



**HAL**  
open science

# Implication du Data Manager dans la Vérification des Systèmes Informatisés (VSI) à l'institut Bergonié

Raphaël Bazenet

► **To cite this version:**

Raphaël Bazenet. Implication du Data Manager dans la Vérification des Systèmes Informatisés (VSI) à l'institut Bergonié. Santé publique et épidémiologie. 2020. dumas-03156925

**HAL Id: dumas-03156925**

**<https://dumas.ccsd.cnrs.fr/dumas-03156925>**

Submitted on 2 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

**Master Sciences, Technologies, Santé**

**Mention Santé Publique**

Parcours

**Systèmes d'Information et Technologies Informatiques pour  
la Santé**

**Promotion 2019-2020**

---

***Implication du Data Manager dans la  
Vérification des Systèmes Informatisés  
(VSI) à l'institut Bergonié***

---

Rapport réalisé dans le cadre d'une mission effectuée du 02/12/2019 au  
30/08/2020

Institut Bergonié

229, Cours de l'Argonne, 33000 Bordeaux

Maitre de Stage : Sandrine Barbotin, Data Manager

Soutenu publiquement le 08/09/2020

Par Raphaël Bazenot

## Table des matières

Glossaire .....	5
Remerciements .....	6
1. Introduction.....	7
1.1. Présentation de la structure d'accueil .....	7
1.1.1. Institut Bergonié : Centre de Lutte Contre le Cancer (CLCC).....	7
1.1.2. Direction de la Recherche Clinique et de l'Innovation (DRCI).....	7
1.1.3. Unité de Recherche et d'Epidémiologie et Clinique (UREC).....	9
1.2. Présentation du sujet d'alternance .....	9
2. Contexte .....	11
2.1. Qu'est-ce qu'un essai clinique ?.....	11
2.1.1. Définition d'un essai clinique .....	11
2.1.2. Quelles sont les réglementations en vigueur dans ce domaine ?.....	12
2.2. Data Management d'essai clinique .....	13
2.2.1. Qu'est-ce que le Data Management en général ?.....	13
2.2.2. Quelles sont les missions du data manager d'essai clinique ?.....	15
2.3. Validation des Systèmes Informatisés (VSI) .....	17
2.3.1. Qu'est-ce que la VSI ?.....	17
2.3.2. Pourquoi et comment valider les systèmes informatisés en santé ?.....	19
2.4. Contributions du Data Manager dans la VSI .....	22
2.4.1. Elaboration des procédures standards.....	22
2.4.2. Validation du logiciel MACRO.....	23
2.4.3. Discussion autour de la mise en place de la VSI.....	26
3. Missions de Data Management à l'institut Bergonié .....	27
3.1. Missions du Data Manager.....	27
3.1.1. Rédaction des documents de Data Management .....	27
3.1.2. Bases de données et des contrôles de cohérence .....	29
3.1.3. Validation des données et autres missions .....	31
3.2. Apport personnel de ces différentes missions et discussion autour du Data Management .....	33
3.2.1. Apport personnel des différentes missions.....	33
3.2.2. Comparaison entre les logiciels de Data Management et les SGBD classiques .....	34

4. Discussion, limites et conclusion .....	36
4.1. Discussion et limites .....	36
4.1.1. Discussion générale .....	36
4.1.2. Impact de l'épidémie de Covid-19 sur mon alternance .....	36
4.2. Conclusion personnelle .....	37
Bibliographie.....	38
Table des annexes .....	42

## Table des figures

Figure 1 - Pilotage et ressources de la recherche clinique à l'institut Bergonié [8] .....	8
Figure 2 : Etapes avant le début d'un essai clinique [17] .....	12
Figure 3 : Gestion et traitement des données à l'institut Bergonié, missions du Data Manager [24]..	16
Figure 4 : Exemple d'étapes de validation d'un système informatisé [35] .....	20
Figure 5 : Démarche générale de la validation présente dans le plan de validation .....	24
Figure 6 : Analyse de risque du logiciel MACRO effectué à l'institut Bergonié.....	24
Figure 7 : Cahier de recette pour le logiciel MACRO .....	25
Figure 8 : Exemple de première page du DVP de BIOPSAR.....	28
Figure 9 : Exemple de contrôle automatique sous MACRO .....	30
Figure 10 : Exemple de contrôle automatique sous RedCap .....	30

## Glossaire

ANSM : Agence Nationale de Sécurité du Médicament

ARC : Attaché de Recherche Clinique

BPC : Bonnes Pratiques Cliniques

CLCC : Centre de Lutte Contre le Cancer

CNIL : Commission Nationale de l'Informatique et des Libertés

COFIL : COmité de PIlotage

CORO : COmité de Recherche Opérationnelle

CPP : Comité de Protection des Personnes

CR : Collège de Recherche

CRF/e-CRF : Case Report Form /electronic Case Report Form

DMP : Data Management Plan

DRCI : Direction de la Recherche Clinique et de l'Innovation

DVP : Data Validation Plan

ESPIC : Etablissement de Santé Privé d'Intérêt Collectif

GIRCI SOHO : Groupement Interregional de Recherche Clinique et d'Innovation Sud-Ouest Outre-mer Hospitalier

LIL : Loi Inforamtique et Liberté

MS : Modification Subsantielle

POS : Procédure Opératoire Standard

QI /QP /QO : Qualification d'Installation/ de Performance/ Opérationnelle

RIPH : Recherche Impliquant la Personne Humaine

RGPD : Règlement Général sur la Protection des Données

SITIS : Systèmes d'Information et Technologies Informatiques pour la Santé

UGRC : Unité de Gestion de la Recherche Clinique

UREC : Unité de Recherche et d'Epidémiologie Clinique

VSI : Validation des Systèmes Informatisés

## Remerciements

Je tiens à remercier l'ensemble des personnes qui ont contribué au succès de mon alternance et à la rédaction de ce rapport.

Dans un premier temps je remercie particulièrement mon maître de stage, Sandrine Barbotin, qui m'a beaucoup appris tout au long de mon alternance, qui a toujours été là pour répondre à mes questions et pour m'accompagner dans l'apprentissage du métier de Data Manager.

Ensuite je remercie, ma tutrice d'alternance à l'ISPED et responsable du Master 2 SITIS, Fleur Mougin, pour l'aide qu'elle m'a apporté tout au long de l'année, que ce soit pour trouver cette alternance ou bien pour ces précieux conseils lors de la rédaction de mon rapport. Je remercie également l'ensemble de l'équipe pédagogique de l'ISPED qui m'a accompagné lors de cette année universitaire.

Je remercie la Professeure Simone Mathoulin-Pelissier d'avoir accepté de me prendre en alternance au sein de l'institut Bergonié et pour les relectures de mon rapport tout au long de mon travail.

Je remercie également l'ensemble de l'équipe de l'institut Bergonié qui m'a accompagné tout au long de ma formation et dans mes tâches au quotidien. Je remercie l'ensemble de l'équipe Data, Antoine, Alexis et Morgane, mais également l'ensemble de l'UREC pour son soutien et sa gentillesse tout au long de cette année.

Enfin, je tiens à remercier toutes les personnes qui m'ont aidé en dehors de l'entreprise et qui m'ont apporté du soutien lorsque j'en avais besoin, que ce soit pour la relecture de mon rapport ou d'un point de vue personnel : ma famille, mes amis (plus particulièrement les nains) et Pierre.

