

Master Bioinformatique  
Génie Logiciel et Gestion de Projet

**Cahier des charges du projet :**

---

Conception et Architecture Logiciel -  
Développement d'Outils Bioinformatiques pour l'Étude  
et l'Analyse du Métabolisme du Virus H5N1

---

**Auteurs :**

Céline Hosteins  
Linda Khodja  
Franck Sanchez  
Ilham Bedraoui  
Maroia Alani

**Encadrante :**

Marie Beurton-Aimar



Université de Bordeaux  
Octobre 2024

# Table des matières

<b>Introduction</b>	<b>2</b>
<b>1 Analyse</b>	<b>3</b>
1 Contexte . . . . .	3
2 État de l'art . . . . .	4
2.1 Bases de données pour l'analyse métabolique . . . . .	4
2.2 Outils de visualisation et d'analyse des données . . . . .	4
2.3 Serveurs Web similaires . . . . .	5
3 Analyse des besoins . . . . .	6
3.1 Besoins fonctionnels . . . . .	6
3.2 Besoins non fonctionnels . . . . .	7
4 Engagements . . . . .	7
4.1 Choix des outils : . . . . .	7
4.2 Outils à développer : . . . . .	8
5 Technologies et langages utilisés . . . . .	9
5.1 Langages et frameworks . . . . .	9
<b>2 Organisation</b>	<b>10</b>
1 Architecture du système . . . . .	10
2 Répartition des tâches . . . . .	10
2.1 Diagramme de Gantt . . . . .	12

# Introduction

Le virus de la grippe aviaire H5N1, une menace globale pour la santé animale et humaine, a suscité des inquiétudes majeures au cours des dernières décennies en raison de sa propagation dans les élevages avicoles et de son potentiel zoonotique. Le H5N1 est responsable de nombreuses épidémies parmi les oiseaux domestiques, entraînant d'importantes pertes économiques et des risques de transmission à l'homme [1].

Ce projet collaboratif vise à approfondir l'étude du virus H5N1 par la conception d'outils dédiés à l'analyse et à la modélisation des voies métaboliques liées à ce virus, afin de mieux comprendre les processus biologiques sous-jacents à sa réplication et à sa survie dans les organismes infectés.

Dans ce cadre, une plateforme web sera développée pour centraliser et intégrer différents outils, permettant ainsi à la communauté scientifique d'accéder aux services bioinformatiques, de charger des données, de réaliser des analyses et des modélisations des voies métaboliques du virus, et d'explorer les résultats de manière interactive. Toutes ces informations seront utiles pour le développement de stratégies de lutte contre ce virus.

**Mots-clés :** virus H5N1, grippe aviaire, épidémies, bioinformatique, métabolisme, modélisation, plateforme web.

# Chapitre 1

## Analyse

### 1 Contexte

Le virus de la grippe aviaire H5N1 est une souche hautement pathogène du virus de la grippe A, responsable d'importantes épidémies dans les populations aviaires à travers le monde, y compris en France. Originaire principalement des oiseaux aquatiques sauvages, ce virus peut se propager aux volailles domestiques, entraînant des taux de mortalité élevés. Depuis le début des années 2000, le H5N1 est reconnu comme une menace mondiale pour la santé en raison de son potentiel à infecter l'homme, bien que de tels cas restent rares [1, 2]. Comme illustré dans la Figure 1.1, le virus possède divers composants qui jouent un rôle dans son cycle de réplication et son pouvoir pathogène.

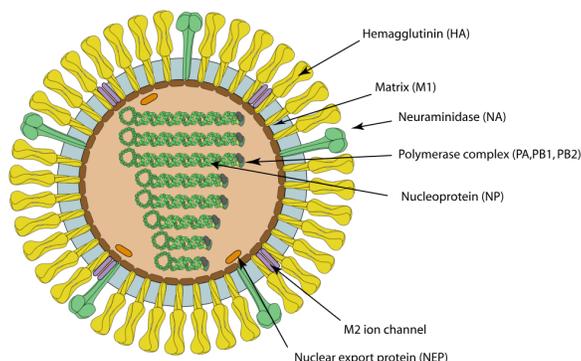


FIGURE 1.1 – Diagramme schématisique du virus de la grippe aviaire illustrant ses divers composants [3].

La transmission du H5N1 se fait principalement par contact direct avec des oiseaux infectés ou des environnements contaminés. Bien que les infections humaines soient rares, elles peuvent entraîner des maladies respiratoires graves avec un taux de mortalité élevé. Le potentiel de mutation du virus soulève des inquiétudes quant à la possibilité qu'il devienne transmissible entre humains, ce qui poserait un risque pandémique important [4]. Le cycle viral du H5N1 au sein des cellules de mammifères, représenté dans la Figure 1.2, montre les étapes clés du processus d'infection et de réplication virale.

L'étude bioinformatique du virus H5N1 joue un rôle crucial dans la compréhension de ses mécanismes biologiques, notamment en ce qui concerne la réplication virale, les interactions hôte-pathogène et les voies métaboliques. L'utilisation d'outils bioinformatiques dans ce projet est essentiel afin d'interpréter les données complexes générées par le séquençage génétique, l'analyse protéomique et la modélisation des réactions biochimiques associées au métabolisme du H5N1. L'objectif du projet est de développer une plateforme web qui centralisera l'accès à divers outils bioinformatiques, permettant de traiter des données biochimiques en entrée, de modéliser les interactions métaboliques spécifiques au H5N1 et de permettre une visualisation des résultats. Cette modélisation pourra servir à identifier des cibles potentielles pour le développement de traitements antiviraux, contribuant ainsi à la lutte contre ce virus hautement pathogène [2].

