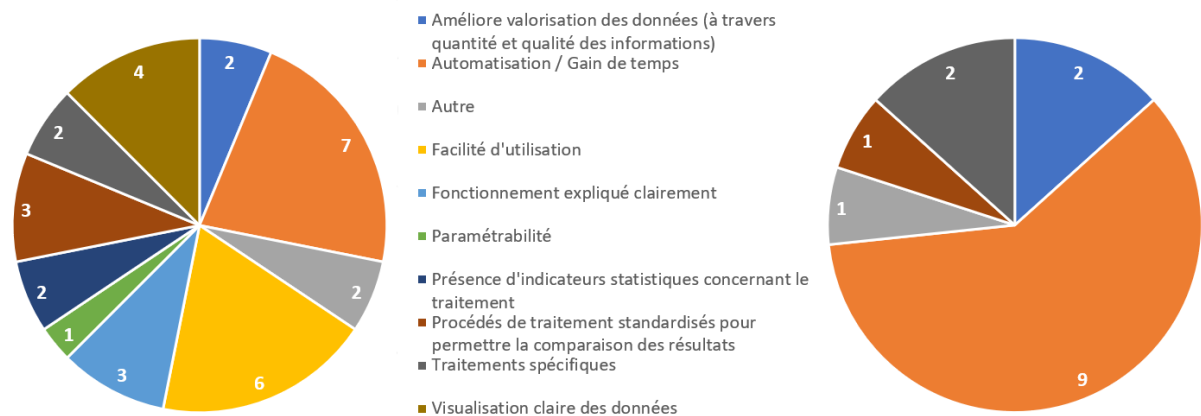


Figure 15. Caractéristiques et fonctionnalités que les personnes enquêtées souhaiteraient retrouver dans un outil. L'intitulé de la question était « Qu'espérez-vous d'un outil idéal ? ». Le graphique de gauche présente les réponses à la version francophone du questionnaire, celui de droite les réponses à la version anglophone. (Jambon, 2019)

A partir de ces différents résultats, nous allons établir les personae qui nous permettront de nous assurer qu'au cours de notre travail nous ne nous éloignerons pas des besoins du milieu du bio-logging révélés par l'enquête.

2. Les personae définis à partir des résultats de l'enquête

L'enquête étudiée précédemment a permis d'établir les personae dont les caractéristiques principales sont présentées dans le tableau ci-dessous (Tableau 2). Il semble cependant important d'apporter quelques précisions sur le contenu de ce tableau. D'abord, bien que le but d'une persona soit de représenter un archétype, il ne faut pas oublier qu'il existe également des profils qui divergent de ceux décrits ici. Par exemple, certains naturalistes sont assez réticents à l'idée d'ouvrir leurs données notamment par crainte de détournements de celles-ci. Une persona est ainsi supposée décrire le profil le plus représentatif d'un groupe ou d'une catégorie d'individus.

Ensuite, un point important n'est pas explicité dans le tableau : la manière dont les différents personae souhaitent diffuser leurs données. En effet, un chercheur qui sera principalement intéressé par l'archivage de ses données afin de pouvoir les associer à un article publié pourra également être intéressé par leur diffusion au grand public. Cependant, un chercheur n'aurait dans ce cas-là pas les mêmes objectifs qu'un écologue ou un naturaliste. Un naturaliste faisant partie d'un parc national français, par exemple, souhaiterait lui plutôt diffuser ses données afin de mettre en avant son travail de collecte, promouvoir le parc et peut-être également sensibiliser le public sur des questions liées à la biodiversité. Ainsi, le naturaliste sera plus attaché à des aspects tels que l'identité graphique ou encore la communication.

Nous verrons plus tard de quelle manière nous pouvons utiliser ces personae et leurs caractéristiques afin de proposer une réponse pertinente à la problématique de ce mémoire.

Tableau 2. Personae établis grâce aux résultats de l'enquête.

| | | Chercheur | Ecologue / Naturaliste |
|--------------------|---|--|--|
| Points clés | Domaine d'expertise | Ecologie | |
| | Ressources à disposition | Budget limité ne permettant pas d'aller chercher les compétences manquantes ailleurs | |
| | Destinée des données | Archivage pour publication scientifique | Diffusion au public |
| | Complexité des traitements réalisés | Elevée | Basique |
| Compétences | R | Se débrouille car régulièrement amené à l'utiliser mais ne le maîtrise pas totalement | Connaissance basique ou inexistante (le poussant alors à utiliser Excel ou un autre outil) |
| | Informatique générale | Compétences basiques mais peut potentiellement disposer d'un service informatique au sein de sa structure | |
| Douleurs | Le traitement des données lui prend trop de temps, temps qu'il préférerait investir pour du travail de terrain ou pour l'interprétation des données | | |
| | | Est contraint d'utiliser de nombreux outils différents pour répondre à l'ensemble de ses besoins | Est contraint d'utiliser des outils peu pratiques (e.g. Excel) ou non maîtrisés (e.g. R) |
| | Doit investir beaucoup de temps pour maîtriser les outils existants | | |
| | Manque de visibilité sur les outils existants | | |
| Souhaits | Gagner du temps, notamment à travers l'automatisation de certaines étapes | | |
| | Disposer d'un outil facile/rapide à prendre en mains | | |
| | | Disposer d'explications complètes et détaillées du fonctionnement de l'outil pour être capable de comprendre les résultats obtenus et donc d'avoir un regard critique dessus | Disposer d'explications claires du fonctionnement de l'outil |
| | Pouvoir adapter l'utilisation de l'outil à ses besoins | | |

B) Étude d'outils existants

Dans cette partie, nous allons présenter certains outils qui assistent aujourd'hui les acteurs du milieu du bio-logging lors de différentes étapes du cycle de vie de leurs données. L'objectif est d'obtenir une vision claire de leur fonctionnement, leurs forces ainsi que leurs faiblesses afin de démontrer qu'ils pourraient constituer les parties d'un environnement applicatif complet et pertinent.

1. Movebank, une plateforme multifonctions phare dans le milieu du bio-logging

Movebank est une plateforme en ligne gratuite accompagnant les chercheurs dans le stockage, la gestion, l'analyse, le partage et l'archivage de données sur les mouvements d'animaux. Débuté en 2007, le projet Movebank est hébergé par l'Institut Max-Planck de comportement animal en coordination avec le musée des sciences naturelles de Caroline du Nord, l'Université d'État de l'Ohio et l'Université de Constance. L'absence de normes à propos des formats, structures et descriptions des données ainsi que l'augmentation de la quantité de données récoltées a poussé la Société Max-Planck pour le développement des sciences (MPG) à lancer ce projet. Sa réalisation a notamment été soutenue par la National Science Foundation (NSF), le Centre allemand pour l'aéronautique et l'astronautique (DLR), la Fondation allemande pour la recherche (DFG) ainsi que la National Aeronautics and Space Administration (NASA). Movebank est désormais financée à long-terme par la MPG et l'Université de Constance. Grâce à ces soutiens de poids, Movebank est devenue une plateforme majeure du milieu du bio-logging (Figure 16) et compte plus de 20 000 utilisateurs. Movebank est ouverte à tous les utilisateurs indépendamment du type de structure qu'ils représentent ainsi qu'à toutes les études indépendamment de l'espèce étudiée, de la zone d'étude ou du financement de l'étude.

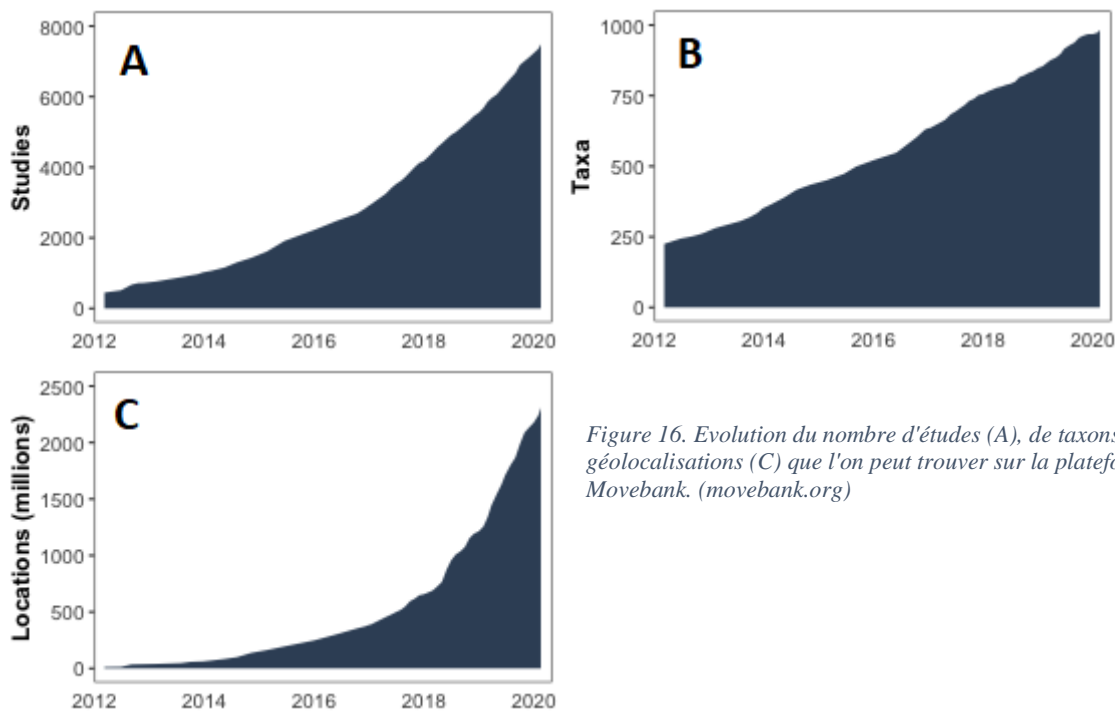


Figure 16. Evolution du nombre d'études (A), de taxons (B) et de géolocalisations (C) que l'on peut trouver sur la plateforme Movebank. (movebank.org)

a. Le stockage, l'archivage et le partage de données dans Movebank

Movebank est donc une plateforme qui propose de stocker des données de bio-logging, c'est-à-dire des géolocalisations mais également toutes informations supplémentaires pouvant y être associées (e.g. des données environnementales, physiologiques, des informations sur l'individu suivi ou la balise qu'il porte...). Le stockage d'une telle diversité de données est rendu possible à travers un Modèle Conceptuel de Données (MCD) composé d'une très longue liste d'attributs. Une version simplifiée de ce modèle est présentée sur la Figure 17. En effet, le MCD a été conçu avec l'objectif de contenir la totalité des champs qu'il soit possible de trouver dans une étude de

bio-logging, et cela inclus les attributs spécifiques aux différentes technologies et aux différents fabricants de balises. Par exemple, on peut retrouver dans le dictionnaire de données de Movebank (Max Planck Institute of Animal Behavior, 2019) 25 attributs dédiés spécifiquement à la technologie Argos ou encore 15 pour le fabricant de balises e-obs GmbH. Ce modèle contient également des attributs réservés à une éventuelle utilisation des outils d'analyse de données proposés sur Movebank comme le champ « Manually marked valid » par exemple. Il n'existe dans ce MCD qu'un seul attribut destiné à une utilisation non prévue par Movebank (« Study specific measurement »).

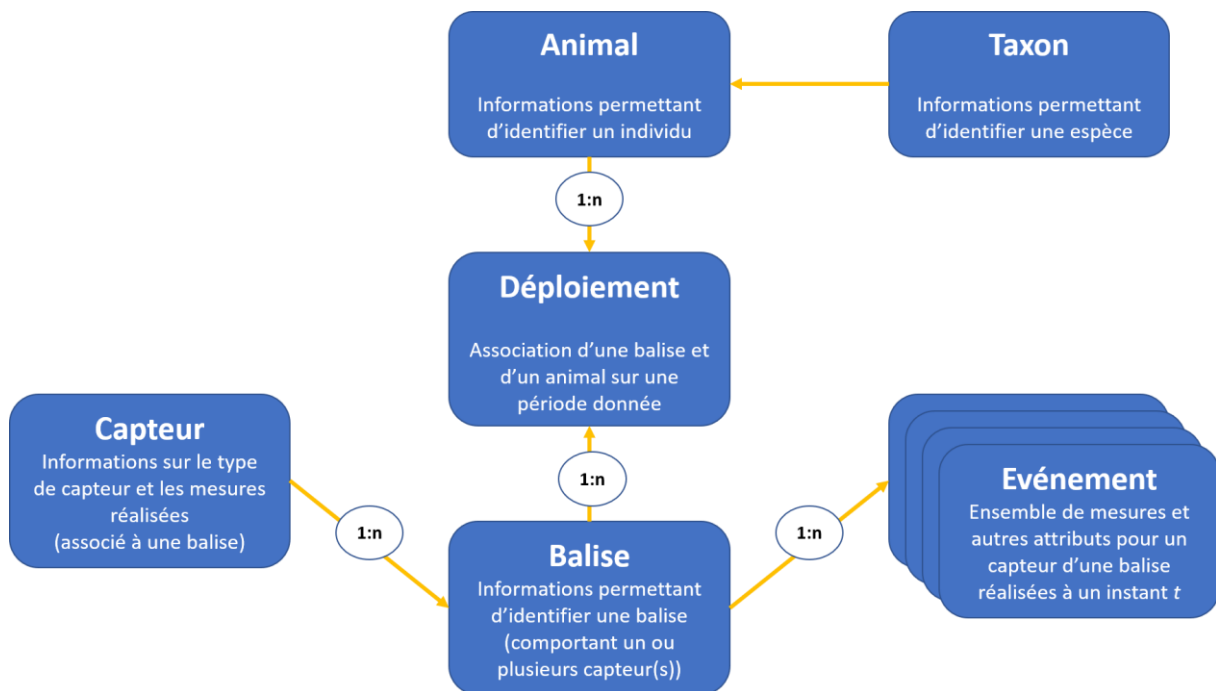


Figure 17. Version simplifiée du Modèle Conceptuel de Données de Movebank.

Avant de pouvoir importer des données, l'utilisateur authentifié doit au préalable créer une étude pour laquelle il renseigne un certain nombre de paramètres et d'informations. Parmi les paramètres en question, l'utilisateur peut définir qui aura la possibilité de voir et/ou de télécharger les données. Une étude peut ainsi être totalement privée et donc consultable uniquement par son créateur. Le propriétaire des données a également la possibilité pour son étude de définir un Contrat de Licence Utilisateur Final (CLUF) que tout utilisateur devra accepter une fois avant de pouvoir télécharger les données de l'étude (dans le cas où elle n'est pas privée). Autre fonctionnalité intéressante, dans le cas d'une étude privée le créateur de l'étude a la possibilité de définir à l'avance une date à laquelle les données deviendront publiques.

L'import de données dans la base de Movebank peut se faire de deux façons différentes : via l'import manuel d'un fichier ou à travers l'établissement d'un flux direct entre la balise ou son fabricant et Movebank. En choisissant l'import manuel, l'utilisateur peut choisir parmi une liste de fabricants de balise et de formats associés pour faciliter l'import ou choisir un format non standard. Les données doivent être sous forme tabulaire, les fichiers importés peuvent donc être au format CSV ou au format TXT. Après avoir sélectionné son fichier contenant les données, l'utilisateur doit alors réaliser le mapping des attributs, i.e. il doit faire la correspondance entre les colonnes présentes dans son fichier et les attributs contenus dans le dictionnaire de données de Movebank.