# 1. Introduction

## 1.1. Présentation de la structure d'accueil

### 1.1.1. Institut Bergonié : Centre de Lutte Contre le Cancer (CLCC)

L'institut Bergonié est le Centre Régional de Lutte Contre le Cancer (CRLCC) de la Nouvelle-Aquitaine. Il constitue avec le CHU de Bordeaux le pôle régional de cancérologie. L'institut est membre du groupe Unicancer, qui est l'unique réseau hospitalier français dédié à 100% à la lutte contre le cancer, avec 18 centres répartis sur 20 sites hospitaliers. [1] Tous ces centres sont des Etablissements de Santé Privé d'Intérêt Collectif (ESPIC) dont les missions sont les soins et la prise en charge globale des patients, mais également l'enseignement et la recherche. [2]

L'institut Bergonié fut le premier centre de cancérologie créé en province dans les années 1920. Son nom lui vient de l'un des fondateurs des CLCC en France, le Professeur Bergonié ayant joué un rôle précurseur dans de nombreux domaines de l'oncologie. Le fonctionnement de l'institut, tout comme celui des autres centres, repose sur le modèle Curie. [3] Ce modèle se définit par la continuité entre la recherche fondamentale et les soins traduisant la phrase « De la recherche aux soins innovants ». Les fondamentaux de ce modèle se basent sur le rapprochement des chercheurs, des soignants et des patients afin de permettre l'accélération de la mise à disposition de diagnostics et de traitements nouveaux. [3, 4] Ainsi, basé sur ce modèle, l'institut fournit, en étroite collaboration avec les autres établissements de santé de Bordeaux, des stratégies diagnostiques et thérapeutiques innovantes dans le but de répondre à tous les besoins des patients et de la cancérologie moderne. [5] Depuis 2016, le directeur de l'institut est le Professeur François-Xavier Mahon qui continue de suivre ce modèle avec un projet phare. Celui-ci consiste en la construction d'un nouveau pôle chirurgical, afin de permettre une réorganisation des activités d'hospitalisation, toujours avec pour objectif d'améliorer la lutte contre le cancer. [4]

### 1.1.2. Direction de la Recherche Clinique et de l'Innovation (DRCI)

Au sein de l'institut différentes structures sont dédiées à la recherche clinique, le pôle promotion où l'institut Bergonié est le promoteur des études cliniques et le pôle investigation dont le promoteur peut être un laboratoire ou un autre centre académique, et où les études sont dirigées et surveiller lors de leur réalisation à l'institut. [6, 7] Ces structures sont en lien avec les départements et groupes impliqués dans la recherche via notamment des comités et structures de pilotage. Ces unités ont pour mission d'assurer le respect des recommandations internationales de Bonnes Pratiques Cliniques (BPC). Cela grâce à des procédures opératoires standards (POS) qui sont régulièrement renouvelées et mises à jour avec l'aide du Département Qualité et Gestion des risques. [6] La structure principale dédiée à la recherche clinique est la Direction de la Recherche Clinique et de l'Innovation (DRCI) qui se compose de deux structures d'appui logistique et méthodologique : l'Unité de Gestion de la Recherche Clinique (UGRC) et l'Unité de Recherche et d'Epidémiologie Clinique (UREC). Ces deux unités ont un rôle aussi bien dans le pôle promotion que le pôle investigation (figure 1). L'UREC a un rôle à jouer dans la mise en place des études en apportant l'expertise des différents experts métiers, tandis que l'UGRC va plutôt jouer sur les aspects réglementaires, juridiques et financiers. [6] Un organigramme détaillé de la recherche clinique à



l'institut Bergonié est disponible en annexe 1. Pour l'organisation opérationnelle de la recherche clinique, plusieurs comités sont impliqués dans les différentes étapes de la vie d'un essai clinique, comme le Collège de Recherche (CR) ou le Comité de la Recherche Opérationnelle (CORO). Le schéma suivant permet de décrire cette organisation opérationnelle. [8]

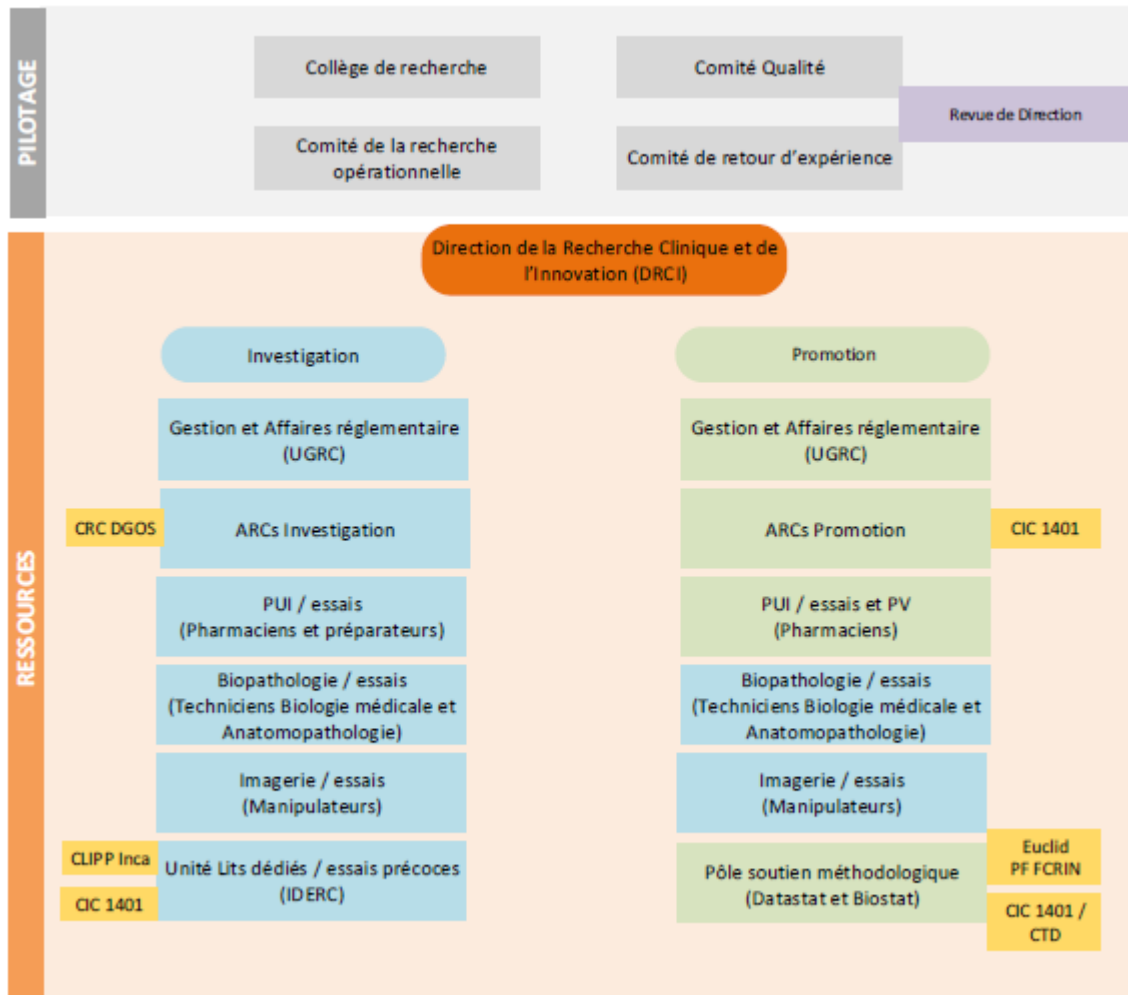


Figure 1 - Pilotage et ressources de la recherche clinique à l'institut Bergonié [8]

Par ailleurs toujours dans une démarche de qualité, l'institut Bergonié est certifié ISO 9001-v2015 dans le département de la recherche clinique. La norme ISO 9001 est un label international et également un gage de haut niveau de qualité qui témoigne de la structuration efficace de la recherche clinique. [9] L'institut Bergonié est également membre du Groupement Interrégional de Recherche Clinique et d'Innovation Sud-Ouest Outre-Mer Hospitalier (GIRCI SOHO) qui a pour missions de créer et d'apporter un soutien mutuel entre établissements de recherche clinique, avec notamment le partage de documents et l'aide à la formation des professionnels de la recherche clinique. [10]