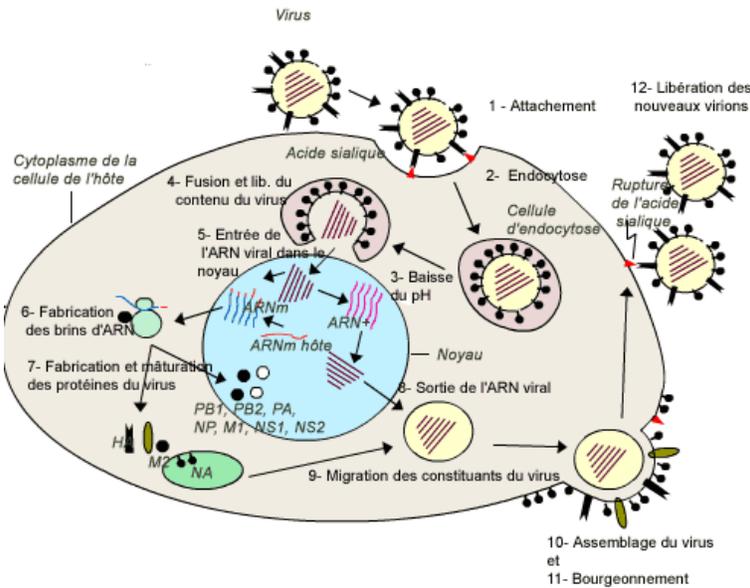


FIGURE 1.2 – Schéma du cycle viral des Orthomyxoviridae (type H5N1) dans une cellule de Mammifère [5].

## 2 État de l’art

La réalisation d’une étude du métabolisme d’une espèce repose sur des outils bioinformatiques permettant de modéliser, simuler et analyser les réseaux métaboliques. Avec le développement d’applications interactives notamment via *RShiny*, il est possible de créer des interfaces conviviales qui facilitent ces analyses. Cet état de l’art examine plusieurs logiciels et APIs qui peuvent être intégrés dans un service web offrant la possibilité d’importer des fichiers, d’effectuer des analyses bioinformatiques standard et de visualiser les résultats de manière intuitive.

### 2.1 Bases de données pour l’analyse métabolique

*KEGG* [6] (Kyoto Encyclopedia of Genes and Genomes) est une base de données largement utilisée dans les analyses métaboliques et biologiques. Elle fournit des informations détaillées sur les voies métaboliques, les interactions moléculaires et les fonctions biologiques ainsi que des modules pour le mappage de données expérimentales. *KEGG* se distingue par sa capacité à fournir des chemins spécifiques à certains organismes ou agents pathogènes, comme le virus H5N1. Son intégration facile avec des API permet une automatisation des analyses, ce qui est crucial pour des projets nécessitant une exploration systématique des réseaux métaboliques. Une autre base de données couramment utilisée est *MetaCyc* [7] qui se concentre également sur les voies métaboliques et les réseaux de réactions. *MetaCyc* couvre une large gamme d’organismes et inclut des informations biochimiques détaillées, utiles pour les études de métabolisme en microbiologie et en biotechnologie. Toutefois, contrairement à *KEGG*, elle propose moins d’options de visualisation interactive ce qui limite l’exploration dynamique des réseaux. Enfin, *Reactome* [8] propose un large éventail de pathways métaboliques, de signalisation et de développement cellulaire. Elle est particulièrement réputée pour sa couverture exhaustive des voies humaines et la qualité de ses annotations. Contrairement à *KEGG*, *Reactome* est centrée sur les organismes humains, ce qui peut la rendre moins appropriée pour les études virales ou d’organismes variés bien qu’elle offre aussi une API pour l’intégration de données.

En résumé, *KEGG* se démarque par son intégration robuste et ses pathways spécifiques à des organismes variés en faisant un choix privilégié pour l’analyse de réseaux métaboliques viraux comme ceux du H5N1. *MetaCyc* et *Reactome* présentent des avantages respectifs en termes de détails biochimiques et de couverture des voies humaines bien qu’ils soient moins adaptés pour les visualisations dynamiques ou la couverture de multiples organismes.

### 2.2 Outils de visualisation et d’analyse des données

Dans l’analyse de données biologiques, la création de graphiques statiques requiert des outils flexibles et adaptés aux exigences de présentation scientifique. Les outils comme *ggplot2* [9] se distinguent par leur struc-

ture intuitive qui permet de superposer des éléments visuels en couches et de créer des graphiques complexes et esthétiques. Cette approche est particulièrement utile dans le domaine bioinformatique où les chercheurs ont besoin de présenter les résultats de manière claire et structurée. À côté de *ggplot2*, des outils tels que *Matplotlib* ou *Seaborn*, bien que performants, demandent souvent des configurations supplémentaires pour obtenir des visuels de même qualité. La syntaxe de *ggplot2* permet aux utilisateurs d’ajuster rapidement leurs graphiques pour des analyses biomédicales et constitue ainsi un atout majeur dans les projets nécessitant des visualisations précises et répétées.

