



HAL
open science

Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?

Laura Mary

► To cite this version:

Laura Mary. Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?. Botanique. 2017. dumas-01799894

HAL Id: dumas-01799894

<https://dumas.ccsd.cnrs.fr/dumas-01799894v1>

Submitted on 25 May 2018

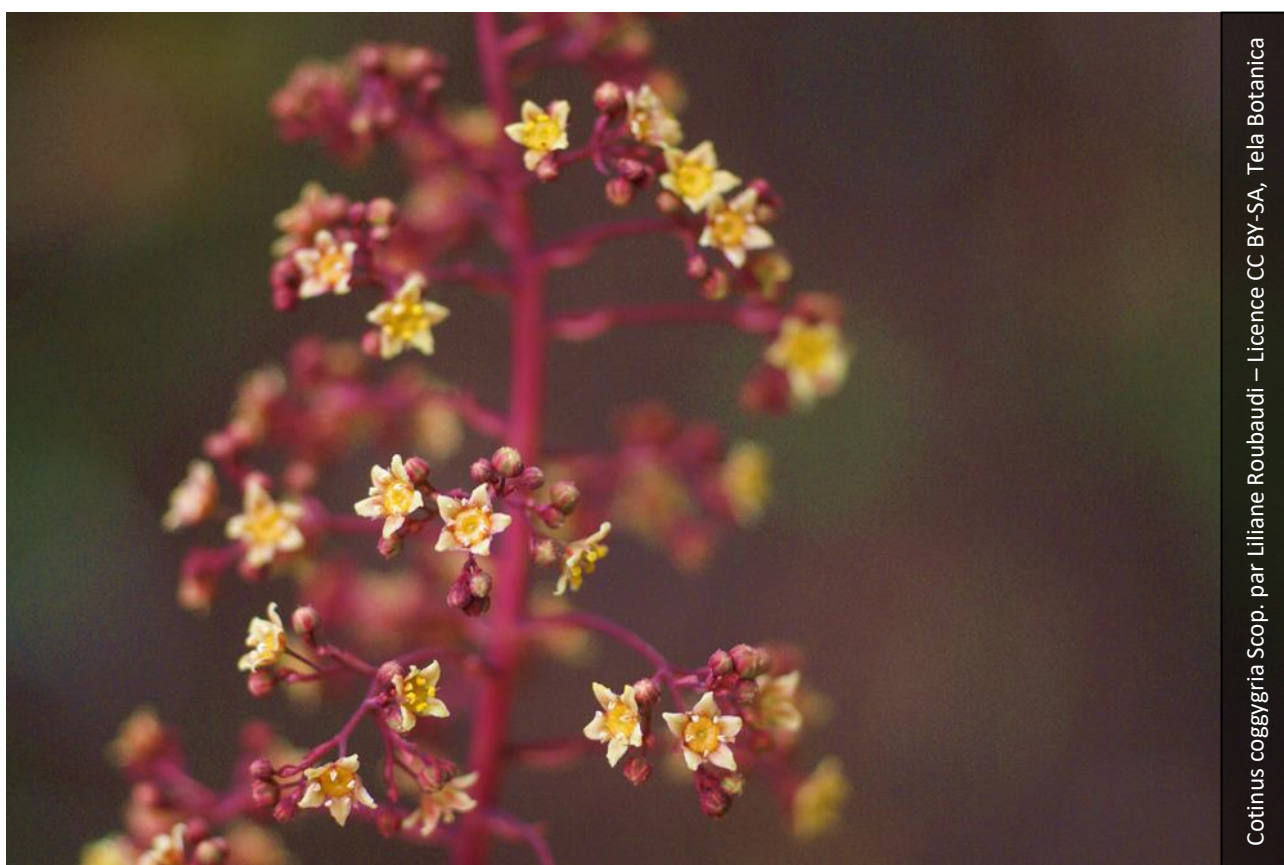
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mémoire de fin d'études

présenté pour l'obtention du diplôme d'ingénieur agronome
Spécialité AgroTIC

Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?



par Laura MARY

Année de soutenance : 2017
Organisme d'accueil : Tela Botanica



Mémoire de fin d'études

présenté pour l'obtention du diplôme d'ingénieur agronome
Spécialité AgroTIC

Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?

par Laura MARY

Année de soutenance : 2017

Organisme d'accueil : Tela Botanica

Mémoire préparé sous la direction de : Bruno TISSEYRE

Présenté le : 18/09/2017

Devant le jury :

- Lionel BOMBRUN
- Guilhem BRUNEL
- Arnaud DUCANCHEZ
- Gilbert GRENIER
- Hazaël JONES
- Nina LACHIA
- Pauline MATRAT
- Léo PICHON
- Bruno TISSEYRE
- Philippe VISMARA

Tuteurs entreprise : Christel VIGNAU, Daniel MATHIEU

Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?

Résumé

Aujourd'hui, Tela Botanica – « le réseau des botanistes francophones » – agrège un grand nombre de données d'observations botaniques (ou occurrences de taxon). Celles-ci ont pour particularité d'avoir été collectées par des botanistes amateurs – au sens non professionnels – sans toujours répondre à un protocole scientifique défini. Pourtant, un certain nombre d'entre elles pourraient venir compléter les données « professionnelles » utilisées en recherche, pour la gestion du territoire ou la préservation de l'environnement.

Une partie de ces données est cependant considérée de « mauvaise qualité », c'est-à-dire incomplète ou incohérente. Afin de valoriser les connaissances et les contributions des amateurs de botanique, que doit mettre en œuvre l'association pour les diffuser ? C'est la question qui a motivé ce mémoire.

Dans un premier temps, l'analyse du système d'information dans son ensemble a permis d'identifier les points critiques, impliqués dans la qualité des données et donc cibles d'améliorations. Cette analyse s'est déroulée en plusieurs étapes : état des lieux de l'existant, veille technologique, enquêtes sur les besoins des utilisateurs, analyse fonctionnelle et enfin spécifications. Parmi ces spécifications, la définition d'un format standard pour les données de Tela Botanica a constitué une base, sur laquelle se sont construits les cahiers des charges des outils et des plateformes de gestion des données d'observation.

L'une des pistes d'amélioration identifiée grâce à cette analyse concerne la localisation des données d'observation. Cette information, indispensable pour décrire l'occurrence d'un taxon, peut être délicate à renseigner correctement. Afin de déterminer les moyens d'optimiser sa saisie, un état de l'art des outils de collecte de données naturalistes a d'abord permis d'identifier les solutions existantes. Un cas d'étude plus spécifique, portant sur la performance du géocodage inversé, a abouti à la définition d'un protocole de saisie de la localisation. Enfin, une étude a permis d'évaluer les conséquences de ce nouveau protocole sur les autres interfaces en lien avec les données d'observation.

Cette étude a montré notamment que la transformation des outils proposés par Tela Botanica (notamment pour la saisie et la révision collaborative), en plus d'enrichir l'expérience des utilisateurs, entraînerait l'amélioration des données d'observation produites. En effet, elle permet d'accompagner la montée en compétence des membres du réseau, tout en limitant la saisie et la diffusion de données incomplètes ou incohérentes.

Ainsi, en adaptant aujourd'hui ses outils au standard d'échange de données naturalistes en passe d'être adopté nationalement, l'association leur assure une certaine pérennité. L'amélioration potentielle et donc la diffusion des données, issues notamment des sciences participatives, permettrait à Tela Botanica de mettre en lumière l'étendue des compétences et des connaissances de son réseau d'amateurs.

Mots-clés

Botanique, donnée d'observation naturaliste, qualité des données, sciences participatives, protocole de saisie, réseau d'amateur

How to improve the quality of botanical observation data collected by amateurs, inside an existing information system?

Abstract

Tela Botanica, the French-speaking botanists network, aggregates nowadays numerous observation data (which describes a taxa occurrence). They have been collected by amateur – meaning nonprofessional – botanists, sometimes missing a rigorous scientific protocol. However, some of this data could be a great addition to “professional” data, used by researchers, for land management or for environmental conservation.

Another part of this data set is considered as “bad quality” data, that is to say incomplete or inconsistent. To promote the knowledge and the contributions of botany lovers, the association must identify what kind of changes are needed. This thematic inspired this memoir.

First, the information system analysis resulted in the identification of some critical aspects, involved in data quality, that should be improved. This study was conducted in several steps: inventory of the existing information system, technology watch, interviews about users’ requirements, functional analysis and eventually specifications. Among these results, the definition of a standard format for Tela Botanica’s data provided a basis to construct specifications for data management platforms and tools.

One of the identified ways to improve data involves data location. This information is essential to describe a taxa occurrence, but it can be difficult to fill it properly. In order to optimize location entry, a study of existing solutions for environmental data collection was made. A more specific case study about reverse geocoding services performance led to the definition of a data entry protocol for location. Finally, the consequences of this new protocol on other interfaces linked to observation data were studied.

This study showed that the renewal of Tela Botanica’s tools for observation data management could lead to enhance users experience, but also to improve data quality. This transformation will, indeed, help users to increase their knowledge and their skills. It will also reduce the entry and the diffusion of incomplete or inconsistent data.

Thus, Tela Botanica’s adaptation to the new national standard format for data exchange will ensure its tools’ sustainability. Thanks to the potential improvement of data quality, the association will be able to highlight the extend of the knowledge of its contributors, including with data generated by citizen science.

Keywords

Botany, biodiversity observation data, data quality, citizen science, data entry protocol, amateur network

Remerciements

Je remercie Bruno Tisseyre, qui a su m'accompagner dans cette alternance, me réorienter lorsque je ne prenais pas le bon chemin et être à l'écoute de mes questions et mes craintes tout en me laissant une grande autonomie. Merci également à Léo Pichon qui m'a encouragée il y a un an à postuler chez Tela Botanica.

Merci à tous les membres de l'équipe AgroTIC, de Montpellier comme de Bordeaux, pour la qualité des cours dispensés cette année et leurs efforts en termes pédagogiques, qui enrichissent énormément notre formation. Merci à ceux d'entre eux que j'ai sollicités pendant mon alternance pour leur réactivité et leur aide.

Merci à Christel Vignau pour sa confiance tout au long de l'année et pour m'avoir accordé du temps à chaque fois que j'en ai ressenti le besoin. Merci également à Daniel Mathieu pour son intérêt pour mon travail, sa disponibilité et ses conseils, sans oublier son engouement contagieux pour la botanique ! Votre accueil au sein de Tela Botanica et les missions que vous m'avez confiées ont contribué à faire de mon alternance une expérience épanouissante.

Je tiens à remercier l'ensemble de l'équipe de Tela Botanica, avec qui j'ai travaillé, échangé et ri cette année : Alexandre, Amandine, Arthur, Audrey, Élodie, Émilie, Delphine, Jennifer, Killian, Mathias et Véro. Merci de votre accueil, votre bonne humeur et votre bienveillance. Je suis heureuse d'avoir partagé avec vous tous ces moments, et je ne doute pas qu'il y en aura encore de nombreux.

A tous ceux qui m'ont soutenu et aidé à produire ce mémoire, par leurs conseils précieux et leur relecture attentive, Alex, Isabelle, Simon, Véro, Zoé, merci !

Table des matières

Table des matières	1
Table des figures.....	3
Table des tableaux.....	3
Glossaire	4
Lexique des abréviations	4
Introduction.....	5
I. Présentation du contexte	6
A. Tela Botanica	6
1. Activités, missions et organisation de l'association	6
2. Un contexte en pleine évolution	7
3. Problématique globale : s'adapter au contexte	7
B. Le programme Flora Data.....	8
1. Gérer des données d'observation	8
2. Gérer la qualité des données.....	10
3. Bilan et présentation de la démarche / mémoire	12
II. L'approche générale : analyser un système d'information existant	13
A. Analyse de l'existant.....	13
1. Méthodologie	13
2. Etat des lieux de l'existant.....	13
B. Veille technologique.....	15
1. Méthodologie	15
2. Résultats	16
C. Analyse des besoins.....	16
1. Méthodologie	16
2. Les besoins des utilisateurs de Flora Data.....	18
D. Analyse fonctionnelle	22
1. Méthodologie	22
2. Résultats	23
E. Spécifications.....	24
1. Méthodologie	24
2. Définition d'un format standard de données.....	24
III. Etude de cas : la (géo)localisation	26

A.	Etat de l'art des solutions existantes.....	26
1.	Méthodologie	26
2.	Etat de l'art : la localisation dans les outils de saisie naturalistes.....	26
B.	Tests comparatifs sur la solution la mieux adaptée : performance du reverse geocoding	31
1.	Définition et intérêt du reverse geocoding	31
2.	Méthodologie	31
3.	Zoom sur un aspect critique : le géocodage inversé	36
C.	Définition du cahier des charges	42
1.	Méthodologie	42
2.	Définition d'un (unique !) protocole de saisie.....	43
D.	Etude des conséquences sur les autres interfaces et outils.....	47
1.	Méthodologie	47
2.	Lien avec les autres outils : quelles conséquences sur la restitution des données ?.....	47
E.	Bilan de l'étude de cas.....	49
IV.	Bilan et perspectives.....	50
A.	Bilan et analyse critique	50
1.	Pertinence de la méthode employée	50
2.	L'organisation du travail et les compétences au sein de l'association.....	50
3.	Les choix stratégiques vis-à-vis de la problématique	51
B.	Perspectives et préconisations.....	53
1.	Quelles compétences nécessaires ?	53
2.	Externaliser les développements : oui mais dans quelle mesure ?.....	53
3.	A quels coûts doit-on s'attendre ?	53
	Conclusion	55
	Bibliographie.....	56
	Table des annexes	58

Table des figures

Figure 1 : Relevé manuscrit d'une observation botanique sur le terrain - Extrait de la vidéo « Enregistrer ses découvertes » du MOOC botanique de Tela Botanica	8
Figure 2 : Le processus Flora Data : parcours de la donnée	10
Figure 3 : Evolution du nombre de données mensuel venant de Pl@ntNet mobile, entre 2015 et aujourd'hui (juin 2017).....	11
Figure 4 : Diagramme de cas d'utilisation : saisir et gérer des données grâce à Flora Data.....	14
Figure 5 : Zone de saisie de l'interface du CEL	30
Figure 6 : Carte interactive illustrant la répartition des données testées (réalisée sur batchgeo.com).....	35
Figure 7 : Les temps d'exécution moyen, minimum et maximum des différents services sur l'ensemble des données	36
Figure 8 : Les taux de réponses complètes des différents services, par jeu de données.....	37
Figure 9 : Les taux de concordance de la localité entre la référence et les différents services pour les données françaises.....	38
Figure 10 : Les taux de concordance de la localité entre la référence et les différents services pour les données étrangères	39
Figure 11 : Justesse de la localité répondue par les différents services pour les données françaises.....	40
Figure 12 : Justesse de la localité répondue par les différents services pour les données étrangères	41
Figure 13 : Diagramme d'activité décrivant les bases du fonctionnement du nouveau protocole de géolocalisation.....	46
Figure 14 : L'un des visuels de la maquette de l'outil de saisie (module de localisation) - Etape 4.....	47

Table des tableaux

Tableau 1 : Caractéristiques et statistiques d'utilisation	27
Tableau 2 : Présentation des services de reverse geocoding envisagés pour le test comparatif	33
Tableau 3 : Avantages et inconvénients de deux modes de localisation : localité vs. géolocalisation.....	44

Glossaire

Licence CC BY SA : Cette licence Creative Commons (CC) permet la libre utilisation d'une œuvre sous réserve de citer l'auteur (BY) ainsi que la transformation sous réserve que le partage se fasse ensuite dans les mêmes conditions (*ShareAlike*, SA) c'est-à-dire avec la même licence que l'œuvre originale.

Référentiel taxonomique : Index listant et caractérisant les taxons d'une région (flore d'une entité géographique définie) ou une thématique (révision d'une famille par exemple). « Ce sont des outils indispensables [qui] mettent à disposition un vocabulaire permettant d'échanger des informations sur les taxons. » (Mathieu & Mathez, 2016).

Taxon : « Unité quelconque (genre, famille, espèce, sous-espèce, etc.) des classifications hiérarchiques des êtres vivants » (INSEE, 2016).

Observation : Ensemble d'informations décrivant la présence d'un taxon (individu ou ensemble d'individus appartenant à ce taxon) en un lieu, à une date et par un observateur donné.

Sciences participatives : « Formes de production de connaissances scientifiques auxquelles des acteurs non-scientifiques-professionnels — qu'il s'agisse d'individus ou de groupes — participent de façon active et délibérée » (Houiller, 2016).

Telabotaniste : Membre du réseau de Tela Botanica, soit toute personne inscrite sur le site internet.

Lexique des abréviations

AMAP : botAnique et Modélisation de l'Architecture des Plantes et des Végétations

API : Application Programming Interface

BDD : Base De Données

BDTFX : Base de Données des Trachéophytes de France métropolitaine et contrées limitrophes

BDTRE : Base de Données des Trachéophytes de la Réunion

BDTXA : Base de Données des Trachéophytes des Antilles Françaises

CBN : Conservatoire Botanique National

CBNMed : Conservatoire Botanique National Méditerranéen de Porquerolles

CEL : Carnet En Ligne

CST : Conseil Scientifique et Technique

DEL : Détermination En Ligne

DWC : Darwin Core

GBIF : Global Biodiversity Information Facility

GPS : Global Positioning System

INPN : Inventaire National du Patrimoine Naturel

Inria : Institut national de recherche en informatique et en automatique

IP : IdentiPlante

LPO : Ligue pour la Protection des Oiseaux

MNHN : Muséum National d'Histoire Naturelle

MOOC : Massive Open Online Course

MQ : MapQuest

obs. : observation(s)

OC : OpenCage

OSM : OpenStreetMap

PF : PictoFlora

PN : Pl@ntNet

SCV : Service Civique Volontaire

SI : Système d'Information

SINP : Système d'Information sur la Nature et les Paysages

SP : Sciences Participatives

TB, Tela : Tela Botanica

TPG : The Plant Game

UML : Unified Modeling Language

UMR : Unité Mixte de Recherche

Introduction

Amateur, amatrice : du latin amator, "celui qui aime" (Larousse, 2017).

L'origine de ce terme, souvent utilisé de manière péjorative, nous rappelle qu'il désigne avant tout quelqu'un qui pratique une discipline par intérêt - voire par passion - sans en faire pour autant son activité professionnelle.

Souvent associées aux valeurs du partage, du libre et du collaboratif, les communautés d'amateurs occupent aujourd'hui une place importante dans la création et la valorisation du "bien commun"¹, notamment grâce au développement et à la démocratisation massif-ve-s du numérique (Lescure et al., 2015). Parmi ces initiatives on trouve de (très) nombreux projets informatiques open source à l'image de GNU/Linux (lancé en 1984), mais on peut également citer, hors du domaine du *hacking*, Wikipédia (2001) ou OpenStreetMap (2004).

C'est sur ce terreau et avec ces valeurs que quelques botanistes, amateurs et professionnels, ont initié ensemble en 1999 Tela Botanica, souhaitant profiter de l'essor du numérique pour créer un réseau réunissant l'ensemble des acteurs de la botanique, afin notamment de promouvoir sa diffusion et sa vulgarisation. Comptant aujourd'hui plus de 38 000 membres, l'association peut se vanter d'avoir réussi, mais il lui faut continuer d'évoluer afin de conserver sa place. Pourtant bien impliquée dans la dynamique des sciences participatives², la structure se voit notamment critiquée à propos de la qualité de ses données d'observation botanique, qui ne répondent pas toutes strictement aux critères d'une donnée « scientifique » (absence de métadonnées, incohérences, doublons, erreurs, etc.). Afin de mieux diffuser les connaissances de ses amateurs de botanique, l'association a montré la volonté de mettre à niveau ses données en s'adaptant aux normes en vigueur.

C'est dans ce cadre que se place ce mémoire d'alternance, qui décrit le travail mené pour la définition et la mise en place de l'amélioration des données d'observation de Tela Botanica. Après avoir décrit le contexte dans lequel se sont déroulées mes missions et les problématiques rencontrées, j'ai présenté la démarche générale suivie pour analyser le système d'information dans son ensemble. Ne pouvant détailler ici l'ensemble des tâches et productions réalisées durant mon alternance, je n'ai sélectionné que certains résultats pertinents vis-à-vis de la problématique. Cette analyse m'a permis d'identifier différentes pistes d'amélioration des données. Parmi elles, les modalités de la localisation d'une observation, présentées sous forme de cas d'étude, font l'objet de la troisième partie. Enfin, un bilan est dressé sur cette expérience et les perspectives que j'envisage pour le projet.

¹ « Les biens communs sont des biens (au sens large : matériels, immatériels, services) qui mobilisent une action collective dans un objectif d'intérêt commun, souvent associé à la soutenabilité (au sens large : respect des limites écologiques et justice sociale). Ils incorporent pour la plupart l'idée d'une construction simultanée d'un bien et d'une communauté. Leur mode de production et/ou de gestion (par l'action collective) les distingue des biens privés (produits par des individus pour eux-mêmes ou pour le marché) et des biens publics (produits et/ou gérés par les pouvoirs publics) » (Dedeurwaerdere & Cassiers, 2012).

² Voir Glossaire

I. Présentation du contexte

A. Tela Botanica

1. Activités, missions et organisation de l'association

Tela Botanica est une association loi 1901, fondée en 1999 par Daniel Mathieu, la Société Botanique de France, la Garance Voyageuse et Biotope. Depuis, elle accompagne la mise en place et l'animation de projets botaniques, en s'appuyant sur un fonctionnement de réseau et des outils collaboratifs ouverts. Son but historique est de mettre en place et gérer un réseau de botanistes francophones, réseau de communication et d'échange ouvert à toute personne intéressée par la botanique. Dans ce cadre, Tela Botanica vise à constituer un centre de ressources scientifiques sur la botanique, à partir notamment des données fournies par les membres du réseau, développer des ressources pédagogiques, organiser et accompagner le montage de projets botaniques, assurer des formations et des prestations dans le domaine de la botanique pour l'ensemble du réseau. Enfin, elle participe aux activités en faveur de la protection de la nature et de l'environnement lorsqu'elles relèvent de ses thématiques d'action (principalement la flore sauvage).

En pratique aujourd'hui, outils et ressources sont mis à la disposition de tous sur le site internet de Tela Botanica www.tela-botanica.org, sur lequel est également menée une grande partie de l'animation des programmes de sciences participatives et l'accompagnement des projets.

D'un point de vue structurel, l'association emploie une dizaine de salariés permanents et périodiquement des services civiques et stagiaires. Les actions et les outils proposés par l'association étant en grande majorité sur internet, les activités salariées sont réparties autour de l'animation (4 salariées + SCV/stagiaire-s), le développement informatique et la gestion du SI (3 salariés), ainsi que la gestion administrative (2 salariées). Le conseil d'administration et le conseil scientifique et technique (CST) participent quant à eux au pilotage de l'association en termes stratégiques pour l'un et scientifiques pour l'autre.

Les missions des salariés sont organisées autour de trois axes, reflétant les objectifs de l'association :

1. **Développer le réseau** : lettre d'actualité, animation sur les réseaux sociaux, maintenance/amélioration du site internet, projets à l'étranger, développement d'applications mobiles, etc.
2. **Constituer des bases de connaissances** : collecte, gestion et restitution des données collaboratives, gestion des serveurs.
3. **Sciences participatives et pédagogie** : MOOCs, programmes de sciences participatives et observatoires.

La conduite des projets informatiques en méthode Scrum a été élargie et adaptée à l'ensemble de l'équipe. Un « product owner » a ainsi été désigné pour chaque projet (généralement la personne chargée du projet), dont le rôle est d'échanger avec l'équipe informatique afin d'organiser les développements. Les fonctionnalités à développer ou « stories », les notifications de bugs et le suivi de l'avancement en général sont centralisés sur la plateforme Taïga. Un « scrum meeting » ou « stand up meeting » hebdomadaire permet à tous les membres de l'équipe de présenter ce qu'ils ont fait la semaine passée et feront la semaine à venir, afin d'améliorer la visibilité pour l'ensemble de l'équipe sur les activités de chacun, et la communication de manière générale.

Méthode Scrum

La méthode Agile est une approche de gestion des projets reposant sur un processus itératif et incrémental. Dans le domaine informatique, les développements sont planifiés sur de courtes périodes (sprints ou itérations), à partir de spécifications générales élaborées en début de projet. A l'issue de chaque sprint on peut éventuellement réévaluer les développements à venir, la planification, les fonctionnalités à développer, etc. Ce fonctionnement permet ainsi d'être plus réactif et de s'adapter rapidement au changement, en construisant étape par étape.

Scrum est une des méthodes Agiles les plus courantes. Elle définit notamment les rôles de « Product Owner », la personne qui porte le projet et de « Scrum Master », qui assure l'application de la méthode Scrum (Lothon, sd).

2. Un contexte en pleine évolution

On constate aujourd'hui que, si le public de Tela Botanica est toujours en majeure partie français (82%) et francophone pour le reste (dont 8,6% en Afrique), le profil des telabotanistes, lui, évolue. En effet, parmi les 38 000 inscrits au réseau, on compte depuis quelques temps plus de 50% d'utilisateurs se déclarant débutants lors de leur inscription (Tela Botanica, 2017). Les membres du réseau ne sont donc plus seulement les botanistes émérites de son origine mais aussi de nombreux curieux et amoureux de la nature sans expertise particulière dans le domaine. Afin d'**accompagner l'évolution de ses utilisateurs**, l'association cherche à adapter son site web, ses outils et ses animations de manière à ce que tous, des plus experts aux débutants en botanique, y trouvent leur compte en termes d'ergonomie comme de contenu. La réalisation d'une plateforme MOOC et le succès du premier cours d'initiation en botanique fin 2016 sont de bons exemples de ce renouvellement.

Dans ce contexte, l'association a initié en 2015 avec différents partenaires dans les domaines de la botanique, l'informatique et l'animation, le **projet Floris'Tic**. Celui-ci a pour objectif de promouvoir la botanique et les sciences du végétal auprès du plus grand nombre grâce au numérique, donnant ainsi une image et une accessibilité nouvelle à un domaine vieillissant. Cela passe notamment par le **développement et le partage d'outils numériques collaboratifs**, adaptés aux différentes communautés d'usagers. La participation à ce projet implique donc également, par la nature de ses objectifs, de s'adapter à un public nouveau pour l'association, dont les attentes et les usages sont en partie inconnus.

3. Problématique globale : s'adapter au contexte

Début 2016, un gros chantier visant à la **refonte totale du site internet de Tela Botanica** a été lancé, impliquant son équipe informatique, une animatrice et un prestataire (en ergonomie et développement web). Le nouveau site doit répondre à un besoin de modernisation mais aussi d'ouverture et d'amélioration des services rendus par l'association, en optimisant l'ergonomie des interfaces, simplifiant le plan du site et homogénéisant l'ensemble des pages. Certains outils comme l'espace projets³ ont été entièrement revus pour l'occasion, d'autres comme la flore en ligne eFlore⁴ simplement adaptés, mais la plupart ont été transposés tels quels.

³ Interface permettant à chacun de créer et suivre des "projets", espaces thématiques d'échange et de discussion (forum, porte-document, wiki, ...) voire de création collaborative. Voir : www.tela-botanica.org/page:projets

⁴ Flore électronique créée et enrichie par les membres du réseau. Voir : www.tela-botanica.org/page:eflore

L'association est ainsi amenée à revoir son site internet et l'ensemble des outils collaboratifs qu'il héberge. L'enjeu est de ne pas se laisser dépasser sur le plan technologique à cause de l'obsolescence de certains de ses outils, tout en maintenant la qualité des services et la participation des membres du réseau.

Dans ce cadre, le **programme Flora Data** occupe une place historique. Il regroupe une grande partie des outils collaboratifs permettant **la saisie et la gestion des données d'observation botaniques**. En initiant une analyse sur l'environnement et le fonctionnement de Flora Data comme système d'information, Tela Botanica souhaite redynamiser la collecte participative de données dans le champ de la botanique, à l'échelle nationale voire internationale.

B. Le programme Flora Data

1. Gérer des données d'observation

L'objectif premier de ce programme est la collecte et la gestion de données d'observation botanique. Il convient donc de commencer par définir cette notion qui sera centrale dans la suite de mon étude :

Une observation botanique qualifie **l'identification de la présence d'un taxon végétal**. Elle est constituée des informations suivantes :

- Quoi :
 - **un nom** caractérisant l'objet de l'observation (taxon, nom commun, « indéterminé », etc.). Dans le cas où le taxon est identifié, celui-ci doit être rattaché à un référentiel taxonomique ;
 - éventuellement **une ou plusieurs photos** illustrant l'espèce observée, surtout dans le cas où le taxon n'est pas identifié par l'observateur ;
- Où : **le lieu** (coordonnées précises, entité géographique, pays...);
- Quand : **la date** de l'observation ;
- Qui : un (ou plusieurs) **observateur-s** ;

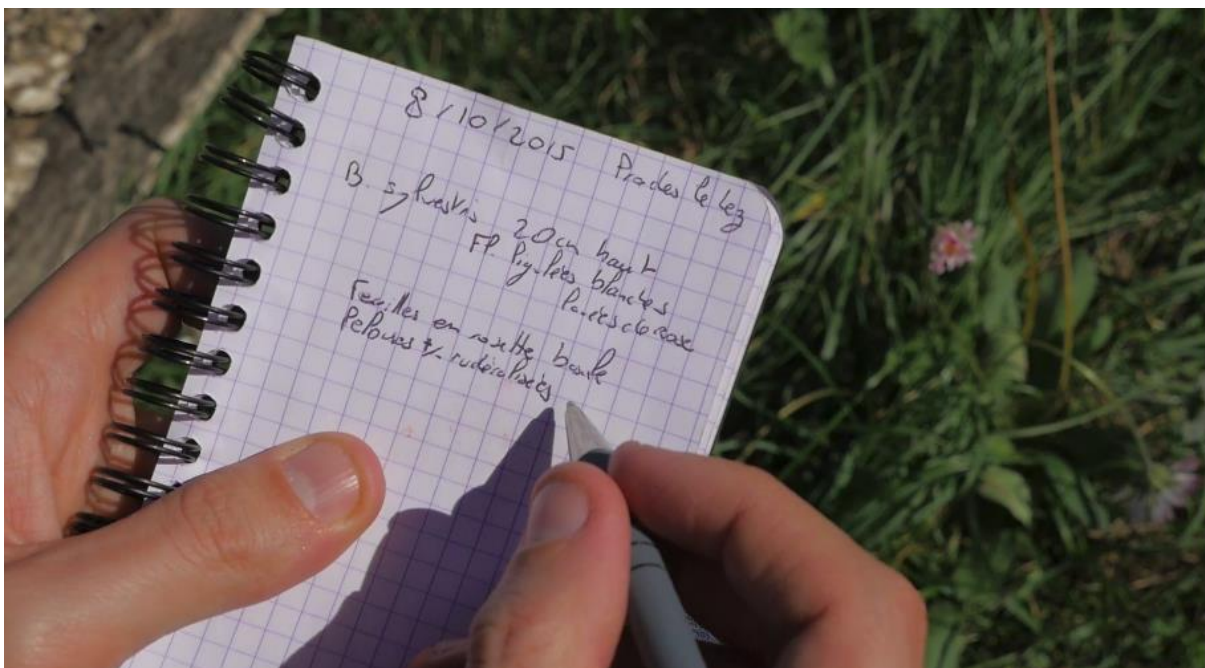


Figure 1 : Relevé manuscrit d'une observation botanique sur le terrain - Extrait de la [vidéo](#) « Enregistrer ses découvertes » du MOOC botanique de Tela Botanica

Des informations complémentaires peuvent accompagner une observation botanique, mais celles qui précèdent sont indispensables à la valorisation des données (Jomier et al., 2016). Dans la suite du document, une **donnée d'observation « de qualité »** désigne une donnée **pour laquelle l'ensemble de ces informations au moins ont été renseignées et sont cohérentes**.

