

Rapport de stage

du 3 avril 2023 au 23 juin 2023

**Direction des Systèmes d'Information du Département du
Territoire de Belfort**

Création d'une application web permettant au service des
archives départementales de référencer leurs documents

Maxime THEVENEAU

BUT2 Informatique (IUT de Belfort-Montbéliard 2022-2023)

Tuteur en entreprise : **PERIARD Béatrice**

Professeur référent : **NOURA Hassan**

Rapport confidentiel

Note de confidentialité

Afin de garantir la confidentialité des agents du département du Territoire de Belfort, les annexes II et III, concernant les organigrammes des services de la Direction des archives départementales et de la Direction des systèmes d'information, ne doivent pas être divulgués en dehors du cadre de l'évaluation pédagogique du stage.

Remerciements

Je remercie ma tutrice de stage, Madame **Béatrice PERIARD**, cheffe du service Innovation et Transformation Numérique de la Direction des Systèmes d'Information, pour son accueil ainsi que pour son suivi et ses conseils apportés lors de mon stage.

Je remercie l'ensemble des agents de ma direction (DSI) ainsi que mon binôme de stage Valentin BRIZARD et Julien GALLET pour leur aide et leurs conseils tout au long du projet. Je remercie également les agents de la Direction des archives départementales (DAD) pour avoir pu collaborer ensemble.

Je remercie mon professeur référent Monsieur **Hassan NOURA**, professeur à l'Institut Universitaire de Technologie de Belfort au département Informatique et chercheur en sécurité informatique et *machine learning*¹, ainsi que l'ensemble du corps enseignant et les intervenants de l'IUT de Belfort-Montbéliard pour l'ensemble des connaissances transmises lors de ces deux ans de formation.

Enfin je tiens à remercier toutes les personnes qui m'ont aidées pour ce rapport et pour le bon déroulement de mon stage.

¹ Le machine learning (ML) est une forme d'intelligence artificielle (IA) qui est axée sur la création de systèmes qui apprennent, ou améliorent leurs performances, en fonction des données qu'ils traitent.

Sommaire

Introduction.....	6
1. Présentation de la collectivité.....	8
1.2. Présentation du Conseil départemental.....	8
1.3. Présentation de la Direction des Systèmes d'Informations (DSI).....	8
1.4. Présentation de la Direction des Archives Départementales (DAD).....	9
2. Présentation du projet.....	11
2.2. Contexte.....	11
2.3. Process actuel.....	12
2.4. Objectifs du projet.....	13
2.4.1. Description des besoins.....	13
2.4.2. Difficultés.....	14
3. Cahier des charges.....	15
3.2. Description fonctionnelle.....	15
3.2.1. Importation des données.....	15
3.2.2. Gestion des données.....	15
3.3. Contraintes techniques.....	16
4. Mise en oeuvre.....	17
4.2. Organisation.....	17
4.3. Environnement de développement.....	17
4.4. Importation des données.....	18
4.4.1. Base de données.....	18
4.4.2. Récupération des données.....	20
4.5. Gestion des données.....	21
4.5.1. Connexion.....	21
4.5.2. Consultation.....	21
4.5.3. Ajout et Modification.....	23
5. Bilan du stage.....	24
5.2. Bilan pour le département.....	24
5.3. Bilan humain.....	24
5.4. Bilan pédagogique.....	25
Conclusion.....	26
Table des figures.....	27
Sitographie.....	28
Table des annexes.....	29
Annexes.....	30

Introduction

Aujourd'hui les archives du Territoire de Belfort ce sont plus de 650 mille fichiers numérisés depuis plus de 20 ans, plus de 1 million de lignes dans des tableaux Excel et seulement 5 % du total à numériser réalisé.

La numérisation est une nouvelle clé essentielle pour les Archives départementales et c'est dans ce cadre que s'inscrit mon stage de fin de 2ème année à l'IUT de Belfort-Montbéliard du 3 avril 2023 au 23 juin 2023 durant lequel j'ai pu participer au développement d'une nouvelle solution pour répondre aux besoins exponentiels de la Direction des Archives Départementales (DAD) concernant la numérisation des archives du Territoire.

En effet, depuis au moins 20 ans, la DAD s'est convertie au numérique et à l'usage de la numérisation dans un objectif de conservation et de partage des archives départementales. Les documents conservés sont pour la plupart tous très anciens, le plus ancien date du XIVème siècle. La numérisation permet aux archivistes de les partager de manière numérique sans avoir à les déplacer et les manipuler afin d'éviter une dégradation importante dans le temps et permettre une consultation libre des usagers. Ainsi les archives sont numérisées par ordre de priorité et selon des critères définis avec un objectif à très long terme qui serait de dématérialiser l'ensemble des documents pour la consultation, sachant qu'aujourd'hui, seulement 5 % des documents ont pu être numérisés. Ces numérisations de documents génèrent plusieurs données qui concernent la répertorisation et l'endroit de stockage des fichiers dans des espaces partagés et qui sont ensuite stockés dans plusieurs fichiers Excel comprenant de nombreuses colonnes.

C'est dans ce cadre que s'inscrit l'objet de mon stage qui consiste en la conversion des données Excel dans une nouvelle base de données dans un souci de sécurité et la création d'un site web pour consulter, gérer et analyser ces données de numérisation. Ce projet est totalement nouveau et nous partons de zéro, l'objectif étant de concevoir et produire une nouvelle manière pour la DAD de manipuler les numérisations d'archives, c'est en partie pour cela que le sujet de stage m'attire particulièrement.

A travers ce rapport, nous découvrirons l'environnement de travail de la collectivité du Département du Territoire de Belfort, l'objectif de ce stage, le cahier des charges, sa mise en œuvre et le bilan tant face aux compétences qu'aux difficultés liées au projet.

1. Présentation de la collectivité

1.2. Présentation du Conseil départemental²

Au sens strict, le Conseil départemental (anciennement conseil général) du Territoire de Belfort est l'assemblée délibérante du département en tant que collectivité territoriale, formée par la réunion des conseillers départementaux élus pour six ans. Dans un sens plus général, ce terme a fini par désigner la collectivité elle-même. Il est dirigé par 18 élus répartis sur 9 cantons et présidé par Monsieur Florian Bouquet. Le Conseil départemental est composé de 18 directions chargées de décliner la politique mise en œuvre par les élus dans des domaines précis comme l'action sociale (protection de l'enfance, Revenu social de solidarité RSA), la voirie, l'éducation (gestion des collèges), la culture, le développement local, le logement, le tourisme ou encore le financement des SDIS (sapeurs-pompiers), des domaines aujourd'hui obligatoires pour tous les Conseils départementaux. Celui du Territoire de Belfort agit dans le but de saisir tous les sujets d'intérêt public local et développe de nouvelles compétences dans différents domaines tels que l'aménagement du territoire, l'environnement, le soutien à l'agriculture, les transports ou l'enseignement supérieur.