Pour ce qui est des flux directs (fonctionnalité « Live Feeds »), cela n'est disponible qu'avec certains fabricants de balises. En effet, Movebank a collaboré avec une liste de constructeurs (18 à compter de début septembre 2020) pour lesquels il est possible de paramétrer un flux direct de données permettant d'alimenter une étude Movebank quasiment en temps réel.

Après avoir réalisé l'une de ces deux étapes, l'utilisateur dispose d'une interface lui permettant d'ajouter des métadonnées, de gérer des individus (création ou suppression d'animaux) et de gérer les déploiements de balises, c'est-à-dire d'associer une balise portant un ou plusieurs capteur(s) à un individu pour une période donnée. L'utilisateur a également la possibilité de modifier certains attributs de ses données à la main.

Movebank propose également aux chercheurs d'archiver les données. Lorsque l'étude d'un utilisateur est terminée, il a la possibilité de soumettre ses données pour qu'elles soient publiées. Les données sont alors revues et reçoivent un Identifiant Numérique d'Objet (DOI) ainsi qu'une licence.

b. L'analyse et la visualisation des données dans Movebank

Movebank ne se limite pas au stockage des données puisqu'elle comporte également un module de visualisation. On peut ainsi afficher les données d'un ou plusieurs individu(s) provenant d'une ou plusieurs étude(s) sur une carte Google Maps ou OpenLayers. Les données sont visualisables sous formes de points, de tracés ou de densités de points. Il est possible de changer le fond cartographique et les fonctionnalités de base des applications cartographiques comme zoomer et se déplacer mais les points ne sont pas sélectionnables, modifiables ou annotables.

Movebank possède également différentes options pour le traitement et l'analyse des données. En effet, plusieurs filtres sont proposés pour la détection de points aberrants. Ces filtres s'appliquent à la totalité des données d'une étude, c'est-à-dire celles déjà importées sur la plateforme et celles qui le seront après création du filtre.

- Filtre de doublons : l'utilisateur peut sélectionner des attributs parmi la liste de ceux présents dans ses données pour qu'ils soient utilisés comme clés de détermination de doublons. Les attributs timestamp et identifiant de balise sont sélectionnés par défaut. A savoir qu'un événement est désigné comme un doublon si l'ensemble des attributs sélectionnés par l'utilisateur sont effectivement dupliqués.
- Filtre sur une plage de valeurs : pour chaque attribut présent dans les données, l'utilisateur a la possibilité de définir un test qui permettra de marquer un point comme aberrant s'il ne passe pas soit l'un soit l'ensemble des tests.
- Filtre sur la vitesse de déplacement de l'individu : en renseignant l'erreur maximale de localisation (en m) du matériel utilisé dans l'étude ainsi que la vitesse maximale plausible (en m/s) de déplacement pour l'espèce étudiée, l'utilisateur peut choisir parmi 3 algorithmes détectant les points aberrants.
 - Algorithme « Simple outlier » : cet algorithme teste pour chaque point si le déplacement de ce point au précédent ainsi qu'au suivant (d'un point de vue temporel) nécessite une vitesse semblant aberrante. Si c'est le cas, le point testé est marqué comme aberrant.

- Algorithme « Valid anchor » : cet algorithme considère que le premier point du jeu de données est correct et teste à partir du second si les points suivants semblent valides. C'est-à-dire qu'à partir d'un point n , l'algorithme teste si le point $n + 1$ est correct (i.e. n implique pas une vitesse de déplacement aberrante). Si ce n'est pas le cas, le point $n + 1$ est marqué comme aberrant, l'algorithme teste alors le point $n + 2$ (toujours par rapport au point n) et ainsi de suite jusqu'à trouver un point correct. Si le point $n + k$ est déterminé comme valide, l'algorithme reprend à partir du point $n + k$, c'est-à-dire qu'il va tester le point $n + k + 1$ par rapport au point $n + k$ et ainsi de suite.
 - Algorithme « Longest consistent track » : cet algorithme parcourt chaque point du jeu de données et commence une nouvelle "piste candidate" pour le premier point et lorsqu'un point ne passe pas les paramètres du filtre par rapport au point précédent. Si le point actuel correspond à plus d'une piste candidate en fonction des paramètres de filtrage, il sera attribué à la piste candidate la plus longue. Après avoir parcouru l'ensemble de la piste, il utilisera la piste candidate la plus longue comme étant la bonne et signalera les points non inclus dans cette piste comme étant des valeurs aberrantes.
- Filtres spécifiques aux données Argos : ces filtres nécessitent que les classes de localisation, qui sont des classes indiquant la qualité des données (telles que définies dans le Manuel Utilisateur Argos), soit présentes dans le jeu de données.
- Filtre « Simple Argos filter » : l'utilisateur peut sélectionner la classe de localisation Argos la moins précise qu'il souhaite conserver dans ses données.
 - Algorithme « Douglas Argos filter » : cet algorithme détecte les points peu plausibles en fonction de seuils définis par l'utilisateur, ce qui permet d'adapter sa performance aux comportements de déplacement des espèces et aux objectifs de l'étude (Douglas *et al.*, 2012).

Pour accompagner ses utilisateurs dans l'analyse de leurs données, Movebank propose également un service appelé « Env-DATA Track Annotation » offrant la possibilité d'ajouter des données environnementales aux données provenant de balises. L'utilisateur peut ainsi, pour un ou plusieurs individu(s) de son étude ou pour une zone géographique, sélectionner des variables environnementales qu'il considère intéressantes pour l'analyse de ses données. Plusieurs centaines de variables environnementales sont disponibles, obtenues grâce à la collaboration entre Movebank et de nombreux organismes disposant de données de ce type tels que la NASA, l'Agence Spatiale Européenne (ASE) ou encore le Centre Européen pour les Prévisions Météorologiques à Moyen Terme (CEPMMT) par exemple. De plus amples explications sur le fonctionnement de ce service peuvent être trouvées dans le manuel utilisateur de Movebank (Max Planck Institute of Animal Behavior, 2020b).

c. Forces et faiblesses de Movebank

Movebank rassemble désormais une grande communauté d'utilisateurs et se propose alors comme une réponse intéressante à la problématique de standardisation des données dans le milieu du bio-logging. La plateforme se place notamment comme intermédiaire entre les fabricants et les

utilisateurs de balises. Ces collaborations constituent une des grandes forces de Movebank puisque la fonctionnalité Live Feeds qui en est issue semble particulièrement pertinente dans le contexte du bio-logging. En effet, de nombreuses structures qui réalisent des suivis d'animaux sauvages, telles que les associations de protection de la nature par exemple, ne possèdent pas en interne les compétences nécessaires pour la gestion de gros volumes de données, qui implique notamment l'administration de serveurs et de bases de données. A cause de la taille et des moyens limités de ce type d'organisations, la récupération, le stockage et la gestion des données peuvent alors représenter une difficulté non négligeable. Ainsi, on observe souvent des ensembles de données stockées dans des fichiers Excel, comme le montre l'enquête où Excel a été largement plus cité que des SGBD comme PostgreSQL (Figure 13), ce qui, au-delà d'être très peu pratique, présente de nombreux risques. Movebank offre donc gratuitement une solution à ce problème, bien que cela implique d'accepter que ses données soient stockées sur des serveurs qui n'appartiennent pas aux utilisateurs de balises.

Movebank est donc un acteur majeur du monde du bio-logging notamment grâce à ses nombreux partenariats qui lui permettent de proposer un grand nombre d'outils et de fonctionnalités intéressantes. Cependant, dans l'enquête étudiée plus tôt, parmi les 11 personnes anglophones ayant explicitement cité Movebank comme faisant partie des outils qu'ils utilisent, seulement 1 a cité uniquement Movebank. En effet, les 10 autres ont déclaré utiliser d'autres outils en complément de Movebank, dont R qui est cité dans chacune de ces 10 réponses ainsi qu'un logiciel de SIG (QGIS et/ou ArcGIS) que l'on peut retrouver dans 4 de ces 10 réponses. Comment expliquer que ces utilisateurs aient recours à d'autres outils pour des fonctionnalités que Movebank propose ? Premièrement, nous pouvons penser que cela est dû à la recherche de fonctionnalités plus poussées que celles présentes dans Movebank. De plus, certaines fonctionnalités de Movebank possèdent quelques limites comme le filtre qui détecte les doublons par exemple. En effet, cet algorithme, qui recherche des duplicatas de données à partir d'une combinaison d'attributs, élimine systématiquement la 2^{ème} occurrence sans prendre en considération les autres valeurs. Nous pouvons donc avoir un enregistrement avec un certain couple timestamp et identifiant de balise et des valeurs de longitude et latitude nulles (dus à un dysfonctionnement de l'appareil par exemple). Si l'algorithme trouve ensuite un autre enregistrement avec le même couple timestamp et identifiant de balise mais avec des coordonnées valides, il va marquer cet enregistrement comme aberrant parce qu'il l'aura rencontré en 2^{ème}. D'autres limites similaires peuvent être trouvées dans les fonctionnalités de Movebank, pourtant, ce ne sont pas elles qui semblent être les premières responsables du recours à d'autres outils par les utilisateurs.

En effet, très rapidement, lors de l'utilisation de la plateforme, l'interface se révèle peu ergonomique et déroutante. L'expérience utilisateur apparaît comme le principal défaut de Movebank, que ce soit au niveau de la visualisation cartographique des données ou au niveau des différentes étapes d'import, de traitement et de gestion. L'apparence est quelque peu vétuste, les différents onglets ne sont pas bien délimités et donc peu visibles et l'espace est globalement mal utilisé.

Movebank est donc un outil très intéressant, en particulier pour le stockage et le partage des données de bio-logging, mais présente certains défauts pouvant bloquer l'utilisateur dans l'exploitation des autres fonctionnalités de cette plateforme.

2. Le Parc National des Écrins, une locomotive pour la diffusion des données de biodiversité en France, et ses outils

Le Parc National des Écrins (PNE) est un parc national français créé en 1973 situé dans les Alpes. Ce parc, qui est à la fois une institution et un territoire, est géré principalement dans le but de protéger les écosystèmes et à des fins récréatives. En effet, son cœur est classé comme aire protégée de catégorie II par la Commission Mondiale des Aires Protégées (CMAA) de l'Union Internationale pour la Conservation de la Nature (UICN, principale organisation non gouvernementale mondiale consacrée à la conservation de la nature). Dans cette optique, le parc collecte depuis sa création de très nombreuses données sur le territoire et, à travers son pôle Système d'Informations (SI), acte depuis plusieurs années pour la diffusion de ces données. Début juillet 2020, le PNE a d'ailleurs publié sous licence libre près de 700 000 observations et descriptions d'espèces de faune et flore (*Observations de biodiversité (faune-flore) du Parc national des Écrins*, 2020). Cette opération s'inscrit notamment dans la démarche d'ouverture des données publiques de l'État et des administrations visant à encourager la réutilisation des données d'intérêt public. Pour le PNE, cette ouverture constitue une étape d'un projet mené depuis plusieurs années déjà et ayant commencé par la modernisation de ses outils et ses bases de données. Dans son attachement à l'Open Source, le parc a ensuite travaillé au développement de ses propres applications et portails web qui sont également publiés sous licence libre comme GeoNature.

GeoNature est une application composée de nombreux modules permettant la saisie, la gestion, la synthèse et la diffusion d'observations de faune et flore. Développé en collaboration avec les autres parcs nationaux français, et notamment le Parc National des Cévennes, cet outil est désormais déployé dans de nombreux territoires français mais également étranger comme en Jamaïque par exemple (*Biodiv'Jamaica*, 2016). Natural Solutions a participé au développement de certains modules de GeoNature et travaille régulièrement avec le pôle SI du PNE comme sur l'outil FollowDem, un site internet permettant de suivre les déplacements des bouquetins du parc équipés de balises.

a. *FollowDem, une application web de suivi des bouquetins du PNE à destination du grand public*

Au cours de l'année 2013, le PNE a lancé, grâce au soutien financier de l'Europe et des régions Rhône-Alpes et Provence Alpes Côte d'Azur, une étude visant à mieux connaître les déplacements et l'utilisation de l'espace des populations de bouquetins des alpes (*Capra ibex*). Ainsi, un groupe de mâles et de femelles adultes a été équipé de colliers transmettant plusieurs localisations quotidiennement. Une application web permettant de visualiser ces données a alors été conçue avec pour premier objectif leur diffusion auprès du grand public (Parc National des Écrins, 2013).

Les données sont transmises depuis les balises vers un serveur mail sous la forme de fichiers textes. Une couche applicative en PHP, le back-end de FollowDem, permet l'importation des données dans la base de données de l'application. Cette couche se connecte à intervalle régulier au serveur mail afin de copier les fichiers sur le serveur de l'application. Ces fichiers texte sont ensuite transformés en fichiers CSV qui sont eux-mêmes lus afin d'importer les localisations dans la base

de données de l'application. Le front-end de l'application quant à lui utilise la bibliothèque JavaScript libre de cartographie en ligne Leaflet ainsi que le framework d'interface Bootstrap.

Back-end : partie d'une application web exécutée côté **serveur** et donc non visible aux yeux de l'utilisateur final.

Front-end : partie d'une application web exécutée côté **client** (i.e. le navigateur) qui constitue donc ce que voit l'utilisateur final.

Bibliothèque logicielle : collection de **routines réutilisables** visant à faciliter le travail des développeurs.

Framework : ensemble cohérent de **composants** logiciels structurels guidant la conception d'une architecture logicielle.

Développée en 2013, cette application n'a connu que très peu de mises à jour depuis. La seule mise-à-jour notable, effectuée en 2018, visait notamment à permettre l'historisation et l'archivage des données des colliers inactifs. En effet, la base de données avait été conçue uniquement pour afficher les positions des colliers et animaux actifs mais pas pour stocker correctement un historique de toutes les données du programme. Natural Solutions a donc réalisé la modification du MCD de la base de données (annexe 3). Cette mise-à-jour a également été l'opportunité de développer un nouvel outil apportant des fonctionnalités qui manquaient alors cruellement à FollowDem pour la gestion des données.

b. FollowDem-admin, un outil de gestion des animaux et des balises qu'ils portent

FollowDem-admin, dont le support est effectué par Natural Solutions, est né de la volonté du PNE et ses partenaires de disposer d'une interface pour l'administration des données avec différents niveaux d'accès et de droits. En effet, avant le développement de cet outil, les opérations telles que l'ajout ou la suppression d'individus ou de colliers ne pouvaient se faire qu'à travers l'utilisation de scripts SQL. Or, ces opérations étant plutôt réalisées par des agents du parc plutôt que par des membres du pôle SI, il devenait nécessaire qu'elles soient possibles à travers une interface facile d'utilisation pour des personnes n'ayant aucune connaissance en SQL.

FollowDem-admin, que l'on peut voir comme un complément à FollowDem, est donc constitué d'un back-end développé en python à l'aide du framework Flask ainsi que d'un front-end (dénommé FollowDem-admin-front) développé avec le framework TypeScript Angular 6 et le framework d'interface Bootstrap 4. Le back-end gère l'authentification des utilisateurs ainsi que les connexions et les échanges avec la base de données contenant les positions des animaux et autres informations. Il permet également l'import de données dans la base de données à partir d'un fichier CSV. Le front-end est lui un simple tableau de bord permettant de réaliser les opérations évoquées auparavant.

