TN09 : Stage assistant ingénieur

De la conception à l’exploitation d’une base de données

sur les exploitations agricoles de l’Océan Indien

CIRAD, 40 chemin du grand canal, Saint Clotilde, la Réunion



Figure  : L'agriculture familiale

Encadrement :

Tutrice entreprise : Sandrine Auzoux

Responsable entreprise : Mialet-Serra Isabelle

Tutrice UTC : Marie-Hélène Abel

Stagiaire : Lucas Le Moine (GI03)

Lucas.le-moine@etu.utc.fr

Du 01/09/2020 au 12/02/2021

Remerciements

Je souhaite remercier en premier lieu toutes les personnes qui ont rendu ce stage possible.

Tout d’abord, je remercie ma suiveuse de stage à l’UTC, madame ABEL pour l’appui qu’elle a pu être pendant cette période.

Ensuite, ma tutrice de stage, Madame AUZOUX mérite, pour son accueil, son aide et sa clairvoyance, mes plus sincères remerciements.

Je remercie également profondément Madame MIALET-SERRA, coordinatrice du projet de l’observatoire des agricultures de l’océan Indien, qui m’a permis d’effectuer ce stage au sein de ce groupe.

Je voudrai aussi remercier Messieurs Bosc et Bélières qui ont été d’excellents interlocuteurs et qui ont permis un contact avec le domaine métier durant la conception de la base de données.

Aussi, je remercie tous les membres du CIRAD à Saint Denis pour leur accueil et les connaissances qu’ils m’ont transmises de par la diversité de leurs domaines.

Enfin, je tiens à remercier madame LY pour ses réponses claires et rapides malgré la charge de travail qu’elle supporte.

Table des matières

[I. Table des illustrations 5](#_Toc60841409)

[II. Introduction 6](#_Toc60841410)

[III. Présentation du CIRAD 7](#_Toc60841411)

[A. Faits et chiffres 7](#_Toc60841412)

[B. Objectifs actuels 8](#_Toc60841413)

[C. Contexte du stage 8](#_Toc60841414)

[D. Département 11](#_Toc60841415)

[E. L’équipe 11](#_Toc60841416)

[IV. Dans les grandes lignes 12](#_Toc60841417)

[A. Missions 12](#_Toc60841418)

[B. Obstacles 13](#_Toc60841419)

[C. Méthodologie 13](#_Toc60841420)

[D. Outils utilisés 14](#_Toc60841421)

[1. CodeIgniter 14](#_Toc60841422)

[2. Leaflet 14](#_Toc60841423)

[3. Git et GitHub 14](#_Toc60841424)

[4. D3.js 14](#_Toc60841425)

[5. Bootstrap 15](#_Toc60841426)

[6. SQL Power Architect 16](#_Toc60841427)

[7. PostgreSQL et PgAdmin 16](#_Toc60841428)

[8. Microsoft Office Access 16](#_Toc60841429)

[9. Gantt Project 17](#_Toc60841430)

[E. Planning 17](#_Toc60841431)

[V. Déroulement du projet 18](#_Toc60841432)

[A. Compréhension des intentions 18](#_Toc60841433)

[B. Exploration des Dataverses 18](#_Toc60841434)

[C. Benchmark 18](#_Toc60841435)

[D. Recueil des besoins des utilisateurs 20](#_Toc60841436)

[E. Recueil des données 21](#_Toc60841437)

[F. Réalisation du cahier des charges 21](#_Toc60841438)

[1. Profils 21](#_Toc60841439)

[2. Fonctionnalités 22](#_Toc60841440)

[3. Charte graphique 22](#_Toc60841441)

[4. Droits des utilisateurs 23](#_Toc60841442)

[G. Conception de la base de données 24](#_Toc60841443)

[1. Général 24](#_Toc60841444)

[2. Contraintes 24](#_Toc60841445)

[3. Problèmes rencontrés 25](#_Toc60841446)

[H. Intégration de 2 bases de données 26](#_Toc60841447)

[1. Généralités 26](#_Toc60841448)

[2. Problèmes rencontrés 27](#_Toc60841449)

[I. Vues 29](#_Toc60841450)

[J. Réalisation du prototype 31](#_Toc60841451)

[1. Base de l’application Web 31](#_Toc60841452)

[2. Avancement de l’interface 32](#_Toc60841453)

[3. Data visualisations 34](#_Toc60841454)

[K. Documentation 35](#_Toc60841455)

[VI. Bilan d’expérience 36](#_Toc60841456)

[A. Difficultés 36](#_Toc60841457)

[B. Apports personnels 36](#_Toc60841458)

[VII. Glossaire 38](#_Toc60841459)

# Table des illustrations

[Figure 1 : L'agriculture familiale 1](#_Toc60841460)

[Figure 2 : Carte des directions régionales du CIRAD 7](#_Toc60841461)

[Figure 3 : Logo de la FAO 8](file:///C:\wamp64\www\OA-OI_git\Autres_livrables\Rapport_de_stage.docx#_Toc60841462)

[Figure 4 : La Commission de l'Océan Indien (COI) 10](file:///C:\wamp64\www\OA-OI_git\Autres_livrables\Rapport_de_stage.docx#_Toc60841463)

[Figure 5 : Diagramme de Gantt du stage à l'origine 17](#_Toc60841464)

[Figure 6 : Benchmark 20](#_Toc60841465)

[Figure 8 : Logo de la Prérad-OI 23](#_Toc60841466)

[Figure 7 : Proposition de logo OA-OI 23](file:///C:\wamp64\www\OA-OI_git\Autres_livrables\Rapport_de_stage.docx#_Toc60841467)

[Figure 9 : Matrice des droits sur une exploitation 23](#_Toc60841468)

[Figure 10 : Extrait du diagramme Power Architect 25](#_Toc60841469)

[Figure 11 : Table matériel manuel d'une des bases de données récupérées 28](#_Toc60841470)

[Figure 12 : Extrait du questionnaire d'une des bases de données récupérées 28](#_Toc60841471)

[Figure 13 : Extrait d'un questionnaire avec codification 29](#_Toc60841472)

[Figure 14 : Vue valeur ajoutée brute 30](#_Toc60841473)

[Figure 15 : Schéma architecture Modèle\_Vue\_Controleur 32](#_Toc60841474)

[Figure 16 : Page de présentation de l'observatoire 33](#_Toc60841475)

[Figure 17: Recherche dans l'annuaire 33](#_Toc60841476)

[Figure 18 : Data visualisation sur la composition des revenus 34](#_Toc60841477)

[Figure 19 : Data visualisation sur l’Afrique 35](#_Toc60841478)

[Figure 20 : Histogramme du CIRAD 39](#_Toc60841479)

# Introduction

Le stage d’assistant ingénieur TN09, dans le cadre du parcours universitaire de l’Université Technologique de Compiègne, branche Génie Informatique a été réalisé au sein du Centre de Coopération Internationale en Recherche Agronomique pour le développement (CIRAD). Ce stage de 24 semaines a été effectué en télétravail entre le 1er et le 18 septembre, puis sur le site de la direction régionale du CIRAD à Saint Denis de la Réunion entre le 18 septembre et le 12 février. Il n’a pas été affecté par le reconfinement fin octobre.

Ce rapport de stage contient une brève présentation de l’institut d’accueil et du contexte du projet. Sont ensuite discutés les objectifs et des livrables attendus ainsi que les technologies employées. Enfin, certaines réflexions personnelles seront évoquées à la fin du rapport.

Le but principal du stage est la conception d’une base de données sur des exploitations agricoles ainsi que l’exploitation de cette dernière. Dans le même temps, un cahier des charges et une interface web étaient à développer selon le temps et les ressources disponibles.

# Présentation du CIRAD

## Faits et chiffres

Le CIRAD est un Établissement Public à caractère industriel et commercial (EPIC) français fondé en 1984. Il est spécialisé dans la recherche agronomique dans les régions aux climats chauds, comme en témoignent les différentes implantations des directions régionales du CIRAD sur la carte ci-dessous. Ce type d’établissement est créé pour répondre à un besoin qui, pour des raisons concurrentielles, ne peut pas être satisfait par une entreprise privée. En raison de la diversité de ses activités, le CIRAD est à la fois placé sous la tutelle du ministère de l’Enseignement supérieur et de la Recherche et du ministère des Affaires étrangères. Cet établissement couvre les deux tiers de ses frais grâce aux financements publics et le dernier tiers est le fruit de ressources contractuelles.

Le CIRAD emploie actuellement 1650 personnes, dont 800 chercheurs, répartis en 3 départements scientifiques et 33 unités de recherche (mixtes ou propres). Le partenariat fait partie de l’identité du CIRAD et les 12 directions régionales (dont celle de la Réunion) permettent une grande coopération. Par ailleurs, le partage des connaissances, la qualité de la recherche, l’ouverture internationale ainsi que l’engagement pour le développement sont les clés de l’action de cet établissement.

Les bureaux de la station « La Bretagne » à Sainte-Clotilde accueillent la direction régionale Réunion Mayotte (DRRM), où le stage s’est déroulé. Il est lié à la station de La Mare, à 4 km, où se trouvent 12 ha de parcelle d’expérimentation sur la canne à sucre. Le CIRAD dispose également d’un site important à Saint-Pierre, au sud de l’île. On y trouve des bureaux, des laboratoires, des serres et des parcelles d’expérimentation. Au total, ce sont 35 ha de terrains d’expérimentation qui permettent aux 60 chercheurs et 110 techniciens de mener à bien leurs essais.



Figure : Carte des directions régionales du CIRAD

## Objectifs actuels

Le CIRAD vient de signer avec ses ministères de tutelle son nouveau Contrat d’Objectifs et de Performances pour la période 2019-2023. Le CIRAD s’engage autour de quatre grandes ambitions pour contribuer à l’atteinte des Objectifs du développement durable au Sud, en particulier ceux sur l’éradication de la faim et de la pauvreté (ODD 2, ODD 1).

Pour se donner les moyens d’accompagner les sociétés du Sud dans des trajectoires de développement durable, le CIRAD s’engage ainsi sur quatre grandes ambitions pour la nouvelle période 2019-2023 : la science, le partenariat, la formation et l’impact.

Le stage est quant à lui réalisé grâce au financement du programme FEDER INTERREG V. Il est attribué à des projets visant à renforcer l’insertion régionale de Mayotte et de la Réunion dans l’Océan Indien.

## Contexte du stage

#### La naissance de l’Observatoire des Agricultures du Monde (OAM)

La décennie 2019-2028 a été proclamée comme celle de l’agriculture familiale par les Nations Unies. Ce type d’agriculture représente pas moins de 90% des exploitations agricoles dans le monde qui sont ainsi au nombre de 500 millions. Elles produisent également 80% de la production alimentaire mondiale alors que près des 3 quarts d’entre elles ne dépassent pas un hectare. Pourtant, la majorité de ces agriculteurs sont touchés par la pauvreté et sont vulnérables aux aléas météorologiques et aux fluctuations des prix des marchés.

Par ailleurs, les Nations Unies ont ciblé 17 objectifs du développement durable (ODD) à réaliser d’ici à 2030. Ces objectifs interconnectés ont pour but d’assurer que chaque être humain connaisse l’égalité, la prospérité, la santé, la justice et la paix plutôt que la faim et la pauvreté. Les deux premiers objectifs, et les plus importants, étant éradiquer la faim et la pauvreté, l’agriculture familiale a un rôle immense à jouer pour atteindre ces objectifs. Au total, ce ne sont pas moins de 10 des ODD qui sont liés directement à l’agriculture familiale.

Aujourd’hui, à l’heure où les effets du changement climatique se font ressentir, les exploitants agricoles doivent produire pour nourrir une population toujours plus nombreuse en limitant leur impact sur l’environnement.

Malgré leur importance, autant en terme de nombre qu’en terme de potentiel, l’agriculture familiale n’est que très peu connue et reconnue. Très peu de données existent sur les structures de ces exploitations familiales, les problèmes qu’elles rencontrent ou le travail qu’elles génèrent.