### 1.1.3. Unité de Recherche et d'Epidémiologie Clinique (UREC)

L'UREC est ainsi une structure de la DRCl divisé en deux pôles, le pôle promotion et le pôle investigation. Le pôle promotion dispose de personnels dédiés pour les activités d'Attaché de Recherche Clinique (ARC) promotion, de data management, de statistiques et de pharmacie promotion. L'intégration de compétences pluridisciplinaires en biostatistique, en informatique ou en monitoring permet un accompagnement structuré des investigateurs. Ainsi le projet de recherche bénéficie d'un apport méthodologique et logistique dès le montage du projet et ce jusqu'à la clôture et la valorisation de l'étude. [6] Pour le pôle investigation, l'UREC comprend des personnels à la fois dédiés et détachés fonctionnellement pour les activités relatives aux essais cliniques. L'unité est liée à d'autres départements comme le secteur essais cliniques du département de pharmacie, la biopathologie, l'imagerie ou encore le département qualité et gestion des risques. Les missions de l'UREC sont donc la mise en place et le suivi de l'étude en partenariat avec l'investigateur principal et les unités associés. [6] On peut détailler ces missions en plusieurs points. Tout d'abord, une aide à l'élaboration et à la formalisation des protocoles et projets de recherche. Ensuite, l'apport d'un soutien méthodologique pour le recueil et l'analyse des données ainsi que pour les traitements informatique et biostatistique. Puis, l'apport de conseils pour la valorisation de la recherche finalisée avec une aide à la publication. Enfin, la participation à la formation en recherche clinique. [11] Aujourd'hui, c'est la Professeure Simone Mathoulin-Pelissier qui est directeur médical de la DRCl et responsable de l'UREC, et qui permet le bon fonctionnement de l'unité en lien avec les autres départements de l'institut. [11]

## 1.2. Présentation du sujet d'alternance

Dans le cadre de mon Master 2 Système d'Informations et Technologies Informatiques pour la Santé (SITIS), j'ai intégré le pôle promotion de l'UREC en tant que Data Manager en alternance, depuis le 2 décembre 2019. Au cours de mon alternance plusieurs missions m'ont été confiées, dont deux principales. La première était d'assurer les fonctions classiques d'un Data Manager auxquelles j'ai été formé tout au long de mon alternance, de la validation du CRF papier à la construction de bases de données sur différents logiciels, en passant par le paramétrage des contrôles de cohérences et la validation des données cliniques. J'ai ainsi eu l'occasion de travailler sur plusieurs études au cours de cette année d'alternance comme l'étude BIOPSAR sur les performances diagnostiques de la biopsie utérine écho-guidée ou encore l'étude IBMC sur la réduction du risque tabac/alcool chez les femmes traitées pour un cancer du sein avec l'implication du conjoint. Ma seconde mission était de participer à la réflexion autour de la mise en place de la Validation du Système Informatisé (VSI) de l'institut Bergonié, plus particulièrement la validation des logiciels de Data Management. Avec le Règlement Général sur la Protection des Données (RGPD) et les différentes normes en vigueur, il est aujourd'hui important de valider les différents logiciels utilisés dans le cadre de la recherche clinique. L'institut Bergonié, avec l'aide du GIRCI SOHO, doit donc mettre en place des procédures de validation afin de prouver la conformité des logiciels informatiques utilisés dans le cadre des essais cliniques. Ma mission était donc, en tant que Data Manager, d'aider à la rédaction des différentes procédures standards de la VSI, puis par la suite d'aider à la mise en place de la validation de certains logiciels utilisés en data management tel que MACRO.

La réalisation de ces deux missions s'est faite en parallèle tout au long de mon alternance, et m'a permis d'apporter les différentes compétences, que j'ai pu acquérir lors de mon Master 2, dans la mise en œuvre de ces tâches. Dans ce rapport je vais dans un premier temps vous expliciter le contexte de cette alternance avec les différentes définitions qui encadre ce travail. Les définitions d'un essai clinique, du Data Management et de la VSI. Ensuite, je vais vous décrire dans la même partie ce que j'ai pu apporter au domaine qu'est la VSI grâce aux différentes tâches que j'ai effectué lors de mon alternance. Ensuite, je vous décrirais plus en détails les missions que j'ai pu réaliser dans le cadre du métier de Data Manager à l'institut Bergonié. Et enfin, je terminerai par une conclusion générale sur mon travail et sur cette année d'alternance.

## 2. Contexte

### 2.1. Qu'est-ce qu'un essai clinique ?

#### 2.1.1. Définition d'un essai clinique

Les essais cliniques font partie des Recherches Impliquant la Personne Humaine (RIPH) et constituent des recherches biomédicales en vue du développement des connaissances biologiques ou médicales. Les essais portant sur les médicaments ont pour but, la plupart du temps, d'établir ou de vérifier certaines données d'un nouveau médicament ou d'une nouvelle façon d'administrer un traitement déjà connu comme ses modalités d'absorption, ses mécanisme d'action ou bien encore son efficacité et sa tolérance. [12] Pour simplifier cette définition on peut dire qu'un essai clinique a pour but d'évaluer de nouveaux traitements ou de nouvelles modalités de soins du cancer ou d'autres maladies. [13] La recherche a donc pour vocation d'améliorer la prise en charge et le traitement des patients atteints de ces maladies. Un essai clinique est donc l'ultime étape avant le lancement ou non d'un médicament ou d'une nouvelle technique thérapeutique. [14] Il existe différentes phases pour un essai clinique. Les essais de phase I ont pour objectif d'évaluer la tolérance de l'organisme et la toxicité d'un nouveau traitement, cela afin de déterminer la dose recommandée pour l'administration de ce nouveau médicament. Ces essais ont généralement peu de patients (entre 10 et 40). Les essais de phase II précisent l'activité clinique ou pharmacologique d'un traitement. Ils nécessitent plus de patients (entre 40 et 80). Enfin les essais de phase III sont quant à eux des essais comparatifs. Ils ont pour but de comparer le nouveau traitement avec le traitement habituellement utilisé afin d'en évaluer son efficacité. Ces essais vont nécessiter l'inclusion d'un grand nombre de patient pour établir une différence significative (plusieurs centaines voire milliers de patients). [13, 15] Lorsque le traitement est commercialisé il entre dans une phase IV. Celle-ci permet d'approfondir la connaissance du médicament dans les conditions réelles d'utilisation et sur une durée plus longue. [14, 15] Les essais cliniques portant sur le médicament ou sur de nouvelles pratiques thérapeutiques sont classés RIPH-1, c'est-à-dire qu'il y a un risque élevé et une contrainte pour le patient. On peut toutefois noter qu'il existe également des essais classés RIPH-2 et RIPH-3 portant respectivement sur les essais interventionnels à risque et contraintes minimales et les essais non interventionnels. [13, 16]

Pour débiter un essai clinique il est nécessaire d'avoir au préalable obtenu un avis favorable (voir figure 2) du Comité de Protection des Personnes (CPP) et une autorisation de l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM). Le CPP est indépendant, il est chargé de vérifier que l'essai clinique proposé est acceptable sur un plan scientifique et éthique. [13] L'ANSM de son côté assure la sécurité et la qualité des produits afin de garantir la sécurité des personnes impliquées. [13]

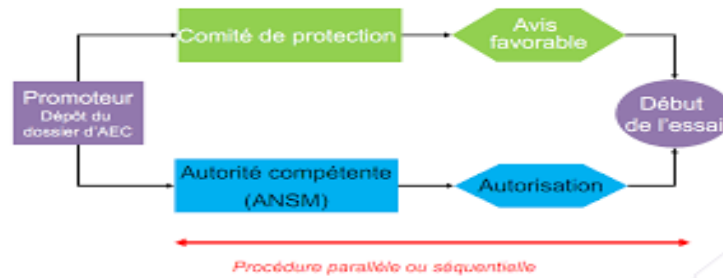


Figure 2 : Etapes avant le début d'un essai clinique [17]

L'initiative de l'essai est prise par le promoteur. Celui-ci est la personne physique ou morale qui propose de mettre en œuvre l'essai et qui assure son financement. Il peut s'agir d'un laboratoire pharmaceutique, d'un établissement de soins, d'une association ou bien d'un médecin. [12, 13] Une fois les différentes autorisations obtenues l'essai clinique peut alors être mis en place. La réalisation sera menée par un investigateur dont le rôle est de diriger et surveiller la réalisation de l'essai. Il peut y avoir plusieurs centres investigateurs lorsque l'essai est dit multicentrique (réalisé dans plusieurs centres de recherche) et dans ce cas un investigateur coordonnateur, ou principal, est désigné pour assurer le bon déroulement général de l'essai. [12] Néanmoins avant toute chose pour débiter un essai clinique il faut établir un protocole de recherche clair et précis respectant les normes et réglementations en vigueur.