SQL (Structured Query Language) : langage informatique servant à exploiter des bases de données **relationnelles** (i.e. rechercher, modifier, ajouter ou supprimer des données).

c. Forces et faiblesses de l'ensemble applicatif FollowDem

L'ensemble applicatif FollowDem, dont l'architecture est représentée de manière simplifiée ci-dessus (Figure 18), est intéressant même si les différentes parties qui le composent sont d'une qualité assez variable. En effet, nous avons d'un côté un module de visualisation et un module de récupération de données datant de 2013 et n'ayant pas ou peu évolué depuis leur création et de l'autre côté un module d'administration de données datant de 2018 et reposant donc sur des technologies plus récentes permettant de proposer une expérience utilisateur qui soit plus proche des standards actuels.

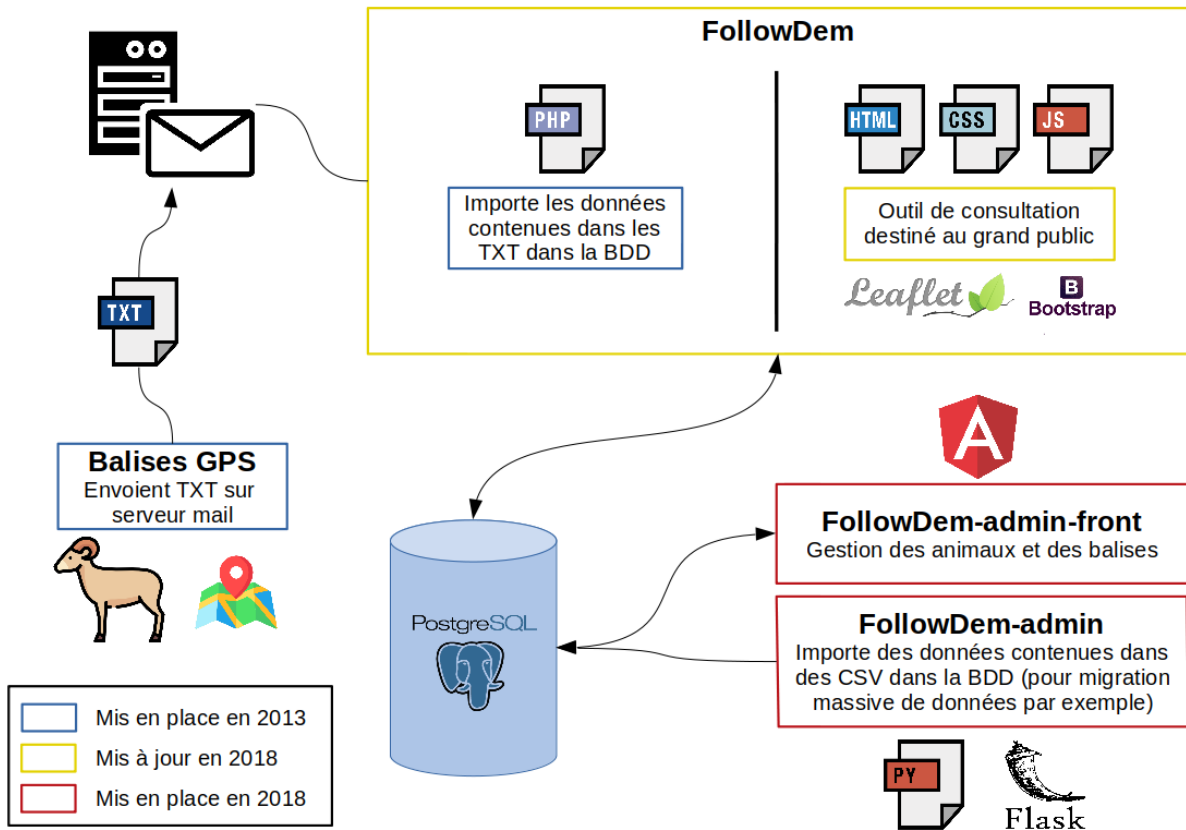


Figure 18. Architecture de l'ensemble applicatif FollowDem.

Le module de visualisation est particulièrement révélateur de cette différence à travers son interface peu recherchée et vieillissante ainsi que l'absence de certaines fonctionnalités dont l'ajout aurait permis de bonifier l'application sans en changer la nature. Il n'est par exemple pas possible de filtrer la liste des animaux qui possèdent pourtant des attributs qui pourraient être utilisés en tant que filtres (e.g. le sexe, l'âge...). Cependant, le module de récupération des données est celui qui pose le plus problème. Cette couche applicative a été spécialement conçue pour récupérer les données telles qu'elles leurs étaient, et sont toujours, fournies par le fabricant des balises utilisées pour les bouquetins du parc (i.e. par l'intermédiaire d'un serveur mail). Ce module n'est donc pas du tout adapté à un changement de méthode de transmission de données et repose ainsi sur la volonté du constructeur de balises de conserver ce fonctionnement. Cela implique également que si le souhait du PNE est d'acquérir de nouvelles balises d'un fabricant différent, il devra très probablement reconstruire un module de récupération de données. Des échanges avec le PNE ont révélé que cette dépendance représente une crainte dont le parc souhaiterait se débarrasser. Cette faiblesse est partiellement compensée par la possibilité d'importer des données contenues dans un fichier CSV grâce à un script python, mais c'est une solution qui présente également certaines

limites. De plus, l'outil étant Open Source, il devrait être adapté pour une utilisation dans des contextes rencontrés fréquemment plutôt qu'à des cas spécifiques, dans un souci de généricité, pourtant mot d'ordre du fonctionnement des outils du PNE.

Quant au module d'administration des données, il présente une certaine simplicité qui le rend efficace dans l'accomplissement de son rôle. Cependant, il a été conçu comme une partie totalement indépendante du reste (en dehors de la base de données et du MCD) ce qui implique l'ajout d'étapes d'installations lors du déploiement de l'ensemble applicatif. Les documentations de ces différentes parties ayant également été réalisées indépendamment les unes des autres, le processus d'installation global peut s'avérer relativement compliqué et peu accessible pour une personne ne possédant pas au préalable une bonne connaissance de l'ensemble. Pour s'attaquer à ce problème, il serait intéressant d'étudier la faisabilité et la pertinence de l'utilisation de la technologie des conteneurs qui permet le déploiement rapide d'applications. Devenue extrêmement populaire ces 5 dernières années avec l'apparition de la société Docker qui a grandement facilité sa mise en œuvre, cette technologie pourrait s'avérer très adaptée à un tel contexte.

L'ensemble applicatif FollowDem constitue donc une option intéressante pour la diffusion des données de bio-logging au grand public mais semble tout de même nécessiter certaines améliorations pour devenir réellement convaincante. La possibilité d'avoir une structure de base de données adaptée au bio-logging et prête à l'emploi semble être un atout de grande valeur pour s'adresser aux organismes ne souhaitant pas délocaliser le stockage de leurs données dans des plateformes telles que Movebank. Il reste cependant à trouver une option viable pour le traitement des données.

3. Animal Movement Cleaner Application, une application de traitement de données de mouvements d'animaux

Le traitement des données est une problématique récurrente dans le monde du bio-logging et à laquelle Natural Solutions est souvent confrontée, notamment à travers son client principal. En effet, Reneco effectue un suivi des outardes élevées puis relâchées afin de mieux connaître leur environnement ainsi que leurs interactions avec celui-ci. Plus de 20 000 individus sont produits par an depuis 2013 (References, 2020) et une certaine partie est relâchée en milieu naturel afin d'assurer la conservation de l'espèce. Parmi ces oiseaux relâchés, environ 400 sont équipés de balises qui permettent de suivre leurs déplacements, ce qui constitue une flotte d'appareils particulièrement importante quand on la compare aux résultats de l'enquête présentés plus tôt (Figure 12). La quantité de données alors générée pour une seule étude est donc également d'un ordre de grandeur plutôt inhabituel dans le milieu du bio-logging. La nécessité d'un outil de traitement de données de mouvements d'animaux n'en devient donc que plus grande. C'est ce besoin qui est à l'origine de l'idée de Natural Solutions de développer Animal Movement Cleaner Application (AMCA), une application web monopage adressée à n'importe quel possesseur de balises. Sa conception ayant été présentée en détails dans (Jambon, 2019), nous allons ici nous limiter à une présentation rapide de l'outil.

AMCA est donc une application web de traitement de données de mouvements d'animaux qui propose également une visualisation 3D. Les données peuvent être importées au format CSV et sont traitées automatiquement par le serveur une fois quelques paramètres renseignés. Parmi les paramètres présents sur l'application, on trouve la technologie utilisée (GPS ou Argos), la date de déploiement de la balise, la vitesse maximale plausible (en km/h) de déplacement pour l'espèce

étudiée et la durée d'immobilité maximale plausible (en h) que l'on peut constater pour l'espèce étudiée. La date de déploiement est utilisée pour déterminer le moment à partir duquel il faut traiter les données afin de ne pas prendre en compte les données correspondant à des tests de la part des écologues ou encore au transport de l'animal jusqu'au lieu de relâcher. La vitesse maximale plausible est utilisée pour filtrer les points qui impliquent un déplacement à une vitesse aberrante, à la même manière que le filtre de Movebank présenté précédemment. La technologie utilisée et la durée d'immobilité maximale plausible servent à détecter une immobilité anormalement longue pouvant indiquer que l'animal suivi est décédé ou encore que sa balise se soit détachée. Le renseignement de la technologie permet de définir une valeur d'erreur de localisation afin de prendre en compte le fait que des déplacements de quelques mètres peuvent être causés par une erreur. D'autres paramètres, comme le nom de l'espèce ou le type de cette dernière (terrestre, aérienne ou aquatique) ne sont pas encore utilisés par les algorithmes de traitement mais ont été laissés sur l'interface en vue de futurs développements. Les données, une fois traitées, sont renvoyées par le back-end, développé en python à l'aide du framework Pyramid, et peuvent alors être modifiées manuellement par l'utilisateur à travers le front-end, qui utilise le framework JavaScript Vue ainsi que la bibliothèque JavaScript Cesium qui permet la création de globes 3D.

AMCA présente deux atouts que sont la visualisation 3D et la simplicité d'utilisation. C'est effectivement le seul outil du milieu du bio-logging permettant la visualisation de données en 3D. Cette fonctionnalité s'avère particulièrement intéressante puisque l'altitude est une information supplémentaire permettant notamment la détection de points aberrants. Cela peut par exemple permettre, grâce à la visualisation du relief, de repérer un mammifère terrestre localisé à une altitude aberrante, ce qui est difficile à détecter automatiquement. La facilité d'utilisation est un autre atout d'AMCA puisque, comme l'a révélé l'enquête évoquée plus tôt, c'est une caractéristique particulièrement recherchée par les utilisateurs de ce type d'outils.

Cependant, cette simplicité s'explique en partie par le nombre limité de fonctionnalités d'AMCA dû au faible temps qui a pu être consacré au développement de cette application ayant été réalisée au cours d'un stage. De plus, le choix de développer les algorithmes de traitement de données en python plutôt que d'utiliser les nombreuses bibliothèques R open source existantes prive l'application d'un nombre important de fonctionnalités intéressantes, à moins de traduire l'ensemble de ces bibliothèques en python ce qui semblerait peu judicieux. Bien sûr, il faut replacer ce choix dans son contexte puisqu'il a été motivé d'abord par les compétences en présence au sein de l'équipe de Natural Solutions, puisque python y est largement utilisé et maîtrisé tout au contraire de R, et ensuite par le manque de temps ne permettant pas d'envisager une période de formation au langage R et aux moyens nécessaires à son inclusion dans une application python. De plus, ces algorithmes étaient destinés à être réutilisés dans d'autres applications de Natural Solutions développées en python. Dans une vision à plus long terme, il semble important d'intégrer les bibliothèques R les plus pertinentes qui sont déjà largement utilisées au sein de la communauté du bio-logging.

Nous venons donc de voir divers outils répondant à différents besoins et possédant chacun leurs forces et faiblesses. Nous allons désormais démontrer comment l'association de ces outils permettrait de proposer une solution complète et pertinente aux problématiques rencontrées dans le milieu du bio-logging.

C) Le potentiel détenu par l'association de ces outils

Revenons aux personae que nous avons définis plus tôt. D'un côté, nous avons un profil type de chercheur qui recherche en priorité le gain de temps sans pour autant vouloir sacrifier la qualité des traitements de données effectués. Ses compétences en R semblent suffisantes pour prendre en main un outil faisant appel à des notions légèrement complexes si une documentation claire est fournie. L'outil ne doit pas pour autant être trop compliqué à prendre en main. La visualisation de ses données lui sert avant tout pour les analyser mais également pour effectuer certains traitements et vérifier les résultats de ceux qui sont automatisés. Son objectif final étant généralement d'utiliser ses données pour une publication scientifique, le stockage organisé et l'archivage de celles-ci est important à ses yeux. De l'autre côté, nous avons un profil type de naturaliste ou écologue qui recherche également le gain de temps en priorité mais se contente de traitements plus basiques. La simplicité d'utilisation de l'outil est primordiale. La visualisation est également importante pour le traitement et l'analyse des données mais doit également servir un autre objectif : la diffusion des données auprès du grand public. Le module de visualisation doit donc également permettre de mettre en avant graphiquement l'identité d'une structure. Pour ce qui est du stockage des données, il serait intéressant de proposer plusieurs options afin d'être en mesure de s'adapter aux ressources assez hétérogènes, sur le plan des finances et des compétences, rencontrées d'une structure à une autre.

Pour répondre à ces besoins, de nombreux outils existent déjà et, bien qu'ils possèdent chacun leurs défauts, il semblerait peu judicieux de vouloir construire un nouvel outil qui viendrait s'ajouter à une liste déjà relativement longue. Il serait ainsi plus intéressant de voir comment nous pourrions essayer de constituer un ensemble idéal à partir de ces pièces existantes.

Movebank est un élément particulièrement intéressant pour le stockage et l'archivage notamment parce qu'il permet de s'affranchir gratuitement de plusieurs difficultés techniques impliquant un coût non négligeable (achat ou location de serveurs et salaires des personnes responsables de l'installation et l'administration). Ce sera donc l'option la plus adaptée pour les structures les plus modestes disposant de ressources économiques et humaines limitées. Pour les organisations disposant de plus de moyens et pour lesquelles la délocalisation de leurs données poserait un problème, l'association d'une base de données et de l'application FollowDem-admin pour sa gestion, installées sur un serveur administré par l'organisation en question, serait une solution pertinente. Nous avons donc deux options intéressantes pour ce qui est du stockage et de l'archivage des données.

Pour ce qui est de leur diffusion, bien que Movebank soit une plateforme intéressante pour le partage de données entre chercheurs, elle est cependant bien moins adaptée pour la diffusion au grand public qui ne fait pas partie de ses raisons d'être. Il faudrait donc une plateforme destinée uniquement à la visualisation des données qui soit adaptée pour une utilisation grand public et qui permette à une organisation de mettre en avant le travail qu'elle réalise autour de la biodiversité. L'application FollowDem existe dans ce but mais nécessiterait une grosse mise à niveau pour être plus en accord avec les standards actuels. Une réunion impliquant différentes personnes du PNE et de Natural Solutions ayant eu lieu à la fin du mois de juin a mené à la décision prise d'un commun accord de développer un nouvel outil plutôt que de repartir de fondations vétustes. Nous détaillerons plus tard les choix techniques qui ont été réalisés pour cette application qu'on appellera « viewer ».

