



RAPPORT DE STAGE

Conception et implémentation d'une base de données et d'un outil permettant l'accès à des données issues d'échantillons biologiques de patients atteints d'hémopathie maligne

Réalisé par :
Mathieu BLANCHET

Encadré par :
Elina ALATERRE
Angélique BRUYER

Année universitaire 2022-2023

Remerciement

Je tiens à remercier l'équipe de Diag2Tec pour m'avoir accueilli et intégré au sein de son entreprise.

Je remercie mon tuteur Elina Alaterre pour m'avoir accompagné et aidé tout au long du stage.

Je souhaite également remercier Petru Valicov pour avoir tutoré ce stage.

Résumé

Dans le cadre d'un stage de douze semaines au sein de la société Diag2tec, j'ai eu pour mission de réaliser une base de données afin d'améliorer le suivi et le traitement de données concernant des caractéristiques biologiques d'échantillons de patients atteints de cancers hématologiques qui étaient jusqu'à présent stockées par l'entreprise sous la forme de multiple fichiers Excel. De plus, pour permettre à l'ensemble des collaborateurs de Diag2Tec d'utiliser la base de données, il m'a été demandé de créer une interface web permettant d'interagir avec cette base de données pour rechercher des échantillons d'intérêt selon des caractéristiques biologiques spécifiques, importer de nouvelles données, exporter des données d'intérêt et gérer l'évolution de la base de données.

Mot clés : base de données, interface web, stockage, import, export, données biologiques, cancers hématologiques

As part of a twelve week internship with Diag2tec, I was tasked to create a database to improve the monitoring and the processing of data concerning the characteristics of samples of patients with hematological malignancies. Until now, these data were stored into Excel files by the company. Also, to allow all Diag2Tec employees to use the database, I was asked to create a web interface to interact with this database to search samples of interest according to specific biological characteristics, import new data, export data of interest and manage the evolution of the database.

Key words : database, web interface, storage, import, export, biological data, hematological malignancies

Sommaire

Introduction	6
1 - Présentation de l'entreprise	7
1.1 - Le Centre Hospitalier Universitaire (CHU)	7
1.2 - Le Centre National de la Recherche Scientifique (CNRS)	7
1.3 - L'entreprise Diag2Tec	8
2 - Analyse	9
2.1 - Présentation du sujet et analyse du contexte	9
2.2 - Analyse de l'environnement technique	9
2.3 - Analyse de l'existant	10
2.4 - Cahier des charges	10
3 - Spécifications techniques	12
3.1 - Conception de la base de données	12
3.2 - Conception du site web	14
3.2.1 – Les objectifs	14
3.2.2 – L'architecture de l'application	15
3.2.3 – Les fonctionnalités d'ajout	16
3.2.4 – La fonctionnalité de recherche	17
3.2.4 – La fonctionnalité de mise à jour	18
3.3 – Dynamisation du site web et de la base de données	19
3.4 – Contrôle qualité et test de la solution	22
4 – Méthodologie et organisation du projet	24
4.1 – Organisation du travail au sein de l'entreprise	24
4.2 – Organisation du projet	25
4.3 – Méthodologies et gestion du projet	25
Conclusions et perspectives	26
Annexes	28
Annexe n°1 : Cahier des charges	29

Table des figures

Figure 1 : Modèle entité-association	13
Figure 2 : Schéma MLD	14
Figure 3 : Diagramme de classe de l'architecture PHP	15
Figure 4 : Schéma simplifier des échanges entre les différentes technologies	16
Figure 5 : Visuel de la page d'importation	17
Figure 6 : Visuel de la page de recherche	18
Figure 7 : Exemple de résultat obtenu après une recherche	18
Figure 8 : Visuel de la page de mise à jour des données	19
Figure 9 : Table « Colonne »	20
Figure 10 : Formulaire permettant l'ajout d'une colonne	20
Figure 11 : Formulaire d'édition d'une colonne	21
Figure 12 : Cas-Utilisation application web	22

Glossaire

Myélome Multiple : Le myélome multiple est un cancer de la moelle osseuse, dû à l'accumulation anormale d'un type de globules blancs, les plasmocytes, responsables de la production et la sécrétion des anticorps.

Lignée cellulaire : Ensemble de cellules issues d'une cellule originelle par mitose successive qui est rendu immortel.

Mitose successive : Processus de division d'une cellule 'mère' résultant en cellules "fille" possédant les mêmes caractéristiques que la cellule 'mère'.

Hémopathie maligne : Maladie du sang.

Cellule primaire : Cellule qui a été isolée directement à partir d'un tissu ou d'un organe vivant d'un patient.

Epigénétique : Discipline de la biologie qui étudie les mécanismes modifiant de manière réversible, transmissible et adaptative l'expression des gènes sans en changer la séquence ADN.

Cytotoxique : Substance ou partie d'une cellule spécialisée capable de détruire des cellules. Ce qui est cytotoxique est toxique pour les cellules vivantes.

Biomarqueur : Caractéristique biologique mesurable utilisée comme indicateur d'une fonction du corps, d'une maladie ou de l'action d'un médicament.

Lymphomes : Tumeur maligne prenant naissance dans les organes lymphoïdes secondaires (lieux de passage d'accumulation et de rencontre des antigènes et des cellules immunitaires impliquant des lymphocytes T et B à différents stades).

Séquençage : Méthode permettant de déterminer la séquence d'ADN du génome humain.

Introduction

Dans le cadre du quatrième semestre de mon BUT Informatique, j'ai l'occasion de réaliser un stage de douze semaines au sein de l'entreprise Diag2Tec, une start-up localisée au sein du département d'hématologie-biologique du CHU de Montpellier qui propose des services précliniques consacrés aux **hémopathies malignes***, et particulièrement au **myélome multiple***. L'entreprise se focalise notamment sur des tests *in vitro* de molécules anti-cancéreuses innovantes dans deux types de modèles : les **cellules primaires*** et les **lignées cellulaires*** issues de **patients atteints d'hémopathies malignes***.

Néanmoins, cette start-up éprouve certaines difficultés concernant l'unification du stockage et la recherche des données relatives aux échantillons de patients dont elle bénéficie. Cela a pour effet de causer un ralentissement sur certains de ses processus métier.

Ainsi, il m'a été demandé de mettre en place une base de données permettant le stockage de toutes les informations relatives aux échantillons de patients que l'entreprise utilise, ainsi qu'un site web accessible à tous les collaborateurs de l'entreprise permettant de **visualiser**, **d'extraire** et de **modifier** ces informations à leur convenance. Ces solutions permettront d'améliorer le système d'information de Diag2Tec et ainsi d'augmenter la productivité de l'entreprise.

Nous allons désormais nous intéresser aux différentes parties qui constituent ce rapport. Pour commencer, nous présenterons l'entreprise Diag2Tec par son historique, son effectif et son activité ainsi que les entités qui gravitent autour, telles que le CHU de Montpellier et le CNRS. Ensuite, nous analyserons en détail le sujet du stage en décrivant la problématique associée, l'environnement technique dans lequel se situe l'entreprise et les solutions à apporter grâce à une analyse de l'existant et au cahier des charges qui m'a été soumis. Aussi, nous nous concentrerons sur les spécifications techniques liées aux solutions apportées. Nous aborderons également la façon dont l'organisation du projet a été menée ainsi que les différentes méthodes utilisées. Pour conclure, nous ferons le bilan sur l'ensemble du projet en rappelant les objectifs, ce qui a été réalisé, que les compétences auxquelles j'ai fait appel durant le projet et que j'ai pu renforcer ainsi que les perspectives d'amélioration du projet.

1 - Présentation de l'entreprise

J'ai réalisé mon stage au sein de l'entreprise Diag2Tec, une entreprise fondée en juillet 2015 et issue du laboratoire hospitalier « Suivi des Thérapies Innovantes » et de l'équipe académique « Modifications épigénétiques et l'instabilité génomique des cellules B normales et tumorales ». Les bureaux et laboratoires de la société Diag2Tec se situent au sein du département d'hématologie-biologique du Centre Hospitalier Universitaire (CHU) de Montpellier.

Cette partie sera consacrée à la présentation de l'entreprise Diag2Tec et de son environnement de travail.

1.1 - Le Centre Hospitalier Universitaire (CHU)

Créé en 1958, le CHU de Montpellier est un établissement exerçant dans le domaine de la santé publique, de la recherche et de l'enseignement.

Parmi les nombreux services que compte le CHU, le département d'hématologie-biologique est consacré au diagnostic et au suivi des patients atteints d'hémopathies malignes (maladies du sang).

Ce département héberge différentes structures telles que l'équipe académique et le laboratoire hospitalier du Pr Jérôme Moreaux et la société Diag2Tec. Le laboratoire hospitalier "Suivi des Thérapies Innovantes" a notamment pour missions de diagnostiquer et de suivre la maladie résiduelle de patients atteints de myélome multiple et de collectionner des cellules de patients atteints de diverses hémopathies malignes.