Pour mettre en œuvre l'ensemble de ces compétences, le Département dispose d'un budget d'un montant de 179,8 millions d'euros pour l'année 2023 et d'environ 1000 agents techniques et administratifs, les assistants familiaux chargés d'accueillir les enfants confiés au département et les ouvriers spécialisés (OS) comme les cuisiniers et les agents d'entretien des collèges.

1.3. Présentation de la Direction des Systèmes d'Informations (DSI)³

La Direction des Systèmes d'Information fait partie de la Direction Générale Adjointe Ressources (DGAR) qui a pour objectif d'apporter un soutien aux autres directions et d'être une fonction support pour l'ensemble de la collectivité et de ses agents. La direction est

² cf. [Annexe I : Organigramme des Services départementaux](#)

³ cf. [Annexe II : Organigramme du Service des Systèmes d'Information](#)

dirigée par Monsieur Lahid LECHEKHAB et est composée de 20 agents répartis en 2 services.

Le premier de ces services est le Centre de Service chargé du support aux usagers tels que les agents départementaux et le réseau des collèges du département. Il est aussi chargé de la sécurité des systèmes d'information sur un ensemble d'environ 2500 postes informatiques répartis sur l'ensemble du Territoire de Belfort.

Le second service de la DSI est le service de l'Innovation et de la Transformation Numérique chargé de dynamiser la transformation des outils numériques de la collectivité et d'accompagner les différents services dans la gestion et l'utilisation des logiciels métiers utiles à leurs missions. C'est au sein de ce service que mon stage se déroule sous la responsabilité de la cheffe de service, Madame Béatrice PERIARD.

Ma mission principale est d'accompagner la Direction des Archives départementales dans la modernisation de leurs outils.

1.4. Présentation de la Direction des Archives Départementales (DAD)⁴

La Direction des Archives Départementales fait partie de la Direction Générale Adjointe Territoire (DGAT) qui regroupe l'ensemble des directions techniques qui permettent d'assurer la gestion des compétences départementales. La Direction est dirigée par Madame Aude SEILLAN et est composée de 12 agents. Les missions⁵ des Archives départementales sont très variées:

Premièrement, le rôle des archives est de conseiller et d'apporter un contrôle scientifique et technique. Elles exercent une mission de conseil et de contrôle des services administratifs et des collectivités locales qui sont de gros producteurs d'archives publiques en raison de leur mission de service à la population.

Deuxièmement, les administrations et les services publics du département sont tenus de transmettre régulièrement leurs archives. Tout ne doit pas être conservé définitivement dans

⁴ cf. [Annexe III: Organigramme du Service des Archives Départementales](#)

⁵ cf. [vidéo youtube: le classement des archives départementales](#)

la multitude de documents produits. Il existe donc des tableaux de gestion qui précisent quels types de documents sont à éliminer après leur Durée d'Utilité Administrative (DUA⁶), certains sont à conserver absolument et d'autres vont faire l'objet d'un tri en vue d'une conservation à titre d'échantillon pour les historiens futurs. Les Archives ont un rôle fondamental car elles sont la mémoire de l'action des administrations et des citoyens.

Troisièmement, les Archives ont été séparées en périodes historiques : avant et après la Révolution, après juillet 1940. Un cadre de classement alphabétique a été imaginé pour organiser et retrouver les documents. Une fois le classement d'un versement effectué, un inventaire est dressé pour que l'administration et le public puissent savoir où trouver les documents qui les intéressent. C'est à ce stade que l'objet de mon stage prend tout son sens en permettant aux Archives d'avoir recours à des outils numériques efficaces pour les assister dans le classement et la répertorisation de l'ensemble des documents.

Quatrièmement, les documents conservés aux Archives départementales le sont sans limite de durée. Ils font donc l'objet d'une politique de conservation qui vise à les préserver des altérations physiques. Cela comprend un contrôle de l'air, de l'humidité, de la lumière, la possibilité de procéder à des restaurations de documents fragilisés ou de faire réaliser des copies de substitution pour éviter la dégradation liée à une consultation trop fréquente d'un document.

Cinquièmement, la communication des archives au public se fait en salle de lecture. La loi sur les Archives de 1979 (modifiée en 2008) définit l'accès aux documents dans un double souci : transparence des décisions administratives pour les citoyens, mais aussi protection de la vie privée de chacun. Ainsi des délais de communicabilité sont mis en place. Cependant la grande majorité des documents d'archives sont accessibles immédiatement à tous.

Dernièrement, la valorisation des archives permet de partager le patrimoine archivistique départemental avec le plus grand nombre. Cette valorisation prend plusieurs formes comme l'action culturelle⁷ ou l'action éducative⁸.

⁶ La DUA est une durée pendant laquelle les archives, les documents et les données doivent être conservés afin de pouvoir être consultés et utilisés par les services administratifs ou par les citoyens.

⁷ L'action culturelle consiste en l'organisation de visites pour des groupes, l'accueil des visiteurs lors des journées du patrimoine, la réalisation d'expositions et de conférences pour mettre en valeur le patrimoine conservé

⁸ L'action éducative permet l'accueil d'élèves pour les confronter aux sources et la réalisation pour les enseignants de dossiers pédagogiques sur des thématiques diverses

Figure 1: 15 km linéaire de documents d'archives du Territoire de Belfort



2. Présentation du projet

2.2. Contexte

Les Archives du Territoire de Belfort ont de nombreuses missions; une partie de leur activité est de répertorier et entreposer les différentes archives et documents matériels de l'ensemble du territoire. Depuis environ 20 à 25 ans, la direction des archives a commencé à numériser ses documents d'archives dans différents formats (bitmap, jpeg, tif) qui sont principalement:

- Les collections de presse depuis la 1ère moitié du 19ème siècle
- Les archives iconographiques
- Les archives contemporaines
- Les archives privées (Avocats, Alstom, etc.)
- Les archives modernes

- Toutes les successions
- Les archives de la préfecture

Il existe actuellement 3 types de numérisation:

- **Interne:** les documents sont numérisés par la DAD pour elle-même et modifiés si besoin.
- **Externe:** les documents sont numérisés par la DAD pour un service ou une personne tierce.
- **Sous-traitance:** les documents ont été numérisés par un service ou une personne tierce puis donnés à la DAD.

Aujourd'hui les référencements des fichiers numérisés sont conservés dans 3 tableaux Excel eux-même contenus dans un disque dur. Ce fonctionnement est fragile et peut entraîner une corruption ou une perte de données conséquentes.

C'est pourquoi un espace numérique dédié, appelé le "partage X", a été créé. Il existe également un dossier protégé nommé "FRAD090_NUM" rangé par ordre alphabétique. Ainsi les données des tableaux Excel permettent de retrouver un document dans ces espaces de stockage partagés.

2.3. Process actuel

Avant même de pouvoir imaginer et concevoir une solution pour répondre aux besoins de la DAD, la première étape de ce projet est de comprendre l'environnement de travail des usagers archivistes et leur processus actuel de numérisation afin de mieux appréhender les enjeux du projet et les points critiques.

