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 1 : 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 et qui m'ont aidé dans la rédaction de ce rapport.

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 Madame MIALET-SERRA, responsable de la PRéRAD dont le projet d'Observatoire de l'agriculture de l'Océan Indien est une des composantes, 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 d'expertise.

Je remercie également Mme Darras pour la collaboration fructueuse que nous avons eue.

Par ailleurs, je remercie Madame Dutour, qui s'occupe notamment des logements du CIRAD, pour sa gentillesse et sa réactivité.

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

II. Résumé technique

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.

Ce rapport de stage contient une brève présentation de l'institut d'accueil et du projet de l'observatoire des agricultures de l'Océan Indien (OA-OI) dans lequel s'intègre le stage. Par la suite, les missions et l'environnement du stage sont présentés. Les réalisations ainsi que les choix effectués sont ensuite développés. Enfin, certaines réflexions personnelles sont évoquées à la fin du rapport.

Le stage consiste, en premier lieu, en la rédaction d'un cahier des charges de l'application Web liée à l'observatoire. Cette dernière s'appuie sur des données concernant les exploitations agricoles de la Réunion et de certains de ses pays voisins. Par la suite, la base de données associée à cette application est à conceptualiser et à implémenter. Cette base est ensuite à alimenter et à exploiter par l'intermédiaire de vues et de data visualisations adaptées. Parallèlement à ces tâches axées sur la partie donnée, une première interface Web est à développer.

L'application évoquée est une preuve de concept (POC) de l'observatoire des agricultures de l'Océan Indien qui a pour but de démontrer sa faisabilité.

Mots clés : cahier des charges, base de données, data visualisation, programmation Web

Note : Ce rapport cherche a été modifié afin de répondre aux exigence du CIRAD qui publiera ce document sur Agritrop, une bibliothèque en ligne.

III. Contexte du stage et projet

A. Présentation de l'institut d'accueil : le CIRAD

Le Centre de coopération Internationale de Recherche Agronomique pour le Développement (CIRAD) est un Établissement Public à caractère Industriel et Commercial (EPIC) français fondé en 1984. 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. Il est spécialisé dans la recherche agronomique dans les régions aux climats chauds. En raison 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 (le CIRAD a des activités dans plus de 100 pays).

Figure 2 : Carte des directions régionales du CIRAD

Le CIRAD emploie actuellement 1650 personnes, dont 800 chercheurs, répartis en 3 départements scientifiques et 33 unités de recherche (l'organigramme du CIRAD est présenté en Annexe 1). Les chiffres placent cet établissement parmi les plus importants dans son domaine en France, mais le CIRAD derrière, en terme d'effectif, des centres de recherches plus diversifiés tels que le CNRS (Centre National de la Recherche Scientifique) dans lequel travaillent plus de 10 000 chercheurs.

D'autre part, le partenariat fait partie de l'identité du CIRAD et les 12 directions régionales, représentées sur la carte ci-dessus, permettent une grande coopération. Ces directions régionales coordonnent les actions des différentes stations et unités de recherche tout en assurant la répartition du budget. Elles supervisent également les missions ayant des domaines d'actions géographiquement étendus.

La Direction Régionale Réunion-Mayotte (DRRM) au sein de laquelle s'est déroulé le stage, se trouve à la station « La Bretagne », près de Saint-Denis de la Réunion. Les chercheurs de cette station étudient principalement la canne à sucre en menant des expérimentations sur des parcelles à proximité.

B. Financement et objectifs actuels

Le CIRAD couvre les deux tiers de ses frais grâce aux financements publics et le tiers restant provient de revenus contractuels. Récemment, il a signé, avec ses ministères de tutelle, son nouveau Contrat d'Objectifs et de Performances pour la période 2019-2023. Le CIRAD s'est ainsi engagé 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 et ODD 1).

C. Projet du stage

1. La naissance de l'Observatoire des Agricultures du Monde (OAM)

a) *Contexte*

La décennie 2019-2028 a été déclarée décennie de l'agriculture familiale par les Nations Unies. Ce type d'agriculture représente presque 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 ont une surface inférieure à un hectare [1]. Pourtant, la majorité de ces agriculteurs sont touchés par la pauvreté et sont vulnérables aux aléas météorologiques et aux fluctuations des prix des marchés.

Parallèlement, 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. Les deux premiers objectifs, 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 sont 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 de plus en plus fortement, les exploitants agricoles doivent produire pour nourrir une population grandissante en limitant leur impact sur l'environnement.

Malgré leur importance, autant quantitative 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 difficultés qu'elles rencontrent ou le travail qu'elles génèrent. Actuellement, les enquêtes LSMS (« Living Standard Measurement Study » menées par la banque mondiale) et les recensements agricoles coexistent. Les premières se concentrent sur les conditions de vie des ménages, tandis que les seconds explorent les structures des exploitations. Cependant, aucun lien de gouvernance n'existe entre ces études. Les durées de livraison de ces dernières sont soit variables soit très longues (le recensement agricole est une collecte

[1] Ces données proviennent de la FAO : http://www.fao.org/family-farming/themes/smallfamilyfarmers/fr/#:~:text=Environ%2090%20pour%20cent%20des,rurales%20de%20pays%20en%20d%C3%A9veloppement

décennale). Dans ce cadre, il est difficile de prendre les décisions pour influencer positivement l'avenir des exploitations agricoles.

C'est dans ce contexte qu'un groupe de travail autour de M. BOSC, chercheur en agroéconomie au CIRAD et 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 principal de caractériser la diversité des exploitations afin d'éclairer les décisions des parties prenantes en matière de politiques publiques. Ces dernières peuvent notamment être des producteurs, des organisations de producteurs ou des agences de développement. Grâce à des données partagées et cohérentes, ces acteurs pourront élaborer des stratégies inclusives et différenciées.

En septembre 2020, le projet repose principalement sur un guide opérationnel dont la rédaction a été coordonnée par M. Bosc. Ce guide regroupe les différentes étapes de la réalisation d'un « observatoire ». Il expose également les définitions des concepts agricoles importants, mais aussi les méthodes pour élaborer un questionnaire ou choisir les échantillons d'exploitations à sonder.

b) *Cadre méthodologique*

Le projet d'observatoire repose sur le cadre Livelihood[2], qui permet une étude des exploitations à partir de 5 capitaux :

- Le capital humain prend en compte la main d'œuvre employée, de manière qualitative et quantitative ;
- 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 éléments physiques mis en œuvre dans le cadre de la production (notamment les animaux, les infrastructures et les équipements) ;
- Le capital social repose 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 et l'entraide entre exploitants.

Les données sur ces capitaux permettent de catégoriser les exploitations et devraient ainsi permettre la formulation de politiques différenciées.

[2] Livelihood Framework : http://www.fao.org/docs/up/easypol/581/3-7-social%20analysis%20session_167en.pdf

2. Création de l'Observatoire des Agricultures de L'Océan Indien (OA-OI)

La Commission de l'Océan Indien (COI) est une organisation intergouvernementale regroupant la France (à travers la Réunion), Madagascar, l'Union des Comores, l'île Maurice et les Seychelles. Cette organisation défend les intérêts de ces Etats insulaires et cherche à favoriser la coopération entre ses membres. Elle a manifesté son intérêt pour la création d'un observatoire des agricultures dans l'Océan Indien, auprès du CIRAD. Les pays de la COI partagent des caractéristiques fortes comme l'insularité ou le climat. Il est donc nécessaire de créer un observatoire, à une échelle régionale, adapté aux contraintes des pays membres. C'est dans ce contexte que la Direction Régionale Réunion Mayotte du CIRAD a entrepris de décliner le projet de l'observatoire des agricultures du monde sur les pays de la COI. L'OA-OI sera déployé par l'intermédiaire de la plateforme régionale en recherche agronomique pour le développement dans l'Océan Indien (PRéRAD-OI).

Figure 3 : Commission de l'Océan Indien (COI)

Le stage a été proposé par la PRéRAD-OI et s'inscrit dans la création d'une preuve de concept de l'observatoire adapté à la région. Il a ainsi été financé par le Fond de Coopération Régionale (FCR) qui soutient les projets visant à renforcer l'intégration de la Réunion dans l'Océan Indien.

D. L'équipe projet

Le projet de l'OA-OI est un projet pluridisciplinaire qui nécessite le travail, la collaboration, les connaissances et l'appui de plusieurs personnes :

- Mme Mialet-Serra est responsable de la coopération régionale au CIRAD à La Réunion et la coordinatrice du projet basée à la station « La Bretagne ».
- Mme Darras est une volontaire service civique (VSC) en agronomie, également basée à « La Bretagne » qui avait en charge l'inventaire et la collecte des données nécessaires pour l'observatoire.
- M. Bosc, basé aux locaux de la FAO à Rome, qui porte 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 », informaticienne spécialisée dans la conception de bases de données et encadrante du stage. Elle 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, est un chercheur en agroéconomie, basé à Madagascar, qui possède une grande expertise des données de terrain. Son rôle a été important pour la conception de la base de données.

IV. Présentation du stage

A. 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 une analyse de l'existant des outils du CIRAD en lien avec le projet ;
- Elaborer le modèle conceptuel de données sur les exploitations agricoles ;
- Créer la base de données sous PostgreSQL ;
- Rédiger le cahier des charges de l'application Web sur la base d'une analyse des besoins ;
- Développer une preuve de concept de l'application Web avec les fonctionnalités décrites dans le cahier des charges sous un environnement Web ;
- 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).

B. Contraintes contextuelles

1. Epidémie de COVID-19

La prise en compte des besoins des futurs utilisateurs potentiels est un prérequis à la rédaction du cahier des charges. Malencontreusement, le recrutement de deux VSC a dû être repoussé en raison de l'épidémie de COVID-19. Ils auraient dû être en charge de la récolte des données existantes à Maurice et aux Seychelles et de faire le lien avec les utilisateurs potentiels. Ainsi, le projet ne disposait pas d'interlocuteurs assimilables à des clients. Plusieurs partenaires du CIRAD, comme les réseaux SOA et RuralStruct, qui sont des organisations de producteurs à Madagascar, ont pu être contactés. Cependant, il n'a pas été possible de travailler directement avec eux sur leurs besoins réels. Par conséquent, le cahier des charges n'a pas pu être rédigé aussi finement qu'envisagé ce qui a également limité la réalisation de l'application Web.

2. Interopérabilité

Par ailleurs, les bases de données de la FAO ne contiennent pas de données à l'échelle de l'exploitation, mais plutôt à des plus grandes échelles (filière, territoire, paysage, …). De plus, les données n'étaient pas publiques. De ce fait, la partie interopérabilité du stage a dû être supprimée.

3. Adaptations du programme du stage

Afin de compenser ces contretemps, une partie data visualisation a été ajoutée au stage. Les data visualisations constituent des livrables pérennes et réutilisables indépendamment des données disponibles. Une analyse de données avec du Machine learning a aussi été réalisée.

C. Planning

Le planning initial du stage est représenté par le diagramme de Gantt ci-dessous.

Les tâches ont été réparties, sur le diagramme, en trois catégories :

- En vert, apparaissent les tâches qui nécessitent davantage des compétences en communication que des compétences techniques ;
- En marron, sont représentées les parties ayant trait à la base de données. La base de données est le livrable prioritaire ;
- Enfin, les tâches en bleu, sont celles concernant la programmation Web.