Dans le double objectif de fournir aux botanistes amateurs des outils performants pour gérer en ligne leurs propres données d'observation et de centraliser et mettre à disposition de tous ces données, Tela Botanica a mis en place le **programme Flora Data**. Celui-ci est souvent désigné comme un processus car il permet de suivre le « parcours de la donnée », de sa saisie à sa restitution en passant par sa révision collaborative. Le processus général intègre différents référentiels taxonomiques⁵, qui permettent son usage en dehors de la métropole (Liban, Afrique du Nord, Afrique de l'Ouest, plusieurs DOM-TOM...). Toutes les données publiées peuvent être réutilisées librement.

Parmi les outils collaboratifs proposés dans le processus Flora Data, le **Carnet en Ligne (CEL)** permet de **saisir, stocker et gérer** (modifier, dupliquer, supprimer, ...) ses observations botaniques et leurs illustrations. Cet outil est puissant pour qui sait s'en servir et son rôle est central, mais il n'a quasiment pas évolué depuis sa création il y a une dizaine d'années. Une interface dédiée à la **saisie rapide** a été développée plus récemment, censée simplifier la procédure de saisie. Elle est plus moderne mais aussi moins précise, puisque pour faciliter la saisie plusieurs champs optionnels ont été supprimés. D'autre part, elle ne permet pas la gestion des observations saisies précédemment comme le CEL.

Les données illustrées (c'est-à-dire associées à une/plusieurs photo-s) et rendues publiques apparaissent ensuite automatiquement sur **IdentiPlante**, la **plateforme de validation participative**. Elle permet, à partir des photos postées avec l'observation, de :

- confirmer ou infirmer l'espèce sélectionnée par l'auteur
- proposer un autre taxon si l'auteur n'a pas sélectionné de nom ou si le nom sélectionné est faux

Un système de votes et de commentaires permet à l'auteur de l'observation de savoir quelle proposition d'identification fait consensus et la valider. Le réseau peut ainsi l'aider à identifier ou vérifier le taxon qu'il a observé.

En parallèle, **PictoFlora** permet de traiter les photos, cette fois-ci indépendamment de l'observation qu'elles illustrent. Cet outil se présente comme une galerie photo et son fonctionnement repose également sur la participation des membres du réseau. Il consiste à **caractériser les photos** en les taguant, par exemple en fonction de l'organe végétal qu'elles représentent, et à **noter leur qualité**. Cela permet notamment de collecter des métadonnées sur les photos, qui pourront servir à les discriminer lors de leur utilisation sur d'autres plateformes.

Les données peuvent être visualisées sur deux **cartographies** (communale et au point) et sur plusieurs interfaces de restitution. Il est également possible de télécharger des jeux de données (de tous les membres) grâce à un **widget d'export**, qui permet de filtrer les observations sur différents critères.

⁵ Index caractérisant les taxons d'une région (flore d'une entité géographique définie) ou une thématique (révision d'une famille par exemple). « Ce sont des outils indispensables [qui] mettent à disposition un vocabulaire permettant d'échanger des informations sur les taxons. » (Mathieu & Mathez, 2016).

Enfin, plusieurs interfaces utilisent des données (en particulier les photos) issues de Flora Data, comme **eFlore** ou **Smart'Flore**, l'outil de création de sentiers botaniques, mais ils n'en font pas partie en tant que tels.

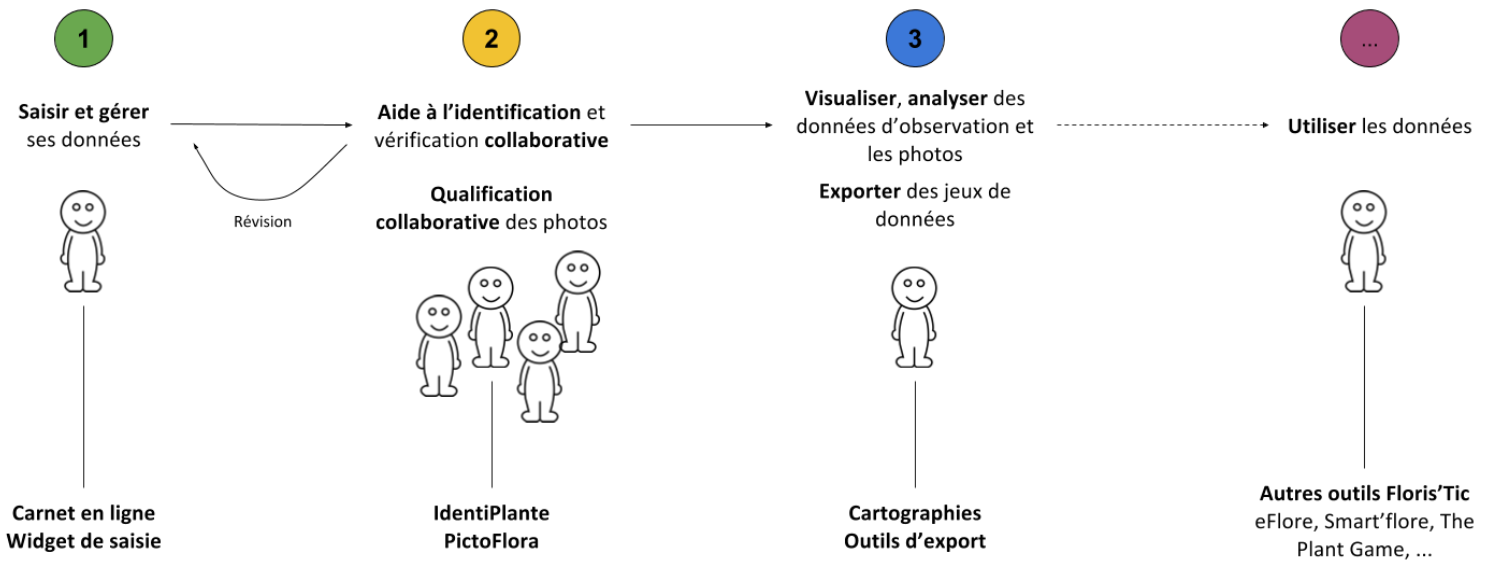


Figure 2 : Le processus Flora Data : parcours de la donnée

Grâce à ces outils, Flora Data vise à rendre des **services à la fois individuels et collectifs**. Ces objectifs ne sont pas contradictoires mais ils obéissent à des contraintes différentes :

- D'une part, proposer des **services individuels d'aide à la gestion de données d'observation** : saisie, aide à l'identification et à la vérification, stockage et visualisation de ses données. Les contraintes sont l'ergonomie, la facilité d'utilisation, la robustesse et la mobilité (usages nomades).
- De l'autre, **permettre le partage des données au sein du réseau**. Cela induit notamment leur intégration dans le cadre de programmes participatifs ou encore leur utilisation par des communautés ou des structures (cartes de répartition collaboratives, missions flore spécifiques, export vers le SINP ou le GBIF, etc.). Il s'agit de la dimension participative et collaborative du processus Flora Data, qui implique la standardisation, la bonne caractérisation, la qualité et une quantité significative de données ainsi qu'une certaine ergonomie pour le partage.

Afin d'optimiser le processus participatif, il convient de faire converger au mieux ces deux objectifs en prenant en compte leurs enjeux et leurs contraintes.

2. Gérer la qualité des données

Historiquement, le processus Flora Data a été construit par accumulation d'outils et de fonctionnalités, sans réelle réflexion sur son architecture globale. **Cela limite à présent l'amélioration du service et de la qualité des données**. L'obsolescence des outils les plus anciens et les problèmes de cohérence au sein du processus en général nécessitent aujourd'hui une révision du système. Différents éléments de contexte présentés ci-dessous viennent également poser la question de la gestion de la qualité des données.

Essor de Pl@ntNet mobile et augmentation du flux de données

Début 2013, l'UMR AMAP et l'Inria ont lancé l'application **Pl@ntNet mobile** qui permet **l'identification d'une plante par reconnaissance d'image**. L'application, par la suite intégrée au projet Floris'Tic, renvoie chez Tela

Botanica l'ensemble ou presque des données publiées. A partir de l'été 2016, l'association a vu une **rapide évolution dans la nature et la quantité des données** envoyées par PI@ntNet, en lien notamment avec le succès de l'application. Le système d'information doit désormais faire face à une croissance exponentielle du flux de données entrant. En plus de la contrainte technique que cela représente (saturation de la base de données, ralentissement du serveur...), la vocation participative de Flora Data et notamment la révision des données illustrées (IdentiPlante) n'ont plus de sens lorsque **la quantité de données dépasse à tel point le nombre d'utilisateurs actifs sur le site**. Le graphique ci-dessous représente bien l'évolution des flux de données : alors qu'en 2015 le nombre maximum de données envoyées depuis PI@ntNet est de 3 900 au mois de juin, il est de près de 50 000 un an plus tard, en août 2016, et on approche déjà les 100 000 en avril 2017, avant même le début de l'été. On constate également qu'en parallèle, le nombre de données envoyées depuis les outils Tela Botanica reste relativement stable, les pics d'activités tournant autour de 6 000 données par mois et jusqu'à 12 000 certaines années (non représenté sur le graphe ci-dessous, se reporter à l'Annexe A).

Evolution de la quantité de données envoyées par mois

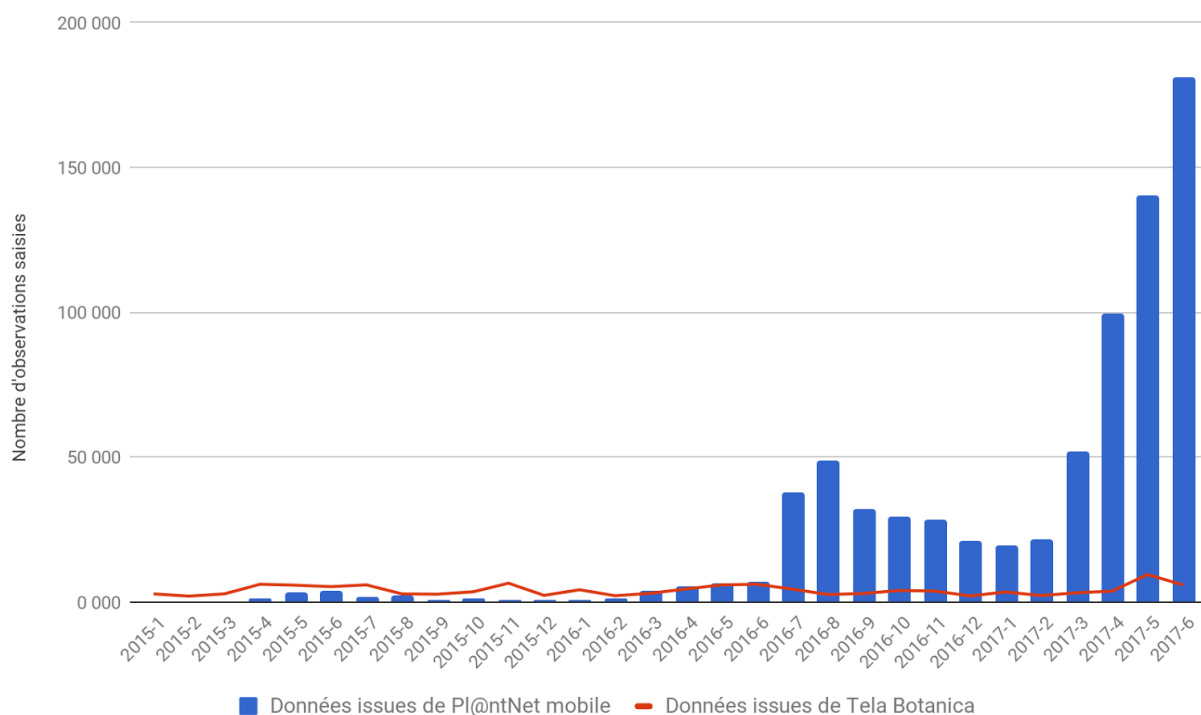


Figure 3 : Evolution du nombre de données mensuel venant de PI@ntNet mobile, entre 2015 et aujourd'hui (juin 2017).

D'autre part, les données envoyées depuis PI@ntNet mobile sont souvent **indéterminées et peu qualifiées** (manque de métadonnées). Elles contiennent une proportion importante de plantes cultivées et de photos inadaptées à une analyse botanique (floues, mal cadrées, etc.), ce qui concourt d'autant plus à décourager les membres du réseau de Tela. Des solutions de « secours » ont été adoptées jusqu'à présent afin de permettre aux utilisateurs de continuer à utiliser Flora Data sans devoir se couper totalement du flux de PI@ntNet, qui reste un outil pertinent pour une partie des utilisateurs, mais le problème mérite une réflexion plus approfondie.

Qualité des données et norme SINP

Une étude a été menée entre 2014 et 2016 par le CBNMed sur la qualité des données de Tela Botanica. L'objectif était d'évaluer **l'apport des données issues des sciences participatives** à la base de données du Conservatoire, en vue de les intégrer au SINP. Cette analyse s'inscrit dans la démarche actuelle du SINP de mise en partage des données naturalistes françaises. Pour Tela, il est important de s'adapter aux normes de qualité mises en place dans ce cadre afin de continuer à valoriser les données collaboratives et à échanger avec d'autres structures.

Le rapport du CBNMed met en évidence « l'apport non négligeable que fourniraient les sciences participatives ». En effet, près d'un quart des données de Tela sont considérées comme issues des sciences participatives⁶ et apportent une information nouvelle⁷ par rapport à la base du CBNMed. Cela représenterait un apport d'environ 12,7% pour le CBNMed soit 2,5% pour le SINP, ce qui est loin d'être négligeable. Cependant, le CBNMed souligne la nécessité pour Tela Botanica d'améliorer ses outils afin de limiter le nombre de données incohérentes ou incomplètes, actuellement en grand nombre. En conclusion, le rapport estime qu'une « **mise en conformité [des données de Flora Data] avec les standards du SINP est une condition préalable avant toute intégration** » (De Barros & Molina, 2017).

Le SINP (Système d'Information sur la Nature et les Paysages) est un dispositif partenarial entre le ministère de l'Écologie et les acteurs de la biodiversité et du paysage. Il consiste en un réseau d'acteurs acceptant de partager l'information naturaliste qu'ils détiennent. Dans ce cadre, le SINP facilite la mise en commun d'informations tout en offrant un cadre méthodologique de référence. Il facilite, si nécessaire, la mise en place d'outils communs et de règles de fonctionnement, dont la définition des critères de qualité des données.

Accessibilité problématique aux outils

De nombreux retours mettent en évidence la **complexité et le manque d'ergonomie** du site internet et notamment des outils de Flora Data. Les utilisateurs (surtout les nouveaux) abandonnent rapidement leur utilisation, trop compliquée au premier abord. Ceux qui connaissent les outils depuis longtemps ou ont été formés ont généralement une bonne vision des fonctionnalités de Flora Data, mais son fonctionnement est beaucoup plus obscur pour les nouveaux inscrits. Cela est dû notamment au **manque d'unité et de lien entre les différentes interfaces** de Flora Data ainsi qu'à leur ergonomie déficiente.

On notera que cet aspect a également une influence sur la qualité des données. En effet, les utilisateurs ayant saisi une observation ne savent pas toujours comment l'éditer ou profiter de l'aide du réseau pour son identification, la laissant incomplète ou erronée. De même, il est difficile d'amener les utilisateurs de Pl@ntNet, application mobile relativement intuitive, à utiliser les outils de Tela Botanica dont la prise en main ne l'est pas. Le lien entre Pl@ntNet et Tela Botanica n'apparaît d'ailleurs pas explicitement dans l'application, ce qui limite d'autant le passage des utilisateurs d'une plateforme à une autre pour suivre leurs données.

3. Bilan et présentation de la démarche / mémoire

Forts de ces constats, nous avons identifié au début de mon alternance deux thématiques majeures pour améliorer Flora Data : **la gestion de la qualité des données et l'accessibilité aux outils pour tous les utilisateurs**. Si j'ai choisi de me concentrer dans ce mémoire sur la question des données, la deuxième problématique n'a pas été écartée pour autant : elle est ici mise au service de l'amélioration des données.

⁶ C'est-à-dire saisies par des particuliers, hors du cadre professionnel

⁷ Données dont la localisation est plus précise, l'observation plus récente ou nouvelle pour la commune

II. L'approche générale : analyser un système d'information existant

Dans cette partie, nous verrons la démarche globale qui a été suivie avant de produire des spécifications concernant Flora Data. Pour chaque étape, la méthode puis les résultats sont présentés. Ne pouvant entrer dans les détails de manière systématique, j'ai choisi de développer les parties les plus pertinentes pour comprendre les prises de décisions et les recommandations qui suivront.

A. Analyse de l'existant

1. Méthodologie

Objectifs : prendre en main le système, comprendre son fonctionnement. Analyser la situation avec du recul, un regard neuf. Construire un document qui servira de base pour le reste de l'analyse.

a) Découverte

Mon statut de novice en botanique m'a permis d'aborder le fonctionnement de base de Flora Data avec un regard de débutante. J'ai porté mon attention sur :

- Le fonctionnement général,
- Les dysfonctionnements, les bugs,
- Ce qui est incompréhensible sans explications supplémentaires.

En parallèle, j'ai parcouru la documentation existante, en grande partie hébergée sur l'intranet de Tela. L'étude de la base de données m'a également permis de mieux comprendre l'organisation du SI.

b) Formalisation

Une deuxième étape de formalisation m'a permis de mieux échanger avec le reste de l'équipe sur le fonctionnement du SI. Dans un premier temps, j'ai cherché à comprendre la répartition des flux de données entrant, sortant et au sein de Flora Data via la réalisation de schémas. Je me suis également appuyée sur la modélisation UML afin de décrire le système, la gestion des droits et la base de données. J'ai réuni le tout dans un document décrivant précisément le système, les différents processus et leur environnement.

Dans un deuxième temps, j'ai entrepris de centraliser l'ensemble des critiques et propositions d'améliorations effectuées en interne, glanées lors de mes recherches dans la documentation, les comptes-rendus de réunions, d'anciennes user story (en attente depuis plusieurs années) et beaucoup en discutant avec l'équipe. De nombreuses remarques et suggestions avaient déjà été faites, mais rarement formalisées.

c) Synthèse

Enfin, j'ai rédigé une synthèse de l'analyse reprenant pour chaque élément de Flora Data les grands constats, les principales qualités et les critiques portées au système d'information et à son fonctionnement. Ceux-ci sont mis en perspectives avec les objectifs à atteindre grâce à l'amélioration du système.

2. Etat des lieux de l'existant

a) Etude des flux de données

J'ai réalisé deux schémas pour illustrer les flux de données, au sein de Flora Data d'une part et avec les autres projets Floris'Tic d'autre part (cf. Annexe B). Ils constituent une base pour expliquer le fonctionnement du SI et donnent un aperçu de l'ensemble des sources de données. Ils ont également permis d'identifier certains points bloquants ou mal connus côté Tela Botanica, qui méritent d'être éclaircis avec l'équipe Floris'Tic. C'est

le cas en particulier du processus de « push » des données de Pl@ntNet vers Tela. Bien que ces schémas aient mené à des échanges sur le sujet, ils n'illustrent pas réellement les problématiques soulevées par la gestion des flux de données. Une description plus précise des processus est donc nécessaire.

b) Diagrammes UML

Décrire le fonctionnement global du système : diagrammes de cas d'utilisation

Les diagrammes de cas d'utilisation mettent en évidence l'organisation des grandes fonctionnalités et présentent dans quels cas elles sont utilisées. Ils permettent également d'étudier les droits accordés aux différents utilisateurs. Un diagramme a été réalisé pour chacune des trois grandes étapes du parcours de la donnée : la **saisie et la gestion des données** (diagramme ci-dessous), l'**aide à l'identification**, la **visualisation et l'analyse des données** (cf. Annexe C).

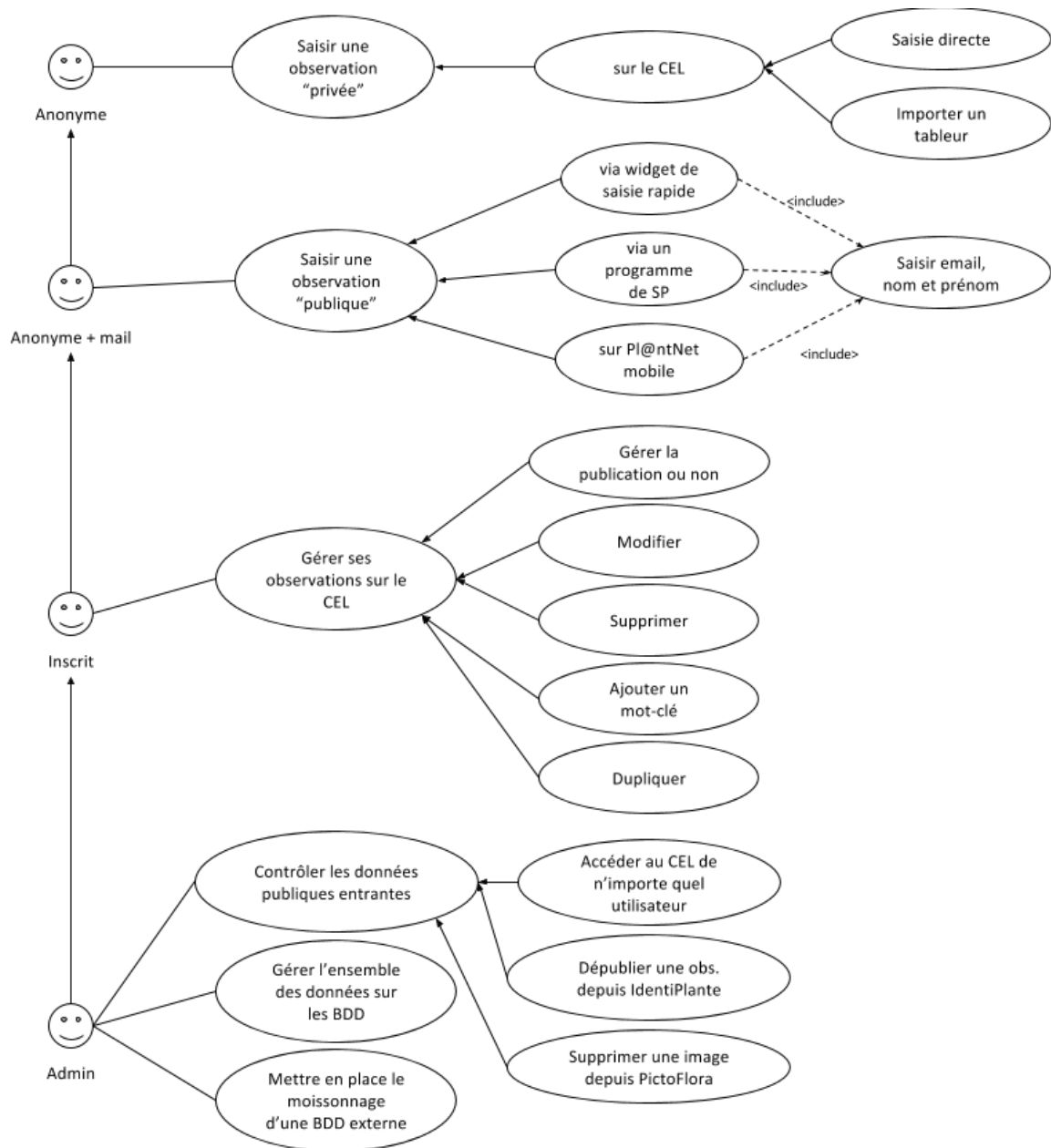


Figure 4 : Diagramme de cas d'utilisation : saisir et gérer des données grâce à Flora Data

Toutes les fonctionnalités ne sont pas illustrées par ces diagrammes, de même que certains cas particuliers ne sont pas traités. Par exemple, un utilisateur anonyme peut retrouver une observation saisie dans le CEL (et donc l'éditer, la modifier, etc.) dans le cas où il n'a pas supprimé les cookies de son navigateur. De même, si un utilisateur non inscrit renseigne son adresse mail lors de la saisie sur le widget de saisie rapide, il pourra retrouver ses observations (sur le CEL) s'il s'inscrit par la suite avec le même email.

Ces exemples illustrent bien **la complexité et la diversité des cas envisagés lors de la conception du système**, dans le but de laisser une grande liberté à l'utilisateur. En effet, l'accessibilité au site et l'utilisation des outils est assez peu restreinte pour les utilisateurs « anonymes » (non identifiés sur le site). Ces éléments font part de la volonté de l'association **d'ouvrir le site et les données**, sans forcer l'inscription lorsque ce n'est pas indispensable pour des questions de gestion. Si cela facilite la vie de quelques utilisateurs ne souhaitant pas s'inscrire, ces choix ont d'autres conséquences : d'une part, les données saisies anonymement ne peuvent être rendues publiques et ne peuvent pas être retrouvées par leur auteur à moyen terme (cookies), elles sont donc en quelques sortes « perdues » puisque personne ne peut les voir. D'autre part, cela contribue à la complexité du système côté gestion (trop de cas différents), voire côté utilisateurs, ceux-ci n'ayant pas l'habitude de ce fonctionnement.

Décrire comment sont gérées les données en interne : diagrammes de classe

Afin de mieux comprendre la structuration des données de Flora Data, j'ai modélisé les bases CEL (Carnet En Ligne) et DEL (Détermination En Ligne). La première **stocke les données d'observation** elles-mêmes, les images et les mots-clés associés. La deuxième permet **la gestion des commentaires et des votes** issus de la révision participative (IdentiPlante et PictoFlora). Les diagrammes de classe sont consultables en Annexe D.

c) Analyse et synthèse

L'analyse de l'existant a fait l'objet d'un document complet, reprenant les productions ci-dessus et décrivant l'ensemble des processus du point de vue de l'utilisateur, le plus précisément possible. Cela a permis l'identification d'un certain nombre de freins au bon fonctionnement du système mais surtout de comprendre son fonctionnement en le décrivant. Afin de mettre en avant les points importants et de simplifier la communication à l'aide d'un document plus court, une synthèse reprend quelques éléments de contexte, les qualités, les critiques et les objectifs de chaque processus. Un tableau récapitulatif des principales conclusions tirées de l'analyse de l'existant est consultable en Annexe E.

Bilan : Une fois le fonctionnement du système globalement compris et un certain nombre de freins identifiés en interne, je me suis intéressée à des points de vue extérieurs à l'association. Pour cela, j'ai entamé une veille technologique ainsi qu'une analyse des besoins des utilisateurs.

B. Veille technologique

1. Méthodologie

Objectifs : connaître ce qui se fait ailleurs, comparer les volumes de données traités. S'inspirer de ce qui existe et qui marche.

J'ai mené une veille technologique sur les **outils collaboratifs de collecte de données naturalistes**, plus ou moins en continu pendant toute l'alternance. Des recherches approfondies m'ont permis d'identifier de nombreux sites, réseaux et structures, proposant ce type d'outil. Pour chaque plateforme, j'ai rédigé une fiche reprenant l'objet du site, un rapide historique, quelques statistiques sur le nombre d'utilisateurs, leur

activité, les volumes de données, une synthèse des fonctionnalités de l'outil de saisie quand j'ai pu le tester, d'éventuelles remarques sur le reste du site, ou encore des articles de presse.

2. Résultats

L'ensemble des notes prises sur ces outils naturalistes ne pouvant être restitué dans ce mémoire, une partie des résultats de la veille technologique seulement est présentée en III.A.2 dans le cadre plus restreint de l'étude du protocole de localisation. Un récapitulatif des sites étudiés est restitué en Annexe F.

C. Analyse des besoins

1. Méthodologie

Objectifs : établir une typologie d'utilisateurs pour aboutir à des personae, déterminer quels sont leurs usages et leurs besoins dans leur pratique de la botanique (en lien avec Flora Data), afin d'améliorer les services proposés en relation avec ces besoins

a) Définition d'un panel d'utilisateur

En étudiant les profils des membres de l'association et après des échanges avec l'équipe, une première liste des profils des utilisateurs a été dressée. Cette classification *a priori* est nécessaire pour orienter la discussion au mieux pendant les entretiens, en ciblant les questions et les sujets les plus adaptés suivant les interviewés. L'idée est aussi de répartir au mieux les entretiens entre ces différents profils, afin de recueillir le besoin de manière représentative (idéalement) parmi les utilisateurs de Tela. **Six profils ont ainsi été définis**, destinés à évoluer au fur et à mesure des entretiens :

- les débutants en botanique
- les utilisateurs avertis
- les botanistes experts
- les enseignants/animateurs
- les professionnels de la botanique
- les utilisateurs des données - finalement traités à part lors de l'analyse

Pour chacun des six profils, un certain nombre d'utilisateurs ont été identifiés comme représentatifs par l'équipe de Tela Botanica, souvent en relation avec eux dans le cadre d'animations ou de projets communs. Cette liste a été complétée en recherchant parmi les inscrits actifs sur le site. Une quarantaine de contacts ont été collectés en tout, plus ou moins répartis sur les six profils.

b) Mise en place et déroulement des entretiens

Le choix du type d'entretien dépend des objectifs et du cadre de l'enquête. Ici, afin de recueillir les usages et les besoins des utilisateurs sans les influencer par des questions trop spécifiques, c'est **l'entretien semi-directif** qui a été choisi. Ce type d'entretien est bien adapté à la collecte de données qualitatives, tout en permettant de comparer les résultats obtenus sur l'ensemble (ou parmi des sous-groupes) des personnes interrogées. Il prend la forme d'une conversation, qui met en confiance la personne interrogée afin de garantir la richesse et l'authenticité de ses propos (Imbert, 2010). La discussion est cependant dirigée en fonction de thèmes choisis au préalable, ici :

- **Cibler le profil de l'utilisateur** (âge, profession, centres d'intérêt, etc.), son rapport à la botanique et aux outils numériques ;
- **Comprendre ses usages** des outils et des données de Tela Botanica ;

- **Identifier ses besoins** dans la pratique de la botanique.