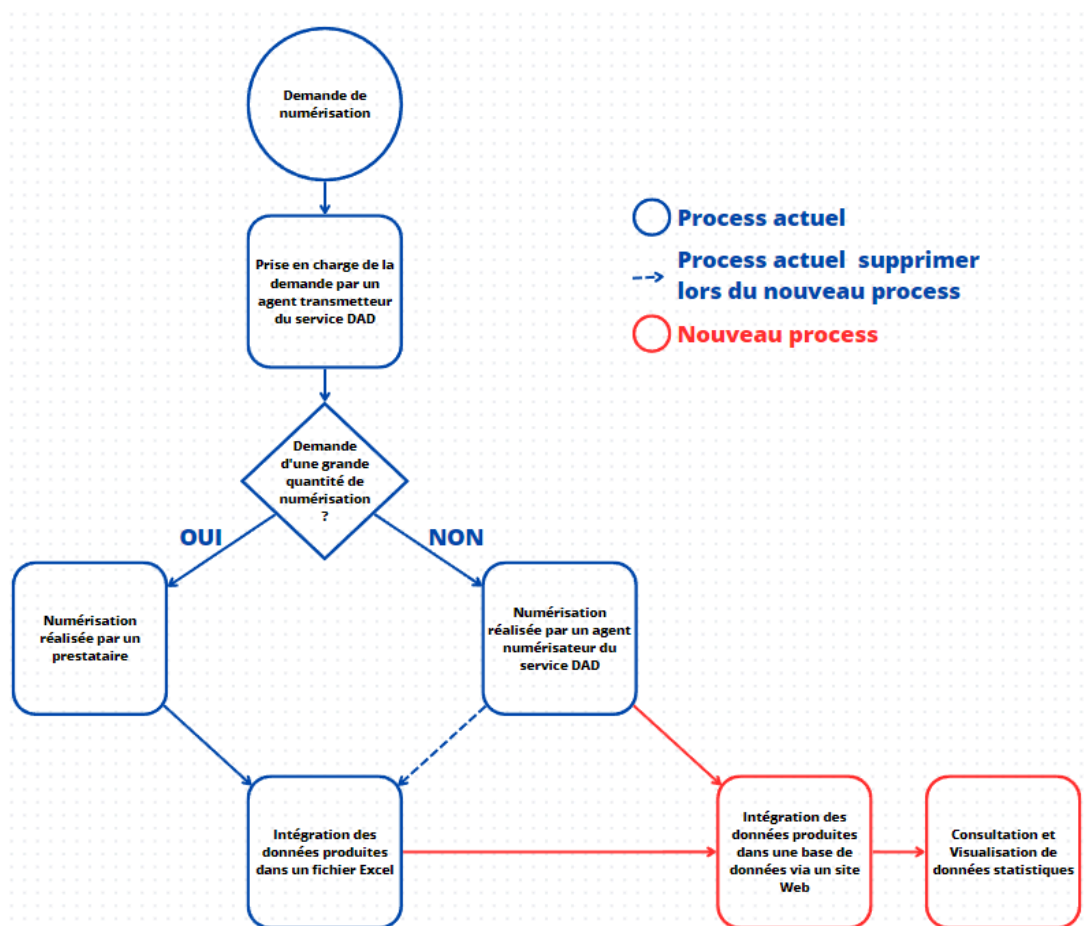
Ce processus est engagé lors d'une demande de numérisation d'archives à la DAD et poursuivi par la prise en charge d'un agent transmetteur du service. A ce moment précis, la numérisation d'archives empreinte 2 branches distinctes en fonction de sa quantité et du volume de documents à numériser. Au-delà de dizaines de milliers de numérisation, la direction des archives fait appel à un prestataire extérieur qui s'occupe de numériser les documents et qui renvoie les données générées dans un tableau Excel. En dessous, la quantité reste acceptable par le service des Archives qui numérise lui-même les documents par l'intermédiaire d'un agent numériseur propre au service. Dans les deux cas, des données

propres à la numérisation concernant le répertoriage des documents sont produites et stockées dans des fichiers Excel.

Il y a actuellement 3 fichiers Excel de plus de 100 000 lignes car tout ne peut rentrer dans un seul fichier et même dans ce cas, la manipulation des fichiers reste instable dû à la taille des fichiers rendant leur utilisation dangereuse (bugs réguliers, temps de réponse important à l'ouverture, etc.).

Ainsi, la DAD exprime une demande de restructuration du processus actuel afin de garantir un stockage des données plus fiable et une manipulation de ces dernières plus facile.

Figure 2: modélisation des processus



2.4. Objectifs du projet

2.4.1. Description des besoins

L'objectif du projet est double afin de répondre aux besoins du service des archives. En effet, dans un premier temps, le but est de concevoir un modèle conceptuel de données respectant les liens entre chaque champ de chaque colonne dans les fichiers Excel et ensuite d'intégrer l'ensemble des 3 fichiers Excel dans la base de données. Pour cela il est important de comprendre la signification de chaque colonne et les différents liens entre elles afin de ne pas errer le modèle. Dans un second temps, l'objectif est de créer un site web permettant aux agents de gérer et consulter les données d'archives de la base de données. A terme, ils pourront rechercher une archive via des filtres, la modifier, en ajouter des nouvelles ou en supprimer et obtenir des données statistiques sur le nombre de consultations de chaque archive ou le pourcentage d'avancement de numérisation de chaque série.

2.4.2. Difficultés

La réalisation de ce projet se confronte à priori à plusieurs difficultés dont principalement :

- La compréhension du besoin métier des Archives

En effet, la première contrainte liée à ce projet est la nécessité de comprendre les besoins métier de la direction des archives départementales. Il faut ainsi s'appliquer à utiliser un langage compréhensible entre chaque service et il est même nécessaire de reformuler chaque besoin pour garantir une mise en œuvre du projet en adéquation avec les besoins exprimés par les usagers.

- La conception d'un Modèle Conceptuel de Données (MCD⁹) avec une vision imprécise de la relation entre ces dernières

⁹ Le schéma conceptuel de données, également appelé MCD pour modèle conceptuel de données, est une représentation claire des données du système d'information à concevoir. Cette représentation en outre figure les relations entre ces données.

Afin de pouvoir lier les données des fichiers Excel, le modèle conceptuel est une étape importante du projet qui permet de définir les différentes relations entre chaque donnée et qui nécessite une vision d'ensemble de celles-ci. Lors de ce projet, une des difficultés est de comprendre ces relations car les fichiers Excel ne s'articulent pas tous exactement de la même manière, les données ayant été rentrées sur plus de 20 ans avec des critères et des besoins changeants. A cela s'ajoute l'usage de données erronées dû au format Excel qui ne permet pas un contrôle d'entrer des valeurs pour chaque colonne et qui se répercute dans l'import des données dans la base de données.