Pour répondre aux besoins croissants de l’exploration de données interactive, plusieurs outils offrent des solutions variées en termes de personnalisation et d’accessibilité. *Plotly* [10] est particulièrement reconnu pour sa capacité à transformer des graphiques statiques en visualisations dynamiques directement intégrables dans des applications web. Utilisé avec *RShiny*, *Plotly* permet de créer des applications interactives dans lesquelles les utilisateurs peuvent explorer les données via des fonctionnalités telles que le zoom, le survol d’informations et la sélection de points de données. Cela est particulièrement précieux dans l’analyse de données biologiques volumineuses, où une exploration approfondie est nécessaire pour comprendre les tendances et identifier des anomalies. Dans le même domaine, *D3.js* [11] offre une flexibilité incomparable pour créer des visuels interactifs sur mesure mais nécessite des compétences en JavaScript et des configurations avancées ce qui peut limiter son adoption pour les utilisateurs recherchant une solution plus intuitive. *Plotly* en revanche, simplifie l’ajout d’interactivité dans des applications *RShiny*, permettant aux chercheurs de produire des visualisations dynamiques sans maîtriser des langages de développement web complexes. Ainsi, *Plotly* et *RShiny* forment une solution équilibrée, bien adaptée aux visualisations interactives en ligne et facilement configurable pour les utilisateurs de R.

Pour la visualisation des pathways métaboliques, intégrer des données omiques dans les réseaux visuels est crucial surtout pour l’étude des voies spécifiques comme celles impliquées dans les infections virales. *Pathview* [12] est un outil efficace permettant d’intégrer les données d’expression dans les pathways de *KEGG*, simplifiant ainsi la compréhension des variations d’expression dans des réseaux complexes. Cette fonctionnalité est essentielle pour les analyses biologiques où la visualisation des données omiques dans leur contexte fonctionnel est un atout majeur. D’autres outils, comme *Cytoscape* [13], sont largement utilisés pour l’analyse et la manipulation des réseaux métaboliques et biologiques. Cependant, ils sont plus adaptés à des explorations manuelles de réseaux avec une interface détaillée pour les études de grande échelle. Par ailleurs, *Cytoscape*, bien qu’utile pour l’analyse de réseaux complexes, n’offre pas la même intégration directe avec les flux de travail R, limitant ainsi son usage dans des pipelines d’analyse omique automatisés. *Reactome Pathway Browser* est aussi un choix pertinent pour les visualisations de pathways, surtout dans les analyses humaines, bien qu’il manque parfois d’adaptabilité pour des études spécifiques à des pathogènes. Avec son intégration dans R et son alignement direct avec les données de *KEGG*, *Pathview* s’avère particulièrement adapté pour des analyses de pathways métaboliques dans des projets nécessitant des visualisations simples et rapides.

Les réseaux de co-expression exigent des outils interactifs pour visualiser les relations entre les gènes ou les métabolites. *visNetwork* [14] est particulièrement bien adapté pour cet usage grâce à son interactivité et à sa personnalisation avancée permettant de modifier des attributs comme la taille ou la couleur des nœuds selon les valeurs d’expression. L’interactivité offerte aide les utilisateurs à naviguer facilement dans les réseaux et à explorer les relations biologiques. Dans ce domaine, des logiciels comme *Gephi* [15] proposent des solutions avancées pour l’analyse des réseaux complexes, mais leur utilisation implique souvent une phase d’import/export de données, alourdissant les processus. De plus, *Gephi* est mieux adapté aux analyses de structure de réseau à grande échelle, mais n’offre pas l’intégration directe avec R ni la même facilité d’utilisation pour des visualisations interactives en ligne. *visNetwork* se distingue donc par sa facilité d’utilisation et son intégration native dans R le rendant particulièrement adéquat pour les études de co-expression nécessitant une exploration fluide et intuitive.

## 2.3 Serveurs Web similaires

Dans le domaine de la bioinformatique, plusieurs plateformes permettent d’analyser les données métaboliques et d’explorer les voies biologiques, offrant des outils polyvalents pour les chercheurs. Trois plateformes particulièrement pertinentes pour ce projet sont *MetaboAnalyst* [16], *WebGestalt* [17], et *MetaboReport* [18] qui fournissent des fonctionnalités avancées et variées pour l’analyse des données omiques.

*MetaboAnalyst*, actuellement en version 6.0, est une plateforme bien établie pour l’analyse des données

métaboliques. Elle propose une gamme complète d'outils pour la transformation des données, l'analyse statistique et la visualisation des données métaboliques. De plus, *MetaboAnalyst* intègre des fonctionnalités de parcours des voies métaboliques en s'appuyant sur des bases de données publiques telles que *KEGG* facilitant ainsi l'exploration de réseaux biologiques complexes. Cette plateforme interactive permet aux utilisateurs de charger des fichiers de données au format CSV, de générer des graphiques variés et d'interagir avec des modèles métaboliques sophistiqués. Grâce à ces caractéristiques, *MetaboAnalyst* est devenu une référence pour l'analyse des voies métaboliques et constitue une ressource précieuse pour les études en métabolomique.