Pour ce qui est du traitement et de l'analyse des données, AMCA représente une option intéressante qui mérite d'être considérée pour ses forces déjà évoquées. Cependant, elle possède également des points négatifs qui imposent la réalisation de certaines améliorations afin de rendre

cette application réellement attrayante. Nous reviendrons dans une partie ultérieure aux pistes d'améliorations envisageables pour AMCA.

Enfin, bien qu'il soit déjà intéressant de disposer de différents outils répondant à différents besoins, la réelle valeur ajoutée se situe dans leur association. En effet, rendre ces outils interopérables apporterait un réel confort aux utilisateurs qui sont actuellement obligés de jongler constamment entre divers systèmes ce qui rajoute une complexité supplémentaire dans leur travail. Pour réaliser cela, nous pourrions notamment utiliser l'API REST de Movebank qui permet de récupérer des informations sur les études présentes sur la plateforme ainsi que les données de ces études. Nous expliquerons son fonctionnement ainsi que les autres moyens employés pour rendre ces outils compatibles dans la suite de ce mémoire. Le schéma ci-dessous (Figure 19) représente de manière simplifiée l'architecture décrite ici.

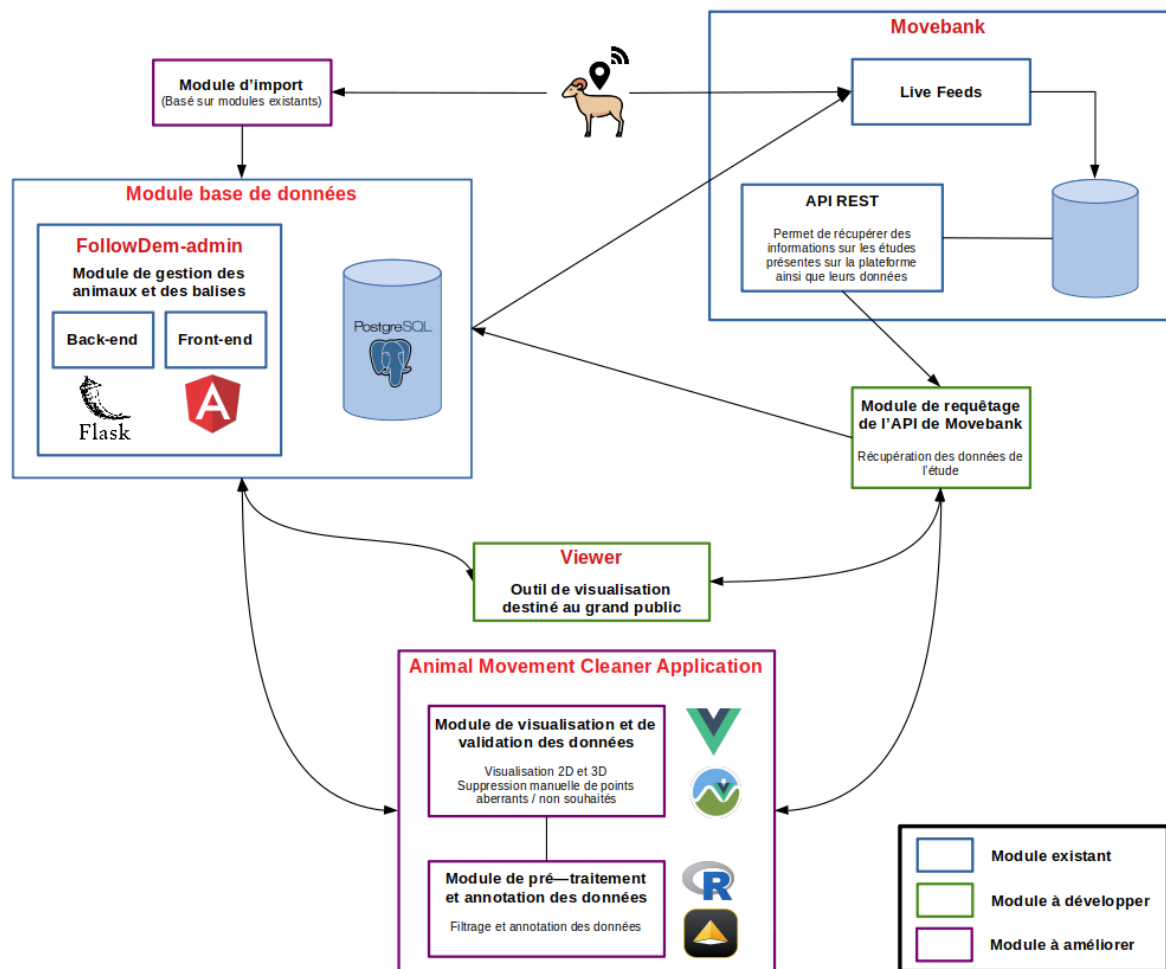


Figure 19. Architecture de l'ensemble applicatif idéal.

Maintenant que nous avons présenté l'approche théorique de notre proposition de réponse à la problématique de la redondance des outils dans le milieu du bio-logging, nous allons désormais nous pencher sur la mise en œuvre pratique de cette proposition.

IV – Les réalisations du stage

Comme indiqué sur le planning présent au début du corps de ce mémoire, la phase d'analyse des besoins dont les résultats viennent d'être présentés a duré relativement longtemps (Figure 4). Ce projet, bien qu'étant interne à Natural Solutions, a été pensé en collaboration avec le PNE, et notamment le responsable de son pôle SI, Camille Monchicourt, qui a l'habitude de travailler avec l'entreprise et qui a donc, d'une certaine manière, joué ici le rôle de client pilote. Cependant, la situation sanitaire actuelle a engendré une certaine inertie dans les échanges avec le PNE qui étaient essentiels pour la bonne compréhension du contexte. La 2^{ème} réunion avec le parc, déterminante dans la conduite du projet, ayant eu lieu fin juin le développement n'a pu être lancé que mi-juillet. Cette inertie n'a pas été totalement négative non plus puisqu'elle m'a permis d'étudier plus en profondeur le milieu du bio-logging et ses problématiques, notamment celle de la standardisation des données que j'ai pu aborder à travers le prisme de Movebank et son MCD. Cette période a également été mise à profit pour travailler sur des tâches indépendantes du projet présenté dans ce mémoire et qui concernaient les applications développées par Natural Solutions pour Reneco. Plus spécifiquement, ces tâches portaient sur la gestion des bases de données de Reneco qui permettent le suivi de l'ensemble des pratiques d'élevage. J'ai par exemple pu réaliser des procédures SQL permettant des vérifications sur l'intégrité des données lors de saisies ou d'imports. Ce fût ainsi une expérience largement bénéfique puisque j'ai pu découvrir et travailler sur un système particulièrement complexe et développé.

Nous allons donc présenter ce qui a pu être développé dans le temps imparti en commençant par expliciter ce qui a été priorisé et les choix techniques qui ont été réalisés.

A) Des priorités et des choix techniques

La priorisation, visant à organiser la période de développement des fonctionnalités en fonction du temps disponible, s'est faite selon plusieurs critères : l'importance de la fonctionnalité pour le produit, l'intérêt de son développement d'un point de vue pédagogique et le risque lié à la difficulté de la tâche. L'évaluation du premier critère est d'abord passée par la détermination de la valeur de chaque outil composant l'ensemble applicatif idéal que nous venons de décrire. En effet, l'importance d'une fonctionnalité est pondérée par la valeur de l'outil dont elle fait partie. Ce travail a notamment reposé sur les personae définis et présentés précédemment et sur les échanges avec le PNE, notamment au cours de la 2^{ème} réunion ayant eu lieu fin juin. Puisque l'interopérabilité des données ainsi que leur ouverture est au cœur des problématiques traitées dans ce mémoire, cette évaluation s'est faite en suivant le modèle de l'échelle de qualité des données ouvertes de Tim Berners-Lee, inventeur principale du World Wide Web (WWW), du langage HTML et fondateur du World Wide Web Consortium (W3C). Cette échelle, établie en 2010 dans le but d'encourager l'ouverture des données dans un format adapté, évalue la qualité de données mises en ligne en leur attribuant un nombre d'étoiles compris entre 1 et 5 (5 étoiles étant la meilleure note).

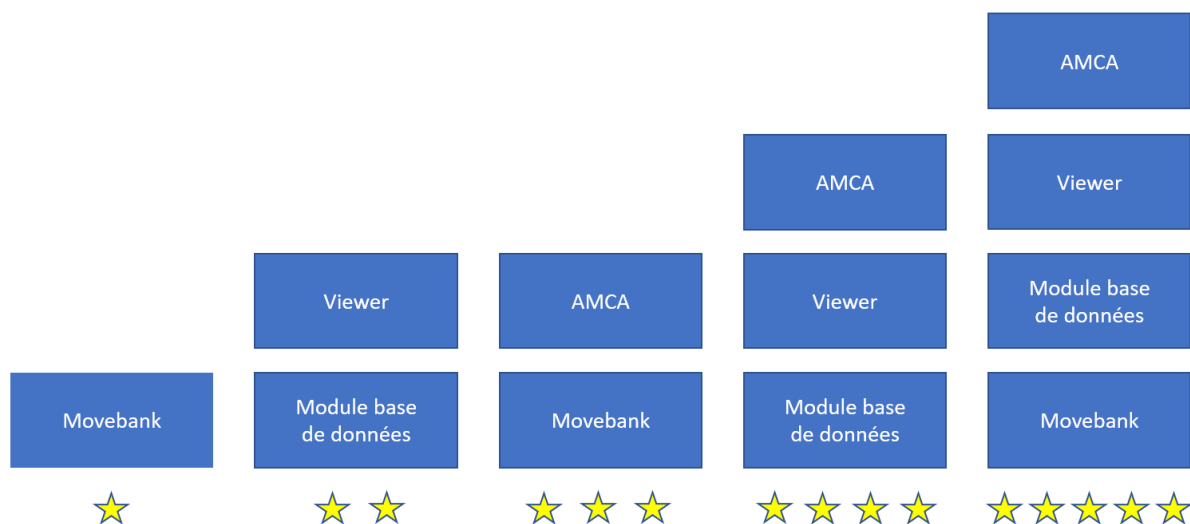


Figure 20. Evaluation de la valeur des différents outils présentés précédemment selon le modèle de l'échelle de qualité des données ouvertes. (Berners-Lee, 2010)

Ainsi, les outils décrits plus tôt atteignent leur valeur maximale lorsqu'ils sont tous associés, chacun profitant de ses interactions avec les autres (Figure 20). A partir de cette évaluation, on considère que l'outil le plus prioritaire est celui qui est le plus en amont de la chaîne de valeur. Ce qui donnerait alors, dans l'ordre, Movebank, le module base de données, le viewer et enfin AMCA. Cependant, le module base de données étant l'élément nécessitant le moins de changements, il sera considéré comme le moins prioritaire. Le module de requêtage de l'API de Movebank est donc en haut de la liste des priorités, position renforcée par le fait que cette API avait déjà fait l'objet d'une étude plus tôt dans le stage, limitant ainsi les risques pour le développement. Or, AMCA étant l'application qui devrait le plus profiter de ce module, elle passe alors devant le viewer dans l'ordre des priorités. Maintenant que cet ordre est établi (i.e. Movebank, AMCA, viewer puis module de base de données, respectivement du plus prioritaire au moins prioritaire), nous pouvons nous intéresser aux fonctionnalités à développer pour chacun d'entre eux afin de les prioriser. Le tableau ci-dessous (Tableau 3) présente ces différentes fonctionnalités ainsi que la priorisation qui a été retenue.

Tableau 3. Fonctionnalités à développer et leur priorisation (où 1 correspond au niveau de priorité le plus élevé).

| Outil | Fonctionnalité | Description | Priorité |
|----------|--|--|----------|
| Movebank | Récupération de données de Movebank à travers l'API REST | Module de requêtage de l'API de Movebank permettant de faciliter la récupération de données présentes sur la plateforme afin de les réutiliser dans une autre application. | 1 |
| AMCA | Visualisation de plusieurs jeux de données | La visualisation simultanée de plusieurs jeux de données peut être utile pour, par exemple, comparer les trajets de différents individus de la même espèce. | 2 |
| | Implémenter de nouveaux algorithmes de traitement de données | Implémenter certaines bibliothèques R libres dans AMCA pour permettre de nouveaux traitements comme le calcul | 5 |

| | | | |
|------------------------|---|---|---|
| | | du domaine vital d'un animal par exemple. | |
| | Importer des données de Movebank | Implémentation du module de requête de l'API de Movebank pour permettre aux utilisateurs d'obtenir directement depuis AMCA des données d'études stockées dans Movebank. | 1 |
| Viewer | Base d'application cartographique | Fonctionnalités de bases d'une application cartographique (déplacement, zoom, choix de fond cartographique...). | 3 |
| | Récupération et affichage de données provenant de différentes sources | Récupérer dynamiquement des données de différentes sources, telles qu'une base de données PostgreSQL ou un fichier CSV, et les afficher. | 3 |
| | Filtrage des données selon différents attributs | Filtrage des données affichées à partir d'attributs associés aux données de localisation (par exemple sexe ou âge de l'animal). | 4 |
| Module base de données | Récupération de données de balises de différents constructeurs | Rendre plus générique la récupération des données de balises afin de pouvoir s'adapter à un changement de fabricant. | 5 |

La priorisation des fonctionnalités à développer ainsi effectuée, nous pouvons nous intéresser aux choix techniques à réaliser, dont le nombre sera relativement restreint puisque l'essentiel du travail s'axe sur des éléments existants. Ces choix portent donc sur le module de requête de l'API de Movebank et sur le viewer.

Pour le premier, le choix fut relativement évident en considérant les technologies déjà utilisées pour les autres éléments. En effet, la partie back-end de AMCA ainsi que celle de FollowDem-admin, composante du module de base de données, ont été développées en python. Le langage python est largement utilisé et donc maîtrisé au sein de Natural Solutions, ce qui me donnait ainsi la possibilité de recevoir du soutien si nécessaire. Un argument de poids dans la prise de décision étant donné le faible temps disponible pour le développement et donc l'importance que pourrait avoir un blocage dû à une difficulté technique. De plus, une semaine avait été prise en début de stage pour suivre une formation python afin de ne pas être en difficulté avec le back-end de AMCA.