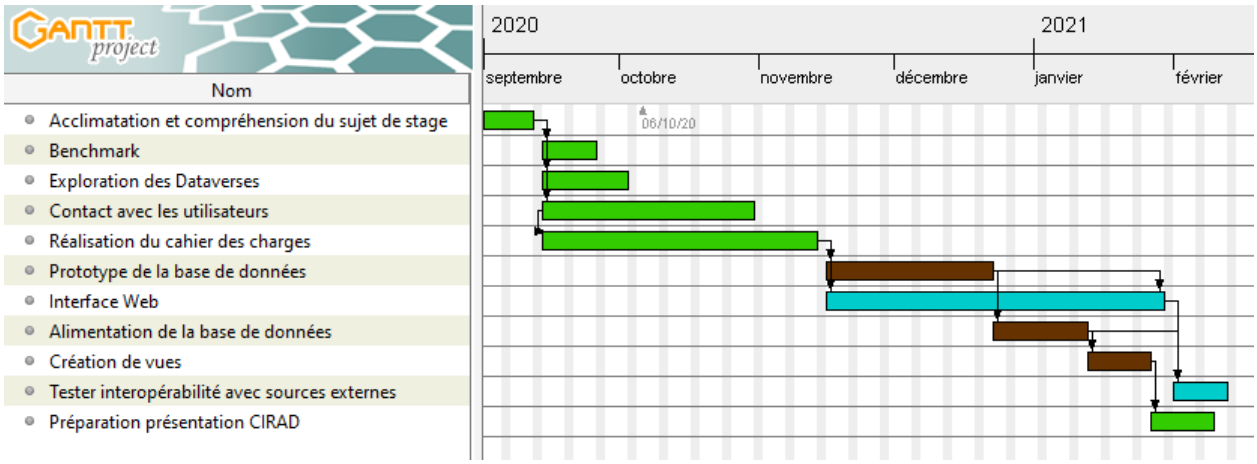


Figure 4 : Diagramme de Gantt du stage à l'origine

Dans le monde de la recherche, planifier les tâches est très difficile. Les idées nouvelles engendrent des modifications de tâches dont l'impact sur les échelles de temps envisagées est difficilement quantifiable. Finalement, et pour les raisons évoquées dans la partie « contraintes contextuelles », le stage a plutôt suivi l'agenda suivant :

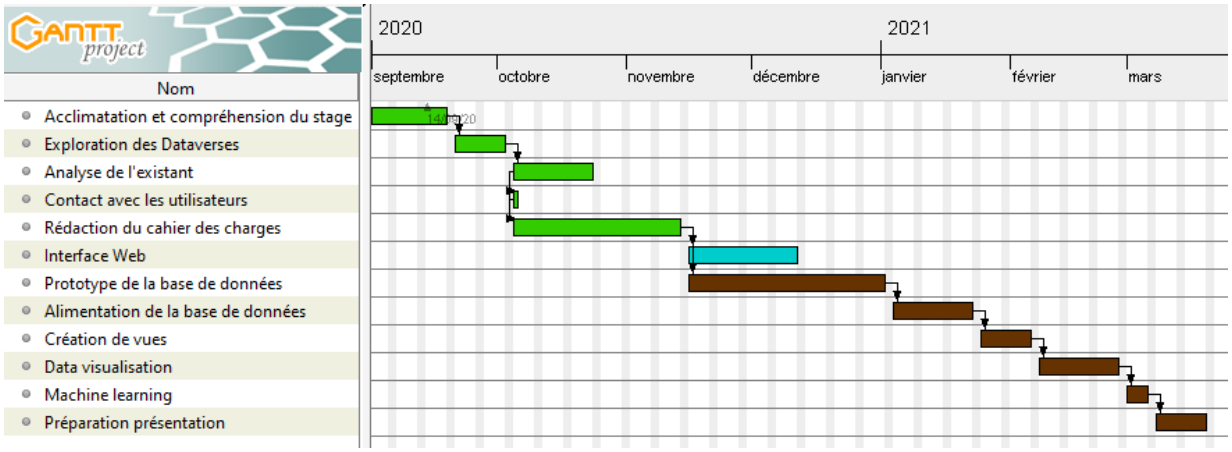


Figure 5 : Diagramme de Gantt final

Ce diagramme illustre un des points forts du stage qui est la diversité des tâches effectuées. L'ensemble des étapes du cycle de vie des données a été couvert.

D. Méthodologie

L'observatoire étant encore un concept avec de nombreuses inconnues, le travail effectué était proche d'un travail de recherche. Les travaux menés étaient modelés par les discussions, au sein de l'équipe, durant lesquelles de nouvelles idées prenaient place. Le workflow s'apparente à une méthode AGILE, à la différence que les révisions sur le projet sont effectuées au sein de l'équipe et non avec des clients. Cette méthodologie permet une grande flexibilité.

Par ailleurs, la majorité des tâches étaient effectuées par groupes de 2 ou 3 personnes et ne disposaient pas de deadlines fortes. Dans ce cadre, j'ai joui d'une grande autonomie.

E. Outils utilisés

De multiples outils ont été utilisés au cours du stage. En voici une liste non-exhaustive.

1. 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 projet git. Git est très utilisé pour développer des fonctionnalités indépendantes simultanément. 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 de l'application Web sur laquelle j'étais seul à travailler.

2. SQL Power Architect

SQL Power Architect est un logiciel open source permettant de réaliser du Forward Engineering de modèles de bases de données. Cela signifie que l'utilisateur crée graphiquement le modèle logique 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 (système de gestion de base de données). Ce logiciel permet également de générer un dictionnaire de la base de données qui intègre les commentaires sur les tables et sur les colonnes.

3. PostgreSQL et PgAdmin

Le choix du SGBDR (système de gestion de base de données relationnelle) a été imposé. PostgreSQL, un SGBDR open source et largement répandu, avait aussi été choisi afin de profiter du module spatial POSTGIS qui possède des fonctionnalités propres aux données géo référencées. L'outil d'administration de la base de donné utilisé était PgAdmin 4 qui permet de créer les tables et des vues de manière accélérée. En outre, PgAdmin donne la possibilité d'ajouter des contraintes avancées aux tables et de créer des triggers.

Figure 6 : Logo PostgreSQL

4. Microsoft Office Access

Access est un logiciel propriétaire de la suite Office qui permet de gérer des bases de données relationnelles. Par l'intermédiaire de pilotes ODBC, Microsoft Access peut accéder à des bases de données distantes créés sur d'autres SGBD comme PostgreSQL, par exemple. Tout changement effectué sur les données depuis Access est automatiquement appliqué sur la base de données PostgreSQL. Ces changements concernent uniquement les données et non pas la structure de la base qui est préservée.

Le filtrage des lignes, la fonction « recherche et remplace » ainsi que l'ajout de colonnes avec des valeurs prédéfinies ont été d'une utilité précieuse lors d'intégrations de données.

5. PHP et CodeIgniter

Le choix du langage PHP a été imposé. CodeIgniter est un Framework PHP dont l'utilisation permet de faciliter et d'accélérer la production de site Web, à travers une programmation objet. Ce Framework se distingue par une grande légèreté associée à une grande rapidité d'apprentissage. Ce Framework a été choisi, car il est utilisé régulièrement par Madame Auzoux. Le développement PHP a été effectuée sur l'environnement PHP Storm.

6. 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 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, utilisé pour placer les composants visuels et leur attribuer une taille. Ainsi, pour chaque composant visuel, nous pouvons 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 dont le code est disponible sur le site. Des barres de navigation ou des menus déroulants sont par exemple présentés et sont déjà prêts à l'emploi. Un style est déjà appliqué à ces éléments ce qui contribue à accélérer le développement Web. De plus, il est toujours possible d'éditer les styles appliqués dans un nouveau fichier CSS supplémentaire.

7. D3.js

D3js est une bibliothèque JavaScript utilisée pour la création de data visualisations. Elle permet de manipuler des éléments HTML. Les SVG (Scalable Vector Graphics), que D3js peut créer, permettent de créer des formes simples telles que des lignes, des rectangles ou des cercles. 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.

Figure 7 : Logo D3js:

Par ailleurs, le site d3js.org exposent des milliers d'exemples de data visualisations ainsi que le code qui permet de les créer. Il reste néanmoins nécessaire de comprendre le fonctionnement de d3js et de ses fonctions, afin de créer ses propres mises en forme, adaptées à ses données et à ses besoins

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.

8. Python et les librairies

Python est très utilisé en machine learning et c'est dans ce but qu'il a été utilisé durant le stage. L'environnement de développement utilisé a été Spyder.

La librairie Pandas permet de traiter les données qui sont sous forme de data frame. Ces derniers sont semblables à des tableaux, possédant de nombreuses fonctions qui sont très pratiques pour sélectionner les données ainsi que pour effectuer des prétraitements. Des fonctions permettent également de lire des fichiers externes (csv notamment) ou de convertir les data frame en fichiers.

Pour débuter avec le machine learning, la librairie scikitlearn est très largement recommandée. Elle a été utilisée pour les modèles qu'elle contient ainsi que les fonctions pour effectuer des tests sur ces dits modèles.

La représentation des données sous forme de graphe s'effectue avec la librairie matplotlib.

9. 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.

V. Déroulement du projet

Les parties suivantes abordent les différentes réalisations dans un ordre chronologique.

A. Appropriation du sujet du stage

La première partie du stage a consisté à intégrer le domaine métier, c'est-à-dire comprendre les vocabulaires agronomique, économique et social utilisés pour caractériser l'OA-OI. De nombreux documents provenant du projet ont été mis à ma disposition.

Ensuite, le stage s'est concentré sur les besoins attendus pour l'observatoire des agricultures de l'Océan Indien. Je n'ai pas hésité à poser de nombreuses questions afin de comprendre le projet au risque que les réponses puissent paraître évidentes. Ces questionnements provenant d'une personne avec un œil externe au projet, ont parfois permis de clarifier certains points ou de mettre en évidence d'autres points dont les problématiques n'avaient pas été soulevées.

Pour résumer cette phase du stage, l'objectif final du projet est la réalisation d'une application Web (appelée « observatoire » dans ce rapport) qui doit répondre aux problématiques soulevées précédemment. Pour cela, elle doit permettre :

- 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 de son projet.

Dans un premier temps, seules les données à l'échelle de l'exploitation sont concernées. A long terme, l'observatoire devrait aussi permettre de collecter et de visualiser des données à l'échelle de la filière et du territoire.

B. Analyse de l'existant

Après avoir intégré les différents objectifs de l'OA-OI, il est nécessaire de savoir si des outils répondent à des besoins similaires. J'ai donc réalisé une analyse de l'existant sur les outils du CIRAD.

Il s'agit d'étudier les outils mis en place par le CIRAD et de savoir s'ils pourraient être intégrés plus ou moins directement dans le cadre de l'observatoire. Leur éventuel succès ou échec est aussi pris en compte.

Tous les outils étudiés n'avaient donc pas vocation à être la base du futur observatoire. J'ai relevé les points forts et les points faibles de chaque outil en lien avec les besoins de l'OA-OI. La conclusion de cette analyse est qu'aucun outil, à l'heure actuelle, ne répond aux besoins formulés pour l'observatoire. En effet, la diversité des échelles, des acteurs impliqués et des objectifs affichés rend l'observatoire unique. Voici un résumé de cette analyse.

Outils	Avantages	Inconvénients
AGREF	-Echelle exploitation -Application Web indépendante -Saisie données -Export données -Application mobile -Différents profils utilisateurs	-Design -Failles de sécurité -Non opérationnel -Nombreuses variables inutilisées -Pas de données agrégées -Pas d'indicateurs -Pas de gestion à l'échelle filière
AEGIS	-Système d'information -Design -Data visualisation -Analyse -Collecte des données -Base de données générique -Multi-échelles et pluridisciplinaire	-Pas de données à l'échelle de l'exploitation actuellement -En cours de développement
E-Watch	-Requête par la cartographie -Développé pour tablette -Statistiques par zone/projet	-En refonte -Peu adapté à divers profils
IDEA RUN	-Echelle exploitation -Méthode d'auto-évaluation rapide -Co-créée avec acteurs locaux	-Adaptée uniquement à la Réunion -Trop de données supplémentaires
FAST	-Méthode et critères de recherche	-Pas de classement des résultats de la recherche par ordre préférentiel
LASER	-Package sous R pour effectuer les calculs poussés	
Phyto'aide	-Différents leviers d'action sont utilisés et explicités	-Modèle très simple, pas adapté à des problèmes complexes
HiH	-Evolution données avec « vidéo » sur la carte -Curseur droite/gauche pour comparaison -Carte et graphique liés	

Figure 8 : Analyse de l'existant

Lors de mon analyse de l'existant, il s'est avéré que de nombreux outils avaient été peu utilisés. La communication sur ces outils ainsi que le suivi (maintenance, apprentissage) sont des éléments clé de la réussite d'un développement informatique.

