

---

# Création d'un module ImageJ destiné à l'analyse cytogénétique

« Rapport de stage »

---



Stage du 11 avril au 17 juin 2022

Dans le cadre du stage de fin de formation obligatoire  
du DUT informatique option génie logiciel

Formation assuré par l'IUT de Lannion – Département Informatique

---

Tuteur : M. Christian JAULIN

Enseignant référent: M. Ludovic LIETARD



# Création d'un module ImageJ destiné à l'analyse cytogénétique

« Rapport de stage »

Stage de fin de formation obligatoire

DUT informatique option génie logiciel

Rapport du 11 avril au 17 juin 2022

IUT de Lannion – Département Informatique

Irset

Tuteur : M. Christian JAULIN

Enseignant référant : M. Ludovic LIETARD



---

# Remerciements

Voici une page dédiée à ceux que je remercie et que j'ai rencontrés durant mon stage.

Tout d'abord, je tiens à remercier tout le personnel de l'Irset pour son accueil et les nombreuses connaissances qu'ils ont partagés avec moi tout au long de ma période de stage.

Ensuite, je tiens à remercier particulièrement mon tuteur de stage M. Christian JAULIN pour m'avoir proposé ce stage au sein de l'Irset ainsi que pour m'avoir fait confiance. Je tiens à remercier aussi le reste des membres de l'équipe 4 de l'Irset. C'est à dire Mme. Laura MAGNAGHI JAULIN, Mme. Fatima SMAGULOVA et les 6 étudiants stagiaires que j'ai rencontrés, pour leurs accueils, leurs bienveillances ainsi que leurs sympathies.

Enfin, je tiens à remercier M. Thierry PECOT pour son workshop\*, son partage et son aide qui m'ont initié au deep learning\* pour les biologistes.

# Sommaire

Remerciements.....	2	3.1. Gestion de projet.....	18
Sommaire.....	3	3.2. Outils utilisés.....	20
Introduction.....	4	3.3. Formations.....	22
1. La structure d'accueil.....	5	4. Réalisation technique.....	25
1.1. L'institut de recherche de santé, environnement et travail (Irset).....	6	4.1. Étude des besoins.....	26
1.1.1. Présentation générale.....	6	4.2. Conception logiciel.....	26
1.1.2. Localisation de l'Irset.....	7	4.2.1. La conception.....	27
1.1.3. Caractéristique de l'Irset.....	8	4.2.2. La programmation.....	29
1.1.4. Environnement de l'Irset.....	10	4.2.3. Les tests.....	32
1.2. Équipe « Méiose, épigénétique et reproduction » (Mer).....	11	4.3. Environnement de travail.....	33
1.3. Objet de la mission de stage.....	12	Conclusion.....	34
2. L'analyse cytogénétique, son besoin et l'existant.....	13	Résumé.....	35
2.1. Contexte.....	14	Abstract.....	37
2.2. Détail de la mission.....	15	Termes techniques et sigles utilisés.....	38
2.3. Aspect juridique.....	16	Bibliographie.....	42
3. Organisation du projet.....	17	Index des figures et dessins.....	43
		Annexe.....	44

# Introduction

Ce document est un rapport de stage.

Ce stage se terminera le 15 juillet 2022. Ce document concerne la période de 10 semaines du 11 avril 2022 au 17 juin 2022 dans le cadre de ma formation en institut universitaire de technologie, en vue d'obtenir le diplôme universitaire de technologie (DUT) en informatique option génie logiciel.

L'objectif de ce stage est de m'introduire dans le monde professionnel et de prendre conscience de son fonctionnement dans le but de mettre en pratique les connaissances et les compétences que j'ai acquises durant mon cursus. De plus ce stage me permet d'acquérir de nouvelles connaissances.

J'ai eu la chance d'intégrer l'Irset et de réaliser mon stage au sein de cette administration publique. J'ai effectué ce stage au sein des locaux de l'équipe 4 sur le Campus de Villejean à Rennes. J'ai choisi cette structure d'accueil pour le sujet de stage proposé. Celui-ci étant en lien avec l'analyse d'images, une des branches de l'imagerie numérique, domaine dans lequel je souhaite poursuivre mes études.

On m'a chargé de la mission de concevoir et mettre en place un logiciel répondant aux deux problématiques suivantes : 1/ Comment automatiser l'établissement de l'index mitotique d'une population cellulaire par reconnaissance de forme des chromosomes condensés?, 2/ Comment mesurer de la proportion de chromatides sœurs non attachées dans une population cellulaire par évaluation de la distance entre centromère appariés ?

Dans ce rapport, tout d'abord je présenterai la structure d'accueil pour exposer le contexte du stage. Ensuite, je décrirai l'objectif et le contexte de la mission du stage ainsi que l'existant. Puis, j'exposerai l'organisation de mon projet et les outils utilisés. Enfin, je finirai par la réalisation technique sans oublier de décrire les problèmes rencontrés et les solutions développées.

# 1. La structure d'accueil

## 1.1.L'institut de recherche de santé, environnement et travail (Irset)

### 1.1.1.Présentation générale

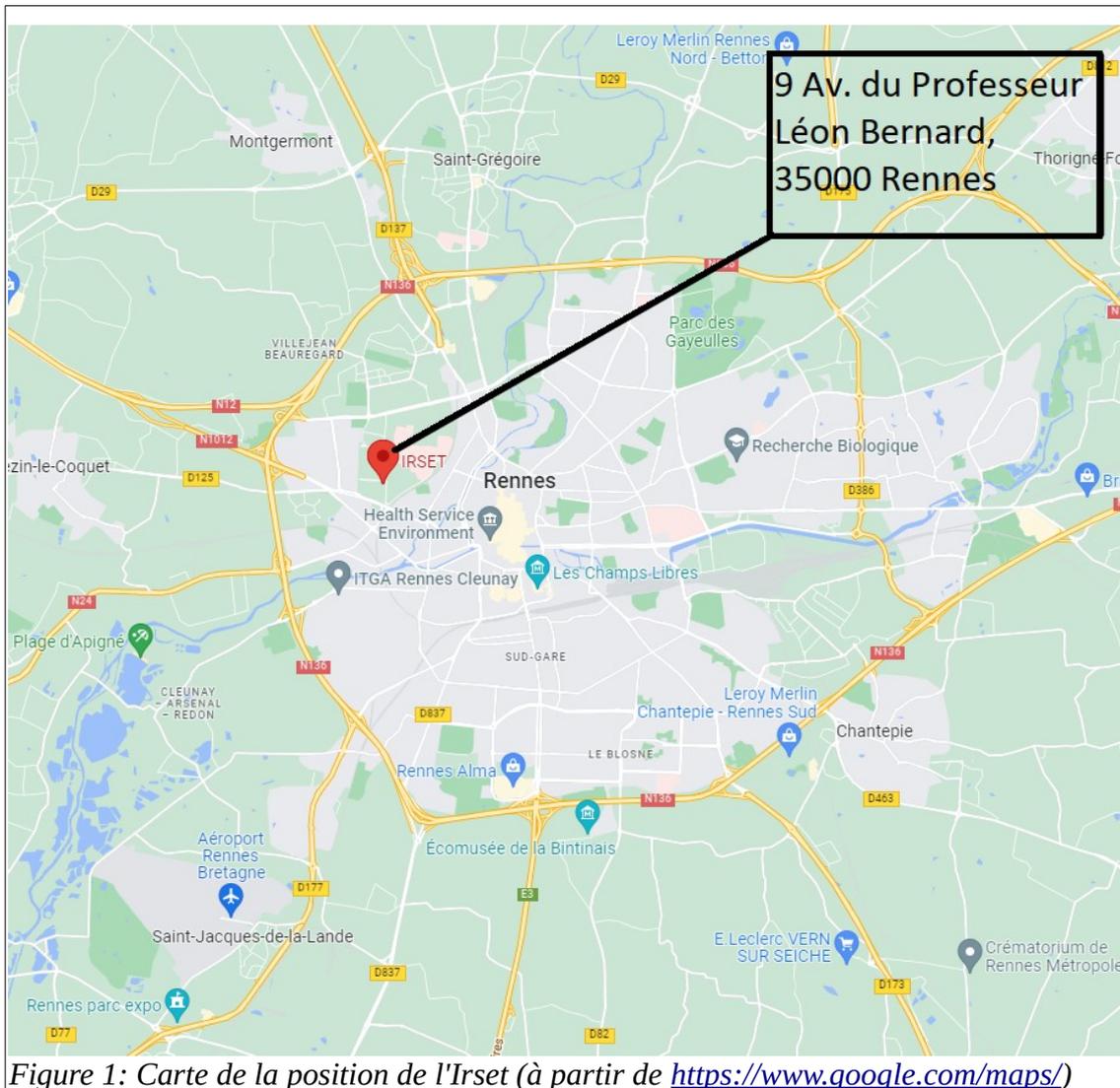
L'Irset ou l'institut de recherche en santé, environnement et travail est une unité mixte de recherches (UMR\_S 1085) sous la triple tutelle de l'Institut national de la santé et de la recherche médicale (Inserm), de l'Université de Rennes 1 et de l'École des hautes études en santé publique (EHESP). L'Irset est en partenariat avec le centre national de la recherche scientifique (CNRS), les centres hospitalier universitaire (CHU) de Rennes, Angers et Pointes-à-Pitre, l'Université d'Angers et l'Université des Antilles.

L'Irset a été créée en 2009 puis labellisée par l'Inserm en 2012. Fin septembre 2016, l'Irset a installée une partie de ses équipes dans le nouveau bâtiment de recherche créé en février 2016 par l'EHESP. Ce bâtiment est situé sur le campus santé de la ville de Rennes proche des locaux de l'EHESP sur le campus de Villejean.

L'Irset est l'un des plus grand pôles internationaux de recherche sur les risques liés à l'environnement et au travail. Ses missions sont au nombre de quatre. La première mission est d'étudier les processus biologiques et les facteurs environnementaux - physiques, chimiques, biologiques, professionnels et socioculturels - qui influencent la santé humaine. La seconde mission est de répondre aux demandes des autorités gouvernementales, nationales et régionales, et des agences de santé en charge de la santé publique, de la santé environnementale, de la santé au travail et de la sécurité des médicaments, en leur fournissant des données scientifiques et en contribuant à des travaux d'expertise. La troisième missions est de contribuer par la recherche à l'élaboration des plans nationaux, notamment dans les domaines de la santé publique, de la santé au travail, de la santé environnementale et de la santé au travail. La quatrième et dernière mission est de répondre aux sollicitations des citoyens, des associations non-gouvernementales, des organismes intermédiaires et des médias.

### 1.1.2. Localisation de l'Irset

L'Irset possède des locaux au sein de la ville de Rennes (35 000). J'ai fait mon stage dans leur bâtiment au 9 Avenue du Professeur Léon Bernard indiqué sur la carte ci dessous (Figure 1).



### 1.1.3. Caractéristique de l'Irset

La grille de caractérisation de l'organisation en Annexe 1 caractérise l'Irset. L'Irset est de nationalité française. Cette organisation est un établissement public à but non lucratif, c'est un établissement public à caractère scientifique et technologique (EPST). Les EPST sont des personnes morales de droit public dotées de l'autonomie administrative et financière et possédant un statut juridique d'établissement public.