- Le nombre très important de données

Aujourd'hui le problème le plus significatif du projet est la masse de données qui ne représente qu'une partie infime de l'ensemble des archives à numériser. La récupération des informations des fichiers Excel ne peut se faire de manière linéaire car elle demanderait un coût en ressource trop important notamment par un nombre excessif de requêtes dans la base de données. Il faut ainsi concevoir une stratégie d'import des données pour minimiser le temps et le coût de l'import.

3. Cahier des charges

3.2. Description fonctionnelle

3.2.1. Importation des données

L'ensemble des données de numérisation d'archives doit pouvoir être importé dans la base de données de manière intègre en respectant ces règles:

- Intégrité de la donnée

Chaque donnée insérée dans la base de donnée doit correspondre en tout point aux données des tableaux Excel et les informations manquantes doivent être au maximum complétées automatiquement si cela est possible afin de minimiser le nombre d'erreurs et maximiser le nombre de données importées.

- Rapport d'erreurs

En cas de manquement d'informations sur une ligne du tableau Excel qui ne peut être ajoutée automatiquement, une erreur doit être produite et reportée afin de connaître les numéros de lignes des données non insérées dans la base de données à la fin du processus d'importation.

3.2.2. Gestion des données

La base de données constitue le fond du projet et celle-ci doit pouvoir permettre aux agents de la DAD de gérer leurs numérisations à travers un site Web respectant certains critères:

- Navigation simple

Le site web doit pouvoir permettre de trouver une information rapidement sans cliquer sur beaucoup d'onglets pour atteindre la donnée voulue. Ainsi l'accueil du site permet une recherche directe via le nom complet d'une archive et un menu permet d'accéder à la consultation sous forme de tableau de bord des archives.

- Tableau de bord

La consultation est l'élément principal du site car il permet de rechercher une archive afin de la consulter, la modifier ou la supprimer. Ainsi il est composé de filtres de recherches précis et organise les résultats sous forme de tableau simple et lisible.

- Droit d'accès

L'objectif de ce site web est de permettre aux archivistes de gérer les numérisations mais également de pouvoir accorder un droit de consultation unique à d'autres usagers et cela à travers un système de droits à différents niveaux pour une lecture seule ou une modification en fonction du grade.

- Statistiques

L'avantage de ce projet est de pouvoir générer des données statistiques automatiques et en temps réel afin de renseigner les agents de la DAD sur des données jusqu'alors incertaines et difficiles à obtenir.

3.3. Contraintes techniques

Dans un souci de constante augmentation des besoins et de continuité du projet, les fonctionnalités développées se doivent d’être au maximum génériques afin d’être exploitées par la suite, notamment l’importation des données en masse qui pourra être intégrée dans le futur au site web avec la possibilité de créer un tableau Excel et de l’importer dans la base de données, surtout lorsque la numérisation s’effectue via un prestataire qui renvoie un tableau Excel.

De plus, le projet s’inscrivant dans une démarche à long terme, les technologies utilisées se doivent d’être modernes et d’actualité pour faciliter la reprise par d’autres développeurs tels que des stagiaires.

4. Mise en oeuvre

4.2. Organisation

Afin de mener à bien ce projet, une organisation suivant la méthode agile a été instaurée à travers un logiciel propre au Département nommé “Gouti¹⁰”. Ce dernier permet de définir dans un premier temps les grandes étapes du projet pour délimiter le cadre général de celui-ci puis, par la suite, de fonctionner sous forme de sprint. Chaque sprint se concentre sur des fonctionnalités à développer et attendues par la DAD. Il aboutit à une réunion avec les agents archivistes pour qu’ils puissent donner leur avis sur les fonctionnalités produites et formuler de nouvelles demandes à intégrer lors du sprint suivant. Cette méthode de travail est en adéquation avec le projet car celui-ci évolue très vite et de nouveaux besoins se créent constamment, ainsi chaque besoin peut être ajouté sans difficultés dans le processus du projet.

La communication entre chaque membre de l’équipe est importante et chaque matin une mise au point d’une dizaine de minutes sur le travail réalisé la veille est organisée par la cheffe du service. Cela permet d’être au courant de l’avancée du sprint en cours et des difficultés rencontrées. De plus, une réunion plus longue est planifiée tous les jeudi dans le but de faire un point sur le projet dans sa globalité et de définir les axes principaux de développement

¹⁰ cf. [Annexe IV: Tableau de bord Gouti du projet](#)

pour la suite, il permet également de faire un point sur les différentes réunions organisées avec la DAD et d'intégrer les nouveaux besoins.

L'ensemble du code est partagé via le gestionnaire Gitlab¹¹ sur le compte du Département du Territoire de Belfort.

4.3. Environnement de développement¹²

L'objectif du choix des technologies utilisées est de permettre de réaliser le projet en un minimum de temps avec des langages simples d'utilisation et facilement modulable pour être repris par d'autres développeurs dans le futur. La première idée était de développer la partie Back-End¹³ en Php et le Front-End¹⁴ en HTML et CSS natif mais le choix a été orienté par une technologie plus moderne et moins contraignante, le NodeJS et le VueJS.

En effet l'avantage d'utiliser la technologie NodeJS est qu'elle est parfaitement adaptée pour les API Rest¹⁵ et permet donc d'exécuter plusieurs requêtes vers le serveur simultanément, car c'est un système « single thread » non bloquant. De plus cette technologie offre de bonnes performances, beaucoup de flexibilité et repose sur le langage Javascript, un langage très largement répandu et donc facile à comprendre. Enfin l'utilisation du VueJS est également basée sur du Javascript ce qui diminue grandement le temps de développement en uniformisant les 2 parties (Back et Front). De plus, Vue JS est un framework relativement léger et très complet.

Concernant la base de données, le choix de MySQL est en partie dû au fait qu'il est totalement open source et gratuit, ses performances sont excellentes et il est en plus

¹¹ Gitlab est une plateforme de développement collaborative open source éditée par la société américaine du même nom. Se basant sur les fonctionnalités du logiciel Git, elle permet de piloter des dépôts de code source et de gérer leurs différentes versions.

¹² cf. [Sitographie](#) - Technologies

¹³ **Back-End** (Back): partie d'un système informatique ou d'une application qui n'est pas directement accessible par l'utilisateur, généralement responsable du stockage et de la manipulation des données

¹⁴ **Front-End** (Front): partie d'un système informatique ou d'une application avec laquelle l'utilisateur interagit directement.

¹⁵ cf. ["Qu'est ce qu'une API Rest ?"](#), IBM

multi-threadé et multi-utilisateurs. De plus, un module est fourni avec NodeJS pour l'interaction entre le Back et la base de données.

Enfin, le projet mettant à contribution plusieurs personnes, la base de données se doit d'être partagée et cela grâce à la conteneurisation via docker¹⁶ afin que chacun traite les mêmes données.