*WebGestalt* (WEB-based Gene SeT AnaLysis Toolkit) est un autre outil notable orienté vers l'enrichissement fonctionnel et l'analyse des interactions biologiques. *WebGestalt* se spécialise dans l'analyse des voies biologiques et des interactions moléculaires en utilisant des données omiques, notamment génétiques et protéomiques. Il propose des fonctionnalités d'enrichissement des cheminements biologiques, permettant aux utilisateurs d'accéder et de visualiser des informations sur les voies métaboliques via des bases de données publiques. *WebGestalt* se distingue également par sa flexibilité d'analyse et sa capacité à interagir avec différentes sources de données biologiques ce qui en fait un outil adapté aux projets nécessitant une analyse approfondie des réseaux biologiques.

*MetaboReport* est une autre plateforme importante pour l'analyse métabolomique, offrant une interface spécialisée pour générer des rapports automatisés des résultats d'analyse métabolomique. Contrairement aux autres plateformes, *MetaboReport* se concentre sur l'exportation rapide et systématique des résultats sous forme de rapports standardisés, ce qui est particulièrement utile pour la documentation et la communication des résultats dans des contextes cliniques ou de recherche. Bien qu'il ne propose pas autant de fonctionnalités interactives que *MetaboAnalyst* ou *WebGestalt*, il est particulièrement apprécié pour sa capacité à structurer des sorties de données claires et faciles à interpréter répondant ainsi aux besoins des chercheurs dans des cadres nécessitant une présentation structurée et synthétique des résultats.

Ces interfaces répondent aux besoins fonctionnels identifiés dans ce projet, notamment l'importation et la validation de fichiers de données, la visualisation de résultats sous forme graphique, et l'intégration avec des bases de données publiques comme *KEGG* pour l'analyse des voies métaboliques. Toutefois, la spécificité de ce projet réside dans le développement d'une plateforme web dédiée à l'analyse du métabolisme du virus H5N1, avec un accent sur la modélisation des interactions biologiques spécifiques à ce virus. L'objectif est de centraliser l'accès à divers outils bioinformatiques permettant de traiter des données biochimiques en entrée, de modéliser les interactions métaboliques spécifiques au H5N1 et de visualiser les résultats. L'outil proposé vise à offrir une interface unique adaptée aux besoins des scientifiques travaillant sur le H5N1, en intégrant des fonctionnalités de personnalisation et de connectivité accrues pour l'étude approfondie des processus métaboliques liés à ce virus.

## 3 Analyse des besoins

### 3.1 Besoins fonctionnels

- **Importation et traitement de fichiers d'analyse métabolique :**  
L'utilisateur doit pouvoir importer des fichiers tabulés contenant des données d'expression génique ou métabolique afin d'y effectuer les analyses métaboliques depuis la plateforme. L'application inclura un contrôle de qualité pour vérifier l'intégrité des données importées, en détectant les valeurs manquantes ou les types de données incorrects.
- **Recherche et affichage des réseaux métaboliques spécifiques au H5N1 :**  
L'application doit permettre à l'utilisateur de rechercher des réseaux métaboliques liés au virus H5N1, en utilisant l'API de KEGG. Une option permettra de sélectionner le pathway d'intérêt spécifique au H5N1 parmi les pathways disponibles dans KEGG. Elle affichera ces réseaux métaboliques de manière visuelle et interactive, avec des options de mise à jour en temps réel des pathways pour visualiser les modifications en fonction des données d'expression fournies par l'utilisateur.
- **Visualisation de graphes de co-expression :**  
L'application doit intégrer visNetwork pour la visualisation de graphes de co-expression. La visualisation inclura une interactivité avancée permettant de personnaliser la taille et la couleur des nœuds selon des valeurs d'expression ou d'autres attributs biologiques, et une légende dynamique pour indiquer la signification de ces options de style. Les utilisateurs pourront également exporter le graphe

sous forme d'image ou de fichier JSON.

— **Analyses statistiques et visualisations des données :**

L'application permettra la visualisation et l'analyse statistique de ces données, en intégrant plusieurs tests statistiques (t-test, ANOVA) et des visualisations comme les *boxplots* et *heatmaps*. Les utilisateurs auront la possibilité de personnaliser les options de visualisation, comme le choix des couleurs et des styles de graphes, pour adapter l'affichage aux besoins de chaque analyse.

## 3.2 Besoins non fonctionnels

— **Gestion d'une interface utilisateur conviviale et interactive :**

L'application doit être facilement navigable grâce à une interface claire, avec des onglets distincts pour chaque fonction (recherche, analyse, visualisation). Les résultats des analyses et des visualisations doivent être affichés de manière claire et compréhensible.

— **Performance :**

— **Taille des Fichiers :** Pour assurer des performances optimales et une expérience utilisateur fluide, une limite de taille de fichier de 10 Mo est recommandée pour les données importées. Les tests de performance seront effectués pour garantir que l'application reste réactive pour les fichiers jusqu'à cette limite.

— **Recommandations pour les utilisateurs :** Les utilisateurs sont encouragés à importer des fichiers de taille inférieure ou égale à 10 Mo pour des performances optimales. Dans le cas de fichiers de grande taille, il est recommandé de subdiviser les jeux de données si des problèmes de performance sont rencontrés.

— **Tests de Performance :** Des tests seront réalisés avec des fichiers de 1 Mo à 10 Mo pour évaluer l'impact sur la réactivité de l'interface. Les résultats de ces tests permettront d'ajuster les limites si nécessaire pour garantir une utilisation efficace de la plateforme.