### 2.1.2. Quelles sont les réglementations en vigueur dans ce domaine ?

Suite aux différentes expérimentations menées sur l'homme lors de la Seconde Guerre Mondiale, les différents pays se sont concertés pour définir des principes éthiques à appliquer lorsqu'on effectue une recherche scientifique impliquant l'homme. Ainsi en 1947, le Code de Nuremberg définissait les principes fondamentaux d'éthique de la recherche biomédicale. Ce code a constitué par la suite la base des textes de loi réglementant les expérimentations humaines et a ainsi donné naissance aux déclarations d'Helsinki en 1964, de Tokyo en 1975 et de Manille en 1981. [18] Sur ces bases ont ainsi été élaborés les Bonnes Pratiques Cliniques (BPC) en 1987, puis unifiés en 1997. Tous les deux ans des conférences internationales d'harmonisation ont lieu pour permettre à l'Europe, aux Etats-Unis et au Japon d'harmoniser leurs pratiques. [18] En France depuis 1988, les essais cliniques sont régis par la loi Huriot-Sérusclat du code de la Santé Publique, renforcé par la suite par la directive européenne de 2001 et entré en vigueur en 2006. Cette loi stipule que tout projet de recherche clinique impose la constitution d'un dossier et sa soumission au CPP. Le projet doit contenir un promoteur, un investigateur principal et être autorisé par l'ANSM. Cette loi institue également une protection obligatoire des individus qui se prêtent à des études de recherche clinique, avec un devoir d'information des personnes avec la nécessité d'obtenir leur consentement éclairé. [19] La déclaration d'Helsinki, qu'on retrouve presque systématiquement en annexe de tous protocoles cliniques, énonce les principes éthiques applicables à la recherche biomédicale. [18, 19] En 2012, en France, la loi Jardé a été votée pour encore mieux protéger les personnes. Cela en changeant d'approche et en opérant une classification des essais en fonction du risque encouru par la personne incluse dans la recherche. C'est ainsi que la notion de recherche biomédicale en 2016, fut remplacé par la recherche impliquant la personne humaine, regroupant les recherches interventionnelles et non interventionnelles. [19] C'est ainsi qu'ont été définies les différentes

catégories de recherche impliquant la personne humaine décrites précédemment. L'ANSM met à disposition une partie de ces textes de loi et du code de la Santé Publique qui font office de réglementation. [20]

Les essais cliniques nécessitent de collecter et de traiter des données de santé. Ces données sont des données à caractère personnelle considérées comme très sensible et méritent ainsi une protection plus élevée. Cette protection est ainsi prévue par plusieurs réglementations et lois comme le Règlement Général sur la Protection des Données personnelles (RGPD), en application depuis 2018, et la Loi Informatique et Liberté (LIL), de 1978 modifiée en 2018. La Commission Nationale de l'Informatique et des Libertés (CNIL) a adopté cinq méthodologies de référence qui offrent un cadre sécurisé pour la mise en œuvre des traitements de recherche dans le domaine de la santé (MR-001, MR-003, MR-004, MR-005, MR-006). Les MR-001 et MR-003 concernent les recherches impliquant la personne humaine, telles qu'elles sont définies par le code de la Santé Publique. [19]

Dans le cadre d'un essai clinique il existe donc de nombreuses normes et lois qui régissent son déroulement et sa mise en place. La première étape d'un essai clinique est ainsi la rédaction du protocole de recherche qui sera soumis aux différents comités afin d'être validé, dans le but d'obtenir les différentes autorisations avant de démarrer cet essai. Ce protocole doit faire une revue représentative de la littérature scientifique sur le sujet de recherche et expliquer pourquoi il est nécessaire d'entreprendre cette recherche. Il doit donc avoir une légitimité scientifique, respecter un rapport bénéfice/risque favorable pour les patients et définir l'objectif de la recherche. Il doit également prendre en compte les cas particuliers bénéficiant d'une protection renforcée et assurer la prise en charge des événements indésirables qui peuvent avoir lieu au cours de l'essai. [19] Ce protocole se doit de respecter une méthodologie précise suivant les différentes normes et réglementations qui ont pour objectif de protéger et garantir la sécurité des patients impliqués dans un essai clinique. Les BPC permettent également de définir les termes, les responsabilités de chacun (promoteur et investigateur), les aspects scientifiques de certains essais, les responsabilités du comité d'éthique et le contrôle de la qualité. [21]

Ainsi de nombreux métiers et domaines vont être impactés par ces différentes lois et réglementations. Du médecin aux statisticiens, tous doivent respecter les normes qui leurs incombent afin de permettre la bonne réalisation de l'essai clinique et de garantir la sécurité des patients. Dans cette large gamme de métier, nous allons retrouver le Data Manager qui assure la création des bases de données et la gestion de celles-ci.

## **2.2. Data Management d'essai clinique**

### **2.2.1. Qu'est-ce que le Data Management en général ?**

Le Data Management, ou gestion des données, est un processus incluant l'acquisition, la validation, le stockage, la protection et le traitement des données nécessaires pour assurer la fiabilité, l'accessibilité et la création des données. Avec les volumes de données toujours plus importants et l'avènement du Big Data, il est de nos jours essentiel de gérer ses données produites toujours plus rapidement. Le Data Management est là pour orchestrer le traitement de cet afflux de

données. Ces données pouvant être exploitées à de nombreuses fins, la gestion de celles-ci permet à une organisation de rendre les données précieuses afin d'obtenir des informations approfondies et prédictives. Ces informations extraites et utiles peuvent ainsi être exploitées dans de nombreux domaines, par exemple pour prendre des décisions commerciales ou dans le cadre de la gestion des essais cliniques. [22]

Les entreprises se voient donc aujourd'hui dans l'obligation de se doter de nouveaux outils et de nouvelles pratiques pour exploiter ces immenses sources d'information. Ainsi le principe du Data Management réside dans la conception et la mise en place de base de données permettant de collecter, stocker, exploiter et sécuriser les données, et ce tout au long du cycle de vie de celles-ci. [23] L'un des premiers enjeux du data management est le besoin d'anticipation, c'est-à-dire prévoir quelles données vont être utilisées dans le secteur d'activité et comment les organiser. Ce déluge de données hétérogènes représente une importante source d'informations. Exploiter ces données peut donc influencer et aider à la prise de décision. [23] Cependant plusieurs composantes apparaissent face à cet enjeu. Tout d'abord la gouvernance des données. Elle fait référence aux politiques et processus utilisés pour assurer l'intégrité, la qualité et la sécurité des données. Il englobe les lignes directrices concernant l'application de ces politiques, la responsabilité globale et l'autorité de gouvernance. C'est l'une des premières étapes à prendre en compte pour le lancement d'un projet utilisant une grande quantité de données. Elle permet de garantir le respect de la vie privée, l'intégrité et l'image de l'entreprise, la qualité des services et assurer la cybersécurité. La gouvernance établit les lois sur les données et la façon dont elles sont appliquées pour encadrer la collecte et le traitement des données. [22] Ensuite, l'intégration des données consiste à déplacer, consolider et éclaircir les données en un seul endroit accessible pour l'entreprise. Cette étape permet d'accéder aux données et de les exploiter facilement et rapidement. Il existe de nombreuses façons d'intégrer ses données que ce soit de façon manuelle ou automatisée. [22] Puis, la gestion des données de références utilise des méthodes d'intégration des données. Le but de cette gestion est d'éviter toutes confusions dans différentes terminologies. Pour cela la gestion des données de références consiste à rendre les données utilisables et de gérer leur qualité. Il s'agit donc de l'ensemble des méthodes, outils et processus permettant d'assurer la qualité des données de référence. Ce sont donc des données avec le minimum d'erreur et utilisables sans risque. [22] Enfin la dernière composante, apparaissant néanmoins comme majeur, est la sécurité des données. Cette sécurité doit être au centre de toute stratégie liée aux données. Cette sécurité fait référence aux mesures mises en place pour assurer la protection des données à tous les stades du cycle de vie de celles-ci. Cette protection concerne les menaces externes (hackers, vol de données, ...) mais également les menaces internes (corruption des données, destruction, ...). [22]