1.2 - Le Centre National de la Recherche Scientifique (CNRS)

Le CNRS est un institut de recherche scientifique dans les domaines du vivant, de la matière et de l'univers, possédant une influence à échelle internationale.

L'équipe académique du Pr Jérôme Moreaux de l'Institut de Génétique Humaine (UMR9002-IGH-CNRS) étudie les modifications épigénétiques* et l'instabilité génomique des cellules impliquées notamment dans les hémopathies malignes, et particulièrement dans le myélome multiple. L'équipe étudie la biologie du myélome multiple et les mécanismes associés à la résistance aux drogues. Elle collabore avec de nombreux groupes français mais aussi internationaux (Pr Goldschmidt, Heidelberg ; Pr Vanderkerken, Bruxelles ; Pr Elemento, New York)

1.3 - L'entreprise Diag2Tec

L'entreprise Diag2Tec est une jeune entreprise innovante issue du laboratoire hospitalier et de l'équipe académique dirigée par le Pr Jérôme Moreaux dans le but de valoriser les travaux effectués par les deux entités.

Elle est fondée en juillet 2015 par six associés dont le Pr Jérôme Moreaux et Angélique Bruyer, actuelle présidente de la société.

Diag2Tec compte à ce jour dix collaborateurs, dont quatre personnes à temps plein en CDI : la présidente qui occupe également la fonction de Directrice du laboratoire, deux chefs de projets et une assistante ingénieur. Concernant le personnel en CDD, il y a une cheffe de projets, une doctorante dans le cadre d'une thèse CIFRE et un assistant ingénieur. La société accueille à ce jour deux stagiaires. Quant à Jérôme Moreaux, il occupe le poste de CSO (Chief Scientific Officer).

Pour mener à bien ses missions, Diag2Tec se positionne comme une société de recherche contractuelle (CRO : Contract Research Organization) préclinique et propose à ses clients des tests innovants via des modèles cellulaires uniques et des analyses bio-informatiques de pointe. Le modèle économique de Diag2Tec repose sur une activité de prestations de services proposées aux clients et une activité interne de Recherche et Développement (R&D) permettant de développer de nouveaux types de services, de modèles cellulaires et d'identifier de nouveaux biomarqueurs* d'orientation thérapeutique. La principale activité de prestations de services de Diag2Tec est l'évaluation de l'effet cytotoxique* de différents agents thérapeutiques en développement préclinique utilisés seuls ou en combinaisons sur différents modèles cellulaires de cancers hématologiques.

Les principaux clients de Diag2Tec pour la vente des différentes prestations de services sont de deux typologies : les laboratoires pharmaceutiques et les sociétés de biotechnologie développant des agents thérapeutiques en préclinique dans les différents cancers hématologiques. Diag2Tec travaille depuis fin 2017 avec plusieurs laboratoires pharmaceutiques localisés aux USA, en Chine et en Corée du Sud et avec plusieurs sociétés de biotechnologies localisées en France, en Allemagne, en Angleterre et en Israël.

La stratégie de développement de Diag2Tec est de se positionner comme leader sur le marché des sociétés proposant des prestations de services pour les tests de molécules précliniques dans les cancers hématologiques. Ainsi, ses objectifs sont de pouvoir travailler avec quelques-uns des dix plus gros laboratoires pharmaceutiques mondiaux qui développent des traitements anticancéreux dans les lymphomes*, le myélome multiple et les leucémies et de devenir leur CRO avec une activité récurrente chaque année sur leurs différents pipelines de développement. On retrouve parmi les clients cibles : Novartis, Merck, Sanofi, Gilead, GSK, Pfizer, AstraZeneca, Janssen, Genentech.

2 - Analyse

Dans cette partie, nous analyserons dans un premier temps, le sujet du stage ainsi que le contexte associé. Ensuite, nous effectuerons une analyse de l'existant. Enfin, nous décrirons le cahier des charges afin de mettre en lumière les différentes activités que j'ai effectuées.

2.1 - Présentation du sujet et analyse du contexte

Dans l'objectif d'améliorer son système d'information concernant le stockage des données biologiques relatives aux informations concernant des échantillons de lignées cellulaires et de cellules primaires de patients, l'entreprise Diag2Tec m'a engagé en tant que stagiaire afin d'y apporter une solution. En effet, les membres de l'entreprise Diag2Tec rencontrent certaines difficultés quant à leur manière de rassembler, stocker, rechercher et partager leurs données concernant les caractéristiques biologiques des échantillons de patients atteints d'hémopathie maligne. Jusqu'à présent, Diag2Tec utilise le logiciel Excel pour stocker ces données sous forme de tableau ce qui amène à certains problèmes tels que l'insertion de doublons ou encore l'apparition de données incohérentes. De plus, ce système de stockage rend la recherche fastidieuse car elle nécessite de parcourir de nombreux fichiers.

Ma mission est de rassembler toutes les données stockées dans ces fichiers Excel et de concevoir puis implémenter une base de données dans le système informatique de l'entreprise. Enfin, pour que les employés puissent interagir avec cette base de données, il m'a été demandé de créer une application web accessible sur plusieurs supports (Windows et macOS) permettant de visualiser, modifier, importer et exporter les données de cette base de données.

Enfin, une fois les solutions mises en place, mon rôle sera d'accompagner les membres de cette société à l'utilisation de ces nouvelles solutions.

2.2 - Analyse de l'environnement technique

L'entreprise Diag2Tec a mis à disposition de ses employés plusieurs postes informatiques ainsi que des ordinateurs portables (intégrant les systèmes d'exploitation Windows / macOS). De plus, Diag2Tec possède un réseau wifi privé pour ses employés ainsi qu'un poste informatique centrale gérant un NAS permettant le dépôt de fichiers par les utilisateurs de ce réseau privé et la sauvegarde de l'ensemble des postes informatiques de l'entreprise.

Enfin, cette start-up utilise le pack Office et notamment 'Excel' pour stocker et gérer l'ensemble des informations liées aux échantillons de lignées ou de cellules primaires de patients qu'elle manipule ainsi que certains résultats de ces manipulations.

2.3 - Analyse de l'existant

Pour estimer le travail à accomplir, il m'a été fourni plusieurs exemples de produits finaux tels que l'outil Biomart disponible sur le site ensembl.org qui offre la possibilité de rechercher des données biologiques parmi diverses bases de données. Biomart permet aussi d'exporter le résultat des données recherchées au format HTML, CSV, TSV et XLS. Néanmoins, cet outil possède un défaut majeur concernant son temps d'exécution très long dû à son important volume de données à traiter. Il sera donc important que notre base de données soit optimisée afin d'éviter des problèmes similaires lors de son utilisation.

En parallèle, l'équipe du CNRS possède une base de données de congélation d'échantillons biologiques liée à une interface web et a proposé de me montrer le fonctionnement et l'architecture de leur outil. Leur base de données est constituée de plusieurs tables contenant chacune plusieurs dizaines de champs n'étant pas systématiquement remplis. Aussi, elle possède un nombre réduit d'associations permettant de faciliter les jointures entre chaque table pour ainsi afficher plus simplement les informations importantes. En effet, cette base de données ne traite pas un volume important de données, il n'a donc pas été nécessaire de normaliser entièrement cette base de données.

Concernant leur interface web, elle permet l'ajout, la suppression, la mise à jour et le visionnage des données. Cette interface offre aussi la possibilité aux utilisateurs de choisir les tables à afficher par l'intermédiaire de cases à cocher et de trier les données grâce à des filtres.

Enfin, ces deux exemples présentent des points communs : un design minimaliste permettant d'aller à l'essentiel, une ergonomie concernant l'affichage des informations, la présence de filtres permettant d'améliorer la précision des recherches, le traitement d'un grand nombre de données.

Par la suite, les membres de l'entreprise Diag2Tec m'ont apporté davantage de détails sur leurs besoins. J'ai pu conclure qu'ils recherchaient un site internet possédant des caractéristiques similaires aux deux précédents sites avec la possibilité :

- de choisir avec précision les données à afficher ainsi que de trier par résultat grâce à la présence de filtres ;
- d'exporter les résultats après une recherche en fichier Excel ;
- de modifier les données présentes dans la base de données depuis ce même site ;
- d'ajouter de nouvelles données via un formulaire et enfin, d'importer des données via des fichiers au format '.CSV'.

2.4 - Cahier des charges

Le cahier des charges ([Annexe 1](#)) a été rédigé par Diag2Tec. Il a pour objectif de détailler les attentes de l'entreprise vis-à-vis de la solution à mettre en place.