Par ailleurs, les outils développés par le CIRAD peuvent être classés en 2 catégories :

- Les outils orientés recherche, utilisés pour des fonctionnalités techniques par des chercheurs ;
- Les outils pour le grand public plus attachés à l'expérience utilisateur.

L'observatoire répond aux deux exigences à la fois, d'où l'intérêt d'étudier ces deux types d'outils.

Par ailleurs, j'ai proposé de pouvoir réutiliser certains des outils évalués en post-traitement si cela apportait une réelle plus-value. Par exemple, Phyto'aide 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'il utilise. Si nous souhaitons guider les exploitants vers de meilleures pratiques d'utilisation de produits phytosanitaires, une passerelle pourrait être créée vers Phyto'aide.

C. Recueil des besoins des utilisateurs par un questionnaire d'enquête

Dans un projet visant à impliquer de nombreux acteurs dans son fonctionnement, intégrer leurs besoins est absolument primordial.

En collaboration avec Mme Darras, nous avons élaboré un questionnaire Google Form afin d'identifier les partenaires possibles, les données dont ils disposent, et d'avoir un aperçu général de leurs besoins potentiels. Ce questionnaire devait, par la suite, être suivi par des interviews pour spécifier les attentes de chacun. Les informations recueillies devaient permettre de rédiger le cahier des charges de l'observatoire. Ce questionnaire a donc été soumis aux partenaires du CIRAD (des agences de développement et des organisations de producteurs principalement). 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, j'ai voulu relancer les partenaires qui n'avaient pas répondu et approfondir les contacts avec les autres. Cela n'a pas été possible durant mon stage et l'enquête se poursuivra après celui-ci.

Cette partie du travail a été particulière, car les acteurs ne sont pas réellement des clients. Il était donc nécessaire de concilier les besoins réels des futurs utilisateurs (organisations de producteurs et agence de développement) avec les intentions des porteurs de projet. Ce point contraste avec les projets informatiques en général.

D. Recueil des données

Afin d'alimenter la future base de données, il faut recueillir des données sur des exploitations agricoles des pays de la COI. Préférentiellement, ces données doivent provenir de sources différentes afin d'observer le panel des variables récoltées. Ces sources peuvent être notamment des producteurs, des organisations de producteurs, ou des enquêtes réalisées par des chercheurs. La collecte de données directement chez l'exploitant étant une opération chronophage et très coûteuse, il est essentiel d'exploiter ces sources.

1. Exploration des Dataverses

Le CIRAD, l'IRD et l'INRA (trois instituts de recherche), stockent de grandes quantités de données de recherche dans leur Dataverse. L'exploration de ces grandes sources de données a été infructueuse. En effet, des données agricoles sont disponibles mais concernent d'autres zones géographiques ou d'autres échelles que celles recherchées.

2. Questionnaire

Dans le questionnaire évoqué précédemment, les partenaires devaient également renseigner les coordonnées des personnes à contacter pour récupérer les données dont ils disposent. En outre, 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. Cependant, ces dernières concernaient directement le projet du réseau (qui consistait à aider des jeunes à se lancer dans l'agriculture) et pas les exploitations agricoles dans leur ensemble.

3. Données de projet de recherche

Par ailleurs, M. Bélières, ayant travaillé à Madagascar sur les exploitations agricoles, a fourni deux bases de données issues de ses précédentes recherches. Ces bases contiennent de nombreuses données à l'échelle de l'exploitation qui sont pertinentes dans le cadre de l'observatoire. Elles couvrent plus de 800 exploitations des réseaux SOA et RuralStruct.

4. DAAF

Dans la continuité de cette recherche de données, des contacts ont été établis avec la DAAF de la Réunion (Direction de l'Alimentation, de l'Agriculture et de la Forêt). Elle dispose d'un département statistique et donc de nombreuses données sur les exploitations agricoles de la Réunion qui pouvaient servir au projet.

La DAAF n'est pas autorisée à transmettre des données brutes, car celles-ci sont personnelles donc elles relèvent du RGPD (Règlement Général sur la Protection des Données) et sont soumises au secret statistique. Elle a tout de même accepté de faire parvenir des données agrégées et anonymisées au CIRAD pour les intégrer à l'observatoire, dans le cadre d'une convention.

5. Perspectives

Dans les prochains mois et les prochaines années, les données déjà collectées devraient être complétées par des enquêtes de terrain selon le montant des financements accordés à l'observatoire. Par la

suite, ces collectes se concentreront sur des échantillons représentatifs des exploitations agricoles dans chaque pays de la COI.

6. Remarques

Des recherches ont probablement été effectuées sur des exploitations agricoles dans les pays concernés par l'observatoire. Le travail d'anonymisation, nécessaire à la publication des résultats, a possiblement constitué un frein pour cette démarche de partage. Faciliter la réutilisation des données devrait être une préoccupation majeure compte tenu du temps nécessaire à leur obtention de ces données et l'intérêt qu'elles peuvent renfermer.

D'autre part, ce faible nombre de données récoltées constitue un obstacle pour l'OA-OI puisque ce dernier a intérêt à regrouper des informations sur le plus grand nombre d'exploitations possibles. Par ailleurs, cela signifie aussi que l'observatoire, lorsqu'il sera opérationnel, permettra de pallier certains déficits en données dans ce domaine.

E. Réalisation du cahier des charges

Afin de clarifier et de consigner l'avancement des réflexions autour de l'OA-OI, j'ai rédigé un document qui sera appelé « cahier des charges » à défaut de trouver un mot plus adapté à cet ouvrage. Il ne s'agit pas d'un réel cahier des charges comportant toutes les spécifications fonctionnelles et techniques, le budget et le planning. En effet, la rédaction d'un tel cahier des charges implique d'avoir une vision détaillée et précise de tous les éléments de l'observatoire, ce qui n'était pas le cas à ce stade du projet.

La rédaction de ce livrable a été le fruit d'une collaboration avec l'équipe projet. Les prochaines parties décrivent succinctement le contenu de ce document d'une vingtaine de pages.

1. Fonctionnalités

Voici une liste non-exhaustive des fonctions essentielles envisagées et qui sont décrites dans le cahier des charges :

- Collecte des données ;
- Création et visualisation d'indicateurs sur des tableaux de bord ;
- Exportation des données pour offrir la possibilité à des chercheurs de réaliser des études ;
- Publication de ces études pouvant contenir des vidéos appelées « paroles d'agriculteurs », dans le but de valoriser les pratiques agricoles durables.
- Publications de bilans réguliers sur l'état des agricultures familiales ;
- Création, suivi et évaluation de projets par des agences de développement ;
- Administration des comptes utilisateurs ;
- Annuaire d'utilisateurs de l'observatoire.

2. Profils

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 autorisant l'accès à des fonctionnalités distinctes. Un utilisateur peut disposer de plusieurs profils. Cela permet, par exemple, à un producteur de s'impliquer dans une étude ou un projet.

3. Droits des utilisateurs

L'observatoire offre un accès public et un accès privé, avec un compte. Pour ces derniers, il est nécessaire de définir des droits utilisateurs sur les données. Ces droits sont renseignés dans des matrices. 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(e)	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 9 : Matrice des droits sur une exploitation

4. Perspectives

La rédaction de ce livrable est à poursuivre. Des informations complémentaires de la part des utilisateurs sont attendues pour finaliser ce cahier des charges. Ainsi, le cahier des charges contient des propositions qui restent à affiner concernant :

- L'architecture de l'application Web ;
- Les fonctionnalités de cette application ;
- La composition des tableaux de bord permettant aux exploitations et aux agences de développement de visualiser les données ;
- Les spécifications techniques ;
- Les améliorations à envisager.

F. Conception de la base de données

L'objectif est de créer une base de données permettant de produire des indicateurs sociaux-économiques et environnementaux à l'échelle de l'exploitation agricole. La construction de cette base va permettre de stocker une multitude d'informations sur des exploitations permettant de générer ces indicateurs. Une première version de la base de données a été conçue et implémentée depuis le modèle logique de données grâce à SQL Power Architect. Elle devrait évoluer vers une « base de données modèle » afin d'être réutilisée dans le cadre de nouvelles enquêtes à l'échelle de l'exploitation.

La base de donnée a été conçue en s'appuyant sur un inventaire détaillé des variables d'intérêt pour l'observatoire[3].

La base de données contient 42 tables dont 31 pour la partie collecte des données des exploitations et 8 pour la partie utilisateur de l'application Web qui est encore en évolution. Les 3 tables restantes stockent les conversions entre les unités, ce point sera développé ultérieurement. L'enregistrement d'une exploitation nécessite le renseignement d'environ 200 variables.

Cette base de données a été conçue en 3NF. Cette structuration de la base de données assure un faible niveau de redondance tout en conservant toutes les informations et les dépendances fonctionnelles. La redondance pouvant induire une perte d'intégrité des données, il est essentiel de la limiter.

La partie suivante expose des exemples représentatifs des choix effectués lors de la conception de la base. Certains de ces choix ont été effectués suite à l'intégration des données de terrain.

1. Choix

a) *Réduction du nombre de variables et limitation du nombre de tables*

Afin que la collecte de données sur le terrain soit réalisable dans un temps raisonnable, il a été primordial de réduire le nombre de variables présents dans l'inventaire.

D'autre part, afin d'optimiser la base de données, le nombre de tables a été limité. Par rapport à l'inventaire, de nombreux regroupements ont été effectués. Par exemple, toutes les productions, animales végétales ou issues de transformations ont été regroupées dans une seule table. Ce travail de regroupement simplifiera notamment la collecte de données et la création de requêtes.

b) *Notion de temps*

La majorité des entités de la base ont des variables qui dépendent du temps. Afin de concevoir une base de données en 3NF, il a été nécessaire de séparer certaines entités en deux tables selon si les variables dépendaient du temps ou non. C'est le cas de l'entité individu.

[3] Darras *et al.*, 2021 : https://agritrop.cirad.fr/597467/

(1) Exemple de l'entité « individu »

Pour cette entité, les dépendances fonctionnelles suivantes sont observées :

Id -> nom, prenom, annee_naissance

Id, annee_observation -> sexe, apprentissage_pro, niveau_scolarisation, alphabetise

J'ai donc créé deux tables appelées « individu_invariable » et « individu » en concordance avec ces dépendances fonctionnelles.. Cette architecture permet de conserver une architecture en 3NF.

Le choix de placer le sexe dans la partie «variable» des données permet de s'adapter à un plus grand panel de possibilités. J'ai jugé les conséquences négligeables sur le temps de collecte et la quantité de données à stocker.

(2) Réflexions personnelles

Certaines variables ne varieront vraisemblablement pas d'une année sur l'autre (comme les infrastructures ou les terres possédées). Pour chaque table, lors d'une collecte de données, il pourrait être avantageux de signifier qu'aucun changement n'a été opéré. Cela permettrait de remplir les champs de la même manière que la collecte précédente en modifiant uniquement l'année. Cette possibilité est évoquée dans le cahier des charges.