Aujourd’hui, coexistent des enquêtes LSMS (« Living Standard Measurement Study ») qui se concentrent sur les conditions de vie des ménages, et les recensements agricoles qui explorent les structures des exploitations. Cependant, aucun lien n’existe entre ces études et les durées de livraison de ces études sont soit variables soit extrêmement longue (le recensement agricole est une collecte décennale). Cela, additionné à des gouvernances non coordonnées génère un gaspillage de ressources considérable.

Figure : Logo de la FAO

C’est dans ce contexte que M. BOSC, ancien chercheur au CIRAD aujourd’hui détaché à la FAO, a lancé le concept de l’Observatoire des Agricultures du Monde (OAM). Le terme agriculture est employé au pluriel pour signifier la diversité des types d’exploitations agricoles. Cet observatoire a pour but de caractériser cette diversité ainsi que de rassembler les parties prenantes autour de données partagées et cohérentes afin d’élaborer des stratégies inclusives et différenciées.

#### Les capitaux pour caractériser les exploitations agricoles

Le but de ce projet est de collecter, de rassembler et de visualiser les données sur les exploitations agricoles selon leur structure. Pour cela, M. Bosc a établi une typologie permettant de catégoriser les exploitations selon 5 capitaux :

* Le capital humain prend en compte la main d’œuvre disponible et employée de manière qualitative et quantitative. Des variables comme l’alphabétisation, l’âge ou le sexe sont ainsi explorées. L’origine de la main d’œuvre (familiale ou non) est aussi un point clé dans la construction de la typologie.
* Le capital naturel se compose des terres et de leur utilisation (culture, bois, aquaculture, jachère, …) mais également de leur mode de faire valoir ;
* Le capital financier se compose des actifs monétaires ou capitalisés mobilisables par le chef d’exploitation ;
* Le capital physique comprend l’ensemble des entités mises en œuvres dans le cadre de la production. Les infrastructures, les aménagements, les équipements ainsi que les animaux et les cultures pérennes font partie de ce capital ;
* Le capital social repose quant à lui sur les relations tissées par les membres de l’exploitation avec les communautés et réseaux professionnels locaux. Ce capital valorise les participations à des organisations de producteurs tout comme l’entraide entre des exploitants.

Caractériser ainsi les exploitations agricoles devrait permettre d’influencer les politiques mises en place dans l’agriculture. Des stratégies différenciées peuvent être mises en place à partir des informations sur les capitaux. En effet, on n’aide pas de la même manière une exploitation de 100 ha et appartenant à un grand groupe et une exploitation de 0.75 ha gérée par une famille. Le but était donc de viser principalement des acteurs avec une capacité d’action comme des agences de développement ou des organisations de producteurs. Par ailleurs, M. Bosc souhaite donner également une coloration développement durable à l’observatoire.

#### Situation en septembre 2020 et création de l’Observatoire des Agricultures de L’Océan Indien (OA-OI)

Au début du stage, en septembre 2020, le projet reposait en grande partie sur un guide opérationnel rédigé par M. Bosc. Ce guide regroupe les différentes étapes de la réalisation d’un « observatoire ». On y trouve les définitions des concepts importants, mais aussi les méthodes pour élaborer un questionnaire ou choisir les échantillons d’exploitations à sonder.

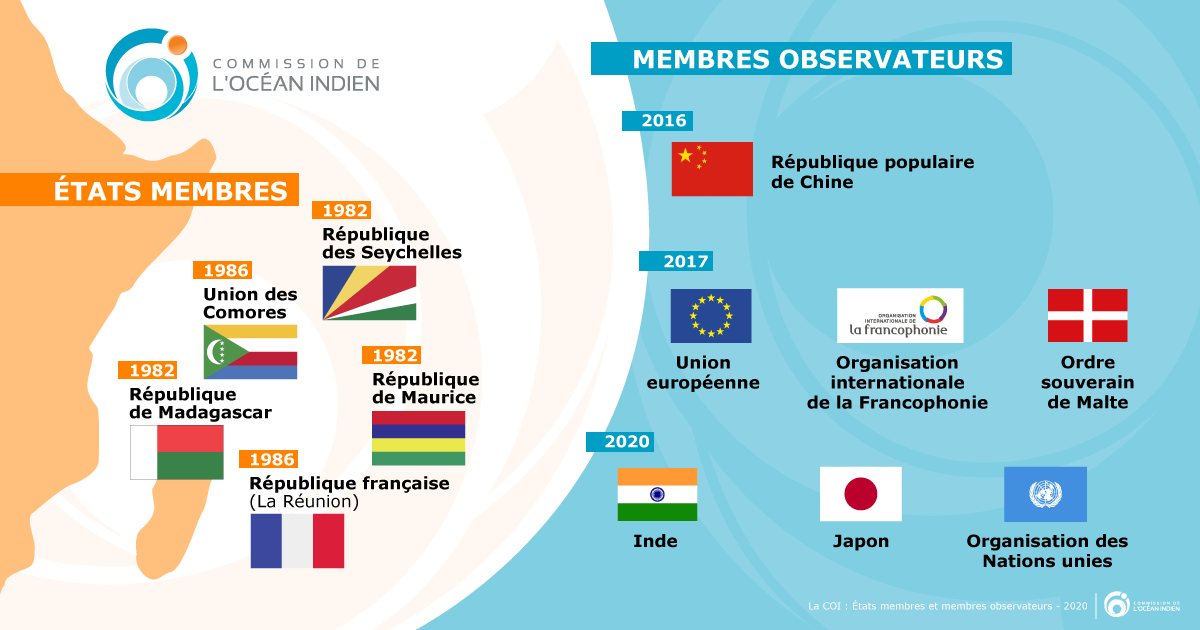
Ce projet, bien qu’embryonnaire, a attisé la curiosité de nombreux partenaires. La Commission de l’Océan Indien a notamment fait part de son intérêt pour un tel observatoire auprès du CIRAD (par la suite, le Mozambique a voulu rejoindre ce projet) . Pour information, la Commission de l’Océan Indien (COI) est une organisation intergouvernementale regroupant la France (à travers la Réunion), Madagascar, l’Union des Comores, Maurice et les Seychelles. Cette organisation défend les intérêts de ces Etats insulaires et cherche à favoriser la coopération entre ses membres. C’est donc assez logiquement que le projet a été remis entre les mains de la Direction Régionale Réunion Mayotte sous le nom d’Observatoire des Agricultures de l’Océan Indien.

Figure : La Commission de l'Océan Indien (COI)

Il est à noter que les pays de la COI partagent des caractéristiques fortes comme l’insularité ou le climat. Il est donc nécessaire de créer un observatoire à cette échelle régionale, adapté aux contraintes des pays membres. De plus, en réduisant le champ géographique de l’observatoire, on réduit également les différences entre les agricultures pratiquées. Des variables communes ont ainsi plus de sens. Pour illustrer ces propos, il serait insensé de demander à un agriculteur Américain combien il possède de charrettes tirées par des bœufs alors que cela est tout à fait adapté dans d’autres cas.

#### Projections

Les prochaines étapes du projet devraient donc se dérouler de la manière suivante.

Après la conception du prototype de la base de données en se basant sur le travail de M. Bosc, l’idée est de collecter un grand nombre de données de sources différentes. Que ce soient des enquêtes réalisées par des chercheurs ou des données collectées par des producteurs ou des organisations de producteurs, chaque donnée est importante. En effet, pour obtenir les données sur une exploitation qui ne les a pas consignées, plusieurs heures sont nécessaires à un technicien pour se rendre sur place, questionner le producteur et réaliser des mesures si besoin. La collecte peut être longue et donc coûteuse.

Cette première collecte est primordiale et ce, pour plusieurs raisons. D’abord, cela permet de confronter le modèle de la base de données aux données réelles pour l’affiner. Par ailleurs, cela permettra de créer les vues et d’alimenter l’interface Web pour la « Proof of concept ». Enfin, ces données seront complétées, selon les financements, par des enquêtes de terrain. En effet, étant donnés les orientations de l’observatoire, il est évident que le modèle établi ne collera pas exactement aux données déjà récoltées.

L’objectif est ensuite de créer une interface Web, dont les fonctionnalités sont à préciser dans un cahier des charges. Les vocations principales étant de pouvoir collecter des données et de les visualiser. A long terme, l’observatoire devrait aussi permettre de collecter et de visualiser des données à l’échelle de la filière. Ces données seraient intéressantes à exploiter d’un point de vue régional. C’est pour cela que l’interface devrait être intégrée au site de la Prérad-OI (Plateforme Régionale de Recherche Agronomique pour le Développement dans l’Océan Indien). Ce dernier, créé par le CIRAD, regroupera à terme d’autres outils comme un inventaire des projets agronomiques et des projets de développement dans la région ou encore des publications de recherche en agronomie.

Aussi, dans ce but de coopération ainsi que réutilisation des données existantes, toutes les données ne seront pas stockées directement sur la base de données de l’observatoire. Grâce à des API, les données non confidentielles de l’observatoire seront accessibles à tous et les visualisations de données de l’observatoire puiseront sur les données d’autres bases éventuelles.

## Département

Le stage s’est donc déroulé au sein de la Direction Régional Mayotte Réunion (DRRM) à la station « La Bretagne ». Les directions régionales ont pour rôle de coordonner les actions des différentes stations et unités de recherche tout en assurant la répartition du budget. Les missions ayant un impact plus globales sont aussi supervisées par ces directions régionales.

Néanmoins, mon encadrante Mme AUZOUX, fait partie de l’unité de recherche AIDA ("Agroécologie et intensification durable des cultures annuelles"). Cette unité se positionne sur l’intensification et la durabilité de la production des cultures annuelles en quantité et, lorsque pertinent, en qualité, en milieu tropical particulièrement contraint. Pour cela, ses recherches visent la pleine valorisation des ressources disponibles, en mobilisant les processus écologiques qui régissent leur dynamique au sein des agrosystèmes. Mme AUZOUX est spécialiste de la gestion des données et était donc parfaitement compétente pour superviser le stage.

## L’équipe

Ce projet d’envergure multidisciplinaire nécessite le travail, la collaboration, les connaissances et l’appui de plusieurs personnes :

* Mme Mialet-Serra est la coordinatrice du projet, basée à la station « La Bretagne ». On peut aussi la qualifier de responsable partenariat puisqu’elle s’occupe de trouver des partenaires et de gérer les interactions avec eux.
* Mme Darras est une volontaire service civique (VSC), également basée à « La Bretagne » qui travaille dans l’agronomie. Sa contribution principale est la réalisation d’un inventaire de toutes les variables pouvant être prises en compte dans l’élaboration des capitaux. Nous avons travaillé ensemble lors de l’élaboration de la base de données. Son travail s’est étendu jusqu’à la sélection de certaines variables de sortie.
* M. Bosc, basé aux locaux de la FAO à Rome, qui porte donc le projet de l’observatoire des agricultures du monde. Il a été mon principal interlocuteur pour tout type de question sur l’orientation de l’observatoire.
* Mme Auzoux, basée à « La Bretagne », ma tutrice, informaticienne spécialisée dans la conception de bases de données. Elle n’a pas contribué directement dans le projet mais a été de très bon conseil tout au long de ce stage. Du choix des technologies à la priorisation des tâches, son appui a été primordial.
* M. Bélières, basé à Montpellier, est un agronome ayant une grande connaissance de la collection des données de terrain. Son rôle a plutôt été celui d’un appui lors du choix des variables et de la conception de la base de données. Certains choix proviennent de son expertise du terrain.