L'Irset répond au besoin sociétal d'une demande de progrès en santé humaine. L'activité principale de l'Irset étant la recherche fondamentale et appliquée dans le domaine du biomédical, son secteur d'activité concerne les services non marchands, soit le secteur tertiaire. La finalité de l'Irset est d'assurer des missions de service public, il s'agit d'investir le champ de la recherche biomédicale fondamentale et appliquée dans le domaine de la santé humaine. Son objectif est donc d'améliorer et d'apporter des connaissances en santé humaine. On notera que dans la pratique, la performance d'un organisme de recherche s'évalue principalement via la qualité et le nombre des publications scientifiques de ses chercheurs.

L'Irset est composée d'environ 300 personnes. Ces personnes sont sur plusieurs postes. Il y a 27 chercheurs EPST de l'Inserm ou du CNRS, 40 enseignants-chercheurs de l'Université de Rennes 1, 52 ingénieurs, techniciens et administratifs, 50 hospitalo-universitaires, 63 doctorants et enfin 12 chercheurs post-doctorants.

Le Président de l'Inserm est nommé par décret en Conseil des ministres pour une durée de 4 ans renouvelable une fois. Il s'appuie sur les délibérations d'un Conseil d'Administration composé de 22 membres nommés par les ministères de tutelle (santé et recherche). Il est conseillé par un Conseil Scientifique composé pour moitié de membres nommés et pour moitié de membres élus par le personnel. Les commissions scientifiques spécialisées (CSS), au nombre de 7, participent à l'évaluation des activités de recherche, au recrutement des chercheurs et à leur promotion. Chaque CSS est spécialisée dans un grand domaine scientifique. Ces commissions scientifiques sont composées pour moitié de membres élus par le personnel et pour moitié de membres nommés par président de l'Inserm.

Comme indique l'organigramme de l'Irset en Annexe 2, les unités de recherche regroupent plusieurs équipes de recherche (11 pour l'Irset) et sont dirigées par un Directeur d'Unité qui s'appuie sur un comité de direction et un Conseil de Laboratoire. La réalisation d'un projet de recherche repose généralement sur une équipe de recherche entre 3 et 30 personnes, éventuellement en collaboration avec une ou plusieurs autres équipes. Au sein d'une équipe, le mode de coordination dépend de nombreux facteurs tels que la personnalité et l'expérience du responsable d'équipe et des membres de l'équipe, le besoin d'expertise extérieure, le caractère standardisé ou non des expériences. En règle générale, la conduite d'un projet de recherche repose sur une combinaison de différents modes de coordination. En soit, il y a supervision directe dans le cas du projet d'un doctorant placé sous la supervision de son directeur de thèse. Sinon il y a standardisation des tâches si il y a des travaux répétitifs nécessitant une reproductibilité normée. Sinon il y a standardisation des compétences lorsque l'équipe est habituée à travailler ensemble. Enfin, il y a ajustement mutuel dans le cas où, compte tenu des compétences de chacun et de la confiance mutuelle, les contributions se font par contact direct plutôt que via une supervision hiérarchique. C'est le cas lors de coopérations internationales entre plusieurs équipes de recherche.

En terme de budget, l'Irset dispose d'environ 10M€ dont 10% proviennent de la dotation d'État et 90% de contrats externes et de financements sur appels à projets. Cet apparent déséquilibre dans la répartition de l'origine des ressources entre l'Inserm et l'Irset s'explique par le fait que les traitements et salaires des personnels est en majorité financé par les organismes de tutelles Inserm, Université ou Hôpital sur dotation d'État tandis que les coûts de fonctionnement et d'entretien sont principalement assurés sur fonds propres. Pour information, le budget de l'Inserm correspond à un tiers de ressources propres contre deux tiers de dotations de l'État.

### 1.1.4. Environnement de l'Irset

Le macro environnement de l'Irset a été diagnostiqué à l'aide du modèle de PESTEL en Annexe 3. L'environnement politique est clairement celui qui, via l'investissement public, impacte de façon dominante la capacité du pays à produire une recherche de qualité. Il est constaté, depuis une vingtaine d'années, une désaffection croissante des étudiants pour les matières scientifiques et il devient difficile de recruter des doctorants performants. La recherche biomédicale est extrêmement dépendante des innovations technologiques. L'environnement écologique est un sujet d'étude de l'Irset, mais n'est pas une contrainte. Les fonctionnaires de l'Inserm sont soumis au droit public tandis que les agents non titulaires sont embauchés, selon les missions, sur des contrats de droit public ou privé. Les nouveaux entrants à l'Irset ne représente pas une menace pour l'organisation bien qu'il puisse exister une forme de concurrence qui évolue avec l'arrivée de nouveaux acteurs.

Le micro environnement de l'Irset a été diagnostiqué à l'aide du modèle de Porter en Annexe 4. La concurrence en matière de recherche est représentée par les laboratoires qui travaillent sur le même sujet et risquent donc d'être les premiers à publier un résultat que l'on s'efforce d'obtenir. Les laboratoires se fournissent auprès d'entreprises privées, le matériel informatique est soumis à un appel d'offre national, pour le reste une mise en concurrence des fournisseur est possible. L'État intervient dans le sens où l'accumulation des tâches administratives liées à son désengagement – et donc l'augmentation du temps passé à répondre aux appels d'offre - conduit à limiter le temps que les chercheurs peuvent consacrer à leur activité scientifique.

L'étude de marché est globalement sans objet pour la recherche fondamentale. La demande est une demande de progrès technologique et social de la part de l'ensemble de la société et l'offre est la production de nouvelles connaissances scientifiques qui pourront (ou pas) se traduire, par exemple, par la mise à disposition de nouveaux traitements ou la création de nouvelles normes sanitaires.

En Annexe 5, se trouve la synthèse du diagnostic de l'Irset.

## 1.2.Équipe « Méiose, épigénétique et reproduction » (Mer)

L'équipe nommée « Méiose, épigénétique et reproduction », « Meiosis, epigenetics and reproduction » en anglais, ou encore MER est la quatrième équipe de l'Irset. J'ai effectué mon stage au sein de cette équipe. Elle est composée de trois membres permanents, à savoir sa responsable Mme. Fatima SMAGULOVA (Inserm CRCN\*), ainsi que de mon tuteur M. Christian JAULIN (Inserm DR2\*), puis de Mme. Laura MAGNAGHI JAULIN (CNRS CRCN). Enfin les autres membres, tous étudiants, que j'ai rencontrés durant la durée de mon stage sont au nombre de six.

L'objectif de l'équipe 4 est d'élucider les effets des toxiques environnementaux sur les mécanismes de régulation épigénétique. L'équipe 4 possède trois axes de recherche. Le premier axe porte sur les effets transgénérationnels promus par des substances toxiques. Le second axe porte sur l'étude des mécanismes épigénétiques impliqués dans la méiose. Et enfin le troisième axe porte sur l'impact des composés toxiques sur les mutations génomiques.

Hormis moi, toute l'équipe 4 de l'Irset est composée de membres permanents et de stagiaires travaillant directement dans le domaine de la biologie. Étant un stagiaire en formation informatique, je suis isolé dans un environnement qui me semble plutôt inconnu.

## 1.3. Objet de la mission de stage

Au cours de mon stage, j'ai étudié et travaillé sur l'analyse d'images de microscopie préalablement acquises.

Pour atteindre ce but, j'ai appris l'utilisation du logiciel ImageJ\* ainsi que du logiciel Fiji\*, ainsi que son développement de plugins. Je me suis ensuite initié puis auto-formé au deep learning et à ses méthodes d'entraînement\* pour créer un modèle, ainsi qu'aux méthodes d'utilisation de ce modèle.

En tant que stagiaire, j'ai utilisé de nombreuses connaissances et compétences acquises durant ma formation informatique à l'IUT de Lannion, et j'ai aussi acquis de nouvelles connaissances et compétences en me documentant et en apprenant dans une démarche d'autoformation.

De mon point de vue, l'objectif de ce stage était d'une part de développer des connaissances dans le domaine de l'analyse et le traitement d'images, et, d'autre part, de découvrir la pratique de la programmation et l'utilisation des langages et outils dans un environnement professionnel.

## 2. L'analyse cytogénétique, son besoin et l'existant

## 2.1. Contexte

La cytogénétique est l'étude des chromosomes des cellules. Nous allons donc suivre cette analyse cytogénétique. L'épigénétique correspond à l'étude des caractères transmissibles sans modification de la séquence d'ADN, nous nous intéresserons plus particulièrement à la division cellulaire au cours de laquelle ces caractères sont transmis. La division cellulaire, appelé mitose, conduit une cellule mère à se diviser en deux cellules filles. À l'intérieur de cette cellule mère, lors de la mitose, chaque chromosome est alors formé de deux éléments nommés chromatides sœurs unies par le centromère. Ce centromère est lui composé de deux complexes de protéines nommées kinétochores

Le dénombrements et l'identification de ces éléments (chromatides et kinétochores) sont des tâches effectuées manuellement. Et cela pose plusieurs problèmes : il y a immobilisation d'un membre pour effectuer les tâches qui sont répétitives et nombreuses. Ainsi un problème de temps s'ajoute. Il s'agit donc d'un point à travailler sans tarder

L'automatisation informatique est une solution qui permet, via un processus de création d'un logiciel, de réduire les interventions manuelles. Actuellement étudiant en informatique, je suis une personne prête à recevoir la mission de création.

## 2.2. Détail de la mission

Le sujet de stage en Annexe 6 détaillait le propos de la mission donnée. De plus en amont du début du stage, j'ai eu un briefing\* de la part de Mme. Laura JAULIN MAGNAGHI sur ce que l'équipe attendait.

Ainsi, la mission est de remplir les besoins issus des problèmes énoncés en 2.1. Contexte : Le nom de cette mission est : « Création d'un module ImageJ destiné à l'analyse cytogénétique ». Il s'agit de créer un logiciel, plugin, faisant office de module pour le logiciel open source\* ImageJ. Ce plugin est destiné à l'analyse d'images de microscopie optique. Il doit remplir deux fonctions. La première fonction consiste à établir l'index mitotique d'une population cellulaire par reconnaissance de forme des chromosomes condensés. La seconde fonction consiste à mesurer la proportion de chromatides sœurs non attachées dans une population cellulaire par évaluation de la distance entre centromère appariés.

Le développement de ce module a pour but de permettre d'automatiser une tâche jusque là effectuée manuellement, et ainsi que de gagner du temps.

## 2.3.Aspect juridique

Durant le projet, le développement des programmes ce sont fait à l'aide de bibliothèques libre et gratuite sous licence GNU General Public License v3.0 (GPL v3.0\*) et sous licence Apache License 2.0\*. Dans ce même sens, les programmes développés sont libres et gratuits sous licence GPL v3.0. De plus le logiciel est considéré comme œuvre composite due au fait de l'utilisation des bibliothèques et de logiciels déjà développés par d'autres personnes. Les droit patrimoniaux sont applicable au logiciel.