Pour le viewer, le choix a plus ou moins été imposé par le contexte. En effet, au moment des décisions techniques Natural Solutions a lancé un projet de développement d'une plateforme de gestion de patrimoine végétal à destination des gestionnaires des espaces. Cette application, nommée ecoTeka (EcoTeka, 2020), repose notamment sur un inventaire cartographique du patrimoine vert public et privé. Ce projet a donc mené au développement d'un outil de visualisation cartographique partageant une grande quantité de caractéristiques avec ce qui était imaginé pour l'outil présenté dans ce mémoire. Cependant, l'architecture d'ecoTeka présente un certain niveau de complexité et son utilisation impliquait que j'utilise des technologies que je ne maîtrise pas du tout. En effet, l'architecture de cette solution repose sur l'utilisation de tuiles vectorielles. Les tuiles sont des carrés de 256 x 256 pixels servant à représenter le monde en utilisant la projection Web Mercator (EPSG3857). Les tuiles vectorielles, qui contiennent des données de géométrie

vectérielles plutôt que des images précalculées, peuvent être mises en cache améliorant ainsi les performances d'une application cartographique. De plus, ces tuiles sont légères et donc faciles à stocker ainsi qu'à télécharger. Lors de la création d'un serveur de tuiles, on génère plusieurs cartes à différents niveaux de zoom permettant ainsi d'adapter l'affichage. Dans le cas d'ecoTeka, ces tuiles sont créées à partir de fichiers GeoJSON grâce à l'outil tippecanoe. Les fichiers GeoJSON peuvent eux-mêmes être générés, grâce à l'outil GDAL, à partir de différents formats de fichiers tels que les CSV ou à partir d'une base de données géographiques comme une base PostgreSQL utilisant l'extension géospatiale PostGIS. Les différentes cartes peuvent ainsi être créées dynamiquement et servies sur un serveur de tuiles. L'affichage cartographique est, dans le cas d'ecoTeka, géré par la bibliothèque JavaScript Mapbox GL JS. Le front-end mettant en place cette cartographie utilise lui la bibliothèque JavaScript React qui est assez similaire à Vue, le framework utilisé dans AMCA. Enfin, le tout est géré à travers la technologie de conteneurs, déjà évoquée plus tôt, qui permet un déploiement rapide de l'application pour son installation chez différents clients.

Tout ceci représente donc un nombre important d'éléments que je ne maîtrise pas et qui impliquent une difficulté technique quasiment incontournable à la vue du temps disponible. Le choix a tout de même été fait de reprendre cette architecture pour le viewer à développer. En effet, en considérant qu'il était assez peu probable que j'ai le temps d'aller au bout de ce développement qui serait alors potentiellement repris par des développeurs de Natural Solutions, et notamment ceux travaillant sur ecoTeka, à l'issue de ce stage, il semblait absurde de développer deux outils similaires en parallèle et de manière totalement indépendante. Ce fût pour moi l'opportunité de découvrir des technologies qui me semblaient intéressantes en ayant l'appui des développeurs de Natural Solutions pour cela.

Une fois ces choix effectués, le développement a ainsi pu démarrer mi-juillet afin de tenter de mettre en œuvre la solution décrite plus tôt dans ce mémoire.

B) Des réalisations diverses

La phase de développement a commencé par la réalisation du module de requêtage de l'API de Movebank, qui avait déjà été suggérée comme une perspective d'amélioration d'AMCA par (Jambon, 2019).

1. Un module de requêtage de l'API de Movebank

Le développement de ce module est passé par une phase d'étude de l'API de Movebank ayant débuté pendant l'analyse des besoins. Cette API permet la récupération d'informations sur les études présentes sur Movebank ainsi que leurs données. Cette récupération est permise en fonction de droits d'accès définis au moment de la création de l'étude sur la plateforme. On constate ainsi 3 groupes utilisateurs : le data manager ou data owner (i.e. le créateur de l'étude), les collaborateurs et le public. L'accès aux informations sur l'étude et aux données peut être défini distinctement pour chaque groupe, que ce soit pour la consultation ou le téléchargement. L'accès peut également être restreint à une partie spécifique des données. Les données peuvent donc être récupérées via l'API uniquement si elles ont été définies comme téléchargeables. Des conditions d'utilisations s'appliquent aux données, le data owner peut définir ses propres conditions, sinon celles de Movebank s'appliquent. Lorsqu'un utilisateur, qu'il soit authentifié ou non, souhaite télécharger les données d'une étude, il doit accepter ces conditions pour l'étude en question s'il ne

l'avait jamais fait auparavant. Ce prérequis peut être désactivé par le créateur de l'étude. Ces conditions d'utilisation peuvent être renvoyées par l'API et il est possible de les accepter à travers l'API en ajoutant un cookie à la requête. L'API peut être requêtée via 3 routes différentes dont les principales caractéristiques sont résumées ci-dessous (Tableau 4).

Tableau 4. Principales caractéristiques des différentes routes de l'API de Movebank.

| | service/direct-read? | service/public/json? | service/json-auth? |
|---|--|--|---|
| Authentification requise | Oui | Non | Oui |
| Format données obtenues | CSV | JSON | |
| Possibilité d'obtenir la liste des entités requêtables | Oui | Non | |
| Possibilité d'obtenir la liste des attributs récupérables et utilisables en tant que filtres | Oui | Non | |
| Possibilité d'obtenir la liste des études et leur description | Oui, en fonction des droits d'accès Possibilité de filtrer sur les attributs suivants : i_am_owner, i_can_see_data et/ou there_are_data_which_i_cannot_see | Oui mais semble limité | |
| Possibilité d'obtenir la description des entités d'une étude | Oui | Oui mais semble limité | |
| Possibilité d'obtenir les données de localisations | Oui si elles sont publiques ou si l'utilisateur authentifié a les droits nécessaires | Oui si elles sont publiques | Oui si l'utilisateur authentifié a les droits d'accès nécessaires |
| Remarques | Route plus intéressante si on veut des informations complètes sur les études | Semble plus lente Destinée à la récupération des données de géolocalisation pour leur utilisation sur des outils externes comme Google Maps | |
| Mise en place | Script python exemple | Gestion des identifiants/certificats en PHP | |

Ce module, développé en python et directement intégré dans le back-end de AMCA, permet actuellement de récupérer la liste des études, la liste des animaux faisant partie d'une étude avec les déploiements de balises associés à ces animaux, ainsi que les données de localisation. La liste des études puis la liste des animaux s'affichent dans un tableau dont les lignes sont sélectionnables avant de passer à l'étape suivante (Figure 21). Lorsque c'est nécessaire, les conditions d'utilisations s'affichent dans une fenêtre pop-up sur AMCA et peuvent être acceptées afin de télécharger les données. Une fois téléchargées les données peuvent être directement utilisées dans AMCA. Des exemples des différents types de requêtes pouvant être effectuées avec cette API et des explications sur son fonctionnement peuvent être trouvés dans la documentation officielle (*Movebank REST API: Description of download interface to build calls to the Movebank database*, 2017).

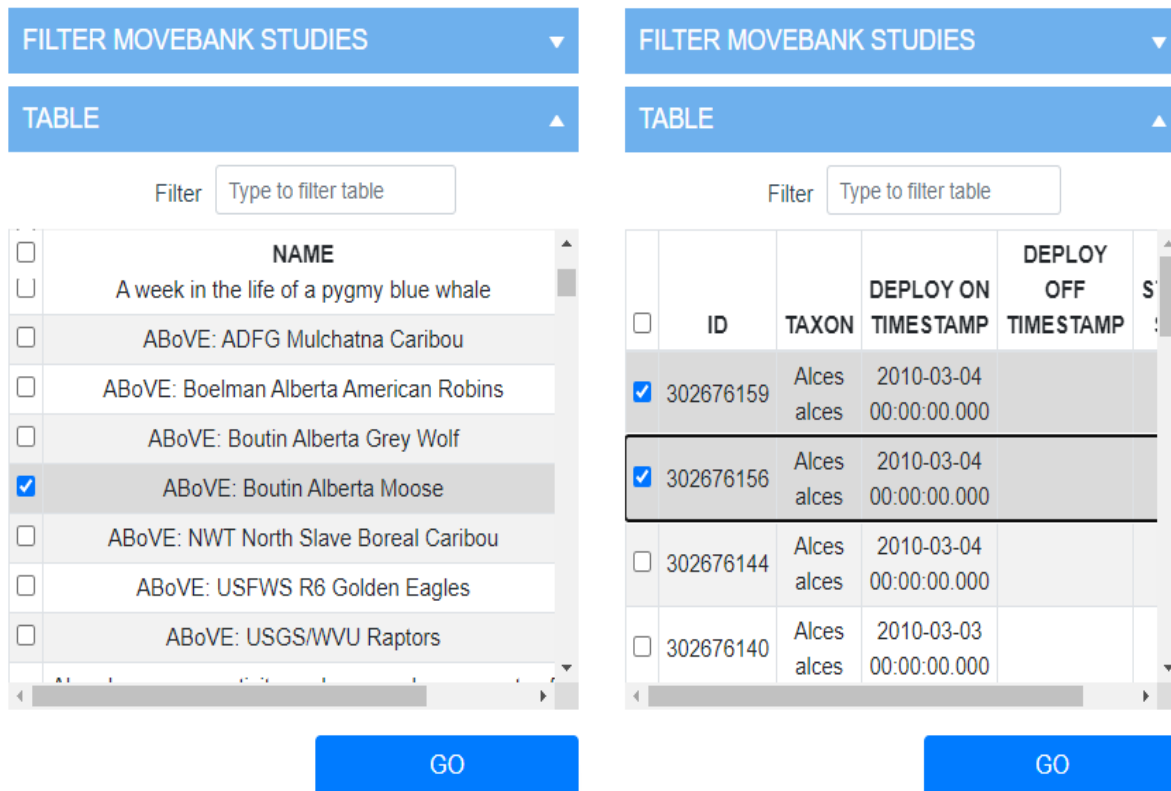


Figure 21. Tableaux de sélection d'études Movebank (à gauche) et d'animaux d'une étude Movebank (à droite).

L'intégration de ce module a rendu nécessaire une modification de AMCA de sorte à ce qu'elle permette l'import de plusieurs jeux de données.

2. La comparaison de jeux de données sur AMCA

Comme expliqué précédemment, AMCA ne permettait pas de visualiser plusieurs jeux de données à la fois, ce qui semble pourtant intéressant pour effectuer des comparaisons. Pour ajouter cette fonctionnalité, il a d'abord fallu améliorer la gestion des fichiers importés par les utilisateurs sur la plateforme qui, jusqu'alors, n'étaient jamais supprimés et s'empilaient donc sur le serveur. Désormais, lorsqu'un fichier est importé pour être traité et visualisé, une copie temporaire est créée sur le serveur puis supprimée lorsque l'utilisateur quitte AMCA. Cela est réalisé grâce à la définition d'un événement sur la page principale de l'application qui est déclenché lorsque l'utilisateur ferme son navigateur par exemple. Cet événement permet également d'afficher une fenêtre pop-up qui demande une confirmation du souhait de quitter AMCA, ce qui entrainera la perte des données. Il a ensuite fallu adapter le front-end pour qu'il conserve les différents jeux de données importés successivement. Cela est passé par une modification de la structure sous laquelle les données étaient stockées côté client. L'ensemble des données est désormais stocké dans un objet JSON qui permet de naviguer facilement dans ces données. Cependant, l'ajout de la possibilité d'importer plusieurs jeux de données ainsi que l'implémentation du module de requêtage de l'API de Movebank ont rendu l'interface de l'application inadaptée à ces nouvelles fonctionnalités.

En effet, l'ajout de ces fonctionnalités implique la présence potentielle d'éléments supplémentaires sur l'interface de l'application qui était initialement conçue pour présenter un seul jeu de données. Il est donc apparu nécessaire d'opérer une refonte de cette interface afin de maintenir une certaine qualité au niveau de l'expérience utilisateur. Ainsi, pour évaluer les modifications à apporter et ainsi être aiguillé pour cette mise-à-jour de l'interface, j'ai réalisé avec l'aide de Angélique Ries, web designer chez Natural Solutions, un test de Bastien et Scapin. Ce test vise à évaluer le respect des critères heuristiques de Bastien et Scapin (Bastien and Scapin, 1993). Ces critères permettent de catégoriser les défauts d'ergonomie, de juger de leur importance et de trouver des solutions pour les résoudre. Les aspects évalués sont l'adaptabilité, la charge de travail, le guidage, l'homogénéité, le contrôle explicite, la gestion des erreurs, la signifiante des codes et dénominations et la compatibilité. Chacun de ces critères est noté sur 10 et la note globale de l'application correspond à la moyenne de celles-ci. AMCA a ainsi obtenu la note de 4/10, les détails du test peuvent être trouvés en annexe 4.

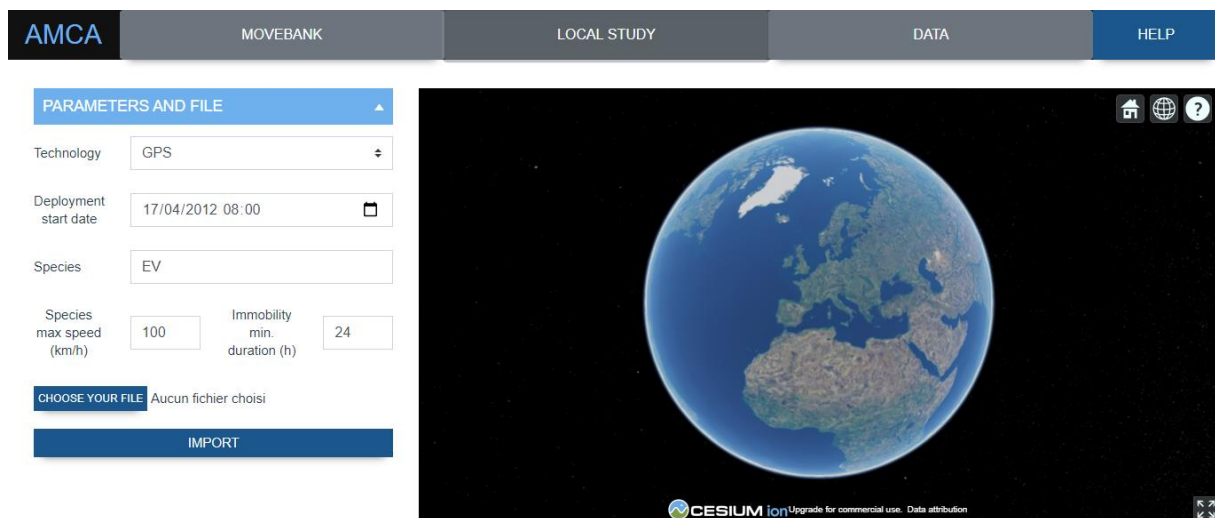


Figure 22. Onglet "Local study" de AMCA.

Les résultats de ce test ont servi à produire des maquettes, réalisées par Angélique Ries, qui ont fait l'objet de plusieurs échanges pour parvenir à une interface qui offre une meilleure expérience utilisateur. Il a ainsi été décidé de diviser cette application monopage en plusieurs onglets permettant de distinguer les différentes fonctionnalités : import d'une étude Movebank, import d'un fichier local et visualisation et traitement des données importées (Figure 22). Les deux premiers onglets cités ont conservé la visualisation mais qui est uniquement destinée à voir les données importées et ainsi s'assurer qu'il n'y a pas eu d'erreur dans l'import. Le 3^{ème} onglet regroupe l'ensemble des données importées et traitées et permet ainsi la comparaison de différents jeux de données. Un écran d'accueil a également été ajouté afin d'aider l'utilisateur dans la compréhension de la navigation.

What would you like to do ?

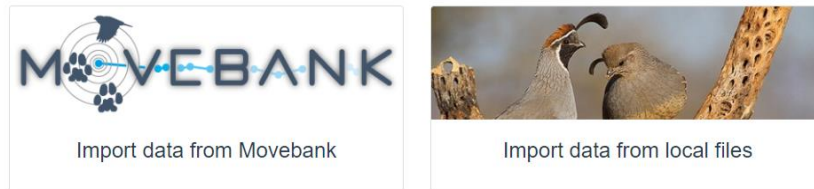


Figure 23. Page d'accueil de AMCA.

Les fonctionnalités prioritaires pour AMCA ayant été implémentées, le développement s'est donc ensuite axé sur le viewer.