Selon les périodes et les tâches courantes, les contacts étaient plus ou moins fréquents avec les différents membres de l’équipe. Néanmoins, la majorité du travail a été effectuée en autonomie.

variables de sortie

Typologie

PGD

# Dans les grandes lignes

## Missions

A l’origine, les missions à réaliser durant le stage étaient les suivantes :

* Explorer la base de données Dataverse du CIRAD. Un Dataverse est une application web permettant de préserver, partager, rechercher, analyser et favoriser la citation de jeux de données de recherches ;
* Réaliser un Benchmark des outils du CIRAD en lien avec le projet ;
* Elaborer le modèle conceptuel de données ;
* Créer la base de données sous PostgreSQL ;
* Rédiger le cahier des charges des spécifications techniques et fonctionnelles de la base de données, sur la base d’une analyse des besoins ;
* Développer les fonctionnalités décrites dans le cahier des charges sous un environnement web (développement d’une interface) ;
* Alimenter le prototype de base de données avec des données existantes au moins sur La Réunion et Maurice, élargies sur Madagascar si possible ;
* Tester l’interopérabilité avec des sources externes de données, sous réserves que celles-ci soient accessibles (bases de données FAO).

## Obstacles

Dès le premier jour, Madame Auzoux annonçait que le programme était trop ambitieux étant donnée l’avancement actuel du projet. En effet, bien que M. Bosc avait une idée relativement avancée de ce qu’il souhaitait, il était nécessaire de prendre en compte les besoins de futurs utilisateurs potentiels. Le problème réside dans le fait que, malgré l’intérêt de la COI, le projet n’avait pas réellement d’interlocuteurs que l’on pourrait qualifier de « clients ». Plusieurs partenaires du CIRAD, comme les réseaux SOA et RuralStruct, qui sont des organisations de producteurs à Madagascar, ont été contactés. Néanmoins, il n’a pas été possible de travailler directement avec eux sur leurs besoins. Par conséquent, la rédaction du cahier des charges n’a pas pu accomplie aussi finement qu’envisagé. Par la suite, la réalisation de l’interface a aussi été grandement freinée par ce manque de connaissance sur les utilisateurs.

Par ailleurs, il s’est avéré que les bases de données de la FAO ne contenaient pas de données qui avaient un intérêt pour l’observatoire. En effet, et c’est le cas pour de nombreuses bases de données auxquelles on a été confrontées, elles renseignent des informations à plus grande échelle. De plus, ces bases de données n’étaient pas publiques et malgré les contacts dont le projet disposait au sein de la FAO (avec M. Bosc notamment), il n’a pas été possible d’y accéder.

En outre, concernant la récolte de données pour alimenter le prototype de la base de données, le COVID-19 a été un gros frein. En effet, deux VSC devaient être engagés pour récupérer des données aux Seychelles et à Maurice mais leurs missions ont été annulées à cause de la pandémie. Par ailleurs, Mme Darras a été dépêchée sur d’autres tâches que la récolte de données qui faisaient partie de ses missions. Ainsi, les seules données à la disposition du projet étaient des données résultant d’un précédent travail de M. Bélières en lien avec les réseaux SOA et RuralStruct, qui dataient de 2015 et 2018.

Afin de compenser ces manques, une partie data visualisation a été ajoutée au stage. Bien que n’ayant pas toutes les données nécessaires, la création de data visualisations sera très utile.

Par ailleurs, certains changements comme les choix des technologies utilisés ou les changements de postes informatiques ont aussi causé certaines pertes de temps. Aussi, à plusieurs reprises, il s’est avéré que le stage était arrivé un peu trop tôt dans la chronologie du projet.

## Méthodologie

Une méthodologie dite AGILE, où les clients et les réalisateurs du projet se réunissent régulièrement aurait pu être très appréciable. En effet, le projet ayant beaucoup d’inconnues, cela permet d’ajuster précisément les besoins des clients et d’avoir une grande flexibilité. Néanmoins, elle n’était pas employable car ce projet ne disposait pas réellement de client.

Le choix a donc été de réaliser un cahier des charges et de s’y tenir.

## Outils utilisés

De multiples outils ont trouvé leur utilisé au cours du stage. Chacun d’eux à nécessiter une période d’apprentissage plus ou moins profonde comme cela est détaillé dans le section suivante.

### CodeIgniter

CodeIgniter est un Framework PHP. L’utilisation de Framework permet de faciliter et d’accélérer la production de site web. Ce Framework se distingue par une grande légèreté associée donc à une grande rapidité d’apprentissage. Néanmoins, avec son orientation objet, il regroupe toutes les fonctionnalités nécessaires à la programmation Web. Madame Auzoux étant habitué à utiliser ce Framework, ce choix était logique.

Il a été préféré à Symfony, qui est un autre Framework PHP car Symfony est bien plus long à apprendre et Mme Auzoux ne maîtrise pas le développement web avec ce Framework. Elle avait d’ailleurs eu plusieurs expériences négatives avec des stagiaires tentant d’apprendre Symfony.

### Leaflet

Leaflet est une bibliothèque JavaScript libre permettant de faire apparaître des cartes interactives sur des pages html. Elle est notamment utilisée par le projet de cartographie libre et ouverte OpenStreetMap.

### Git et GitHub

Git est un outil de gestion de version et GitHub est un service d’hébergement de projets git. C’est-à-dire qu’il est possible d’enregistrer sur GitHub les différentes versions d’un code au fur et à mesure du temps. Cela est très pratique pour retrouver une erreur qui s’est glissée dans le code. Git est également très utilisé pour développer des fonctionnalités indépendantes en même temps. Il est donc primordial afin de travailler en équipe. Les commandes init, push, pull, commit, add et checkout sont les plus basiques et celles qui ont le plus servi au cours du développement du prototype.

### D3.js

D3js est aussi une bibliothèque JavaScript. Elle permet de manipuler des éléments HTML et est donc complémentée par le langage CSS. Les SVG (Scalable Vector Graphics) qui sont disponibles à partir de D3js offrent la possibilité de créer des formes simples telles que des lignes, des rectangles ou des cercles. cf wikipédia . En combinant ces formes et en y intégrant les données, cette bibliothèque se révèle être un outil puissant pour représenter les données avec des visualisations originales et attrayantes. Les possibilités sont très nombreuses et le contrôle sur le résultat final est très précis.

Par ailleurs, le site d3js.org exposent des milliers d’exemples de data visualisations ainsi que le code qui permet de les créer. Cela n’empêche néanmoins pas de devoir comprendre le fonctionnement de d3js et de ses fonctions afin de créer ses propres mises en forme adaptées à ses données. Parmi les plus utilisées on retrouve :

* D3.select(« sélecteur ») et D3.selectAll(« sélecteur ») permettent de sélectionner des éléments HTML de la même manière qu’avec du code CSS. C’est-à-dire que l’on peut sélectionner un élément par son id en utilisant d3.select(« #id ») ;
* D3.append(« child ») crée un élément « child » à l’intérieur de l’élément HTML sélectionné ;
* D3.data(données) lie les données aux éléments du SVG ;
* D3.attr(« attribut », « valeur ») et en particulier des fonctions de la forme D3.attr(« transform », « translate(x, y) »). Ces dernières permettent de situer le SVG ;
* D3.scale() applique un zoom au SVG ;
* D3.transition() crée une transition à laquelle on va ajouter les modifications à effectuer ;
* D3.scaleLinear().domain([]).range([]) crée une échelle à partir du « domain » (qui est l’étendue des valeurs possibles en entrée) et d’un range (qui correspond au nombre de pixels sur lesquels cette valeur va être représentée) ;
* D3.on(« event », function(d, i) {}) lie des évènements, tels qu’un clique ou le passage de la souris sur un élément, à des fonctions ;
* D3.style(« attribut », « valeur ») est une alternative à l’écriture de code CSS.

Un autre grand avantage de d3js est que chaque fonction renvoie l’objet sur lequel la dernière fonction a été appliquée ou le dernier élément créé. Cela offre la possibilité de condenser le code en chaînant les fonctions les unes aux autres.

### Bootstrap

Bootstrap est un Framework CSS qui permet de faciliter et d’accélérer la présentation des pages d’un site web. Il est reconnu pour l'adaptation extrêmement simplifiée d’un site web en format téléphone (responsive design). Bootstrap est dit orienté smartphone. La grande force de ce Framework repose sur son système de grille de 12 colonnes pour placer les composants visuels et leur donner une taille. Ainsi, pour chaque composant visuel, on peut spécifier une taille d’écran et un nombre de colonnes occupées par l’objet.

Bootstrap classe les écrans selon 4 tailles :

* Xs qui correspond à une taille de smartphone ;
* S pour les tablettes ;
* M pour les petits écran d’ordinateurs ou les ordinateurs portables ;
* Lg pour les grand écran d’ordinateur.

Par ailleurs ce Framework dispose de nombreux éléments que l’on peut simplement copier depuis le site. On trouve par exemple des barres de navigation ou des menus déroulants déjà prêts à l’emploi. L’avantage, c’est que ces éléments ont déjà un style qui leur est appliqué ce qui contribue à son succès. Néanmoins, cela n’empêche pas de pouvoir éditer les styles appliquées dans un nouveau fichier CSS.

### SQL Power Architect

SQL Power Architect est un logiciel permettant de réaliser du Forward Engineering. Cela signifie que l’utilisateur crée graphiquement les tables et les relations de la future base de données. Le logiciel peut alors créer automatiquement le code SQL de la base ou directement implémenter la base de données grâce à des drivers permettant de faire des liens avec les SGBD.

### PostgreSQL et PgAdmin

Le SGBDR (système de gestion de base de données relationnelle) utilisé a été PostgreSQL afin de profiter du module POSTGIS qui peut être très utile dans le cadre de données géo référencées. Finalement, aucune donnée géo référencée n’a pu être intégrée au prototype mais le choix de ce SGBD pourrait s’avérer payant plus tard. L’interface utilisé était PgAdmin. Cette dernière permet de créer les tables et des vues sans avoir à écrire tout le code SQL. En outre, et à la différence de Power Architect, PgAdmin donne la possibilité d’ajouter des contraintes aux tables.

Bien que le modèle de base de données a été créé sur Power Architect, de nombreuses modifications ont été réalisées à partir de PgAdmin.

### Microsoft Office Access

Access permet de gérer des bases de données relationnelles. Ce logiciel n’a pas été utilisé pour la conception de la base de données car ce n’est pas le plus adapté pour cet usage. Néanmoins, pour intégrer des données, il est extrêmement pratique et c’est donc dans cet objectif qu’il a été utilisé.

Access peut importer des bases de données venant de SGBD grâce à des drivers. Ces derniers permettent également d’appliquer directement les changements effectués sur les données depuis Access sur la base de données PostgreSQL. Néanmoins, la structure de la base de données est préservée quoiqu’il arrive.

Il dispose notamment de fonctions qui permettent de créer des colonnes avec des valeurs prédéfinies. D’autre part, la fonction « recherche et remplace » a aussi été d’une grande aide.

### Gantt Project

Gantt Project est un outil de création de diagramme de Gantt. Ces diagrammes sont très utilisés dans le cadre de la gestion de projet afin d’estimer le temps nécessaire à la livraison d’un projet. Cela permet aussi de mettre en évidence le chemin critique.

## Planning

Le stage s’est déroulé en plusieurs phases qui sont regroupées dans le diagramme de Gantt ci-après. Celui-ci a été réalisé avec le logiciel Gantt Project. Ce planning montre l’intérêt du stage qui couvre l’ensemble du cycle de vie des données.

Les tâches ont été réparties en trois catégories :

* En vert, apparaissent les tâches qui ne nécessitent pas de compétences techniques mais demandent plutôt des qualités de communication ;
* En marron, sont représentées les parties ayant trait à la base de données ;
* Enfin, les tâches en bleu, sont celles concernant la programmation Web.

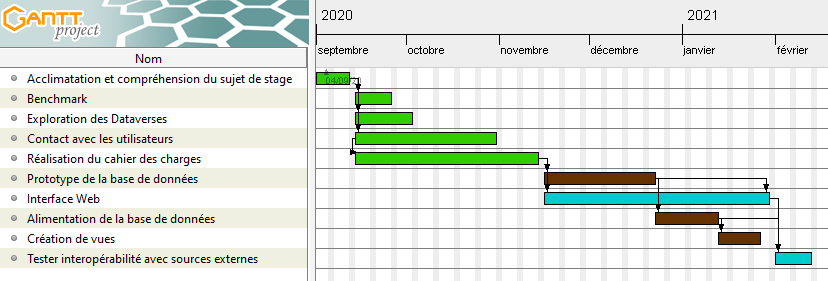


Figure : Diagramme de Gantt du stage à l'origine

Ce qui apparaît avec ce diagramme, c’est la diversité des tâches et la continuité qui s’opère. C’est un des vrais points positifs du stage. Par ailleurs, le livrable indispensable du stage étaient la base de données.