Un guide puis des canevas d'entretien ont été préparés en amont des interviews, listant les informations à collecter dans le but de nourrir si besoin la conversation (exemple en Annexe G). Ils ont été améliorés au fur et à mesure des entretiens, certaines questions n'ayant pas lieu d'être et d'autres demandant à être approfondies. Le canevas aide également à la prise de note, mais dans la mesure du possible la retranscription informatique a suivi directement l'entretien. Tous les utilisateurs du panel ont été avertis par mail, puis les volontaires ont été recontactés pour convenir d'une date d'entretien (téléphonique ou sur Skype).

c) Analyse - de l'extraction à la priorisation des besoins

La première phase de l'analyse consiste à séparer, parmi les informations collectées, celles qui permettent l'identification du profil utilisateur et celles qui concernent les besoins.

Dans un deuxième temps, les besoins sont identifiés et classés par grandes catégories, à partir de l'analyse des entretiens. Cette étape permet notamment de dénombrer les évocations d'un même besoin et de mettre en évidence les profils d'utilisateurs associés. Une note de 0 à 3 leur est ensuite attribuée selon l'importance que chaque personne lui a accordée :

- 3 : **besoin fondamental, prioritaire**
- 2 : **besoin secondaire**
- 1 : **besoin accessoire**
- 0 : **besoin inexistant** (et exprimé comme tel)
- Pas de note : le besoin n'a **pas été évoqué**

L'avantage de cette méthode de priorisation est qu'elle se fait totalement *a posteriori* : lors des entretiens, la conversation est orientée vers certaines thématiques rejoignant les catégories de besoins mais l'utilisateur n'a pas à donner lui-même de notes, ce qui alourdirait l'entretien. D'autre part, ce système de notation est facile à pondérer en fonction du nombre d'occurrences du besoin et de l'importance que l'on souhaite accorder éventuellement à certains utilisateurs/groupes d'utilisateurs. Cependant, cette méthode implique une part de subjectivité dans l'attribution de la note, qui est donnée en fonction du ressenti de l'interrogateur et pas directement par l'utilisateur.

Une fois chaque besoin noté, différents indicateurs ont été calculés, qualifiant cette fois-ci l'importance de chaque besoin au sein de groupes d'utilisateurs (et plus individuellement) :

- la somme des notes des utilisateurs de **chaque profil d'utilisateurs** (une note par profil et par besoin).
- deux notes **globales** :
 - la somme de toutes les notes (moyenne sur les répondants x nb d'occurrences). Dans ce cas l'absence de prise de position sur un besoin ne compte pas (une note par besoin).
 - la moyenne sur tout l'échantillon x nb d'occurrences. Dans ce cas l'absence de prise de position sur un besoin compte comme une note de 0 (une note par besoin).

Les deux notes globales présentées ci-dessus classent les besoins sensiblement dans le même ordre, mais leur répartition se fait de manière différente : la première méthode « étale » plus les besoins alors que la deuxième discrimine mieux les besoins dont la note est plus élevée (donc ceux pour lesquels il y a eu le plus de réponse) et est très sensible à l'absence de réponse, qui compte alors comme une note de 0. Dans la suite

du document, la « note globale » désigne la première des deux méthodes (somme des notes), choisie pour prioriser les besoins.

La note par profil met en évidence **quels sont les profils d'utilisateurs impliqués dans la note globale**. On peut ainsi déterminer si la note globale d'un besoin est élevée car il est important pour tous les utilisateurs ou si c'est à cause d'un groupe ayant beaucoup de poids. Cette note fait ressortir les différences entre les besoins des groupes, tout en prenant en compte les variations d'effectifs.

Une fois les notes attribuées, on priorise les besoins ainsi :

- les **besoins prioritaires**, dont la note est supérieure aux deux tiers de la note maximale,
- les **besoins secondaires**, dont la note est supérieure au tiers de la note maximale,
- les **besoins accessoires**, dont la note est inférieure au tiers de la note maximale.

Enfin, les besoins prioritaires et secondaires de chaque groupe d'utilisateurs ont été analysés en fonction de leur profil et leurs usages.

d) *Réalisation des personae*

A partir de l'analyse des besoins, j'ai construit cinq personae abordant les thématiques suivantes :

- Le rapport à la botanique
- Le rapport à la technologie et au numérique
- Le cadre d'utilisation de Flora Data
- L'investissement sur le site
- Les envies et attentes
- Les freins et les craintes

Pour chaque persona, j'ai imaginé une histoire illustrant les témoignages recueillis lors des entretiens. Les items présentés ci-dessus permettent d'aborder les thématiques importantes et de comparer les différents profils sur une base commune et parlante.

Deux personae ont été rajoutés *a posteriori* afin de prendre en compte des profils un peu différents qui auraient pu être oubliés. Cependant, à la différence des cinq autres, ils ne sont pas issus de témoignages directs d'utilisateurs mais de l'expérience de Daniel Mathieu et moi-même.

2. Les besoins des utilisateurs de Flora Data

a) *Résultats de l'analyse*

Au total entre janvier et février, 16 personnes et structures sur la quarantaine sollicitée ont été interviewées. Les cinq profils d'utilisateurs définis au préalable ont été conservés pour l'analyse. Certains utilisateurs ont cependant changé de groupe à l'issue des entretiens. Six catégories de besoins ont été identifiées lors de l'analyse des entretiens :

- **Les objectifs de la participation** : l'enrichissement personnel, se rendre utile ou trouver du contenu,
- **Le fonctionnement en réseau**,
- **Le stockage de données**,
- **L'utilisation de données** (point de vue outils),
- **L'aide à l'identification taxonomique**,

- **L'accessibilité aux outils** en général,

Au sein de ces catégories, huit besoins sont ressortis comme étant prioritaires, c'est-à-dire qu'ils cumulent des notes hautes et une forte occurrence lors des entretiens. La moitié des besoins concerne l'identification de taxon, thématique qui regroupe tous les profils d'utilisateurs. Ces besoins sont :

- **Vérifier l'une de ses identifications**
- **Identifier une espèce observée et photographiée**
- **Chercher dans un recueil en ligne**
- **Être aidé par le réseau**

Le deuxième besoin le mieux noté est le **stockage d'observations illustrées**, cependant ce n'est pas un besoin unanime : il est prioritaire pour les utilisateurs experts et les collectifs, c'est-à-dire lorsque les usages sont (ou se rapprochent) du domaine professionnel. C'est un besoin secondaire pour les utilisateurs avertis (et les animateurs), accessoire pour les débutants.

Les motivations à participer font également partie des besoins prioritaires avec :

- **Trouver des informations botaniques de source sûre**, avec là encore un poids très fort des utilisateurs experts, moindre pour les botanistes avertis chez qui ce besoin est secondaire.
- **Se faire plaisir**, prioritaire pour les utilisateurs débutants et avertis, secondaire pour les botanistes experts.

Enfin, le besoin d'**appartenir au réseau et d'être en lien avec les autres** se retrouve chez les utilisateurs experts (prioritaire) et avertis (secondaire).

Les besoins des débutants en botanique

Deux utilisateurs débutants seulement ont été interrogés. Amplifiés par ce faible échantillon, deux biais peuvent fausser le classement des besoins : le premier lié au fort poids de chaque témoignage, le deuxième lié à l'interprétation du témoignage de l'utilisateur lors de l'analyse.

Les besoins (prioritaires et secondaires) des utilisateurs débutants concernent l'enrichissement personnel, l'aide à l'identification et la dimension pratique/facile dans l'utilisation des outils :

L'apprentissage et le **plaisir** font partie des besoins prioritaires, et la **satisfaction de la curiosité** des besoins secondaires : ce sont les premières raisons de la participation à Tela Botanica des utilisateurs débutants en botanique. Ils sont souvent arrivés sur le site en cherchant des informations sur une plante, et ont porté un grand intérêt au MOOC. Intéressés par le monde naturaliste, ils possèdent des connaissances de base en botanique et souhaitent continuer à découvrir ce domaine en utilisant de temps en temps les ressources et outils de Tela Botanica.

Dans ce cadre, les utilisateurs ont besoin de trouver de l'aide pour **identifier des plantes observées et photographiées** (*a posteriori* plutôt que sur le terrain), en **comparant leur observation à des photos**, en **cherchant dans un recueil en ligne**. Toutefois, ils ne sont pas toujours dans l'attente d'être accompagnés par le réseau, n'ayant pas encore découvert (ou osé découvrir ?) toutes les fonctionnalités du site.

En effet, la compréhension du fonctionnement du site et des outils est un élément bloquant dans l'usage des utilisateurs débutants. Ils attendent des **outils faciles à utiliser**, plus **ergonomiques** et **mieux reliés entre eux**.

Les besoins des utilisateurs avertis, participant aux sciences participatives

Quatre utilisateurs interrogés ont été classés dans ce groupe.

Les utilisateurs répondant à ce profil ont une bonne pratique de la botanique, parfois dans un domaine précis, et sont souvent des passionnés du domaine naturaliste, de botanique ou encore de photo. On retrouve donc le **plaisir** dans les besoins prioritaires de ces utilisateurs. **L'apprentissage** et la **curiosité** sont eux aussi importants, bien que dans les besoins secondaires. Certains ont participé au MOOC, ce qui témoigne d'une volonté de continuer à se former malgré un niveau assez avancé en botanique. En lien avec cette démarche, les utilisateurs avertis souhaitent avoir accès à des **informations en botanique de source sûre**.

Leur pratique est régulière et concerne principalement l'identification, qui est quasiment l'unique thématique de leurs besoins prioritaires. Tous viennent chercher sur le site de Tela la possibilité de **vérifier la validité de leurs observations**, en s'appuyant sur un **recueil en ligne** ou grâce à **l'aide du réseau**, qui a ici une place très importante. Ils cherchent également à **identifier des espèces photographiées**, chez eux en les **comparant à d'autres photos** ou **directement sur le terrain**. Une partie de ces utilisateurs se sert également de **flores papiers** et **aident les autres membres du réseau** à l'identification lorsqu'ils le peuvent.

Le **stockage d'observations illustrées** fait également partie des besoins des utilisateurs de ce groupe, avec une grande place de la **photo** dans cette démarche. En effet, c'est elle qui leur permet ensuite de retrouver et réutiliser plus facilement une observation faite précédemment, pour vérifier une nouvelle identification par exemple.

Ces utilisateurs sont souvent en lien avec d'autres personnes dans le cadre d'associations, de groupes d'activité et/ou participent à des inventaires (pour des communes par exemple). Dans ce cadre ils ont besoin d'**exporter leurs propres données** pour les partager, en faire des comptes-rendus ou les mettre en forme.

Enfin, on retrouve l'**ergonomie** des outils dans les besoins secondaires. Les utilisateurs avertis arrivent à se servir des outils mais leur manque d'ergonomie, leur lenteur ou encore la présence de bugs posent parfois problème. Ils souhaitent que les outils soient **mieux liés entre eux** et plus **agréables à utiliser**.

Les besoins des experts en botanique

Public historique de Tela Botanica, les botanistes experts ont été les plus disponibles pour les entretiens. L'échantillon se compose de cinq personnes dont le profil correspond à ce groupe.

On constate que le cadre d'utilisation des utilisateurs experts en botaniques est souvent en lien avec leur activité professionnelle ou associative. De ce fait, leurs motivations sont différentes des deux groupes précédents : alors qu'on ne retrouve le **plaisir** et la **curiosité** qu'en besoins secondaires, les objectifs sont principalement la **recherche d'informations botaniques de sources sûres** ainsi que la **contribution à la richesse et l'amélioration du site**. Dans ce cadre, **l'appartenance au réseau** et le **partage de données et de connaissance** prennent une place très importante parmi les motivations des utilisateurs.

Les activités de ces botanistes les amènent à chercher des ressources pour **vérifier leurs identifications** mais aussi **identifier des observations photographiées**, en cherchant dans un **recueil en ligne** ou en **s'appuyant sur le réseau**. **L'identification sur le terrain** se retrouve dans les besoins secondaires, toutefois moins importante que pour les utilisateurs « curieux des plantes » - moins experts.

D'autre part, leurs connaissances leur permettent d'**apporter de l'aide aux autres utilisateurs** concernant l'identification d'espèces photographiées. Cela dit, d'après leurs témoignages cette pratique est fortement

influencée par leur disponibilité et par l'activité qu'ils observent sur le réseau : s'ils n'ont pas l'impression d'**être utile** à l'auteur de l'observation ou au réseau, ils participent moins.

Le **stockage de données** fait partie des besoins prioritaires, avec là encore la nécessité presque systématique d'y associer des photos. En effet, elles constituent un support de partage adapté aux connaissances et aux données botaniques. Ces utilisateurs sont également amenés à **exporter leurs données** pour les réutiliser, et ont notamment besoin d'**exporter des photos**.

Enfin, l'**ergonomie** et la **simplicité** des interfaces reviennent comme des besoins respectivement prioritaires et secondaires.

Les besoins des structures et groupes utilisant Flora Data dans le cadre professionnel ou associatif

Trois structures correspondant à ce profil ont répondu à l'entretien. En plus du petit échantillon, la diversité d'activités et de fonctionnement de ces trois organismes n'a permis d'identifier qu'un faible nombre de besoins prioritaires et secondaires.

Le besoin de stockage et de mutualisation des données est toutefois très bien mis en évidence par ces structures, qui doivent gérer un grand nombre de données naturalistes. Cela implique le besoin de **stocker des observations non illustrées** et parfois accompagnées de photos, de **les rendre accessibles par un groupe de personnes** et de pouvoir y accéder ensuite pour les **exporter** (souvent en masse).

La **fiabilité des outils de saisie** comme **d'export** font donc partie des préoccupations des structures qui les utilisent.

Les besoins des animateurs et enseignants utilisant Flora Data dans le cadre professionnel

Pour ce dernier groupe, seule une personne correspondant vraiment au profil a pu être interrogée. Cependant, son activité professionnelle se répartissant entre de l'animation dans le cadre d'une association d'une part et de la formation dans une classe de BTS de l'autre, son témoignage représente déjà plusieurs facettes des besoins de ce profil d'utilisateurs. L'apport de certains témoignages d'autres utilisateurs a permis de cerner un peu mieux les besoins. Il ne faut pas oublier toutefois que l'analyse qui suit repose en grande partie sur les usages d'une seule personne.

Les besoins ressortant de manière évidente sont l'accès à des **supports pédagogiques et d'animation** en lien avec la botanique, et la dimension **libre** du contenu. L'animateur effectue ses recherches d'informations sur la flore **en ligne**, plutôt que dans une flore papier.

D'autre part, dans le cadre de l'éducation à l'environnement, l'animateur cherche à promouvoir auprès de son public la **contribution aux sciences participatives**. De ce fait, son approche n'est pas orientée vers le stockage de données destinées à être réutilisées mais plutôt vers l'apport de données à la communauté scientifique dans le cadre de programmes locaux ou nationaux, lui permettant de sensibiliser le grand public.

Enfin, l'**existence et l'appui du réseau** a une certaine importance, notamment dans le cadre de **l'aide à l'identification**. En effet, tous les animateurs n'étant pas nécessairement experts en botanique, le réseau constitue une ressource rassurante.

Cette analyse a permis d'identifier plusieurs profils d'utilisateurs de Flora Data. Cependant, sans parler de leur faible nombre, les répondants ont pour point commun une certaine implication dans leur participation

à Tela Botanica, et ne sont de ce fait pas tout à fait représentatifs de l'ensemble des utilisateurs. C'est d'ailleurs visible au niveau de la répartition des entretiens entre les différents profils : beaucoup plus d'utilisateurs avertis (et souvent passionnés !) et experts en botaniques ont répondu que de débutants, qui connaissent mal l'association et sont peu impliqués.

b) *Personae*

Les résultats précédents ont conduit à la construction de personae, qui permettront par la suite de mieux confronter l'analyse des besoins avec les choix en termes de développement. Deux exemples sont présentés en Annexe H.

Bilan : L'objectif de cette étude était de se rapprocher des utilisateurs dans le cadre du travail de réflexion autour du processus Flora Data et de ses données. L'identification de leurs usages permettra à l'association d'orienter ses choix en termes d'amélioration du système et des données, sans perdre de vue les besoins réels des utilisateurs de Tela Botanica. Concrètement, cette analyse va permettre d'identifier et de prioriser les fonctionnalités à mettre en place ou à améliorer pour optimiser Flora Data.

L'analyse des besoins concernant les données a finalement fait l'objet d'une autre étude, visant à définir les critères auxquelles elles doivent répondre pour être valorisables (présentée au I.A.2 - p. 24).

D. Analyse fonctionnelle

1. Méthodologie

Objectifs : identifier les fonctionnalités répondant aux besoins des utilisateurs et les comparer avec l'existant. Définir des axes de travail prioritaires à partir de cette analyse.

Un certain nombre de fonctionnalités ont été identifiées en raisonnant à partir du fonctionnement actuel, des cas d'utilisation évoqués par les utilisateurs sondés ou encore des remarques de l'équipe.

Pour chacun des profils d'utilisateurs, les besoins prioritaires et secondaires ont été associés aux fonctionnalités correspondantes. Celles-ci ont été réparties en trois groupes : **inexistante**, **améliorable**, **satisfaisante**, selon leur état dans le système actuel. Cette répartition est illustrée par cinq diagrammes (un par profil d'utilisateurs) mettant en évidence quels besoins sont les mieux ou les moins bien satisfaits actuellement.

Un tableau de synthèse réunit ces informations pour l'ensemble des profils et des besoins. Chaque fonctionnalité a ensuite été notée en fonction de :

- son **importance pour les utilisateurs**, de 0 - absence de besoin à 3 - besoin prioritaire.
- son **influence sur la qualité des données**, de 0 - pas d'influence à 3 - influence directe, rôle déterminant dans la construction et la structuration des données.
- des **risques en cas d'absence**, de 0 - pas de conséquence sur le système et l'utilisation à 3 - risque fort : stoppe l'utilisation, entraîne la perte de l'utilisateur immédiatement ou à long terme.

Les fonctionnalités dont la note totale est supérieure ou égale à 7 sont considérées comme **indispensables**, celles dont la note est comprise entre 4 et 6 sont **nécessaires** et celles dont la note est inférieure à 3 sont **complémentaires**. On obtient ainsi un classement permettant d'identifier les fonctionnalités les plus

critiques : à la fois prioritaires pour les utilisateurs et pour la qualité des données, dont l'absence présente un risque important et qui sont absentes ou à améliorer.

2. Résultats

Un exemple de diagramme ayant servi à mettre en lien besoins et fonctionnalités est consultable en Annexe I. Le classement final (cf. Annexe J) confronte une note sur l'importance de chaque fonctionnalité (pour l'utilisateur, pour les données et vis-à-vis des risques en cas d'absence) à l'état de celle-ci dans le système.

Parmi les fonctionnalités « indispensables », on ne retrouve que deux fonctionnalités actuellement inexistantes : le **signalement d'une photo non identifiable** et le **tag des observations et photos de mauvaise qualité**. Ces deux fonctionnalités se rejoignent puisqu'elles visent toutes deux à caractériser les données de mauvaise qualité afin de mieux les gérer. On retrouve d'ailleurs également la thématique de la fiabilité des illustrations des taxons (sur eFlore) dans les fonctionnalités indispensables avec **l'affichage en premier sur les fiches espèces de photos de référence**, validées par le réseau, et le **filtre des photos douteuses** (grâce aux tags, aux retours du réseau sur IP (IdentiPlante), etc.).

Un certain nombre de fonctionnalités indispensables sont qualifiées d'améliorables. C'est le cas notamment de celles ayant trait à la saisie d'observations. On comprend aisément que ces fonctionnalités soient les mieux notées puisqu'elles constituent le cœur de Flora Data. Il est donc prioritaire d'améliorer l'ergonomie et la simplicité de la saisie et des fonctionnalités associées : **associer/dissocier une ou plusieurs photo(s)** à une observation, **modifier et dupliquer** une observation, **choisir parmi plusieurs niveaux de fonctionnalité** pour la saisie, disposer de **liens directs vers l'aide et les tutoriels** depuis les outils. Tout ce qui concerne **l'échange avec les autres utilisateurs pour l'identification d'un taxon** (vote, proposition de nom, commentaires) fait également partie des fonctionnalités indispensables.

Les fonctionnalités « nécessaires » comptent un nombre un peu plus élevé de fonctionnalités n'existant pas sur le site, qui concernent surtout des détails du fonctionnement de certains outils (**annuler un vote sur IP**, **flouter une donnée d'observation** lors de la saisie, avoir des **critères de recherche supplémentaires** dans eFlore). On ne retrouve pas parmi ces fonctionnalités de grandes thématiques comme précédemment, ce sont essentiellement des fonctionnalités facilitant le fonctionnement général des outils comme **ajouter plusieurs photos à la fois**, **gérer la diffusion de ses données**, les **caractériser à l'aide de tags** ou encore **créer des champs de saisie personnalisés**.

Enfin, on constate que la majorité des fonctionnalités absentes du site sont de moindre importance (note inférieure à 3/9). Celles-ci concernent le système de notifications, l'affichage de statistiques mettant en avant la participation, la possibilité de diffuser ses données au sein d'un groupe restreint... Ces fonctionnalités sont généralement mal notées car elles n'ont pas d'impact direct sur la qualité des données et (en partie car elles sont absentes !) ne sont pas ou peu envisagées par les utilisateurs.

Bilan : Ce classement permet de mettre en évidence les fonctionnalités et les processus sur lesquels se concentrer et d'identifier les thématiques prioritaires : la saisie, la gestion de la qualité des données et l'aide à l'identification.

Ces résultats ont donné lieu à une discussion sur les priorités en termes de spécifications. A ce stade, le nouveau site internet qui devait sortir plusieurs mois auparavant ne l'était toujours pas, retardant les

autres développements prévus. Ce facteur ajouté aux objectifs stratégiques de l'association m'a amenée à m'intéresser au format d'une donnée d'observation « idéale » et aux solutions pour y arriver.

E. Spécifications

1. Méthodologie

Objectifs : définir les modalités de développement des fonctionnalités à développer.

Une fois les fonctionnalités prioritaires identifiées, un travail de spécifications est nécessaire. Plusieurs approches ont été envisagées pour décrire les fonctionnalités. C'est finalement un travail à l'échelle du processus qui a été adopté, à l'aide des méthodes suivantes : diagrammes UML, user story, fiche descriptive de processus, maquettage, ... La méthode et les réalisations ont varié en fonction des processus étudiés mais un exemple représentatif est développé en troisième partie.

2. Définition d'un format standard de données

A partir des recommandations faites par le SINP concernant le format des données destinées à être échangées, j'ai cherché à **définir les concepts de base** employés lorsqu'on parle d'observation botanique, dans l'objectif de fixer le format standard des données de Tela Botanica. Pour cela, il a fallu tenir compte des spécificités des données de l'association vis-à-vis du format proposé par le SINP.

a) Les spécificités des données de Tela Botanica

Afin de définir le format des données d'observation, un certain nombre de particularités propres à l'association doivent être prises en compte :

Diversité dans l'origine des données : Tela Botanica peut être considérée comme une structure productrice de données, cependant l'association agrège déjà des données d'origines différentes : des structures (associations, bureaux d'étude, etc.), des professionnels de la botanique, des débutants, des amateurs... saisissent des observations botaniques dans la base de Tela. **Ce public a donc des exigences et des connaissances hétérogènes, ce qui produit des données hétérogènes.** Une partie des données est aussi issue de partenariats, en particulier dans le cadre du projet Floris'Tic, et constitue un flux plus ou moins automatisé de données dont le format d'origine n'est pas contrôlé par Tela. Tout cela souligne l'importance de **qualifier la source et l'origine de la donnée** dans le but de ne pas « mélanger » toutes ces observations hétérogènes.

Diversité du public visé : Comme nous l'avons vu, le public visé par l'association est hétérogène. Afin de convenir aux différents profils d'utilisateurs, les interfaces (de saisie notamment) doivent être **adaptables aux différents niveaux**, en botanique comme en informatique. Cette contrainte pèse aussi sur la quantité et la précision des informations demandées, que tous les utilisateurs ne sont pas à même de renseigner.

Processus de validation participative : Une grande partie des données étant saisie par des particuliers dans le cadre d'activités de loisir, elles sont rarement associées à un « validateur » référent comme dans une structure professionnelle. En revanche, une validation *a posteriori* peut être effectuée par le réseau sur IdentiPlante, lorsque l'observation est illustrée de photos. Dans ce cas, les validateurs sont les membres du réseau qui proposent ou confirment l'identification, et c'est l'auteur de la donnée qui est responsable du nom sélectionné au final.

La notion de donnée validée repose donc, dans le cadre de l'association, sur le nombre de membres du réseau en accord avec une identification. Cette information peut avoir sa place dans les métadonnées puisqu'elle qualifie en partie la qualité de la donnée.

Portée internationale : le fait de pouvoir saisir des observations depuis le monde entier a plusieurs conséquences :

- sur les **données de localisation** : la notion de commune avec son code INSEE est française. Elle a pour avantage de couvrir l'ensemble du territoire et d'être bien référencée. Mais on ne peut pas l'utiliser pour les observations à l'étranger et le terme employé jusqu'à présent est localité ou zone géographique, même si le concept est un peu différent. D'autre part, plusieurs localités pouvant être synonymes, une autre information doit y être associée : en France c'est généralement le département, mais on peut penser au pays, aux coordonnées géographiques, etc.
- sur les **référentiels taxonomiques** : Tela Botanica propose différents référentiels, couvrant plusieurs régions du globe. Le référentiel doit être renseignable lors de la saisie et obligatoire si l'utilisateur renseigne un nom scientifique. En effet, dans le cas où un utilisateur saisit un nom complet avec auteur sans sélectionner de référentiel, celui-ci ne sera pas considéré en tant que tel (pas reconnu dans eFlore, pas affiché dans les filtres par taxon, etc.).

Licence CC-BY-SA 2.0 : en lien avec les valeurs de Tela Botanica, cette licence s'applique notamment aux photographies, saisies via les outils de l'association et publiées par leur auteur. Les conditions d'utilisation selon cette licence sont les suivantes : les images peuvent être reproduites, distribuées, communiquées et modifiées, sous réserve de citer le nom de l'auteur original ainsi que celui de Tela Botanica qui édite l'œuvre. Dans le cas de modifications, transformations ou adaptations, le résultat ne peut être distribué que sous un contrat identique à celui-ci.

b) Proposition de définition

A partir des concepts proposés par le SINP et en prenant en compte les spécificités de Tela Botanica, il est possible de proposer un format standard pour les données d'observation botaniques. Celui-ci est repris comme définition d'une observation botanique en première partie (p. 8).

Bilan : Les préconisations pour la mise en place de ce format font l'objet d'un document complet ne pouvant être détaillé ici. Il a cependant permis d'identifier un élément clé de la saisie, difficile à résoudre simplement : la localisation. Pour cette raison, j'ai décidé de me concentrer sur les modalités du processus de localisation, afin de déterminer un protocole assurant à la fois la qualité des données (observations correctement localisées) et la facilité d'utilisation pour les utilisateurs, ceci, dans un contexte international.

Ainsi, la partie suivante sera consacrée à la description de la méthode employée et des résultats obtenus dans le cadre de l'étude de la mise en place d'un protocole de localisation. Elle sert également d'exemple afin d'illustrer la démarche suivie lors de la production de spécifications des différents processus.

III. Etude de cas : la (géo)localisation

Une observation botanique peut être localisée plus ou moins précisément, généralement à la commune ou au point. Le point est caractérisé par un couple de coordonnées géographiques, que l'utilisateur a enregistré directement sur le terrain ou qu'il a pointé sur une carte lors de la saisie. L'intérêt de localiser les données est multiple :

- Retrouver le spécimen observé (pour vérifier plus tard qu'il n'y a pas eu erreur d'identification, effectuer un suivi temporel, etc.),
- Retourner sur les lieux pour étudier l'habitat,
- Vérifier la cohérence de l'identification *a posteriori* (la localisation donne une indication supplémentaire sur ce que peut ou ne peut pas être l'individu observé),
- Cartographier la répartition des espèces,
- Suivre l'évolution géographique de certains taxons (dans le cadre d'études sur le changement climatique par exemple),
- Caractériser des habitats à partir d'observations existantes.

Ainsi, l'information de localisation est indispensable pour former une donnée botanique de bonne qualité et il convient de proposer aux utilisateurs des interfaces performantes et adaptées à la saisie de données correctement localisées. Cependant les modalités du protocole de saisie sont complexes à définir, d'où une étude dédiée. Dans ce cadre, on souhaite définir :

- Quelles sont les informations indispensables pour qualifier la localisation d'une observation botanique ?
- Quelles sont les fonctionnalités indispensables ?
- Comment optimiser la saisie de la localisation pour l'utilisateur ?
- Quelles seront les conséquences sur les interfaces de restitution des données ?

Par la suite, nous distinguerons **localiser**, qui consiste à donner une information géographique sur l'endroit, quelle que soit l'échelle (point GPS, rue, localité, pays, ...), et **géolocaliser**, qui implique d'associer une paire de coordonnées géographiques à la donnée.

A. Etat de l'art des solutions existantes

1. Méthodologie

La veille technologique décrite en deuxième partie m'a permis d'identifier des protocoles de localisation variés parmi les différents outils et sites dédiés à la saisie d'observations naturalistes. Après avoir testé les outils pour la saisie, j'ai sélectionné trois exemples me paraissant pertinents, à la fois en raison de l'importance du site en termes d'activité et pour la pertinence et/ou l'originalité de leur processus de localisation. Les protocoles permettant la localisation ont été décrits et comparés, y compris avec ceux du CEL et du widget de saisie rapide.

2. Etat de l'art : la localisation dans les outils de saisie naturalistes

Les trois exemples sélectionnés sont les suivants :

- **Biolovision** est un outil conçu par l'entreprise Suisse du même nom, initialement pour les observations ornithologiques et aujourd'hui utilisé par de nombreuses associations naturalistes (en Europe de l'Ouest surtout). C'est notamment l'outil qu'utilisent les sites départementaux de la LPO (36 bases locales) qui recensent la faune locale. Biolovision est également utilisé par un

site de collecte de données floristiques pour les orchidées, OrchiSauvage. L'outil est disponible sur mobile, via l'application Naturalist.