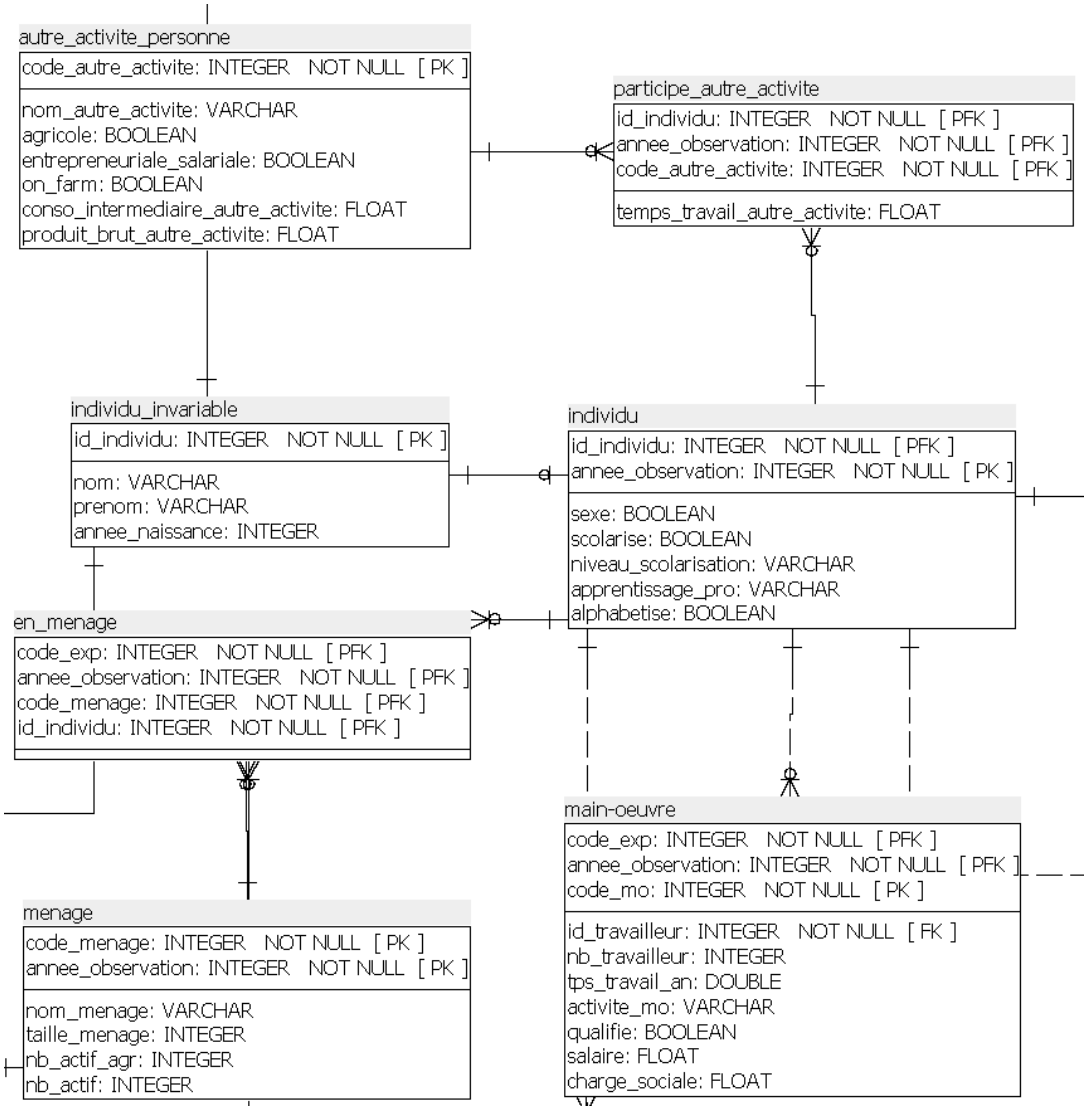


Figure 10 : Extrait du diagramme Power Architect de la base de données de l'OA-OI

2. Modifications pour le respect du Le Règlement Général sur la Protection des données (RGPD)

Le RGPD est applicable à partir de 2018 pour tous les citoyens Européens mais également pour les enquêtes réalisées par des établissements Européens hors de l'Europe. L'observatoire doit donc s'y conformer.

Afin de respecter le RGPD, la base de données qui servira pour l'application Web a été anonymisée. Aussi, la localisation des exploitations a été généralisée : seul le pays a été conservé. D'autre part, l'année de naissance, généralisation de la date de naissance, est gardée afin de pouvoir calculer les indicateurs liés à l'âge (travail des enfants).

Une table de correspondance entre les identifiants et les champs personnels a été conservée hors de cette base. En effet, si des enquêtes de terrain collectent de nouvelles données sur une exploitation déjà présente dans la base, il faut pouvoir réutiliser cette exploitation.

3. Contraintes

Les contraintes sont essentielles dans une base de données, car elles permettent d'assurer un certain niveau d'intégrité. Les futurs utilisateurs pourront ainsi se baser sur des données relativement sûres pour prendre des décisions.

a) *Héritage*

Les contraintes liées à la transformation de relation d'héritage ont été appliquées à la base de données. Par exemple, la table « débouché produit » qui renseigne sur le devenir des productions peut être intéressante à étudier.

Parmi les débouchés possibles, on peut trouver par exemple l'autoconsommation, la mise en stock, le don ou la vente. Le fait que le débouché soit une vente implique la collecte d'autres attributs qui sont le type de vente et le prix de vente. On s'attend donc à avoir un modèle logique de données similaire à celui présenté ci-contre.

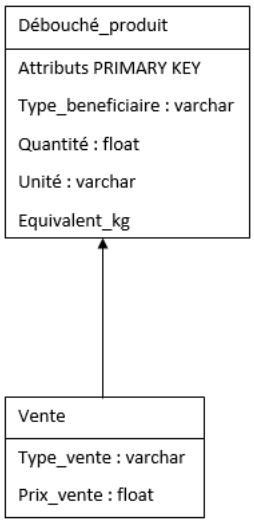


Figure 11 : Modèle logique de l'entité débouché produit

Le passage en relationnel se fait par une transformation par classe mère, car celle-ci n'est pas abstraite, et l'héritage est presque complet. De plus, la classe fille ne possède pas de clé propre. Cela donne la table ci-dessous.

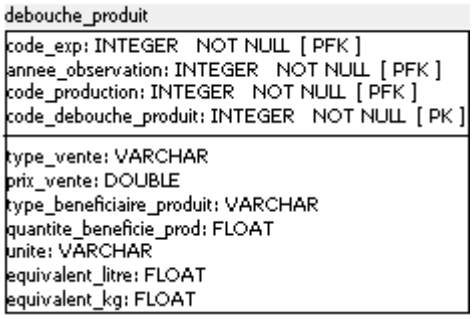


Figure 12 : Table "debouche_produit"

Avec cette modélisation, il faut s'assurer que les champs « prix_vente » et « type_vente » ne sont renseignés que dans le cas d'une vente :

```
ADD CONSTRAINT check_vente CHECK (type_beneficiaire_produit::text = 'vente'::text
                                  OR (prix IS NULL AND type_vente IS NULL));
```

b) *Compositions*

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.

c) *Remarques*

Des contraintes simples ne suffisent pas pour éviter l'ensemble des erreurs et notamment celles de saisie des nombres qui ne peut pas être géré par des listes de valeurs. Ces dernières n'ont pas été implémentées, car les listes sont encore en évolution.

4. Particularités

a) *Unités et conversion*

La question des unités à utiliser pour les quantités s'est révélée problématique. Les agriculteurs utilisent des unités distinctes. Que ce soit pour les consommations intermédiaires ou les quantités issues de la production, il faut pouvoir mesurer les quantités pour les comparer. J'ai donc décidé d'ajouter une variable unité et une variable « équivalent_kg » dans les tables concernées. De plus, une table « conversion_masse », regroupant les conversions des masses en kg, a été créée.

Lors de l'insertion d'une donnée, la colonne « équivalent_kg » est remplie automatiquement grâce à l'utilisation d'un trigger. Un trigger est une opération qui est réalisée sur la base de données avant ou après un évènement donné (insertion, mise à jour ou suppression). Le même système a été utilisé pour les liquides avec une colonne « équivalent_litre ».

Cependant, l'utilisation, par les agriculteurs, d'unités telles que les flacons ou les charrettes, dont les conversions ne sont pas connues, est un problème qui reste à résoudre. Il sera nécessaire d'établir ces conversions.

De même, pour pouvoir comparer les montants d'un pays à un autre, il faut une devise référence. On utilise une stratégie similaire avec une colonne « equivalent_euro » et une table « conversion_devise ». La colonne « equivalent_euro » est remplie automatiquement à partir d'un trigger qui récupère la devise au niveau de l'exploitation concernée. En effet, pour une exploitation donnée, tous les montants sont donnés dans la même devise. Cela permet d'éviter un surplus d'informations. Le code du trigger est présenté en Annexe 3 : Code trigger.

b) *Flexibilité*

Par ailleurs, comme les données qui seront intégrées proviennent de diverses sources, une grande flexibilité de la base est requise. Par exemple, certaines enquêtes enregistrent l'utilisation de consommations intermédiaires à l'échelle de l'exploitation alors que d'autres le font à l'échelle de la parcelle.

Afin de gérer cette granularité, j'ai créé plusieurs variables similaires rattachées à différents niveaux de l'exploitation. Cela permet de capitaliser les données agrégées ou(exclusif) les données unitaires. La redondance des variables ne se traduit pas par une redondance des enregistrements. Par exemple, la surface d'une exploitation peut être obtenue en additionnant la surface de ses terres mais elle peut aussi être renseignée directement au niveau de l'exploitation. Selon les données disponibles, seule une de ces deux possibilités est utilisée.

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 est pris en compte dans 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 tables.

5. Remarques

J'ai eu la chance de travailler sur un projet sur lequel aucune base de données n'avait été réalisée. Créer le modèle entièrement est très stimulant car les choix effectués sont d'une importance capitale. Aussi, il me semble plus simple de créer une nouvelle base de données que s'approprier le travail d'une autre personne.

Par ailleurs, travailler avec SQL Power Architect, qui m'a été conseillé par Mme Auzoux a été particulier. Avec ce logiciel, nous créons graphiquement le modèle logique de données sans passer par le modèle conceptuel de données. Contrairement à ce qui a été appris en cours, les deux étapes sont fusionnées. Cela permet de gagner un temps non négligeable surtout avec une base de données importantes et avec de nombreux changements. Néanmoins, la réflexion doit être plus profonde et l'expérience doit beaucoup aider en ce sens.

En outre, afin de garder une trace de tous les choix effectués, j'ai entrepris de les consigner dans un document intitulé « guide de la base de données ». Il comporte :

- Une définition explicite pour les noms de variable pouvant porter à confusion ;
- Les unités utilisées pour certaines variables m^2 ;
- Les listes de valeurs possibles pour chaque variable ;
- Les contraintes implémentées dans la base de données ;
- Des remarques sur des évolutions possibles

Enfin, j'ai pu ressentir, notamment au cours de formations à distance sur le RGPD, que la partie administrative qui est engendrée par cette législation peut être perçue comme un frein à la recherche.

G. Intégration de données de terrain

Après l'implémentation de la base de données, il est très instructif de l'alimenter avec des données de terrain pour confronter la théorie derrière l'observatoire avec la réalité du monde agricole. De surcroît, cela

permet d'enrichir la base créant ainsi la matière sur laquelle vont s'appuyer les vues représentatives des indicateurs. Les données permettront de valider les requêtes de création de ces vues.

Les deux bases de données fournies par M. Bélières (avec les questionnaires d'enquête respectifs) ont été intégrées. Cette partie du stage, bien qu'essentielle a été fastidieuse et chronophage pour les raisons évoquées dans la partie suivante.

1. Problèmes rencontrés

a) *Problèmes transversaux*

Les bases de données intégrées ne possèdent ni les mêmes tables, ni les mêmes clés primaires, ni les mêmes noms de variables (parfois écrites en malgache). La phase de compréhension de ces bases a été aussi importante que la phase de transposition des enregistrements dans la base de données de l'OA-OI.