Le Data Management dans le milieu académique est très différent du Data Management en entreprise ou dans l'industrie pharmaceutique même si les missions générales et les différents enjeux restent similaires voire identiques. Dans le cadre de la gestion des données d'essais cliniques, le Data Management donne la garantie de la restitution totale et fiable des données en respectant les BPC et la réglementation en vigueur dans chaque pays en vue de leur exploitation ; le respect des droits d'accès et de la confidentialité des données au cours de l'étude et au-delà ; et la sécurité des données et leur conservation. [24] Le Data Management d'essai clinique est défini par l'ensemble des actions de gestion et de traitement des données contribuant à assurer la documentation et la qualité de la base de données clinique. Dans cette base sont reportées toutes les informations recueillies au cours de l'étude. Ces études sont régies par la loi Santé Publique de 2004 et doit

répondre aux obligations légales en vigueur (confidentialité des données, qualité des données et traçabilité des interventions). [24]

Ainsi afin de répondre à toutes ces obligations et respecter les réglementations en vigueur, le Data Manager a de nombreuses missions à accomplir tout au long d'une étude.

## **2.2.2. Quelles sont les missions du data manager d'essai clinique ?**

### *2.2.2.1. Missions principales du data manager à l'institut Bergonié*

Il est possible de diviser les principales missions du Data Manager en trois parties : les missions avant l'ouverture de l'étude, au cours de l'étude et en fin d'étude. [24] Pour généraliser quelle que soit l'étape le Data Manager se doit de garantir la qualité, la confidentialité et la sécurité des données selon les BPC de Data Management ; de vérifier que le traitement des données est réalisé en conformité avec les réglementations ; d'assurer la traçabilité de toutes les interventions sur la base de données ; et de participer à la création et à la mise à jour des documents qualité de Data Management. [25]

Ainsi les missions du Data Manager avant l'ouverture de l'étude consiste à valider le CRF (Case Report Form ou Cahier d'Observation) par rapport au protocole afin de s'assurer que les variables nécessaires pour l'analyse statistique sont présentes par exemple, rédiger le plan de Data Management (DMP), mettre en place la structure informatique de la base de données et en définir sa gestion : annoter le CRF papier ; construire, tester et valider l'e-CRF (CRF électronique) ; rédiger le plan de validation des données (DVP) qui sera suivi pour le paramétrage des contrôles de cohérence sous le logiciel e-CRF (pour les contrôles automatiques) et sous un logiciel de programmation statistique (pour les contrôles dit manuels par exemple le logiciel SAS) ; et élaborer la fiche d'identification de l'étude (voir figure 3, avant T0 : Saisie des données patients). [24]

Tout au long de l'étude le Data Manager doit assurer la gestion de la base de données, c'est-à-dire maintenir la structure de la base et sa documentation suite aux Modifications Substantielles (MS) et demande ponctuelle de modification de la base de données, s'assurer de la présence et de la cohérence des données dans la base grâce à la validation la base de données clinique en suivant le DVP, poser des demandes de clarifications si nécessaire, réconcilier les bases de pharmacovigilance et de relecture, participer aux COPIL (COmité de PILotage), transmettre la liste des événements indésirables pour validation du codage des événements indésirables par la Pharmacovigilance (PV) et poser des demandes de clarification si besoin, réaliser la randomisation lorsque le statisticien est dans l'impossibilité de l'effectuer, réconcilier la base clinique et la base de randomisation. [24]

Enfin à la fin de l'étude, il faut s'assurer que les étapes avant le gel de la base sont respectées, envoyer la demande de fermeture de la base à l'administrateur et s'assurer que l'extraction est bien envoyée au statisticien, et archiver les données, les programmes de contrôle des données et les documents papiers de l'étude concernant le Data Management. [24] Le schéma suivant résume ces différentes missions.



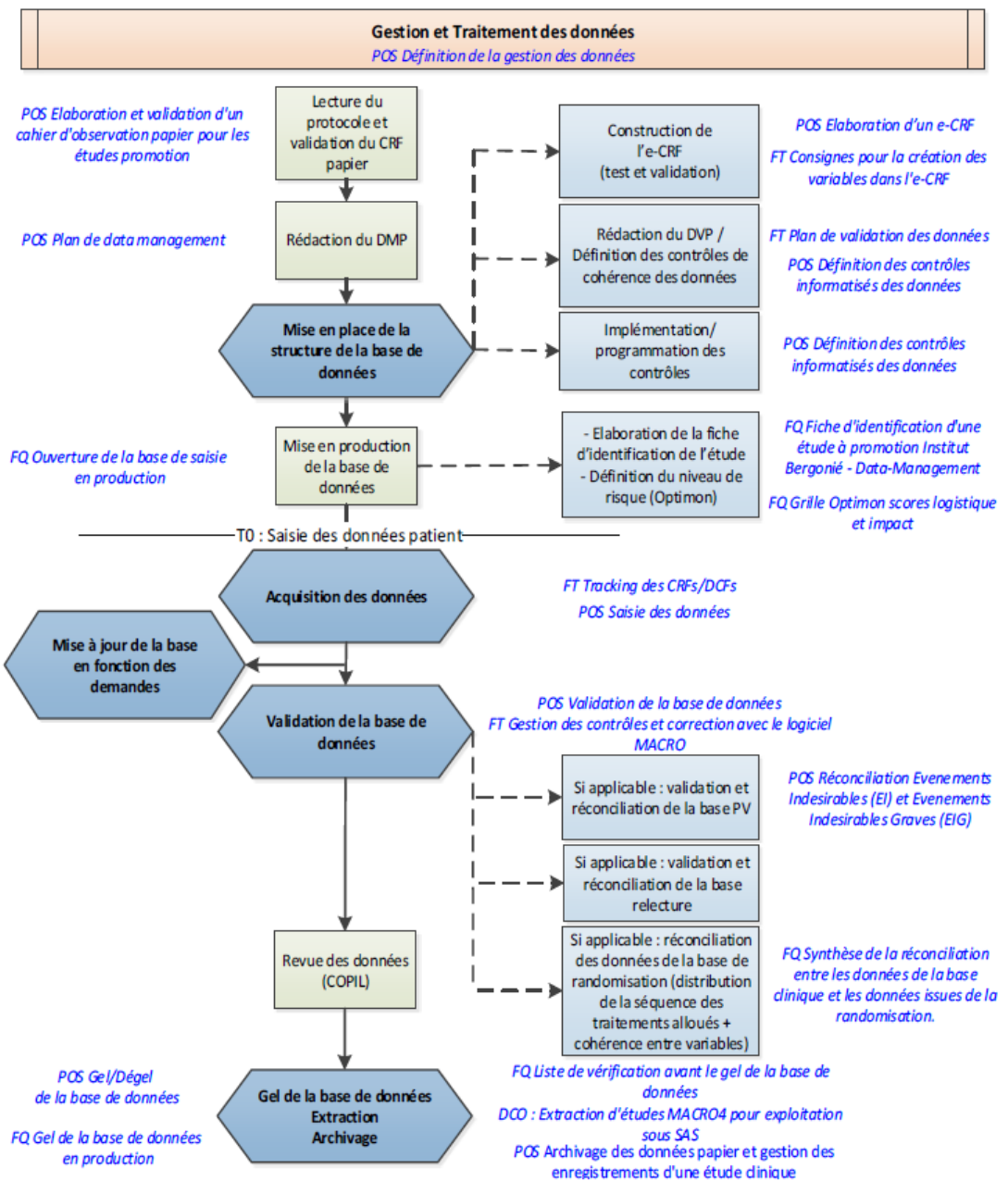


Figure 3 : Gestion et traitement des données à l'institut Bergonié, missions du Data Manager [24]

Pour résumer le Data Manager a pour mission de créer, de valider et de mettre à jour la base de données en assurant la rédaction des documents qualité de Data Management.