- **Artportalen**, site Suédois de collecte de données naturalistes, est développé et géré par le Centre Suédois d'Information sur les Espèces (ArtDatabanken) à l'Université Suédoise des Sciences Agricoles, au nom de l'Agence suédoise pour la conservation de l'environnement.
- **iNaturalist**, un site de la California Academy of Sciences, a une portée internationale et vise un public très large.

Le tableau ci-dessous résume les principales caractéristiques des trois outils. A titre de comparaison, Tela Botanica et ses deux outils de saisie ont été intégrés à l'état de l'art.

	Biolo vision	Artportalen	iNaturalist	Tela Botanica
Date de mise en ligne	2003	2001	2011	2000
Type d'observation	Faune surtout Flore : orchidées	Faune et flore	Faune et flore	Flore
Portée des données collectées	Départementale / LPO Nationale	Nationale (Suède)	Internationale	Internationale
Appli mobile	Naturalist	/	iNaturalist	/
Statistiques d'utilisation <i>(NB : ces chiffres ont été relevés entre avril et août 2017, afin de donner un ordre d'idée de l'activité des structures - mais ils évoluent vite !)</i>				
	Faune-Isère (Biolo vision)	Artportalen	iNaturalist	Tela Botanica
Inscrits	3 350	?	375 000	38 000
Contributeurs	?	30 000	129 500	?
dont en botanique	0	5 650	60 000	(tous)
Observations	1, 9 millions	57,8 millions	5,2 millions	1 million
dont botaniques	0	8,7 millions	1,6 millions	1 million
Espèces	?	?	115 500	8 900
dont floristiques	0	6 500	38 300	8 900

Tableau 1 : Caractéristiques et statistiques d'utilisation

Biolo vision

La saisie se fait en deux grandes étapes, la première consistant à choisir un lieu, la deuxième à saisir le reste des informations sur l'observation. De nombreuses alternatives de localisation sont possibles, aboutissant toutes plus ou moins au même résultat, ce qui entrave la compréhension et la fluidité à la première

utilisation. Il est possible notamment de saisir des coordonnées GPS ou de cliquer sur une carte, mais également de sélectionner un des (nombreux) « lieux-dits » proposés par le site. Ceux-ci sont l'équivalent de stations, ponctuelles et associées à une commune. Il est possible de naviguer sur la carte et parmi les lieux-dits mais aussi de saisir une commune ou un mot libre. Finalement, l'observation est toujours géolocalisée (coordonnées géo) et automatiquement associée à une commune française, non modifiable. Si un lieu-dit est à proximité (<710 m) du point saisi, il y est automatiquement associé. L'utilisateur n'a d'action possible que sur l'emplacement du point sur la carte. Sont restitués le lieu-dit, la commune et le département.

L'API de reverse geocoding semble être interne.

Avantages / inconvénients :

- + Visualisation des limites communales et administratives sur la carte zoomée, (calcul de l'altitude à partir des coordonnées + correction possible). A l'échelle locale, l'existence de lieux-dits facilite l'interprétation de la localisation.
- La saisie en deux pages successives (géoloc puis saisie) complexifie. La notion de lieu-dit n'est pas claire, on ne sait pas si on peut en créer d'autres. Interface fonctionnelle mais assez complexe.

Caractéristiques :

Qualité des données : protocole strict, comité de validation qui surveille les données

Facilité d'utilisation : moyenne à faible

Public visé : professionnel et amateur

Artportalen

L'interface de saisie est en une page, divisée en plusieurs sections, dont deux seulement affichées par défaut. La première section permet de saisir les informations sur l'observation, la deuxième permet de la localiser. La localisation se fait par « site » : un site = un nom, des coordonnées, une précision entre 5m et 5km. Chacun peut définir des sites (privés, publics, liés à un projet, à un autre site...), et toute observation est obligatoirement associée à un site (pas d'autre mode de localisation). Deux champs permettent la recherche : parmi les sites existants, sur la carte librement. L'utilisateur peut renseigner/modifier le site sélectionné (donc les coordonnées GPS), le nom du site et la précision de la localisation.

Les informations restituées sur la localisation sont : le nom du site, ses coordonnées, la précision, le comté, la municipalité et d'autres découpages administratifs (plutôt propres aux Etats-Unis).

Avantages / inconvénients :

- + Gestion par site permet de retrouver facilement un lieu déjà visité pour y ajouter d'autres observations, interface assez claire, renseignement explicite de la précision
- La notion de site peut compliquer la saisie, et oblige à effectuer des actions supplémentaires (créer un site)

Caractéristiques :

Qualité des données : protocole strict, certaines observations doivent être validées par des experts

Facilité d'utilisation : bonne

Public visé : professionnel et amateur

iNaturalist

L'interface permet de saisir plusieurs observations à la fois en complétant seulement les informations majeures. La localisation se fait via une pop-up permettant de placer un point sur une carte, de saisir des coordonnées, de renseigner une précision et d'éditer les « locality notes ». Ces notes sur la localité sont remplies automatiquement à partir des coordonnées GPS mais elles peuvent être modifiées ensuite. L'observation peut être soumise sans localisation mais un avertissement apparaît lors de la validation.

Cet outil utilise l'API Google pour le reverse geocoding.

Sont restituées principalement les infos saisies mais aussi un certain nombre de détails calculés à partir des coordonnées sur les « compassing places » soit littéralement les lieux englobant l'observation (« standards » : continent, pays, région, comté et « validés par la communauté » tels qu'un massif de montagne, une zone géologique, ...). On note toutefois que si la localité a été modifiée, le nom calculé par l'API n'apparaît pas.

Avantages / inconvénients :

- + Interface très simple, facile à prendre en main, renseignement explicite de la précision, saisie libre possible en cas de réponse imprécise ou fausse du reverse geocoding
- Pas d'information obligatoire, peu de garde-fou contre la saisie de n'importe quoi

Caractéristiques :

Qualité des données : protocole assez libre, vérification par les pairs

Facilité d'utilisation : très bonne

Public visé : amateur et grand public

Ces trois exemples donnent un aperçu de la diversité des protocoles que l'on peut mettre en place afin de décrire une observation naturaliste. Contrairement aux deux autres, iNaturalist fait le choix de ne pas utiliser la notion de station, lieu-dit ou site, ce qui simplifie la saisie mais limite potentiellement la précision de la localisation. En effet, si cela offre plus de souplesse à l'utilisateur qui peut modifier par exemple le nom de la localité, ce choix entraîne également un plus grand risque de saisie incomplète ou incohérente. Cependant, cela est à mettre en lien avec la portée internationale d'iNaturalist : la saisie des observations avec Artportalen et Biolovision étant limitée à un pays voire un département, il est plus facile de bien caractériser chaque lieu (avec par exemple le système commune / département en France), alors qu'un outil international doit s'adapter à tous les modes de découpage administratif possibles !

Widget de saisie rapide

L'interface permet uniquement de géolocaliser l'observation, via le pointage sur une carte ou la saisie de coordonnées. La saisie ou la modification de la localité ne sont pas possibles, mais l'utilisateur peut renseigner des infos plus précises sur l'emplacement comme le lieu-dit ou la station (« Lieu précis de l'observation définissant une unité écologique homogène »).

L'application utilise le web service de Tela « Nom-commune » qui donne, pour un point en France métropolitaine, la commune correspondante et son code INSEE. Dans cette interface, le géocodage inversé n'est pas fonctionnel pour des observations à l'étranger, bien qu'il ne soit pas mentionné que son usage est spécifique à la Métropole.

Les données utilisées par ce service sont celles d'OpenStreetMap (OSM), stockées dans une table de la base eFlore, en interne. Une routine quotidienne effectue la mise à jour d'une base intermédiaire (tb_osm) sous

la forme « brute » d'OSM. Cependant, le script qui devrait réaliser le passage au bon format des données mises à jour vers la table interrogée par le service n'est pas exécuté régulièrement et les données n'ont pas été actualisées depuis plusieurs années.

Le carnet en ligne

Le CEL offre de nombreuses alternatives pour localiser l'observation, qui sont redondantes et *a priori* difficilement lisibles. Notamment, deux cartes différentes (sur l'interface ci-dessous « Localiser la zone » et « localiser sur la carte ») permettent de définir un point précis, éventuellement après avoir saisi une localité auparavant. Le pointage sur une carte et la saisie de coordonnées entraînent la suggestion d'une zone géographique qui est reportée plus ou moins automatiquement dans le champs localité. Celle-ci est modifiable ensuite, il est d'ailleurs possible de valider son observation sans saisir aucune localisation valide.

The screenshot shows a web form titled 'Saisie' with the following fields and controls:

- Localité:
- Lieu-dit:
- Station:
- Milieu:
- Notes:
- Coordonnées: [Saisie Lat/Lon \(WGS84\)](#) [localiser sur la carte](#)
- Lat:
- Lon:
- Alt. (mètres):
- Identification:
- Date:
- Phénologie:
- Référentiel:
- Espèce:
- Abondance:
- [Ajouter un champ étendu](#)
- Buttons: Créer, Modifier, Supprimer, Réinitialiser

Figure 5 : Zone de saisie de l'interface du CEL

Le CEL envoie la requête de reverse geocoding à un troisième service : CoordSearch, également développé en interne. Celui-ci vérifie si les coordonnées saisies sont en France et appelle nom-commune si oui, zone-admin sinon. Zone-admin est un service qui fonctionne comme nom-commune mais pour l'étranger (également sur des données d'OSM hébergées en interne).

Bilan : Parmi l'ensemble des outils de saisie étudiés, on constate que souvent le lieu est une donnée obligatoire pour valider l'observation - ce qui n'est pas le cas pour les outils de Tela. La géolocalisation a globalement pris le dessus sur les autres formes de localisation, avec des variantes comme le passage par des stations enregistrées. Ce type de choix est conditionné par le public visé (en termes de niveau comme de répartition géographique), par l'outil et les objectifs du site en termes de qualité de données. Dans tous les cas, l'utilisation presque systématique de coordonnées comme localisation, parfois sans plus pouvoir renseigner soi-même la localité, implique l'utilisation d'un mécanisme performant pour le calcul des informations géographiques qualifiant le lieu et faisant l'objet de l'analyse suivante.

Les outils de l'association doivent continuer à satisfaire les utilisateurs habitués, mieux répondre aux besoins des nouveaux tout en garantissant la qualité de la donnée finale. De plus, s'il faut conserver la dualité saisie rapide vs. outil de gestion de données, le widget et le CEL doivent fonctionner de manière plus semblable et cohérente - d'où la nécessité de définir un protocole de saisie unique.

B. Tests comparatifs sur la solution la mieux adaptée : performance du reverse geocoding

1. Définition et intérêt du reverse geocoding

Le point, défini par un couple de coordonnées, est l'information la plus précise que l'on puisse obtenir pour renseigner un lieu mais aussi l'une des plus pratiques en termes de gestion. On peut notamment en déduire les entités géographiques dans lesquelles le point est inclus. Pour cela, on utilise un procédé appelé **géocodage inversé** ou **reverse geocoding**, qui consiste à calculer **une adresse à partir de coordonnées GPS** (le geocoding étant l'inverse, le calcul des coordonnées GPS avec une adresse).

On a constaté que, lors de la saisie, l'étape de géocodage inversé est importante à la fois pour l'amélioration de la qualité des données et pour la simplicité d'utilisation. Celle-ci fait appel à une API chargée de faire le lien entre les coordonnées et les entités administratives contenant le point.

2. Méthodologie

Afin d'évaluer les performances du service actuel de reverse geocoding de Tela Botanica et de le comparer avec plusieurs API existantes sur le marché (libres ou non), j'ai effectué des tests comparatifs (Goldberg et al., 2013) portant sur :

- leur **rapidité**, en mesurant les vitesses moyenne, minimale et maximale d'exécution de la requête,
- leur **exhaustivité**, en mesurant le taux de réponses non vides, et le taux de réponses complètes, c'est-à-dire avec la localité et le pays.
- leur **justesse** (les infos calculées correspondent-elles à la réalité ?) en comparant les réponses des services entre eux et en les comparant avec une référence.

Cette étape est réalisée en amont de la construction d'un cahier des charges afin de pouvoir prendre en compte les spécificités des services dans celui-ci, notamment leurs performances dans les zones peu peuplées à l'étranger.

a) Services de géocodage évalués

Il existe un grand nombre de services de géocodage inversé, pas tous recensés ici. Ceux qui sont listés ci-dessous sont comparés sur différents critères permettant de faire un premier choix pour les tests.

Fournisseur - nom du service	Geocoder si autre service	Source de données	Limite de gratuité (r = requête)	Libre	Testé (motif)
Google - Google Maps Geocoding API	–	Gmaps	2 500 r/jr, 50 r/sec + 0,5\$ /1000 r suppl Premium	Non	Oui (un des leaders du marché)
MapQuest - Geocoding API et Open Geocoding API	–	OpenStreetMap seul pour la version « Open ». Inconnus pour l'autre version	15 000 r/mois Plans payants	Non ~ Open data	Oui (1 geocoder / 2 sources de données)
Unwired Labs - LocationIQ	–	OpenStreetMap	10 000 r/jr, 60 r/min, 1r/s Autres tarifs (> 50\$/mois)	Open data	Oui
OpenCage - OpenCage Geocoder API	Nominatim Data Science Toolkit Twofishes	OpenStreetMap Yahoo! GeoPlanet Natural Earth Data GeoNames Flickr's shapefiles Thematic Mapping Quattroshapes Ordnance Survey OpenSpace MaxMind What3words (+ autres sources)	2 500 r/jr, 1 r/sec Autres tarifs (> 50\$/mois)	Open data	Oui (variété des sources de données)
Mapzen - Mapzen Search	Pelias	OpenAddresses Who's on First OpenStreetMap GeoNames	25 000 r/mois + 0,5\$ /1000 r suppl Premium support : 500\$ /mois	Open source Open data	Oui (geocoder libre, politique de confidentialité)

Gisgraphy - Reverse geocoding service	–	OpenStreetMap GeoNames Quattroshapes	Gratuit (API ou installation) Premium : à partir de 80\$ / mois	Open source Open data	Non (API non fonctionnelle au moment du test)
Data Science Toolkit - Coordinates to Politics	TwoFishes (GeoIQ)	OpenHeatMap Zillow US census information	Gratuit (API ou installation)	Open source Open data	Oui (Totalemt libre)
OpenStreetMap - Nominatim	–	OpenStreetMap	Gratuit (API ou installation)	Open source Open data	Non (utilisé par TB)
Tela Botanica - Service CEL : CoordSearch	Zone-admin Nom-commune	OpenStreetMap	Développés et hébergés en interne	Open source Open data	Oui

Tableau 2 : Présentation des services de reverse geocoding envisagés pour le test comparatif

Afin de comparer les différents services, il convient d'établir un jeu de coordonnées à « reverse geocoder ». Ces données doivent répondre aux critères suivants :

- Être **réparties sur tout le globe**, de préférence sur les **zones couvertes par Tela** (France puis Europe et pays francophones)
- Représenter des **régions très diversifiées** afin de pouvoir comparer la couverture et la précision des geocoder, dans des zones plus ou moins urbanisées par exemple,
- Être **assez nombreuses** pour être représentatives des performances du service

b) Construction du jeu de données test

Deux méthodes pour réunir ces coordonnées ont été suivies : générer automatiquement un jeu de données aléatoires en ligne ([Cagy](#)) ou les sélectionner par pointage sur une carte grâce à [Geoplaner](#). Dans le premier cas, la procédure est rapide et l'ensemble des points sont répartis de manière aléatoire (même s'il a fallu l'effectuer plusieurs fois pour ne pas avoir la moitié des points dans un océan ou à un pôle, ce qui a très peu de sens pour une observation botanique...). Dans le second cas, j'ai choisi l'emplacement des points de manière à m'assurer qu'ils sont répartis de manière plus pertinente.

Afin de mieux correspondre aux usages des telabotanistes, le jeu de test a été complété par les coordonnées de réelles observations botaniques issues de la base de données de Tela. Ces données ont également l'avantage d'être accompagnées d'une « référence » puisque la localité a normalement été saisie (ou au moins vérifiée) par l'utilisateur.

Une première requête a permis de sélectionner les données du CEL saisies en 2016 répondant à une série de critères assurant dans la mesure du possible leur fiabilité. Un peu plus de 12 000 observations ont été ainsi isolées, dont 95% sont situées en France. Deux jeux de test distincts ont été créés à partir de ces données, d'une soixantaine de données chacun : les observations faites à l'étranger d'une part et les données françaises de l'autre.

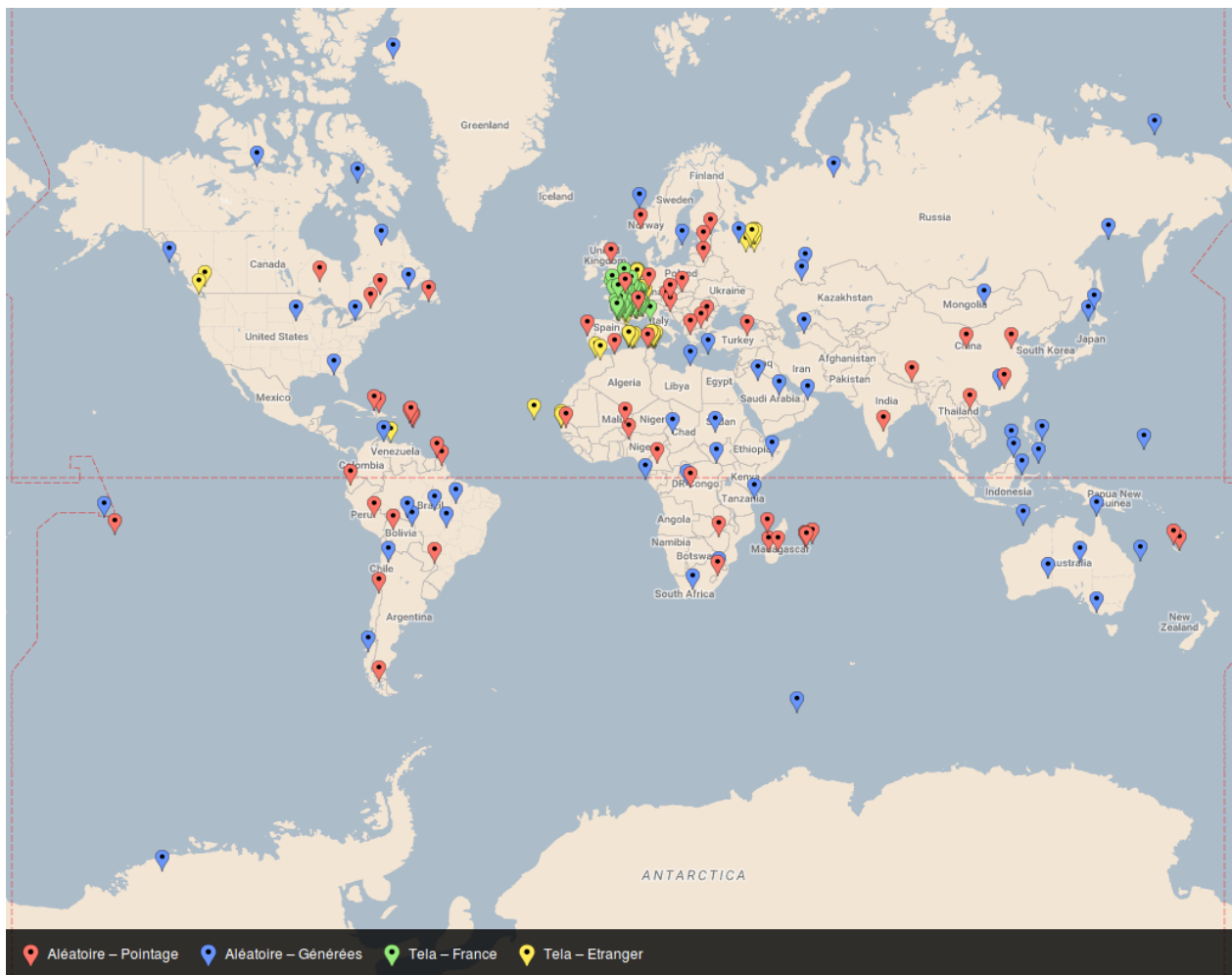


Figure 6 : Carte interactive illustrant la répartition des données testées (réalisée sur batchgeo.com)

La répartition de l'ensemble des données testées est illustrée par la carte ci-dessus. Les quatre jeux de données sont donc complémentaires car ils permettent de :

- **couvrir une large zone géographique** (cf. points bleus et rouges sur la carte) pour tester l'efficacité générale des services,
- **couvrir plus précisément la France, les pays frontaliers et francophones**, qui sont les plus concernés par l'activité de Tela (cf. points verts et jaunes)
- **disposer des données de référence** avec lesquelles comparer les résultats,
- avoir un **jeu de données assez important** (235 au total).

c) Interrogation des API et comparaison des réponses :

Un script PHP interroge les API pour chaque paire de coordonnées, puis les réponses et les temps de réponses sont enregistrées au format csv (une ligne = une paire de coordonnées). Une fois mises au même format, les réponses sont comparées deux à deux afin d'évaluer les différences de résultats d'une API à l'autre, puis il est fait de même avec la référence (dans le cas des données du CEL).

Après une remise en question de la fiabilité de la référence, un passage manuel sur les données de Tela a finalement été effectué pour vérifier la justesse « réelle » des données, à l'aide de divers services cartographiques en ligne. Cela a également permis de prendre en compte les variabilités d'orthographe ou de langue, qui peuvent être considérées comme des erreurs par le script mais qui n'affectent en réalité pas la précision du service. Les réponses des API ont été notées de 0 à 2 :

- 0 - réponse fausse : localité trop loin du point cible (plusieurs kilomètres), nom inconnu.
- 1 - réponse vague : échelle plus large que localité (région, département, comté, etc.), localité proche mais ne contenant pas le point.
- 2 - réponse juste : localité contenant le point ou frontière très proche (< 200 m ?).
- L'absence de réponse n'est pas prise en compte.

3. Zoom sur un aspect critique : le géocodage inversé

a) Résultat des tests comparatifs

Rapidité d'exécution et taux de réponse

Le **temps de réponse du service** joue un rôle important dans la performance de l'outil de saisie. Entre autres, sa lenteur peut entraîner l'abandon de l'utilisateur. Il en est de même pour le **taux de réponse** : à partir d'un certain taux de non-réponses, l'outil devient pénible à utiliser. Ce taux limite peut être différent selon la zone géographique concernée : la France, les pays européens et francophones étant la cible de l'association, il est légitime d'être plus exigeant pour ces zones qu'ailleurs.

L'API de Data Science Toolkit donnant quasiment toujours le pays mais pas de localité (cf. Annexe K), probablement en lien avec le fait que ses sources de données sont essentiellement américaines, il n'est pas pris en compte dans la suite de l'étude.

Ci-dessous les **temps d'exécution de la requête**, en secondes, ont été calculés sur les 235 requêtes.

Temps d'exécution de la requête (classés par temps moyen croissant)

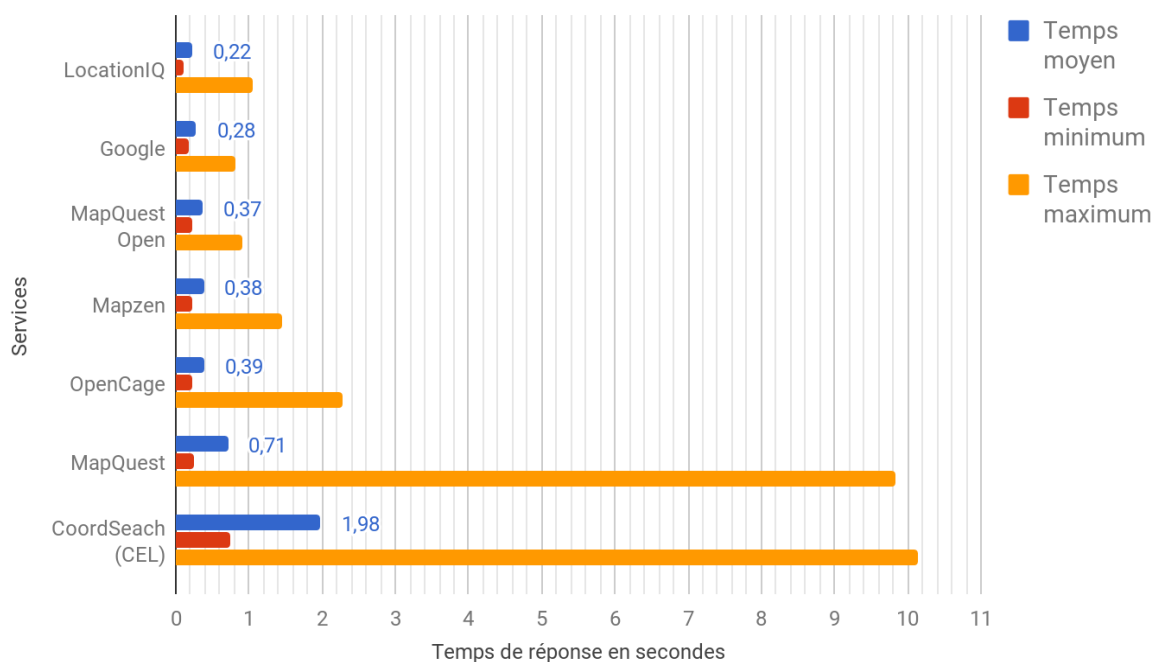


Figure 7 : Les temps d'exécution moyen, minimum et maximum des différents services sur l'ensemble des données

CoordSearch est de loin le service le plus lent, suivi par MapQuest dont le temps moyen est augmenté par plusieurs réponses assez longues (7 à 9 secondes). Les autres services ont des temps moyens et maximums satisfaisants.

Le graphique qui suit permet de détailler pour chaque jeu de données les **taux de réponses complètes** (c'est-à-dire contenant le pays et la localité au moins) de chaque service. Un graphique équivalent pour le taux de réponse (non vides) est présenté en Annexe L.

Taux de réponses complètes (au moins pays-localité) par service et par jeu de données

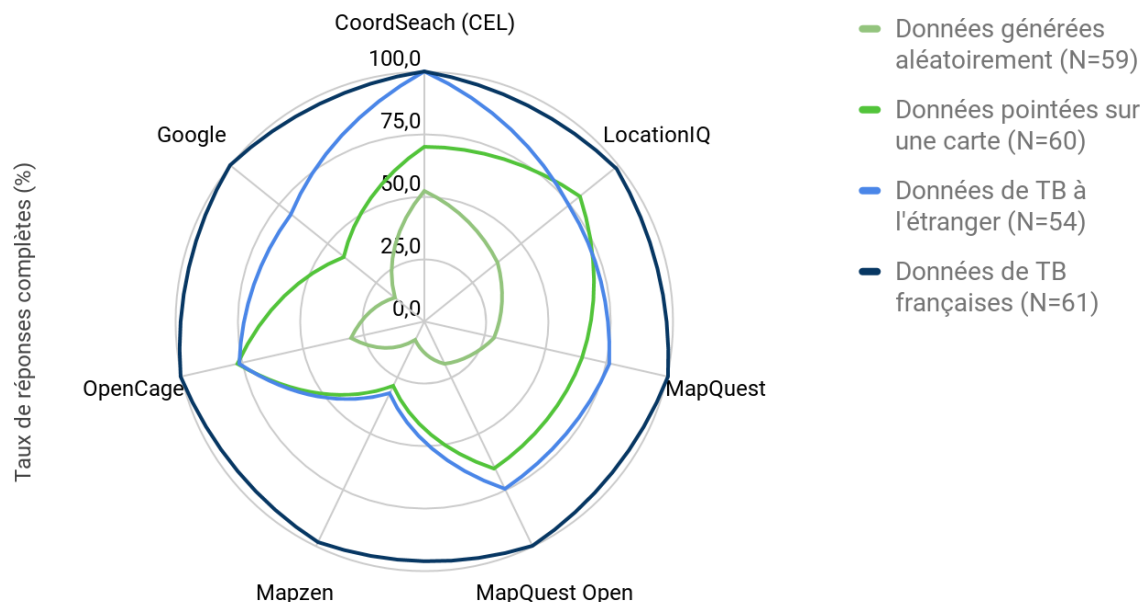


Figure 8 : Les taux de réponses complètes des différents services, par jeu de données

De manière globale, **tous les services renvoient 100% ou presque de réponses complètes pour les données françaises** de Tela. Pour les données CEL étrangères en revanche, on remarque que **CoordSearch est le seul à atteindre 100%** de réponses complètes. Mapzen est en dernière position avec moins d'une réponse complète sur trois, alors que les autres sont plutôt à 2 sur 3 (Google) voire 3 sur 4 (LocationIQ, MapQuest Open, MapQuest, OpenCage). Le jeu de données étant représentatif des lieux visités par les telabotanistes à l'étranger (voyages ou utilisateurs étrangers), ces résultats constituent un indicateur important pour choisir l'outil le mieux adapté aux besoins des utilisateurs.

Pour les données pointées, LocationIQ et OpenCage sont ceux qui renvoient le plus de réponses complètes, suivis de CoordSearch puis de MapQuest/Open. Pour le jeu de coordonnées « aléatoires », CoordSearch atteint également de bons scores relativement aux autres (52%), mais son taux de lignes vides est bien plus important que les autres (52% aussi). Il est d'ailleurs avant-dernier lorsqu'on classe les services par taux de réponse, avec près de 15% de lignes vides en moyenne. Google et Mapzen sont les moins complets pour ces deux jeux de données, avec bien moins de la moitié des réponses donnant pays et localité.

On constate ainsi que le service du CEL est déjà bien adapté aux besoins considérés : il est plus basique que les autres dans la mesure où il ne renvoie que la localité et le pays, mais il semble plus complet pour les données cibles de Tela Botanica. Cela dit, il est beaucoup plus lent et manque tout de même d'exhaustivité sur le pays dans le cas général des données aléatoires (information que les autres services renvoient).

Justesse des réponses

Les résultats précédents ne font que mesurer la présence ou l'absence d'une réponse quelconque, il faut maintenant **vérifier que ces réponses correspondent à la réalité**. Dans cet objectif, les réponses des services de géocodage inversé ont été comparées avec les infos saisies dans le CEL (localité et pays). Ces informations servent ici de référence, cependant on l'a vu, elles peuvent provenir soit de l'utilisateur qui les a saisies ou modifiées, soit du service de géocodage inversé du CEL. On peut considérer que l'intervention humaine dans ce procédé assure une certaine vérification des données de localisation, d'autant plus qu'il s'agit de données sélectionnées pour être complètes. Cela dit il faut garder à l'esprit que ces données peuvent être faussées si l'utilisateur s'est trompé de localité ou si CoordSearch a retourné une réponse fausse et que l'utilisateur ne l'a pas corrigé.

L'ensemble des services présentés ici se révélant fiables pour identifier le pays (pour les données de Tela françaises comme à l'étranger), nous nous intéressons ici à la justesse de la localité uniquement.