Finalement, et pour les raisons évoqués dans la partie « obstacles », le stage a plutôt suivi l’agenda suivant : Mettre le diagramme

On note notamment l’apparition de la partie « data visualisation » et la suppression de la partie « interopérabilité avec les sources externes. »

# Déroulement du projet

## Compréhension des intentions

Dans un premier temps, le stage s’est concentré sur la compréhension des intentions et des objectifs voulus par le CIRAD et par M. BOSC. Il est apparu à de nombreuses reprises que les différentes parties prenantes du projet n’avaient pas exactement la même manière de voir ce projet. Il a donc été nécessaire de se concerter pour concentrer nos efforts dans la même direction.

Les objectifs se résument à la réalisation d’un outil informatique permettant :

* Au grand public d’accéder aux données sur les exploitations agricoles et en particulier les agricultures familiales,
* Aux techniciens de saisir des données sur le terrain concernant les exploitations étudiées,
* Aux décideurs politiques d’élaborer des politiques différenciées d’investissement, censées être plus efficaces car plus appropriées,
* Aux producteurs d’avoir une évaluation de leur exploitation agricole selon les 5 capitaux (financier, humain, social, naturel et physique),
* Aux chercheurs de publier des études basées sur des données extraites de l’outil, et pouvant contenir des « paroles d’agriculteurs ».
* Et éventuellement à tout chef de projet de stocker ses données et d’avoir un aperçu de l’impact du projet.

## Exploration des Dataverses

Une fois les objectifs assimilés, l’objectif a été de rechercher dans les Dataverses du CIRAD, de l’IRD et de l’INRA (un Dataverse est une application web à code source ouvert permettant de préserver, partager, citer, rechercher et analyser des données de recherche), des éventuelles données qui pourraient être utiles dans le cadre de l’observatoire. Malgré les quantités importantes de données de recherche, il s’est avéré qu’aucune base de données ou fichier Excel ne contenait des informations utilisables. Cela représentait un point négatif pour l’observatoire qui n’allait pas disposer de beaucoup de données au démarrage. Néanmoins, cela signifiait également que l’observatoire allait pouvoir subvenir, par la suite, à d’éventuels besoins en données.

## Benchmark

Après avoir saisi et incorporé les différentes intentions derrière l’OA-OI, un benchmark a été réalisé sur les différents outils du CIRAD. Habituellement un Benchmark est utilisé pour comprendre les points forts et les points faibles de concurrents sur un marché. Ici, la situation est un peu différente. Il s’agit d’étudier les outils mis en place par le CIRAD et de savoir s’ils pourraient servir dans le cadre de l’observatoire. Leur éventuel succès ou échec est aussi pris en compte.

Tous les outils étudiés ici n’avaient donc pas vocation à être la base du futur observatoire. Néanmoins, il est intéressant de noter les points forts et les points faibles de chaque outil en lien avec nos besoins. L’inventaire présenté ici n’est qu’une ébauche ; un inventaire plus détaillé est fourni en annexe. Ce qu’il faut retenir, c’est qu’aucun outil, à l’heure actuelle, ne répond aux besoins formulés par l’observatoire. En effet, la diversité des échelles voulues, des acteurs impliqués et des objectifs affichés fait de l’observatoire un cas à part.

Par ailleurs, l’hypothèse a été émise de pouvoir réutiliser certains de ces outils en post-traitement si cela apportait une réelle plus-value. Par exemple, si on se rend compte qu’une exploitation agricole a des efforts à faire concernant l’utilisation de produits phytosanitaires, on l’orientera vers Phyto’aide. Ce dernier est un outil qui permet de cerner les différents leviers qu’un agriculteur a, à sa disposition, pour limiter l’impact environnemental des produits qu’ils utilisent.

|  |  |  |
| --- | --- | --- |
| Outils | Avantages | Inconvénients |
| AGREF | -Plateforme  -Saisie données  -Export données  -Application mobile  -Différents profils utilisateurs | -Design  -Sécurité  -Non opérationnel  -Nombreuses variables inutilisées  -Pas de données agrégées  -Pas d’indicateurs  -Pas de gestion filière |
| AEGIS | -Design  -Analyse  -Collecte des données et côté ludique  -Echelle parcelle | -Pas échelle exploitation  -En cours de développement |
| E-Watch | -Requête/cartographie  -Données par projet  -Développé pour tablette  -Statistiques par zone/projet | -En refonte  -Peu adapté à divers profils |
| IDEA RUN | -Auto-évaluation rapide  -Co-créée avec acteurs locaux | -Adaptée uniquement Réunion  -Juste une méthode  -Trop de données supplémentaires |
| FAST | -Critères de recherche | -Pas de classement par ordre préférentiel |
| LASER | -Package sous R pour effectuer les calculs poussés |  |
| Phyto’aide | -Différents leviers d’action explicités | -Modèle très simple |
| HiH | -Evolution données avec « vidéo » sur la carte  -Curseur droite/gauche pour comparaison  -Carte et graphique liés |  |

Figure : Benchmark

## Recueil des besoins des utilisateurs

#### Questionnaire

Cette partie du travail a été assez particulière car les acteurs ne sont pas réellement des clients. En effet, les acteurs sont les producteurs, les agences de développement, les ONG, les chercheurs ou encore les décideurs politiques. Ces derniers n’ont pas demandé directement l’observatoire. Il était donc nécessaire de faire des compromis entre les besoins réels des futurs utilisateurs et les intentions des porteurs de projet.

Un questionnaire a donc été réalisé sur Google Form (qui facilite grandement l’exploitation des résultats car ces derniers sont visualisables sur un tableur Google Sheet) afin d’identifier les partenaires possibles, les données dont ils disposent, et d’avoir un aperçu général de leurs besoins potentiels. Ce questionnaire devait ensuite être suivi par des interviews pour spécifier les attentes de chacun. Le cahier des charges devait ensuite être élaboré à partir de ces informations. Ce questionnaire a donc été soumis aux contacts du CIRAD (des agences de développement et des organisations de producteurs principalement). Malheureusement, seuls 4 organismes contactés sur 15 ont répondu.

Conscient que ce travail était d’une importance capitale pour le bon déroulement et la pérennité du projet, le stagiaire a voulu relancer les partenaires qui n’avaient pas répondu et approfondir les contacts avec les répondants, comme cela été prévu. Cependant, malgré de nombreuses tentatives sollicitations, les prochains contacts ont dû attendre le début d’année 2021.

#### DAAF

Durant le stage, un rendez-vous a été pris avec la DAAF de la Réunion (Direction de l'Alimentation, de l'Agriculture et de la Forêt) qui dispose d’un département statistique et donc de données sur les exploitations agricoles de la Réunion qui pouvaient s’avérer intéressantes. En effet, la DAAF récupère les données des recensements agricoles qui sont des études approfondies des exploitations (hors revenus) effectuées tous les 10 ans. De plus, elle dispose des résultats des études RICA qui sont réalisées annuellement et se concentre sur les résultats économiques des exploitations.

Le but du département de statistiques de la DAAF est de réaliser des études statistiques afin de publier des études sur les exploitations agricoles. Ils réalisent aussi des études sur des demandes précises.

Malgré les demandes, la DAAF n’était pas autorisée à nous transmettre des données car celles-ci étaient confidentielles même après anonymisation. Par ailleurs, le directeur du département statistique, M. Létoublon, a signifié que le CIRAD avait déjà reçu des données de leur part, sans recevoir la contrepartie convenue.

Développer RGPD

## Recueil des données

Dans le questionnaire cité plus haut, les participants devaient également renseigner les coordonnées des personnes à contacter pour récupérer les données qu’ils collectent. En plus de cela, ils ont pu exprimer le niveau de partage qu’ils souhaitaient pour leurs données. De cette manière, les données d’un projet du réseau SOA ont été récupérées. Néanmoins, elles concernaient directement le projet (qui consistait à aider des jeunes à se lancer dans l’agriculture) et pas les exploitations agricoles dans leur ensemble.

Néanmoins, M. Bélières, un membre du CIRAD ayant travaillé à Madagascar sur les exploitations agricoles a pu fournir deux bases de données au moins partiellement exploitables, à l’échelle de l’exploitation. Elles contenaient des données intéressantes sur un total de plus de 800 exploitations. Ces données n’étaient pas présentes dans le Dataverse du CIRAD. Cela démontre la nécessité, pour les chercheurs, de partager leurs travaux et leurs données sur des plateformes communes.

## Réalisation du cahier des charges

Afin de cadrer la suite du travail et notamment l’interface web, la réalisation d’un cahier des charges est nécessaire. Ce dernier implique, à la différence d’une méthodologie AGILE, d’avoir une vision d’ensemble dès le démarrage du projet. La rédaction de ce livrable a été le fruit d’une collaboration avec M. Bosc, bien qu’il ne soit pas l’utilisateur final du produit.

### Profils

Comme cela a pu être expliqué précédemment, l’observatoire a, comme ambition, de servir des acteurs divers. Cela se traduit, dans l’interface web, par la différenciation de plusieurs types de profils. Il est aussi à noter que l’interface web devant être accessible à tous, il est possible d’y accéder sans créer de compte. Néanmoins, les utilisateurs plus impliqués auront des profils avec des spécificités :

Le contrôle du champ d’action d’un utilisateur est cadré par son profil. Un utilisateur peut disposer de plusieurs profils. En effet, il serait contre-productif d’interdire à un producteur de créer des projets.

Les objectifs se résument à la réalisation d’un outil informatique permettant :

* Au grand public d’accéder aux données sur les exploitations agricoles et en particulier les agricultures familiales,
* Aux techniciens de saisir des données sur le terrain concernant les exploitations étudiées,
* Aux décideurs politiques d’élaborer des politiques différenciées d’investissement, censées être plus efficaces car plus appropriées,
* Aux producteurs d’avoir une évaluation de leur exploitation agricole selon les 5 capitaux (financier, humain, social, naturel et physique),
* Aux chercheurs de publier des études basées sur des données extraites de l’outil, et pouvant contenir des « paroles d’agriculteurs ».
* Et éventuellement à tout chef de projet de stocker ses données et d’avoir un aperçu de l’impact du projet.

### Fonctionnalités

M. Bosc avait auparavant regroupé les différentes fonctionnalités qui pouvaient être intéressantes à implémenter dans le cadre de l’observatoire. Parmi ces dernières se trouvaient, de manière évidente, la collecte et la visualisation de données mais pas seulement. Voici un résumé des fonctions envisagées et qui sont décrites dans le cahier des charges :

* Exportation des données pour offrir la possibilité à des chercheurs de réaliser des études ;
* Publication de ces dites études ;
* Publications de bilan réguliers sur l’état des agricultures familiales ;
* Création, suivi et évaluation de projets par des agences de développement ;
* Recherche d’utilisateur ou de projet via un annuaire ;
* Publication d’études de cas pouvant contenir des vidéos dites « paroles d’agriculteurs », dans le but de faire valoir les pratiques considérées comme durables.

### Charte graphique

Concernant la partie charte graphique, le but est de reprendre les principaux éléments de la plateforme Prérad-OI (Plateforme Régionale de recherche Agronomique pour le Développement dans l’Océan Indien) depuis laquelle l’observatoire sera accessible. Le code couleur ainsi que le bas de page seront les mêmes.