Par ailleurs, les bases de données intégrées n'ont pas une structure optimale. En effet, sur les deux bases de données, l'une est constituée d'une seule table et des centaines de colonnes alors que l'autre ne respecte pas certaines contraintes de clés étrangères. Avant d'insérer les données dans la base de données nouvellement conçue, il a donc été nécessaire de les nettoyer.

Ces difficultés et le travail de nettoyage qu'il a été nécessaire de réaliser sont illustrés à travers les deux exemples suivants.

b) *Exemple de la table « Mat_Manuel »*

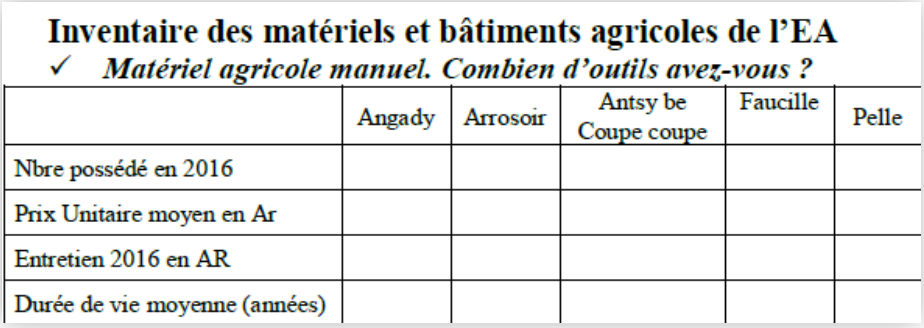
Voici ci-dessous un extrait d'une table récupérée dans une des bases de données intitulée « A06a_Mat_Manuel » pour matériel manuel. La clé primaire est le « Num_Enreg ».

A06a_Mat_Manuel					
Num_EA	Num_Enreg	Nb_Angady	PU_Angady	V_Angady_Entre	DurVie_Angady
101	162	2	15000	1 000,00	2
102	171	3	10000	0,00	1
103	284	2	10000	1 000,00	2
104	280	2	12000	2 000,00	2
105	265	2	12000	1 000,00	1
106	75	4	12000	1 000,00	2
107	188	3	10000	6 000,00	2
108	85	2	11000	2 000,00	1
109	83	3	15000	2 000,00	2
110	295	1	10000	6 000,00	4

Figure 13 : Table matériel manuel d'une des bases de données récupérées

Dans ce cas, plusieurs problèmes sémantiques, syntaxiques et structurels apparaissent. Avant tout, il faut traduire « angady » du Malgache qui signifie « bêche ». Ensuite, les enregistrements ne sont pas datés alors que les équipements possédés varient au cours du temps. Si une nouvelle enquête sonde ces exploitations, il pourrait y avoir des confusions sur les équipements possédés par l'exploitation à un instant donné. Une colonne année est donc ajoutée et renseignée à partir de la date du questionnaire d'enquête.

Par ailleurs, il faut comprendre que PU correspond à prix unitaire. Aussi, la signification de V_Angady_Entre est problématique. Il faut donc de nouveau se référer au questionnaire d'enquête.



Inventaire des matériels et bâtiments agricoles de l'EA

✓ ***Matériel agricole manuel. Combien d'outils avez-vous ?***

	Angady	Arrosoir	Antsy be Coupe coupe	Faucille	Pelle
Nbre possédé en 2016					
Prix Unitaire moyen en Ar					
Entretien 2016 en AR					
Durée de vie moyenne (années)					

Figure 14 : Extrait du questionnaire d'une des bases de données récupérées

Cette variable enregistre la valeur de l'entretien (titre de la 4e ligne du tableau), 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 de l'OA-OI.

c) *Exemple de la table « Crédits »*

Enfin, pour certaines tables, des variables sont stockées sous forme de codification comme sur l'exemple ci-après. Il s'agit de la partie du questionnaire liée à la table « Crédits ». Ces codes sont utilisés pour accélérer les collectes de données. Avec une application Web disposant de valeurs prédéfinies, cette pratique, qui complexifie la compréhension de la base de données, n'est plus justifiée. Le remplacement de chaque code par sa signification (présente dans une table intermédiaire) est requis.

Crédit en 2016 : I___I 0=Non ; 1=Oui ; …Si oui, remplir le tableau suivant

Type de crédit (A)	Origine (B)	Objet (C)	Utilisation réelle (D)	Valeur/Montant emprunté (Ar)	Durée de l'emprunt (mois)	Mode de Remboursement		Observations (préciser la nature du contrat surtout informel)
						Remboursement en (date)	Montant /Valeur remboursée (Ar)	
I__I	I__I	I__I						
I__I	I__I	I__I						
I__I	I__I	I__I						

(A) Type de crédit : (0) Informel, (1) Formel
(B) Origine Crédit : (1) Banque, (2) IMF, (3) Organisme de développement, (4) Organisation paysanne, (5) Commerçant, (6) Autre ménage non famille, (7) Famille, (8) Usuriers, (10) Autres à préciser
(C) Objet du crédit : (1) Crédit de campagne, (2) Crédit d'investissement, (3) Crédit de consommation, (4) Evènement familial, (5) Fournitures scolaires, (10) Autres à préciser
(D) Utilisation réelle : (1) Financement campagne, (2) Investissement agricole, (3) Evènement familial, (4) Fournitures scolaires, (5) PPN non alimentaire (huile, bougie…), (6) PPN alimentaire (paddy, manioc…) (10) Autres à préciser dans colonne

Figure 15 : Extrait d'un questionnaire avec codification

2. Remarques

L'abondance de données dans certains domaines, notamment en informatique, ne doit pas faire oublier que la collecte peut constituer un frein pour d'autres aires de recherche. En outre, je me suis rendu compte du temps nécessaire à l'alimentation et au nettoyage des données de terrain. En agronomie, cette partie du travail peut être bien plus conséquente que l'exploitation des données.

Aussi, afin de récupérer les données, certaines hypothèses ont donc dû être formulées comme cela a été illustré précédemment. Ces dernières sont de potentiels risques qui pèsent sur la validité des données. Il y a donc un équilibre à trouver entre l'intégrité des données et le nombre de données collectées. Une documentation claire éliminerait cette problématique.

Par ailleurs, le nettoyage a engendré de nombreuses pertes de données imputables à des manques de contraintes de clés étrangères. Cela souligne l'importance de la phase de conception de la base de données.

H. Création de vues

Des données brutes sur des exploitations agricoles ne constituent pas une valeur ajoutée pour les parties prenantes du projet. En créant des vues, j'ai combiné ces données pour construire des indicateurs qui sont les variables de sortie de l'observatoire. Ces indicateurs ont été conçus et validés en collaboration avec M. Bosc et Mme Darras. Ils 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
- La production (part de la production vendue, part de la production qui est bio) ;
- Les pratiques culturales (nombre d'associations de cultures, nombre de cultures différentes).

Les capitaux évoqués dans la présentation du concept d'observatoire feront également l'objet de vues.

a) *Exemple de la vue « vab » pour valeur ajoutée brute*

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 obtenues par des agrégations préalables.

Data Output | Explain | Messages | Notifications | Query History

	exploitation integer	annee_observation integer	pb real	cout_total_conso_inter double precision	vab double precision
1	101	2018	1.735403e+06	283309	1452094
2	102	2018	2.156575e+06	144401	2012174
3	103	2018	2.413946e+06	434754	1979192
4	104	2018	2.357689e+06	129654	2228035
5	105	2018	1.49884e+06	[null]	[null]
6	106	2018	3.209435e+06	278762.5	2930672.5
7	107	2018	1.47342e+06	158408	1315012
8	108	2018	1.674152e+06	314650	1359502
9	109	2018	1.043852e+06	433076	610776
10	110	2018	247654	135877	111777

Figure 16 : Vue valeur ajoutée brute

La fonction SQL COALESCE(liste) permet de renvoyer la première valeur non nulle de la liste qui lui est transmise. Nous aurions ainsi pu avoir :

```
(pb_exploitation.pb - coalesce(total_cout_conso_inter.cout_total_conso_inter,0) AS vab
```

Cela aurait permis d'avoir une valeur pour la ligne 5. Seulement, donner une valeur ajoutée brute pour une exploitation dont nous ne connaissons pas le coût des consommations intermédiaires n'a pas de sens. En effet, si le cout des consommations intermédiaires n'est pas renseigné, il n'est vraisemblablement pas null, mais n'a probablement 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.

b) *Exemple de la vue « total_travail_par_sexe »*

Par ailleurs, certaines vues, notamment celles concernant le travail des individus, montrent des répartitions selon l'âge et le sexe (voir la vue ci-dessous). Pour illustrer le manque de connaissance que nous avons, les individus aux âges ou aux sexes non renseignés sont comptés en tant que « non renseigné ». Cela montre le travail de collecte de données qu'il reste à faire pour enrichir la base de données. En outre, cela permet de quantifier notre ignorance et donc de ne pas tirer de conclusions hâtives sur les données. Ce type de vue nécessite l'utilisation de sous-requêtes qui ont été particulièrement utiles. Le code de la vue est donné en Annexe 2 : Code de la vue ""Travail_par_sexe". Ce dernier donne le résultat suivant :

	code_exp integer	annee_observation... integer	total_travail_homme double precision	total_travail_femme double precision	total_travail_non_renseigne... real
1	101	2018	2.5	1.5	2.3
2	102	2018	3	2.25	0.59000003
3	103	2018	1	1	1.5
4	104	2018	1.5	2.25	4.89
5	105	2018	2.5	1	1.0799999
6	106	2018	2	1	6.1000004

Figure 17 : Vue "Travail_par_sexe"

I. Réalisation du prototype de l'application Web

La création de l'application Web représentera l'aboutissement du concept d'Observatoire des Agricultures de l'Océan Indien. Il s'agit, pour l'instant, d'une POC. Cela permet d'avoir un premier aperçu de ce que pourrait être l'observatoire et proposer des idées de fonctionnalités et d'architecture. D'après la première ébauche de cahier des charges, le développement de l'application Web a été initié.

1. Architecture MVC avec CodeIgniter

L'application Web utilise une architecture MVC (Modèle-Vue-Contrôleur) sur laquelle se base CodeIgniter. 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, en général, un fichier pour la vue, un fichier pour le modèle et un fichier pour le contrôleur. Chaque page a une utilité bien précise comme cela est illustré ci-dessous.

Le Modèle :

- Le Modèle regroupe toutes les fonctions liées aux interactions directes à la base de données.
- CodeIgniter facilite les interactions avec la base de données à travers des classes implémentées à la création du projet.

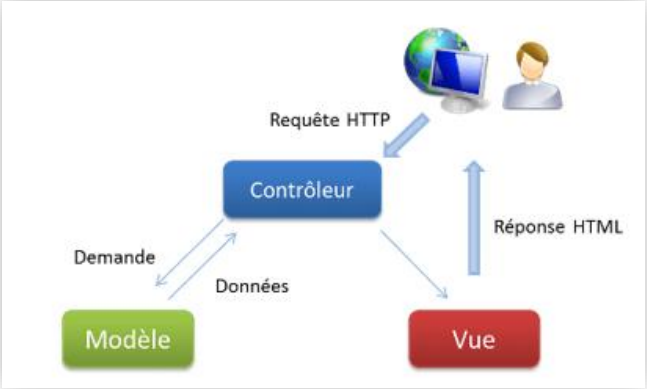


Figure 18 : Schéma architecture MVC

La Vue :

- La vue est la partie du code qui produit le document html.
- Dans CodeIgniter, il est possible de passer des variables du contrôleur à la vue au moment de l'appel à la vue.

Le Contrôleur :

- Le contrôleur est 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 nous instancions les modèles et que nous renvoyons les vues.