Ci-dessous les résultats de la comparaison des services à la référence. Le « taux de concordance » est calculé en comparant chaque couple localité d'un service - localité de référence, et en comptant le nombre de correspondances par rapport au nombre de données comparées (N).

Taux de concordance de la localité avec la référence en fonction du nombre de données comparées - Données françaises de Tela

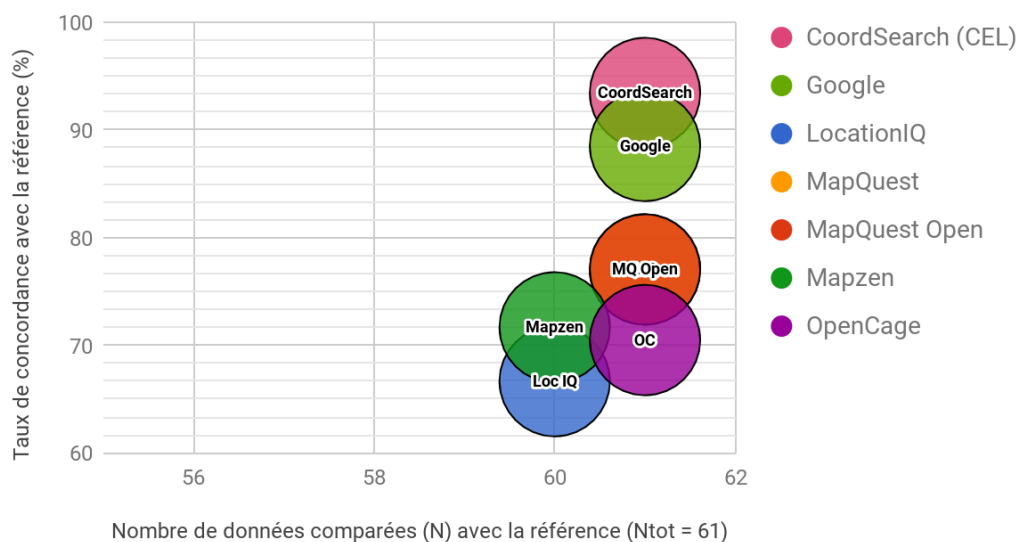


Figure 9 : Les taux de concordance de la localité entre la référence et les différents services pour les données françaises

CoordSearch montre, comme attendu, une forte correspondance avec la référence (> 93%). Google se détache également du lot avec plus de 88% de correspondance, puis MapQuest et MQ Open avec 77%.

Taux de concordance de la localité avec la référence en fonction du nombre de données comparées - Données étrangères de Tela

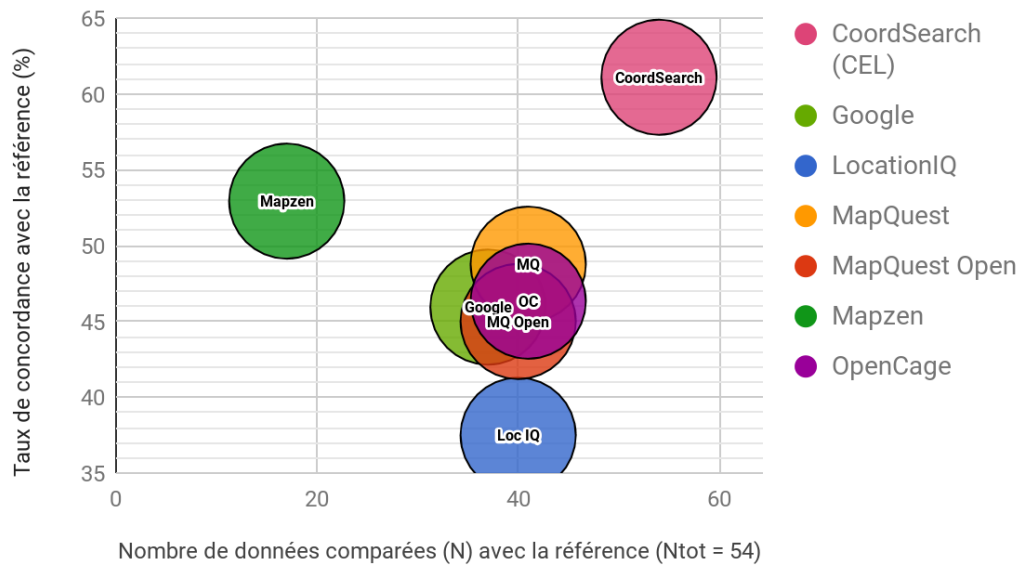


Figure 10 : Les taux de concordance de la localité entre la référence et les différents services pour les données étrangères

Le second graphique présente les mêmes résultats mais sur les données de Tela à l'étranger. On constate également que les réponses de CoordSearch sont celles qui s'accordent le plus avec la référence (61%). Mapzen a cette fois-ci le deuxième taux de correspondance le plus élevé mais sur un petit nombre de données comparées seulement. Les autres services ont globalement le même taux de concordance avec la référence (entre 45 et 50%), sauf LocationIQ qui est en-dessous.

En se penchant sur les données afin d'expliquer ces résultats, on distingue **plusieurs types de non concordance du service avec la référence** :

- **noms totalement différents** (ex : Roumoules vs. Moustiers-Sainte-Marie)
- **différence de langue** (ex : Geneva vs. Genève) ou même **d'alphabet** (Russe, Arabe, ...)
- fautes ou **variabilité dans l'orthographe d'un même nom**, notamment dans la gestion des accents et caractères spéciaux

Pour les données françaises, c'est le premier cas qui domine, avec jusqu'à trois localités différentes pour un même point. Difficile de savoir dans ce cas la cause de l'erreur ni quel service se trompe. Mapzen fait exception avec un taux de concordance sur les données françaises en grande partie influencé par l'absence d'accents dans ses réponses et non lié à des erreurs de localisation. Afin de pallier les biais d'écriture de ce type et de vérifier la fiabilité de la référence, **l'ensemble des données de Tela ont été vérifiées manuellement**. Ce travail a notamment permis d'identifier les causes de non concordance avec la référence :

- **le point est proche des limites administratives entre plusieurs localités**, il y a ambiguïté sur la localisation du point. Ce cas est beaucoup plus courant qu'on pourrait le penser (plus d'un tiers des données françaises !), particularité des données botaniques à prendre en compte.
- **l'utilisateur s'est visiblement trompé** en saisissant les coordonnées ou la commune. Dans l'un des cas, le point est placé à un peu moins de... 200 km de la localité renseignée, Limoges.
- **Le web service se trompe** par manque de précision ou d'exhaustivité de la source de données.

Pour les données françaises, les nouveaux résultats sont présentés sous la même forme que les précédents :

Justesse de la localité en fonction du nombre de données comparées - Données françaises de Tela

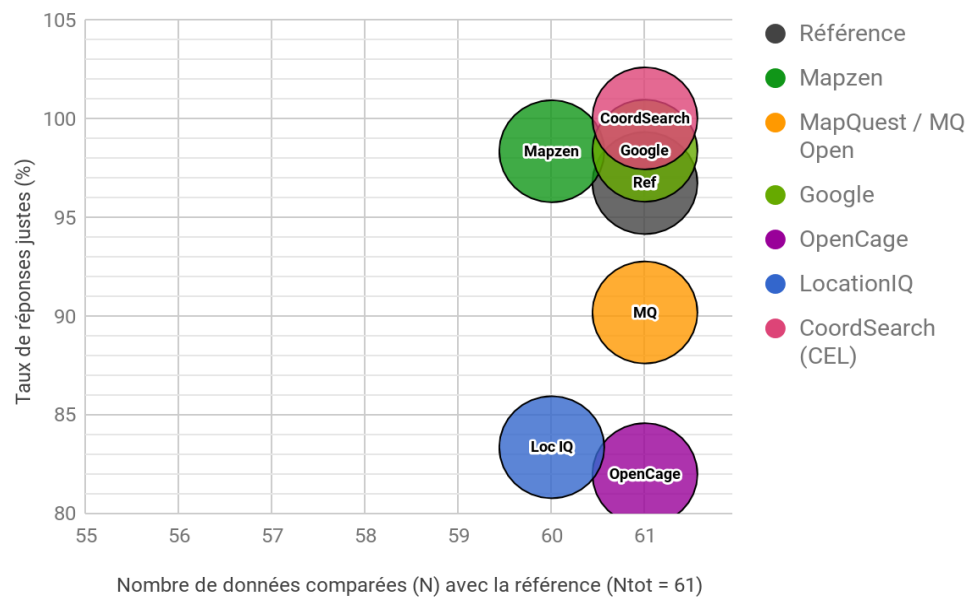


Figure 11 : Justesse de la localité répondue par les différents services pour les données françaises

Tout d'abord, l'évaluation de la référence montre que les données saisies/modifiées par les utilisateurs pour la France, sans être aberrantes (près de 97% de réponses exactes) sont **moins fiables sur ce jeu de données que les réponses « brutes » de CoordSearch**. Le service du CEL est en effet réellement performant sur la France avec 100% de localités bien identifiées. Mapzen, qui était dans les derniers pour les données françaises, apparaît en fait comme très fiable (cf. gestion des accents). MapQuest conserve sa position moyenne, Location IQ et OpenCage restent en dernière place avec tout de même plus de 80% de réponses justes.

Pour les données étrangères (cf. graphique ci-dessous), l'interprétation de la justesse des réponses est plus complexe à cause d'une différence du nombre de données comparées pour chaque service, d'où une représentation différente. D'autre part, on prend ici en compte les réponses « approximatives » : dans certains cas (localité proche, zone géographique plus large qu'une localité, cf. méthode p. 36), la réponse n'est pas considérée comme fautive (elle ne désigne pas un autre lieu) mais comme trop vague.

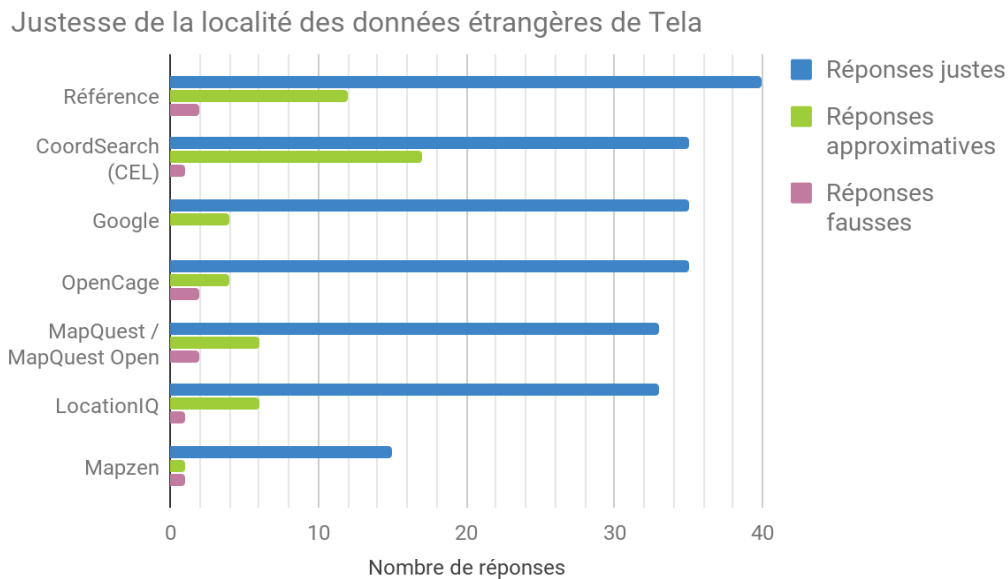


Figure 12 : Justesse de la localité répondue par les différents services pour les données étrangères

On constate qu'ici **les données de référence sont les plus justes**, avec 40 réponses exactes sur 54. On peut supposer que dans le cas des données étrangères, la « localité » proposée n'a pas satisfait les utilisateurs, qui l'ont modifiée pour qu'elle soit plus précise et/ou corresponde à leur connaissance du lieu. CoordSearch, Google et OpenCage ont tous les trois autant de réponses correctes (35), mais **CoordSearch les complète par un plus grand nombre de réponses approximatives**. Cela est en fait lié en grande partie à la manière dont CoordSearch extrait les informations que zone-admin lui envoie : lorsque le zone-admin n'a pas trouvé pas de localité, CoordSearch récupère le nom d'une entité de l'échelle supérieure. Par exemple pour une donnée située au Cap-Vert, zone-admin ne renvoie pas d'information plus précise que le nom de l'archipel et CoordSearch donne « Cap-Vert » comme nom de localité. Cela explique son exhaustivité quand les autres services ne trouvent pas de localité.

OpenCage, LocationIQ et MapQuest restent de précision équivalente, alors que Mapzen est dernier en lien avec son faible taux de réponse.

b) Recommandations sur le choix d'une solution

En conclusion, le service du CEL CoordSearch semble déjà bien adapté aux besoins de Tela Botanica. Il est réellement performant pour la France et fait partie des services les plus justes pour les lieux cibles de Tela à l'étranger. Il faut faire attention toutefois à bien comprendre son fonctionnement lors du traitement des réponses, et ne pas considérer l'ensemble des « noms » qu'il renvoie comme des localités (comme c'est fait dans le CEL) mais bien comme la zone géographique la plus fine que le geocoder ait pu trouver. Ceci explique son fort taux de réponse non vide de manière générale, et dans ce cadre la comparaison avec les autres services est faussée puisque, pour eux, seule la localité a été extraite. Cependant, ce biais n'a que peu d'impact dans l'étude finale de la justesse des données et on constate que la précision de CoordSearch n'est pas inférieure à celle de Google par exemple, qui est parmi les plus performants.

La raison du faible taux de réponse de CoordSearch pour certains pays (cf. résultats sur les données générées aléatoirement en Annexe L) s'explique également par le fonctionnement de CoordSearch, qui ne renvoie pas de données dans un cas particulier, non détaillé ici. Corriger ce « bug » permettrait de rendre le service plus exhaustif à l'étranger, au moins sur les pays.

Le principal défaut du service de Tela est sa lenteur et il conviendrait d'optimiser son fonctionnement pour le rendre plus rapide. C'est déjà pour cette raison qu'il interroge nom-commune et pas zone-admin pour la France, le premier service étant plus rapide (car sur un territoire restreint et bien connu) et les observations en France plus courantes. La vitesse de CoordSearch dépend également de la performance des serveurs de l'association, sur lesquels il est hébergé.

Enfin, il faut veiller à maintenir à jour les sources de données géographiques, qu'OSM actualise régulièrement, en exécutant les scripts prévus à cet effet.

L'objectif de ce test était également de vérifier si certains services permettent une meilleure précision en Afrique et dans les espaces naturels peu peuplés, où se font de nombreuses observations et qui sont souvent mal caractérisés par le reverse geocoding. L'analyse des données étrangères montre qu'actuellement, les services étudiés ne sont pas plus précis que celui déjà utilisé par le CEL et donc que pour ces zones, les utilisateurs sont encore les plus à même de saisir une information exacte. Cela souligne la nécessité de laisser aux utilisateurs une certaine souplesse dans la saisie, même si celle-ci est encadrée.

Dans le même ordre d'idée, il serait intéressant d'effectuer ce type de tests pour le calcul des données d'altitude, autre information importante en botanique et que l'on peut calculer à partir des coordonnées géographiques. Dans cet objectif, les premiers éléments de méthode ont été réunis en Annexe M.

Bilan : L'étude des performances du reverse geocoding a permis d'identifier des opportunités et des contraintes pour la saisie, qu'il est maintenant temps d'intégrer au sein d'un cahier des charges plus large. Celui-ci doit définir le protocole de saisie de la localisation.

C. Définition du cahier des charges

1. Méthodologie

Objectifs : déterminer les modalités de mise en place d'un nouveau protocole de saisie, c'est-à-dire notamment répondre aux deux premières questions posées en introduction : quelles informations et fonctionnalités doivent être gérées pour optimiser la saisie de la localisation.

a) Définition du protocole de saisie

Afin d'**homogénéiser les données saisies via les outils de Tela Botanica**, on souhaite s'accorder sur un protocole de saisie unique. Celui-ci définit l'ensemble des informations à saisir, leur statut obligatoire ou non et la manière de les renseigner (on se concentre toujours ici sur les données en lien avec la localisation). Il s'agit par exemple de trouver des alternatives afin de pallier les limites du reverse geocoding.

La construction du protocole de saisie est réalisée de manière à correspondre au format de données défini précédemment, après avoir étudié les possibilités les mieux adaptées au contexte (portée internationale mais centrée sur la France, utilisateurs de différents niveaux, sources de données variées, etc.).

b) Description de la procédure (diag UML)

Afin de décrire le fonctionnement de l'interface de saisie, j'ai à nouveau utilisé la modélisation UML. Ici les diagrammes d'activité permettent de décrire le fonctionnement du processus tel qu'il peut satisfaire le protocole de saisie.

c) *Maquette et scénario de saisie*

Les maquettes et scénarios de saisie permettent de donner un aperçu des fonctionnalités de l'interface proposée, mais aussi de vérifier que les utilisateurs, modélisés par les personae, sont capables de s'en servir et en ont l'usage.

2. Définition d'un (unique !) protocole de saisie

a) *Quelle(s) information(s) pour localiser une observation ?*

Les informations sans lesquelles il n'est pas possible de soumettre une observation doivent être assez précises pour permettre l'utilisation de la donnée ensuite (pour aider à l'identification d'une plante, signaler un taxon nouveau pour la zone, effectuer des études de répartition, etc.) mais pas trop contraignantes pour que la grande majorité des utilisateurs soient capables de les renseigner, y compris si la donnée est ancienne ou l'observation effectuée par quelqu'un d'autre par exemple. La solution proposée ici est de rendre obligatoire la saisie d'**une paire de coordonnées géographiques décrivant un point**. Dans ce cas, il faut :

- Utiliser une API, héberger ou développer un service de **géocodage inversé** (ce qui implique un coût, un temps de développement, des tests d'efficacité...). Nous avons vu pour ce point les modalités d'amélioration du service existant CoordSearch (cf. III.B.3.b)).
- Gérer la saisie d'un **indicateur de précision** : dans les cas où la localisation exacte n'est pas connue, l'utilisateur doit pouvoir renseigner un indicateur de précision (à 10, 20, 100 m près).
- Proposer une **interface cartographique performante** pour la géolocalisation.

Rendre possible ou non **l'édition du champ localité** est un choix important. Pour la France comme on l'a vu, CoordSearch suffit à déterminer la commune à partir des coordonnées. En revanche, étant données ses performances inégales à l'étranger, les utilisateurs doivent pouvoir modifier ou ajouter la localité de leur choix, si possible avec autocomplétion des noms. Cependant, comme pour le service de géocodage inversé, **les services d'autocomplétion ne sont pas toujours exhaustifs**. Gérer au cas par cas (en France/pas en France, précision ou non de la réponse, ...) rajouterait de la complexité, mais on souhaite tout de même **limiter la saisie totalement libre de données**. Un moyen intermédiaire est d'alerter les utilisateurs lorsqu'ils saisissent une information potentiellement erronée. Par exemple, un message d'avertissement peut s'afficher sous le champ localité si l'utilisateur choisit de se passer de l'autocomplétion, expliquant les risques et inconvénients de cette pratique. Une étude concernant les services d'autocomplétion, équivalente à celle menée sur les services de géocodage inversé, est à envisager pour décider si la saisie totalement libre doit être permise ou non.

Par ailleurs, on peut choisir de **conserver dans la base les deux réponses** : celle de CoordSearch et celle saisie par l'utilisateur. Cela permettrait d'une part d'effectuer des vérifications quant à la fiabilité des données saisies (toujours dans l'objectif de qualifier les données) mais également de mettre à jour régulièrement les données de localisation. Par exemple, si une zone auparavant mal couverte est mieux décrite dans OSM, on veut pouvoir profiter de cette amélioration pour préciser la localisation issue du reverse geocoding, sans pour autant écraser celle saisie par les utilisateurs.

Cette proposition de protocole impose également de **bien gérer l'information de précision**, en particulier : si une observation est géolocalisée à proximité de la frontière entre plusieurs communes (ce qui est un cas courant, comme montré précédemment), le rayon d'incertitude peut amener à placer l'observation dans une commune ou dans une autre. Dans ce cas, il faut déterminer la commune choisie pour la localisation : celle qui contient le point géolocalisé (le plus simple) ou au choix de l'utilisateur (mais cela implique de détecter

automatiquement le conflit, ce qui est plus compliqué). Afin à nouveau d'éviter d'augmenter la complexité du système, il est préférable d'opter pour la première solution, tout en gardant à l'esprit que d'une part l'utilisateur peut modifier la localité s'il le juge pertinent, d'autre part l'incertitude de la localisation est précisée dans les métadonnées et l'observation est donc **bien qualifiée**.

Le tableau suivant permet de comparer les avantages et les inconvénients d'utiliser les coordonnées géographiques plutôt que la localité seule :

"Localité" (et éventuellement pays) obligatoire	Coordonnées obligatoires
Moindre précision	Localisation très précise
Les noms de localité sont redondants, parfois même au sein d'un seul pays	Données robustes : non ambiguës (si référentiel géographique unique)
Les noms évoluent dans le temps	Données robustes : invariables dans le temps
Mise à jour des noms : à partir des noms déjà saisis	Mise à jour des noms : automatique à partir des coordonnées (cf. màj OSM)
La localité est presque toujours connue de l'auteur de l'observation donc facile à renseigner	Donnée très précise pouvant « faire peur » à l'utilisateur s'il n'est pas sûr de la localisation
L'auteur peut retrouver facilement ses observations par la suite car la localité est souvent liée à des souvenirs (vacances, randonnée, domicile, etc.)	Si on déconseille la saisie de la localité, on prend le risque de perdre l'utilisateur qui n'est pas d'accord avec les informations calculées
Si on permet à l'utilisateur de saisir à la fois une localité et des coordonnées, on prend le risque d'incohérence entre les deux	
Nécessite un index géographique fiable, complet, mis à jour régulièrement, fournissant des données non ambiguës pour une autocomplétion performante	Nécessite un index géographique fiable, complet, mis à jour régulièrement pour un reverse geocoding performant
L'absence de donnée de précision pose problème lorsque l'utilisateur connaît la localisation plus précisément qu'à la commune mais pas assez pour placer un point. Alternative actuelle : champs lieu-dit et station, qui posent des problèmes d'homogénéité.	Les données de précision entraînent des contraintes : <ul style="list-style-type: none"> • pour le choix d'une localité si proche d'une limite communale par exemple • pour la restitution des données (carto notamment)

Tableau 3 : Avantages et inconvénients de deux modes de localisation : localité vs. géolocalisation

b) Gestion du floutage

Dans le cas de données sur des espèces sensibles, les auteurs peuvent souhaiter rendre leur observation publique pour de nombreuses raisons (aide à la détermination, contribution à la richesse du site, espèce peu illustrée, etc.), tout en **masquant la localisation précise** afin d'éviter la dégradation des stations. Ce floutage est envisageable à la commune ou à la maille (5x5 ou 10x10 km).

Si floutage il y a, la localisation précise est conservée dans la base, accessible seulement par l'auteur de l'observation via son CEL. L'information de floutage est stockée elle aussi et sert pour la restitution dans toutes les autres interfaces (cf. D.2.c).

c) Champs complémentaires

Les champs complémentaires existants peuvent être conservés, afin de mieux qualifier la localisation et de permettre aux utilisateurs de conserver certaines habitudes :

- **Lieu-dit** : permet surtout à l'utilisateur de mieux se rappeler du lieu, ou aux autres utilisateurs de mieux l'identifier même sans carte.
- **Commentaires sur la localisation** (actuellement « station ») : autres précisions sur le lieu de l'observation. Ce champ est libre, d'où l'intitulé de "commentaires".
- **Milieu** : le type d'habitat, actuellement « plus ou moins standardisé ». Devrait être branché sur un vrai référentiel d'habitats (Corine ou autre, cf. [HABREF](#)) et bénéficier d'autocomplétion pour homogénéiser au maximum les données saisies, ou être supprimé.
- **Altitude** : cette information importante en botanique (influence sur les conditions climatiques, le milieu, etc.) pourrait être calculée par un web service à partir des coordonnées géographiques, et/ou complétée par l'auteur.

d) Description de la procédure : diagramme, maquette et scénario

On a décrit le protocole de saisie, il faut maintenant définir le fonctionnement de l'interface :

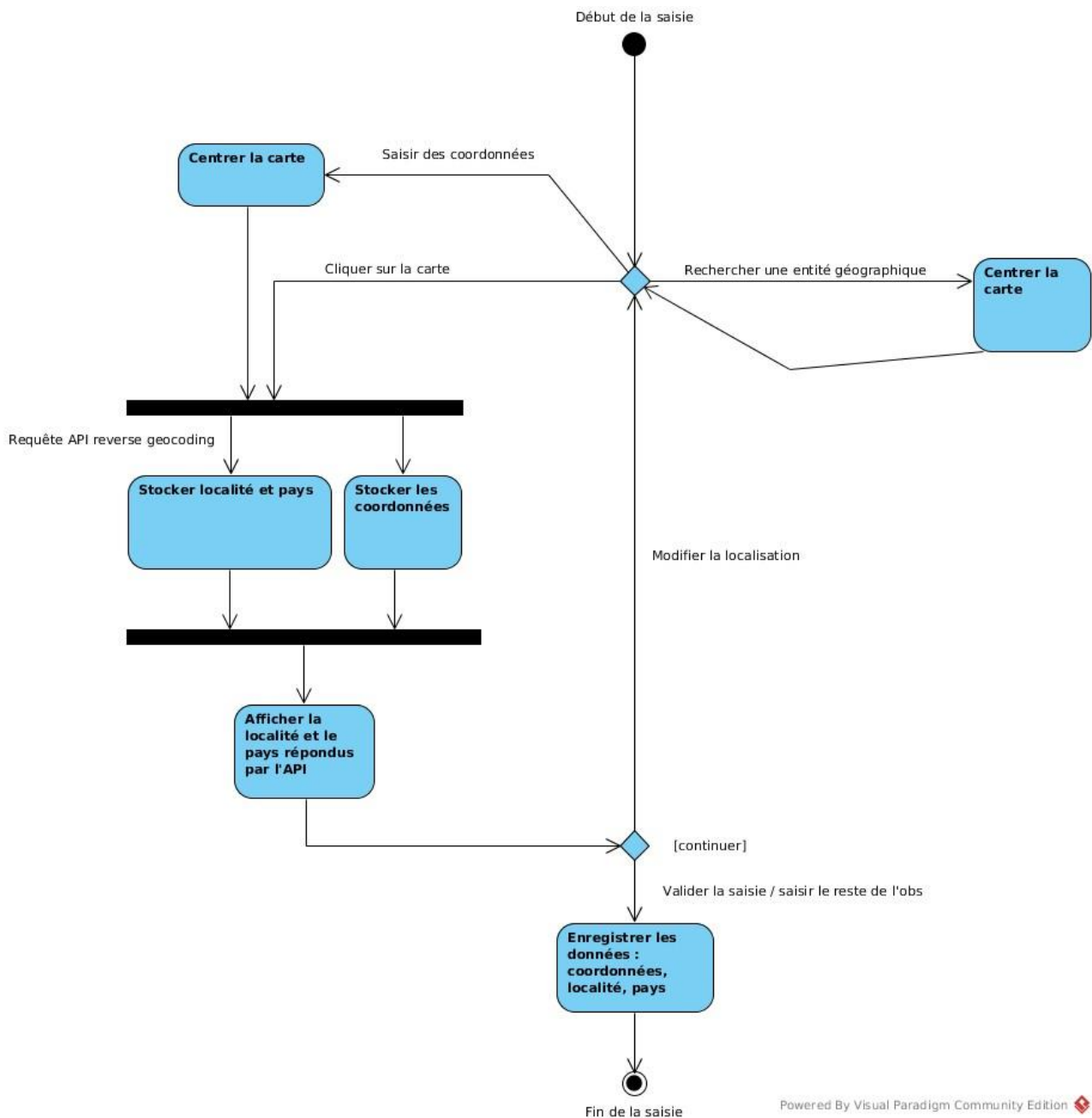


Figure 13 : Diagramme d'activité décrivant les bases du fonctionnement du nouveau protocole de géolocalisation

Ce diagramme d'activité se concentre sur le processus de **géolocalisation**, en décrivant le comportement de l'interface de saisie selon les actions (représentées par les nœuds) de l'utilisateur. Dans cette proposition, il n'a pas la possibilité de saisir ou modifier la localité ni le pays mais seulement de renseigner une paire de coordonnées géographiques en cliquant sur une carte ou en les saisissant directement. La saisie d'une entité géographique a pour seul effet de modifier la vue sur la carte, afin de rechercher plus rapidement l'endroit de l'observation. Ainsi, les informations de localité et pays reposent entièrement sur la performance du service de reverse geocoding qui les calcule. Ce type de schéma a permis d'identifier les problèmes de fonctionnement potentiels et de construire plus facilement par la suite les maquettes illustrant l'interface.

e) Maquettes et scénarios

La maquette présentée ici est un aperçu de l'aspect que pourrait avoir l'interface de saisie du carnet en ligne. L'ensemble des visuels de la maquette est consultable en Annexe N.

Figure 14 : L'un des visuels de la maquette de l'outil de saisie (module de localisation) - Etape 4

Dans le cas présenté ci-dessus, l'utilisateur a commencé par rechercher la ville de Grenoble grâce au champ « Rechercher un lieu », puis il a déplacé la carte jusqu'à trouver l'endroit où il se souvient avoir effectué l'observation. En cliquant sur la carte, les coordonnées géographiques, l'altitude, la zone géographique ainsi que le pays ont été calculés automatiquement grâce à différents services web (dont CoordSearch). Le fonctionnement de cette interface est un peu plus complexe que celui présenté précédemment (cf. diagramme d'activité en Figure 13) puisqu'en cliquant sur le bouton « Vous n'êtes pas d'accord ? », l'utilisateur a la possibilité de saisir une localité dans un second champ.

D. Etude des conséquences sur les autres interfaces et outils

1. Méthodologie

L'ensemble des outils et des interfaces de Flora Data étant liés (au moins par la présence de données), la modification d'un processus a potentiellement des conséquences sur les autres. Il convient donc d'évaluer ces effets et de proposer des solutions si nécessaire.

2. Lien avec les autres outils : quelles conséquences sur la restitution des données ?

Les choix concernant le protocole ont une influence sur le contenu des données d'observation (c'est d'ailleurs en partie l'objectif) et donc sur ce qui est affiché dans les autres interfaces. Un rappel des interfaces et outils concernés :

- Métadonnées affichées dans **IdentiPlante**, **PictoFlora**, **eFlore** :

Ces interfaces restituent les observations illustrées seulement. Actuellement, leur affichage est à homogénéiser afin qu'il soit cohérent entre eFlore, IdentiPlante, PictoFlora et éventuellement d'autres interfaces telles que la page d'accueil du site. L'information de localisation systématiquement affichée est la localité, parfois le département (pour la France), le pays ou encore les infos complémentaires telles que le lieu-dit ou la station. Ces interfaces ne permettent pas d'accéder à l'emplacement précis de l'observation.