Par ailleurs, le choix concernant le logo n’a pas été tranché. Ci-contre se trouve une proposition d’un logo pour l’observatoire. Il est également possible de reprendre celui de la Prérad-OI (ci-dessous)

Figure : Logo de la Prérad-OI

Figure : Proposition de logo OA-OI

### Droits des utilisateurs

Les utilisateurs du site pouvant être des utilisateurs avec un compte, il est nécessaire de prévoir quel utilisateur possède les droits sur quelles données. Ces droits sont renseignés dans des matrices des droits où l’on retrouve des informations assez classiques. Voici ci-dessous l’exemple des droits attribués pour les informations d’une exploitation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Administrateur/trice ou technicien(ne) associé(e) à l’exploitation | Exploitant(e) agricole correspondant | Utilisateur/trice enregistré(e) ou non |
| Ajouter une exploitation | X |  |  |
| Supprimer une exploitation | X |  |  |
| Ajouter des données sur une exploitation | X |  |  |
| Supprimer des données sur une exploitation | X |  |  |
| Consulter les infos publiques de l’exploitation | X | X | x |
| Consulter les infos privées d’une exploitation | x | X(la sienne) |  |

Figure : Matrice des droits sur une exploitation

Ce livrable attend des informations complémentaires de la part des utilisateurs pour être finalisé. De nouveaux contacts sont prévus pour début 2021. Ainsi, le cahier des charges contient des propositions qui restent à affiner concernant :

* L’architecture du site ;
* La composition des tableaux de bord pour les exploitations mais aussi pour les agences de développement ou les gestionnaires de projet ;
* Les spécifications techniques ;
* Les améliorations à envisager.

## Conception de la base de données

### Général

La conception de la base de données avait un double objectif. En effet, elle devait servir, dans un premier temps, pour l’OA-OI. Dans un deuxième temps, elle devait être réutilisée en tant que « base modèle » pour les futures études concernant les exploitations agricoles réalisées par le CIRAD.

La conception de la base de données a été réalisée suite au travail de Madame Darras. Cette dernière ayant élaboré, en contact avec Messieurs Bosc et Bélières, un inventaire détaillé des variables devant être collectées et stockées dans la base. Son travail a grandement facilité la conception de cette dernière.

Une notion à absolument prendre en compte durant cette conception est la notion de temps. Presque toutes les entités présentes ont des variables qui en dépendent. Afin de concevoir une base de données en 3NF (3ème forme normale), il a donc été nécessaire de séparer certaines entités en deux selon si elles dépendaient du temps ou non. Certains choix comme le sexe ou le nom de famille sont discutables et ne sont pas fixés dans le marbre.

Pour rappel, la 3e forme normale est une manière de structurée la base de données qui assure un faible niveau de redondance (bien que non nul), tout en conservant toutes les informations et les dépendances fonctionnelles. Cette dernière pouvant mener à l’invalidité de données dans le cas de modifications. Développer 3NF et choix

Finalement, la base de données contient 39 tables dont 31 directement pour la partie collecte des données des exploitations et 8 pour la partie utilisateur de l’interface Web qui reste à modifier. Ce sont ainsi quelques 200 variables qui doivent être renseignées lors d’une collecte. Ces 200 incluent des variables qui ne varieront vraisemblablement pas d’une année sur l’autre (comme les infrastructures ou les terres possédées). Pour chaque table, il sera possible, lors de la collecte de données, de signifier qu’aucun changement n’a été opéré. Cela permettra de remplir les champs de la même manière que la collecte précédente en changeant simplement l’année.

### Contraintes

Les contraintes liées à des relations d’héritage ont été appliquées à la base de données contrairement aux contraintes concernant les listes de valeurs possibles. En effet, ces dernières étant encore en évolution, elles ont simplement été spécifiées dans un document à part.

D’autres contraintes concernant des compositions ont aussi été ajoutées (par exemple, la superficie d’une parcelle est inférieure à celle de la terre sur laquelle elle se trouve).

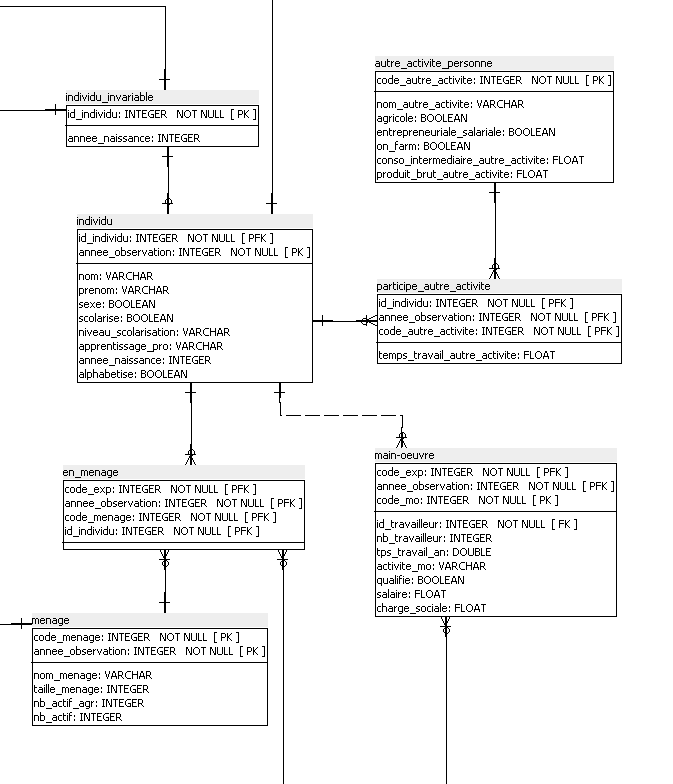


Figure : Extrait du diagramme Power Architect

### Problèmes rencontrés

Afin que la collecte de données sur le terrain soit réalisable dans un temps raisonnable, il a été nécessaire de réduire cette liste de variables. Cela a occasionné de nombreux changements, supprimant parfois des tables entières. Ainsi, cette partie a pris un temps considérable. Plusieurs versions de la base de données ont été conservées selon les quantités d’informations nécessaires ou présentes.

Un autre aspect qui a été nécessité une certaine réflexion est celui des unités à utiliser pour les quantités. Que ce soit pour les consommations intermédiaires ou les quantités issues de la production, il faut pouvoir mesurer les quantités pour les comparer. Il a donc été décidé d’ajouter une variable unité et une variable « équivalent\_kg ». De plus, une table « conversion », regroupant les conversions des masses en kg, a été créée. Grâce à l’utilisation d’un trigger, lors de l’insertion d’une donnée, la colonne « équivalent\_kg » est remplie automatiquement. Les unités utilisées sont d’ailleurs renseignées dans un document « dictionnaire » généré automatiquement par SQL Power Architect à partir du schéma mais aussi des commentaires.

DEVISE

Le problème est que les agriculteurs mesurent parfois leurs quantités en charrettes, en sceaux ou en flacons et ne connaissent pas nécessairement la masse que cela représente. Pour ces cas-là, il sera nécessaire d’estimer le volume de ces unités afin de trouver la masse grâce aux masses volumiques des produits. Il s’agira bien évidemment d’une estimation.

Par ailleurs, les données potentiellement récoltées provenant de diverses sources, il fallait faire preuve de flexibilité. Par exemple, certaines exploitations ou certaines enquêtes enregistrent l’utilisation de consommation intermédiaires à l’échelle de l’exploitation alors que d’autres le font à l’échelle de la parcelle. Certaines informations paraissent ainsi redondantes alors qu’il s’agit d’une réelle volonté de pouvoir récupérer des données plus précises si elles sont disponibles et plus générales dans le cas contraire.

D’autre part, étant donnés les pays dans lesquels l’observatoire a vocation à s’implanter, certaines situations particulières doivent être envisagées. En effet, en France, un ménage peut posséder et/ou vivre sur une ou plusieurs exploitations agricoles. Dans les pays de la COI (Commission de l’Océan Indien), il est possible d’avoir plusieurs ménages sur une seule exploitation. Ce type de situations compliquent la base de données. Les deux entités « ménage » et « exploitation » étant séparées, il est parfois difficile de savoir si une variable doit être liée à l’une ou à l’autre de ces entités.

Enfin, les mots employés par les Messieurs Bélières et Bosc n’étaient pas toujours connus de tous. Le monde agricole possède un vocabulaire spécifique et c’est dans ce type de cas qu’une bonne communication facilite le travail.

## Intégration de 2 bases de données

### Généralités

Après la conception de la base de données, le but était de récupérer les données disponibles et de les intégrer à notre base. Cela permettait plusieurs choses :

* De confronter notre base de données « théorique » à des données collectées sur le terrain. De cette manière, de multiples modifications ont été effectuées, notamment sur l’aspect flexibilité évoqué plus haut. Par ailleurs, certaines variables ont été ajoutées et d’autres retirées, en discussion avec Messieurs Bosc et Bélières.
* De « remplir » la base de données afin de réaliser les vues

### Problèmes rencontrés

Cette partie du stage, bien qu’essentielle est extrêmement fastidieuse. En effet, les bases de données récupérées ne possèdent ni les mêmes tables, ni les mêmes clés primaires, ni les mêmes noms de variables (qui sont parfois difficiles à comprendre lorsqu’aucune documentation n’existe ou que les mots utilisés sont en malgache). Parfois, il était même nécessaire de remonter aux questionnaires utilisées lors de la collecte de données pour comprendre ce que les noms des variables impliquaient.

Cela a d’ailleurs permis de souligner l’importance de la clarté des questionnaires d’enquête. Un questionnaire demandait d’abord les informations du chef d’exploitation, puis, des membres du ménage. Certains enregistrements faisaient ainsi apparaître le chef d’exploitation dans les membres de ménage et pas d’autres. Afin de ne pas corrompre les données, la décision a été prise de ne pas comptabiliser les chefs d’exploitation dans les membres du ménage pour de ne pas avoir de doublons.

Les bases de données témoignent parfois d’un manque de connaissance en gestion des données. En effet, sur les deux bases de données, l’une était constituée d’une seule table et des dizaines de colonnes alors que l’autre ne respectait pas certaines contraintes de clés étrangères. Il est par exemple arrivé d’avoir des engrais utilisés sur un champ qui n’avait jamais été enregistré. Avant d’insérer les données, il a donc été nécessaire de réaliser un long nettoyage. De nombreuses données ont ainsi été perdues.

Pour illustrer certaines de ces difficultés et le travail de nettoyage qu’il a été nécessaire de réaliser, il est intéressant de prendre un exemple. Voici ci-dessous un extrait d’une table récupérée dans une des bases de données intitulée « A06a\_Mat\_Manuel » pour matériel manuel. La clé primaire est le « Num\_Enreg ».

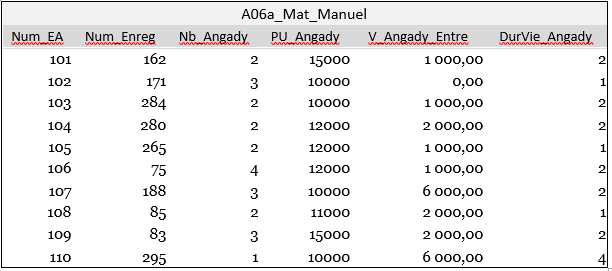


Figure : Table matériel manuel d'une des bases de données récupérées

On trouve ici plusieurs problèmes. Avant tout, il faut traduire « angady » du malgache qui signifie « bêche ». Ensuite, les enregistrements ne sont pas datés alors qu’ils devraient l’être. En effet, une exploitation ne possède pas forcément les mêmes équipements en 2010 et en 2020. Certains ont pu être cassés, perdus, volés ou d’autres ont pu être achetés. Si une nouvelle enquête sonde ces exploitations, il y aura un problème de gestion de la temporalité. Une colonne date est donc ajoutée en récupérant la date de sondage de l’exploitation présente sur le questionnaire.

Par ailleurs, il faut comprendre que PU correspond à prix unitaire. Aussi, il n’est pas possible en l’état de comprendre ce que V\_Angady\_Entre. On se réfère donc au questionnaire lié à la base de données s’il existe (seul le questionnaire de l’une des 2 bases de données récupérées était disponible).