4.4. Importation des données

4.4.1. Base de données

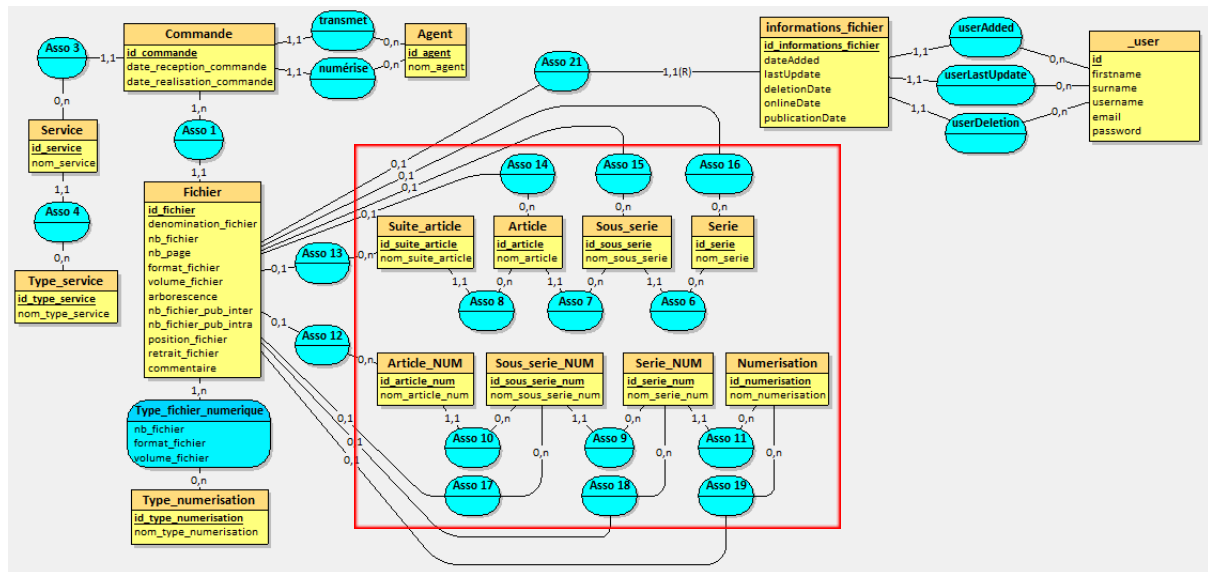
La première phase du projet est l'importation des données de numérisation des archives des fichiers Excel dans la base de données. Cette phase débute par la création d'un modèle conceptuel de données qui évolue en fonction des besoins mais également de la compréhension des relations entre les données des fichiers Excel qui s'articulent comme tel:

Figure 3: échantillon d'un tableau de données de numérisations Excel

Ainsi plusieurs versions de modèle de conception des données (MCD) ont été réalisées après avoir constaté des erreurs de conception pour correspondre au mieux au besoin du projet.

¹⁶ La dockerisation est une forme de virtualisation utilisée pour les applicatifs. Le principe repose sur la création de plusieurs espaces isolés les uns des autres sur un noyau commun. Dans ce cadre, Docker assure la création et la manipulation de conteneurs.

Figure 4: modèle de conception des données



La partie du MCD concernant les données de série, sous-série, article, suite-article (cf. [Figure 4: MCD](#), **cadre rouge**) a été plusieurs fois modifiée car malgré le fait que chaque donnée s'imbrique (une suite-article appartient à un article, un article appartient à une sous-série, une sous-série appartient à une série), connaître un fils ne permet pas de déduire un père. Concrètement la série A peut contenir une sous-série C mais la série B peut également contenir une sous-série C qui n'est pas la même que la précédente. Ainsi savoir que le fichier appartient à une sous-série C ne permet pas de déduire sa série et cela pour chaque donnée de cette partie.

4.4.2. Récupération des données

La prochaine étape est de développer la partie serveur (Back-End) afin de pouvoir se connecter à la base de données pour la créer et la remplir. Un dossier permet de référencer le code concernant la connexion à la base de données et l'importation des données et contient différents modules développés dont le principal permettant l'import des fichiers Excel dans la base de données : *dataFetcher.js*

Celui-ci renvoie un tableau d'erreurs contenant les numéros de ligne des fichiers dont les données sont erronées et le code erreur. Cela permet de transmettre ses informations à la DAD qui pourra corriger les lignes concernées.

Figure 5: log de l'insertion des données

```
{
  "success": 1,
  "data": {
    "time": null,
    "insertServices": {
      "time": 933,
      "types_services": [ ... ], // 16 items
      "services": [ ... ], // 140 items
      "errors": [
        581,
        582
      ]
    },
    "insertAgents": { ... }, // 13 items
    "insertSeries": {
      "time": 10928,
      "series": [ ... ], // 36 items
      "sous_series": [ ... ], // 148 items
      "articles": [ ... ], // 767 items
      "suites_articles": [ ... ], // 1078 items
      "errors": [
        1701
      ]
    }
  },
  "insertSeriesNum": {
    "time": 108,
    "num": [ ... ], // 3 items
    "series_num": [ ... ], // 4 items
    "sous_series_num": [ ... ], // 7 items
    "articles_num": [ ... ], // 10 items
    "errors": [ ... ] // 36 items
  },
  "insertCommandes": {
    "time": 3494,
    "commandes": [ ... ], // 880 items
    "errors": [
      581,
      582
    ]
  },
  "insertFichiers": {
    "time": 8011,
    "fichiers": [ ... ], // 1496 items
    "errors": [ ... ] // 340 items
  }
}
```

L'insertion des données ne peut se faire de manière linéaire ligne par ligne car cela impliquerait un nombre de requêtes beaucoup trop important dont la majeure partie peut être évitée. En effet, il faut concevoir une stratégie d'insertion des données limitant le nombre de requêtes à la base de données et pour cela, il faut filtrer les données d'entrée du tableau Json. Chaque doublon doit être supprimé et inséré dans les tables dans un ordre précis.

Effectivement, l'ordre d'insertion des données est important car la plupart d'entre-elles en référence une autre. Ainsi les séries doivent être insérées avant les sous-séries car une sous-série appartient à une série. Cela implique d'itérer plusieurs fois à travers le tableau et montre donc l'intérêt de la conversion des données au format Json car lire des données dans ce type de tableau est beaucoup plus rapide que lire des données dans un tableau Excel.

4.5. Gestion des données

4.5.1. Connexion

L'objectif de ce projet est de pouvoir centraliser les données de numérisation du service des Archives mais également de mieux contrôler leur accès. En effet l'ancien process ne permettait pas de décider quel agent pouvait ou non accéder à tel ou tel donnée car

l'ensemble du fichier Excel était partagé. Ainsi l'intérêt de ce site est de pouvoir mettre en place un système de connexion et d'accès restreints afin de limiter de manière nuancée l'accès aux données de chaque utilisateur.

De plus, afin de garantir la sécurité du site, il est nécessaire de vérifier les connexions grâce à des clés (json web token) et de respecter les règles de protection des données (RGPD¹⁷) comme le cryptage des informations sensibles tels que les mots de passe (bcrypt).