- **Cartographies** :

La restitution cartographique des observations dépend des informations saisies. D'une part, la localisation n'étant pour l'instant pas une information obligatoire, toutes les données ne peuvent pas être représentées sur une carte. D'autre part, on distingue deux cartographies : la cartographie communale, qui représente les observations associées à une localité valide, et la cartographie au point sur laquelle on ne peut voir que les observations géolocalisées.

- **Export** :

Deux interfaces d'export sont actuellement proposées par Tela Botanica. La première est accessible depuis le CEL et permet de télécharger les données contenues dans celui-ci uniquement. La seconde est un widget permettant de télécharger n'importe quelles données parmi les observations publiques.

a) Choix de la localité à afficher en cas de conflit avec la valeur saisie par l'utilisateur

Dans la mesure où l'on permet à l'utilisateur de saisir une localité, celle-ci peut être en conflit avec la zone géographique calculée par CoordSearch. Dans ce cas et si l'auteur s'est servi de l'autocomplétion, on dispose de deux informations sur la localité et il faut choisir laquelle restituer. Dans un souci de clarté il semble plus approprié d'utiliser, le cas échéant, la **localité saisie par l'utilisateur**.

b) Représentation des données « imprécises »

Ici « donnée imprécise » désigne toute observation associée à une distance de précision. Cette information pousse à représenter la donnée dans l'espace non plus comme un point mais comme un disque. Ainsi, plusieurs questions se posent, notamment pour la représentation cartographique : faut-il représenter ou non ces observations sur la carte au point, et si oui comment ? Une solution consiste à **représenter la zone d'incertitude en affichant un disque** plutôt qu'un point, en veillant à **adapter leur affichage au niveau de zoom** pour ne pas rendre la carte illisible.

c) Gestion du floutage

Il faut veiller à ce que la restitution des données floutées se fasse à l'échelle spécifiée par leur auteur dans l'ensemble des interfaces. Afin de ne pas compliquer l'affichage de la carte ni de risquer la divulgation de données floutées, celles-ci **ne doivent jamais être affichées en clair**, même pour leur auteur lorsqu'il est identifié. L'accès à la position précise de ces données ne doit être possible que via son propre CEL.

De même pour le téléchargement, la localisation doit être dégradée lors de l'export, c'est-à-dire pas de coordonnées géographiques mais seulement la maille ou la commune. Si l'auteur des données souhaite les télécharger sans floutage, là encore il peut le faire depuis son CEL.

Enfin, si le floutage à la maille est rendu possible, il faut envisager un nouveau mode de représentation cartographique. Ce problème est cependant facile à régler en passant à une « vraie » web carto interactive, permettant de gérer plusieurs couches sur la même interface.

E. Bilan de l'étude de cas

En partant sur un protocole de localisation basé sur les coordonnées géographiques, les outils de saisie de Tela Botanica s'aligneraient sur les pratiques actuelles en termes d'observation naturaliste. Les données gagneraient en robustesse et en qualification.

Il y a cependant un effort de communication à faire pour sensibiliser les utilisateurs et ne pas perdre les habitués des anciennes interfaces : d'une part, en expliquant ce nouveau protocole (les infos attendues, obligatoires, les bonnes pratiques de saisie, le fonctionnement et les limites du reverse geocoding, comment améliorer le processus, etc.), de l'autre, et toujours dans un but pédagogique, en instaurant un indicateur de qualité lors de la saisie pour sensibiliser les utilisateurs aux bonnes pratiques et les encourager à s'appliquer. Ce type d'indicateur permettrait ensuite de qualifier les données en fonction de leur qualité potentielle.

Dans cette étude de cas nous nous sommes focalisés sur la partie localisation, mais il convient de traiter l'ensemble du protocole de saisie. En plus de répondre à un format standard de données, l'un des objectifs de cette démarche est d'homogénéiser la saisie du CEL avec celle du widget de saisie rapide. En effet, il est nécessaire qu'ils soient cohérents du point de vue des utilisateurs et qu'ils produisent des données répondant aux mêmes règles, même si la deuxième interface est simplifiée.

Après avoir décrit ma démarche, à la fois dans le cas général et pour une étude de cas plus spécifique, j'ai abordé avec autant de recul que possible le bilan de mon expérience au sein de Tela Botanica, ainsi que la situation de l'association.

IV. Bilan et perspectives

Dans cette partie, il s'agit dans un premier temps d'analyser et de dresser un bilan de mon expérience et de ma perception du fonctionnement de Tela Botanica dans le cadre de cette alternance. En seconde partie, j'ai produit des recommandations en fonction des constats précédents, selon ma vision des perspectives pour Flora Data.

A. Bilan et analyse critique

1. Pertinence de la méthode employée

La méthodologie générale a été choisie au début de l'alternance, alors que la problématique et les missions apparaissaient comme très larges. Son but était avant tout de **préciser l'organisation du travail grâce à un cadre méthodologique clair**, sur lequel je pourrai trouver assistance auprès de l'équipe pédagogique et dont je maîtrisais (ou du moins connaissais) les principes.

Dès le début, **l'approche a montré sa pertinence** dans la mesure où je n'avais pas de connaissances ou presque en botanique : elle m'a permis de m'approprier très finement les outils et leur fonctionnement. Un peu plus tard lors de l'analyse des besoins j'ai pu « rencontrer » les utilisateurs, demander leur avis, mieux connaître leurs usages. Peu pratiqué par l'association jusqu'à présent, les utilisateurs apprécient ces échanges directs et l'opportunité de donner leur avis. Par ailleurs, ces retours me semblent importants pour l'équipe qui n'a pas souvent l'occasion d'avoir des retours précis sur son travail (dont de nombreux compliments). Afin de travailler sur un pool d'utilisateurs plus représentatif, un questionnaire écrit aurait cependant pu venir compléter l'analyse des besoins.

Un regret est de ne pas avoir plus approfondi les aspects techniques du SI, en partie car l'équipe informatique était très prise par le développement du nouveau site. J'ai pu toutefois en avoir un aperçu lors des réunions Floris'Tic, puis les aborder cet été en prenant part à la formation d'un nouvel informaticien.

Enfin, la méthode a montré des limites dans le cadre de l'amélioration du SI **existant**. En effet, si elle est bien adaptée à la conception de nouveaux outils, pour certains aspects il a été difficile de l'appliquer à un système déjà existant. Lors des entretiens par exemple, il était aisé de questionner les utilisateurs sur leurs usages (puisqu'ils connaissaient déjà les outils de Tela) mais difficile dans un premier temps d'en faire ressortir les besoins, d'autant plus les besoins non satisfaits par les outils de Tela. Dans la même veine, la priorisation des fonctionnalités à développer ne pouvait être effectuée uniquement sur la base des besoins des utilisateurs (ceux-ci, donc, étant majoritairement satisfaits).

Même si j'ai dû parfois adapter la méthodologie, celle-ci a rempli son rôle en cadrant efficacement les différentes tâches et en me permettant facilement d'échanger avec mes encadrants sur l'avancée de mon travail.

2. L'organisation du travail et les compétences au sein de l'association

Les compétences - à l'instar des activités - de l'équipe de Tela Botanica sont **fortement orientées vers l'animation**, en nombre de salariées, stagiaires et services civiques : animation de réseau, conduite de projet, accompagnement et formation de relais, communication. L'équipe informatique se réduit à trois personnes, dont deux développeurs web et une gestionnaire des bases de données.

La **priorisation des développements** est jusqu'à présent réalisée conjointement par la direction et l'équipe informatique, sur la base des échéances imposées par des appels à projets, en fonction de critères financiers et des objectifs globaux de l'association (développer le réseau par exemple). Des tâches plus ou moins imprévues viennent souvent s'ajouter, en cas de bug ou de dysfonctionnement du site internet par exemple.

Le déséquilibre entre les équipes d'animation et informatique surcharge les développeurs, qui ne peuvent travailler sereinement sur l'ensemble des projets. Avec le départ de l'un d'eux, ce phénomène risque de beaucoup s'accroître : même si un développeur de bon niveau le remplace rapidement, il lui faudra du temps avant de maîtriser suffisamment le SI pour être pleinement efficace. Faire appel à des prestataires pour compenser le manque de moyens humains semble être une bonne solution, mais cela ne peut remplacer l'équipe interne et ses connaissances du système.

Sur Flora Data plus spécifiquement, mis à part quelques patchs⁸ apportés aux outils, on observe peu de renouvellement. Suivant le fonctionnement de l'association, en l'absence d'animateur et donc de product owner attaché à ce programme, personne ne pousse pour que Flora Data soit enrichi dans sa globalité. Ce programme est donc maintenu mais pas réellement amélioré, alors même que la majorité des processus collaboratifs de Tela reposent sur la saisie de données d'observations. Ce constat est d'ailleurs l'une des raisons de mon recrutement sur le sujet.

3. Les choix stratégiques vis-à-vis de la problématique

a) Recruter un-e étudiant-e plutôt qu'ouvrir un poste

La problématique de l'amélioration des données et de Flora Data est centrale, mais la gestion des données d'observation ne constitue qu'une partie des activités de l'association. Jusqu'à présent, elle n'était pas gérée de manière globale mais au cas par cas lorsqu'un problème concret se présentait (ex : trop de données PI@ntNet mobile l'été 2016). En embauchant une alternante AgroTIC sur cette thématique, l'idée était notamment d'ajouter à l'équipe une personne capable d'assurer l'interface entre l'équipe informatique et les utilisateurs des réalisations au sens large, des salariés de l'association aux membres du réseau. Cette place ainsi que la compréhension des technologies employées par les informaticiens, du fonctionnement et des enjeux de l'association devaient permettre à cette personne de mener à bien la définition des spécifications fonctionnelles pour un certain nombre d'outils.

Ce choix impliquait une mise en route peut-être plus lente qu'avec quelqu'un d'expérimenté, mais l'association n'étant pas dans l'urgence pour le travail sur Flora Data, cela n'a pas énormément porté préjudice au déroulement du projet. Cette décision, également économique pour une structure dont les charges salariales s'élèvent à plus de 70% des charges totales (rapport d'activité 2016) et dans un domaine peu financé, permet aussi de tester et former quelqu'un pendant plus d'un an.

b) Les choix stratégiques concernant Flora Data

Concernant les données d'observation, avec la **recherche d'intégration au SINP** l'association s'engage à s'adapter à certaines normes. Pour moi, cela représente une opportunité : profiter d'un cadre prédéfini pour améliorer les données, en présence d'experts prêts à évaluer leur qualité (et donc capables de mesurer les

⁸ Un patch ou correctif, est une section de code que l'on ajoute à un logiciel, pour y apporter des modifications, par exemple pour corriger un bug (Wikipédia).

effets d'éventuelles améliorations), et enfin confirmer la place de Tela Botanica et des données d'observation collectées par des amateurs dans le monde naturaliste.

La **position de l'association vis-à-vis de Pl@ntNet** aussi comporte une dimension stratégique. Le choix de conserver dans Flora Data les observations venant de l'application mobile ne fait pas l'unanimité, notamment pour des raisons de qualité des données. De mon point de vue, l'enjeu pour l'association est d'être associé à un outil innovant, qui peut profiter à l'image de Tela Botanica (même si celle-ci a tendance à rajeunir, avec le MOOC et bientôt le nouveau site). Plus simplement, le partenariat avec Pl@ntNet permet de profiter d'une technologie assez jalousement gardée et en constante amélioration : la reconnaissance taxonomique à partir de l'analyse d'image. Celle-ci peut constituer une valeur ajoutée aux processus de détermination, dans la mesure où elle ne remplace pas la dimension collaborative, les échanges, la transmission de savoir qui constituent le point fort de Tela. Par exemple, un outil réellement dédié à la saisie (mobile ou non) proposant une aide à l'identification automatisée serait à la fois innovant et pratique pour les utilisateurs.

Enfin, début 2017 l'association a lancé la réalisation par un prestataire (botaniste et phytosociologue, utilisateur des outils de l'association, membre du CST et développeur à ses heures perdues) d'un **outil de saisie pour les données de phytosociologie**⁹ (relevés de végétation). En effet, si le CEL est bien adapté à la saisie d'observations ponctuelles, il n'est pas prévu pour la saisie de listes de plantes relatives à la description d'un milieu (relevé phytosociologique). D'autre part, l'association possède déjà de nombreuses données de végétation qu'elle n'a pas l'occasion de valoriser. Le développement d'un outil dédié doit permettre de disposer d'un outil plus professionnel, plus complexe, destiné à la collecte et la gestion de relevés phytosociologiques. L'ensemble des données, issues du CEL et de ce nouvel outil, est complémentaire et pourra être affichée de façon unique.

Justifiée par le positionnement assez unique de Tela Botanica dans cette discipline - aujourd'hui peu vulgarisée - et par son ouverture potentielle à un public plus large, l'investissement est cependant risqué. En effet, c'est un domaine complexe qui requiert de bonnes connaissances en botanique mais aussi en écologie, voire en géographie, puisqu'il s'agit de décrire les relations des plantes entre elles et avec leur milieu, dans l'espace et dans le temps. D'autre part, pour des raisons aussi politiques que scientifiques l'approche de la phytosociologie adoptée par les membres du réseau, dite « synusiale intégrée », n'est pas soutenue au niveau national (et donc pas financée). Le contexte et la situation économique de l'association ne se prêtant pas vraiment à une telle prise de risque, cette décision a fait l'objet de nombreuses discussions, remettant en question l'intérêt de développer un outil complexe qui ne répondra aux besoins que d'un petit nombre d'utilisateurs, même s'il permet par la même occasion d'exploiter des données aujourd'hui laissées de côté. La disponibilité d'un botaniste, à la fois qualifié en phytosociologie et capable de développer l'outil de saisie en question, est toutefois apparue comme une opportunité à ne pas manquer. La dimension innovante de cette démarche a également constitué un argument en sa faveur, positive pour l'image de Tela Botanica et afin de contribuer à l'évolution de la botanique.

⁹ La phytosociologie est une branche de l'écologie, dont l'objet est l'analyse des communautés végétales spontanées, « en se basant sur des listes floristiques les plus exhaustives possibles », pour « les définir et les classer selon des critères floristiques et statistiques, caractériser leur structure et leur organisation, leur origine, leur genèse, leur déterminisme et leur évolution ainsi que leurs habitats. » (Tela Botanica, Wikipédia)

B. Perspectives et préconisations

Le travail de réflexion mené sur Flora Data m'a permis d'envisager de nombreuses perspectives pour le programme. Je reste convaincue qu'il est central, utile et utilisé mais surtout qu'il gagnerait à être mieux valorisé.

1. Quelles compétences nécessaires ?

Cette mise en valeur passe par le (re)développement des interfaces et des outils, il est donc important qu'un développeur web au moins connaisse assez le SI de l'association pour y travailler. Des développements mobiles pourraient être envisagés à moyen terme, mais semblent pour le moment loin d'être prioritaires.

Afin d'accompagner le travail sur les interfaces, l'appel à un spécialiste de l'ergonomie permettrait de ne pas retomber dans une trop grande complexité, souvent lié à un désir d'exhaustivité. De nombreux utilisateurs sont des passionnés, et mis à part ceux qui utilisent les outils dans un cadre professionnel, ils sont là pour leur plaisir. Il est important donc que les interfaces soient pratiques à utiliser mais aussi esthétiques.

Enfin, je trouverai enrichissant d'impliquer les utilisateurs lors des développements, afin de vérifier justement la lisibilité et la fonctionnalité des outils. En établissant plus de liens entre les développeurs et les usages et usagers réels, en les mettant à contribution pendant les phases de test, on peut envisager d'aboutir à un produit bien adapté aux utilisateurs. C'est notamment ce qui est promu par la méthode Agile, que l'on pourrait mettre mieux à profit chez Tela Botanica, en constituant par exemple un ou plusieurs groupes d'utilisateurs "bêta-testeurs". La personne chargée de coordonner ces tests peut faire partie de l'équipe informatique ou de l'équipe d'animation avec une certaine affinité pour l'informatique.

2. Externaliser les développements : oui mais dans quelle mesure ?

Vue la force de développement actuelle, l'appel à des prestataires semble indispensable pour les développements à venir. La plupart des projets informatiques peuvent être à la fois internalisés et externalisés comme c'est le cas pour le nouveau site, la partie ergonomie notamment étant assurée par le prestataire.

Actuellement, il est difficile d'envisager externaliser la maintenance des plateformes web et le suivi utilisateurs (réponse aux questions et signalements de bugs). Ces tâches constituent selon moi un lien important entre l'équipe, l'état du site et les utilisateurs, qu'il faut conserver.

Les choix stratégiques et les spécifications doivent évidemment être faits en interne et en consultation avec l'ensemble de l'équipe.

3. A quels coûts doit-on s'attendre ?

Les coûts sont surtout fonction du temps de développement et de l'externalisation ou non de certains développements.

Les développements sont longs et difficiles à planifier, avec ou sans prestation, d'autant plus avec un SI aussi complexe. Il serait intéressant de tester différentes méthodes de planification afin d'estimer au mieux les temps de développement. Ceci explique en partie les retards pour le nouveau site, de nombreux imprévus s'étant présentés au fur et à mesure. Les retards portent forcément préjudice à l'association et pas seulement financièrement : les délais peuvent impliquer des partenaires qui attendent un service, d'autres choix peuvent reposer sur ces délais (ex : recruter des services civiques pour couvrir la sortie du nouveau site), l'équipe est sous pression et doit travailler au plus urgent et non plus au plus enrichissant.

Enfin il faut être prêt à améliorer d'ici quelques années les serveurs et leur capacité de stockage si l'association souhaite continuer à agrandir le réseau. En effet, il faut pouvoir faire face à un nombre et un flux de données croissant pour éviter des ralentissements comme le site a connu récemment, avant la suppression d'une grande partie des données venant de Pl@ntNet mobile (le site et les outils n'étaient alors quasiment plus utilisables).

Bilan : Malgré un point de vue encore partiel sur le fonctionnement de l'association, cette expérience m'a permis de me faire une idée de sa situation et de ses perspectives. Ce bilan est toutefois à replacer dans un contexte économique, social et politique particulier pour les structures associatives (Tchernonog, 2016), qui doivent faire face à de nombreux changements et adapter leur mode de gestion.

Conclusion

Le travail sur les données d'observation semble un enjeu majeur, à l'heure où le monde scientifique s'intéresse aux données produites par le « Tiers État de la recherche » (AKRICH *et al.*, 2017). La diffusion des données issues notamment des sciences participatives permet à Tela Botanica de mettre en lumière l'étendue des compétences et des connaissances de son réseau d'amateurs (Mathieu, 2017).

En travaillant à la transformation de ses outils (de saisie, de révision collaborative, de restitution), elle s'assure par ailleurs de pouvoir développer encore ce réseau - pourquoi pas à l'international - en lui proposant des outils plus robustes et adaptés à différents niveaux en botanique. En plus d'enrichir l'expérience des utilisateurs, l'amélioration des plateformes de Flora Data entraînerait nécessairement celle des données d'observation produites : en encourageant d'une part la montée en compétence des membres du réseau et en limitant d'autre part la saisie et la diffusion de données incomplètes ou incohérentes.

L'association assure une certaine pérennité de ses outils de saisie en les adaptant aujourd'hui au standard d'échange de données naturalistes, en passe d'être adopté au niveau national. L'étude du fonctionnement de Flora Data dans ce cadre a permis de mettre le doigt sur différents points critiques qui méritent une attention particulière. C'est le cas du protocole de saisie et en particulier de la localisation.

L'étude de différents services de géocodage inversé a montré que les outils développés en interne à Tela n'ont rien à envier à certaines solutions commerciales. Ils sont fonctionnels et, dans ce cas, plutôt adaptés aux besoins du réseau en termes de couverture géographique. Comme c'est souvent le cas pour les autres plateformes de Tela, c'est leur lenteur qui fait leur faiblesse. Cela dit, ce problème est bien connu de l'équipe informatique qui a commencé à travailler dessus de manière globale.

A l'échelle de l'association, essentiellement financée par des fonds publics (Rapport d'activité 2016), le contexte politique, économique, social et sociétal a de nombreuses répercussions. La réduction des subventions nationales et territoriales pousse les associations à se mettre en concurrence avec des entreprises pour répondre aux appels d'offre des marchés publics. D'autre part, d'après Viviane Tchernonog, chercheuse au CNRS au Centre d'Economie de la Sorbonne, ces structures "ont du mal à attirer et conserver des emplois salariés. [Les salariés] ont tendance à acquérir ou améliorer leurs qualifications au sein des associations, avant de rejoindre d'autres structures publiques ou privées dans les lesquelles les conditions de l'emploi – rémunérations, stabilité, carrières – apparaissent plus attrayantes" (Tchernonog, 2016).

A l'heure actuelle, l'association accomplit pleinement ses missions : diffusion de connaissances et formation de débutants avec le succès, la répétition et la diversification de MOOCs, développement des programmes de sciences participatives, formation de relais locaux ou encore extension progressive à l'international par l'agrégation de référentiels d'autres espaces géographiques. Tela Botanica se retrouve toutefois dans la nécessité de s'adapter rapidement, en cherchant de nouveaux modèles économiques. Entre autres, la vente des produits et des services en lien avec les activités de l'association, à destination des particuliers et des collectivités, mais aussi le développement d'une cotisation régulière d'adhésion sont envisagés. En effet jusqu'à présent, les "membres du réseau" n'ont pas le statut d'adhérent.

Les échanges, le travail effectué et les choix stratégiques de l'association m'ont permis de réaliser qu'il est nécessaire que son fonctionnement évolue d'ici les prochaines années, tout en constatant qu'elle possède déjà les ressources qui favoriseront ce changement.

Bibliographie

Akrich Madeleine, Barré Rémi, Bentz Émeline, Bontema François, Delannoy Marie, Evrard Francine, Goussault Bénédicte, Grasset Lucile, Joly Pierre-Benoît, Larqué Lionel, Le Bars Yves, Lenouvel Damien, Lhoste Evelyne, Millier Claude, Ottolini Lucile, Porcheron Patrick, Porte Emmanuel, Roturier Christophe, Schneider Mathieu, 2017. *Prendre au sérieux la société de la connaissance, Livre Blanc*. ALLISS Sciences Société, p. 39.

Audibert Laurent, 2006 - 2009. *UML 2, de l'apprentissage à la pratique*, developpez.com.

URL : <http://laurent-audibert.developpez.com/Cours-UML/>

Bollier D., Cassiers I., Dedeurwaerdere T., De Moor T., El Ghabri M., Errebault G., Knops L, Lechat B., Marechal A., Peugeot V., Roels M., Servigne P., Zacharie A., 2012. *Les Biens Communs : comment (co)gérer ce qui est à tous ?* Actes du colloque coorganisé par Etopia, Oikos et Green European Foundation, 9 mars 2012, Bruxelles, p. 51.

URL : http://www.etopia.be/IMG/pdf/biens_communs_fr_web-2.pdf

Champavier Kelly, 2016. *Mémoire de fin d'étude : Comment impliquer les utilisateurs dans la vie et l'évolution d'un outil d'aide à la décision ? Mise en place du support client de Vintel*, p. 43-50.

De Barros G., Molina J., 2017 - *Analyse des observations des sciences participatives pour un apport au SINP Languedoc-Roussillon*. Conservatoire Botanique National Méditerranéen de Porquerolles.

Folding Burritos, 2016. *20 Product Prioritization Techniques : A Map and Guided Tour*.

URL : <http://foldingburritos.com/product-prioritization-techniques/>

Goldberg D.W., Ballard M., Boyd J.H., Mullan N., Garfield C., Rosman D., Ferrante A.M., Semmens J.B., 2013. *An evaluation framework for comparing geocoding systems*. International Journal of Health Geographics 2013 12:50.

URL : <https://doi.org/10.1186/1476-072X-12-50>

Houiller F., 2016. *Les Sciences participatives en France. Etats des lieux, bonnes pratiques & recommandations*. p.6.

Imbert Geneviève, 2010. *L'entretien semi-directif : à la frontière de la santé publique et de l'anthropologie*, Recherche en soins infirmiers, 3/2010 (N° 102), p. 23-34.

URL : <http://www.cairn.info/revue-recherche-en-soins-infirmiers-2010-3-page-23.htm>

INSEE, 2016. *Définitions, méthodes et qualité - Définitions - Taxons*.

URL : <https://www.insee.fr/fr/metadonnees/definition/c1239>

Jomier R., Chataigner J., Poncet L., Lebeau Y., Robert S., Bourgoïn T., Chagnoux S., Saltré A., Borremans C., Archambeau A-S., Lecoq M-E, Pamerlon S., Just A., Milon T., Cousin J-L, Viel N., Barreau S., 2016. *Standard de données SINP Occurrences de taxons*, version 1.2, MEDDE/MNHN, Paris, 93 pp, p. 12.

Lescure P., Chaudret B., Podalydes D., Latour B., Thieulin B., Tessier M., Bensoussan A., Wiewiorka M., et al. (Tribune), 2015. *Favorisons la libre diffusion de la culture et des savoirs*. Le Monde.fr | 10.09.2015.

URL : http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs_4751847_3232.html

Lothon Florent, sd. *Introduction aux méthodes agiles et Scrum*, L'Agiliste.

URL : <http://www.agiliste.fr>

Mathieu Daniel, Mathez Joël, 2016. *Les référentiels de Tela Botanica - Notions de base permettant de comprendre comment sont constitués les référentiels de Tela Botanica et comment sont organisées les connaissances qui s'y rattachent*. 14 pp, p. 9.

URL : <https://docs.google.com/document/d/1hDc7dCTXual8UM8A93kUqcrctgVllnXGPsbOYwh4pU3U/>

Mathieu Daniel, 2017. *Dossier - Ouvrons nos carnets : Intégrer les données de réseaux amateurs*. Espaces naturels n° 59 juillet - septembre 2017, p. 32.

Tela Botanica, sd. *Introduction à la phytosociologie - Notion d'association végétale*.

URL : <http://www.tela-botanica.org/page:Introduction>

Tela Botanica, 2017. *Statistiques sur les inscrits au réseau Tela Botanica*.

URL : http://www.tela-botanica.org/page:membres_du_reseau_statistiques

Tela Botanica, 2017. *Rapport d'activité 2016*. p. 44.

URL : http://www.tela-botanica.org/sites/reseau/fr/documents/activite_tb_2016.pdf

Tchernonog Viviane, 2016. *Quel est l'état et le contexte économique actuel des associations ? – Regards Croisés*. Ecosocioconso, les associations en France.

URL : <http://www.ecosocioconso.com/2016/01/11/les-associations-en-france/>

Wikipédia, 2016. *Patch (informatique)*.

URL : [https://fr.wikipedia.org/wiki/Patch_\(informatique\)](https://fr.wikipedia.org/wiki/Patch_(informatique))

Wikipédia, 2017. *Phytosociologie*.