3. Le viewer, plateforme de diffusion des données destinée au grand public

Le viewer est un outil principalement destiné à la diffusion de données de bio-logging auprès du grand public afin de mettre en avant le travail des naturalistes et des structures comme des parcs nationaux ou des associations de protection de la nature. Cette application doit donc proposer une expérience utilisateur satisfaisante à travers une interface simple. Devant la complexité de l'architecture choisie, le développement a été découpé de plusieurs étapes au cours desquelles j'ai reçu le soutien de Javier Blanco, développeur chez Natural Solutions. La 1^{ère} étape, dans l'ordre logique du flux de données, était donc la création de fichiers GeoJSON à partir de fichiers CSV et d'une base de données PostgreSQL. La bibliothèque GDAL (Geospatial Data Abstraction Library) possède un sous-ensemble, la bibliothèque OGR, qui permet notamment de nombreuses opérations de manipulation des formats courants de données vectorielles. Il est ainsi possible de créer un fichier GeoJSON à partir d'un fichier CSV ou encore en extrayant les données d'une base de données PostgreSQL. Cela se fait via l'intermédiaire d'un fichier VRT (Virtual Format) dans lequel on indique quels attributs on veut récupérer de la source de données (Figure 24).

```
<OGRVRTDataSource>
  <OGRVRTLayer name="other_test">
    <SrcDataSource>PG:dbname=testdb</SrcDataSource>
    <SrcLayer>test</SrcLayer>
    <GeometryField name="pg_geom_field_1" />
    <GeometryField name="vrt_geom_field_2" field="pg_geom_field_2">
      <GeometryType>wkbPolygon</GeometryType>
      <SRS>EPSG:4326</SRS>
      <ExtentXMin>-180</ExtentXMin>
      <ExtentYMin>-90</ExtentYMin>
      <ExtentXMax>180</ExtentXMax>
      <ExtentYMax>90</ExtentYMax>
    </GeometryField>
    <Field name="vrt_field_1" src="src_field_1" />
  </OGRVRTLayer>w
</OGRVRTDataSource>
```

Figure 24. Modèle de fichier VRT permettant d'extraire des données d'une base PostgreSQL.

A partir du fichier GeoJSON ainsi généré, l'outil Tippecanoe permet de créer des tuiles qui peuvent ensuite être envoyées sur un serveur. Ces étapes ont nécessité un certain temps pour être réalisées à cause de difficultés rencontrées dans l'utilisation de ces outils au sein de conteneurs. Un conteneur est une enveloppe virtuelle qui permet notamment d'isoler chaque partie d'un système avec l'ensemble des éléments dont elle a besoin pour fonctionner (environnements d'exécution, bibliothèques, fichiers...). Les conteneurs peuvent ainsi être déployés sur un serveur qui n'a alors besoin de rien d'autre pour que l'application fonctionne. Cependant leur configuration et leur mise en œuvre peut parfois présenter une certaine complexité. Nous sommes parvenus à utiliser les tuiles vectorielles à travers une application cartographique utilisant la bibliothèque Mapbox GL JS mais les difficultés rencontrées avec Docker et les conteneurs ne nous ont pas permis de disposer de suffisamment de temps pour développer une interface reprenant ce module cartographique.

V – Bilan et perspectives

Le développement des fonctionnalités évoquées tout au long de ce mémoire et ayant été menées plus ou moins proche de leurs termes, nous allons maintenant essayer de prendre un peu plus de recul sur le travail réalisé jusqu'à présent.

A) Prise de recul sur le travail effectué

1. Bilan sur les réalisations techniques

Revenons sur les différents éléments présentés tout au long de ce mémoire. Premièrement, le module de requête de l'API Movebank est celui qui semble le plus abouti. En effet, il est fonctionnel et permet bien de récupérer des informations sur les études présentes sur Movebank, des informations sur les animaux composant ces études mais surtout les données de mouvements de ces animaux. L'acceptation des conditions d'utilisation des données à travers l'API est également permise par le module. Cependant, il n'exploite qu'une seule des 3 routes proposées par l'API Movebank et cette route impose de gérer l'authentification dans AMCA. En effet, via cette route, il n'y a pas de service qui permette de s'authentifier dans Movebank sans avoir à indiquer ses identifiants dans AMCA, ce qui peut constituer un vrai frein pour certains utilisateurs qui ne souhaiteraient pas renseigner leurs identifiants dans AMCA. Une autre route permet ce service mais elle est limitée dans les informations qu'elle permet d'obtenir, il est par exemple impossible d'obtenir des informations détaillées à propos d'une étude en l'utilisant. Il serait intéressant d'étudier une éventuelle combinaison des différentes routes proposées par l'API qui permette de contourner ces problématiques.

Pour ce qui est de AMCA, l'outil n'est pas totalement fonctionnel. En effet, la partie de l'application qui permet la comparaison de différents jeux de données n'a pas pu être totalement développée par manque de temps. Au-delà de ce problème majeur évidemment, quelques bugs liés aux nouvelles fonctionnalités ajoutées n'ont pas pu être réparés non plus. Notamment, le player de Cesium, qui permet une visualisation temporelle des données, a été laissé de côté pendant le développement et ne fonctionne désormais plus. Dans la globalité, AMCA a tout de même été améliorée notamment du point de vue de l'expérience utilisateur grâce au test de Bastien et Scapin

réalisé. Cependant, seuls les aspects les plus critiques (c'est-à-dire les moins bien notés) ont été traités et de nombreuses améliorations peuvent encore être apportées. Par exemple, les différents jeux de données importés peuvent devenir difficiles à distinguer les uns des autres visuellement s'ils sont trop nombreux. Autant de choses qui font qu'il est difficile de considérer l'application comme actuellement prête à l'emploi.

Enfin, le viewer représente l'élément le moins abouti. En effet, les difficultés techniques rencontrées, ajoutées au temps très faible disponible pour son développement n'ont pas permis de le mener à terme. Cependant, nous pouvons certainement affirmer que la partie la plus compliquée d'un point de vue technique a été réalisée ce qui permet alors d'envisager une potentielle suite du développement plus sereinement.

2. Bilan global sur le déroulement du stage

Bien que le bilan des réalisations techniques soit mitigé, notamment à cause du temps très limité qui a pu être consacré au développement, le bilan global est lui bien plus positif. Même si elle fût relativement courte, cette phase de développement m'a tout de même permis de progresser techniquement et notamment en développement web. J'ai pu profiter de l'expérience et de la compétence de l'équipe de développeurs de l'entreprise.

La brièveté de la phase de développement s'explique en grande partie par une phase d'analyse qui s'est prolongée pour différentes raisons. Or cette phase d'analyse fût particulièrement intéressante puisqu'elle m'a permis de plonger pleinement dans le milieu du bio-logging et d'en saisir les subtilités.

Cette période m'a également permis de travailler sur des tâches annexes et donc de découvrir encore d'autres domaines. Natural Solutions a ainsi pris le temps de me former sur des problématiques telles que la création de procédures stockées et leur déploiement sur plusieurs bases de données faisant partie d'un système de grande échelle.

Ce projet fût également l'opportunité de progresser en gestion de projet, qui est certainement le point sur lequel j'ai le plus besoin de m'améliorer, bien que ce stage se soit déroulé dans un contexte très particulier. L'équipe de Natural Solutions a permis que ce stage soit une bonne expérience malgré ce contexte et a constitué un environnement de qualité.

B) Perspectives d'évolution

L'ensemble applicatif présenté dans ce mémoire est déjà intéressant en tant que réponse aux problématiques rencontrées dans le milieu du bio-logging cependant il semble nécessaire de considérer l'apport d'améliorations et de fonctionnalités supplémentaires pour qu'il devienne réellement pertinent et puisse ainsi s'imposer comme un élément fort dans son milieu.

1. Amélioration des fonctionnalités existantes

L'interaction avec Movebank semble devenue la clé pour le succès d'un outil dans le monde du bio-logging tant cette plateforme s'est imposée dans la communauté. Dans cette optique il serait intéressant de développer les interactions entre Movebank et les outils présentés plus tôt.

a. Echanger avec Movebank pour développer les interactions avec la plateforme

L'API de Movebank permettant déjà de récupérer les données stockées sur la plateforme depuis une autre application, il serait intéressant de pouvoir également envoyer des données à Movebank depuis une application externe. Des échanges ont été lancés avec la plateforme pour discuter de cette éventualité mais n'ont pas été menés jusqu'à leur terme. Ces contacts ont tout de même permis d'entrevoir la possibilité d'une telle interaction devienne possible. En effet, en établissant une relation de confiance avec la plateforme, il semble envisageable de mettre en place une transmission de données à la façon du service Live Feeds qui permet aux utilisateurs de balises de recevoir leurs données directement sur la plateforme grâce à une transmission directe de fabricant de balise à Movebank. Cela permettrait par exemple d'envoyer ses données traitées à l'aide de AMCA sur Movebank.

b. Intégration d'algorithmes de traitement et d'analyse de données supplémentaires

AMCA possède l'avantage de sa simplicité d'utilisation cependant cette simplicité est aussi le signe d'un nombre de fonctionnalités restreint. Comme nous en avons déjà parlé dans ce mémoire, il existe de nombreuses bibliothèques R libres qui sont réellement intéressantes pour le traitement et l'analyse de données de mouvements d'animaux. Se priver de ce genre de ressources serait assez dommageable notamment dans l'optique de rendre l'application attrayante pour un public tel que les chercheurs qui recherchent des outils relativement pointus.

2. Apport de nouvelles fonctionnalités

a. L'intelligence artificielle pour étudier le comportement animal

Les mouvements des animaux reflètent leur comportement et avec la quantité de données dont nous disposons aujourd'hui il semble pertinent d'envisager leur utilisation pour mettre en œuvre des intelligences artificielles qui nous aideraient dans la compréhension du comportement des animaux et de leurs mouvements. Cette perspective, déjà citée dans (Jambon, 2019) mais n'ayant pas pu être mise en place, constitue une réelle voie d'évolution pour AMCA.

b. GeoNature, un outil qui pourrait aider le milieu du bio-logging

GeoNature est un outil composé de multiples modules destiné à la saisie, la gestion, la synthèse et la diffusion d'observations de faune et flore. Son association à AMCA pourrait être particulièrement intéressante et en particulier dans le cadre de la mise en place d'un module d'intelligence artificielle pour l'étude comportementale des animaux. En effet, on sait que l'annotation des données constitue un frein majeur pour l'étape d'apprentissage notamment car elle est très chronophage. Une utilisation couplée de GeoNature et AMCA permettrait de contourner cet obstacle en utilisant les observations réalisées dans les différentes applications GeoNature.

GeoNature possède des modules qui pourraient également être intéressants à combiner avec AMCA. Par exemple l'application TaxHub, qui permet de gérer des taxons en se basant sur le référentiel TAXREF du Muséum national d'Histoire naturelle (MNHN) [référence]

, pourrait être intégrée à AMCA afin de prendre en compte des informations supplémentaires sur les taxons pour le traitement des données. Une utilisation attrayante d'un tel système serait par exemple la prise en compte du type de l'espèce (i.e. aérien, aquatique ou terrestre) pour détecter des points aberrants. En couplant cela avec l'utilisation de données d'occupation du sol, comme CORINE Land Cover par exemple, nous serions alors capables de détecter une localisation suspecte plaçant un mammifère terrestre au milieu de l'océan.

De nombreuses perspectives d'évolution similaires sont imaginables et cela démontre tout le potentiel que possèdent les outils de bio-logging pour l'accompagnement de la communauté scientifique dans la résolution de diverses problématiques et la réponse à des enjeux majeurs.

Conclusion

Dans un monde qui contient d'importantes mutations, l'étude de notre environnement paraît plus importante que jamais. Les changements climatiques amplifiés par la forte croissance démographique bouleversent les écosystèmes qui ont du mal à faire face à ces perturbations. La réduction de la biodiversité observée actuellement est ainsi sans précédent. Pour endiguer ce phénomène, la communauté scientifique s'efforce de collecter toujours plus d'informations puisque l'élaboration de solutions passera forcément par le développement de nos connaissances.

Les progrès technologiques réalisés au cours des dernières décennies ont permis l'explosion du nombre de capteurs utilisés sur des animaux et donc de la quantité de données récoltées. Le spectre des espèces que nous sommes en mesure d'étudier s'élargit au fil du temps tout comme le spectre des types de données que nous sommes en mesure de récolter. Cependant, au milieu de l'enthousiasme qui a petit à petit gagné le monde du bio-logging, des problématiques difficilement occultables ont fait leur apparition.

Les volumes de données phénoménaux générés ont ainsi pris de court le milieu du bio-logging qui ne s'est pas organisé et rencontre désormais les plus grandes difficultés dans l'établissement de standards internationaux pour ces données. L'impact de l'absence de normes se fait d'autant plus ressentir que, tout comme la quantité de données récoltées, le nombre de fabricants de balises a également connu une nette augmentation. Le matériel devenant alors de plus en plus accessible, l'utilisation de balises n'est plus l'apanage des chercheurs.

Les profils d'utilisateurs de balises se diversifiant, les outils destinés à les accompagner florissent renforçant ainsi le problème de standardisation puis d'interopérabilité. Nous avons tenté de démontrer, à travers ce travail, qu'il est possible de répondre aux besoins, aussi diversifiés soient-ils, d'un milieu sans construire un énième nouvel outil qui viendrait étoffer une collection déjà bien garnie. Via l'association de Movebank, AMCA, un module de visualisation et un module de base de données, nous avons proposé un ensemble applicatif qui apporte une piste de réponse aux problématiques rencontrées par la communauté du bio-logging. Cette démarche vise également à promouvoir l'ouverture des données et l'utilisation de solutions libres.

Bibliographie

- Andreassen, H. P. and Ims, R. A. (1998) 'The effects of experimental habitat destruction and patch isolation on space use and fitness parameters in female root vole *Microtus oeconomus*', *Journal of Animal Ecology*, 67(6), pp. 941–952. doi: 10.1046/j.1365-2656.1998.6760941.x.
- Banks, S. C. and Lindenmayer, D. B. (2014) 'Inbreeding avoidance, patch isolation and matrix permeability influence dispersal and settlement choices by male agile antechinus in a fragmented landscape', *Journal of Animal Ecology*, 83(2), pp. 515–524. doi: 10.1111/1365-2656.12128.
- Bastien, J. M. C. and Scapin, D. L. (1993) 'Critères Ergonomiques pour l'Évaluation d'Interfaces Utilisateurs', *INRIA Technical report*, 156.
- Bernard, C. and Dagognet, F. (2001) *Introduction à l'étude de la médecine expérimentale*. Paris: Flammarion.
- Berners-Lee, T. (2010) *Linked Data, W3C*. Available at: <https://www.w3.org/DesignIssues/LinkedData.html> (Accessed: 3 September 2020).
- Beyer, H. L. *et al.* (2013) 'The effectiveness of Bayesian state-space models for estimating behavioural states from movement paths', *Methods in Ecology and Evolution*. Edited by R. Freckleton, 4(5), pp. 433–441. doi: 10.1111/2041-210X.12026.
- Biodiv'Jamaica* (2016). Available at: <http://jamaica.geonature.fr/>.
- Blackburn, T. M. and Gaston, K. J. (1994) 'The Distribution of Body Sizes of the World's Bird Species', *Oikos*, 70(1), p. 127. doi: 10.2307/3545707.
- Boyd, I. L., Kato, A. and Ropert-Coudert, Y. (2004) 'Bio-logging science: sensing beyond the boundaries', *Memoirs of the National Institute of Polar Research, Special Issue*, 58.
- Campbell, H. A. *et al.* (2016) 'A plea for standards in reporting data collected by animal-borne electronic devices', *Animal Biotelemetry*, 4(1), p. 1. doi: 10.1186/s40317-015-0096-x.
- Castle, K. T. *et al.* (2015) 'Using sutures to attach miniature tracking tags to small bats for multimonh movement and behavioral studies', *Ecology and Evolution*, 5(14), pp. 2980–2989. doi: 10.1002/ece3.1584.
- Chen, I.-C. *et al.* (2011) 'Rapid Range Shifts of Species Associated with High Levels of Climate Warming', *Science*, 333(6045), pp. 1024–1026. doi: 10.1126/science.1206432.
- Conroy, J. W. H. and Twelves, E. L. (1972) 'Diving depths of the Gentoo penguin (*Pygoscelis papua*) and le-eyed shag (*Phalacrocorax atriceps*) from the South Orkney Islands', *British Antarctic Survey Bulletin*, 30, pp. 106–108. Available at: <http://nora.nerc.ac.uk/id/eprint/526178/> (Accessed: 22 August 2020).
- Cooke, S. J. *et al.* (2017) 'Troubling issues at the frontier of animal tracking for conservation and management', *Conservation Biology*, 31(5), pp. 1205–1207. doi: 10.1111/cobi.12895.
- Croxall, J. P. *et al.* (1991) 'Diving patterns and performance in the Antarctic blue-eyed shag *Phalacrocorax atriceps*', *Journal of Zoology*, 225(2), pp. 177–199. doi: 10.1111/j.1469-7998.1991.tb03810.x.