4.5.2. Consultation

L'élément central du projet concernant la gestion des données est la consultation des informations de numérisation afin de pouvoir visualiser et modifier une donnée précise. Ainsi l'élaboration d'un tableau de bord complet permettra aux agents de la DAD de contrôler l'ensemble des actions depuis une seule et même interface.

Le tableau de bord est divisé en 3 parties:

1. Les filtres de recherche et d'affichage

Plusieurs filtres permettent à l'utilisateur de pouvoir filtrer les données d'entrée par rapport à des critères de recherche tels que la série ou le service du fichier recherché, mais également de personnaliser le tableau de bord afin d'accéder plus rapidement et efficacement à la donnée voulue. En effet, il est possible d'afficher uniquement certaines colonnes (certains champs), de consulter les données supprimées (archivage) ou d'activer la mise à jour automatique du tableau lors du changement de filtres.

2. Le tableau de données

Le tableau permet d'afficher toutes les données de manière organisée et linéaire afin d'avoir un visuel clair et de ne pas perdre de temps pour obtenir l'information voulue. Il permet également de trier les colonnes (champs) de manière alphabétique, numérique ou chronologique et possède un système de pagination permettant de ne pas importer directement toutes les données de la base de données mais seulement par groupe.

¹⁷ **RGPD** ([Règlement Général sur la Protection des Données](#)): règlement qui encadre le traitement des données personnelles sur le territoire de l'Union européenne.

3. Les actions

Chaque ligne du tableau possède une colonne d'actions permettant d'obtenir des informations de gestion (date et utilisateur d'ajout, de modification ou de suppression), de supprimer (archiver) ou de modifier les données de numérisation d'un fichier. Il est également possible de gérer les données par groupe pour éviter une action répétée sur plusieurs lignes.

Figure 6: tableau de bord du site

TYPE SERVICE	SERVICE	DATE RECEPTION	DATE REALISATION	SERIE	SOUS-SERIE	ARTICLE	SUIVE-ARTICLE
Ad090	numérisation patrimoniale	29/09/2029	29/09/2029	Fi	20	10	
Ad090	numérisation patrimoniale	29/09/2029	29/09/2029	J	1	1297	
Usager	Elodie Lambert	29/09/2029	29/09/2029	Q	3	1	250bis (page 99)
Usager	Elodie Lambert	29/09/2029	29/09/2029	Q	3	3	117 (n°125)
Usager	SBE - Michel Rilliot	29/09/2029	29/09/2029	J	5	fi	Vézelois, d-XII93, 19970320
Usager	Marie-Antoinette Vacelet	29/09/2029	03/10/2022	W	99	172	
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Giromagny, d-XI70, 19930630
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Giromagny, d-XI72, 19930630
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Offemont, d-III97, 19931204
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Roppe, d-IV84, 19840506
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Roppe, d-IV87, 19840506
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Roppe, d-IV94, 19840506
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Roppe, d-XII75, 19970313
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Roppe, d-XIII2, 19970313
Usager	SBE - Michel Rilliot	28/09/2029	29/09/2029	J	5	fi	Roppe, d-XIII6, 19970313
Usager	LISA - Jean-Dominique Pellegrini	17/04/2023	17/03/2023	M	3	12	
Mairie	Châtenois-les-Forges - Christophe L. Antoinette	04/04/2023	11/04/2023	E-Dépôt	22	1 O	12

4.5.3. Ajout et Modification

Ce nouveau tableau de bord permet de mieux visualiser les données et ainsi de déceler de potentielles erreurs dans le renseignement des informations de numérisation de certaines archives. Ainsi un formulaire de modification donne la possibilité aux utilisateurs de modifier les informations. Ce formulaire se base sur le même que celui pour ajouter de nouvelles données mais place directement les données de fichier sélectionné dans les bons champs.

Avant tout ajout ou modification, chaque champ est contrôlé et affiche un message d'erreur s'il est obligatoire mais n'a pas été renseigné. De plus, les listes déroulantes se débloquent automatiquement lorsque l'utilisateur choisit un champ parent, par exemple il faut choisir une série pour avoir accès à la liste des sous-séries de cette dernière.

Figure 7: page de modification et création de données de numérisation

The screenshot displays the 'Archives du Territoire de Belfort' web application interface. At the top, a blue header bar contains the logo on the left, the title 'Archives du Territoire de Belfort' in the center, and the date 'Jeudi 1 juin 2023 12 09 46' on the right. Below the header is a large form for data entry. The form is organized into several sections: the top section contains date fields for 'Date réception' (01/06/2023), 'Date réalisation' (01/06/2023), 'Date mise en ligne' (jj/mm/aaaa), and 'Date publication' (jj/mm/aaaa); the middle section contains dropdown menus for '*Type Service', '*Service', '*Agent transmetteur', '*Agent numérisation', and '*Type numérisation'; the bottom section contains dropdown menus for '*Numérisation', 'Série NUM', 'Sous-série NUM', 'Article NUM', '*Série', 'Sous-série', 'Article', and 'Suite articles'. Below these are input fields for 'Nombre de fichier' (1), 'Nombre de page' (1), 'Nb pub* INTERNET', 'Nb pub* INTRANET', 'Volumétrie (Mo)', and 'Format'. At the bottom of the form are fields for 'Dénomination du fichier', 'Arborescence du fichier', 'Position fichier', and 'Retrait fichier', followed by a 'Commentaire' field. A legend indicates that fields marked with an asterisk are mandatory. At the bottom of the form are two buttons: 'EFFACER' and 'ENREGISTRER'. A 'Retour' button is located at the bottom center of the page.

5. Bilan du stage

5.2. Bilan pour le département

Ce stage de 12 semaines a permis de créer une nouvelle manière de renseigner les informations de numérisation des archives et de transformer un besoin utilisateur en une véritable solution concrète. Ainsi les agents du service des Archives peuvent maintenant mieux appréhender le projet et son envergure et même développer de nouveaux besoins jusqu'alors impossibles avec leur ancien process.

Cette première version du projet a permis de mettre en place une base solide regroupant les fonctionnalités impératives et attendues par les agents et a su satisfaire les utilisateurs concernés. Elle signe alors le début d'une phase d'amélioration et d'enrichissement afin de satisfaire à terme l'ensemble des attentes des archivistes notamment concernant l'importation en masse de données et la visualisation de données statistiques.

5.3. Bilan humain

Durant ce stage, la communication fût l'élément clé afin de mener à bien le projet. En effet, le service informatique du Département est un support qui doit comprendre le

fonctionnement des autres pour mieux les aider. Ainsi il a fallu découvrir le service des Archives et comprendre pourquoi ce projet a vu le jour et quels en étaient les intérêts. Chaque agent a sa propre manière de concevoir le projet, il faut alors discuter avec chacun.

Ce schéma est également valable dans le service informatique lui-même car le projet doit être construit avec une cohésion entre chaque membre de l'équipe. Les difficultés des uns sont les facilités des autres et inversement. Il est alors important de prendre le temps de s'aider mutuellement. Certains arbitrages doivent être pris ensemble et la voix de chacun est importante.