### 2.2.2.2. E-CRF et Documents qualité de Data Management

Ainsi l'une des missions majeures du Data Manager est l'élaboration de l'e-CRF. Il est une interface de saisie et de contrôle des données à distance permettant de garantir la confidentialité des données et l'accessibilité des données pour les personnes habilitées. L'e-CRF facilite le contact entre les différents partenaires de l'étude. Il permet entre autres de saisir, visualiser et contrôler les données en temps réel en améliorant la qualité des données grâce aux contrôles interactifs (défini dans le DVP), et d'offrir une meilleure maîtrise du circuit des données en assurant la traçabilité des

données. [26] L'e-CRF est le reflet du CRF papier intégrant un remplissage intelligent et des contrôles de cohérence à la saisie. Pour le mettre en place il est nécessaire de disposer d'outil de création et de gestion de bases de données. [26]

Une autre mission importante du Data Manager est la rédaction des documents de Data Management qui sont notamment le DMP et DVP. Le DMP est un document obligatoire dans la gestion des données d'un essai clinique. Ce document regroupe les informations nécessaires à la gestion d'une étude et planifie les tâches relatives à la gestion des données. Il permet de faire le lien entre les différents intervenants d'un essai clinique. Il reprend ainsi l'organisation générale du Data Management de l'étude. [27] Le DVP (Plan de validation des données) recense l'ensemble des contrôles de cohérence mis en place pour l'étude. Il contient les spécificités de chaque contrôle et le détail des données à valider. Il doit être validé par l'équipe projet, c'est-à-dire l'ARC promoteur, l'investigateur et le statisticien de l'étude notamment. [28] L'objectif de définir des contrôles est de vérifier la rigueur du suivi du protocole et aboutir à une base de données fiable. Ainsi le DVP contient des contrôles automatiques (paramétré directement dans le logiciel) permettant de vérifier les incohérences en temps réel lors de la saisie sur l'e-CRF, et des contrôles manuelles (paramétrées sous SAS) où les incohérences sortiront sous forme de listing qui seront revu par le Data manager manuellement lors de sa validation de données. [29] Ces contrôles permettent donc de valider les données et d'assurer leur qualité tout au long de l'étude.

Ainsi la mise en place de contrôles et la validation des données permettent d'assurer la qualité, la traçabilité et la fiabilité des données. Cette validation est très importante pour respecter les réglementations et les BPC du Data Management. Néanmoins aujourd'hui avec les nouvelles normes et la nécessité de protéger les données personnelles, il est essentiel de valider également les logiciels utilisés pour traiter et valider ses données, et cela passe notamment par la Validation des Systèmes Informatisés.

## **2.3. Validation des Systèmes Informatisés (VSI)**

### **2.3.1. Qu'est-ce que la VSI ?**

#### *2.3.1.1. Définition générale de la validation*

Dans le dictionnaire on peut trouver que la validation est le fait de valider quelque chose, c'est-à-dire de prouver que quelque chose présente les conditions requises pour produire son effet. La validation est donc une opération qui a pour objectif de démontrer, avec des documents à l'appui, qu'une procédure, un procédé ou une activité conduit bien aux résultats attendus. Il s'agit de fournir des preuves documentées afin de justifier qu'un processus donne constamment le résultat souhaité conforme aux spécifications et aux caractéristiques de qualité requises. Ainsi il faut vérifier que toutes les étapes de fabrication ou du fonctionnement d'un produit sont conformes aux exigences et aux normes en vigueur. [30] On peut généralement diviser la validation en trois phases : la phase de prévalidation ou de qualification qui couvre le développement du produit, la phase de validation vérifiant toutes les limites établies, et la phase de maintien de la validation qui assure une vérification fréquente du bon état de fonctionnement du système. La validation est un processus très strict qui nécessite la création d'un protocole de validation (tout comme le protocole susmentionné

pour un essai clinique). Ce protocole décrira toutes les étapes et toutes les phases de la validation. [30]

Quel que soit le produit ou le système à valider il existe une étape importante et incontournable de la validation qui est l'analyse des risques. L'analyse de risque est utilisée comme première étape d'un processus d'évaluation des risques et de validation. Son résultat permet de mettre en évidence les différents dangers potentiels pouvant résulter de l'utilisation d'un produit ou d'un système. Un danger ou risque peut à lui seul ou combiner à d'autres risques entraîner l'échec de fonctionnement du produit ou du système. [31] La gravité des risques va être évaluée en prenant pour référence la condition finale la plus grave possible. A cette gravité s'ajoute la probabilité d'occurrence du risque qui une fois combinés permettent de déterminer la criticité du risque. Cette criticité permet d'établir une notion d'acceptation du risque. Ainsi l'objectif principal de l'analyse de risque sera à terme de mettre en place des mesures de minimisation du risque afin de contrôler ou d'éviter le danger. [32, 33] L'analyse des risques est une démarche relativement récente (fin du 20<sup>e</sup> siècle), c'est pourquoi il n'existe pas aujourd'hui réellement de normes mais uniquement des démarches et des guides de recommandations tel que le GAMP 5. [31]

Toutes ces étapes de la validation d'un système sont importantes et aujourd'hui prouver la conformité et le bon fonctionnement des systèmes est devenu très important. C'est pourquoi la validation s'applique désormais de plus en plus aux systèmes informatisés.

### *2.3.1.2. La validation appliquée aux systèmes informatisés*

Comme définit précédemment la validation est un processus documenté qui fournit un degré élevé d'assurance. [34] On peut définir un système informatisé comme un système informatique supportant un processus de travail, une fonction. Ainsi, un système informatique fait exactement et de manière cohérente ce pour quoi il a été conçu. [34, 35] La VSI, ou Computerized System Validation en anglais, consiste à documenter qu'un système répond bien de manière cohérente et logique à un ensemble de conditions prédéfinies. La VSI doit donc garantir l'exactitude, la fiabilité, la performance attendue et la capacité à discerner des enregistrements non valides ou altérés. [35, 36] Il faut donc préciser que chaque application remplit son objectif tel qu'il le devrait. [34] Le processus de validation est utilisé notamment lorsque le papier est remplacé par des données électroniques dans des environnements hautement réglementés pour s'assurer que le système est transparent, robuste et inviolable. [36] La documentation associée à la VSI est extrêmement détaillée. Il est nécessaire que n'importe qu'elle personne qui lit ces documents puissent refaire le processus dans son intégralité. Ces documents peuvent être par exemple utilisés lors d'un audit pour s'assurer du respect des normes du logiciel. [36]

On peut résumer en 5 points ce qu'apporte la VSI :

- La précision : les résultats des tests sont régulièrement vérifiés par rapport aux résultats attendus ;
- La sécurité : les processus de la VSI assurent une traçabilité forte et indiquent clairement chaque entrée dans le système ;
- La fiabilité : les processus garantissent que les sorties du système peuvent être utilisées tout au long du cycle de vie du logiciel ;

- La cohérence : les processus garantissent également que les sorties correspondent aux entrées tout au long du cycle de vie du logiciel ;
- L'optimisation : les systèmes informatisés sont optimisés de par les mesures de prévention des risques rendant le logiciel plus efficace et efficient. [36]

La VSI est donc un processus qui se déroule tout au long du cycle de vie d'un système informatisé et qui permet de garantir au maximum la conformité de celui-ci. Cette validation est d'autant plus importante dans les secteurs sensibles tels que la santé qui traitent des données personnelles. Il est néanmoins légitime de se demander pourquoi il est si important de valider et comment cela est possible pour les systèmes informatisés en santé.