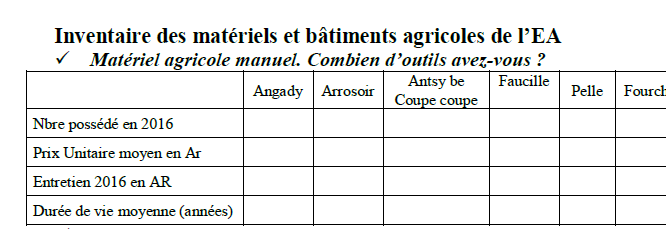


Figure : Extrait du questionnaire d'une des bases de données récupérées

Cela signifie donc la valeur de l’entretien, on suppose que ce coût est un coût total pour l’ensemble des bêches. Cette information est à stocker dans la table consommations intermédiaires de la base de données prototype.

(En outre, et ce n’est pas visible sur cet extrait de base de données, cette dernière stockait les informations pour chacun des 13 types de matériel agricole manuel listés dans le questionnaire. Le problème est que de nombreux champs restaient ainsi vides + clé locale utilisée)

Enfin, certains enregistrements étaient stockés sous forme de codification comme sur l’exemple ci-dessous.

Travail sur les unités standardisées

Galère ménage exploitations

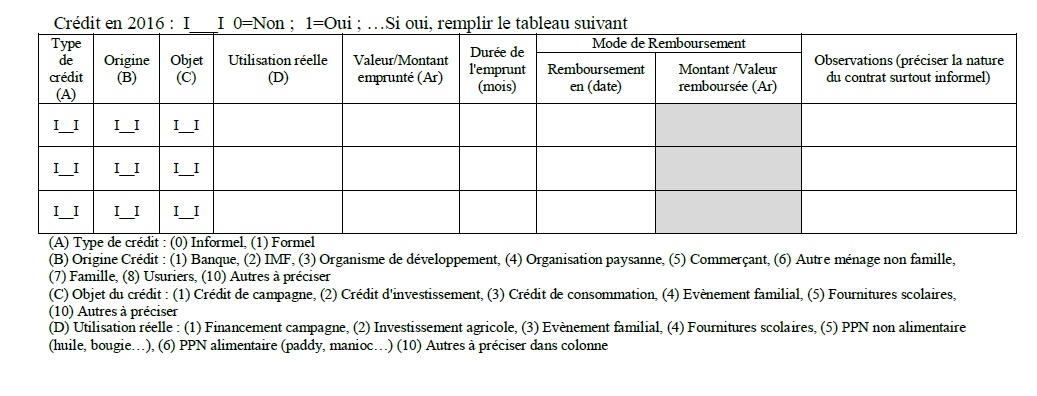


Figure : Extrait d'un questionnaire avec codification

Dans ce type de cas, il a fallu remplacé chaque code par sa signification pour ensuite le rentrer dans la base de données prototype. Les correspondances étaient renseignées dans des tables intermédiaires. La possibilité de garder ces tables intermédiaires et ces codes n’a pas été retenue car les listes de valeurs possibles. Elles avaient été principalement utilisées pour accélérer le remplissage des bases de données qui étaient des fichiers Access. Avec une interface web qui dispose de valeurs prédéfinies, cela n’est plus nécessaire et entraîne uniquement une plus grande complexité non justifiée.

Toutes ces opérations ont dû être réalisées pour environ 70 tables pour la première base de données qui était la plus fournie.

## Vues

Une fois la base de données alimentée, il a été possible de réaliser des vues pour faire émerger des informations essentielles. Ces variables de sortie ont été travaillées par Mme Darras en collaboration avec M. Bosc et moi-même. Elles concernent entre autres :

* L’exploitation (capital, surfaces totale, surface cultivée) ;
* Les performances économiques (valeur ajoutée brute et nette, valeur ajoutée par unité de capital, par unité de travail annuel, par unité de surface) ;
* Les performances environnementales (valeur ajoutée par utilisation de carburant) ;
* La composition de la main d’œuvre (travail familial, des femmes et des enfants notamment) ;
* La composition des revenus du ménage ;
* La composition des dépenses du ménage ;
* Le ménage (taille du ménage, ratio nombre d’actifs par inactifs) ;
* L’utilisation de produit phytosanitaire et d’engrais (part des surfaces agricoles recevant ces traitements) ;
* La production (part de la production vendue, part de la production qui est bio) ;
* Les pratiques culturales (nombre de parcelles avec des associations de cultures, nombre de cultures différentes).

Voici ci-dessous, l’exemple de la vue « vab » (valeur ajoutée brute). Cette dernière renseigne le numéro de l’exploitation et l’année pour laquelle cette exploitation a été observée. La valeur ajoutée brute (vab) est obtenue en soustrayant le produit brut(pb) par le coût des consommations intermédiaires (cout\_total\_conso\_inter). Ces deux dernières ayant été obtenue par des agrégations préalables.

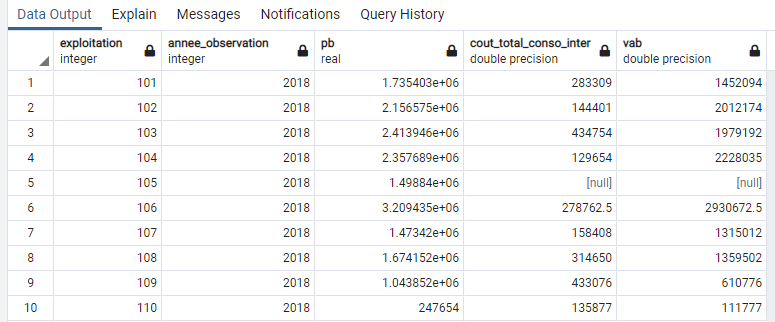


Figure : Vue valeur ajoutée brute

On aurait ici pu utiliser la fonction SQL COALESCE(liste). Cette dernière permet de renvoyer la première valeur non nulle de la liste qui lui est transmise. On aurait ainsi pu avoir :



Cela aurait permis d’avoir une valeur pour la ligne 5. Néanmoins, donner une valeur ajoutée brute pour une exploitation dont on ne connaît pas le coût des consommations intermédiaires n’aurait pas eue trop de sens. En effet, si le cout des consommations intermédiaires n’est pas renseigné, il n’est certainement pas null mais n’a juste pas pu être collecté. Garder une vab à null est donc préférable. Ce type d’opération a conditionné un certain nombre de vues ainsi que certaines jointures. En effet, lors de certaines jointures, une des deux tables pouvait contenir une information jugée bien plus importante que l’autre. Les jointures « right join » ou « left join » ont alors été appliquées pour éliminer les enregistrements ne disposant pas de réelle valeur ajoutée.

Par ailleurs, certaines vues, notamment celles concernant le travail des individus, montrent la répartition selon l’âge et le sexe. Pour illustrer le manque de connaissance que nous avons, les individus aux âges ou au sexe non renseignées sont compter en tant que « non renseigné ». Cela montre le chemin qu’il reste à parcourir pour compléter la base de données en plus de nous raisonner quant à des conclusions hâtives sur les données.

Utilisation de sous requête de partition by et over et avg bien utile car valeur null

## Réalisation du prototype

### Base de l’application Web

Développer création application

L’application Web est basée sur une architecture MVC (Modèle-Vue-Contrôleur). Il s’agit d’un design pattern permettant de séparer le code pour une plus grande lisibilité. Pour chaque page html, il y a donc au minimum un fichier pour la vue, un fichier pour le modèle et un fichier pour le contrôleur. Chaque partie a donc une fonction bien précise.

Le Modèle :

* Le Modèle regroupe toutes les fonctions liées aux interactions directes à la base de données.
* L’utilisation de CodeIgniter, comme de beaucoup de Framework, facilite sensiblement ces interactions. Dans CodeIgniter, le modèle est en effet un objet qui hérite d’une classe AbstractController présente dans le Framework dès la création du projet. Cette classe implémente déjà de nombreuses fonctions (comme ->find (), ->where (), ->, …) qui rendent la création de requêtes SQL bien plus rapide.

La Vue :

* La vue est la partie du code qui produit le document html.
* Dans CodeIgniter, il est possible de passer directement des variables du contrôleur à la vue au moment de l’appel à la fonction render (« view\_name », $data). Chaque variable passée dans l’ensemble « $data » peut-être utilisée entre des balises PHP dans la vue appelée.

Le Contrôleur :

* Le contrôleur est quant à lui chargé de gérer les requêtes http qui lui sont envoyées. Par la suite, il appelle les modèles et les bonnes vues en fonction des informations dont il dispose.
* Dans CodeIgniter, c’est dans ce fichier que l’on instancie les modèles et que l’on renvoie les vues (grâce à la fonction render (« view\_name »).

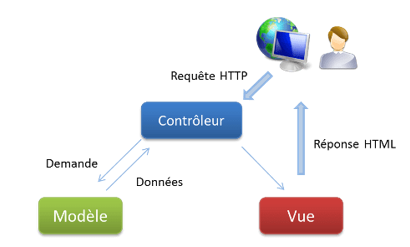


Figure : Schéma architecture Modèle\_Vue\_Controleur

### Avancement de l’interface

Malgré les manques concernant les besoins des utilisateurs, l’interface a pu être commencée. L’avancement est décrit ci-dessous.

Le Template général composé de la barre de navigation ainsi que le pied de page est implémenté. Le site contient actuellement diverses pages qui ont vocation à présenter l’observatoire, et visualiser des données sur une exploitation ou sur une région donnée

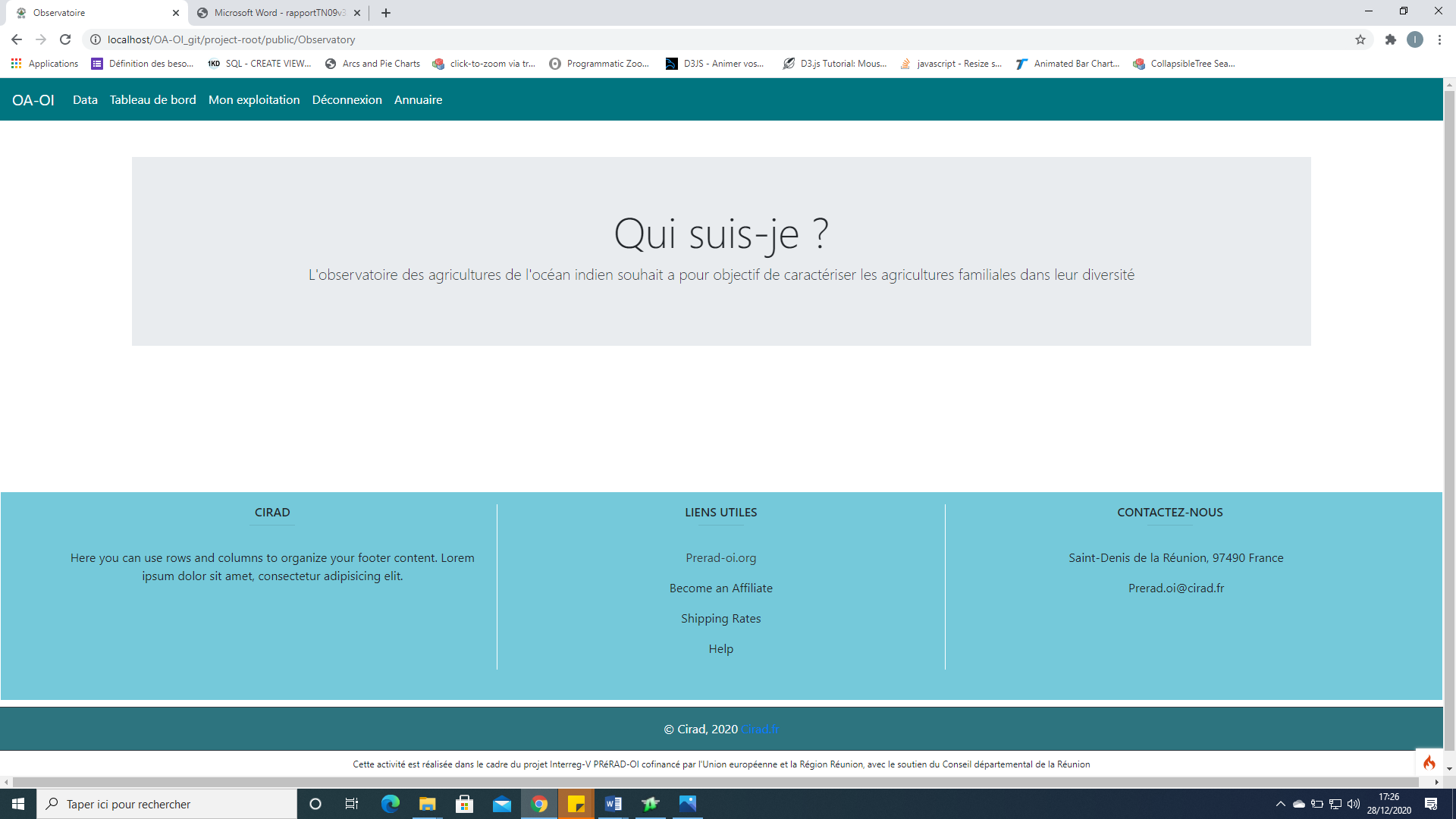


Figure : Page de présentation de l'observatoire

La possibilité de s’enregistrer, de se connecter et de déconnecter est également gérée. Cet enregistrement inclut un ou plusieurs types de profil. Cela permet, par la suite, d’avoir une barre de navigation différente selon ce profil. Un producteur aura par exemple une page « Mon exploitation » que n’aura pas une agence de développement.