5.4. Bilan pédagogique

La réalisation de ce stage a permis de mettre en œuvre les connaissances et les compétences acquises lors de ces 2 années à l'IUT telle que l'utilisation de NodeJS pour la partie Back du projet et VueJS pour le Front. De plus, une application pratique de la méthode agile de gestion de projet permet de rendre compte de son importance surtout lorsque les besoins client ne sont pas figés mais évoluent avec l'avancement du projet.

Certains points peuvent malgré tout être améliorés notamment sur la partie importation des données qui pourrait être gérée automatiquement par un ETL¹⁸ tel que Pentaho¹⁹ et qui pourrait diminuer le temps d'introduction des données dans la base de données.

¹⁸ **ETL** (extraction, transformation et chargement): processus qui consiste à combiner les données provenant de plusieurs sources dans un grand référentiel central appelé entrepôt des données.

¹⁹ cf. [Pentaho](#)

Conclusion

Ce stage s'inscrit dans une démarche de changement de process de la part des agents du service des Archives départementales concernant le stockage et la gestion des données de numérisation de leurs archives numérisées.

L'objectif était de mettre en place un nouveau process afin de mieux gérer toutes ces données et le projet réalisé lors de ce stage correspond aux attentes des utilisateurs permettant ainsi de consulter et modifier toutes les données renseignées dans les anciens tableaux Excel et ajouter de nouvelles données de manière unitaire.

Ce projet constitue une base sur laquelle pourra s'appuyer chaque agent des Archives afin d'être amélioré dans le futur par d'autres développeurs en ajoutant de nouvelles fonctionnalités telles que l'importation en masse de données ou la génération de données statistiques.

Ainsi ce stage clôture cette deuxième année de formation BUT à l'IUT de Belfort-Montbéliard avec une application pratique des connaissances et une montée en compétences d'un point de vue technique et professionnel.

Table des figures

Figure 1: 15 km linéaire de documents d'archives du Territoire de Belfort.....	11
Figure 2: modélisation des processus.....	13
Figure 3: échantillon d'un tableau de données de numérisations Excel.....	19
Figure 4: modèle de conception des données.....	20
Figure 5: log de l'insertion des données.....	21
Figure 6: tableau de bord du site.....	23
Figure 7: page de modification et création de données de numérisation.....	24

Sitographie

Département

Archives départementales du Territoire de Belfort:

<https://archives.territoiredebelfort.fr/>

Département du territoire de Belfort:

<https://www.territoiredebelfort.fr/>

Youtube: chaîne du Département du Territoire de Belfort:

<https://www.youtube.com/@territoiredebelfort>

Youtube: vidéos des Archives du Territoire de Belfort:

https://www.youtube.com/results?search_query=Archives+du+Territoire+de+Belfort

Technologies

API Rest: “Qu’est ce qu’une API Rest ?”, IBM: <https://www.ibm.com/fr-fr/topics/rest-apis>