- Cunningham, K. A. *et al.* (2014) ‘Auditory detection of ultrasonic coded transmitters by seals and sea lions’, *The Journal of the Acoustical Society of America*, 135(4), pp. 1978–1985. doi: 10.1121/1.4868371.
- Curatolo, J. A. 1986. Evaluation of a satellite telemetry system for monitoring movements of caribou. *Rangifer*, Spec. Issue No. 1:73-80.
- Davis, M. B. (2001) ‘Range Shifts and Adaptive Responses to Quaternary Climate Change’, *Science*, 292(5517), pp. 673–679. doi: 10.1126/science.292.5517.673.
- Delong, R. L., Stewart, B. S. and Hill, R. D. (1992) ‘DOCUMENTING MIGRATIONS OF NORTHERN ELEPHANT SEALS USING DAY LENGTH’, *Marine Mammal Science*, 8(2), pp. 155–159. doi: 10.1111/j.1748-7692.1992.tb00375.x.
- Douglas, D. C. *et al.* (2012) ‘Moderating Argos location errors in animal tracking data’, *Methods in Ecology and Evolution*, 3(6), pp. 999–1007. doi: 10.1111/j.2041-210X.2012.00245.x.
- EcoTeka* (2020). Available at: <https://ecoteka.org/>.
- Fahrig, L. (2007) ‘Non-optimal animal movement in human-altered landscapes’, *Functional Ecology*, 21(6), pp. 1003–1015. doi: 10.1111/j.1365-2435.2007.01326.x.
- Fancy, S. G. *et al.* (1988) ‘Satellite Telemetry: A New Tool for Wildlife Research and Management’,.
- Fraser, K. C. *et al.* (2018) ‘Tracking the Conservation Promise of Movement Ecology’, *Frontiers in Ecology and Evolution*, 6. doi: 10.3389/fevo.2018.00150.
- Gardezi, T. and da Silva, J. (1999) ‘Diversity in Relation to Body Size in Mammals: A Comparative Study’, *The American Naturalist*, 153(1), pp. 110–123. doi: 10.1086/303150.
- Gargominy, O. *et al.* (2019) ‘TAXREF v13, référentiel taxonomique pour la France : méthodologie, mise en œuvre et diffusion.’, *Muséum national d’Histoire naturelle, Paris*, Rapport Patrinat, p. 63.
- Gebauer, K. *et al.* (2013) ‘Matrix Matters: Differences of Grand Skink Metapopulation Parameters in Native Tussock Grasslands and Exotic Pasture Grasslands’, *PLOS ONE*, 8(10), p. e76076. doi: 10.1371/journal.pone.0076076.
- Hickling, R. *et al.* (2006) ‘The distributions of a wide range of taxonomic groups are expanding polewards’, *Global Change Biology*, 12(3), pp. 450–455. doi: 10.1111/j.1365-2486.2006.01116.x.
- International Bio-Logging Society* (2020). Available at: <https://www.bio-logging.net/>.
- Jacoby, D. M. P. and Freeman, R. (2016) ‘Emerging Network-Based Tools in Movement Ecology’, *Trends in Ecology & Evolution*, 31(4), pp. 301–314. doi: 10.1016/j.tree.2016.01.011.
- Jambon, A. (2019) *Animal Movement Cleaner Application : Un outil de pré-traitement des données de bio-logging*. Bordeaux Sciences Agro.
- Jennings, J. G., and W. F. Gandy. 1980. Tracking pelagic dolphins by satellite. Pages 753-755 in C. J. Amlaner, Jr., and D. W. MacDonald, eds. A handbook on biotelemetry and radio tracking. Pergamon Press, Oxford, England. 804 pp

- Joo, R. *et al.* (2020) ‘Navigating through the r packages for movement’, *Journal of Animal Ecology*, 89(1), pp. 248–267. doi: 10.1111/1365-2656.13116.
- Kays, R. *et al.* (2015) ‘Terrestrial animal tracking as an eye on life and planet’, *Science*, 348(6240), pp. aaa2478–aaa2478. doi: 10.1126/science.aaa2478.
- Kolz, A. L., J. W. Lentfer, and H. G. Fallek. 1980. Satellite radio tracking of polar bears instrumented in Alaska. Pages 743–752 in C. J. Amlaner, Jr., and D. W. MacDonald, eds. A handbook on biotelemetry and radio tracking. Pergamon Press, Oxford. 804 pp.
- Kooyman, G. L. (1965) ‘Techniques used in measuring diving capacities of Weddell Seals’, *Polar Record*, 12(79), pp. 391–394. doi: 10.1017/S003224740005484X.
- Kooyman, G. L., Gentry, R. L. and Urguhart, D. L. (1976) ‘Northern fur seal diving behavior: a new approach to its study’, *Science*, 193(4251), pp. 411–412. doi: 10.1126/science.935876.
- Krogh, A. (1929) ‘The Progress of Physiology’, *Science*, 70(1809), pp. 200–204. doi: 10.1126/science.70.1809.200.
- Larsen, T., Jonkel, C. and Vibe, C. (1983) ‘Satellite Radio-Tracking of Polar Bears between Svalbard and Greenland’, *Bears: Their Biology and Management*, 5, pp. 230–237. doi: 10.2307/3872542.
- Lister, N.-M., Brocki, M. and Ament, R. (2015) ‘Integrated adaptive design for wildlife movement under climate change’, *Frontiers in Ecology and the Environment*, 13(9), pp. 493–502. doi: 10.1890/150080.
- Lundberg, J. and Moberg, F. (2003) ‘Mobile Link Organisms and Ecosystem Functioning: Implications for Ecosystem Resilience and Management’, *Ecosystems*, 6(1), pp. 0087–0098. doi: 10.1007/s10021-002-0150-4.
- Max Planck Institute of Animal Behavior (2019) ‘Movebank Attribute Dictionary’. Natural Environment Research Council. Available at: <http://vocab.nerc.ac.uk/collection/MVB/current/> (Accessed: 3 September 2020).
- Max Planck Institute of Animal Behavior (2020a) *About Movebank, Movebank*. Available at: <https://www.movebank.org/cms/movebank-content/about-movebank> (Accessed: 3 September 2020).
- Max Planck Institute of Animal Behavior (2020b) *The Env-DATA System, Movebank*. Available at: <https://www.movebank.org/cms/movebank-content/env-data> (Accessed: 3 September 2020).
- Meeuwig, J. J., Harcourt, R. G. and Whoriskey, F. G. (2015) ‘When Science Places Threatened Species at Risk: When science places threatened species at risk’, *Conservation Letters*, 8(3), pp. 151–152. doi: 10.1111/conl.12185.
- Movebank REST API: Description of download interface to build calls to the Movebank database* (2017). Available at: <https://github.com/movebank/movebank-api-doc/blob/master/movebank-api.md> (Accessed: 3 September 2020).
- Muff, S., Signer, J. and Fieberg, J. (2020) ‘Accounting for individual-specific variation in habitat-selection studies: Efficient estimation of mixed-effects models using Bayesian or frequentist computation’, *Journal of Animal Ecology*. Edited by E. V. Wal, 89(1), pp. 80–92. doi: 10.1111/1365-2656.13087.

Naito, Y. *et al.* (1990) 'Estimation of Migration Route of the Loggerhead Turtle *Caretta caretta* around the Nesting Ground', *Nippon Suisan Gakkaishi*, 56(2), pp. 255–262. doi: 10.2331/suisan.56.255.

Natural Solutions (2020). Available at: <https://www.natural-solutions.eu/>.

Nuijten, R. J. M. *et al.* (2020) 'Less is more: On-board lossy compression of accelerometer data increases biologging capacity', *Journal of Animal Ecology*. Edited by L. Börger, 89(1), pp. 237–247. doi: 10.1111/1365-2656.13164.

Observations de biodiversité (faune-flore) du Parc national des Écrins (2020). Available at: <https://www.data.gouv.fr/fr/datasets/observations-de-biodiversite-faune-flore-du-parc-national-des-ecrins/> (Accessed: 3 September 2020).

Parc National des Écrins (2013) *Bouquetins du Parc National des Écrins*. Available at: <http://bouquetins.ecrins-parcnational.fr> (Accessed: 3 September 2020).

Patin, R. *et al.* (2020) 'Identifying stationary phases in multivariate time series for highlighting behavioural modes and home range settlements', *Journal of Animal Ecology*. Edited by G. Street, 89(1), pp. 44–56. doi: 10.1111/1365-2656.13105.

Priede, I. G. 1982. Tracking of basking sharks. Page 73 in C. Cote, R. Taylor, and E. Gilbert, eds. Nimbus-6 random access measurement system applications experiments. NASA SP-457, Washington, D.C. 99 pp.

References (2020) *Reneco for wildlife preservation*. Available at: <http://www.reneco.net/references/>.

Roport-Coudert, Y. *et al.* (2009) 'Diving into the world of biologging', *Endangered Species Research*, 10, pp. 21–27. doi: 10.3354/esr00188.

Rutz, C. and Hays, G. C. (2009) 'New frontiers in biologging science', *Biology Letters*, 5(3), pp. 289–292. doi: 10.1098/rsbl.2009.0089.

Schweinsburg, R. E. and Lee, L. J. (1982) 'Movement of Four Satellite-Monitored Polar Bears in Lancaster Sound, Northwest Territories', *ARCTIC*, 35(4), pp. 504–511. doi: 10.14430/arctic2357.

Stoneburner, D. C. 1982. Sea turtle (*Caretta caretta*) migration and movements in the South Atlantic Ocean. Page 74 in C. Cote, R. Taylor, and E. Gilbert, eds. Nimbus-6 random access measurement system applications experiments. NASA SP-457, Washington, D.C. 99 pp.

Stonehouse, B. (1967) 'Feeding Behaviour and Dmng Rhythms of Some New Zealand Shags, Phalacrocoracidae', *Ibis*, 109(4), pp. 600–605. doi: 10.1111/j.1474-919X.1967.tb00028.x.

Strikwerda, T. E., H. D. Black, N. Levanon, and P. W. Howey. 1985. The bird-borne transmitter. Johns Hopkins APL Tech. Dig. 6:60-67.

Strikwerda, T. E., M. R. Fuller, W. S. Seegar, P. W. Howey, and H. D. Black. 1986. Bird-borne satellite transmitter and location program. Johns Hopkins APL Tech. Digest 7:203-208.

Subra, J.-P. (2019) *Scrum: une méthode agile pour vos projets*.

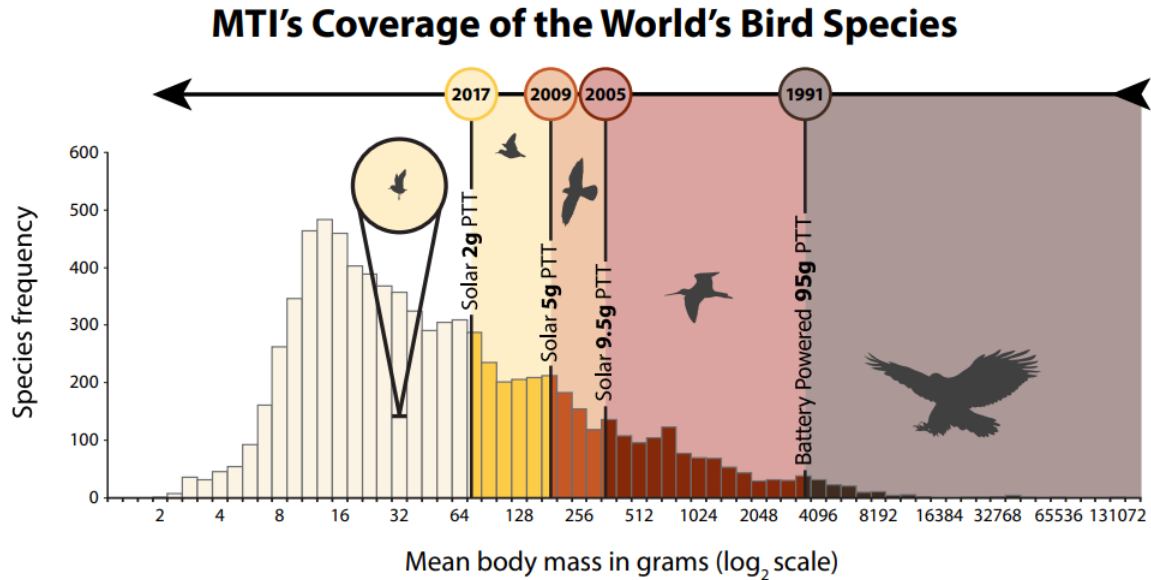
Taylor, M. K. 1982. The distribution and abundance of polar bears (*Ursus maritimus*) in the Beaufort and Chukchi seas. Ph.D. thesis, University of Minnesota, Minneapolis. 353 pp.