## **2.3.2. Pourquoi et comment valider les systèmes informatisés en santé ?**

### *2.3.2.1. Le RGPD appliquée au domaine de la santé*

En Europe, Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union Européenne (UE). Le contexte juridique s'adapte pour suivre l'évolution des technologies et de nos sociétés. Ce règlement européen s'inscrit dans la continuité de la LIL et renforce le contrôle par les citoyens de l'utilisation qui peut être faite de leurs données. Il permet d'harmoniser les règles au sein de l'UE en offrant un cadre juridique unique aux différents professionnels. Il permet de développer leurs activités numériques en se basant sur la confiance des utilisateurs. [37] Tout organisme public ou privé qui traite des données personnelles est concerné dès lors qu'elle est établie sur le territoire européen ou que son activité cible directement des résidents européens. [37] Le RGPD impacte fortement le Data Management dans le domaine de la santé et plus généralement les activités liées au traitement des données de santé. Par exemple, dans certains pays européens comme la Serbie candidat pour entrer dans l'Union européenne, le RGPD a obligé la création de nombreuses procédures pour permettre à ce pays de s'harmoniser avec le reste de l'Europe. Dans le but que celui-ci s'intègre avec les différentes pratiques en vigueur dans le reste de l'union et puisse ainsi échanger des données en assurant la sécurité et la fiabilité de celles-ci. [38]

Sur le site de la CNIL plusieurs recommandations ont été publiées afin d'aider au mieux les organismes à se préparer au RGPD. La CNIL précise que ces organismes doivent assurer une protection optimale des données à chaque instant et être en mesure de la démontrer en documentant leur conformité. Ainsi, il conseille 6 étapes à suivre pour se préparer qui sont : désigner un pilote, cartographier le traitement de données personnelles, prioriser les actions à mener, gérer les risques, organiser les processus internes et documenter la conformité. [39] Ce sont ces 3 dernières étapes qui impliquent grandement la nécessité d'une VSI. En effet, la démarche de la VSI permet de gérer les risques, d'organiser les processus et de documenter la conformité. Bien évidemment le RGPD n'est pas le seul règlement à prendre en compte. En effet, les BPC imposent également de valider les systèmes informatisés en santé tout comme les différents accords et lois existants. Par exemple en Corée du Sud en 2019, un groupe de chercheurs a publié un article pour prouver que leur système informatisé de gestion de l'imagerie des essais cliniques était bien conforme aux bonnes pratiques. Ils se sont basés pour cela des recommandations de la Food and Drug Administration (FDA) des Etats-Unis ainsi que des BPC pour prouver la conformité de leur système. Ils ont donc pour cela effectué une validation de leur système en interne et en externe par une société mondiale d'assurance qualité. Ils ont démontré que leur système répondait à toutes les

exigences réglementaires et fonctionnelles, avec une documentation soignée afin de prouver la bonne validation. [40]

### 2.3.2.2. La VSI en santé et la méthodologie du GAMP 5

L'un des premiers articles traitant de la VSI dans le domaine de la santé date de 2015. Des chercheurs chinois ont cherché à démontrer le bon fonctionnement de leur système informatisé pour la gestion des essais cliniques. Leur objectif était de prouver que leur système répondait bien aux différentes exigences réglementaires. Pour cela ils ont décrit le processus de validation et démontré qu'il était essentiel de mettre en place des procédures de validation complètes et de maintenir une documentation adéquate accompagnée de documents de formation. [34]

La VSI est donc majeure dans le domaine de la santé, que ce soit pour respecter toutes les normes et lois existantes, particulièrement dans un cadre strict et réglementé comme celui des essais cliniques, ou bien pour assurer la sécurité et la qualité du traitement des données. La VSI permet d'éviter de graves conséquences négatives directes ou indirectes sur les patients. [35] Aujourd'hui ne pas effectuer de validation pourrait être considéré comme contraire aux réglementations. Pour valider un système informatisé on peut distinguer cinq phases distinctes : la préparation, la définition des besoins (Blueprint Definition : figure 4), la création du système (Configuration/Coding : figure 4), la vérification du système et le reporting. [35] Le schéma suivant illustre ses phases distinctes mais liées les unes aux autres.

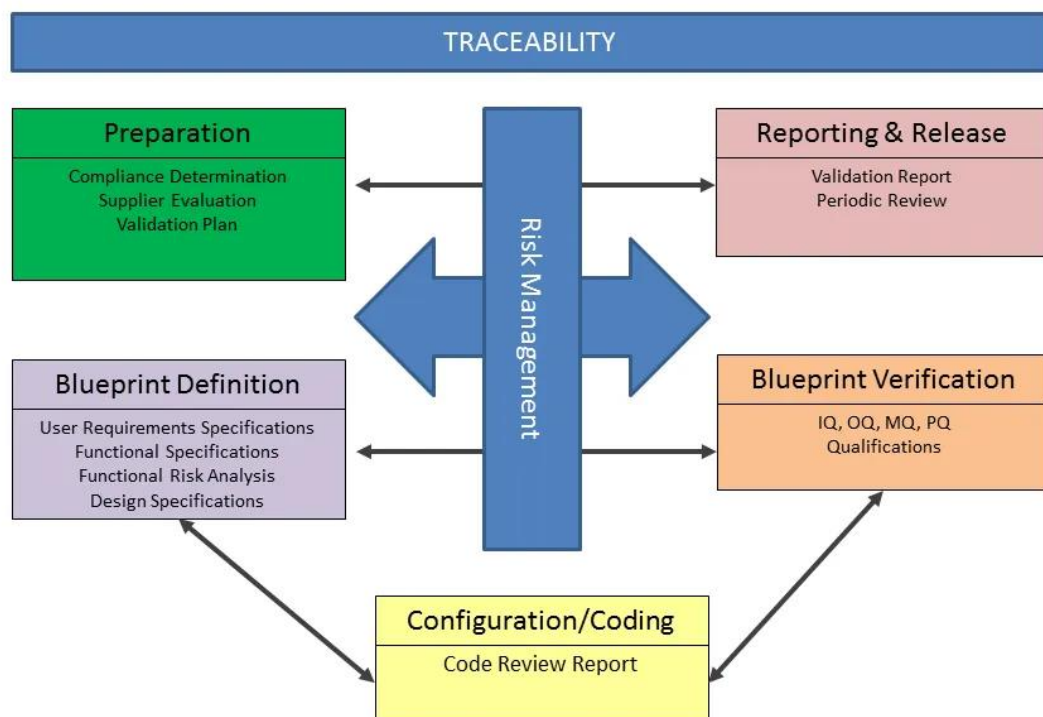


Figure 4 : Exemple d'étapes de validation d'un système informatisé [35]

Néanmoins, il n'existe pas réellement de méthodes ou de règlements fixes pour effectuer une validation. Aujourd'hui le document majeur faisant office de guide dans le milieu pharmaceutique et de la santé afin d'effectuer une validation des systèmes informatisés est le GAMP 5. Le GAMP 5 ou Good Automated Manufacturing Practices version 5 est produit par l'International Society for Pharmaceutical Engineering (ISPE) et répond aux questions relatives à la validation et à la conformité réglementaire des systèmes informatisés. Il se réfère aux bonnes pratiques et aux différentes réglementations européennes [41] Ainsi bien qu'il ne soit pas un texte réglementaire, le GAMP 5 est à prendre en compte dans la validation et même dans la conception et le développement des systèmes informatisés en santé. [42] Dans ce document la validation informatique est définie comme un effort basé sur l'analyse des risques. Les facteurs à considérer sont les risques pour le patient, pour l'intégrité des données et pour la structure (ou risque business). [43] Ainsi généralement dès la fabrication du système le GAMP 5 recommande un système en V qui impose une vérification à chaque étape du processus afin de s'assurer de la qualité de celui-ci. L'objectif global est de garantir que le système informatisé fonctionne conformément aux résultats attendus et aux besoins de l'utilisateur. [36] La validation selon le GAMP 5 réside donc sur une approche basée sur les risques et la première étape majeure est donc d'effectuer l'analyse des risques.