2. Avancement de la preuve de concept de l'application Web

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 19 : Page de présentation de l'observatoire

a) *Connexion et profils*

Il est également possible de s'enregistrer, de se connecter et de déconnecter. 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.

b) *Annuaire*

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

Figure 20 : Recherche dans l'annuaire

c) *API*

Afin de partager les connaissances accumulées avec tous les acteurs, l'observatoire doit posséder une API. Cela permet la coopération ainsi que la réutilisation des données.

CodeIgniter offre la possibilité de créer une API rapidement. J'ai créé une vue qui regroupe certains indicateurs classiques sur les exploitations (notamment la surface, le capital financier, ou encore le produit brut). L'API permet d'obtenir ces variables agrégées pour toutes les exploitations dans un format JSON. Il est également possible d'accéder uniquement aux indicateurs d'une exploitation en renseignant son identifiant.

J. Data visualisations

Les vues précédemment réalisées contiennent des indicateurs quantitatifs. La production de « data visualisations » va permettre de faciliter la lecture et la compréhension de ces données chiffrées en les présentant sous forme de graphique, d'images ou de cartes. Elle est donc indispensable pour l'OA-OI qui se veut être un outil accessible à tous et orienté vers la prise de décision. Ces visualisations ont donc vocation à être intégrées à l'interface Web. De nombreuses données étant manquantes, j'ai créé une exploitation fictive pour laquelle j'ai rempli les champs les plus significatifs. Voici quelques exemples de Data visualisations qui ont été créées sur les données contenues dans la base.

1. Composition

Les compositions sont le premier type de data visualisation intéressant à utiliser pour l'OA-OI. En effet, grâce à des compositions, on peut, par exemple, représenter la répartition du travail selon l'âge le sexe ou l'origine de la main d'œuvre.

Dans ce but, j'ai produit la data visualisation ci-dessous qui représente la composition des revenus de l'exploitation fictive évoquée précédemment. La partie de gauche est le visuel à l'origine. Un clic provoque un zoom et l'apparition des légendes (partie droite).

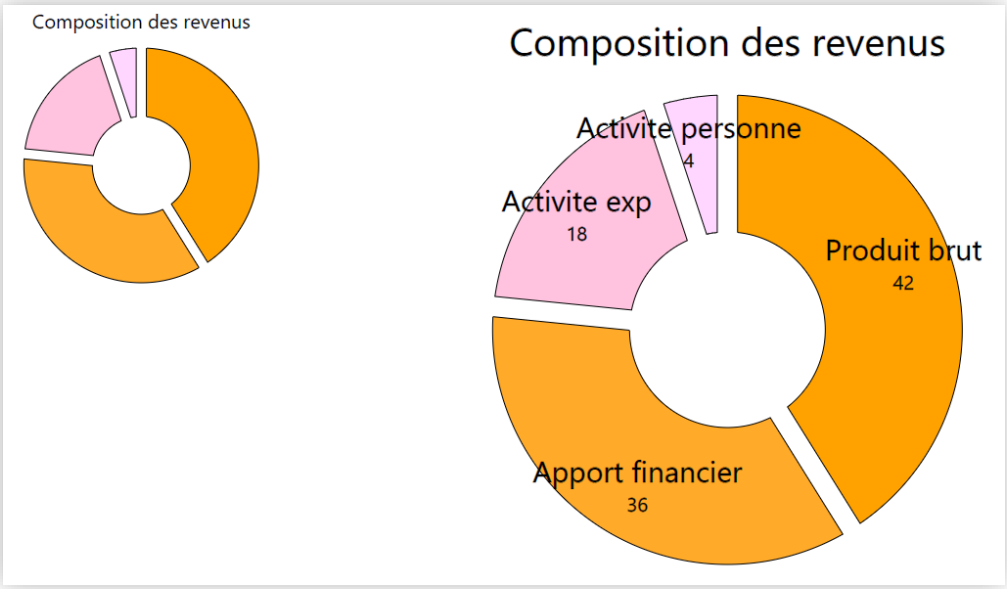


Figure 21 : Data visualisation sur la composition des revenus

2. Carte chloroplèthe

L'observatoire devant s'implanter dans plusieurs territoires, observer des tendances territoriales apparaît profitable aussi bien pour le grand public que pour les parties prenantes de l'observatoire. J'ai donc créé une carte chloroplète qui permet de colorer les pays en fonction d'une variable donnée (le nombre d'exploitations dans l'observatoire dans ce cas).

Maurice, la Réunion ou les Seychelles étant des petits pays en terme de superficie, ils sont peu visibles. Pour résoudre ce problème, ces pays ont été identifiés à partir d'un seuil de surface et zoomés jusqu'à atteindre une taille raisonnable.

Par ailleurs, une simple carte ne permet pas d'accéder à un grand nombre d'informations. Afin de visualiser plus de données, deux méthodes ont été utilisées. D'abord, un tooltip est ajouté et apparaît lors du passage de la souris. Deuxièmement, il est possible de cliquer sur les pays de la COI ce qui donne le rendu ci-dessous, présentant un plus grand nombre d'informations.

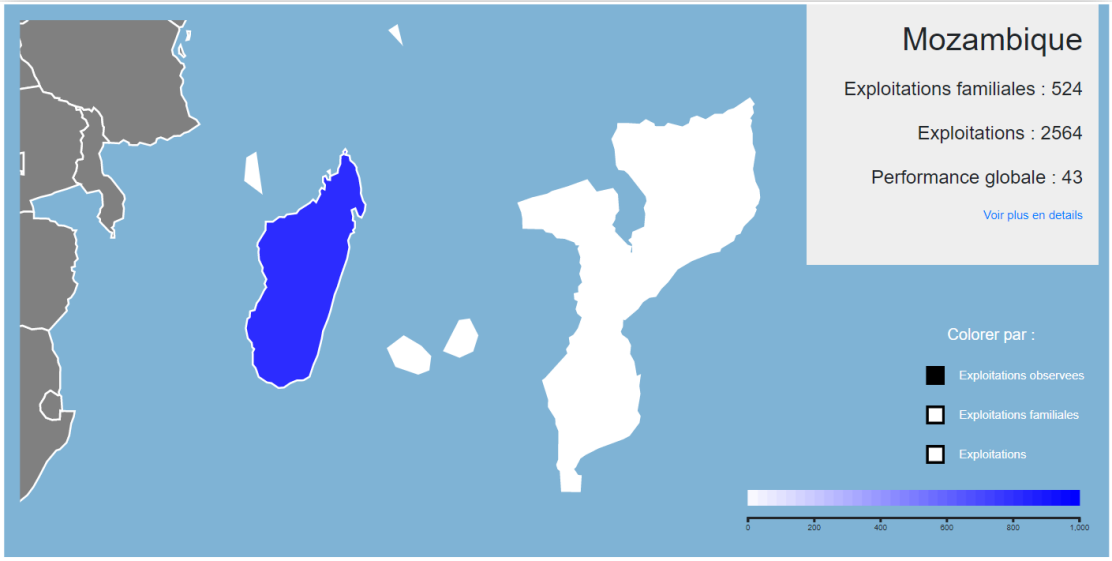


Figure 22 : Data visualisation sur l'Afrique

Pour réaliser cette carte, il est nécessaire de s'appuyer sur des fichiers JSON et plus particulièrement des fichiers geojson qui sont gérées par la bibliothèque d3js. Ces derniers sont adaptés pour les données géographiques et associer des données à des territoires (cf. annexe). Je n'ai pas réussi à trouver un fichier de bonne qualité regroupant l'Afrique et les îles concernées par l'observatoire. C'est un point d'amélioration possible.

3. Comparaison multicritères

Ensuite, afin de comparer des exploitations entre elles ou par rapport à une moyenne, une comparaison multicritère peut être intéressante. C'est dans ce but qu'un diagramme radar a été réalisé. Il permet d'avoir un aperçu rapide des performances ou des caractéristiques des exploitations sur un grand nombre de critères.

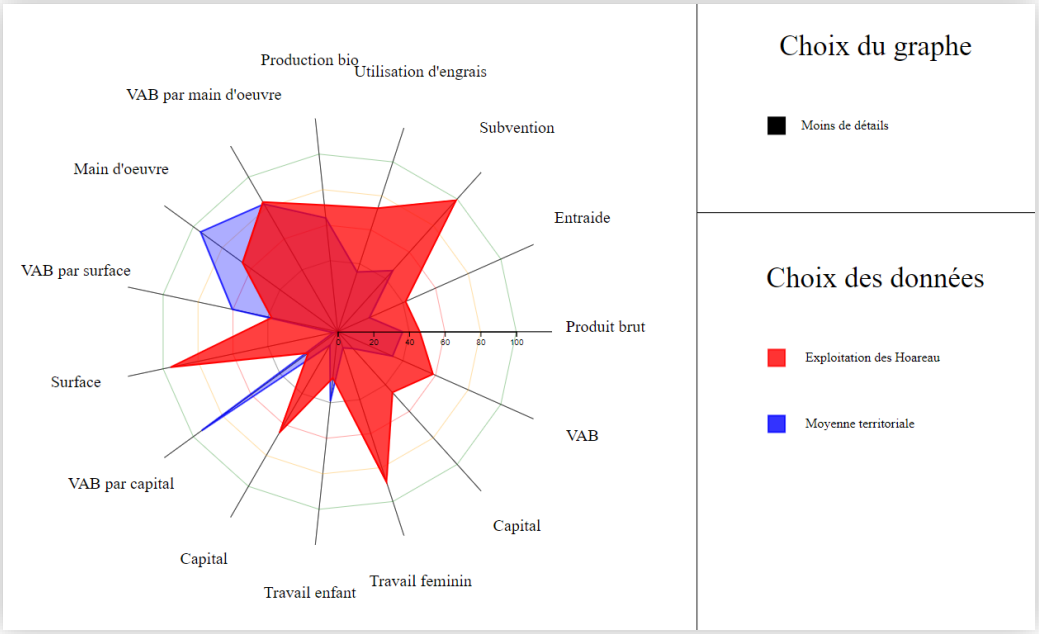


Figure 23 : Radar Chart

4. Comparaison et évolution temporelle

Enfin, les indicateurs générés dans le cadre de l'observatoire varieront au cours du temps. L'étude de ces dynamiques aide à la compréhension des exploitations agricoles. Afin de répondre à ce besoin, j'ai créé un diagramme d'aires, présenté ci-après.

Ce type de data visualisation permet de représenter une variable donnée au cours du temps. Plusieurs jeux de données peuvent y être inclus et ceux qui doivent apparaître sont sélectionnés à partir des carrés à droite du diagramme. De plus, le passage de la souris sur une des aires permet de se focaliser sur un seul jeu de données en faisant apparaître les indicateurs chiffrés et en adaptant l'opacité des différentes aires.

La sélection de la variable représentée peut être une des améliorations à envisager pour cette data visualisation.

