**SOUTENANCE**

1.Introduction

Bonjour à toutes et à tous,

Je m’appelle Merlin Simsen Baratault, étudiant en BUT3 Sciences des données parcours EMS, en alternance au CEREDIH à l’hôpital Necker.

J’ai effectué cette alternance dans le cadre de ma formation, pendant la deuxième et troisième année. J’ai été sous la supervision de Mickaël Alligon et Nizar Malhaoui au CEREDIH, et de François Montestruc en tant que tuteur universitaire.

L’objectif de cette soutenance est de vous présenter premièrement le contexte dans lequel j’ai fait cette alternance, mes principales missions, les compétences que j’ai pu développer, et en quoi cette expérience a été déterminante pour mon projet professionnel.

2.Présentation du CEREDIH et du service Data

Le CEREDIH, ou Centre de Référence des Déficits Immunitaires Héréditaires, à été fondé en 2005 dans le cadre du Plan national maladies rares. Il gère le registre national des déficits immunitaires héréditaires, qui compte aujourd’hui plus de 10 000 patients, ce qui en fait la base la plus importante en nombre de patients sur ces maladies rares.

Son rôle est multiple :

* Améliorer la connaissance épidémiologique
* Soutenir la recherche médicale
* Et contribuer directement ou indirectement à la prise en charge et la qualité de vie des patients

Le service Data où j’ai travaillé, joue un rôle clé car il assure la qualité, la structuration, et l’analyse des données, contrairement ou groupe des ARC (attachés de recherche clinique), qui elles collectent les données des patients dans les hôpitaux, nous au service data on fait le nettoyage, la gestion, les contrôles qualité, les analyses statistiques, et la production de rapports.

Concrètement, j’ai occupé un poste de datamanager et « biostatisticien », en soutien direct du statisticien principal, et en lien direct avec les ARC et les médecins.

3.Aperçu global de mes missions

Pendant ces deux années, j’ai eu trois grandes missions, ou trois grands projets.

Le premier était la refonte complète du code de Data Management, le deuxième a été la modernisation du rapport d’activité, et le troisième à été la rédaction d’un article méthodologique sur le modèle de Cox.

Je vais commencer par détailler les deux premiers projets, plus techniques, puis je mettrai l’accent sur la troisième, qui m’a permis d’approfondir mes compétences méthodologiques et scientifique.

4.Mission 1 – Refonte du Data Management

Ma première mission a été de réécrire entièrement le script de Data Management, qui prépare la base avant toute analyse.

Le script initial faisait près de 9 000 lignes. Et il fallait environ deux heures pour le faire compiler, et il s’interrompait souvent en cas d’erreur dans les données, de nouvelle variable ou nouvelle modalité. Cela posait plusieurs problèmes : une difficulté à réagir rapidement à une demande, blocage en cas d’erreur, qui obligeait à tout relancer, et une pèrte de temps et manque de fiabilité.

Les objectifs de ces projets étaient donc de réécrire ce code, afin de gagner en rapidité de compilation, d’améliorer la robustesse face aux erreurs, et de le structurer afin de le rendre plus lisible et maintenable.

Pour faire cela j’ai commencé par lire et analyser le code existant, et d’identifier les parties les plus coûteuse en termes de temps de compilation. Nous avons au final trouvé que ce qui consommait le plus de temps était les boucles for, qui pour la plupart des créations de nouvelles variables, parcourais la base ligne par ligne et prenait donc beaucoup de temps à répétition.