Une fonction d’annuaire a également été implémentée. Elle permet de rechercher les utilisateurs selon leur type de profil.



Figure : Recherche dans l'annuaire

### Data visualisations

La production de data visualisations permet de comprendre plus facilement les chiffres. Elle est donc indispensable pour un outil qui se veut orienté vers l’information et la décision. Grâce aux vues précédemment réalisées, il a été possible de récupérer des données agrégées. Néanmoins, de nombreuses données étant manquantes, une exploitation fictive a été créée en remplissant les champs les plus significatifs.

La data visualisation ci-contre est basée sur la vue « Composition\_revenus » et donne des renseignements sur les revenus de l’exploitation fictive. La partie de gauche est le visuel à l’origine. Lorsque l’on clique dessus, un zoom est effectué et les légendes apparaissent (partie droite).

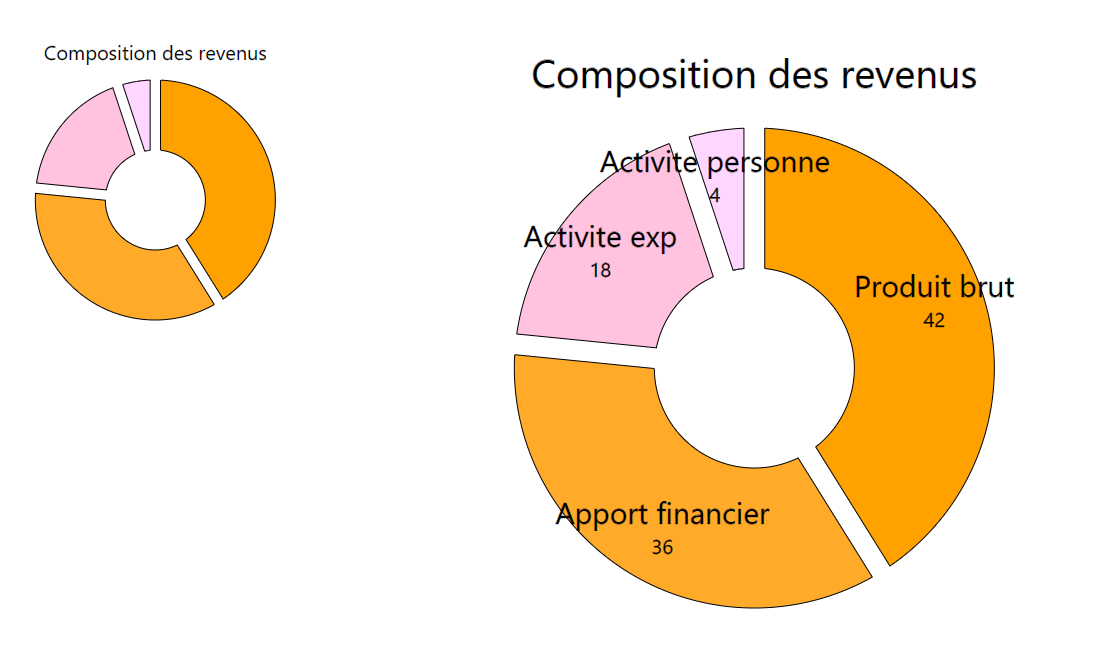


Figure : Data visualisation sur la composition des revenus

La data visualisation suivante est une carte permettant d’accéder aux informations sur les différents pays de la Commission de l’Océan Indien et du Mozambique. En cliquant sur ces derniers, un zoom est effectué. Bien sûr, plus un pays est grand, moins le zoom est important. Pour le grand public, cette carte peut être intéressante pour connaître les tendances territoriales. Ici, il s’agit d’une carte chloroplète puisqu’elle est colorée en fonction d’une échelle. D’ailleurs, il est possible de changer le critère choisi pour la coloration en bas à gauche de la carte. Un tooltip est également ajouté (visible au niveau du Mozambique sur la capture d’écran ci-dessous). Ce dernier apparaît lorsque la souris passe sur un pays pour donner des informations sur ce territoire. Il disparaît ou change de données lorsque la souris change de position.

Pour réaliser cette carte, il est nécessaire de s’appuyer sur des fichiers JSON et plus particulièrement des fichiers geojson qui sont d’ailleurs bien gérées par la bibliothèque d3js. Ces derniers sont adaptés pour les données géographiques et associer des données à des territoires (cf annexe).

Malheureusement, trouver un fichier regroupant l’Afrique et les îles concernées par l’observatoire qui soit de bonne qualité a été impossible. Cela a pour conséquence que les Seychelles ne sont représentées que par 4 points. Il sera donc nécessaire de gérer ce désagrément.

Json

Plus faut-il zoomer avant de commencer sur les petites iles ?

Geojson

Plus zoom sur pays petits parce que ce sont eux qui nous intéressent

Gérer les zooms changements de coloration

Bibliothèque supplémentaire ajoutées ?