Ainsi, ce rapport se compose en plusieurs rubriques. La première se concentre sur l'identification des besoins de Diag2Tec en matière de stockage et de traitement des données. En effet, l'entreprise utilise de nombreux fichiers Excel pour stocker ses données ce qui ne facilite pas la recherche, l'ajout, l'édition et le contrôle qualité des données.

Pour résoudre ces problèmes, deux objectifs majeurs sont donnés :

- Conceptualiser une base de données et l'implémenter en harmonisant au préalable les données, et notamment les noms de colonnes, contenus dans les fichiers ;
- Créer une interface web permettant la visualisation, l'édition, l'exportation et l'importation de données.

En plus de ces indications, un extrait d'une table est présenté dans le cahier des charges mettant en avant les critères sur le nom, la catégorie (catégorie d'activité à laquelle cette donnée appartient), le type de valeur (boolean, chaîne de caractères, nombre) et le degré de liberté sur la saisie des données. Ainsi, chaque valeur possède ses propres caractéristiques devant être retranscrites dans la base de données.

Enfin, les nombreuses fonctionnalités de l'interface web sont détaillées. Cette interface web devra dans un premier temps, permettre une recherche précise des données en donnant le choix à l'utilisateur : du modèle (lignée cellulaire ou cellule primaire de patient), du type de pathologie (« Multiple Myeloma », « Acute Myeloid Leukemia » ...) des attributs à afficher (colonnes de données) et enfin, des filtres agissant sur les résultats attendus.

Après qu'une recherche ait été effectuée, l'utilisateur aura la possibilité d'exporter ses résultats en fichier CSV.

Ensuite, cette interface devra proposer deux méthodes d'insertion. La première méthode d'insertion se fera grâce à un formulaire et permettra l'insertion d'un échantillon un à un, tandis que la seconde utilisera les données contenues dans des fichiers CSV pour simplifier les insertions multiples.

Finalement, l'interface permettra d'éditer les données afin qu'elles soient ensuite mises à jour directement dans la base de données.

3 - Spécifications techniques

Cette partie abordera les spécifications techniques de la solution apportée au problème de système d'information de l'entreprise. En première partie, nous nous concentrerons sur la conception de la base de données. Enfin, nous nous pencherons sur la conception du site web associé à cette base de données.

3.1 - Conception de la base de données

Jusqu'à présent, l'entreprise Diag2Tec stockait ses données dans des fichiers Excel. Pour concevoir cette base de données, je me suis référé aux données contenues dans ces fichiers Excel et sur les informations apportées par mon tuteur de stage pour établir des relations entre ces données.

Mon premier objectif fut de déterminer les redondances présentes dans les différentes tables et de trouver de potentielles solutions afin de les éliminer pour accélérer les temps d'exécution des requêtes et limiter les potentielles incohérences de données en prévision de son utilisation future. Pour ce faire, j'ai d'abord déterminé les clés primaires puis formé les tables contenant les informations en relation avec ces clés. Rapidement, j'ai remarqué qu'une grande majorité des données étaient liées à la table « Samples » ([Figure 1](#)), ne possédant aucune caractéristique les rendant uniques, et ainsi ne me permettant pas de les séparer. Pour alléger la table « Samples » et organiser les données de manière plus cohérentes, j'ai décidé de faire appel aux notions d'héritages pour séparer les données en sous-table afin de rendre les données plus lisibles et éviter une table « Samples » constitué en grande partie de valeur NULL.

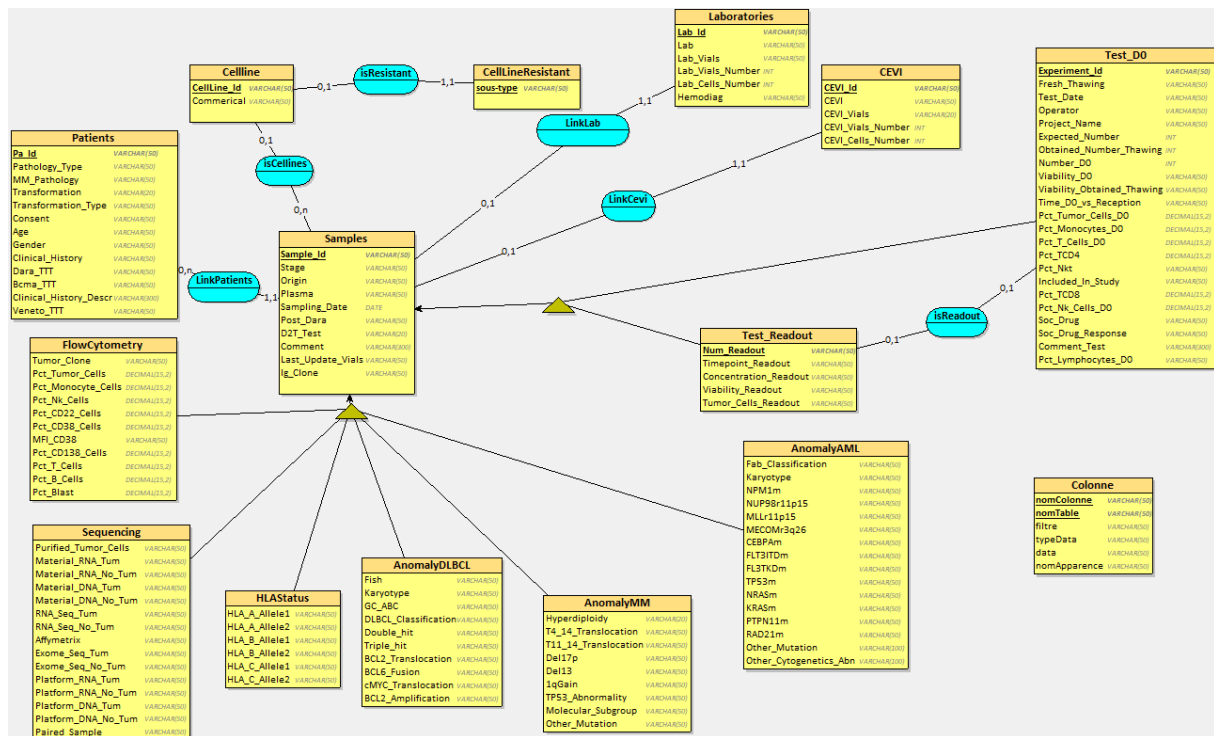


Figure 1 : Modèle entité-association

Ensuite, pour faciliter les imports des données formatées au format .CSV et la lecture des données par l'intermédiaire de requêtes SQL, j'ai choisi de ne pas séparer les données de certaines tables telles que « Sequencing ». En effet, la table « Sequencing » répertorie les informations de séquençage* sur les différents échantillons tumoraux et non tumoraux. J'ai choisi de ne pas séparer les données de ces deux types d'échantillons car il était plus intéressant d'avoir accès à ces deux informations par le biais d'une seule jointure.

Aussi, dans ces mêmes objectifs, j'ai fait le choix de ne pas utiliser de table de « jointure intermédiaire » (table contenant uniquement des clés primaires entre deux ou N tables) et par conséquent, de n'avoir que des jointures unique table par table ([Voir Figure 2](#)).