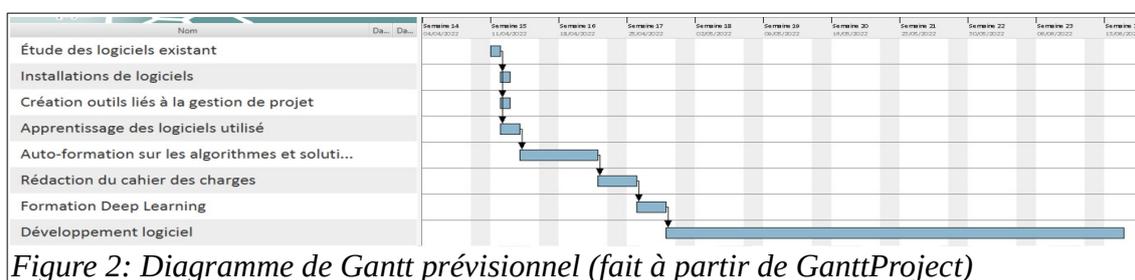
Il n'y a pas eu de contrainte confidentialité retenue. D'autre part le RGPD\* n'entre pas en vigueur dans le projet.

## 3. Organisation du projet

## 3.1. Gestion de projet

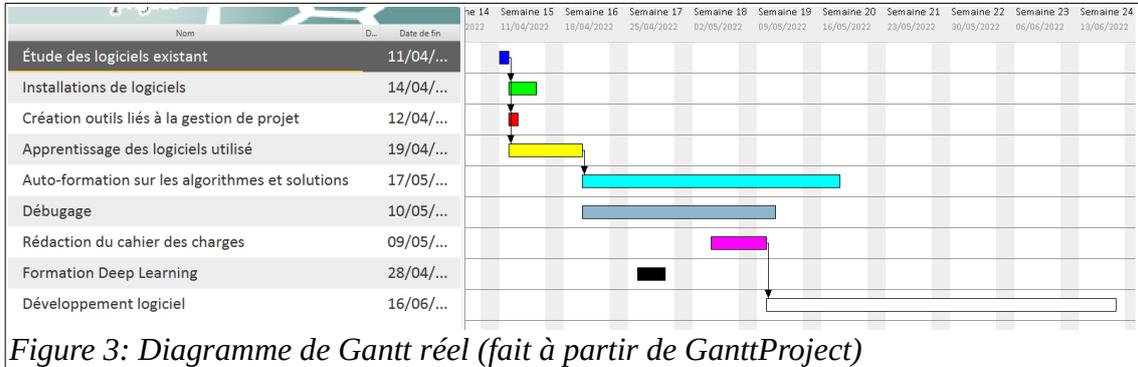
Un journal de bord en ligne que j'ai établi avait pour but de sauvegarder toutes traces de mes actions lorsque j'écrivais dedans. J'ai ensuite découpé le projet en différentes tâches dans un diagramme de structure de décomposition du travail (WBS\*), voir Annexe 7.

Rapidement, avec les différentes tâches, j'ai établis un diagramme de Gantt prévisionnel que vous pouvez trouver ci-dessous (Figure 2) ou en Annexe 8 pour pouvoir organiser globalement mes tâches dans le temps et me donner une idée de leur état d'avancement. J'avais planifié des tâches consécutives, c'est à dire que l'exécution de la tâche précédente est nécessaire pour la tâche qui la suit.



Durant le reste du stage j'ai tenu un autre diagramme de Gantt, il s'agit du diagramme réel qui se trouve ci dessous (Figure 3) ou en Annexe 9. Ce diagramme retrace la réalité des tâches dans le temps. D'autres tâches qui n'apparaissent pas sur le diagramme prévisionnel sont apparues sur le diagramme réel. De plus, certaines tâches du diagramme réel ne sont plus consécutives. Les raisons de ces modifications sont, d'une part que j'avais fait une mauvaise estimations de temps des tâches dû au fait que je n'avais pas réellement d'expérience dans chacune de ces tâches, et d'autre part parce qu'il y a eu des obstacles imprévus allongeant parfois le temps consacré à une tâche. Ces obstacles ont été des installations d'ordinateurs non adaptées

ou manquant de puissance, une formation d’outils et de logiciels non calculée, ainsi qu’une formation dans le domaine du deep learning, c’est à dire l’étude puis l’expérimentation de celui-ci.



J’ai rédigé un cahier des charges définissant le projet, son contexte, sa finalité, les contraintes, les besoins, puis les prestations attendues. Ce cahier des charges avait pour but de servir de fil rouge durant le projet.

## 3.2. Outils utilisés

Conformément au sujet préalablement établi en Annexe 6, le projet s'est effectué sur un ordinateur récent sous un environnement Windows 10\*. Il a été développé en langage Java\* fonctionnant avec le logiciel open source Fiji.

Suite à l'étude des logiciels existant et à l'analyse liée au bon déroulement du projet, j'ai conclu qu'il était nécessaire d'avoir une connexion internet.

Côté gestion de projet, les outils de bureautique en ligne Drive\*, Docs\*, Sheets\* et Draw\* que propose Google ont été utilisés. Ils m'ont permis de m'organiser, de laisser un historique et une trace écrite de ce que j'ai fait chaque jour. Je les ai choisis parce qu'ils sont des outils en ligne, contrairement à la suite Office365\*. J'ai aussi utilisé le site planhammer\* et le logiciel GanttProject\* afin de planifier les tâches et activités prévues. J'ai aussi utilisé le logiciel Bouml\*.

L'outil de développement principal était le logiciel et environnement de développement intégré (IDE\*) nommé Eclipse. J'ai fait le choix de cet IDE parce que c'est un logiciel que j'ai déjà utilisé pour développer et programmer en langage Java, contrairement à d'autres IDE tel que IntelliJ\*. De plus Eclipse possède les outils Maven et Gradle qui permettent d'automatiser la construction d'un projet sur Eclipse, en l'occurrence des projets Java. Bien que Eclipse possède l'outil Git\* intégré, j'ai préféré utiliser ce logiciel indépendamment de l'IDE parce que, auparavant, j'avais seulement utilisé Git en ligne de commande. Cet outil a été utilisé pour initialiser et mettre à jour des dépôts sur le site web GitHub\*. Ces dépôts me permettent de stocker l'avancement de mon travail ainsi que de garder un historique de ce qui a été fait, à l'instar de GitLab\*.

LabelImg\* est un logiciel que j'ai brièvement utilisé, il permet d'annoter et de labelliser des images. J'ai choisi ce logiciel parce que contrairement au logiciel SuperAnnotate\*, il est gratuit.

Le module développé est un plugin qui utilise plusieurs bibliothèques\*. La bibliothèque Java open source IJ-OpenCV\* permet une utilisation de la bibliothèque OpenCV\*, initialement développée en C/C++\*, elle ajoute la possibilité d'utiliser des algorithmes\* de vision par

ordinateur et des algorithmes d'apprentissage automatique aussi appelé machine learning. La bibliothèque Java open source Deep Java Library (DJI\*) permet une utilisation de frameworks comme MXNet\* et une utilisation de deep learning, une branche du machine learning. Les bibliothèques d'exécution Microsoft C et C++ sont indispensables pour faire fonctionner la bibliothèque DJL dans un environnement sous Windows 10.

## 3.3. Formations

Le logiciel ImageJ, le logiciel Fiji, le deep learning, les outils Maven et Gradle ainsi que les bibliothèques IJ-OpenCV et DJL m'ont demandé un temps de formation total de près d'un mois, en comptant l'étude, l'installation, puis la manipulation. J'ai rencontré des difficultés dans les installations ainsi que dans les manipulations, ce qui m'avait beaucoup ralenti.

Je me suis formé en priorité au logiciel ImageJ. J'ai d'abord appris à l'utiliser, puis je me suis formé ) la façon dont les biologistes l'utilisent dans le cadre d'une analyse d'image, et enfin j'ai étudié le développement de plugin et macro. Je me suis auto-formé durant près d'une semaine grâce au site web d'ImageJ (<https://imagej.nih.gov/ij/>). Celui-ci regroupe la documentation, des plugins et surtout des ressources liés au logiciel ImageJ pour les développeurs.

J'ai été initié au deep learning lors d'un workshop sur trois jours proposé par Biosit\* dans le cadre de l'utilisation du deep learning pour biologistes. J'ai appris différentes méthodes de deep learning et j'ai aussi appris l'utilisation de différents outils et logiciels, la plupart en langage Python . Je me suis par la suite auto-formé sur l'utilisation du deep learning en langage Java.

En ce qui concerne les bibliothèques IJ-OpenCV et DJL, elles m'ont posé plusieurs difficultés allongeant considérablement leur temps de formation initialement estimé. Ainsi à propos de l'auto-formation sur les algorithmes et solutions, il y a près de 4 semaines de différence entre l'Annexe 8 et l'Annexe 9.

L'installation des bibliothèques n'était pas évidente, il fallait utiliser les outils Maven et Gradle, qui m'étaient plus au moins nouveaux. J'ai donc commencé par apprendre à utiliser ces outils dans l'IDE Eclipse, à l'aide de tutoriels.

Après l'installation de la bibliothèque IJ-OpenCV, celle-ci avait une intégration difficile au logiciel ImageJ. Je me suis dirigé vers le logiciel Fiji qui implémente déjà cette bibliothèque. Je me suis auto-formé sur ce logiciel avec leur site web <https://imagej.net/>. Il regroupe la

documentation, des plugins et des ressources pour les développeurs pour le logiciel Fiji ainsi que pour le logiciel ImageJ.