- Tilman, D. *et al.* (2011) 'Global food demand and the sustainable intensification of agriculture', *Proceedings of the National Academy of Sciences*, 108(50), pp. 20260–20264.
- Timko, R. E., and A. L. Kolz. 1982. Satellite sea turtle tracking. *Mar. Fish. Rev.* 44:19-24.
- Tucker, M. A. *et al.* (2018) 'Moving in the Anthropocene: Global reductions in terrestrial mammalian movements', *Science*, 359(6374), pp. 466–469. doi: 10.1126/science.aam9712.
- United Nations, 2015. World population prospects: The 2015 revision, key findings and advance tables. *Working Paper No. ESA/P/WP. 241*.
- Wilmers, C. C. *et al.* (2015) 'The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology', *Ecology*, 96(7), pp. 1741–1753. doi: 10.1890/14-1401.1.
- Wilson, R. *et al.* (1992) 'Estimation of location: global coverage using light intensity', *Wildlife Telemetry: remote monitoring and tracking of animals*, pp. 131–134.
- Woods, E. G., and A. Kemmerer. 1982. Satellite-linked porpoise tracking system. Pages 70-71 in C. Cote, R. Taylor, and E. Gilbert, eds. *Nimbus-6 random access measurement system applications experiments*. NASA SP-457, Washington, D.C. 99 pp

ANNEXES

Annexe 1ii
Annexe 2ii
Annexe 3iii
Annexe 4iv

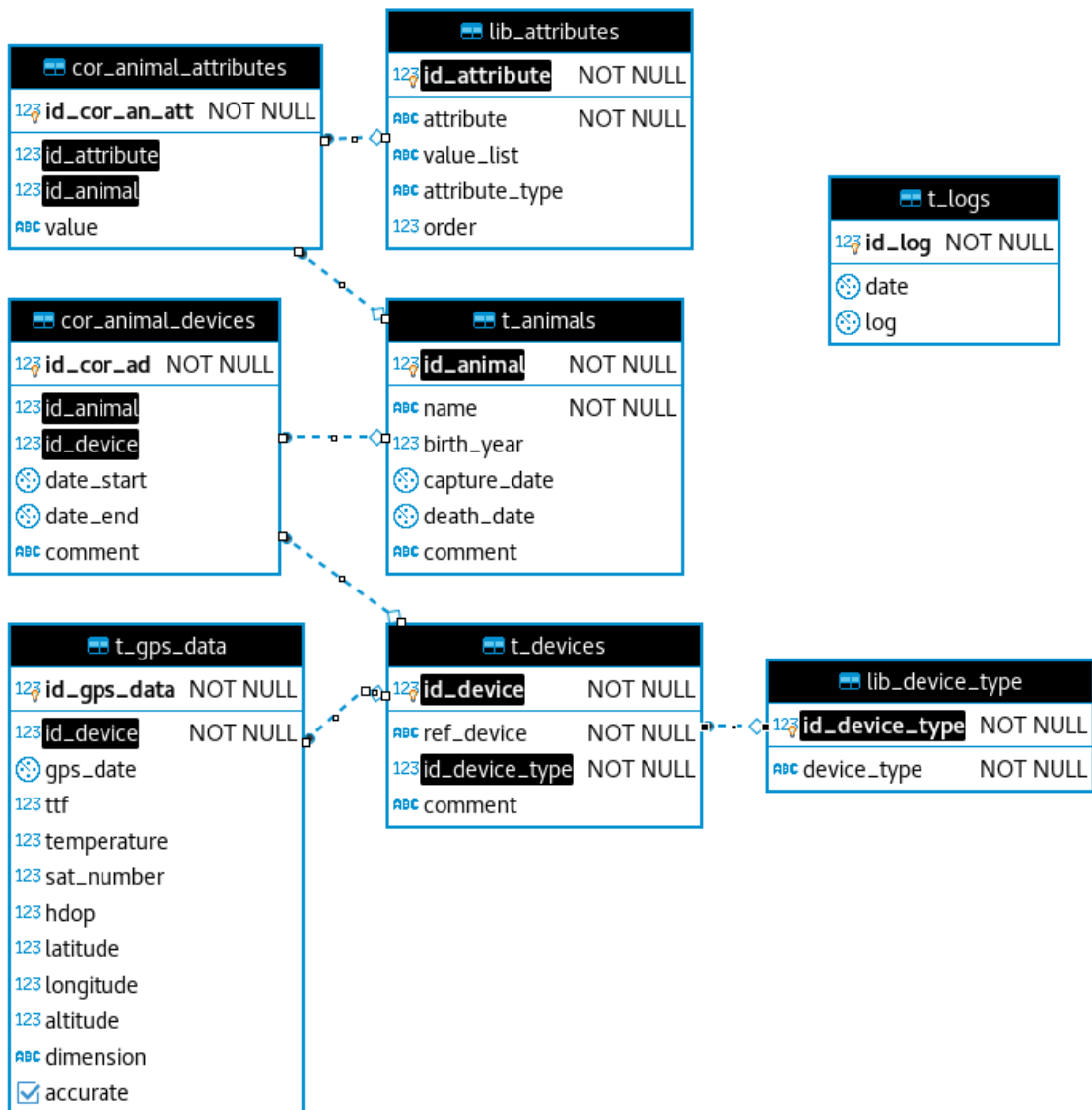
Annexe 1 : Distribution et évolution de la proportion des espèces d’oiseaux pouvant être suivies avec des balises Argos (35%). 1. Indique la limite de masse pour les balises GPS. 2. Indique la limite pour tout suivi en bio-logging. Source : (MTIs Coverage of the World’s Bird Species 2018, 2018)



Annexe 2 : Liste non exhaustive de fabricants de balises

| Entreprise | Année de création |
|---------------------------------|-------------------|
| GeoTrak | 1999 |
| Lotek | 1984 |
| Microwave Telemetry Inc | 1991 |
| Cellular Tracking Technologies | 2007 |
| Druid Technology | ? |
| Ecotone Telemetry | 1993 |
| e-obs GmbH | 2001 |
| Followit Wildlife | 2000 |
| Intricode | ? |
| KoEco | ? |
| MoveTech Telemetry | ? |
| North Star Science & Technology | ? |
| Ornitela | ? |
| Advanced Telemetry Systems | 1981 |
| CLS Telemetry | ? |
| Holohil | ? |
| Iridium | ? |

Annexe 3 : Modèle Conceptuel de Données de FollowDem



Méthode de recette Bastien & Scapin

Projet : AMCA

Note finale : 4/10

ADAPTABILITE : 2/10

CAPACITE A REAGIR SELON LE CONTEXTE ET LES BESOINS/PREFERENCES DU USER

Flexibilité : moyen de personnaliser l'interface par rapport à la stratégie et/ou habitudes de travail et des exigences des tâches. capAussi : les différentes possibilités pour les user d'atteindre l'objectif donné.

- Menu "parameters and file" pliable/dépliable. Pas d'autre signe de **flexibilité**

Prise en compte de l'expérience de l'utilisateur : senior/junior, expert/débutant

- **Pas de prise en compte de l'expérience de l'utilisateur // Idée : Guider l'utilisateur sur le travail qu'il souhaite réaliser (bouton au naming clair + texte indicatif etc.)**

CHARGE DE TRAVAIL : 6/10

L'ENSEMBLE DES ELEMENTS DE L'INTERFACE QUI ONT UN ROLE DANS LA REDUCTION DE LA CHARGE PERCEPTIVE OU AMNESIQUE DES UTILISATEURS

Brièveté : réduire les suites d'actions nécessaires à l'atteinte d'un but ou l'accomplissement d'une tâche. Limiter le travail de lecture et d'entrée de données.

- Concision : limiter les entrées et sorties
- Actions minimales : limiter autant que possible les étapes pour atteindre le but

- **RAS**

Densité informationnelle : charge de travail perceptive, affichage 7 +/- 2 informations

- **Écran d'accueil bien au niveau charge informationnelle // En revanche lors de l'import d'un document ça devient limite**

- **Note : Lors de l'import de plusieurs jeux de données, ne pas les afficher les uns à la suite des autres, favoriser un affichage en onglet (par exemple)**

GUIDAGE : 2/10

CONSEILLER, ORIENTER, INFORMER, ET CONDUIRE L'UTILISATEUR LORS DE SES INTERACTIONS

incitation : pages individuelles, liens, mécanismes de recherche, navigation

- **Actuellement tout se situe sur une page**
- **Idée : avoir plus de possibilités de filtrer dans la partie "Search Studies" (ex : filtrer par espèce, filtrer par date de l'étude etc...)**

Groupement/distinctions entre items (items, localisation, format) : localisation = positionnement sur la page des informations les unes par rapport aux autres => appartenance ou non à une même classe (fieldsets, groupement de boutons...) Format = caractéristiques graphiques => appartenance ou non à une même classe

- **Il y a logique de positionnement des éléments (carte à droite // données à gauche)**
- **Peu de caractéristiques graphiques concernant les boutons/ménu etc...**
- **Différenciation des points présents sur la carte (légende, point sélectionné, point valide, point incorrecte, point qui vient d'un autre jeu de données)**
- **Introduire le bouton "Download template" à l'utilisateur pour qu'il ait l'information que son document .csv doit prendre forme bien précise**

Feed-back immédiat : réponses de la machine suite aux actions utilisateurs (délais brefs, appropriés et homogènes)

- **Pas de feedback (lié au chargement) quand on clique sur "Search Studies" + pas de feedback quand on tape dans la barre de recherche (délais de 5 secondes)**
- **Pas de retours si le fichier a bien été importé**
- **Pas d'indication si un mauvais format de données a été importé**
- **Pas de message informatif pour la suppression de données**
- **La carte n'affiche rien à l'import des données (il faudrait pouvoir tout voir puis filtrer sur les données que l'utilisateur souhaite afficher)**
- **Avoir un retour sur l'import du jeu de données + pop-up indiquant le résultat du travail de l'algo**

Lisibilité : luminance caractères, contraste avec le fond, dimensions des lettres, espacement, longueur des lignes

- **Taille des boutons trop importante : Même largeur que les menus - Confusant**

HOMOGENEITE/COHERENCE : 4/10

CHOIX DE CONCEPTION DE L'INTERFACE CONSERVES POUR DES CONTEXTES IDENTIQUES ET DIFFERENT POUR DES CONTEXTES DIFFERENTS.

- **Les menus dépliant sont homogènes**
- **Les boutons sont très hétérogènes**

CONTROLE EXPLICITE : 5/10

PRISE EN COMPTE PAR LE SYSTEME DES ACTIONS EXPLICITES DES UTILISATEURS ET LE CONTROLE QU'ONT LES UTILISATEURS SUR LE TRAITEMENT DE LEURS ACTIONS

Actions explicites : le système doit exécuter seulement les opérations demandées par l'utilisateur au moment où il les demande

- **RAS**

Contrôle utilisateur : pouvoir contrôler le déroulement (interrompre, reprendre) des traitements informatiques en cours. Anticipation des actions du user, options appropriées pour chaque cas.

- **Idee : Pour les données supprimées, mettre en place une légende "Deleted data" qui permettra d'afficher ces données + un bouton "Undelete data" Les faire "réapparaître" (ne pas supprimer les données dans les documents .csv, uniquement les mettre sous le label associé à la légende "Deleted data")**

GESTION DES ERREURS : 1/10

CONCERNE LES MOYENS D'EVITER OU REDUIRE LES ERREURS OU LES CORRIGER SI ELLES SURVIENNENT, ÇA CONCERNE LES SAISIES INCORRECTES, FORMATS INADEQUATS, SAISIES DE COMMANDES ERRONEES

Protection contre les erreurs : détecter, prévenir les erreurs d'entrées de données ou commandes, ou les actions à conséquences néfastes

- **L'utilisateur est mal averti du format de fichier qu'il doit importer + que ce fichier doit respecter une nomenclature précise**

- Si l'utilisateur importe un fichier qui n'est pas exploitable, il n'est pas averti
- Pas de message lors de la suppression d'un point
- Si l'utilisateur veut recharger sa page, il est averti que toutes ses modifications ne seront pas enregistrées
- Dans les paramètres, il n'y a aucun contrôle d'erreur des données entrées par l'utilisateur

Qualité des messages d'erreurs : pertinence, facilité de lecture, exactitude de l'information sur la nature des erreurs (syntaxes, format...) et actions correctrices à entreprendre. Fournir des messages d'erreur orientés-tâches. Utiliser des termes spécifiques. Vocabulaire neutre, non personnalisé, non réprobateur, pas d'humour.

- RAS

Corrections des erreurs : moyens mis à disposition pour permettre la correction des erreurs.

- **Note : Toujours avoir la possibilité de revenir en arrière/d'annuler l'action qui vient d'être réalisée**

SIGNIFIANCE DES CODES ET DENOMINATIONS : 5/10

ADEQUATION ENTRE L'OBJET OU L'INFORMATION AFFICHEE OU ENTREE, ET SON REFERENT

Vérifier la symbolique des icônes et les termes utilisés avec les utilisateurs, Les titres doivent véhiculer ce qu'ils représentent, et être distincts. Rendre les règles d'abréviation explicites.



- Sur la carte : ce bouton n'est pas du tout explicite
- Revoir les naming des boutons
- Le terme "collection" est à vérifier (revoir)

- | | | |
|--|--|---|
| <input checked="" type="checkbox"/> Raw data | <input type="checkbox"/> Impossible data | <input type="checkbox"/> Prefiltered data |
| <input type="checkbox"/> Eliminated data | <input type="checkbox"/> Filtered data | <input type="checkbox"/> Immobility data |

- revoir le naming des filtres + un bouton "?" pour en savoir plus

COMPATIBILITE : 6/10

LA LOGIQUE D'UTILISATION DU SYSTEME DOIT DONC CORRESPONDRE A LA LOGIQUE DE L'UTILISATEUR

Parler le langage de l'utilisateur. Présenter les informations de façon cohérente par rapport aux autres supports de travail. Rendre compatible l'accès aux commandes avec la tâche de l'utilisateur

- **Globalement les termes sont génériques (accessibles à tous)**
- **Note : Avoir le site en français/anglais**

Résumé

Le bio-logging est un terme qui désigne l'utilisation d'appareils électroniques placés sur des êtres vivants afin d'enregistrer et/ou transmettre des données portant sur la physiologie, le comportement, le mouvement et/ou l'environnement des individus. Cette discipline connaît actuellement son âge d'or grâce aux progrès technologiques réalisés au cours des dernières décennies, notamment la miniaturisation des composants et les avancées au niveau de la transmission de données.

Ces avancées ont rendu ces appareils plus accessibles entraînant alors une très forte augmentation de leur utilisation et donc de la quantité de données récoltées. Cela a mené à des problématiques de standardisation et d'interopérabilité qui nuisent à la bonne valorisation des données de bio-logging. Ce problème est renforcé par la prolifération d'outils supposés accompagner les possesseurs de balises dans le traitement de leurs données, qui est une étape particulièrement chronophage.

Ce mémoire tente de démontrer qu'il est possible de répondre aux différents besoins exprimés par la communauté du bio-logging en utilisant des outils déjà existants, ce qui permet d'apporter une solution aux problématiques de standardisation et d'interopérabilité. Dans ce travail, les outils utilisés sont Movebank, une plateforme de stockage et d'archivage de données, AMCA, une application de traitement de données, un outil de gestion de base de données géographiques et un outil de visualisation cartographique.

Mots clés : *bio-logging, standardisation, interopérabilité, stockage, traitement, diffusion*

Abstract

Bio-logging is a term that refers to the use of animal-borne electronic devices to record and/or transmit data about the physiology, behaviour, movement and/or environment of individuals. This discipline is currently experiencing its golden age thanks to the technological advances made in recent decades, including the miniaturization of components and advances in data transmission.

These advances have made these devices more accessible, resulting in a very large increase in their use and therefore in the amount of data collected. This has led to standardization and interoperability issues that hinder the proper use of bio-logging data. This problem is reinforced by the proliferation of tools that are supposed to support tag owners in the processing of their data, which is a particularly time-consuming stage.

This thesis attempts to demonstrate that it is possible to respond to the various needs expressed by the bio-logging community by using existing tools, thus providing a solution to the issues of standardization and interoperability. In this work, the tools used are Movebank, a data storage and archiving platform, AMCA, a data processing application, a geographic database management tool and a cartographic visualization tool.

Keywords : *bio-logging, standardization, interoperability, storage, processing, distribution*