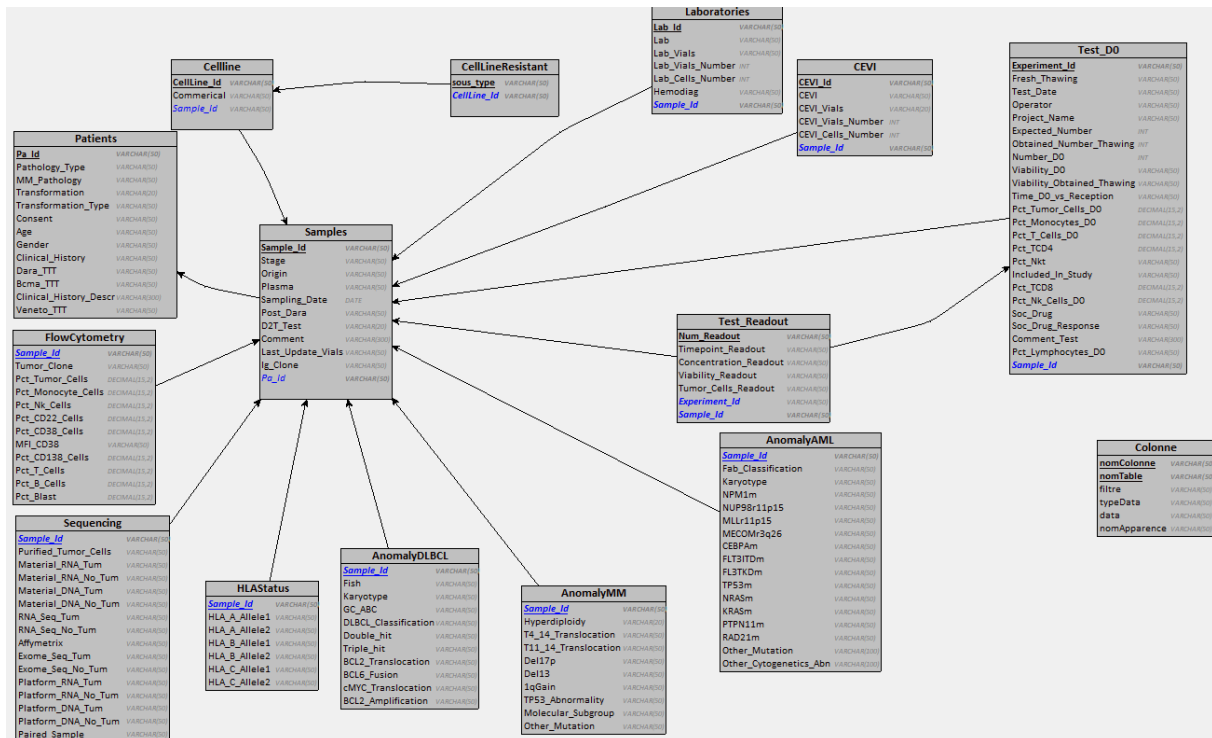


Figure 2 : Schéma MLD

Enfin, pour rendre possible la création de filtres sur les données, il m'a été nécessaire de m'entretenir avec les membres de Diag2Tec afin d'harmoniser les données contenues dans les fichiers Excel fournis puis établir des énumérations de « VARCHAR » pour toutes les données. Par exemple, toutes les données de type booléen (oui, non) telles que le consentement des patients, se sont transformées en énumération de type : Yes, No, NA permettant ensuite de retrouver plus aisément tous les patients qui ont donné leur consentement (Yes) ou qui sont en attente de réponse (NA).

3.2 - Conception du site web

3.2.1 – Les objectifs

Pour commencer, à l'aide des recherches décrites dans l'analyse de l'existant, ma mission fut de créer un site internet consultable seulement par les membres de l'entreprise et permettant de :

- Visionner n'importe quelles données dans la base de données en choisissant les colonnes à afficher ;
- Trier les données par résultat grâce à des filtres ;
- Éditer les données ;
- Ajouter de nouvelles données par l'intermédiaire de formulaires ;

- Importer de nouvelles données par l'intermédiaire de fichiers au format .CSV ;
- Exporter les données voulues au format .CSV.

3.2.2 – L'architecture de l'application

Dans un premier temps, j'ai mis en place l'architecture de l'application. Pour ce faire, j'ai utilisé les langages HTML/CSS pour la création des pages web mais aussi PHP et Javascript pour les dynamiser et communiquer avec la base de données. De plus, j'ai utilisé le model MVC afin de séparer les différents modules de mon application et de pouvoir les tester séparément. (Diagramme de classe PHP [Figure 3](#))

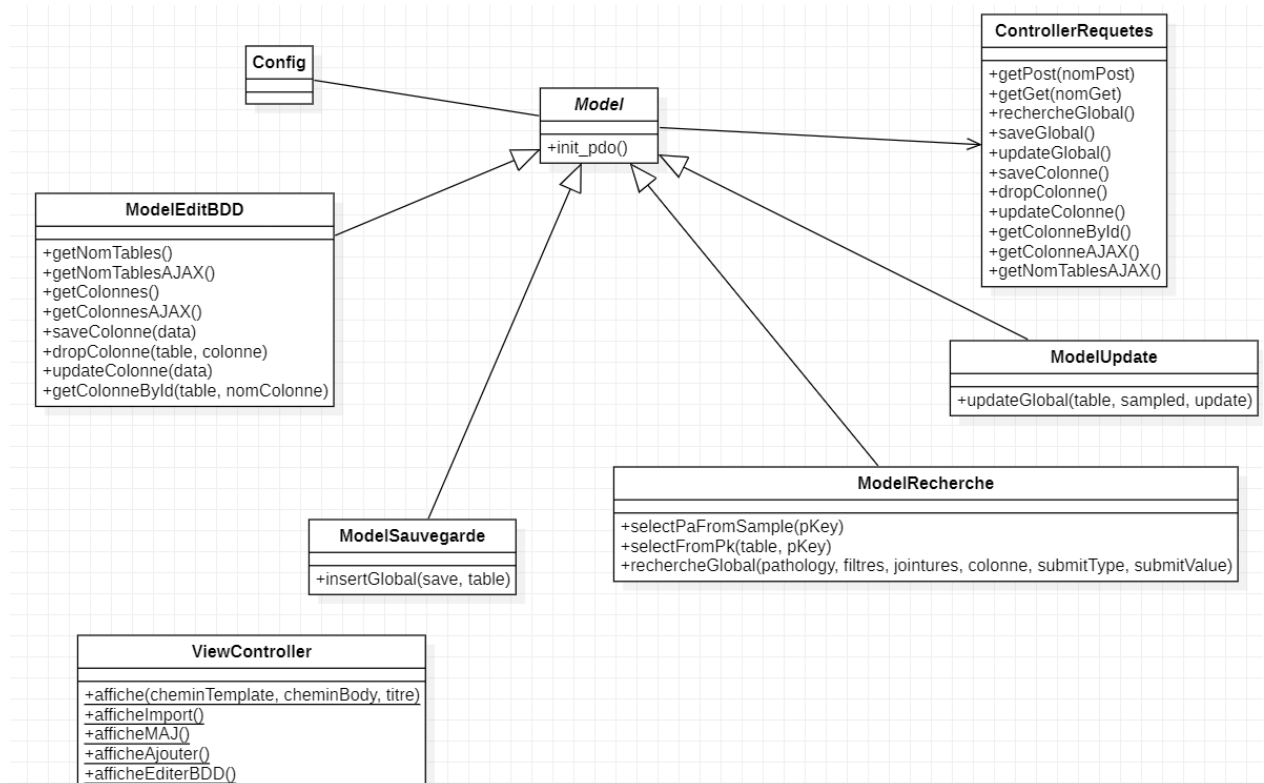


Figure 3 : Diagramme de classe de l'architecture PHP

Le diagramme ci-dessus présente les différentes classes PHP qui compose l'application. Pour commencer, la classe « ControllerRequetes » est utilisé pour récupérer les informations envoyées par les requêtes AJAX. Elle joue un rôle de passerelle entre JavaScript et PHP. En fonction de la méthode d'envoi (GET ou POST), les méthodes `getPost()` ou `getGet()` sont utilisées afin de récupérer les différentes informations tout en vérifiant que les valeurs reçue ne soit pas « NULL », vide ou « undefined » (undefined étant la valeur par défaut de JavaScript pour les valeurs non défini). Cette vérification permet ensuite d'insérer des valeurs « NULL » dans la base de données lorsque l'information est manquante.

Outre cela, la classe « ControllerRequetes » fait appel aux différentes méthodes des classes « Model » permettant l'exécution de requêtes préparées (voir schéma de communication [Figure 4](#)).

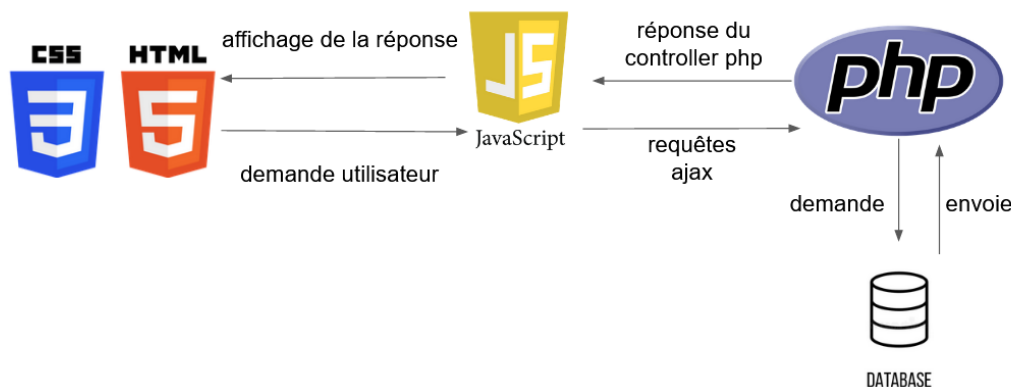


Figure 4 : Schéma simplifié des échanges entre les différentes technologies

Enfin, la classe « ViewController » couplée à un fichier index.php, permet d'effectuer le routage des différentes pages du site web. La méthode « affiche() » de « ViewController » permet l'utilisation d'un fichier templateView.php contenant tous les éléments à afficher sur chaque page (header, footer, CSS pour le style global des pages...). Le contenu des différentes pages est ensuite affiché dans le « <body> » de cette page template.