Lors de ma formation à la bibliothèque DJL j'ai eu les plus grosses difficultés à surmonter. Dans un premier temps, des problèmes liés aux composants des ordinateurs sur lesquels je travaillais - tel que le manque d'espace mémoire vive (RAM\*) - interrompaient l'installation de la bibliothèque DJL via la construction Maven et Gradle dans les projets Java d'Eclipse. Dans un second temps, après la construction de DJL l'encodage des caractères n'était ni homogène ni correct, ce qui rendait inutilisable la bibliothèque DJL. Par exemple sur la capture d'écran ci-dessous (Figure 4) à la ligne 150, la chaîne de caractère en vert « ' â-^ ' » est censé être un rectangle vertical plein : █.

```
107
108     currentPercent = percent;
109     StringBuilder sb = new StringBuilder(100);
110     sb.append("\n").append(message).append(':');
111     for (int i = 0; i < 12 - message.length(); ++i) {
112         sb.append(' ');
113     }
114     sb.append(String.format("%3d", percent)).append("% |");
115     for (int i = 0; i < TOTAL_BAR_LENGTH; ++i) {
116         if (i <= percent * TOTAL_BAR_LENGTH / 100) {
117             sb.append(progressChar);
118         } else {
119             sb.append(' ');
120         }
121     }
122     sb.append('|');
123     if (additionalMessage != null) {
124         sb.append(' ').append(additionalMessage);
125     }
126     if (percent == 100) {
127         System.out.println(sb);
128     } else {
129         System.out.print(sb);
130     }
131 }
132
133 private String trimMessage(String message) {
134     int len = message.length();
135     if (len < 13) {
136         return message;
137     }
138     return message.substring(0, 4) + "... " + message.substring(len - 5);
139 }
140
141 private static char getProgressChar() {
142     if (System.getProperty("os.name").startsWith("Win")) {
143         return '=';
144     } else if (System.getProperty("os.name").startsWith("Linux")) {
145         String lang = System.getenv("LANG");
146         if (lang == null || !lang.contains("UTF-8")) {
147             return '-';
148         }
149     }
150     return 'â-^';
151 }
152 }
```

Figure 4: Imprime écran du logiciel Éclipse possédant le mauvais encodage

Enfin dans un quatrième et dernier temps j'ai repris des exemples et écrits d'autres programmes afin de comprendre l'utilisation et la manipulation de la bibliothèque DJL, c'est lors de l'exécution de ces programmes que j'ai eu d'autres problèmes. Le premier problème était une perte de rapidité due à la non utilisation de la carte graphique (GPU\*), actuellement toujours non résolue. Le second problème était un manque de mémoire vive (RAM), il n'était plus possible d'allouer de l'espace mémoire. Le troisième problème était de même nature que pour la RAM - une saturation du processeur (CPU\*). Il n'était plus possible d'allouer de l'espace mémoire. Le quatrième problème était dû à une mauvaise version du framework MXNet qui engendrait des erreurs. La majorité de ces problèmes a été résolue.

## 4. Réalisation technique

## 4.1. Étude des besoins

Après avoir bien pris connaissance des besoins énoncé en 2. L'analyse cytogénétique, son besoin et l'existant, le cahier des charges a été rédigé. Les livrables\* qui ont été définis sont un plugin, ainsi qu'une documentation de ce plugin. La conception d'un logiciel et la conception de plusieurs programmes et script ont été spécifiés pour aboutir à la conception du plugin.

Pour être plus clair, j'ai déclaré l'élaboration de scripts de conversion de format et d'extension d'images via Fiji qui permet la préparation des données d'entraînement. J'ai déclaré l'élaboration d'un programme qui permet d'entraîner un modèle de deep learning. Enfin, j'ai déclaré l'élaboration du logiciel qui permet l'interface homme machine (IHM\*), qui permet de réaliser l'inférence\* du deep learning et qui permet d'afficher les résultats.

## 4.2. Conception logiciel

La conception logiciel se déroule en trois phases, la première regroupe la spécification, la conception et la définition de l'architecture du logiciel. La réalisation est la seconde phase, elle fait référence à l'écriture et aux tests des programmes. Enfin la dernière phase est celle de la livraison, elle regroupe l'intégration, la validation ainsi que la documentation du logiciel. Ce rapport n'encadre pas la totalité de la durée de mon stage énoncé en Introduction. en conséquence, cette partie conception logiciel ne regroupe partiellement que les 2 premières phases.

### 4.2.1. La conception

Durant mon auto-formation j’ai commencé à réfléchir aux structures et architectures du logiciel et à les formaliser. Ainsi, Le plugin importe les bibliothèques IJ-OpenCV et DJL, il utilise un modèle de deep learning préalablement entraîné par un programme qui est aussi développé en important la bibliothèque DJL, comme le montre le diagramme de classe ci-dessous (Figure 5).

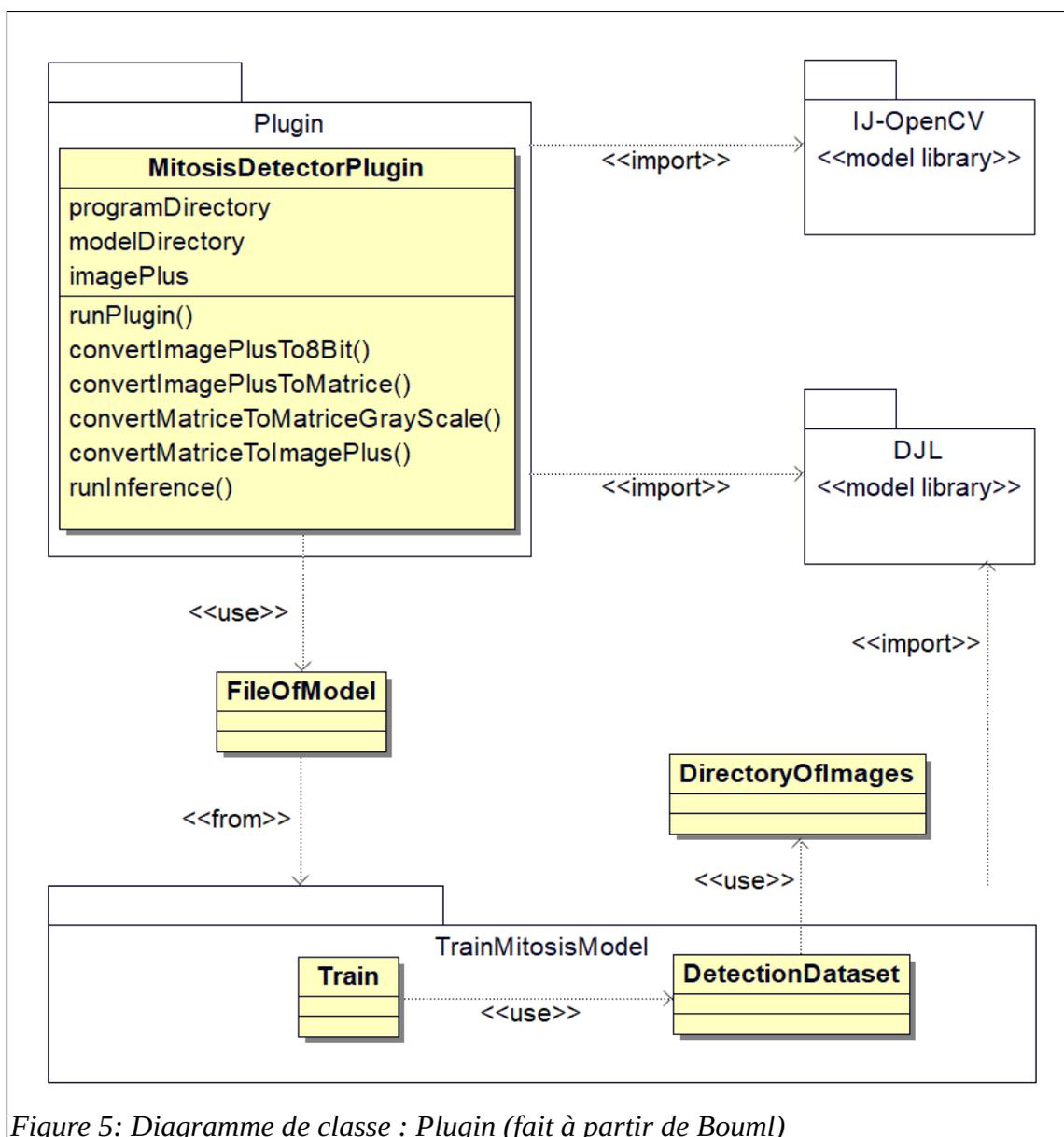
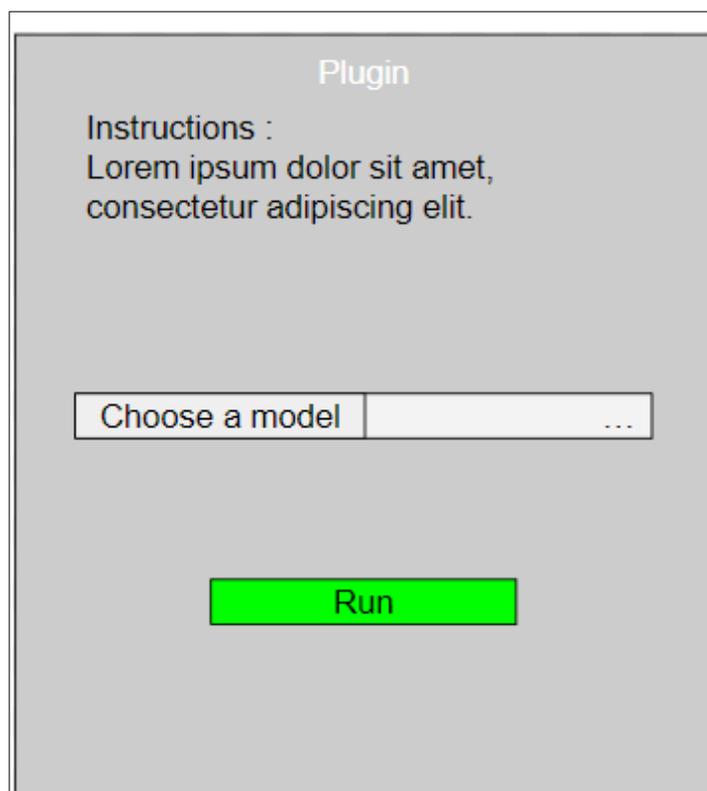


Figure 5: Diagramme de classe : Plugin (fait à partir de Bouml)

Bien qu'il n'y ai pas de charte graphique définie, afin d'avoir une idée visuelle du plugin, j'ai pensé très brièvement à une interface permettant le choix d'un modèle de deep learning. Il s'agit de la maquette ci-dessous (Figure 6).



*Figure 6: Maquette de l'interface homme machine (fait à partir de Google Draw)*

## 4.2.2. La programmation

Tout d'abord, j'ai rédigé et programmé - à l'aide du langage de macro ImageJ - des macros Fiji pour préparer les données de manière à ce qu'elles soient ensuite utilisables par mes autres programmes. En exemple, le script ci-dessous (Dessin 1) permet de sauvegarder à un endroit spécifique sous format png\* un certain nombre d'images spécifiques initialement ouverte sur le logiciel Fiji.

```
path = "C:\\Users\\Robin\\work\\eclipse-
workspaceinternship\\my_work\\src\\test\\
resources\\Photos Noco 09-05-2022\\
Hela_Noco_png\\"

i = 1
j = 1

while (i < 114) {
    selectWindow("Hela-Noco (" + i + ").zvi");
    saveAs("PNG", path + "Hela-Noco (" + i + ").png");
    close();
    i++;
}

print("1st part done, save in : " + path);

while (j < 185) {
    selectWindow("Hela ctrl (" + j + ").zvi");
    saveAs("PNG", path + "Hela-ctrl (" + j + ").png");
    close();
    j++;
}

print("2nd part done, save in : " + path);
```

*Dessin 1: Script en langage macro ImageJ*

Ensuite, je suis en train de développer petit à petit un programme permettant de créer un modèle entraîné. En fonction du modèle que je veux obtenir, ce programme possède des paramètres et des initialisations spécifiques. Il fonctionne en plusieurs étapes, ces étapes sont créées sous forme de méthodes que je crée ou que j'utilise depuis la bibliothèque DJL.