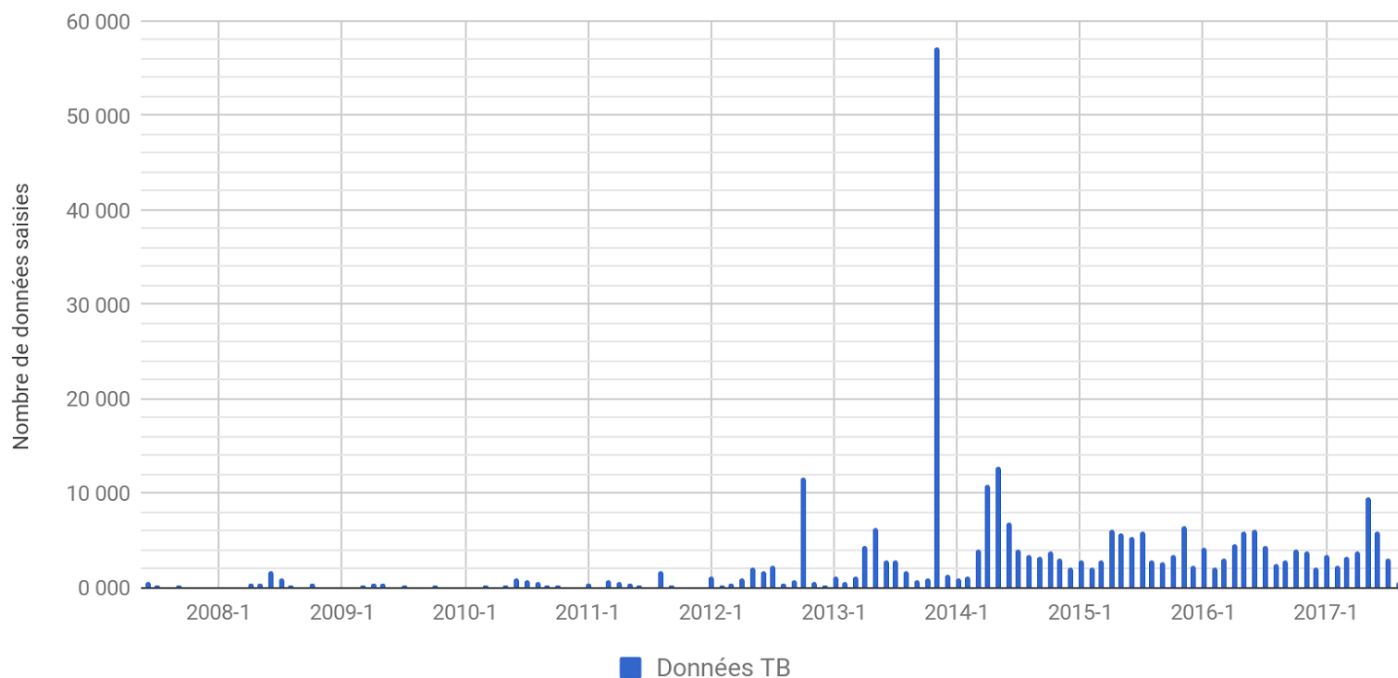
URL : <https://fr.wikipedia.org/wiki/Phytosociologie>

Table des annexes

Annexe A.	Evolution de la quantité d’observations saisies depuis la création de Tela Botanica	59
Annexe B.	Flux de données au sein de Flora Data et avec les partenaires Floris’Tic	60
Annexe C.	Diagrammes de cas d’utilisation des outils de Flora Data.....	61
Annexe D.	Diagrammes de classes : CEL et DEL.....	63
Annexe E.	Synthèse de l’analyse de l’existant.....	65
Annexe F.	Tableau récapitulatif des sites internet ayant fait l’objet de veille	67
Annexe G.	Canevas d’entretien de l’analyse des besoins des utilisateurs – Exemple du Profil 3 : Amateurs éclairés et botanistes experts.....	68
Annexe H.	Exemples de personae.....	70
Annexe I.	Diagramme de définition des fonctionnalités répondant aux besoins des utilisateurs – Exemple du profil débutant	71
Annexe J.	Classement et notation des fonctionnalités.....	72
Annexe K.	Taux de réponse moyens.....	75
Annexe L.	Taux de réponse par jeu de données	76
Annexe M.	Calcul de l’altitude d’une observation à partir de ses coordonnées géographiques : premiers éléments pour une étude comparative des services web.....	77
Annexe N.	Maquettes : saisie de localisation	78

Annexe A. Evolution de la quantité d'observations saisies depuis la création de Tela Botanica

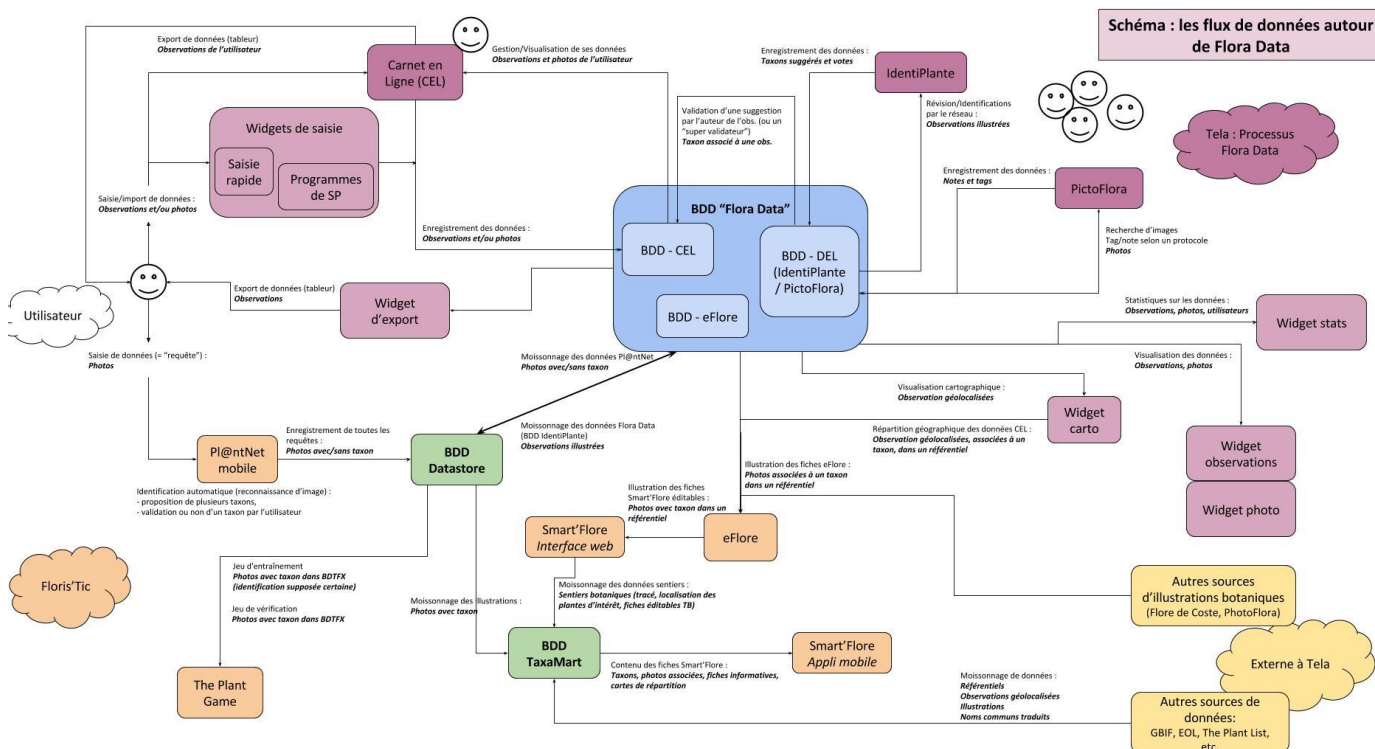
Evolution de la quantité de données Tela Botanica ("non PI@ntNet") saisies depuis sa création



NB : La présence d'un pic de plus de 57 000 données entrées dans Flora Data fin 2016 correspond à un import massif de données effectué en internet, indépendant de l'activité régulière des utilisateurs.

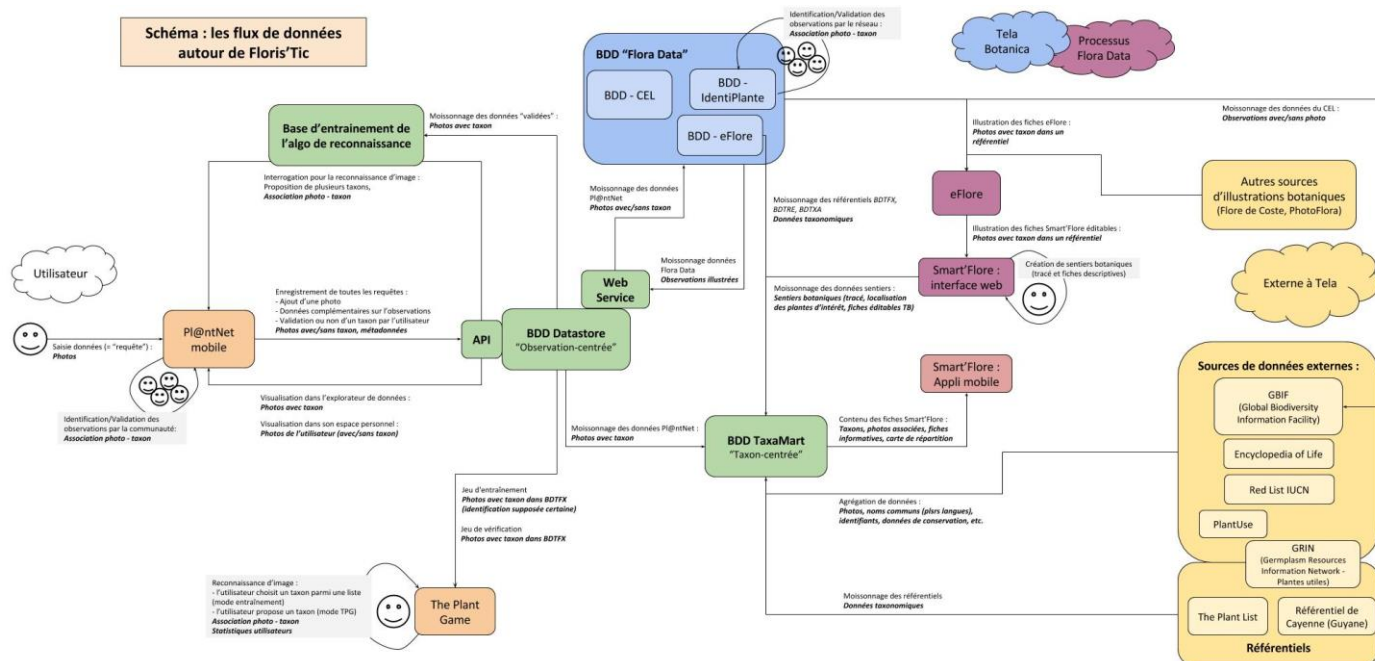
Annexe B. Flux de données au sein de Flora Data et avec les partenaires Floris'Tic

Ces deux schémas étant trop grands pour être présentés ici, un aperçu général seulement en est donné. Ils peuvent être consultés en suivant les liens en légende de chacun.



Les flux de données au sein de Flora Data :

https://docs.google.com/drawings/d/1Er_ZPBU8BNFMXola4K6LjnD18XdCh0oMTYUYkO6Q4Os



Les flux de données entre Flora Data et les partenaires de Floris'Tic :

https://docs.google.com/drawings/d/1aQtGc_sr0AXXf-6rCCQBkY_i4w76wP6rC3hj9kFyQs8

Annexe C. Diagrammes de cas d'utilisation des outils de Flora Data

Diagramme de cas d'utilisation :
Les processus d'aide à l'identification

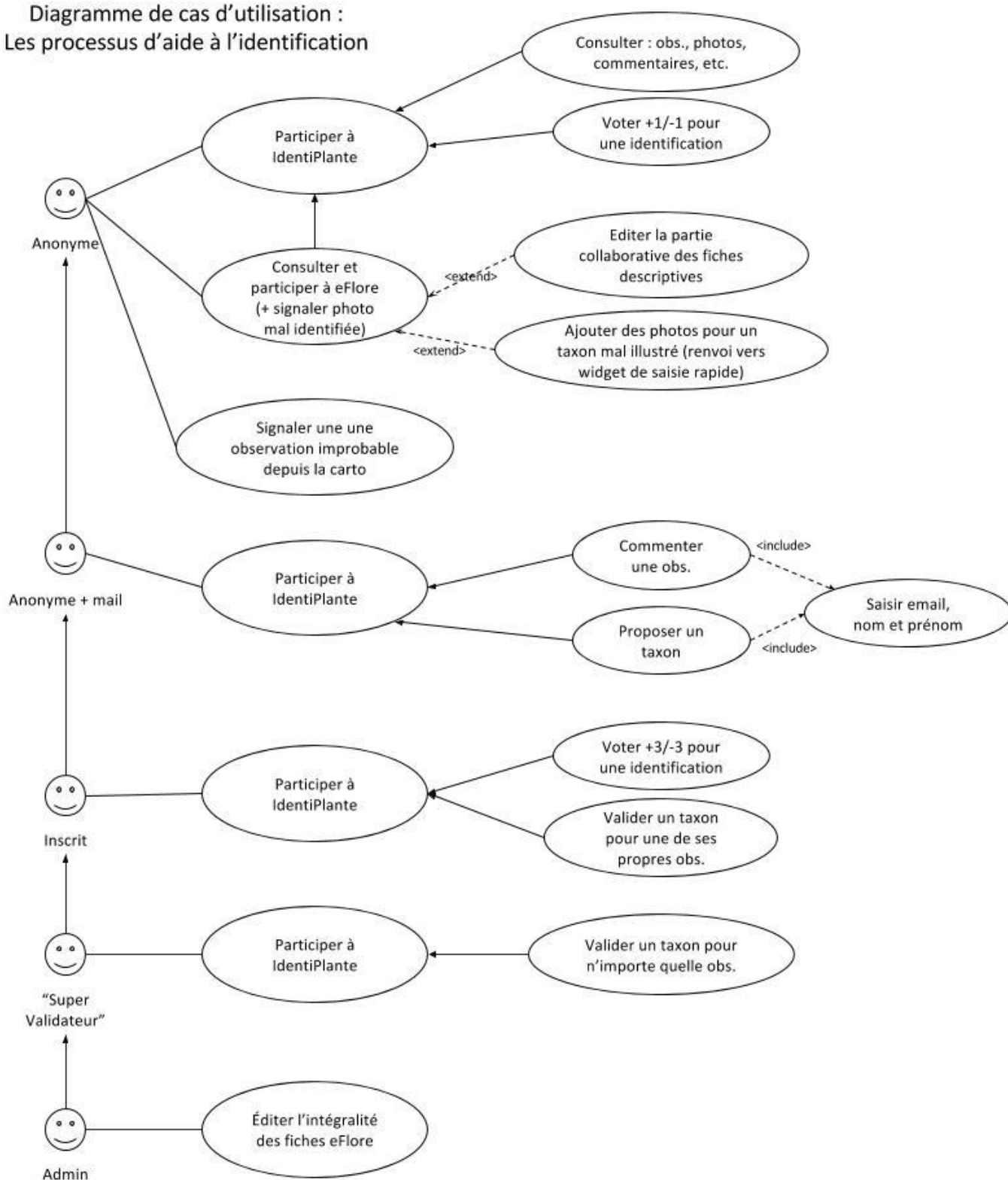
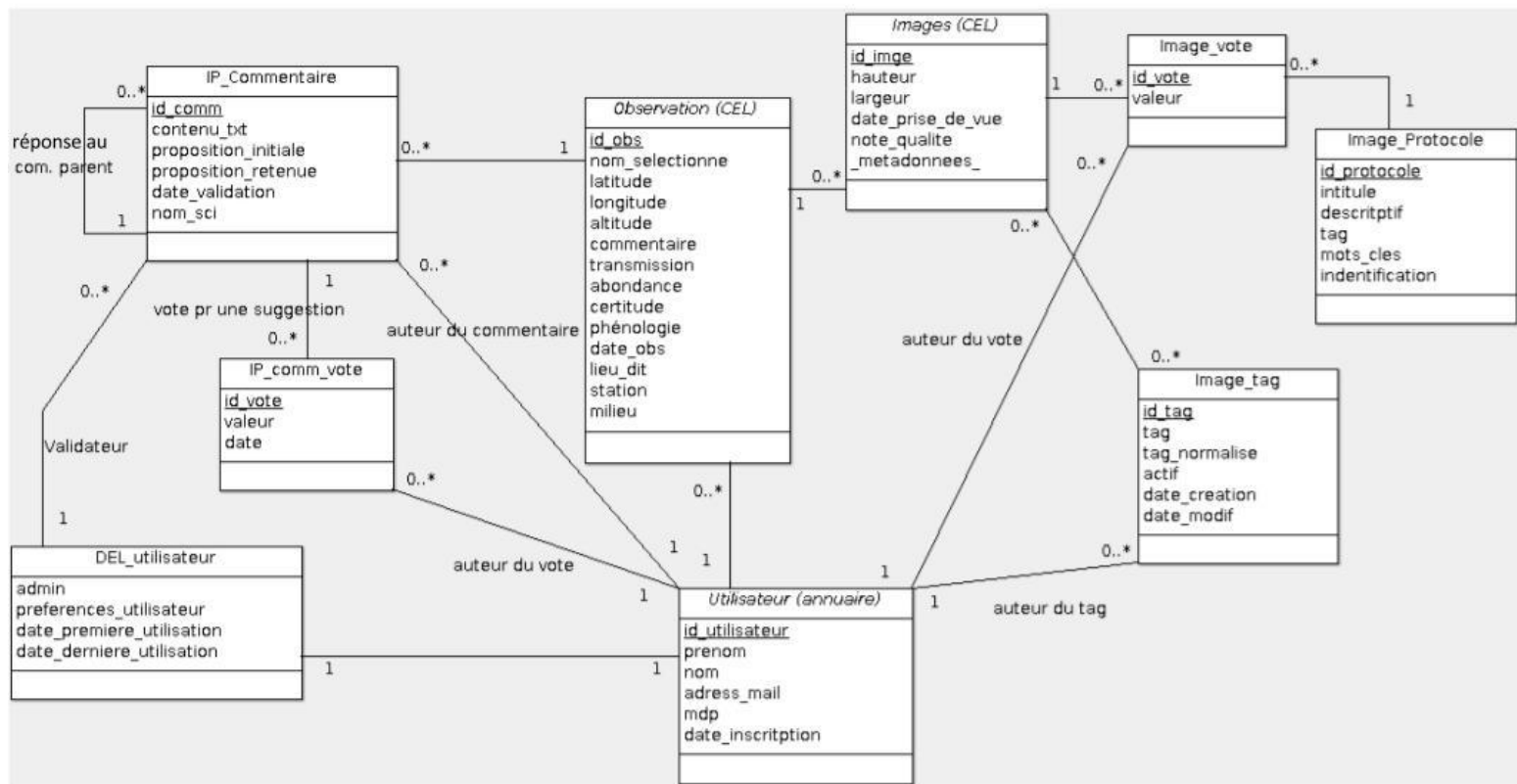


Diagramme de cas d'utilisation :
Visualiser, récupérer, utiliser
des données



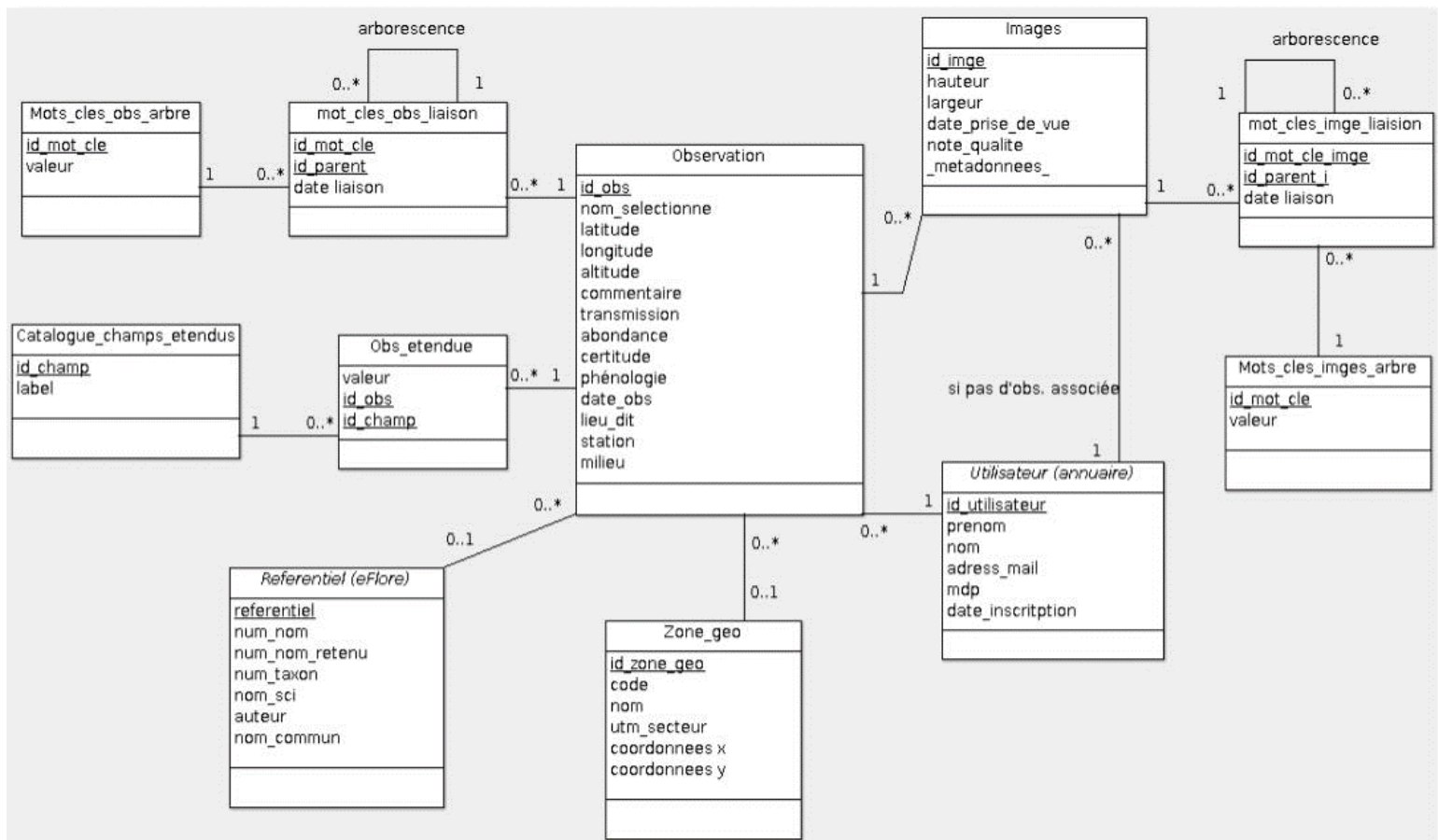
Annexe D. Diagrammes de classes : CEL et DEL

Diagramme de classes : IdentiPlante et PictoFlora (DEL)



Remarques : les tables dont le nom est en italique sont des vues de tables d’une autre base de données (dont le nom est entre parenthèses)

Diagramme de classes : Carnet en ligne (CEL)



Même remarque concernant la représentation des vues.

Annexe E. Synthèse de l'analyse de l'existant

La synthèse complète est disponible à ce lien : https://docs.google.com/document/d/1kUW1gQF1ch-337RuKPnlzIJWnlKwI3-YBUOeKPnW_EA

Objet d'étude	Qualités	Critiques / freins
L'environnement de travail	<ul style="list-style-type: none"> - Richesse des contenus et des outils hébergés. - Richesse et diversité du réseau. - Valeurs de partage et d'échange. - Complémentarité des objectifs individuels et collectifs. - Équipe impliquée et motivée. 	<ul style="list-style-type: none"> - Peu voire pas d'animation de Flora Data. - Accumulation d'outils et de fonctionnalités qui brouille la lisibilité. - Beaucoup de travail.
	<p>Objectifs : Redynamiser la collecte de données participative :</p> <ul style="list-style-type: none"> - En répondant au mieux à une communauté plus jeune et moins experte en botanique, en leur donnant envie de participer au réseau, - En proposant aux amateurs éclairés/botanistes experts des outils performants et originaux, en adéquation avec leurs besoins, - En positionnant les anciens usagers en accompagnement des nouveaux. 	
Les données - Nature des données	<ul style="list-style-type: none"> - Licence libre CC BY-SA - Volume important de données - Un quart des données illustrées - Diversité des données - Présence de données de qualité 	<ul style="list-style-type: none"> - Pas de protocole de saisie - Faible caractérisation des données : difficile d'identifier les observations de « qualité ». - Augmentation incontrôlée du nombre de données et de photos
	<p>Objectifs : Mieux connaître les données de Flora Data : ce qu'elles contiennent, d'où elles viennent, où elles vont - pour être capable de mieux les gérer.</p> <p>Mieux caractériser les données pour faire émerger celles de qualité.</p> <p>Accompagner et former les utilisateurs afin qu'ils soient capables d'entrer des données de qualité et d'identifier une donnée incomplète ou inexploitable.</p>	
Les données - Stockage	<ul style="list-style-type: none"> - Structure des BDD bien connue par l'équipe info - L'organisation est fonctionnelle et relativement simple 	<ul style="list-style-type: none"> - Mauvaise connaissance du fonctionnement et du contenu des flux de données entre PI@ntNet et Tela. - Structure des BDD internes non optimale
	<p>Objectifs : Revoir la structure des BDD pour optimiser le stockage et la caractérisation des données.</p> <p>Mieux comprendre les flux de données entre Tela et ses partenaires (Floris'Tic) pour tirer le meilleur du travail collectif, tout en conservant un contrôle sur les outils et les données.</p>	
Les processus - Saisie et gestion des données	<ul style="list-style-type: none"> - CEL puissant et efficace pour qui sait (bien) s'en servir 	<ul style="list-style-type: none"> - Ergonomie et esthétique du CEL déficientes, long à prendre en main, faible lisibilité des fonctionnalités

	<ul style="list-style-type: none"> - Applications et widgets de saisie des programmes de SP faciles à utiliser pour les moins experts. - Succès et simplicité de PI@ntNet mobile => visibilité, renouvellement du public et nouvelle source de donnée - PI@ntNet mobile seul outil mobile pour saisir des données vers TB actuellement 	<ul style="list-style-type: none"> - Hétérogénéité des protocoles de saisie => hétérogénéité des données entrées - Outils de collecte inadaptés à certains usages (groupes, phytosocio, champignons, mousses) et sources de données (herbiers, biblio). - Mauvaise lisibilité des interactions entre les différents outils (en particulier CEL) - Nouveaux usages mobiles pour lesquels TB n'est pas adapté - Pas d'outil mobile dédié à la collecte de données pour TB
<p>Objectifs : Adapter au mieux les outils de saisie à un public et des usages diversifiés et accompagner leur utilisation dans le but de produire des données de qualité et de les conduire à monter en compétence.</p>		
<p>Les processus - Aide à l'identification</p>	<ul style="list-style-type: none"> - La communauté TB est habituée à IdentiPlante, une partie l'utilise régulièrement - eFlore est relativement complète - Plusieurs référentiels qui sont complémentaires - L'identification automatique de PI@ntNet va en s'améliorant 	<ul style="list-style-type: none"> - Trop de données arrivent (de PI@ntNet) sur IP, et le réseau n'est plus capable de les absorber - Les photos illustrant les taxons dans eFlore ne sont pas filtrées en fonction des révisions d'IP : malgré la vigilance du réseau, on retrouve des erreurs - Processus de « validation » des données dans IP obscur et déficient.
<p>Objectifs : Préserver le fonctionnement et la dynamique du réseau, en optimisant le processus d'aide à l'identification via une meilleure gestion des données entrantes et l'amélioration de la plateforme IdentiPlante.</p> <p>Veiller à ce que les photos illustrant les espèces dans eFlore contiennent le moins d'erreur possible afin de conserver la crédibilité scientifique de Tela Botanica.</p>		
<p>Les processus - Visualisation et analyse des données</p>	<ul style="list-style-type: none"> - PF efficace pour parcourir les photos - Widget d'export simple d'utilisation - Engouement du public pour Smart'Flore et The Plant Game 	<ul style="list-style-type: none"> - Fonctionnement de PictoFlora manque de clarté - Widgets carto peu ergonomiques - Filtres de données pour l'export limités
<p>Objectifs : De manière générale, besoin d'homogénéisation des interfaces et des outils.</p> <p>La cartographie pourrait être un bon outil de communication et de parcours des données si elle était réellement interactive et plus performante.</p>		

Annexe F. Tableau récapitulatif des sites internet ayant fait l'objet de veille

Nom / institution	Institution(s) utilisatrice(s)	Faune/flore	Echelle (locale/nationale)	Application mobile
Biolovision http://data.biolovision.net	LPO (36 bases locales), Station Ornitho Suisse, DDA, LIPU, ICO, Birdlife Austria, OTOP, Orchisauvage, ... => surtout Europe de l'ouest	Faune (Flore)	Toutes échelles ...	NaturaList
iNaturalist https://www.inaturalist.org	(Autres)	Les deux	Mondiale	iNaturalist
Artportalen https://www.artportalen.se		Les deux	Nationale (Suède)	
Biolit / Planète Mer http://www.biolit.fr		Faune et flore littorale	Mondiale (France + Europe et monde)	
E-Observations www.bourgogne-nature.fr	Bourgogne Nature et L'observatoire de la faune de Bourgogne	Faune	Régionale (Bourgogne)	
Noah http://www.projectnoah.org		Les deux	Mondiale	My Noah
Observation.org	Plusieurs sites pour différents pays	Les deux	Nationale (Pays-Bas, Belgique) Mondiale (<i>Other countries</i>)	
Sigogne http://www.sigogne.org		Les deux	Régionale (Franche-Comté)	
SpiPOLL http://spipoll.org		Faune - insectes pollinisateurs (Flore comme support d'observation)	Nationale (France métropolitaine)	
ClicNat http://obs.picardie-nature.org	Picardie Nature Faune Normandie	Faune	Régionale (Normandie et Picardie)	
SIRF http://www.sirf.eu	Groupement Ornithologique et Naturaliste du Nord Pas-de-Calais	Faune	Régionale (Nord-Pas-de-Calais)	

Annexe G. Canevas d'entretien de l'analyse des besoins des utilisateurs – Exemple du Profil 3 : Amateurs éclairés et botanistes experts

Le profil utilisateur

Nom, Prénom, mail	
Activité pro. / loisirs	Étudiant, actif, retraité ? Sciences, biologie, jardinage... ?
Spécialité en relation avec la botanique	Phytosocio, mousses, lichens, etc.
Niveau en botanique	Initié à expert
Découverte TB (Qd, cmt ?)	___ 1 an ___ 5 ans ___ 10 ans ___
Date participation / utilisation outils (Qd, prquoi ?)	___ 1 an ___ 5 ans ___ 10 ans ___
Objectifs participation à TB	

Les usages

Principaux usages	Collecte de données, détermination collaborative, participation au réseau sur des thématiques +/- précises <ul style="list-style-type: none"> • Outils de saisie et de gestion (CEL) • IdentiPlante • PictoFlora • eFlore • Widget carto • Widget d'export • eVeg 	
Cadre, contexte	Loisir, famille, association, professionnel...	
Fréquence	___ 1/an ___ 1/mois ___ 1/semaine ___ 1/jour ___	
Durée des sessions	___ 5 mn ___ 30 mn ___ 1h ___ 4h ___	
Hiérarchisation usages		

Les besoins pour les utilisateurs des outils

Besoins - Élargir au max - Formuler les besoins à partir des usages - <u>Distinguer outils et contenu</u>	<ul style="list-style-type: none"> • Entrer/gérer/partager ses données de manière efficace et rapide • Disposer d'outils complets pour entrer des données de qualité • Faire partie d'une communauté active (et experte ?) • Visualiser un contenu adapté à leurs usages • Exporter ponctuellement ses propres données • Disposer d'outils adaptés à la collecte et la gestion de données de végétation • Disposer d'autres référentiels (bryo, lichénofonge, ...) • Disposer du contenu et des outils traduits • Disposer des référentiels élargis à d'autres zones géographiques 	
Lacunes de Flora Data (vis à vis de ces besoins)		
Hiérarchisation besoins	<i>Formuler : fondamentale, secondaire, accessoire..</i> <i>Note de 1 à 5</i>	

Compléments et ouverture

Outils pas utilisés (lesquels, prquoi ?)	Saisie, aide à la détermination, flore, etc.
Outils ext. à TB (lesquels, prquoi, cmt ?)	
Remarques, suggestions	
Questions	

Participation au panel d'utilisateurs

Suivi de l'évolution des outils, ateliers tests...	Oui / non Tous les outils / certains slmt
Temps à y consacrer (pas avant mars)	
Lieu	Montpellier / ailleurs Sur le terrain / à distance

Annexe H. Exemples de personae

L'ensemble des personae est consultable au lien suivant :

<https://docs.google.com/presentation/d/1kyExlhwrUwvva106r56dAE3NmCludNGjmP2EoAYq6CLg/>



Muriel
61 ans

Prof de chimie à la retraite

"Je souhaite découvrir le monde de la botanique et apprendre à reconnaître les plantes"

Rapport à la botanique

- Connaissances de base
- "J'aime les plantes"

Rapport à la technologie :

- Moyen
- Utilise de temps en temps son smartphone
- Fait des recherches sur internet

Cadre d'utilisation :

Après une balade, Muriel apprécie de pouvoir retrouver ou découvrir le nom d'une plante qu'elle a vue grâce à des photos.

Chez elle quand elle a du temps libre, elle aime découvrir de nouvelles plantes et apprendre des détails sur leur famille, leur utilité, etc.



Lucie
35 ans

Chargée de mission biodiversité dans une association naturaliste

"J'ai besoin d'outils fiables pour saisir et stocker des données"

Rapport à la botanique

- Niveau avancé
- Travail avec des experts
- Collaborations avec des bénévoles amateurs

Rapport à la technologie :

- Très bon
- Utilise quotidiennement les outils informatiques et le web

Cadre d'utilisation :

Dans le cadre de ses missions d'inventaire, l'association de Lucie utilise depuis quelques années les outils de Tela Botanica comme base de données. C'est principalement Lucie qui effectue la saisie sur le site, et ponctuellement des collaborateurs lui reversent leurs données.