3.2.3 – Les fonctionnalités d'ajouts

En ce qui concerne les fonctionnalités, pour permettre à l'utilisateur l'insertion multiple de données, j'ai conçu une page permettant de sélectionner un fichier .CSV et d'importer les données contenues à l'intérieur ([Figure 5](#)). Pour réaliser ce module, j'ai d'abord créé un algorithme de conversion des données CSV au format « Array » de Javascript. Pour réaliser cet algorithme, je me suis intéressé à la composition des fichiers « CSV ». Les fichiers « CSV » (pour « Comma-separated values ») sont des fichiers utilisant des virgules ou points-virgules ainsi que des retours à la ligne pour séparer les données et distinguer les lignes des colonnes. J'ai donc utilisé ces deux particularités pour générer des « Array » de la forme « ['nomColonne', 'valeurLigne'] ». Ensuite, j'ai réalisé que dans certains cas, les conversions de format .XLS à .CSV pouvaient générer parfois des erreurs lors des sauts de lignes nécessaire à la distinction des colonnes. Afin de gagner du temps, je me suis intéressé aux bibliothèques javascript gérant ce genre de conversion. Ainsi, j'ai utilisé la bibliothèque « Papa-parse » qui permet de paramétrer avec précision un algorithme de conversion « Array to CSV » et qui gère ce genre de problème.



Figure 5 : Visuel de la page d'importation

Après avoir résolu ce problème, j'ai ensuite réalisé une requête AJAX qui permet l'insertion des données du tableau de données généré par Papa-parse dans la base de données.

Outre cette méthode d'insertion, j'ai aussi réalisé un formulaire qui offre à l'utilisateur la possibilité d'insertion unique et manuelle des données dans les tables qu'il souhaite.

3.2.4 – La fonctionnalité de recherche

Une fois cette tâche effectuée, j'ai créé une page de recherche interagissant avec la base de données. Cette page permet à l'utilisateur de choisir : les colonnes à afficher, les jointures sur les tables disponibles et d'afficher les données selon un résultat voulu ([Figure 6](#)). Après une recherche effectuée, l'utilisateur a le choix d'exporter le résultat de cette recherche en fichier au format « CSV » ([Figure 7](#)).



Choisir :

Humans Cells Lines FL

Recherche :

☒ CEVI ☐ Laboratories ☐ HLAStatus ☐ Sequencing ☐ FlowCytometry
☐ Tout cocher

Colonnes à afficher :

☒ Stage ☒ Origin ☐ Plasma ☐ Sampling_Date ☐ D2T_Test ☒ Comment
☐ Pathology_Type ☐ MM_Pathology ☐ Tansformation ☐ Transformation_Type ☒ Consent ☐ Age ☐ Gender ☐ Clinical_History ☐ Dara_TTT ☐ Bema_TTT

Autre Affichage :

☒ CEVI_Id ☒ CEVI_Cond ☒ CEVI_Vials ☐ CEVI_Vials_Number ☐ CEVI_Cells_Number

Filtres :

Consent Clinical_History Stage Origin Plasma D2T_Test
Consent Clinical_History Stage Origin Plasma D2T_Test
CEVI_Cond CEVI_Vials
CEVI_Cond CEVI_Vials

Entrer un sample_id rechercher Entrer un pa_id rechercher

rechercher Exporter


Figure 6 : Visuel de la page de recherche

rechercher	Exporter	Sample_Id	Pa_Id	Stage	Origin	Comment	Consent	CEVI_Id	CEVI	CEVI_Vials
		FL1509	1515512	Diagnosis	Lymph node		Yes	CeVi060092	Yes	Yes
		FL1606	2928534	Diagnosis	Lymph node		Yes	CeVi060154	Yes	Yes
		FL16148	1272054	Relapse	Lymph node		Yes	CeVi060295	Yes	Yes
		FL17193	3048902	Diagnosis	Lymph node		Yes	CeVi060497	Yes	Yes
		FL17200	2487413	Relapse	Lymph node	2nd Progression after 2 lines of TTT	Yes	CeVi060502	Yes	Yes
		FL17216	1077077	Diagnosis	Lymph node		Yes	CeVi060509	Yes	Yes
		FL17299	1103938	Diagnosis	Lymph node		Yes	No	No	No
		FL17401	1569633	Diagnosis	Blood		Yes			
		FL1757	2713451	Relapse	Lymph node	contrat DRI 2021-19_D2T, Perdu de vue depuis mars 2017. Cession possible sans CST	Ongoing			
		FL18202	2467053	Relapse	Lymph node	contrat DRB 2020-5, pas CST HEMO mais CALYM	Yes	CeVi060745	Yes	Yes
		FL18451	3180320	Diagnosis	Blood		Yes	CeVi060897	Yes	Yes
		FL18465	2043100	Diagnosis	Lymph node		Yes			
		FL1928	3180320	Diagnosis	Lymph node		Yes	CeVi060915	Yes	Yes

Figure 7 : Exemple de résultat obtenu après une recherche

3.2.5 – La fonctionnalité de mise à jour

Enfin, j'ai mis en place une page, similaire au formulaire d'ajout, permettant de mettre à jour les données en modifiant les données champs par champs, table par table avec une prévisualisation des données actuelles et une actualisation automatique après mise à jour (Figure 8). Cette page est accessible par click sur un « Sample_Id » dans l'onglet de recherche ou d'ajout et n'est pas disponible autrement.

 Recherche Importer PhpMyAdmin

Patient :

MM_Pathology	Transformation	Transformation_Type	Consent	Age	Gender	Clinical_History	Dara_TTT	Bema_TTT	Clinical_History_Descr	Veneto_TTT		
Pa_Id	Pathology_Type	MM_Pathology	Transformation	Transformation_Type	Consent	Age	Gender	Clinical_History	Dara_TTT	Bema_TTT	Clinical_History_Descr	Veneto_TTT
1515512	FL			NA	Yes	0	NA	No				

Update Patients

Samples :

Stage	Origin	Plasma	Sampling_Date	Post_Dara	D2T_Test	Comment		
Sample_Id	Stage	Origin	Plasma	Sampling_Date	Post_Dara	D2T_Test	Comment	Pa_Id
FL1509	Diagnosis	Lymph node	NA	0000-00-00		No		1515512

Update Samples

Laboratories :

Lab	Lab_Vials	Lab_Vials_Number	Lab_Cells_Number	Hemodiag		
Lab_Id	Lab	Lab_Vials	Lab_Vials_Number	Lab_Cells_Number	Hemodiag	Sample_Id
HEMO15_09	Yes	Yes	7	10	Yes	FL1509

Update Laboratories

CEVI :

CEVI	CEVI_Vials	CEVI_Vials_Number	CEVI_Cells_Number		
CEVI_Id	CEVI	CEVI_Vials	CEVI_Vials_Number	CEVI_Cells_Number	Sample_Id
CeVi060092	Yes	Yes	0	0	FL1509

Update CEVI

Figure 8 : Visuel de la page de mise à jour des données

3.3 – Dynamisation du site web et de la base de données

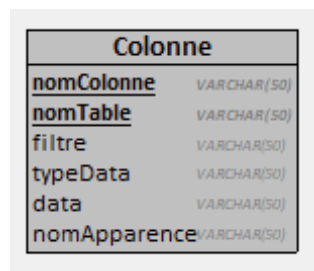
Après avoir conçu et réalisé une version fonctionnelle de la base de données et de l'application, nous nous sommes intéressés sur de potentielles solutions permettant de rendre évolutif cette application. C'est ainsi que deux fonctionnalités ont été retenues, la première permettant d'ajouter de nouvelles colonnes dans la base de données et la seconde, d'ajouter de nouvelles tables pour qu'elles fonctionnent ensuite automatiquement avec toutes les fonctions présentes dans l'application (recherche, import, export, ajout, mise à jour). L'ajout de nouvelles tables étant soumise à trop d'aléatoire (gérer les contraintes d'intégrité sur les clés de manière dynamique, le comportement en cas de suppression d'une table, les potentiels bugs générés par l'application sur les diverses fonctionnalités...). C'est pourquoi, nous avons décidé de commencer par l'ajout de nouvelles colonnes et de réfléchir plus tard à la faisabilité de l'ajout dynamique de table.

Pour rendre possible l'ajout, la modification et la suppression de colonnes dans notre application, j'ai décidé de stocker toutes les informations sur ces colonnes dans notre base de données. L'ensemble des caractéristiques liées aux colonnes dans notre application sont définis par :

- Les colonnes possèdent un nom dans la base de données et un nom « d'apparence » sur l'interface ;

- Chaque colonne doit contenir un certain type de valeur (Chaîne de caractères, nombre entier, nombre flottant, date...) et doit être définie par un type de balise en HTML (<input> pour les nombres, les dates ou les zones de texte ; <select> pour les valeurs définies) ;
- Chaque colonne doit appartenir à une table dans la base de données ;
- Il ne peut exister qu'une seule colonne de même nom dans une table ;
- Une colonne peut être affichée dans une recherche et/ou être utilisée comme filtre
- L'utilisateur doit pouvoir assigner ou mettre à jour les valeurs de chaque colonne (sauf exception) depuis l'application.