Avec le chemin d'accès, le programme doit récupérer les données. Suite à cela, le programme initialise le set de données comme le montre la méthode de classe que j'ai développée ci-dessous (Dessin 2), en l'occurrence il s'agit d'images.

```
private static ImageFolder initDataset
    (String datasetRoot) throws IOException,
        TranslateException {

    ImageFolder dataset = ImageFolder.builder()
        // retrieve the data
        .setRepositoryPath
            (Paths.get(datasetRoot))
        .optMaxDepth(10)
        .addTransform(new Resize(1384, 1030))
        .addTransform(new ToTensor())
        // random sampling
        .setSampling(BATCH_SIZE, true)
        .build();

    dataset.prepare();
    return dataset;
}
```

*Dessin 2: Méthode d'initialisation d'un set d'images*

Ce set de données est constitué d'images ainsi que d'annotations et de labels. Les images sont les objets principaux du set de données. Les labels et les annotations sont des coordonnées de points formant des boîtes englobantes ainsi que de noms, ils sont rangés dans des fichiers XML\* associés à chaque image et c'est au niveau de l'appel de la méthode « prepare() » qu'ils sont extraits. La

structure du programme est prête, de l'initialisation du modèle jusqu'à l'enregistrement du modèle entraîné dans un dossier dédié. En revanche je suis encore en pleine programmation de la méthode permettant l'entraînement du modèle.

Du côté du plugin Fiji, j'ai réalisé la méthode de choix du modèle, ainsi qu'une méthode de conversion d'image en un format spécifique pour ne pas avoir d'erreurs lors de l'inférence comme indiqué sur la méthode de classe ci-dessous (Dessin 3). Par ailleurs je suis en train de réaliser l'interface homme machine.

```
/* method to convert many colorspace to grayscale
   without interrupt */
public Mat convert2Gray(Mat matSrc, Mat matGray) {
    try {
        opencv_imgproc.cvtColor
            (matSrc, matGray,
             opencv_imgproc.COLOR_BGR2GRAY);
        return matGray;
    }
    catch (Exception e){}

    try {
        opencv_imgproc.cvtColor
            (matSrc, matGray,
             opencv_imgproc.COLOR_RGB2GRAY);
        return matGray;
    }
    catch (Exception e){}

    {...}

    try {
        opencv_imgproc.cvtColor
            (matSrc, matGray,
             opencv_imgproc.COLOR_BGR5552GRAY );
        return matGray;
    }
    catch (Exception e){}

    return matSrc;
}
```

*Dessin 3: Méthode de conversion d'image*

### 4.2.3. Les tests

J'ai réalisé des tests même si les programmes et logiciels ne sont pas terminés dans leur totalité. J'ai effectué des tests d'entraînement de modèles, et des tests d'intégrations entre le plugin et le logiciel Fiji. Je n'ai pas fini la programmation de la méthode de création du modèle que j'attends, en revanche je peux utiliser les exemples de méthodes d'entraînement de modèles fournis par la bibliothèque DJL. Le modèle de sortie ne correspond pas à celui que je cherche, mais les tests m'ont permis de confirmer que, actuellement, le programme fonctionne dans son ensemble. Ainsi comme vous pouvez le voir dans l'affichage de la console ci dessous (Dessin 4), il n'y a pas d'erreurs apparente.

```
[WARN ] - No matching cuda flavor for win found: cu92mkl/sm_75.
[INFO ] - Training on: cpu().
[INFO ] - Load MXNet Engine Version 1.7.0 in 0,049 ms.

Training:      7% |===                               |
  Accuracy:  _, SoftmaxCrossEntropyLoss:  _
Training:     100% |=====                           |
  Accuracy: 0,97, SoftmaxCrossEntropyLoss: 0,14

Validating:   27% |=====                           |
Validating:  100% |=====                           |
[INFO ] - Epoch 1 finished.
[INFO ] - Train: Accuracy: 0,97, SoftmaxCrossEntropyLoss: 0,13
[INFO ] - Validate: Accuracy: 0,98, SoftmaxCrossEntropyLoss: 0,11

Training:      7% |===                               |
  Accuracy: 0,97, SoftmaxCrossEntropyLoss: 0,14
Training:     100% |=====                           |
  Accuracy: 0,98, SoftmaxCrossEntropyLoss: 0,09

Validating:   27% |=====                           |
Validating:  100% |=====                           |
[INFO ] - Epoch 2 finished.
[INFO ] - Train: Accuracy: 0,98, SoftmaxCrossEntropyLoss: 0,09
[INFO ] - Validate: Accuracy: 0,98, SoftmaxCrossEntropyLoss: 0,10
[INFO ] - forward P50: 2505,591 ms, P90: 2558,794 ms
[INFO ] - training-metrics P50: 0,017 ms, P90: 0,023 ms
[INFO ] - backward P50: 5,053 ms, P90: 7,151 ms
[INFO ] - step P50: 14,155 ms, P90: 16,774 ms
[INFO ] - epoch P50: 180,290 s, P90: 180,290 s
```

Dessin 4: Affichage de console après test d'entraînement d'un modèle

## 4.3. Environnement de travail

Un ordinateur standard sous environnement Windows 10 n'est pas suffisant. Pour développer et procéder à l'exécution des méthodes d'entraînement, j'ai changé à trois reprises d'ordinateur. En effet une utilisation excessive de mémoires rendait inutilisables les machines. J'ai dernièrement monté un nouvel ordinateur et j'y ai ajouté 8 Gb de RAM supplémentaires, c'est actuellement dessus que je développe et teste mes programmes.

J'utilise beaucoup l'outil Git qui m'a permis d'enregistrer les différentes modifications effectués sur mon projet. Ainsi, bien que je programme localement, je travaille continuellement avec une connexion internet.

# Conclusion

Pour résumer, l'objectif était de créer un logiciel pour répondre à une demande d'automatisation d'analyse d'images. En conséquence, durant dix semaines j'ai réalisé la conception du logiciel et j'ai commencé à le développer. J'ai utilisé mes compétences en gestion de projet et mes compétence en programmation. J'ai également développé mes compétences en auto-formation, et, enfin, j'ai développé des connaissances en informatique et plus particulièrement en analyse d'images et en deep learning.

Mon travail n'est pas finis par manque de temps, c'est ainsi que, avec l'accord de mes encadrants, j'ai pris l'initiative de prolonger ma période de stage.

# Résumé

J'ai réalisé un stage de fin d'études dans le cadre de ma formation à l'institut universitaire de technologie (IUT) de la ville de Lannion, en vue d'obtenir un diplôme universitaire de technologie (DUT) en informatique option génie logiciel. J'ai réalisé ce stage de dix semaines au sein de l'institut de recherche en santé, environnement et travail (Irset), dans l'objectif de concevoir et développer un logiciel.

L'Irset est un établissement public à caractère scientifique et technologique (EPST). Créé en 2009, l'Irset possède des locaux dans la ville de Rennes au sein du campus de Villejean, sur le site l'école des hautes études en santé publique (EHESP). En effet, l'Irset est une unité mixte de recherche (UMR\_S 1085) de l'Inserm, de l'Université de Rennes 1 et de l'EHESP. L'Irset répond au besoin sociétal d'une demande de progrès en santé humaine L'Irset est composée d'environ 300 personnes parmi lesquelles se trouvent des chercheurs, des enseignants-chercheurs, de ingénieurs, des techniciens, des administratifs, des hospitalo-universitaires, des doctorants et des post-doctorants. L'Irset regroupe 11 équipes de recherche qui travaillent éventuellement en collaboration avec d'autres équipes. L'Irset est dirigé par un directeur d'unité qui s'appuie sur un comité de direction et un conseil de laboratoire.

J'ai effectué mon stage au sein de l'équipe 4, en anglais elle est nommée « Meiosis, epigenetics and reproduction » (MER). Lors de la durée de mon stage elle était composée de 3 membres permanents, et de 7 stagiaires. L'équipe 4 possède trois axes de recherche. Le premier axe porte sur les effets transgénérationnels promus par des substances toxiques. Le second axe porte sur l'étude des mécanismes épigénétiques impliqués dans la méiose. Et enfin le troisième axe porte sur l'impact des composés toxiques sur les mutations génomiques.

Au cours de mon stage j'ai étudié et travaillé sur l'analyse d'images de microscopie préalablement acquise.

La mission donnée correspondait à la création d'un module ImageJ destinée à l'analyse cytogénétique. Le développement de ce module à pour but d'automatiser une tâche répétitive jusque-là effectuée manuellement et ainsi gagner du temps.

Durant le projet de développement, les programmes ont été fait à partir de bibliothèques sous licences libres et gratuites. Les droits patrimoniaux sont applicable au logiciel.

Pour commencer le projet j'ai fait de la gestion de projet, j'ai établi un journal de bord en ligne permettant de sauvegarder les traces de mes actions. J'ai établis un diagramme de Gantt prévisionnel pour visualiser et m'organiser dans le temps. J'ai par la suite établi au fur et à mesure un diagramme de Gantt réel. J'ai rédigé un cahier des charges définissant le projet, son contexte, sa finalité, les contraintes, les besoins, puis les prestations attendues. Ce cahier des charges avait pour but de servir de fil rouge durant le projet.

J'ai utilisé différents outils et logiciels, à savoir un ordinateur récent sous Windows 10, le langage de programmation Java et le langage macro ImageJ, les bibliothèques C/C++, IJ-OpenCV et DJ les logiciels Eclipse, GanttProject, ImageJ, Fiji et LabelImg, les outils Git, Gradle et Maven et les outils de bureautique Drive, Docs, Sheets et Draw.

Je me suis également auto-formé sur l'utilisation de logiciels et de plusieurs outils cités ci-dessus. Je me suis auto-formé sur les bibliothèques IJ-OpenCV et DJL. En outre, suite au workshop sur le deep learning, je me suis auto-formé dans ce domaine. J'ai surmonté de nombreuses difficultés liées à des problèmes d'installations, des problèmes d'encodages et des problèmes de performances.

J'ai étudié les besoins, j'ai fait la conception du logiciel demandé, j'ai développé des scripts et des programmes et je les ai testés. Cependant par manque de temps je n'ai pas fini le projet de développement. J'ai donc demandé, et obtenu, la prolongation de la durée de mon stage pour un mois supplémentaire.

Pour conclure, j'ai conçu et commencé à développer un logiciel. J'ai utilisé mes compétences en gestion de projet, en programmation, et enfin j'ai développé des connaissances en informatique, en analyse d'images et en deep learning.

# Abstract

I did a ten-week internship at the Institute for Research Institute for Environmental and Occupational Health (Irset UMR\_S 1085), with the objective of designing and developing a software. Irset is a public scientific and technological institution in France in the city of Rennes. The Irset responds to the societal need for progress in human health. The given mission was to create an ImageJ module for cytogenetic analysis. The development of this module aims to automate a task that was previously done manually, and thus save time. To start the project I did some project management, I established an online logbook, a forecasted Gantt chart and a real Gantt chart. I have self-taught myself on the use of software and on the use of several tools and libraries. I wrote specifications defining the project, its context, its purpose, the constraints, the needs, and the expected services. So I have studied the needs, I have designed the requested software, I have developed scripts and programs and I have tested them. However, I encountered several and computational issues, whose solving quite time consuming and I could not finish the development project in the planned time frame. This is why I requested and obtained an extension of my internship. I used my skills in project management, in programming, and finally I developed knowledge in computer science and image analysis and in deep learning.