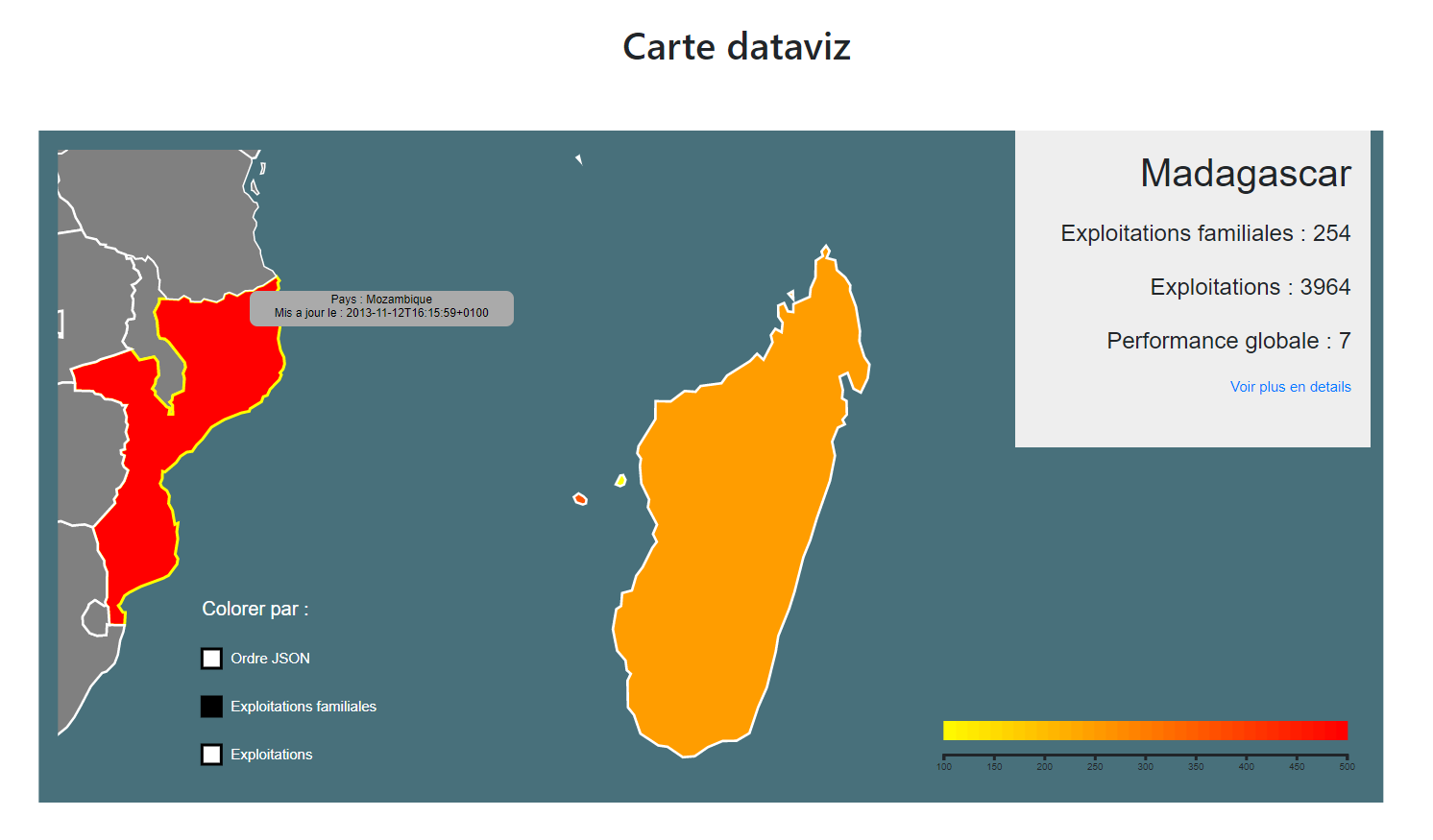


Figure : Data visualisation sur l’Afrique

Enfin, la dernière data visualisation intéressante

Travail sur la généricité

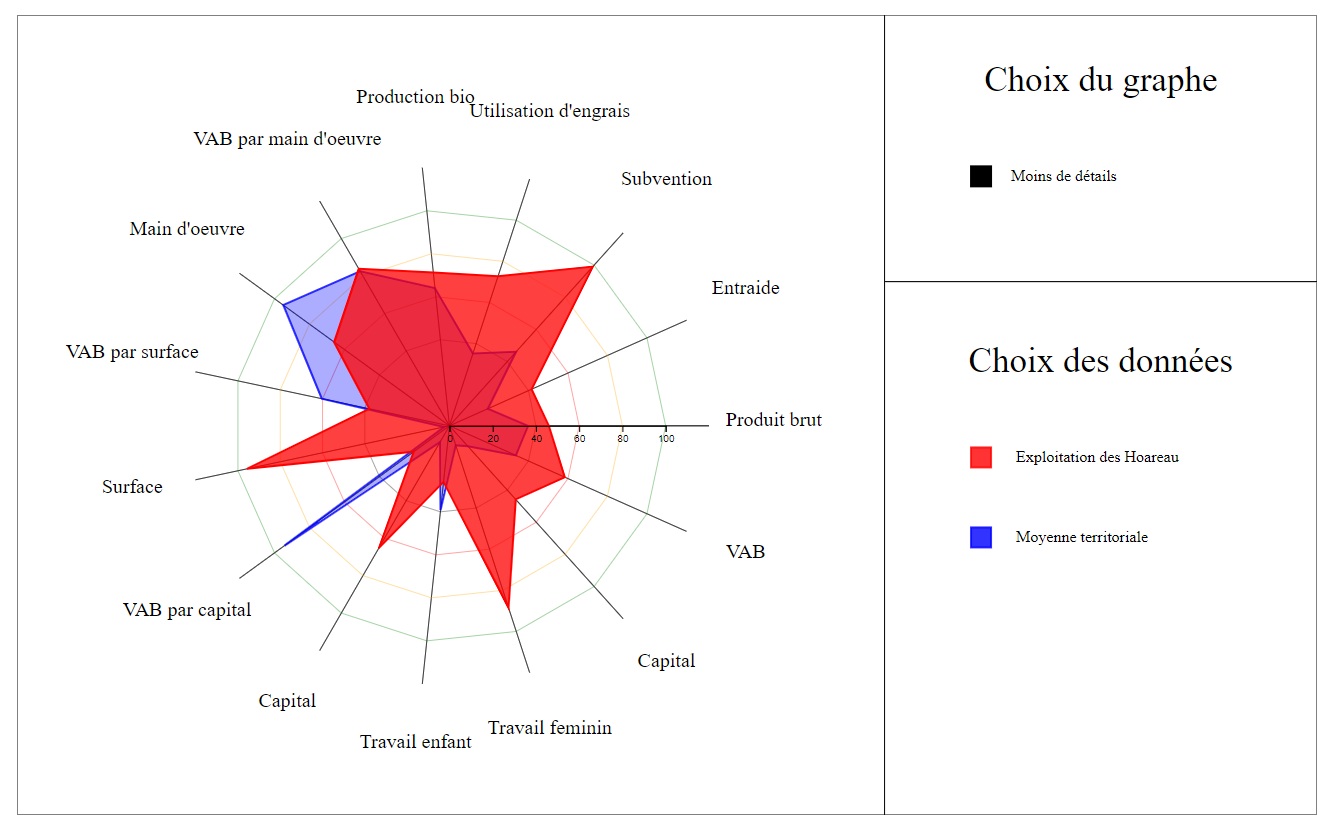


Figure : Radar Chart

## Analyse de données

Export d’un fichier csv et traitement. A venir

## Documentation

Le stage dure simplement 24 semaines et n’est donc pas suffisamment long pour réaliser le prototype complet de l’observatoire. Il est donc primordial de rédiger une documentation fournie et détaillée sur les différents travaux menés.

Suite par un autre stagiaire qui sera chargé spécialement de récupérer des données et

Un guide conséquent concernant la base de données, les vues et les variables a été rédigé. Par ailleurs, un dictionnaire des variables a été généré par grâce à SQL Power Architect. Dans le cahier des charges, les parties à améliorer ou à poursuivre sont signifiées.

# Bilan d’expérience

## Difficultés

Ce stage de 24 semaines m’a offert la possibilité d’être confronté à de nombreuses difficultés. Cela m’a permis de déceler les obstacles auxquels un ingénieur doit faire face au cours de l’exercice de son métier.

Coopération pas toujours évidente avec des acteurs dont les connaissances sont très diverses : agronome et informaticien possèdent des vocabulaires très différents. Par ailleurs, tous les acteurs impliqués dans le projet n’étaient pas présents sur place et la communication n’a pas toujours été parfaite.

Aussi, l’apprentissage constant de nouvelles technologies n’a pas été une difficulté en soi mais cela représente un temps de travail considérable. Ce travail est évidemment plus important en période de stage mais les technologies évoluent si vite qu’il faut savoir consacrer une partie de son temps de travail à la veille technologique.

Par ailleurs, un des plus gros problèmes pour la réalisation de l’OA-OI est le manque de données existantes concernant les variables désirées. En effet, afin de conceptualiser la base de données, il serait préférable d’avoir une idée des données qui sont déjà collectées et celles qui sont facilement récupérables. Sans ces informations, il est à prévoir que les variables présentes dans le MCD puissent changer car elles ne correspondent pas à une réalité du terrain.

D’autre part, la synchronisation des différents stagiaires, VSC et autres personnes travaillant sur le projet n’est pas une mince affaire. Madame Mialet-Serra n’étant pas particulièrement habituée à traiter avec des projets qui contiennent de l’informatique, certaines étapes n’étaient pas parfaitement coordonnées. Sur certains aspects, le stage aurait pu être effectué quelques mois plus tard. Le choix des variables utilisées dans le cadre de l’observatoire a été effectué assez tardivement et ne concernaient que l’exploitation agricole. Initialement, elles devaient aussi inclure des données à l’échelle de la filière. De plus, les objectifs précis de l’observatoire se sont aussi préciser au fur et à mesure du stage. Ainsi, le stage s’est apparenté pendant la première partie à faire de l’assistance à maîtrise d’ouvrage selon Sandrine Auzoux. Concilier des objectifs ambitieux, des données manquantes et des faisabilités techniques s’est révélé être très formateur.

Manque de réponse

Etre dans le flou

Mauvais ordre des choses

Smart gender Importance femme dans CDC

Data analytics sport

## Apports personnels

D’un point de vue plus personnel, ce stage a été d’un grand intérêt. En effet, la diversité des tâches et la continuité entre ces dernières a pu être une grande source de motivation. De plus, cela m’a permis de découvrir un très grand nombre de technologies que j’ignorais. En effet, étant arrivé en Génie Informatique un peu par hasard, je n’avais pas fait beaucoup d’UV d’informatique ce qui a impliqué certaines lacunes en terme de culture informatique. Ces dernières n’ont sans doute pas complètement disparu mais je ressens un certain gain de légitimité dans l’optique de l’obtention d’un diplôme d’ingénieur informatique.

Par ailleurs, durant cette période de 24 semaines dont une grande partie à rechercher, disséquer, analyser et examiner les données, j’ai pris conscience de l’intérêt qu’elles peuvent avoir. Aussi, j’ai pris goût à tenter de faire émerger des informations exploitables de ces données. Etant passionné par le sport et ses statistiques, j’ai ainsi décidé de m’orienter vers la data science appliquée au sport. J’ai d’ailleurs d’ores et déjà contacté une personne travaillant à l’INSEP (Institut National du Sport de l’Expertise et de la Performance) dans le but d’étudier les possibilités de TN10 dans le laboratoire de recherche de cet institut.

# Glossaire

3NF : 3e Forme Normale

AIDA : Unité de recherche « Agroécologie et intensification durable des cultures annuelles »

CIRAD : Centre de Coopération international de recherche agronomique pour le développement

COI : Commission de l’Océan Indien

DAAF : Direction de l'Alimentation, de l'Agriculture et de la Forêt

Dataverse : Dataverse est une application web à code source ouvert permettant de préserver, partager, citer, rechercher et analyser des données de recherche.

DRRM : Direction Régional Mayotte Réunion

FAO : Organisation des nations unies pour l’alimentation et l’agriculture (Food and Agriculture Organisation)

LSMS : Living standard Measurement Study

MCD : Modèle Conceptuel de Données

ODD : Objectifs du développement durable

OA-OI : Observatoire des Agricultures de l’Océan Indien

OAM : Observatoire des agricultures du monde (WAW : World Agriculture Watch)

Prérad-OI : Plateforme Régionale de recherche Agronomique pour le Développement dans l’Océan Indien

SGBDR : Système de Gestion de Base de Données Relationnelle

SVG : Scalable Vector Graphics (graphique vectoriel adaptable)

VSC : Volontaire Service Civique

Annexe

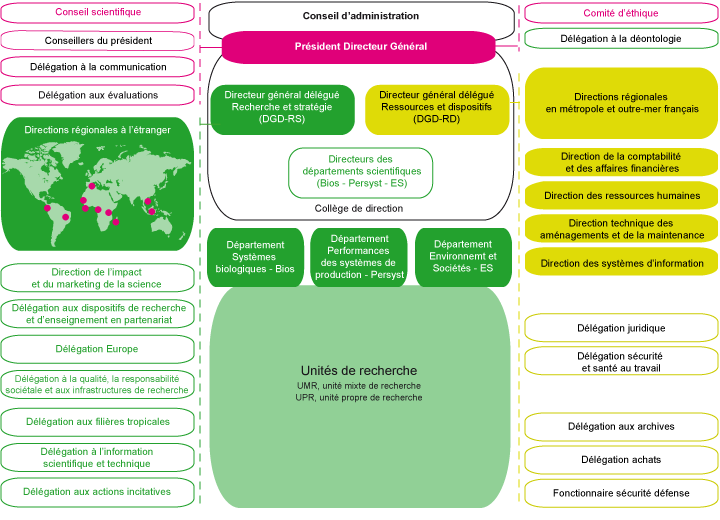
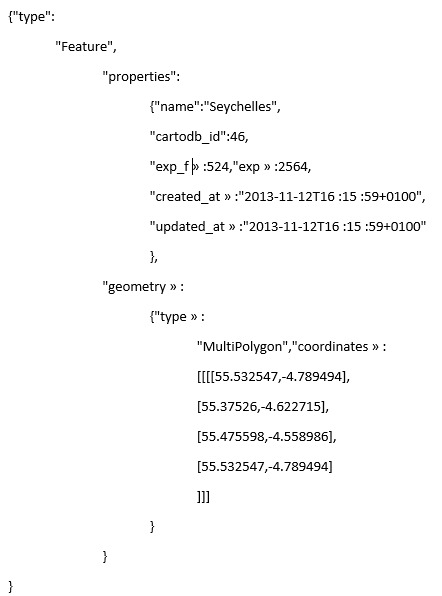


Figure : Histogramme du CIRAD

Les fichiers geojson sont des fichiers JSON ayant une structure regroupant une partie « properties » et une partie « geometry » qui contient les coordonnées.



Idées de choses à développer dans le rapport de stage :

Culture générale que je n’avais pas car je n’avais fait que peu d’uv de gi avant

Le plus simple mais qui rassemble tout, tient compte de ce qui existe mais avec des objectifs nouveaux.

Données =pb

Récupération de données :

Confidentialité des données

Compliqué de transférer des données entre deux bases de données différentes

On se rend compte de l’impact qu’ont les questionnaires et leur clarté. Exemple du chef d’exploitation qui apparaît ou non dans la population de l’exploitation. Choix de garder les données même si risque doublons. Le but ici est de commencer à réaliser un prototype et non d’avoir un produit fini. De plus, un stagiaire sera chargé de récupérer toutes les données existantes.

Nettoyage à effectuer

Données qui proviennent d’agronomes dont les bases de données ne sont pas propres avec des relations non existantes. Lien à rajouter, certains enregistrements ne peuvent pas être gardées car ne remplissent pas les conditions de clés primaires ou étrangères.

Problème de matériel, 2 changements de postes informatiques

PGD grosse galère pour récupérer des données individuelles à la daaf. Ne peuvent donner que des données agrégées ce qui est bien mais celles-ci ne sont pas complètes.

Echantillonnage pour récupérer données mais comment faire dans le cadre d’une OP. Récupérer les données pour leur donner des indicateurs ?

Module de Q supplémentaires ?

Partie politique entre le cirad et la daaf avec des contreparties des antécédents, on est loin d’avoir un niveau de partage optimal alors que nous travaillons tous pour le bien des agriculteurs.

Travail de longue haleine = récupération de données en 2022-23 mais sûrement retardée encore

Difficile jeu avec plein d’acteurs avec données différentes

Réunion riche en données mais pas les autres.

Baladé entre les technologies, les envies des uns et des autres, les ordres contradictoires, … projet vraiment instable et précaire. On ne sait pas sur qui on va pouvoir compter ni qui va compter sur nous. Pas établi, besoin de cette proof of concept, vraiment ?

Manque de motivation car je ne vois pas le marché derrière.

Désaccord avec le chef de projet (entretien)

DESACCORD SUR LA TECHNOLOGIE donc apprentissage dans le vide, perte de temps

UTILISE comme un informaticien et non comme un ingénieur

Faire des requêtes pour montrer le manque de remplissage de la base de données

Importance CodeIgniter vs Symfony combat

Manque de structure

Trop d’acteurs pour contenter tout le monde