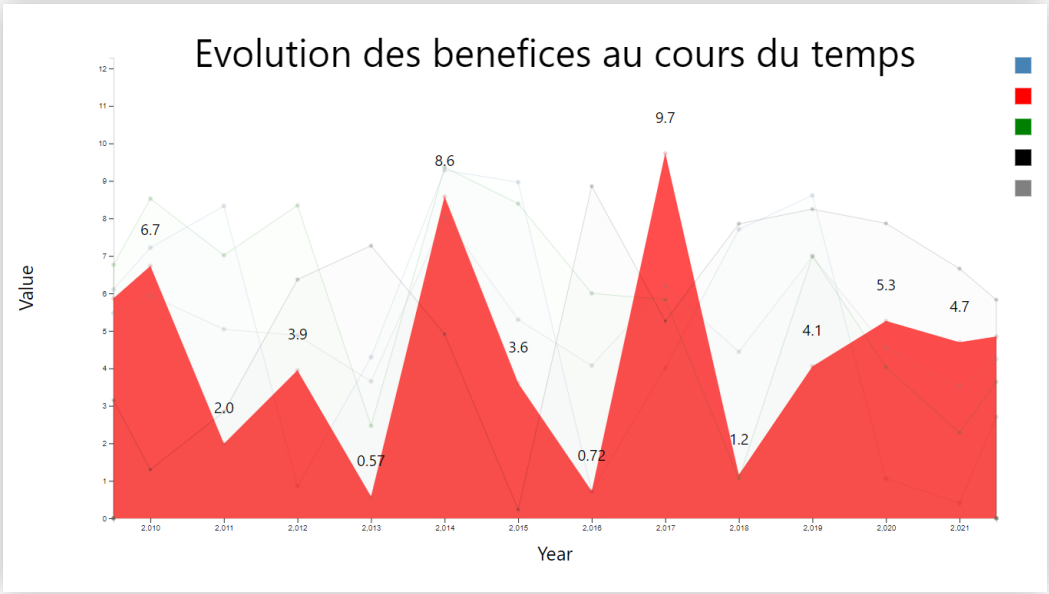


Figure 24 : Area Chart

5. Caractéristiques communes

Toutes ces data visualisations sont animées pour être plus agréables pour les utilisateurs. Les évènements de « clique », « mouseover » (arrivée de la souris sur un élément), « mouseout » (sortie de la souris d'un élément) ont été exploités pour donner vie aux data visualisation. Par ailleurs, étant donné qu'elles apparaîtront sur l'application Web, ces dernières ont été rendues responsives grâce à l'attribut « viewbox ».

Par ailleurs, un grand travail a été effectué sur la généricité. En effet, ces diagrammes pouvant avoir de nombreux usages, certains paramètres peuvent varier. L'utilisation de fonctions prenant en compte certaines caractéristiques des jeux de données permet de créer dynamiquement les visualisations. Par exemple, sur le diagramme radar, le nombre d'axes varie automatiquement en fonction de la taille des jeux de données en paramètres. Etant donné l'ensemble des sorties envisageables, cette généricité est particulièrement pertinente et peut engendrer un gain de temps considérable.

K. Analyse de données

Le but de l'observatoire étant de favoriser l'action politique, il m'a paru essentiel d'estimer l'impact des mesures sur les exploitations. J'ai donc décidé, en accord avec ma tutrice, d'ajouter une partie de Machine learning au stage. J'ai suivi des enseignements sur le site kaggle pour me former dans ce domaine.

1. Cadre de l'étude

Nous pouvons considérer le « produit brut » comme la variable d'intérêt et qui peut donc être intéressante à prédire. Pour établir des prédictions, il faut d'abord trouver les variables qui déterminent cette variable d'intérêt. Le capital (somme des valeurs des équipements), la surface, la main d'œuvre, les apports financiers ainsi que les consommations intermédiaires sont retenus. Ces données ont été récupérées dans la vue utilisée pour l'API et transformées en data frame grâce à la librairie Pandas.

J'ai utilisé le modèle RandomForest qui est un ensemble d'arbres de régression. Ce modèle ne s'applique qu'à des ensembles de valeurs non null. Nous utilisons donc une fonction permettant de remplacer les valeurs null par 0. Nous divisons ensuite l'échantillon en 2 parties :

- La première partie sert à entraîner le modèle avec les données ;
- La seconde sert à tester la précision le modèle sur des données qu'il n'a jamais rencontrées. Nous comparons ainsi les prédictions réalisées par le modèle avec les vraies valeurs. Plusieurs métriques permettent de mesurer la performance du modèle. J'utilise la MAE (mean absolute error) qui est l'écart moyen entre la valeur prédite et la valeur réelle.

2. Prédictions

Les acteurs publics peuvent principalement aider les exploitations via des subventions. Cela se traduit par une augmentation du capital. Nous pouvons alors tenter de prédire quelle sera l'évolution du produit brut en fonction du montant des subventions.

Pour l'ensemble des exploitations, nous prédisons le produit brut et nous calculons le pourcentage de variation entre ces prédictions avec les valeurs réelles (j'appelle ce gain, le gain_par_défaut).

Ensuite, les valeurs du capital sont augmentées de x. Une nouvelle prédiction sur le produit brut est réalisée. De la même manière que précédemment, nous calculons le pourcentage de variation entre ces prédictions et les valeurs réelles (je l'appelle le gain_estimé). Pour ajuster ce gain avec le gain par défaut, nous calculons simplement :

gain_ajusté = gain_estimé – gain_par_défaut

De cette manière, nous éliminons une partie de l'erreur générée par le modèle. Ce gain ajusté, qui est en pourcentage, est représenté sur le graphique ci-après, en fonction du montant de subventions.

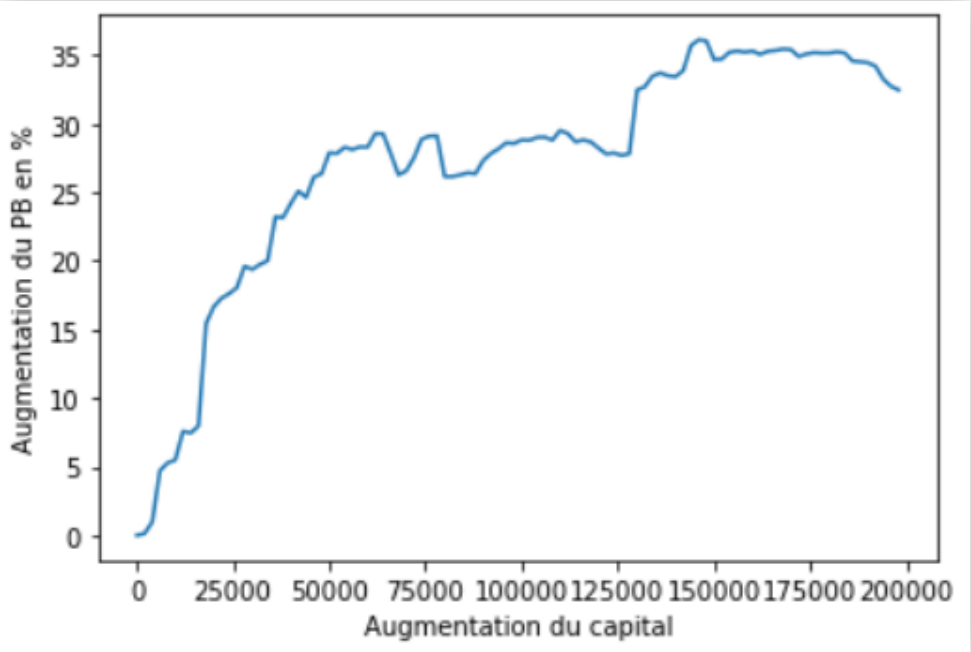


Figure 25 : Evolution du produit brut en fonction de subventions

Nous observons que des subventions d'environ 50 000 euros entraîneraient une augmentation de 25% du produit brut. De plus, à partir de 130 000 euros de subventions, le produit brut n'augmente plus.

3. Critique du modèle

Cet outil peut être très intéressant. Néanmoins, cette estimation reste très sommaire. En effet, la MAE tourne autour de 50% du produit brut avec ces données. De plus, on observe sur la courbe que des subventions de 50 000 euros sont aussi impactantes que des subventions de 125000 euros. Le produit brut augmente de nouveau après ce seuil. Cela ne semble pas cohérent. Il faudrait grandement améliorer cette précision pour donner confiance aux investisseurs.

4. Pistes d'amélioration de la performance du modèle

Pour remédier à la performance médiocre du modèle, plusieurs pistes sont à envisager :

- L'augmentation du nombre d'exploitations dans la base ;
- La collecte des variables non renseignées pour les exploitations déjà observées ;
- L'essai de stratégies de traitement différentes des valeurs null ;
- L'optimisation des paramètres du modèle (notamment le nombre d'enregistrements minimum dans les feuilles des arbres de régression afin d'éviter l'overfitting ou l'underfitting) ;

- Le choix d'autres variables utilisées pour les prédictions. Ces dernières doivent néanmoins être indépendantes entre elles ;
- Le choix d'un autre modèle.

5. Perspectives

Cette utilisation du Machine learning n'est qu'un premier pas dans cette direction et représente un potentiel à exploiter pour l'observatoire. De multiples applications peuvent être envisagées. L'OA-OI étant axé sur la différenciation des exploitations, il serait intéressant de réaliser des prédictions en fonction de certaines de leurs caractéristiques.

D'autre part, il serait avantageux prédire l'évolution d'autres indicateurs des exploitations. Par exemple, les performances environnementales ou le salaire des employés pourraient être explorés.

VI. Bilan d'expérience

A. Le projet

Tout d'abord, j'espère amplement que le projet de l'observatoire aboutira, quelle que soit sa forme finale. En travaillant étroitement avec les partenaires, je suis persuadé qu'il peut avoir un impact positif pour ces derniers. Je suis curieux d'étudier les changements qui seront survenus entre mon stage et le lancement de l'observatoire.

Le stage s'étalant sur simplement 24 semaines, transmettre le travail effectué et faciliter sa réutilisation est d'une importance capitale pour la poursuite du projet. Dans cette optique, une documentation fournie et détaillée a été rédigée. Le guide de la base de données, le dictionnaire généré par SQL Power Architect, l'ébauche de cahier des charges ainsi que les commentaires sur les codes devraient permettre cette réutilisation qui est primordiale. De cette manière, la base de données ainsi que les data visualisations devraient être réutilisées et c'est une réelle satisfaction de permettre au projet d'avancer. Un stagiaire sera recruté en fin 2021 pour poursuivre ces travaux.

D'autre part, un diaporama de présentation de la base de données et de son champ d'application a été conçu en collaboration avec Mme Darras. Ce dernier a pour but de présenter le travail effectué sur l'observatoire à des personnes extérieures au projet. Aussi, de nombreuses personnes ont souhaité avoir un aperçu de l'avancement du projet. J'ai donc effectué une présentation de mon travail aux personnes du CIRAD, différente de la soutenance UTC.

Enfin, j'ai eu la chance d'expérimenter mes propres choix notamment lors de la création des data visualisations et de l'application Web. Être force de proposition est intéressant, car cela permet souvent de stimuler les idées des autres membres de l'équipe. Cela donne aussi un encrage à partir duquel construire de nouvelles réflexions et imaginer de possibles améliorations.

B. 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.

Premièrement, il est apparu que coopérer avec des acteurs dont les domaines d'expertise sont très divers n'est pas toujours évident. Agronomes et informaticiens utilisent parfois des termes similaires pour désigner des concepts très différents. Par ailleurs, tous les acteurs impliqués dans le projet n'étaient pas présents sur place et la communication a été un enjeu clé du stage.

D'autre part, le projet de l'observatoire réunit plusieurs acteurs qui ont tous des intérêts distincts. Entre les besoins des producteurs, les demandes des financeurs, les questions du CIRAD et la coopération avec la DAAF, certains jeux politiques influent sur la conduite de projet. Ces contraintes extérieures font partie du travail d'ingénieur.

Aussi, l'apprentissage constant de nouvelles technologies ne constitue pas une difficulté en soi mais la veille technologique représente un temps de travail considérable. Ce travail est plus important en période

de stage mais les technologies évoluent à une vitesse telle que cette nécessité perdure tout au long d'une carrière.

Enfin, étant d'ordinaire très attaché à l'efficacité de mon travail, il a été difficile de prendre part à un travail de recherche où les modifications sont nombreuses et impactent grandement la temporalité du projet.

C. Apports personnels

D'un point de vue 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 surcroît, la portée relativement large de mes missions a aussi été l'opportunité de faire preuve d'esprit critique.

J'ai également pu découvrir durant la première moitié du stage le travail d'assistant à maîtrise d'ouvrage. Concilier des objectifs ambitieux avec des faisabilités techniques et des données éparses s'est révélé très difficile donc formateur. Il a été enrichissant d'observer l'importance de la communication au sein d'une équipe pluridisciplinaire. En effet, les interprétations des enjeux de l'observatoire sont partiellement différentes pour chacun.