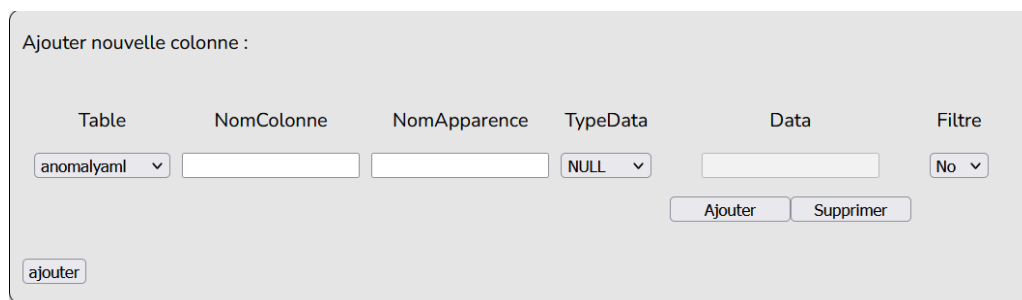
Ainsi, toutes les colonnes utilisées seront stockées dans une table « Colonne » possédant comme données, un nom de colonne, un nom de table, un booléen permettant de savoir si la colonne doit être utilisée ou non comme un filtre de recherche, le type de données représenté par la colonne et un nom d'apparence (nom sur l'application) ([Figure 9](#)).



Colonne	
<u>nomColonne</u>	VARCHAR(50)
<u>nomTable</u>	VARCHAR(50)
filtre	VARCHAR(50)
typeData	VARCHAR(50)
data	VARCHAR(50)
nomApparence	VARCHAR(50)

Figure 9 : Table « Colonne »

Les nouvelles colonnes pourront être ajoutées grâce à des formulaires disponibles sur l'application et des requêtes qui permettront d'ajouter et d'éditer des colonnes dans la table demandée et dans la table « Colonne » ([Figure 10](#) et [Figure 11](#)).



Ajouter nouvelle colonne :

Table	NomColonne	NomApparence	TypeData	Data	Filtre
anomalyaml ▾	<input type="text"/>	<input type="text"/>	NULL ▾	<input type="text"/>	No ▾
			<input type="button" value="Ajouter"/> <input type="button" value="Supprimer"/>		
<input type="button" value="ajouter"/>					

Figure 10 : Formulaire permettant l'ajout d'une colonne

Editer une colonne

Table	NomColonne	NomApparence	TypeData	Data	Editer	Supprimer	Filtre
anomalyaml	CEBPAm	CEBPAm	select	Yes	<input type="checkbox"/>	<input type="checkbox"/>	Yes
				No	<input type="checkbox"/>	<input type="checkbox"/>	
				Ongoing	<input type="checkbox"/>	<input type="checkbox"/>	

Ajouter Supprimer

Editer Supprimer

Figure 11 : Formulaire d'édition d'une colonne

Après avoir modifié la base de données, j'ai ensuite revu l'intégralité des fonctionnalités de l'application afin que les interfaces soient liées aux données de la table « Colonne ». Ainsi, l'application se met automatiquement à jour à la suite d'un ajout ou d'une modification d'une colonne par l'intermédiaire des formulaires.

Cette nouvelle fonctionnalité permet ainsi aux utilisateurs de bénéficier de l'ajout ou la modification de la base de données et sans connaissances requises. De plus, cette solution offre la possibilité aux utilisateurs d'avoir un contrôle précis sur les différents filtres mis à disposition dans l'onglet de recherche de l'application.

Pour conclure sur les fonctionnalités, la [Figure 12](#) présente un diagramme de cas-utilisation résumant l'intégralité des fonctionnalités disponibles pour les utilisateurs.

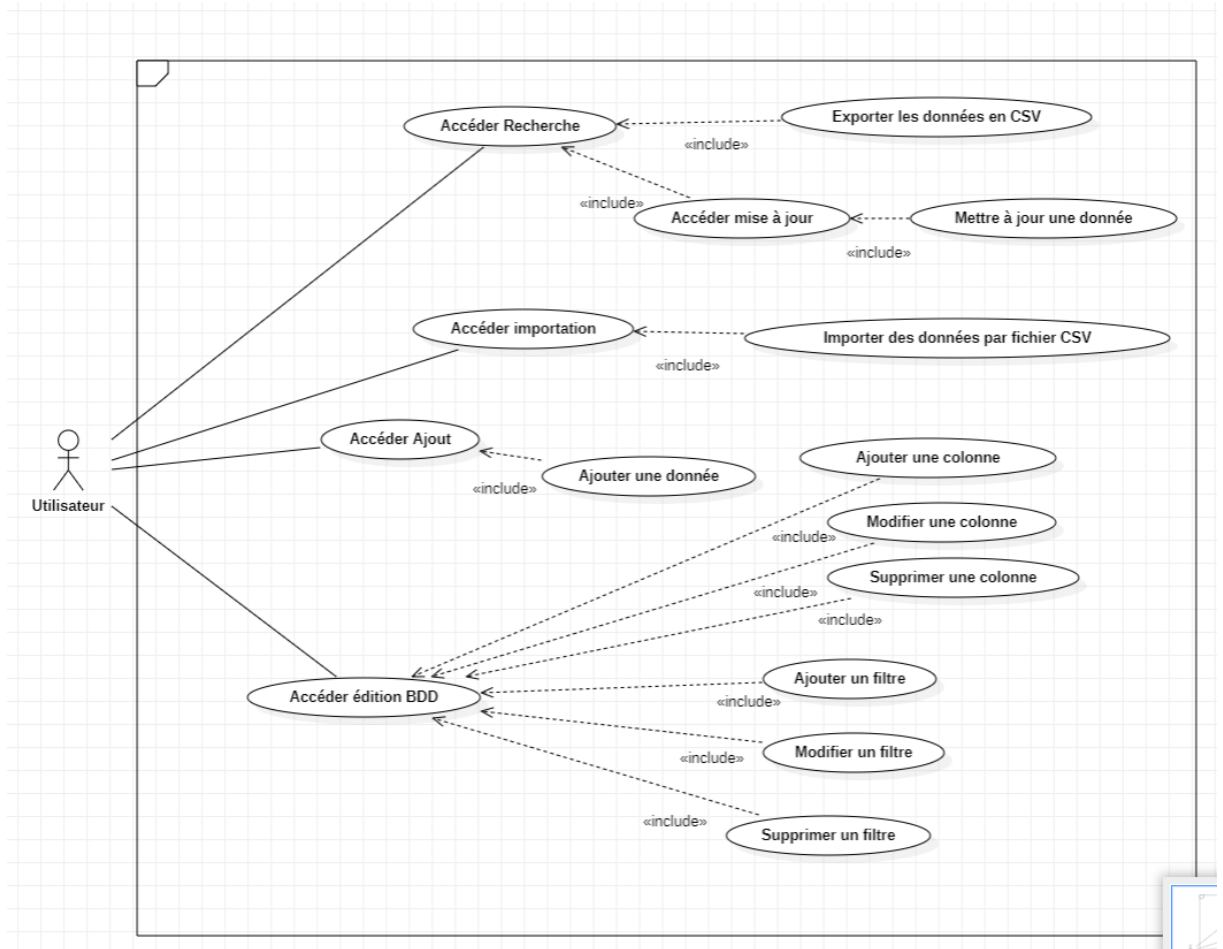


Figure 12 : Cas-Utilisation application web

3.4 – Contrôle qualité et test de la solution

L'application a été testée visuellement tout au long de son développement notamment grâce aux outils de visualisation du trafic réseau des navigateurs web me permettant de déterminer les potentielles erreurs dans les requêtes SQL, le code PHP, ou les requêtes AJAX. Le code de l'application étant partitionné entre plusieurs technologies, il était facile de déterminer où se situaient les éventuels bugs et de les corriger.

Concernant la base de données, pour tester la rapidité de traitement de la base de données, j'envisage au cours des 4 dernières semaines de mon stage d'utiliser l'outil « Apache JMeter » permettant de saturer une base de données de requêtes afin d'évaluer son temps de réponse et ses limites.

Pour finir, afin de s'assurer que l'application répond bien aux différents critères et besoins de l'entreprise, l'application web sera soumise à une phase de « bêta test » durant laquelle les employés auront accès à l'application et pourront tester les fonctionnalités et noter les défauts et bugs rencontrés ainsi que les ajustements nécessaires. Cette phase de test se déroulera durant les quatre dernières semaines du stage.