— **Compatibilité :**

L'application doit être compatible avec différents systèmes d'exploitation (Windows, Mac, Linux) via un navigateur web moderne (Chrome, Firefox). Elle doit aussi être compatible avec des formats de fichiers standards (*CSV*, *TSV*, *SBML*, *SIF*, *GML*).

— **Accessibilité :**

L'interface doit être intuitive et accessible pour des utilisateurs avec peu d'expertise en bioinformatique. Les visualisations doivent être claires, et les interactions faciles à comprendre.

— **Maintenance et évolutivité :** Le code de l'application doit être maintenable et évolutif afin d'intégrer de nouvelles fonctionnalités à l'avenir. Les connexions avec des bases de données externes, comme KEGG ou Cytoscape, doivent être robustes et ajustables en cas de changement de services.

## 4 Engagements

### 4.1 Choix des outils :

#### Base de données :

Dans ce projet, nous avons choisi d'utiliser la base de données **KEGG** (Kyoto Encyclopedia of Genes and Genomes) pour l'analyse des voies métaboliques spécifiques au virus H5N1. Ce choix est motivé par plusieurs raisons.

Tout d'abord, **KEGG** est une ressource largement reconnue et régulièrement mise à jour, regroupant des informations sur les interactions moléculaires, les réactions chimiques, les génomes et les voies métaboliques [19]. KEGG est particulièrement pertinent pour des analyses comme la nôtre, car il permet d'explorer les voies métaboliques spécifiques des virus et de leurs interactions avec l'hôte, tout en facilitant la visualisation

des données grâce à ses cartes de pathways métaboliques [20].

Ensuite, **KEGG** s'intègre parfaitement avec les outils que nous utilisons dans notre projet, notamment **Cytoscape** pour la modélisation des réseaux biologiques et **RShiny** pour l'interface utilisateur. Grâce à des outils comme **Pathview**, il est possible d'analyser et de visualiser efficacement les données métaboliques sur la plateforme KEGG [21]. De plus, des extensions comme **KEGG\_Extractor** facilitent l'extraction et l'analyse des orthologues fonctionnels, ce qui est crucial pour l'étude des processus métaboliques du virus H5N1 [20].

Enfin, l'utilisation de KEGG pour ce type d'analyses est appuyée par de nombreuses études récentes. KEGG est non seulement une base de données fiable, mais elle est également utilisée pour l'exploration de données à grande échelle, notamment dans les domaines de la virologie et de la génomique [19, 21].

### Visualisation et analyse :

- **ggplot2 pour la visualisation de données :**

*ggplot2* est un package de visualisation graphique flexible et largement adopté dans la communauté R pour sa capacité à produire des graphiques esthétiques et personnalisables. Il repose sur la "grammaire des graphiques" pour simplifier la création de visualisations complexes et intuitives, facilitant ainsi l'interprétation des données biologiques [9, 22].

- **plotly pour la visualisation interactive :**

Nous avons choisi d'utiliser **plotly** pour la création de graphiques interactifs dans l'application. **plotly** permet de transformer des graphiques statiques en visualisations dynamiques, offrant aux utilisateurs des fonctionnalités telles que le zoom, le survol pour afficher des détails, et la sélection de données. Par rapport à d'autres outils de visualisation, **plotly** est particulièrement avantageux pour les applications web, car il s'intègre facilement avec R et *ggplot2*, tout en conservant une expérience utilisateur fluide et intuitive [23, 10].

Les graphiques interactifs produits par **plotly** facilitent l'exploration de grandes quantités de données, ce qui est essentiel dans le contexte des analyses biologiques, où les utilisateurs ont besoin de comprendre rapidement les tendances et les anomalies [22].

- **Pathview pour la visualisation de pathways métaboliques :**

Nous avons sélectionné *Pathview* pour mapper les données d'expression génique et métabolique directement sur les pathways KEGG. *Pathview* est reconnu pour sa capacité à intégrer les données omiques sur des pathways métaboliques et à afficher les variations d'expression, ce qui est essentiel pour interpréter les réseaux métaboliques du H5N1 [12].

- **visNetwork pour la visualisation de graphes de co-expression :**

Le package *visNetwork* offre une interface interactive pour explorer les réseaux de co-expression. Basé sur *vis.js*, il est particulièrement utile pour des visualisations de réseaux où l'exploration des relations entre gènes ou métabolites est nécessaire. *visNetwork* permet aussi une personnalisation de la visualisation, comme la taille et couleur des nœuds selon les valeurs d'expression [14].

## 4.2 Outils à développer :

- **Interface utilisateur (UI) :**

Créer une interface utilisateur avec *RShiny*, divisée en plusieurs onglets : recherche de pathways, analyse de données, visualisation dynamique des réseaux, et résultats des analyses statistiques. L'interface doit être intuitive, avec des graphiques clairs et des tableaux interactifs, permettant une navigation fluide entre les différentes sections de l'application.

- **Outil d'importation de fichiers :**

Créer un outil permettant aux utilisateurs d'importer des fichiers tabulés d'expression génique ou métabolique. Seuls les formats CSV, TSV, XLS, XLSX seront supportés pour garantir la compatibilité des données. Un contrôle de qualité analysera les fichiers importés pour s'assurer de la validité des données, en vérifiant la cohérence des colonnes et la compatibilité des formats.