De plus, ce stage m'a permis de découvrir un très grand nombre de technologies. En effet, étant arrivé dans le domaine informatique tardivement, j'avais des lacunes autant techniques que théoriques. Cette période m'a permis de me sentir légitime pour l'obtention du diplôme d'ingénieur informatique.

Aussi, j'ai pris conscience de l'intérêt du mode d'enseignement à l'UTC qui est davantage axé sur la capacité d'apprentissage que sur la maîtrise des technologies qui évoluent très rapidement. Ces technologies ne constituent pas une finalité, mais un outil permettant de mettre en œuvre une solution. Le cœur du métier d'ingénieur réside dans l'élaboration de cette solution qui nécessite de grandes capacités d'écoute, de communication et de compréhension.

Durant ce stage, j'ai aussi pris davantage conscience du rôle de chef de projet et de son importance dans la réussite du projet. Je suis impatient de réaliser le TN10 (stage ingénieur) pour approfondir mes savoir-faire en gestion de projet et avoir plus de poids dans les décisions. En effet, créer un cadre de travail efficace et épanouissant tout en exploitant les connaissances variées de chacun me paraît passionnant.

En outre, durant cette période de 24 semaines dont une grande partie à rechercher, disséquer, analyser et examiner les données, j'ai pris conscience de la valeur et de l'utilité qu'elles peuvent avoir. Aussi, l'extraction d'informations significatives et l'utilisation de Machine learning m'ont beaucoup plu de par leur application très concrète. Etant passionné par le sport et ses statistiques, j'ai décidé de m'orienter vers la data science appliquée au sport. J'ai 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 cet institut.

Enfin, avoir l'opportunité de vivre plusieurs mois à la Réunion a été une réelle chance surtout dans le contexte actuel. Le climat tropical y a forgé des paysages très divers mais tous magnifiques. Mon envie de vivre sur cette île, en compagnie de ses habitants, n'est freinée que par mon goût de la découverte.

VII. Glossaire

3NF : 3e Forme Normale

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

API : Application Programming Interface. Dans le cas de l'observatoire, il s'agit d'une interface grâce à laquelle l'application Web peut offrir des services à d'autres sites Web.

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)

FCR : Fond de Coopération Régionale

LSMS : Living standard Measurement Study

MAE : Mean Absolute Error (erreur absolue moyenne)

MCD : Modèle Conceptuel de Données

MVC : Modèle Vue Contrôleur

ODD : Objectifs du développement durable

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

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

POC : Proof of concept. Il s'agit d'un prototype permettant de valider ou non un concept. Il peut notamment s'agir d'une interface Web.

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

RGPD : Règlement Général sur la Protection des Données

SGBD(R) : Système de Gestion de Base de Données (Relationnelle)

SVG : Scalable Vector Graphics (graphique vectoriel adaptable)

VSC : Volontaire Service Civique

VIII. Bibliographie

Darras Adèle, Bélières Jean-François, Bosc Pierre-Marie, Auzoux Sandrine, Le Moine L, Mialet-Serra Isabelle. 2021. Variables et indicateurs du cadre harmonisé de l'Observatoire des Agricultures du Monde (OAM) : Définitions et descriptions à l'échelle de l'exploitation agricole et du ménage. s.l. : CIRAD, 77 p. Référence https://agritrop.cirad.fr/597467/

Kaggle, https://www.kaggle.com/

Agritrop, Référence https://agritrop.cirad.fr/

CIRAD, Référence de https://www.cirad.fr

AIDA, Référence de https://ur-aida.cirad.fr/

CodeIgniter4, Référence de https://codeigniter.com/user_guide/intro/index.html

PostgreSQL version 13.1. Référence de https://www.postgresql.org/

PostGIS version 3.1.0, Référence de https://postgis.net/

Bootstrap version 4.5, Référence de https://getbootstrap.com/

D3js version 6.2.0 Référence de https://d3js.org/

SQL Power Architect Référence de http://www.bestofbi.com/page/architect

Prérad-OI, Référence https://www.prerad-oi.org/

FAO, Référence http://www.fao.org/home/fr/

DAAF, Référence https://daaf.reunion.agriculture.gouv.fr/

PhpStorm, version 2020.3 Référence https://www.jetbrains.com/fr-fr/phpstorm/?rss

Gantt Project, version 2.8.11 Référence https://www.ganttproject.biz/

Python, version 3.8.1 Référence https://www.python.org/

Scikitlearn, version 0.24 Référence https://scikit-learn.org/

RGPD, Référence https://www.cnil.fr/fr/comprendre-le-rgpd

DRRM, Référence https://reunion-mayotte.cirad.fr/

LSMS, Référence https://www.worldbank.org/en/programs/lsms

COI, Référence https://www.commissionoceanindien.org/

Dataverse, Référence https://dataverse.org/

Wikipédia, Référence https://fr.wikipedia.org/

VSC, Référence https://www.service-civique.gouv.fr/

Annexes

Annexe 1 : Organigramme

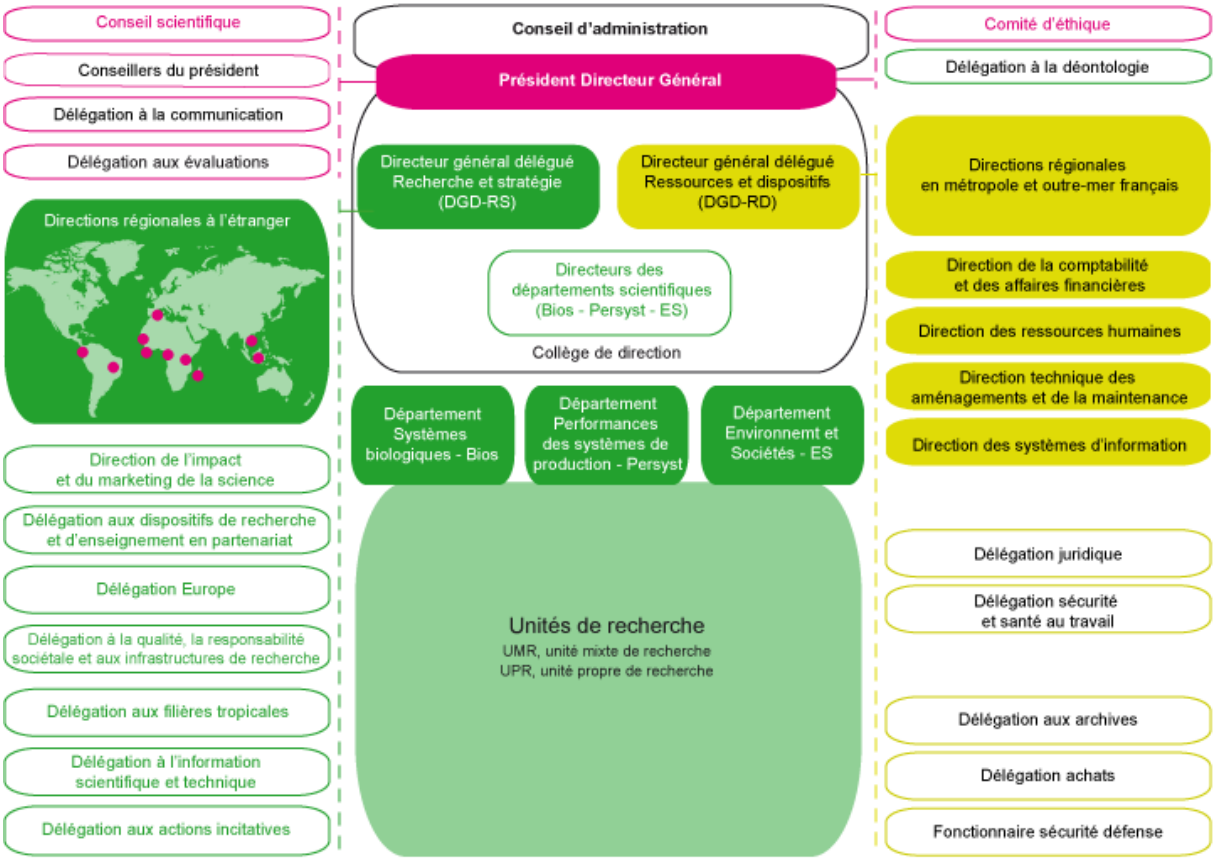


Figure 26 : Organigramme du CIRAD

Annexe 2 : Code de la vue ""Travail_par_sexe"

```
CREATE OR REPLACE VIEW public.total_travail_par_sexe
 AS
 SELECT main_oeuvre.code_exp,
    main_oeuvre.annee_observation,
    (sum(m.tps_homme))::double precision AS total_travail_homme,
    (sum(f.tps_femme))::double precision AS total_travail_femme,
    sum(nr.tps_nr) AS total_travail_non_renseigne
   FROM ((((main_oeuvre
     JOIN individu ON ((individu.id_individu = main_oeuvre.id_travailleur)))
     left JOIN ( SELECT individu2_1.id_individu AS id_femme,
            main_oeuvre_1.tps_travail_an AS tps_femme
           FROM (individu individu2_1
           JOIN main_oeuvre main_oeuvre_1
                 ON ((individu2_1.id_individu = main_oeuvre_1.id_travailleur)))
          WHERE (individu2_1.sexe = false)) f
           ON ((main_oeuvre.id_travailleur = f.id_femme)))
     left JOIN ( SELECT individu2_1.id_individu AS id_nr,
            main_oeuvre_1.tps_travail_an AS tps_nr
           FROM (individu individu2_1
           JOIN main_oeuvre main_oeuvre_1
                 ON ((individu2_1.id_individu = main_oeuvre_1.id_travailleur)))
          WHERE (individu2_1.sexe IS NULL)) nr
          ON ((main_oeuvre.id_travailleur = nr.id_nr)))
     left JOIN ( SELECT individu2_1.id_individu AS id_homme,
            main_oeuvre_1.tps_travail_an AS tps_homme
           FROM (individu individu2_1
           JOIN main_oeuvre main_oeuvre_1
                 ON ((individu2_1.id_individu = main_oeuvre_1.id_travailleur)))
          WHERE (individu2_1.sexe = true)) m
         ON ((main_oeuvre.id_travailleur = m.id_homme)))
  GROUP BY main_oeuvre.code_exp, main_oeuvre.annee_observation
  ORDER BY main_oeuvre.code_exp;
```

Annexe 3 : Code trigger

```
CREATE FUNCTION public.conversion_devise()
    RETURNS trigger
    LANGUAGE 'plpgsql'
    COST 100
    VOLATILE NOT LEAKPROOF
AS $BODY$
DECLARE
equivalent float;
coef float;
BEGIN
if exists(
    select *
    from conversion_devise
    join
    (select code_exp, devise from exploitation10 where exploitation10.code_exp=new.code_exp)as a
on a.devise=conversion_devise.devise)
then
    coef=(select equivalent_euro from conversion_devise join
    (select code_exp, devise from exploitation10 where exploitation10.code_exp=new.code_exp)as b
    on b.devise=conversion_devise.devise);
    new.equivalent_euro=coef*new.prix;
    return new;

else return new;
    end if;
END;
$BODY$;
```

Figure 27 : Trigger conversion_devise

Annexe 4 : Fichier geojson

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


```
{"type":
        "Feature",
                "properties":
                        {"name":"Seychelles",
                        "cartodb_id":46,
                        "exp_f» :524,"exp » :2564,
                        "created_at » :"2013-11-12T16 :15 :59+0100",
                        "updated_at » :"2013-11-12T16 :15 :59+0100"
                        },
                "geometry » :
                        {"type » :
                                "MultiPolygon","coordinates » :
                                [[[[55.532547,-4.789494],
                                [55.37526,-4.622715],
                                [55.475598,-4.558986],
                                [55.532547,-4.789494]
                                ]]]
                        }
                }
}
```