Au départ, seul Mme Alaterre aura accès à la solution et pourra la tester avec le regard d'un informaticien. Ensuite, Mme Bruyer pourra tester la solution avec une utilisation professionnelle. Enfin, l'ensemble des membres de l'entreprise auront accès à la solution et pourront faire part de leurs retours.

4 – Méthodologie et organisation du projet

Dans cette partie, nous aborderons la manière dont Diag2Tec gère ses projets en interne. Ensuite, nous nous intéresserons à l'organisation et les méthodes qui ont été employées pour faciliter l'avancement de mon projet.

4.1 – Organisation du travail au sein de l'entreprise

Diag2Tec est une start-up travaillant dans le domaine de la biologie cellulaire. Les projets de l'entreprise sont gérés par les quatre chefs de projets qui compose l'entreprise. Chaque chef de projet travaille en collaboration avec un assistant ingénieur et/ou un stagiaire. Pour un projet type, la durée varie selon le nombre d'échantillons et la complexité des expériences à réaliser.

Étant une petite entreprise, Diag2Tec n'emploie aucune méthode de gestion traditionnelle. Néanmoins, chaque équipe s'organise en interne. En effet, chaque équipe met généralement en place des réunions durant lesquelles chaque membre témoigne de leurs avancés.

Aussi, la CEO Mme Bruyer organise des réunions avec chacun des chefs d'équipe pour suivre en temps réel l'évolution des projets en cours.

Enfin, Mme Bruyer organise aussi des réunions hebdomadaires durant lesquelles tous les employés participent et peuvent faire part de certaines remarques concernant l'organisation en général et les problèmes rencontrés.

En ce qui concerne les projets informatiques, l'entreprise Diag2Tec n'ayant pas pour habitude de gérer ce genre de projet, je me suis donc adapté à leur méthode pour organiser mon projet.

Concernant le versionnage du code, l'entreprise ne possédant pas au début de mon stage de repository Git, j'ai donc dans un premier temps stocker ma progression sur un git personnel privé. À la suite d'échanges avec ma tutrice, Mme Alaterre, j'ai entrepris la démarche de créer un Git pour l'entreprise Diag2Tec.

4.2 – Organisation du projet

L'ensemble des douze semaines qui compose ce stage ont été organisées en phases. La première semaine a servi de phase d'introduction. Durant cette phase, Mme Alaterre m'a introduit aux bases (en matière de biologie) nécessaires à la compréhension des différentes données grâce auxquelles je devais ensuite réaliser la conception de la base de données.

Ensuite, le premier mois était réservé à la réalisation d'une première conception de la base de données et à l'intégration des données dans cette dernière.

Une fois la base de données fonctionnelle, la troisième phase consistait en la création d'une application web permettant l'interaction avec la base de données précédemment conçue.

Enfin, la dernière phase qui n'a pas encore commencé, sera une phase de « bêta test ». Elle se déroulera pendant les dernières semaines du stage et permettra l'affinage et la résolution de bugs grâce aux retours des membres de Diag2Tec sur leur utilisation de la solution.

4.3 – Méthodologies et gestion du projet

Dans l'objectif d'organiser au mieux mon temps de travail, j'ai tout d'abord réalisé un backlog à l'aide de l'outil « Trello » répertoriant l'ensemble des User stories représentant les fonctionnalités attendues par l'entreprise puis j'ai énuméré l'ensemble des tâches à accomplir. Cependant, étant seul à travailler et en autonomie sur le développement de l'application, j'ai entrepris d'utiliser un carnet de bord et d'abandonner le Trello. Dans ce carnet de bord, je notais au fur et à mesure toutes les tâches restantes à réaliser, certains détails concernant les diverses fonctionnalités de l'application ainsi que toutes les questions et problèmes rencontrés et leurs méthodes ou pistes de résolutions.

Toutes ces questions et ces remarques étaient ensuite abordées lors de réunions avec Mme Alaterre se déroulant de plusieurs fois par jour lors de la première semaine de stage à plusieurs fois par semaine par la suite, gagnant ainsi en autonomie. Lors de ces réunions, nous discutons de l'avancement du projet, des problèmes rencontrés, des potentielles ajustements nécessaires, des potentielles nouvelles fonctionnalités et parfois des conditions de travail et de mon ressenti sur le déroulement du stage.

Conclusions et perspectives

Dans le cadre de mon stage et afin d'accélérer les processus métier en lien avec l'utilisation d'échantillons de patients atteints d'hémopathie maligne, j'ai réalisé à l'aide des compétences acquises durant mes deux années de BUT, une base de données de caractéristiques biologiques de ces échantillons. Cette base de données a été conçue à partir de données répartis dans plusieurs fichiers au format « CSV ». Ensuite, j'ai réalisé une application web connectée à cette base de données qui permet d'effectuer des recherches précises permettant de choisir les données à afficher, de les trier à l'aide de filtres et de les exporter en fichier « CSV ». Cette application intègre aussi des fonctionnalités d'ajout et de mise à jour de données par l'intermédiaire de formulaires et permet l'importation de gros volumes de données grâce à des fichiers au format « CSV ».

Enfin, l'application permet aux utilisateurs d'avoir un contrôle sur les tables de la base en leur donnant accès à des outils d'édition, d'ajout et de suppression de colonnes dans les différentes tables qui compose la base de données. Cette dernière fonctionnalité permet aussi aux utilisateurs d'ajouter, supprimer ou modifier les nombreux filtres disponibles dans l'onglet de recherche de l'application web.

Néanmoins, certains points concernant l'application web et ses fonctionnalités restent encore à améliorer.

En effet, une phase de conception sera nécessaire afin de finaliser la base de données pour la prise en charge des données de type « lignée cellulaire ».

Ensuite, il sera nécessaire de mettre en place des protections sur la suppression des colonnes ou sur les fonctionnalités modifiant les données en général. Des solutions tels que des pop-ups d'alerte ou une gestion de compte utilisateur avec permissions ont été envisagé.

Aussi, l'application n'ayant pas encore été testé dans son intégralité, il sera nécessaire de mettre en place une phase de « bêta test » pour identifier et corriger les bugs ou faire des ajustements.

Enfin, il sera nécessaire de tester la fiabilité de la base de données, notamment à l'aide de l'outil « Apache JMeter », afin de déterminer si la base de données pourra supporter l'utilisation de Diag2Tec.

Ce stage m'a permis de grandement consolider la compétence C3 (Gérer des données de l'information) par la conception et l'implémentation d'une base de données en partant de zéro. J'ai aussi grandement progressé sur le développement backend et frontend notamment avec javascript ou j'ai progressé sur l'utilisation de « requêtes AJAX », l'utilisation de « promise », la dynamisation de page HTML et l'intégration de librairies.

Aussi, j'ai pu consolider la compétence C5 (Conduire un projet). En effet, cerner les besoins de l'entreprise, s'adapter en fonction des disponibilités de chacun, identifier la faisabilité d'un projet, proposer des solutions sont des éléments nécessaires à la réussite d'un projet.

Enfin, travailler auprès de biologistes m'a permis d'avoir une première approche avec les domaines de la biologie et de la santé ainsi que le métier de bio-informaticien, qui m'intéresse tout particulièrement.

Vu le 14/06/2023

A handwritten signature in black ink, appearing to be 'M. Blanchet', with a stylized, sweeping flourish extending from the end.

Bibliographie

- [1]
Matt Holt, « Documentation - Papa Parse », 2019. <https://www.papaparse.com/docs>
- [2]
S. de leucémie et lymphome du Canada, « Le lymphome », 14 juin 2023.
<https://www.cancersdusang.ca/jai-un-cancer-du-sang/le-lymphome>
- [3]
S. de leucémie et lymphome du Canada, « Le myélome », 14 juin 2023.
<https://www.cancersdusang.ca/jai-un-cancer-du-sang/le-myelome>
- [4]
S. de leucémie et lymphome du Canada, « Leucémie », 14 juin 2023.
<https://www.cancersdusang.ca/jai-un-cancer-du-sang/leucemie>
- [5]
Matt Cowley, Tyler Caslin, Alex, Thomas Davis, Sven Sauleau, Drew Freyling, Ryan Kirkman,
« Libraries - cdnjs - The #1 free and open source CDN built to make life easier for
developers », *cdnjs*, 2023. <https://cdnjs.com/libraries>
- [6]
The PHP group, « PHP: PDOStatement::fetchAll - Manual », 2023.
<https://www.php.net/manual/en/pdostatement.fetchall.php>

Annexes

Annexe n°1 : Cahier des charges



CAHIER DES CHARGES TECHNIQUE ET DETAILLE

Conception d'une base de données des caractéristiques biologiques d'échantillons de patients atteints d'hémopathies malignes

Ce document technique a pour objectif de résumer les besoins et les spécifications techniques associés à la création d'une base de données des caractéristiques biologiques d'échantillons de patients atteints d'hémopathies malignes. Étant donné la complexité des modèles, de nombreux échanges seront effectués avec le stagiaire pour la compréhension du sujet et des besoins. L'ensemble des informations décrites dans ce cahier des charges ainsi que des informations complémentaires, notamment concernant la description des tables, auront été préalablement présentés et détaillés au stagiaire lors d'entretiens individuels.