# Termes techniques et sigles

## utilisés

Algorithmes : En informatique, l'algorithme est une suite d'instructions et d'opérations décrivant un processus permettant de résoudre un problème.

Apache License 2.0 : Licence de logiciel libre et open source. Le texte de la licence doit être distribué avec la clause d'exclusion de garantie.

Bibliothèques : En informatique, la bibliothèque ou librairie logiciel est un ensemble de fonctions mises à disposition afin de pouvoir être utilisable par d'autres programmes sans avoir à les réécrire.

Biosit : Unité mixte de service reconnue par l'Université de Rennes 1, le CNRS et l'Inserm, Biosit est une structure de recherche en biologie et santé.

Bouml : Logiciel de création de diagrammes.

Briefing : Anglicisme décrivant une brève réunion d'information.

C/C++ : Basé sur le langage de programmation C, le C++ est l'un des langages les plus utilisés dans le développement des jeux, des systèmes d'opérations, des navigateurs web, etc .

CRCN : Le chargé de recherche de classe normale, est un des grades du corps des chargés de recherche.

CPU : De l'anglais « central processing unit », le CPU est un microprocesseur, il est considéré comme le cerveau d'un ordinateur.

Deep learning : Anglicisme décrivant un apprentissage profond. En informatique, le deep learning est une technologie qui provient de l'intelligence artificielle, elle est composée d'algorithmes

capables de mimer les actions du cerveau humain grâce à des réseaux de neurones artificielles, souvent accompagné de grandes quantités de données

DJL : Deep Java Library est une bibliothèque open source qui permet de construire et utiliser du deep learning en langage Java.

Docs : Outil de Google permettant l'édition de texte en ligne.

Draw : Outil de Google permettant de créer des dessins et diagrammes en ligne.

Drive : Outil de Google permettant de stocker et synchroniser des fichiers en ligne.

Entraînement : En deep learning, l'entraînement consiste à façonner un modèle de deep learning à partir de données d'entraînement.

Fiji : Fiji is just ImageJ est une version étendue de ImageJ livrée avec des plugins.

GanttProject : Logiciel de gestion permettant la planification d'un projet en réalisant un diagramme de Gantt.

Git : Projet open source permettant le travail collaboratif de développement.

GitHub : Site web et service en ligne permettant le stockage, la mémorisation et le partage de codes.

GitLab : Plateforme de développement collaborative basée sur les fonctionnalités de Git.

GPL v3.0 : GNU General Public License est une licence libre qui permet de modifier le travail, de l'étudier et de redistribuer le travail ou un travail dérivé.

GPU : De l'anglais « graphic processing unit », le GPU est une puce informatique d'une carte graphique. Son parallélisme massif est idéal pour la prise en charge de tâches répétitives et hautement parallèles.

IDE : De l'anglais « integrated development environment », l'IDE est un environnement de développement intégré, il s'agit d'un logiciel de création d'application qui rassemble des outils de développement en une seule interface graphique.

IHM : L'interface homme-machine est un tableau de bord permettant à l'utilisateur de communiquer avec une machine ou un programme.

IJ-OpenCV : Bibliothèque Java qui permet la communication d'ImageJ et d'OpenCV.

ImageJ : Logiciel libre et open source de traitement et d'analyse d'images.

Inférence : En deep learning, l'inférence consiste à réaliser des prédictions efficaces à partir d'un modèle entraîné.

DR2 : Directeur de recherche de deuxième classe.

IntelliJ : IDE pour logiciels développés en Java, Kotlin, et d'autres langages de programmation.

Java : Langage de programmation orienté objet, il permet de créer des applications, et des modules d'application.

LabelImg : Logiciel d'annotation et d'étiquetage d'image.

Livrables : En gestion de projet informatique, un livrable est un résultat inscrit dans le planning du projet, il peut s'agir d'un logiciel, d'un programme, d'un document de conception, ou autre éléments produits.

MXNet : Infrastructure logicielle de deep learning écrit en langage Python.

OpenCV : Open Computer Vision est une bibliothèque graphique spécialisée dans le traitement d'images.

Open source : En informatique, un logiciel open source est un code conçu pour être accessible à n'importe qui, c'est à dire le voir, le modifier et le distribuer.

Planhammer : Site web combinant les outils de gestion de projet agiles et classiques.

Png : Portable Network Graphics est un format de fichier graphique utilisant une compression sans perte.

RAM : De l'anglais « random access memory », la RAM est considérée de « mémoire vive », elle permet de stocker provisoirement des données.

RGPD : Le règlement général sur la protection des données encadre le traitement des données personnelles sur le territoire de l'Union Européenne.

WBS : De l'anglais « work breakdown structure », la structure de découpage du travail est un outil de gestion de projet permettant de visualiser chaque tâche.

Sheets : Outil de Google permettant la création de tableurs en ligne.

Suite Office365 : Collections de programmes informatiques intégrant plusieurs outils de bureautique.

SuperAnnotate : Logiciel de labellisation de données.

Windows 10 : Système d'exploitation de la société Microsoft.

Workshop : Atelier collaboratif sous forme d'une réunion de groupe.

XML : Langage de description de données, il permet la structuration de données.

# Bibliographie

Les sources ci-dessous ont été consultées pour l'élaboration de mon travail.

Irset :

- Irset : <https://www.irset.org/fr>
- Équipe 4 : <https://www.irset.org/fr/equipe-f-smagulova-avenir-meiose-epigenetique-et-reproduction-mer>

Deep learning :

- Biosit : <https://biosit.univ-rennes1.fr/workshop-deep-learning-based-segmentation>
- DJL : <https://djl.ai/>
- D2L : <https://d2l.djl.ai/>

ImageJ :

- ImageJ : <https://imagej.nih.gov/>
- ImageJ / Fiji : <https://imagej.net/>

Documentation :

- Java : <https://docs.oracle.com/javase/8/docs/api/>
- ImageJ : <https://imagej.nih.gov/ij/docs/index.html>
- OpenCV : <https://docs.opencv.org/4.x/javadoc/index.html>
- DJL : <https://docs.djl.ai/>
- Stack OverFlow : <https://stackoverflow.com/>

# Index des figures et dessins

## Index des figures

Figure 1: Carte de la position de l'Irset (à partir de <a href="https://www.google.com/maps/">https://www.google.com/maps/</a> ).....	7
Figure 2: Diagramme de Gantt prévisionnel (fait à partir de GanttProject).....	18
Figure 3: Diagramme de Gantt réel (fait à partir de GanttProject).....	19
Figure 4: Imprime écran du logiciel Éclipse possédant le mauvais encodage.....	23
Figure 5: Diagramme de classe : Plugin (fait à partir de Bouml).....	27
Figure 6: Maquette de l'interface homme machine (fait à partir de Google Draw).....	28

## Index des dessins

Dessin 1: Script en langage macro ImageJ.....	29
Dessin 2: Méthode d'initialisation d'un set d'images.....	30
Dessin 3: Méthode de conversion d'image.....	31
Dessin 4: Affichage de console après test d'entraînement d'un modèle.....	32

# Annexe

## **Index des annexes**

<b>Annexe 1 : Grille de caractérisation l'organisation.....</b>	<b>46</b>
<b>Annexe 2 : Organigramme de l'Irset.....</b>	<b>53</b>
<b>Annexe 3 : Modèle de PESTEL associé à l'Irset.....</b>	<b>54</b>
<b>Annexe 4 : Modèle de Porter associé à l'Irset.....</b>	<b>57</b>
<b>Annexe 5 : Diagnostic de l'Irset, matrice SWOT.....</b>	<b>60</b>
<b>Annexe 6 : Sujet de stage établi entre l'IUT et l'Irset.....</b>	<b>61</b>
<b>Annexe 7 : Diagramme WBS (planhammer).....</b>	<b>62</b>
<b>Annexe 8 : Diagramme de Gantt prévisionnel (GanttProject)..</b>	<b>63</b>
<b>Annexe 9 : Diagramme de Gantt réel (GanttProject).....</b>	<b>64</b>

## Annexe 1 : Grille de caractérisation l'organisation

Critères	
Forme :	Établissement Public à but non lucratif
Type d'organisation :	EPST (Établissement Public à caractère Scientifique et Technologique)
Statut juridique :	Les EPST sont des personnes morales de droit public dotées de l'autonomie administrative et financière. Leur objet principal n'est ni industriel ni commercial. Ils relèvent du même régime administratif, budgétaire, financier et comptable que les établissements publics administratifs (Code de la Recherche, articles L321-1 et L321-3).
Finalité :	Assurer une mission de service public : investir le champ de la recherche biomédicale fondamentale et appliquée dans le domaine de la santé humaine.
Objectifs :	Améliorer la santé humaine. Dans la pratique, la performance d'un organisme de recherche s'évalue principalement via le nombre et la qualité des publications scientifiques de ses chercheurs.

Nationalité :	Française
Activité principale :	Recherche fondamentale et appliquée dans le domaine biomédical.
Secteur d'activité:	Secteur tertiaire (services non marchands)
Taille (critères capital n/a quantitatifs) :	chiffre d'affaire n/a parts de marché n/a
	Inserm :
	Personnels (2021) :
	Environ 14000 personnes dont:
	- 5000 fonctionnaires ( 2100 chercheurs et 2900 techniciens, ingénieurs et administratifs (ITA)).
	- 3200 contractuels et vacataires.
	- 5500 hospitalo-universitaires et universitaires.
	Budget (2021) :
	967 M€ dont 67% proviennent de la dotation d'Etat et 33% de ressources propres (contrats externes, exploitation de brevets, etc.)
	263 unités de recherche (l'Irset est l'une d'entre-elles), 40 unités de service, 34 centres d'investigation clinique).
	Irset :
	Personnels (2021) :

Environ 300 personnes parmi lesquelles se trouve :

- 27 chercheurs EPST (Inserm ou CNRS)
- 42 enseignants-chercheurs (Université de Rennes 1)
- 52 ITA
- 50 Hospitalo-universitaires
- 63 doctorants
- 12 chercheurs post-doctorants

Budget (2021) :

Environ 10M€ dont 10% proviennent de la dotation d'Etat et 90% de ressources propres (contrats externes, financements sur appels à projets, etc.). Cet apparent déséquilibre dans la répartition de l'origine des ressources entre l'Inserm et l'Irset s'explique par le fait que les traitements et salaires des personnels est en majorité financé par les organismes de tutelles (Inserm, Université, Hôpital) sur dotation d'Etat tandis que les coûts de fonctionnement et d'entretien sont principalement assurés sur fonds propres.

Origine des capitaux : La dotation d'Etat (qui comprend le traitement des fonctionnaires) provient évidemment de fonds publics. Les fonds propres peuvent être publics (Agence Nationale de la Recherche, Appels d'offre de la Région, Cancéropôle, etc.), privés (contrats avec l'industrie pharmaceutique, cosmétique, etc.) ou associatifs (ARC, Ligue contre le cancer, etc.).