Lucie exporte régulièrement une partie des données afin de les exploiter sur SIG. Tous les ans elle exporte également les données de son association pour les transmettre au conservatoire botanique.

Lucie a parfois besoin de chercher des infos sur des espèces, notamment concernant leur répartition.

La débutante

Investissement sur le site :

Muriel a participé au MOOC botanique et s'est prise au jeu ! C'est comme ça qu'elle a eu l'idée de se lancer dans l'inventaire des plantes sauvages de son jardin.

Au printemps, elle compte saisir plus d'observations sur le site.

Envies et attentes :

Apprendre et se faire plaisir.

Disposer d'outils faciles à utiliser, plus ergonomiques et mieux reliés entre eux.

Freins et craintes :

La compréhension du fonctionnement du site et des outils est un élément bloquant pour Muriel. Elle trouve les interfaces trop compliquées et n'est pas à l'aise avec leur utilisation.

A cause de cela, Muriel n'a pas encore découvert la possibilité d'être aidée par le réseau, pensant que ces outils sont trop compliqués pour elle.

La structure

Investissement sur le site :

Les données de la structure sont saisies sur le site de Tela Botanica, mais, dans leur grande majorité, elles ne sont pas rendues publiques.

La structure veille toutefois à effectuer un don tous les ans à Tela Botanica.

Envies et attentes :

Pouvoir renseigner l'observateur en même temps qu'une donnée lui semble indispensable lorsque c'est une structure qui la saisit.

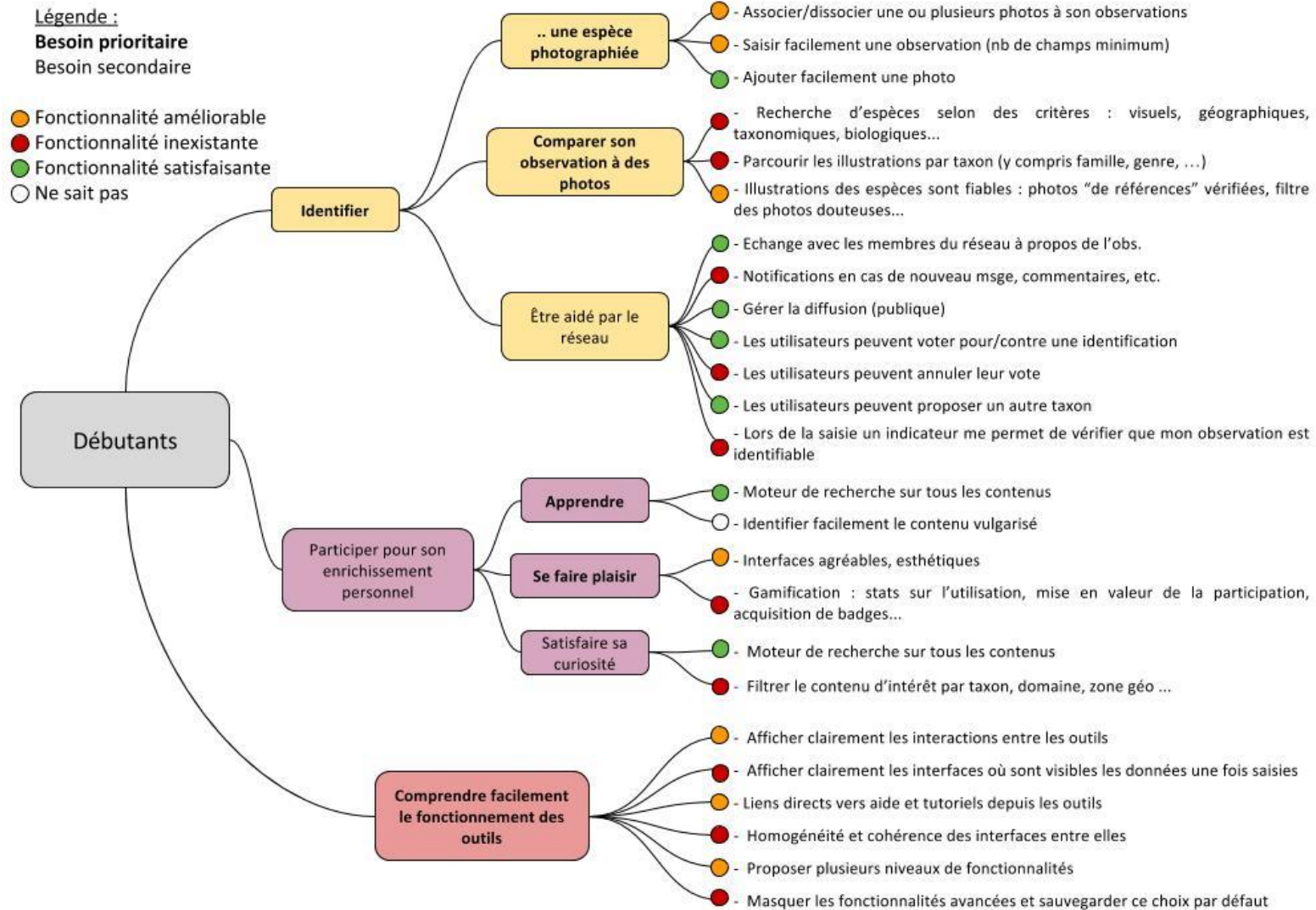
Un outil de saisie de relevés phytosociologiques intéresserait l'association de Lucie.

Freins et craintes :

La compatibilité du format de données avec les structures publiques pose parfois problème, notamment au niveau du référentiel.

Certains bugs font perdre à Lucie beaucoup de temps, la mise en forme lors de l'export notamment.

Annexe I. Diagramme de définition des fonctionnalités répondant aux besoins des utilisateurs – Exemple du profil débutant



Annexe J. Classement et notation des fonctionnalités

Légende :

« Fonctionnalités » (par rapport à l'existant)
Fonctionnalité satisfaisante
Fonctionnalité améliorable
Fonctionnalité inexistante
Fonctionnalité disponible dans le nouveau site

« Note totale »
Fonctionnalité additive
Nécessaire
Indispensable

Les fonctionnalités ci-dessous sont classées par Note totale décroissante. Cette note est la somme des trois critères : importance pour les utilisateurs, risques si absence de la fonctionnalité, influence sur la qualité des données.

Groupe de fonctionnalités	Fonctionnalités	Importance utilisateurs	Risque si absence	Influence sur les données	Note totale
Echanger et s'entraider	Voter pour ou contre le nom proposé pour une observation illustrée	3	3	3	9
Echanger et s'entraider	Proposer un autre nom pour déterminer une observation illustrée	3	3	3	9
Saisir et gérer ses données	Saisir une observation	3	3	3	9
Saisir et gérer ses données	Ajouter facilement une photo	3	3	3	9
Saisir et gérer ses données	Associer une ou plusieurs photos à une observation	3	3	3	9
Saisir et gérer ses données	Modifier une/plusieurs observations	3	3	3	9
Saisir et gérer ses données	Modifier une/plusieurs photos	3	3	3	9
Saisir et gérer ses données	Supprimer une/plusieurs observations	3	3	3	9
Saisir et gérer ses données	Supprimer une/plusieurs photos	3	3	3	9
Echanger et s'entraider	Signaler une photo mal identifiée (depuis n'importe quelle interface)	3	2	3	8
Saisir et gérer ses données	Dissocier une ou plusieurs photo d'une observation	2	3	3	8
Echanger et s'entraider	Echanger avec les membres du réseau à propos de la détermination d'une observation illustrée	3	3	1	7
Echanger et s'entraider	Signaler une photo non identifiable	2	2	3	7
Echanger et s'entraider	Tagger les observations et photos de mauvaise qualité	2	2	3	7
Parcourir le contenu, utiliser le site et les outils	Recherche d'espèces selon des critères taxonomiques	3	3	1	7
Parcourir le contenu, utiliser le site et les outils	Liens directs vers aide et tutoriels depuis les outils	3	3	1	7
Saisir et gérer ses données	Proposer plusieurs niveaux de fonctionnalités pour la saisie	3	2	2	7

Saisir et gérer ses données	Dupliquer une observation (pour ne pas avoir à ressaisir tous les champs)	3	2	2	7
Utiliser des données	Assurer la fiabilité des illustrations des espèces : afficher d'abord des photos de référence , vérifiées	3	3	1	7
Utiliser des données	Assurer la fiabilité des illustrations des espèces : filtrer les photos douteuses	3	3	1	7
Parcourir le contenu, utiliser le site et les outils	Afficher et expliquer les interactions entre les outils	3	2	1	6
Parcourir le contenu, utiliser le site et les outils	Liens directs vers le contact de l'équipe en cas de bug	2	3	1	6
Saisir et gérer ses données	Ajouter plusieurs photos à la fois	2	2	2	6
Saisir et gérer ses données	Gérer la diffusion de ses données : publique/privée	2	2	2	6
Saisir et gérer ses données	Choisir d' afficher ou masquer les fonctionnalités avancées et conserver ce choix par défaut	2	2	2	6
Saisir et gérer ses données	Caractériser ses propres observations grâce à des tags	2	2	2	6
Echanger et s'entraider	Annuler le vote pour/contre un nom	3	1	1	5
Saisir et gérer ses données	Créer des champs personnalisés et les conserver par défaut	2	1	2	5
Utiliser des données	Exporter des données	2	3	0	5
Utiliser des données	Filtrer les données à exporter selon des critères géo, taxonomiques, date, etc.	2	3	0	5
Parcourir le contenu, utiliser le site et les outils	Afficher et expliquer les interfaces où sont visibles les données une fois saisies	2	2	1	5
Echanger et s'entraider	Echanger avec les membres du réseau sur des forums thématiques	2	2	0	4
Parcourir le contenu, utiliser le site et les outils	Recherche d'espèces selon des critères géographiques	2	1	1	4
Parcourir le contenu, utiliser le site et les outils	Recherche d'espèces selon des critères visuels	2	1	1	4
Parcourir le contenu, utiliser le site et les outils	Moteur de recherche sur tous les contenus	2	2	0	4
Saisir et gérer ses données	Gérer la diffusion de ses données : floutage géographique	1	1	2	4
Saisir et gérer ses données	Filtrer ses propres données par date, taxon, zone géographique, etc.	2	1	1	4
Echanger et s'entraider	Echanger avec les membres du réseau à propos de contenus grâce à des commentaires	2	1	0	3
Parcourir le contenu, utiliser le site et les outils	Identifier facilement le type de contenu : vulgarisé, professionnel, sur une thématique particulière, etc.	1	2	0	3
Parcourir le contenu, utiliser le site et les outils	Filtrer les contenus par centre d'intérêt : zone géographique, taxon, domaine botanique, etc.	2	1	0	3

Saisir et gérer ses données	Vérifier que l'observation est identifiable par les autres membres, grâce à un indicateur lors de la saisie		1	2	3
Utiliser des données	Choisir le format d'export	2	1	0	3
Utiliser des données	Afficher le type de licence et les conditions d'utilisation associées	1	2	0	3
Animer le site	Notification en cas de réponse à une demande d'aide à l'identification		1	1	2
Animer le site	Mettre en valeur la participation : affichage de statistiques	1	1	0	2
Saisir et gérer ses données	Gérer la diffusion de ses données : diffusion à un groupe restreint	1	0	1	2
Saisir et gérer ses données	Retrouver ses observations et photos dans son espace personnel		1	1	2
Utiliser des données	Choisir les champs à exporter	1	1	0	2
Utiliser des données	Télécharger une photo depuis n'importe quelle interface	1	1	0	2
Utiliser des données	Exporter plusieurs photos à la fois	1	1	0	2
Utiliser des données	Filtrer les photos par auteur, taxon, qualité, date, etc.	1	1	0	2
Animer le site	Notification en cas de réponse à un commentaire		1	0	1
Animer le site	Notification en cas de nouveau message personnel		1	0	1
Animer le site	Notification en cas de nouveau message sur un forum suivi		1	0	1
Animer le site	Mettre en valeur la participation : badges à gagner	1	0	0	1
Animer le site	Mettre en valeur la participation : dernières photos ajoutées	1	0	0	1
Parcourir le contenu, utiliser le site et les outils	Parcourir les fiches espèces (en particulier les illustrations) par taxon		1	0	1
Utiliser des données	Choisir le format d'export (taille/qualité) de l'image		1	0	1
Echanger et s'entraider	Echanger avec les membres du réseau grâce à un service de messagerie		0	0	0
Parcourir le contenu, utiliser le site et les outils	Recherche d'espèces selon des critères biologiques		0	0	0
Utiliser des données	Afficher un aperçu des données filtrées		0	0	0

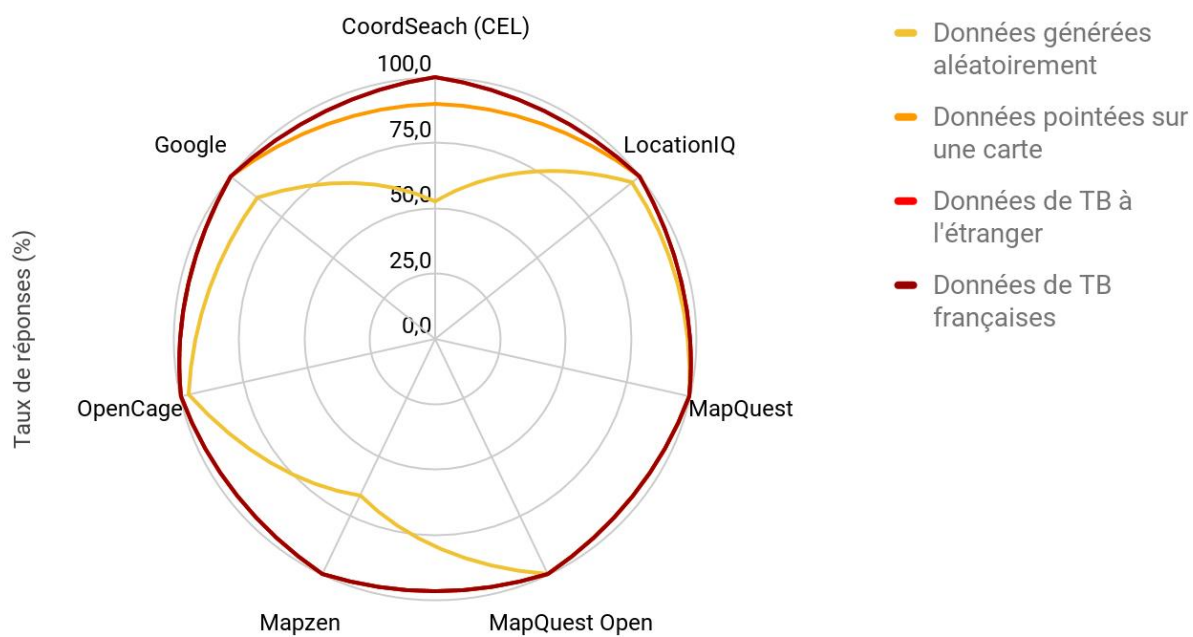
Annexe K. Taux de réponse moyens

Ci-dessous, les taux de réponse en pourcentage, classés par taux de réponses complètes croissants (Δ). Ces taux de réponse sont issus de la moyenne des taux de réponse des 4 jeux de données.

	Taux de réponses (non vides)	Taux de réponses "complètes" (au moins pays-localité) Δ
CoordSearch (CEL)	85,64	80,64
LocationIQ	99,15	72,43
OpenCage	99,15	70,78
MapQuest	100	67,43
MapQuest Open	100	64,43
Google	96,61	56,36
Mapzen	91,53	41,66
Data Science Toolkit	91,53	0,83

Annexe L. Taux de réponse par jeu de données

Taux de réponses (non vides) par service et par jeu de données



Annexe M. Calcul de l'altitude d'une observation à partir de ses coordonnées géographiques : premiers éléments pour une étude comparative des services web

L'**altitude** d'une observation botanique est une information intéressante à partir des coordonnées, dans l'objectif par exemple de calculer des profils altitudinaux de la répartition des espèces – y compris pour des observations déjà saisies.

La principale source de données libres sur l'altitude est la NASA, grâce à la **Shuttle Radar Topography Mission (STRM)** qui a "collecté des données topographiques sur près de 80% de la surface terrestre, créant ainsi la première base de données altitudinales (presque) mondiale." Dans ce cadre, les données du CGIAR-CSI (Consultative Groups on International Agricultural Research - Consortium for Spatial Information), sont celles initialement produites par la NASA dans le cadre de la STRM mais re-traitées par le CGIAR-CSI afin de [compléter les données manquantes dans certaines zones](#). Ce jeu de données fournit une information d'altitude pour (presque) tout le globe, à une **résolution de 90m** et est disponible en libre téléchargement.

L'analyse précédentes des services de geocoding a permis d'identifier les trois API suivantes à tester :

Fournisseur - nom du service	Source de données	Prix (r = requête)	Libre ?
Google - Google Maps elevation API	Gmaps	2 500 r/jr, 50 r/sec • 0,5\$ /1000 r suppl Premium	Non
Mapzen - Elevation Service	SRMT (NASA) GMTED NED ETOPO1 DEMs	20 000 r/jr (pas encore de plan payant)	Open Source (Joerd) Open Data
Data Science Toolkit - Coordinates to Statistics	NASA and the CGIAR CSI	Gratuit (API ou installation sur serveur local)	Open Source Open Data

Avec les mêmes jeux de données et une méthode semblable à celle utilisée pour l'étude du reverse geocoding, on peut comparer la performance des services web précédents pour le calcul de l'altitude à partir des coordonnées géographiques.

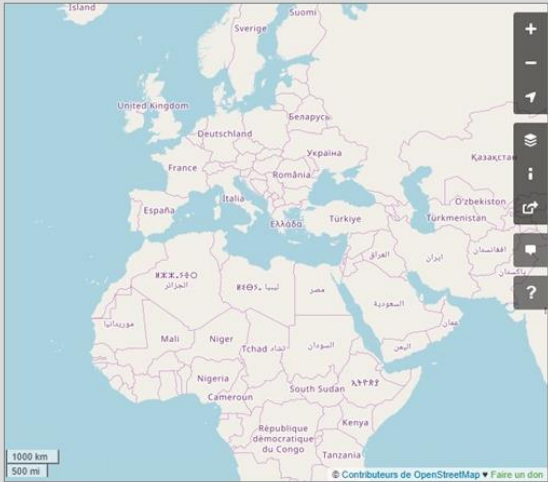
Annexe N. Maquettes : saisie de localisation

Le scénario d'utilisation est rapidement décrit en III.C.2.e). L'ensemble des étapes de la maquette est également disponible au format PDF : https://drive.google.com/file/d/0B2chGfR_7k9DbC1GZWdLSGw3cms.

Etape 1 : vue principale, aucune action

Localisation ▽

Rechercher un lieu :



Latitude : Longitude :

Altitude (m) :

Zone géographique :

Vous n'êtes pas d'accord ?

Pays :

Précision (m) :

Floutage : ▼

Lieu-dit :

Milieu (Corine biotope) :


Commentaires sur la localisation :

Etape 2 : recherche d'un lieu en utilisant l'autocomplétion

Localisation ▽

Rechercher un lieu :

Grenoble (38)
 Grenois (58)
 Grenola
 Grenora



Latitude : Longitude :

Altitude (m) :

Zone géographique :

Vous n'êtes pas d'accord ?

Pays :

Précision (m) :

Floutage : ▼

Lieu-dit :

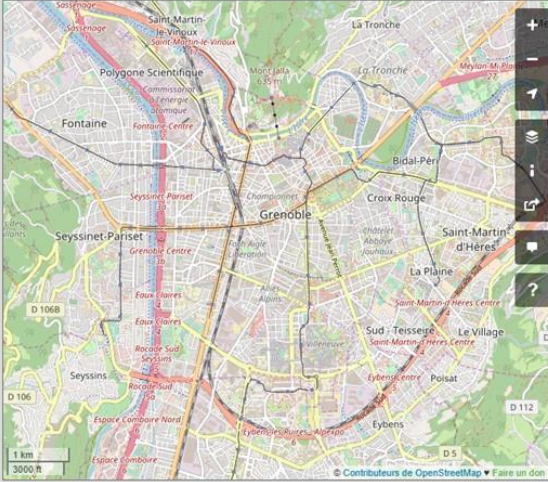
Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 3 : La carte zoom sur le lieu recherché

Localisation ▼

Rechercher un lieu :



Latitude :

Longitude :

Altitude (m) :

Zone géographique :

Vous n'êtes pas d'accord ?

Pays :

Précision (m) :

Floutage :

Lieu-dit :

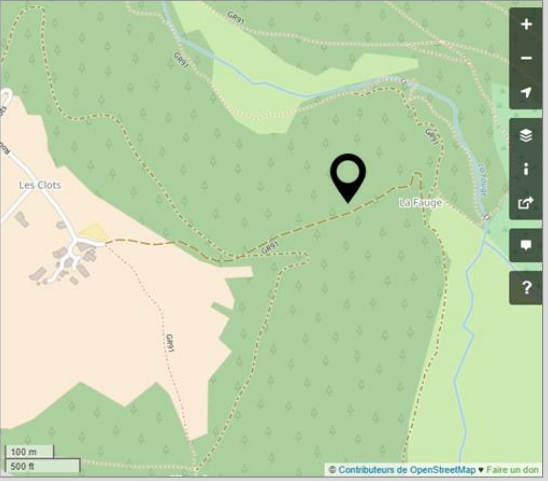
Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 4 : l'utilisateur déplace la carte manuellement pour rechercher l'endroit précis de l'observation. En cliquant sur celui-ci, les coordonnées géographiques sont extraites et l'altitude et la zone géographique en sont déduites grâce à différents web services. Le bouton « Vous n'êtes pas d'accord ? » devient cliquable.

Localisation ▼

Rechercher un lieu :



Latitude :

Longitude :

Altitude (m) :

Zone géographique :

Vous n'êtes pas d'accord ?

Pays :

Précision (m) :

Floutage :

Lieu-dit :

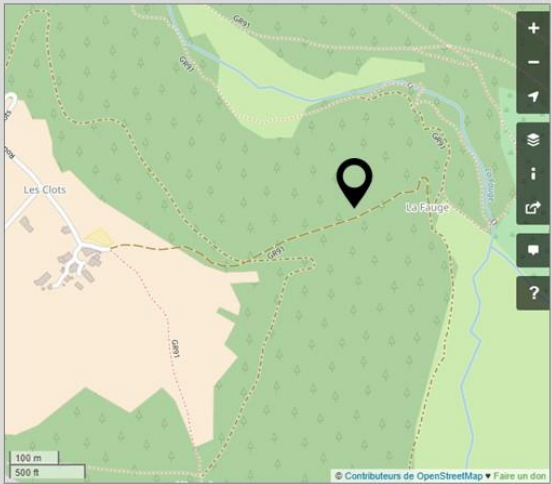
Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 5 : En cliquant sur le bouton « Vous n'êtes pas d'accord », un nouveau champ apparait. L'utilisateur peut ainsi saisir une autre localité s'il estime que le service de reverse geocoding se trompe. Dans ce cas, un message d'alerte encourage l'utilisateur à utiliser l'autocomplétion. Le lien « En savoir plus » renvoie vers une pop-up ou un wiki, expliquant l'importance de renseigner soigneusement ces champs, et notamment le rôle de l'autocomplétion.

Localisation ▼

Rechercher un lieu :



Latitude : Longitude :

Altitude (m) :

Zone géographique :

Localité :

Attention : vous avez saisi une localité sans autocomplétion, êtes-vous sûr de vous ? [En savoir plus](#)

Pays :

Précision (m) :

Floutage : ▼

Lieu-dit :

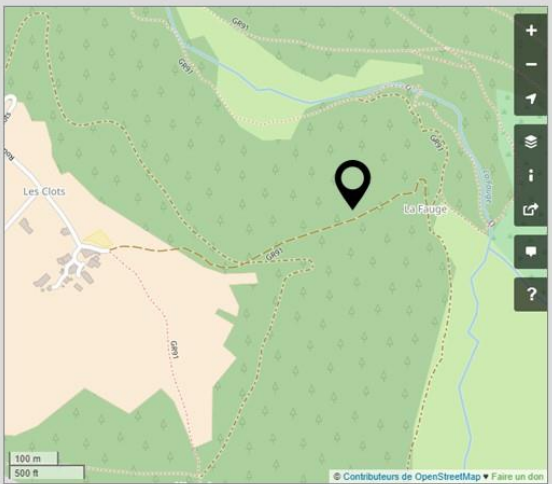
Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 6 : Utilisation de l'autocomplétion

Localisation ▼

Rechercher un lieu :



Latitude : Longitude :

Altitude (m) :

Zone géographique :

Localité :

Attention : vous avez de vous ? [En savoir pl](#)

Pays :
Correns (83)
Correntes
Correntina

Précision (m) :

Floutage : ▼

Lieu-dit :

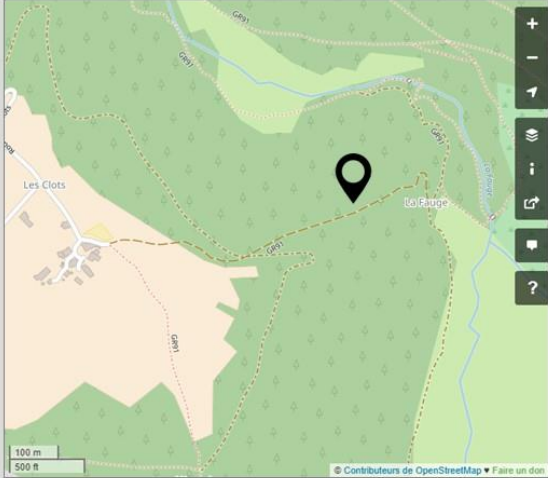
Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 7 : Après autocomplétion : plus de message d'alerte, mais toujours le lien d'information.

Localisation

Rechercher un lieu :



Latitude : Longitude :

Altitude (m) :

Zone géographique :

Localité :

Merci d'avoir utilisé l'autocomplétion ! [En savoir plus](#)

Pays :

Précision (m) :

Floutage :

Lieu-dit :


Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 8 : Saisie d'une distance dans le champ « précision » et adaptation du cercle d'incertitude correspondant sur la carte

Localisation

Rechercher un lieu :



Latitude : Longitude :

Altitude (m) :

Zone géographique :

Localité :

Merci d'avoir utilisé l'autocomplétion ! [En savoir plus](#)

Pays :

Précision (m) :

Floutage :

Lieu-dit :


Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 9 : Choix d'un mode de floutage

Localisation

Rechercher un lieu : Grenoble



Latitude : 45.05692 Longitude : 5.57456

Altitude (m) : 1300

Zone géographique : Villard-de-Lans (38)

Localité : Corrençon-en-Vercors (38)

Merci d'avoir utilisé l'autocomplétion ! [En savoir plus](#)

Pays : France

Précision (m) : 150

Floutage : Aucun

Lieu-dit :


Milieu (Corine biotope) :

Commentaires sur la localisation :

Etape 10 : Complétion des champs complémentaires.

Localisation

Rechercher un lieu : Grenoble



Latitude : 45.05692 Longitude : 5.57456

Altitude (m) : 1300

Zone géographique : Villard-de-Lans (38)

Localité : Corrençon-en-Vercors (38)

Merci d'avoir utilisé l'autocomplétion ! [En savoir plus](#)

Pays : France

Précision (m) : 150

Floutage : A la commune

Lieu-dit : Vallon de la Fauge

Milieu (Corine biotope) :

Commentaires sur la localisation :

Proche chemin sur la gauche

Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?

Résumé

Aujourd'hui, Tela Botanica – « le réseau des botanistes francophones » – agrège un grand nombre de données d'observations botaniques (ou occurrences de taxon). Celles-ci ont pour particularité d'avoir été collectées par des botanistes amateurs – au sens non professionnels – sans toujours répondre à un protocole scientifique défini. Pourtant, un certain nombre d'entre elles pourraient venir compléter les données « professionnelles » utilisées en recherche, pour la gestion du territoire ou la préservation de l'environnement.

Une partie de ces données est cependant considérée de « mauvaise qualité », c'est-à-dire incomplète ou incohérente. Afin de valoriser les connaissances et les contributions des amateurs de botanique, que doit mettre en œuvre l'association pour les diffuser ? C'est la question qui a motivé ce mémoire.

Dans un premier temps, l'analyse du système d'information dans son ensemble a permis d'identifier les points critiques, impliqués dans la qualité des données et donc cibles d'améliorations. Cette analyse s'est déroulée en plusieurs étapes : état des lieux de l'existant, veille technologique, enquêtes sur les besoins des utilisateurs, analyse fonctionnelle et enfin spécifications. Parmi ces spécifications, la définition d'un format standard pour les données de Tela Botanica a constitué une base, sur laquelle se sont construits les cahiers des charges des outils et des plateformes de gestion des données d'observation.

L'une des pistes d'amélioration identifiée grâce à cette analyse concerne la localisation des données d'observation. Cette information, indispensable pour décrire l'occurrence d'un taxon, peut être délicate à renseigner correctement. Afin de déterminer les moyens d'optimiser sa saisie, un état de l'art des outils de collecte de données naturalistes a d'abord permis d'identifier les solutions existantes. Un cas d'étude plus spécifique, portant sur la performance du géocodage inversé, a abouti à la définition d'un protocole de saisie de la localisation. Enfin, une étude a permis d'évaluer les conséquences de ce nouveau protocole sur les autres interfaces en lien avec les données d'observation.

Cette étude a montré notamment que la transformation des outils proposés par Tela Botanica (notamment pour la saisie et la révision collaborative), en plus d'enrichir l'expérience des utilisateurs, entraînerait l'amélioration des données d'observation produites. En effet, elle permet d'accompagner la montée en compétence des membres du réseau, tout en limitant la saisie et la diffusion de données incomplètes ou incohérentes.

Ainsi, en adaptant aujourd'hui ses outils au standard d'échange de données naturalistes en passe d'être adopté nationalement, l'association leur assure une certaine pérennité. L'amélioration potentielle et donc la diffusion des données, issues notamment des sciences participatives, permettrait à Tela Botanica de mettre en lumière l'étendue des compétences et des connaissances de son réseau d'amateurs.

Mots-clés

Botanique, donnée d'observation naturaliste, qualité des données, sciences participatives, protocole de saisie, réseau d'amateur

Pour citer ce document : MARY Laura, 2017. *Comment améliorer, au sein d'un système d'information existant, la qualité des données d'observation botanique collectées par des amateurs ?* Mémoire de fin d'étude, diplôme d'ingénieur agronome, spécialité AgroTIC, Montpellier SupAgro.