- **Transformation et extraction de données pour visNetwork :**  
Développer un script qui transformera les données importées en tables de nœuds et d'arêtes pour visNetwork, en fonction des critères de co-expression ou de relation métabolique.
- **Recherche et intégration de données KEGG :**  
Créer un outil utilisant l'API KEGG pour extraire les informations de pathways spécifiques au H5N1. Cette fonction importera les données de réseau pour les visualiser et les annoter dans visNetwork.
- **Section de visualisation dynamique avec Pathview :**  
Intégrer Pathview pour superposer des données d'expression sur les pathways et les visualiser. Un système de mappage permettra de relier les valeurs d'expression aux métabolites dans le pathway. Chaque mise à jour des données d'expression relancera automatiquement la visualisation.
- **Statistiques et visualisations :**  
Intégrer des options d'analyse statistique de base (t-test, ANOVA) et des visualisations personnalisables comme les boxplots et heatmaps. Les utilisateurs pourront configurer les paramètres des tests et personnaliser les graphiques pour une meilleure interprétation des résultats.

## 5 Technologies et langages utilisés

### 5.1 Langages et frameworks

- **Importation des données :**  
Utilisation de R avec *read.csv* pour les fichiers CSV et TSV, et *readxl* pour les fichiers XLS et XLSX.
- **Visualisation des données :**  
Utilisation de R avec *ggplot2* pour la création de graphiques personnalisables, et *plotly* pour l'interactivité des graphiques, permettant des analyses visuelles plus approfondies.
- **Interfaçage avec l'API KEGG :**  
Utilisation de R avec le package *KEGGREST* pour l'interaction avec l'API KEGG et la récupération des pathways.
- **Visualisation de pathways métaboliques :**  
Utilisation de R avec *Pathview* pour mapper les données d'expression sur les pathways KEGG et visualiser les variations.
- **Interface web :**  
Utilisation de *RShiny* pour créer une interface interactive, permettant aux utilisateurs de naviguer facilement entre les différentes sections de l'application.

# Chapitre 2

## Organisation

### 1 Architecture du système

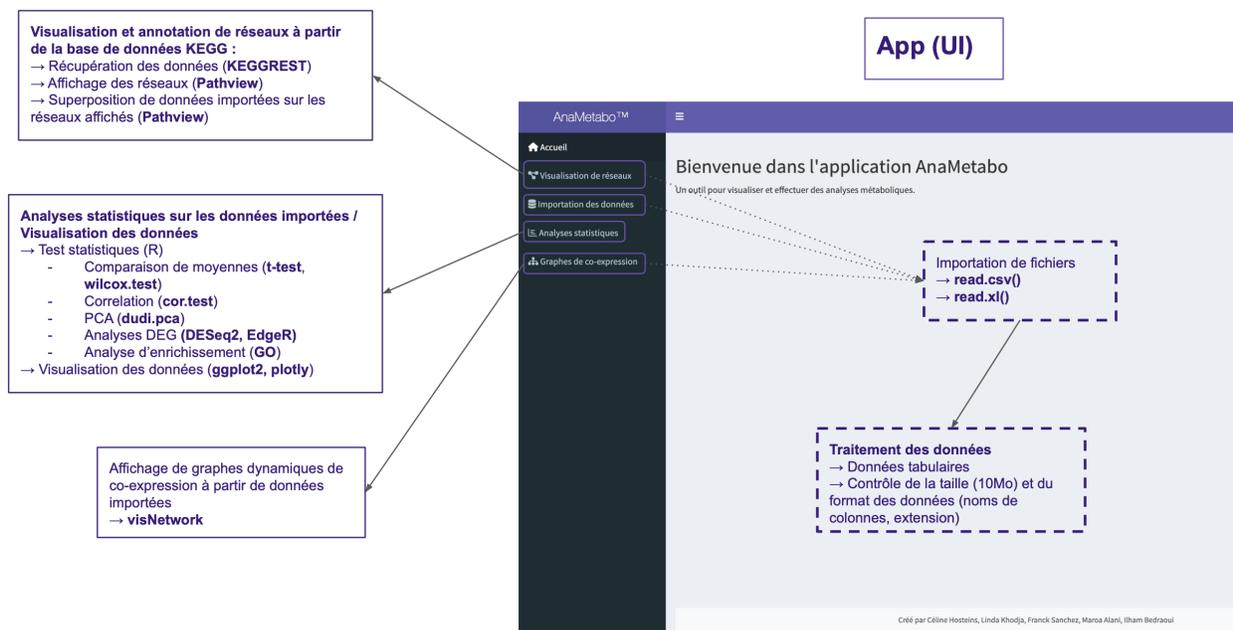


FIGURE 2.1 – Représentation de l'architecture du projet

### 2 Répartition des tâches

- Chaque membre de l'équipe sera responsable d'une partie spécifique de l'outil
- Importation des fichiers CSV :
- Visualisation des données :
- Analyse statistique :
- Interfaçage avec l'API KEGG :
- Interface web avec RShiny :