Besoin : Besoin sociétal (nécessaire au fonctionnement de la vie en société)

Service public qui répond à une demande sociétale de progrès en santé humaine.

Domaine d'intervention ou champ d'action : La recherche scientifique biomédicale a pour but d'améliorer les connaissances dans le domaine de la santé humaine sans objectif lucratif. En conséquence, son domaine d'intervention ne peut être défini comme un marché. Son champ d'action est international.

Performances (notion d'efficacité) : La performance d'un organisme de recherche est et généralement évaluée par le nombre et la qualité des publications issues de ses laboratoires. Selon ces critères, L'Inserm est au 1er rang européen des institutions académiques de recherche dans le domaine biomédical et au 2e rang mondial derrière les National Institutes of Health (NIH) des États-Unis.

L'Inserm est très impliqué dans la recherche en lien avec les besoins de la société et, à ce titre, développe de nombreux projets scientifiques – particulièrement à l'Irset – destinés à explorer l'impact de la qualité de l'environnement sur la santé humaine.

Sur le plan technologique, les chercheurs de l'Inserm sont encouragés à valoriser leurs travaux. Avec un portefeuille de 2 069 familles de brevets en 2020, l'Institut est le 1er organisme académique en recherche biomédicale en Europe, le 1er déposant en Europe dans la catégorie pharmaceutique et le 2e dans la catégorie biotechnologie.

Ressources:	<p>Pour l'Inserm :</p> <ul style="list-style-type: none"><li>• ressources matérielles : Foncier (siège, Délégations Régionales, laboratoires, etc.), Équipements scientifiques.</li><li>• ressources immatérielles : Brevets, notoriété de l'Institut, relations privilégiées avec le monde médical hospitalo-universitaire. Réseaux scientifiques internationaux.</li><li>• ressources financières : Dotation de L'État (2/3 du budget) et ressources propres (1/3 du budget).</li><li>• ressources humaines (compétences) : Personnel hautement qualifié:<ul style="list-style-type: none"><li>- 8000 chercheurs, enseignants chercheurs ou hospitalo-universitaires titulaires d'un doctorat et ayant souvent effectué plusieurs années de post-doctorat.</li><li>- 3000 ingénieurs et techniciens.</li><li>- 3000 contractuels et vacataires de niveau d'études comparable voire identique à celui des personnels titulaires (post-doctorants, ingénieurs contractuels, etc.).</li></ul></li></ul>
Mode de coordination :	<p>La réalisation d'un projet de recherche repose généralement sur une équipe de recherche (entre 3 et 30 personnes) – éventuellement en collaboration avec une ou plusieurs autres équipes. Au sein d'une équipe, le mode de coordination dépend de nombreux facteurs tels que la personnalité et l'expérience du responsable d'équipe et des membres de l'équipe, le besoin d'expertise extérieure, le</p>

caractère standardisé ou non des expériences, etc. En règle générale, la conduite d'un projet de recherche repose sur une combinaison de différents modes de coordination.

Quelques exemples :

- Supervision directe dans le cas du projet d'un doctorant placé sous la supervision de son directeur de thèse.
- Standardisation des tâches pour des travaux répétitifs nécessitant une reproductibilité normée.
- Standardisation des compétences lorsque l'équipe est habituée à travailler ensemble.
- Ajustement mutuel dans le cas où, compte tenu des compétences de chacun et de la confiance mutuelle, les contributions se font par contact direct plutôt que via une supervision hiérarchique. C'est, par exemple, le cas lors de coopérations – souvent internationales – entre plusieurs équipes de recherche.

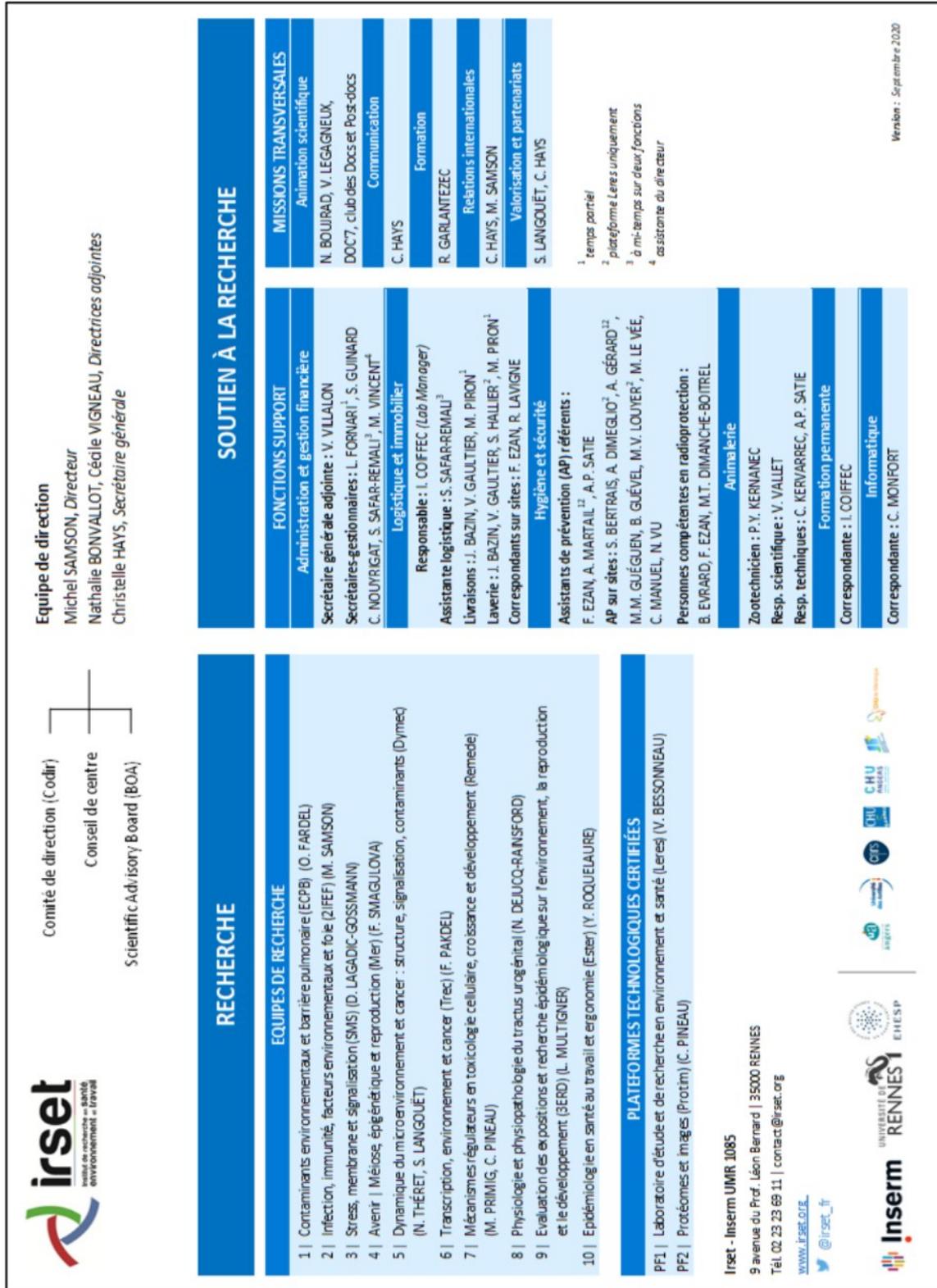
Mode de direction :

Le Président de l'Inserm est nommé par décret en Conseil des ministres pour une durée de 4 ans renouvelable une fois. Il s'appuie sur les délibérations d'un Conseil d'Administration composé de 22 membres nommés par les ministères de tutelle (santé et recherche). Il est conseillé par un Conseil Scientifique composé pour moitié de membres nommés et pour moitié de membres élus par le personnel. Les commissions scientifiques spécialisées (CSS), au nombre de 7, participent à l'évaluation des activités de recherche, au recrutement des chercheurs et à leur promotion. Chaque CSS est spécialisée dans un grand domaine scientifique. Ces commissions scientifiques sont composées pour moitié de membres élus par le personnel et pour moitié de membres

nommés par président de l'Inserm.

Les unités de recherche regroupent plusieurs équipes de recherche (11 pour l'Irset) et sont dirigées par un Directeur d'Unité qui s'appuie sur un comité de direction et un Conseil de Laboratoire.

## Annexe 2 : Organigramme de l'Irset



### Annexe 3 : Modèle de PESTEL associé à l'Irset

#### Critères

##### Environnement Politique

La part des dépenses de recherche en France (publique et privée confondus) est estimée à 2,2% du PIB, ce qui est très en retrait par rapport aux investissements moyens des pays de l'OCDE. De plus, une partie importante des financements publics est orientée vers les entreprises via le Crédit Impôt Recherche (CIR, environ 6 milliards d'euros par ans, à comparer au budget du CNRS qui n'est que de 3 milliards par an). Ce sous-investissement dans la recherche publique, chronique depuis quatre décennies, conduit à une diminution des postes de chercheurs ouverts au concours (-30% en dix ans) et à un vieillissement de la population des chercheurs. De plus, les politiques publiques d'austérité en matière de recherche ont conduit l'Etat à se désengager du financement récurrent des laboratoires au profit d'appels à projets pour lesquels le taux de succès est très bas (15% de succès aux appels d'offre de l'Agence Nationale de la Recherche, par exemple). L'environnement politique est

	<p>clairement celui qui impacte de façon dominante la capacité du pays à produire une recherche de qualité.</p>
Environnement Économique	<p>Sans objet pour l'Inserm, à l'exception de l'inflation qui peut peser sur le budget de fonctionnement des laboratoires..</p>
Environnement Socioculturel	<p>Comme mentionné plus haut, la pyramide des âges des chercheurs tend à se rétrécir à sa base en raison de la baisse tendancielle du soutien de l'État à la recherche publique. Par ailleurs, il est constaté, depuis une vingtaine d'années, une désaffection croissante des étudiants pour les matières scientifiques et il devient difficile de recruter des doctorants performants.</p>
Environnement Technologique	<p>La recherche biomédicale est extrêmement dépendante des innovations technologiques (qu'elle produit souvent elle-même). Chaque innovation majeure (PCR, séquençage d'ADN à haut débit, microscopie con-focale, édition du génome par Crispr/Cas9, pour n'en citer que quelques unes) permet de faire des "bonds" qualitatifs dans la compréhension des mécanismes qui contrôlent le vivant.</p>

### Environnement Écologique

Pour l'Inserm, et particulièrement pour l'Irset, la protection de l'environnement n'est pas une contrainte mais un sujet d'étude. A travers ses travaux, l'Irset est, par exemple, amené à formuler des recommandation concernant les expositions admissibles à différentes substances. Ces recommandations peuvent ensuite être traduites en normes par des organismes tels que l'Anses (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail).

### Environnement Légal

Les fonctionnaires de l'Inserm sont soumis au droit public tandis que les agents non titulaires sont embauchés, selon les missions, sur des contrats de droit public ou privé (et sont donc, dans ce dernier cas, soumis au Code du Travail). La durée légale du travail est de 35h.