I. Identification des besoins

La société Diag2Tec propose des services pré-cliniques de tests de drogues « à la carte » pour ses clients sur différents modèles cellulaires : les cellules primaires de patients et les lignées issues de patients atteints d'hémopathies malignes.

Parmi les hémopathies malignes étudiées au sein de la société, nous retrouvons les leucémies, différents lymphomes de cellules B et le myélome multiple. Chaque échantillon, parmi les différentes hémopathies, est caractérisé par divers critères clinico-biologiques plus ou moins importants pour leur sélection pour les tests de drogues pour nos clients.

A ce jour, l'ensemble des données est regroupé dans de nombreux fichiers Excel. Ces fichiers sont mis à jour régulièrement (importation de nouveaux échantillons et édition d'échantillons déjà implémentés) impliquant de nouvelles versions des documents et pouvant entraîner des incohérences des données. De plus, ces fichiers rendent difficile la recherche d'informations.

Les contraintes des outils actuellement utilisés montrent la nécessité de la création d'une base de données pour implémenter, éditer, rechercher les informations sur les échantillons d'intérêt.

II. Objectifs

Priorité 1 : Conceptualiser la base de données et harmoniser les jeux de données (donner les mêmes noms de colonnes quand cela est possible) et implémenter la base de données.

Priorité 2 : Création d'une interface web pour importer, visualiser, éditer et exporter les données.

III. Spécifications et caractéristiques fonctionnelles

1. Description des différentes tables

a. Les tables existantes

Toutes les informations clinico-biologiques connues pour chaque échantillon ont été regroupés dans diverses tables au format Excel pour chacun des modèles cellulaires (lignées cellulaires et cellules primaires de patients) :

- Cellules primaires de patients :
 - Table « Multiple Myeloma »
 - Table « Acute Myeloid Leukemia »
 - Table « Diffuse Large B-Cell Lymphoma »
 - Table « Mantle Cell Lymphoma »
 - Table « Follicular Lymphoma »
 - Table « Chronic Lymphocytic Leukemia »
- Lignées cellulaires :
 - Table « Multiple Myeloma »
 - Table « Multiple Myeloma – resistant cell lines »
 - Table « Acute Myeloid Leukemia »
 - Table « Diffuse Large B-Cell Lymphoma »
 - Table « Mantle Cell Lymphoma »
 - Table « T-cell Lymphoma »

Toutes les tables ne seront pas décrites dans ce cahier des charges, les échanges entre les tuteurs et le stagiaire permettront la description détaillée de l'ensemble de ces tables. Ici, nous décrirons uniquement les attributs de la table « Multiple Myeloma » pour le modèle des cellules primaires de patients.

b. Exemple de la table « Multiple Myeloma » des cellules primaires

Le tableau suivant résume les différents critères existants pour la table « Multiple Myeloma » pour le modèle des cellules primaires de patients. Il est composé des **attributs**, des **catégories** de chaque attribut (pour aider à la conceptualisation de la base de données), du **type** de chaque attribut (« numérique », « caractère », « booléen » ou « date »), ainsi que la **saisie** si le type est « caractère » : définie ou libre (e.g. si l'attribut est « Stage », le type est « Caractère », la saisie sera définie avec un choix entre « Diagnosis », « RR » ou « Follow-up »).

Attribut	Catégorie	Type	Saisie caractère
SAMPLE.ID	General criteria	Caractère	Libre
Lab.ID	General criteria	Caractère	Libre
PA.ID	General criteria	Caractère	Libre
Vials	General criteria	Booléen	
Consent	General criteria	Booléen	
Hemodiag	General criteria	Booléen	
Stage	General criteria	Caractère	Définie

Origin	General criteria	Caractère	Définie
Pathology.MM	General criteria	Caractère	Définie
Age	General criteria	Numérique	
Gender	General criteria	Caractère	Définie
Plasma	General criteria	Booléen	
Ptg.CD138	General criteria	Numérique	
Ptg.CD38	General criteria	Numérique	
MFI CD38	General criteria	Numérique	
Ig.clone	General criteria	Caractère	Définie
Gener.HMCL	General criteria	Booléen	
HMCL.ID	General criteria	Caractère	Libre
date.sampling	General criteria	Date	
CD138.purif	Sequencing data	Booléen	
RNA.CD138+	Sequencing data	Booléen	
DNA.CD138+	Sequencing data	Booléen	
Non.tumoral.fraction cells	Sequencing data	Booléen	
RNA.CD138-	Sequencing data	Booléen	
DNA.CD138-	Sequencing data	Booléen	
Affymetrix	Sequencing data	Booléen	
RNAseq.tum	Sequencing data	Booléen	
Platform.RNA	Sequencing data	Caractère	Définie
Exome.Seq.tum	Sequencing data	Booléen	
Platform.DNA	Sequencing data	Booléen	
RNAseq.Non.Tum	Sequencing data	Booléen	
Exome.Seq.Non.Tum	Sequencing data	Booléen	
t4.14.transloc	cytoG abnorm	Booléen	
t11.14.transloc	cytoG abnorm	Booléen	
del17p	cytoG abnorm	Booléen	
del 13	cytoG abnorm	Booléen	
Hyperdiploide	cytoG abnorm	Booléen	
TP53abnorm	cytoG abnorm	Booléen	
Molecular.subgroup	cytoG abnorm	Caractère	Définie
1q gain	cytoG abnorm	Booléen	
OTHER MUTATIONS GENES	cytoG abnorm	Caractère	Libre
HLA status A Allele 1	cytoG abnorm	Caractère	Libre
HLA status A Allele 2	cytoG abnorm	Caractère	Libre
HLA status B Allele 1	cytoG abnorm	Caractère	Libre
HLA status B Allele 2	cytoG abnorm	Caractère	Libre
HLA status C Allele 1	cytoG abnorm	Caractère	Libre
HLA status C Allele 2	cytoG abnorm	Caractère	Libre
Number.vials	Banking	Numérique	
Number.cells	Banking	Caractère	Libre

D2T.Test	Banking	Booléen	
Ptg.T.cell	CMF STI	Numérique	
Ptg.NK.cell	CMF STI	Numérique	
Ptg.monoc.cell	CMF STI	Numérique	
Dara.TTT	TTT	Booléen	
Post.dara	TTT	Caractère	Libre
BCMA.TTT	TTT	Booléen	
Clinical.history	TTT	Booléen	
Clinical.history.descr	TTT	Caractère	Libre
Comment		Caractère	Libre

2. Moteur de recherche et interface

Un moteur de recherche permettra d'effectuer des recherches parmi les échantillons comme parmi les patients, les modèles cellulaires (lignées ou cellules primaires) ou les hémopathies. Les recherches pourront être multicritères. Le module de recherche devra permettre de rechercher un échantillon selon un modèle cellulaire, une hémopathie, un nombre d'ampoules, une translocation, etc.

L'interface sera sous forme d'un site web utilisé en local. Elle permettra de séparer les deux modèles cellulaires (lignées et cellules primaires) avec des onglets et séparer l'édition de l'extraction. Des listes déroulantes permettront de sélectionner une des hémopathies malignes et d'afficher en fonction du choix de l'hémopathie, les différents critères (filtres) liées aux échantillons de cette hémopathie (**affichage dynamique**).

L'outil en ligne Biomart du site ENSEMBL (<https://www.ensembl.org/biomart/martview/7ba04e7312496fc9adcf89aedc594458>), dédié à l'édition de séquences d'ADN selon divers critères de sélection, est un exemple pertinent pour la création de l'interface. En effet, l'outil sous forme de page web, permet de sélectionner un jeu de données « dataset » dans un premier temps qui correspondrait à nos hémopathies malignes. Selon le jeu de données choisi, il affiche divers filtres permettant de faire la sélection, et différents attributs à afficher lors de l'édition. Les filtres et attributs sont affichés dans différentes catégories, à l'aide de listes déroulantes, case à cocher, etc.

3. Alimentation de la base de données

Les données de la base de données pourront être importées dans la base grâce à l'import d'un fichier au format CSV si l'importation est multiple, ou *via* un formulaire si l'importation d'un seul échantillon est réalisée. Dans les deux cas, l'importation se fera sur la page web.

4. Visualisation et extraction des données

La page web permettra de visualiser en direct les données selon les critères sélectionnés. Ces données pourront être extraies sous forme de table au format CSV.

Contacts

Elina Alaterre, Cheffe de projets, e.alaterre@diag2tec.com

Angélique Bruyer, Directrice de la société, a.bruyer@diag2tec.com

Tel : 04 67 55 77 73