#	Tâche	Responsable	Date de début	Date de fin
1	Version 1 cahier des charges	Equipe	30/09/2024	14/10/2024
2	Presentation cahier des charges	Equipe	14/10/2024	14/10/2024
3	Version 2 cahier des charges	Equipe	14/10/2024	20/10/2024
4	RDV 1 Remise cahier des charges	Equipe	21/10/2024	21/10/2024
5	Version finale cahier des charges	Equipe	21/10/2024	29/10/2024
6	Architecture du code	Equipe	29/10/2024	04/11/2024
7	RDV 2 remise cahier des charges et architecture	Equipe	05/11/2024	05/11/2024
8	Importation des fichiers CSV	Ilham		
9	Visualisation des données	Maroa		
10	Analyses statistiques	Linda		
11	Interface avec l'API KEGG	Céline		
12	interface web avec Rshiny	Franck		
13	RDV Version 1 du code	Equipe	05/11/2024	11/11/2024
14	RDV Version 2 du Code	Equipe	12/11/2024	25/11/2024
15	Début écriture Rapport	Equipe	18/11/2024	02/12/2024
16	Finalisation du code	Equipe	25/11/2024	02/12/2024
17	Finalisation du Rapport	Equipe	03/12/2024	06/12/2024
18	Ajout besoins non fonctionnels	Equipe	06/12/2024	12/12/2024
19	Préparation présentation plus entrainement	Equipe	06/12/2024	15/12/2024
20	Soutenance	Equipe	16/12/2024	16/12/2024