## Annexe 4 : Modèle de Porter associé à l'Irset

### Critères

#### L'intensité concurrentielle

La concurrence en matière de recherche est représentée par les laboratoires qui travaillent sur le même sujet et risquent donc d'être les premiers à publier un résultat que l'on s'efforce d'obtenir. Cette situation se retrouve dans d'autres secteurs d'activité comme le journalisme. Par ailleurs, la frontière entre concurrent et collaborateur est floue et mouvante. Il n'est pas rare qu'un laboratoire concurrent devienne collaborateur le temps d'un projet et vice-versa. Le laboratoire dans lequel j'ai effectué mon stage a connaissance d'une cinquantaine d'équipes au niveau mondial susceptibles d'être soit compétiteurs, soit collaborateurs. Les notions de part de marché et de taux de croissance du secteur ne sont pas applicables à la recherche fondamentale.

#### Le pouvoir de négociation des fournisseurs

Les laboratoires de l'Inserm se fournissent auprès d'entreprises privées pour leurs équipements, leur fonctionnement et pour certains services. Pour ce qui concerne le fonctionnement (consommables) et les

services (travaux de relecture, séquençage d'ADN à façon, par exemple), les fournisseurs sont relativement nombreux et la mise en concurrence est possible. Elle est souvent effectuée au niveau de l'unité de recherche afin d'augmenter les volumes négociés et donc de peser sur les prix. Une exception notable concerne le matériel informatique qui est soumis à un appel d'offre national (conformément au Code des Marchés Public). Pour ce qui concerne les gros équipements scientifiques (microscopes, par exemple), le nombre de fournisseurs est beaucoup plus restreint et chaque fournisseur présente des spécificités techniques qui lui sont propres. Cette situation ne laisse pas beaucoup de marge de manœuvre pour négocier les prix.

#### La menace de nouveaux entrants

Les nouveaux entrants dans la recherche fondamentale ne représente pas réellement une menace dans la mesure où le produit de cette recherche est destiné, à terme, à être mis à la disposition de la communauté pour répondre à une demande sociétale de progrès. Toutefois, selon la granulométrie utilisée, il peut exister une forme de concurrence qui évolue avec l'arrivée de nouveaux acteurs.

	<p>A titre d'exemple, il devient évident que l'Europe, et particulièrement la France, perd du terrain par rapport aux pays d'Asie qui ont récemment fortement investi dans d'ambitieux projets de recherche.</p>
La menace de produits de substitution	Sans objet pour l'Inserm.
L'intervention de l'État	<p>Comme indiqué plus haut, le rôle des politiques publiques en matière de recherche fondamentale est particulièrement important pour le développement et le maintien d'une recherche performante. Il semble malheureusement qu'il ne s'agisse pas d'une priorité pour les différents gouvernements qui se sont succédés depuis 1986. De plus, l'accumulation des tâches administratives liées au désengagement de l'État (réponse aux appels à projet, justification des contrats, etc.) conduit à limiter le temps que les chercheurs peuvent consacrer à leur activité scientifique.</p>

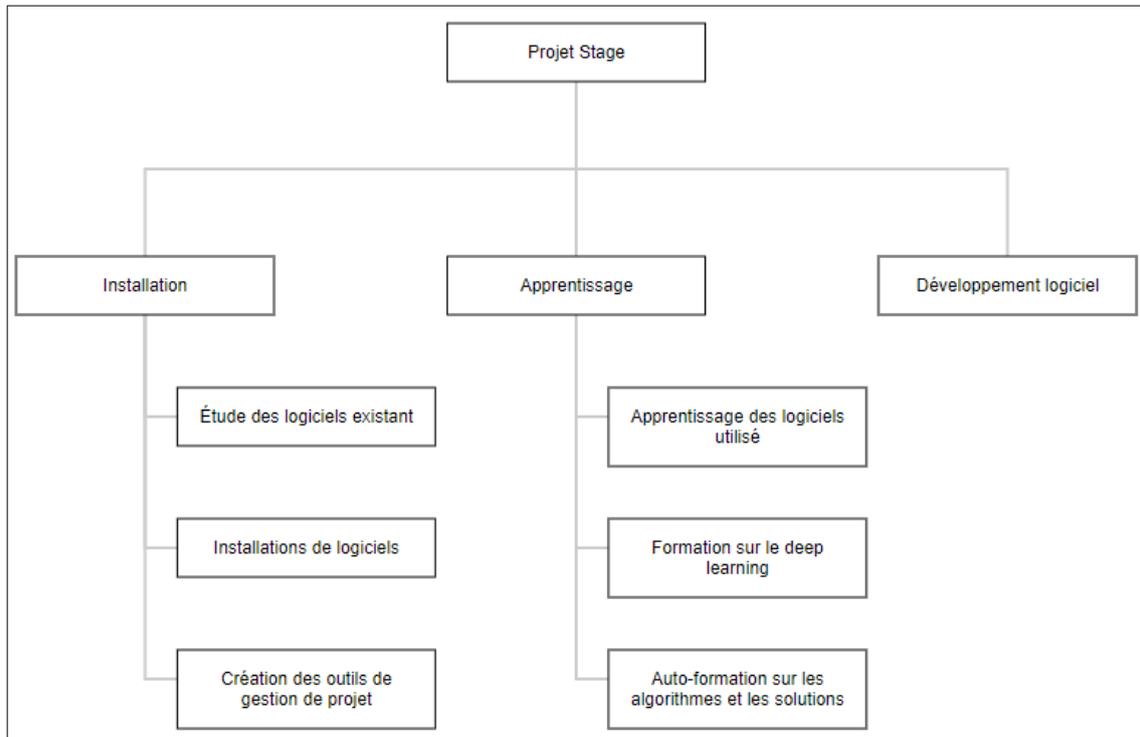
## Annexe 5 : Diagnostic de l'Irset, matrice SWOT

	POSITIF (pour atteindre l'objectif)	NÉGATIF (pour atteindre l'objectif)
INTERNE	<p>Forces :</p> <p>Motivation des personnels. Réseau de laboratoires. Forte capacité d'innovation. Bonne image auprès du public.</p>	<p>Faiblesses :</p> <p>Poids des tâches administratives. Survivance ponctuelle d'une forme de "mandarinat". Poids politique des réseaux de médecins. Injonction croissante à la valorisation.</p>
EXTERNE	<p>Opportunités :</p> <p>Forte demande sociale de progrès. Statut de la fonction publique. Appropriation rapide des nouvelles technologies. Bonne coopération avec les autres acteurs de la recherche à l'échelle nationale ou internationale.</p>	<p>Menaces :</p> <p>Sous-financement chronique de la recherche française. Déplacement du centre de gravité de la recherche scientifique depuis l'Occident vers l'Asie. Désintérêt des étudiants pour les carrières scientifiques. Compétition internationale accrue.</p>

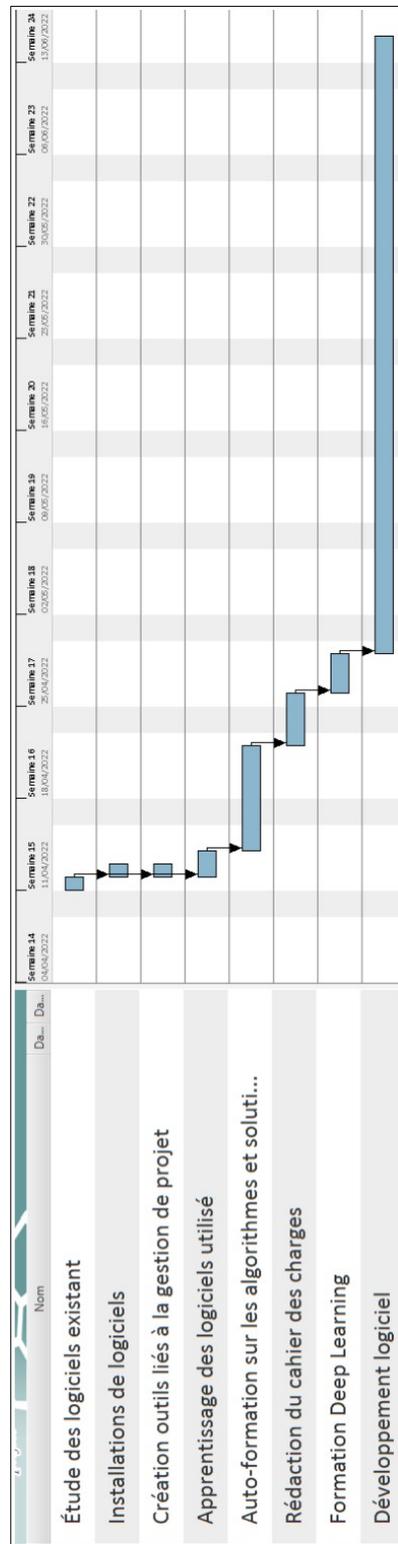
## Annexe 6 : Sujet de stage établi entre l'IUT et l'Irset

<b>Sujet de stage pour DUT Informatique</b>	
<b>Période de 10 semaines continues (minimum) du 11 avril au 17 juin 2021</b>	
Étudiant :	Yang Robin
<b>ORGANISME D'ACCUEIL</b>	
Numéro SIRET :	193509361.00013 code APE / NAF : 8542Z (Enseignement Supérieur)
Nom organisme :	<b>Irset - Inserm UMR_S 1085</b>
Adresse :	9 avenue du Prof. Léon Bernard 35000 RENNES FRANCE
<b>SERVICE D'AFFECTATION</b>	
Service :	Équipe "Meiosis, epigenetics, reproduction"
Tuteur :	Dr. Christian Jaulin
Fonction :	Directeur de recherche INSERM
Tél :	02 23 23 43 29 E-mail : christian.jaulin@univ-rennes1.fr
Le tuteur de stage est-il un informaticien ?	<input type="radio"/> Oui <input checked="" type="radio"/> Non
<b>SUJET DE STAGE</b>	
Intitulé :	"Création d'un module ImageJ destiné à l'analyse cytogénétique"
Description :	Le stagiaire développera un module destiné à être intégré dans le logiciel open source ImageJ (analyse d'images de microscopie optique). Ce module devra remplir 2 fonctions: 1) Établir l'index mitotique d'une population cellulaire par reconnaissance de forme des chromosomes condensés 2) Mesurer la proportion de chromatides sœurs non attachées dans une population cellulaire par évaluation de la distance entre centromères appariés Le développement de ces modules permettra d'automatiser une tâche jusque là effectuée manuellement
Environnement :	PC récent sous Windows 10 / Logiciel Open Source ImageJ / Langage Java <environnement technique (matériel, logiciel, langages, etc.)>
Gratification :	150 euros par semaine (minimum : 3.90 € par heure, à raison de 7h/jour)
Nombres de stagiaire(s) susceptible(s) d'être accueilli(s) :	1

## Annexe 7 : Diagramme WBS (planhammer)



## Annexe 8 : Diagramme de Gantt prévisionnel (GanttProject)



## Annexe 9 : Diagramme de Gantt réel (GanttProject)