Bcrypt: <https://www.npmjs.com/package/bcrypt>

Docker: <https://www.docker.com/>

Git: <https://git-scm.com/>

Gitlab: <https://about.gitlab.com/fr-fr/>

Javascript: <https://www.javascript.com/>

Json: <https://www.json.org/json-fr.html>

Json Web Token: <https://jwt.io/>

MySQL: <https://www.mysql.com/fr/>

MySQL2: <https://www.npmjs.com/package/mysql2>

NodeJs: <https://nodejs.org/fr>

Pentaho: <https://www.hitachivantara.com/en-us>

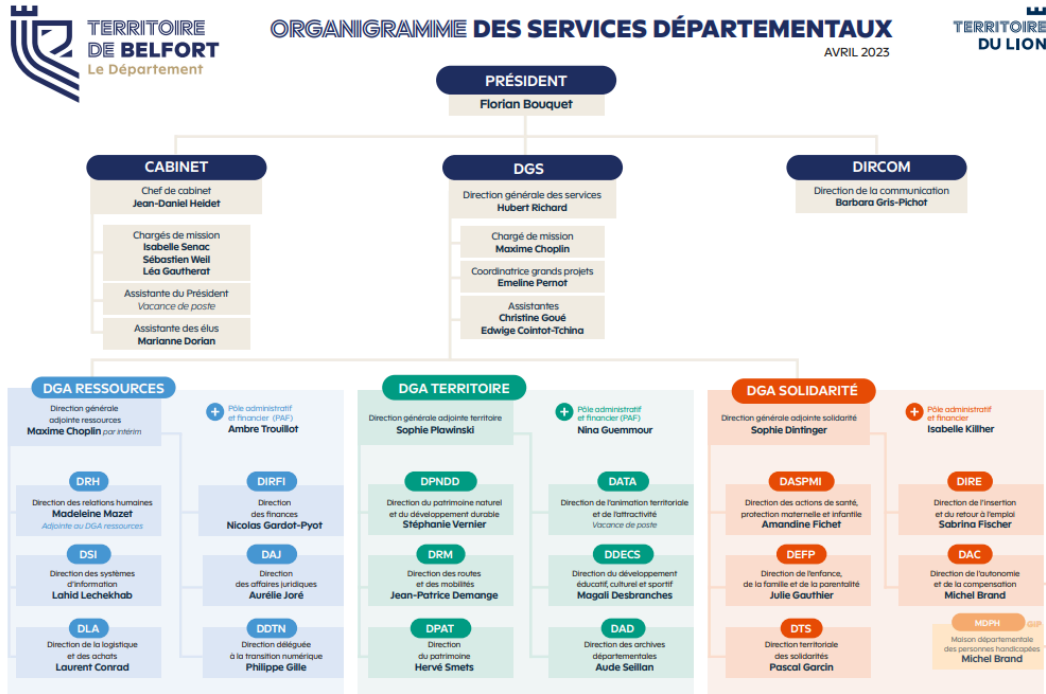
VueJs: <https://fr.vuejs.org/>

Table des annexes

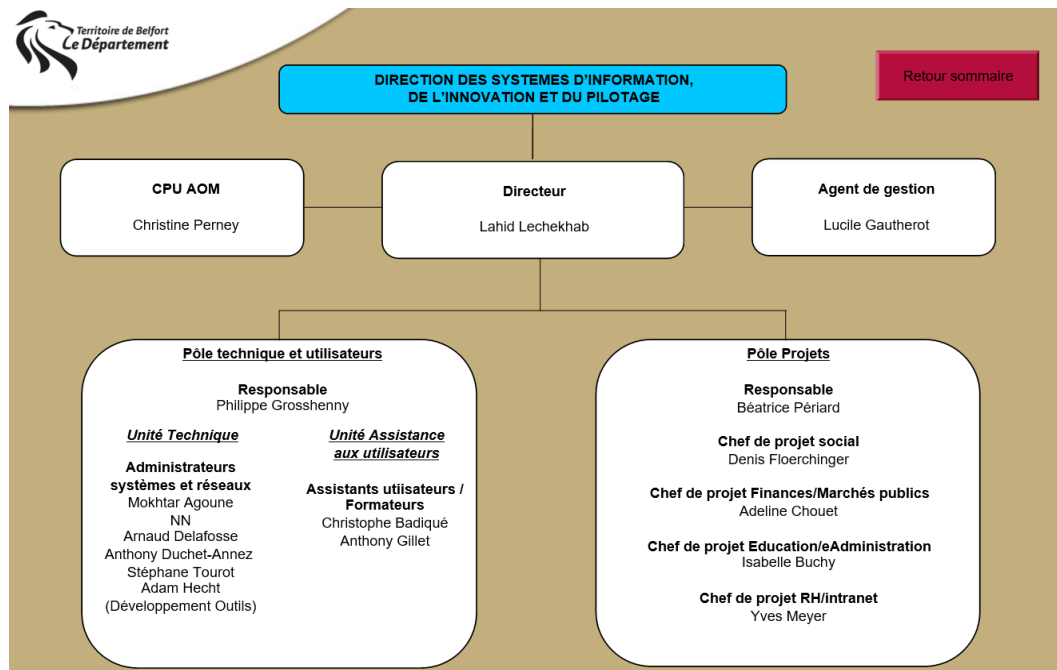
Annexe I : Organigramme des services départementaux.....	30
Annexe II : Organigramme du Service des Systèmes d'Information (DSI).....	30
Annexe III : Organigramme du Service des Archives Départementales (DAD).....	31
Annexe IV : Tableau de bord Gouti du projet.....	31

Annexes

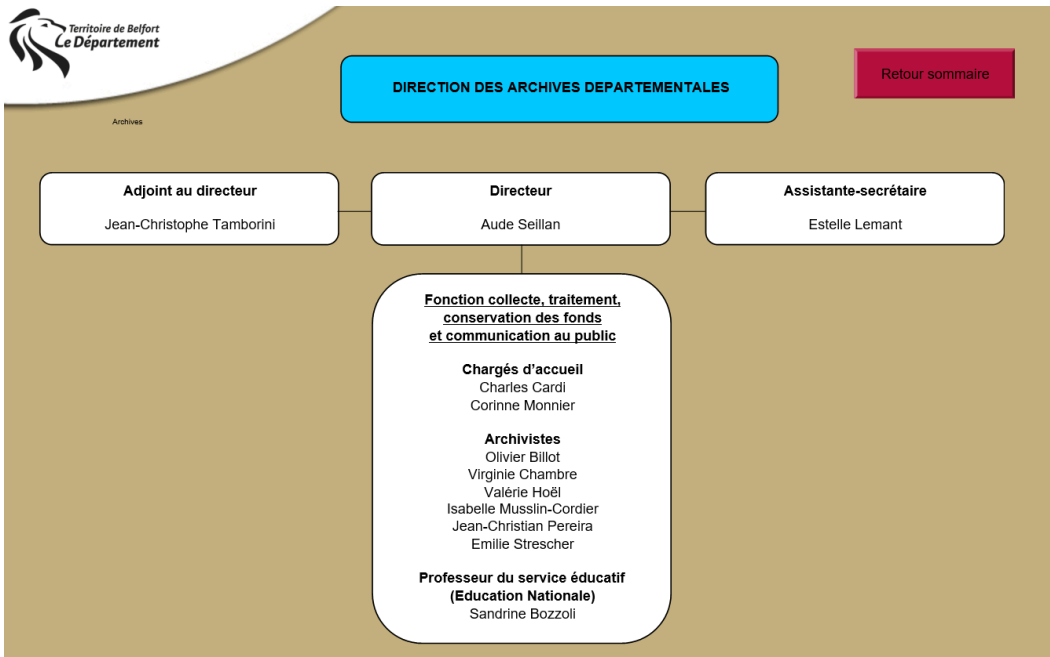
Annexe I : Organigramme des services départementaux



Annexe II : Organigramme du Service des Systèmes d'Information (DSI)



Annexe III: [Organigramme du Service des Archives Départementales \(DAD\)](#)



Annexe IV: [Tableau de bord Gouti du projet](#)

Tableau de bord Gouti du projet

Liste des tâches

Appuyez sur **F11** pour quitter le mode plein écran.

Remarque: Les dates de début et de fin sont indiquées en format JJ/MM/AAAA.

Id	Nom	Responsable	Jalon	Date de début	Date de fin	%	Statut	Pr	i
1	CADRAGE			13/03/2023	30/06/2023	100 %	Terminé		
1.1	Lettre de cadrage			13/03/2023	30/06/2023	100 %	Terminé		
1.2	Description de la demande			13/03/2023	30/06/2023	100 %	Terminé		
1.3	DOC de cadrage			14/04/2023	30/06/2023	100 %	Terminé		
2	ANALYSE DE LA DEMANDE			13/03/2023	30/06/2023	56 %	En cours		
2.1	Réunion de lancement			13/03/2023	30/06/2023	100 %	Terminé		
2.2	modélisation process actuel			13/03/2023	30/06/2023	100 %	Terminé		
2.3	définition des spécifications fonctionnelles			13/03/2023	30/06/2023	100 %	Terminé		
2.4	proposition nouveaux process			13/03/2023	30/06/2023	50 %	En cours		
2.5	Catégorisation des spécifications			20/04/2023	20/04/2023	0 %	À faire		
2.6	DOC d'analyse			10/05/2023	30/06/2023	0 %	À faire		
3	DEVELOPPEMENT DE LA SOLUTION			17/03/2023	09/06/2023	64 %	En cours		
3.1	sprint 1			17/03/2023	10/05/2023	100 %	Terminé		
3.1.1	1er maquette	VBR		17/03/2023	03/05/2023	100 %	Terminé		
3.1.2	Création Modèle Conceptuel de données	MTH		17/04/2023	21/04/2023	100 %	Terminé		
3.1.3	Import services et agents (Excel)	MTH		24/04/2023	10/05/2023	100 %	Terminé		
3.1.4	Import sites et sites NUM (Excel)	MTH		25/04/2023	10/05/2023	100 %	Terminé		
3.2	sprint 2			26/04/2023	11/05/2023	87 %	En cours		
3.2.1	Import commandes (Excel)	MTH		26/04/2023	10/05/2023	100 %	Terminé		
3.2.2	Import fichiers (Excel)	MTH		27/04/2023	10/05/2023	96 %	En cours		
3.2.3	Formulaire ajout archive	MTH		03/05/2023	05/05/2023	100 %	Terminé		
3.2.4	Tableau de consultation	VBR		04/05/2023	11/05/2023	60 %	En cours		
3.3	sprint 3			02/05/2023	19/05/2023	0 %	À faire		
3.4	Finalisation intégration solution			22/05/2023	09/06/2023	0 %	À faire		
4	PILOTAGE			13/03/2023	30/06/2023	20 %	En cours		
4.1	Réunion suivi de projet			13/03/2023	30/06/2023	5 %	En cours		
4.2	Pilotage projet			13/03/2023	30/06/2023	0 %	À faire		