FIGURE 2.2 – Répartition

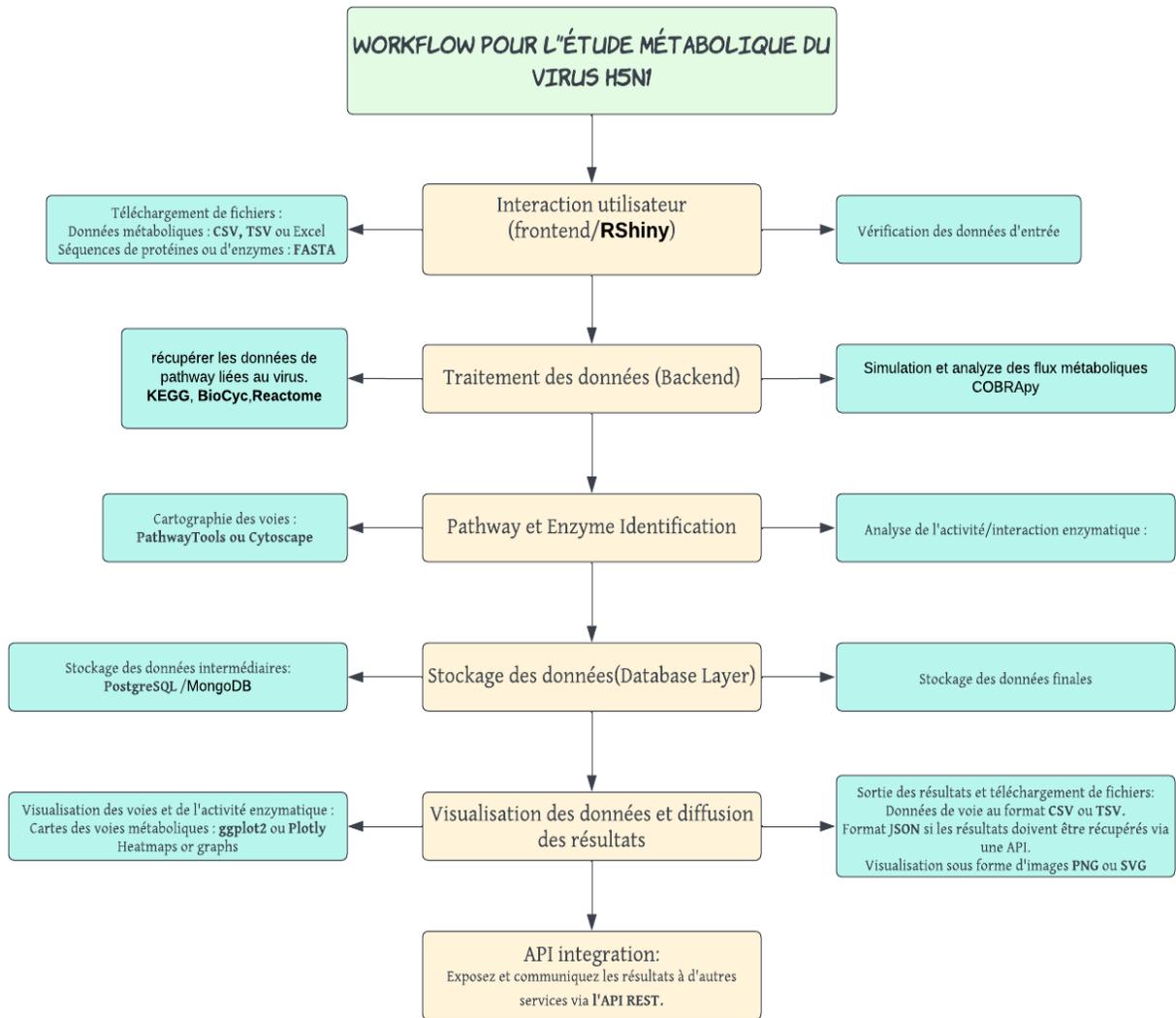


FIGURE 2.3 – Workflow d'analyses métabolique pour le virus H5N1

## 2.1 Diagramme de Gantt

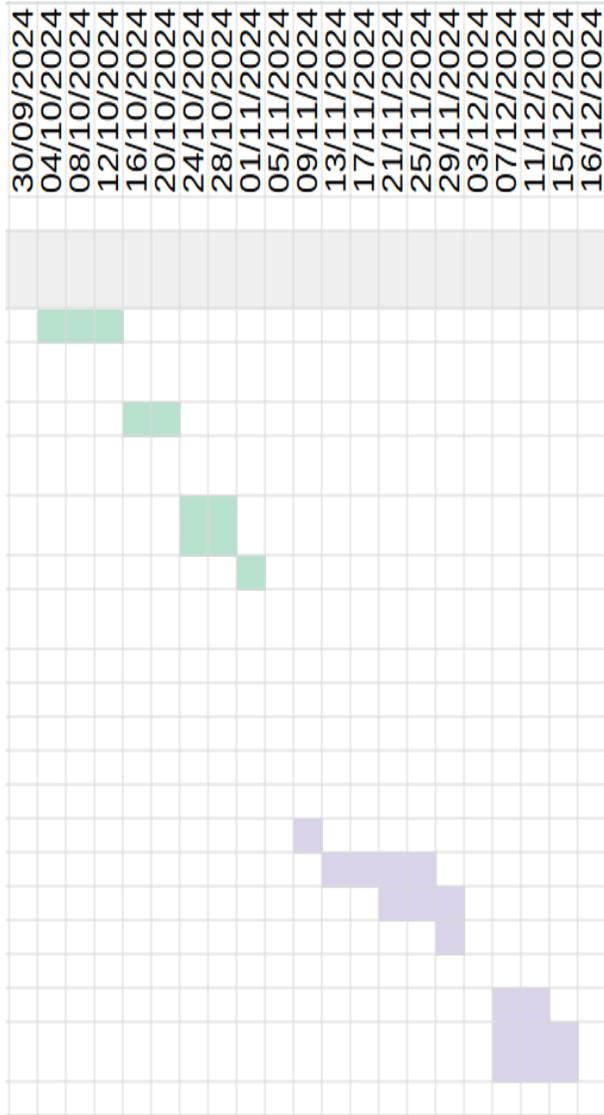


FIGURE 2.4 – Diagramme de Gantt

# Bibliographie

- [1] Organisation mondiale de la Santé. Grippe a(h5n1), 16 mai 2024.
- [2] S. Lambert, B. Durand, M. Andraud, et al. Two major epidemics of highly pathogenic avian influenza virus h5n8 and h5n1 in domestic poultry in france, 2020-2022. *Transboundary and Emerging Diseases*, 2022.
- [3] BRC. H5n1 outbreak overview, 2024.
- [4] PassionSanté.be. Grippe h5n1 : quels risques de contamination pour l’homme ?, 2024.
- [5] Chantal Loevenbruck. Cycle du virus de la grippe aviaire, 2024.
- [6] Miho Furumichi Minoru Kanehisa and other. Kegg : new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*, 2017.
- [7] Monica Riley Peter D Karp et al. The metacyc database. *Nucleic Acids Research*, 2002.
- [8] Lisa Matthews Karen Rothfels, Marija Milacic et al. Using the reactome database. *Curr Protoc*, 2023.
- [9] Hadley Wickham. *ggplot2 : Elegant Graphics for Data Analysis*. Springer, New York, 2nd edition, 2016.
- [10] Konstantinos H. Sidiropoulos, Hiroyuki Tamura, Darren W. Davis, et al. Interactive data analysis with r and plotly. *Journal of Open Source Software*, 2(8) :141, 2017.
- [11] Trey Ideker Keiichiro Ono, Barry Demchak. Cytoscape tools for the web age : D3.js and cytoscape.js exporters. *F1000Res*, 2014.
- [12] Weijun Luo and Cynthia Brouwer. Pathview : An r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 29(14) :1830–1831, 2013.
- [13] Owen Ozier Paul Shannon, Andrew Markiel. Cytoscape : a software environment for integrated models of biomolecular interaction networkss. *Genome Res*, 2003.
- [14] Almende B.V. and Contributors. visnetwork : Network visualization using vis.js library, 2023. R package version 2.1.2.
- [15] Bianca H Habermann Jose M Villaveces, Prasanna Koti. Tools for visualization and analysis of molecular networks, pathways, and -omics data. *Adv Appl Bioinform Chem*, 2015.
- [16] Zhiqiang Pang, Yao Lu, Guangyan Zhou, Fiona Hui, et al. Metaboanalyst 6.0 : towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Research*, 2024.
- [17] John M Elizarraras, Yuxing Liao, Zhiao Shi, et al. Webgestalt 2024 : faster gene set analysis and new support for metabolomics and multi-omics. *Nucleic Acids Research*, 2024.
- [18] Yonghui Dong and Sergey Malitsky. MetaboReport : from metabolomics data analysis to comprehensive reporting. *Bioinformatics*, 40(6), 06 2024.
- [19] M. Kanehisa and S. Goto. Kegg : Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 51(D1) :1–10, 2023.
- [20] S. Jia et al. Kegg\_extractor : An effective extraction tool for kegg orthologs. *Genes*, 14(2) :386, 2023.
- [21] W. Luo and C. Brouwer. Pathview : an r/bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, 39(2) :101–110, 2023.
- [22] Roger D. Peng. Reproducible research in computational science. *Science*, 334(6060) :1226–1227, 2011.
- [23] Carson Sievert. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. CRC Press, Boca Raton, 2020.