Ainsi après l'analyse des risques, on pourrait diviser la validation en 4 étapes successives permettant d'assurer l'objectif global de celle-ci :

- La définition des besoins des utilisateurs : que doit faire le système pour répondre à ses besoins ;
- La rédaction d'un plan de validation : détaillé et qui explique chaque étape à suivre pour garantir la validation du système ;
- La rédaction d'un plan de test ou cahier de recette : quels tests seront effectués, qui doit les effectuer et quels sont les résultats attendus ;
- Les fiches de test : comprenant les qualifications d'installation (QI), les qualifications opérationnelles (QO) et les qualifications de performances (QP) qui vérifient que le système répond à toutes les spécifications et fonctionne comme prévu. [36]

Toutes ces étapes sont ainsi documentées afin de prouver la validité du logiciel ou du système. [36]

De nos jours, avec les nouvelles réglementations et l'importance d'assurer la sécurité et la protection des données personnelles, la VSI apparait comme essentielle dans le domaine de la santé. Elle permet de prouver la conformité du système informatisé et le respect des lois et normes en vigueur. La VSI semble donc bien être un enjeu majeur dans ce secteur et il est évident que cette notion de validation et de conformité devient de plus en plus importante pour les patients comme pour les professionnels. Il faut continuer à avancer dans le domaine pour permettre d'avoir toujours plus de transparence, de sécurité, de fiabilité et de confiance dans les systèmes informatisés en santé.

## 2.4. Contributions du Data Manager dans la VSI

L'institut Bergonié s'est associé avec le GIRCI afin de mettre en place les différentes procédures relatives à la VSI (voir annexe 2). Au sein de celui-ci, on peut diviser la VSI en deux grandes étapes. La première étape consiste à réunir une équipe d'experts métiers afin de réaliser les différentes procédures dites standards, qui serviront par la suite pour la validation des différents logiciels utilisés à l'institut dans le cadre des essais cliniques. La seconde étape quant à elle réside dans la validation en elle-même de tous ces logiciels, et la réalisation des tests sur ceux-ci, par les experts métiers qui utilisent ces logiciels au quotidien.

### 2.4.1. Elaboration des procédures standards

Pour l'élaboration des procédures standards, la première étape a été de se réunir avec une équipe d'experts de différents domaines au sein de l'institut afin d'étudier et de comprendre les attentes du GIRCI et les enjeux de cette VSI. Cette équipe était composée dans un premier temps d'une responsable qualité, d'une pharmacienne, d'une informaticienne, d'une Data Manager et de moi-même. Le GIRCI a partagé plusieurs drafts de procédures nécessaires à la mise en place des différentes étapes de la VSI. L'objectif était par la suite de travailler ensemble sur ces procédures dans le but de les adapter à l'institut, tout en les partageant avec le reste du GIRCI. Ces réunions internes ont permis de créer la procédure VSI générale de l'institut et ainsi de se répartir la rédaction des différents documents standards pour la suite du projet. Nous avons ainsi établi la nécessité de faire tout d'abord un plan de validation du logiciel accompagné d'une analyse de risque de celui-ci. Ensuite, nous avons déterminé qu'il fallait concevoir un cahier de recette comportant l'intégralité des tests à réaliser sur le logiciel, puis une fiche de test décrivant les tests à effectuer avec les preuves associées. Enfin, nous avons conclu qu'il était nécessaire de terminer par un rapport final sur l'état de la validation du logiciel.

Les premiers documents à créer étaient donc le plan de validation ainsi que l'analyse de risque. Pour cela nous sommes partis des documents fournis par le GIRCI ainsi que de la méthodologie du GAMP 5. Le plan de validation contient l'intégralité des informations nécessaires sur le logiciel et les différentes étapes de la validation de celui-ci. Il s'agit du fil directeur à suivre lors de la validation d'un logiciel. Ce document contient l'objectif principal de la VSI, une liste des documents de référence, une description du SI, la méthodologie de validation, le circuit de signatures ou bien encore les règles de classement et d'archivage des documents. Un plan de validation doit être créé pour chaque logiciel à valider. Comme dit précédemment, ce plan définit ainsi la ligne directrice de la validation du logiciel. L'analyse de risque a été conçue en parallèle du plan de validation, celle-ci constituant la base des documents qui seront créés lors de la suite de cette validation. Dans un premier temps, nous nous sommes inspirés des différentes approches que l'on a pu trouver sur internet et de ce qu'il était possible de trouver dans le GAMP5. Ainsi, une première version de l'analyse de risque a pu être mise en place, harmonisée avec ce qui existait déjà au sein de l'institut. Pour accompagner le remplissage de cette analyse de risque, un document a été créé pour expliquer le fonctionnement de la trame et la manière de la remplir. Ces deux documents constituent la base de la VSI, mais d'autres documents sont bien évidemment nécessaires pour la suite de la validation. Ainsi, ont été créés, le modèle du cahier de recette et la fiche de test standard.

Pour le cahier recette nous sommes partis du modèle fourni par le GIRCI, tout comme pour la rédaction de la fiche de test. Le cahier de recette comprend l'intégralité des tests à effectuer avec le profil de l'utilisateur à utiliser lors de chaque test (si la personne doit avoir un profil par exemple de Data Manager ou d'informaticien). La fiche de test contient l'ensemble des tests réalisés avec la résolution de chacun (OK ou KO), ainsi que toutes les preuves associées à la réalisation du test. Ces documents constituent la base à remplir pour valider un logiciel. Pour terminer, un rapport final de la validation a été créé. Celui-ci représente l'état final de la validation, il résume l'ensemble du processus et indique si le logiciel est conforme ou si des anomalies majeures ont été détectées.

Toutes ces procédures standards serviront par la suite à la validation des différents logiciels de l'institut, comme par exemple pour le logiciel de Data Management MACRO.

## **2.4.2. Validation du logiciel MACRO**

### *2.4.2.1. Rédaction des documents pour la validation*

MACRO est un logiciel fourni par Elsevier. Ce logiciel est défini comme une solution de collecte de données électroniques permettant à ces utilisateurs de saisir, de surveiller et d'exécuter rapidement des rapports sur les données dans le but de recueillir des données précises et fiables pour analyse. Ce logiciel constitue une ressource fiable pour la recherche clinique, car celui-ci est régulièrement audité par les BPC afin de démontrer sa conformité. [44] MACRO est l'un des logiciels utilisés pour la création et la gestion de bases de données dans le cadre des essais cliniques. Il contient par conséquent de nombreuses données personnelles concernant les patients de ces essais. C'est pourquoi il est important de valider un tel logiciel. Celui-ci est principalement utilisé par les Data Manager, c'est pourquoi c'est un Data Manager qui sera en charge de cette validation. MACRO a été choisi comme premier logiciel à valider au sein de l'institut et va permettre d'effectuer un premier test pour la validité et l'utilisabilité des différentes procédures standards rédigées précédemment.

La première étape consistait donc à rédiger les différents documents relatifs à la validation du logiciel. Le premier document à rédiger est le plan de validation qui constitue le fil conducteur de la validation. Il décrit les fonctionnalités du logiciel, ce qui est attendu lors de la validation, avec notamment la démarche générale, comme on peut la voir sur la figure 5, et les différents modules qui vont être validés. En effet, tous les modules du logiciel ne sont pas utilisés, ou très peu, et ceux-ci ne rentreront pas dans le scope de la validation.



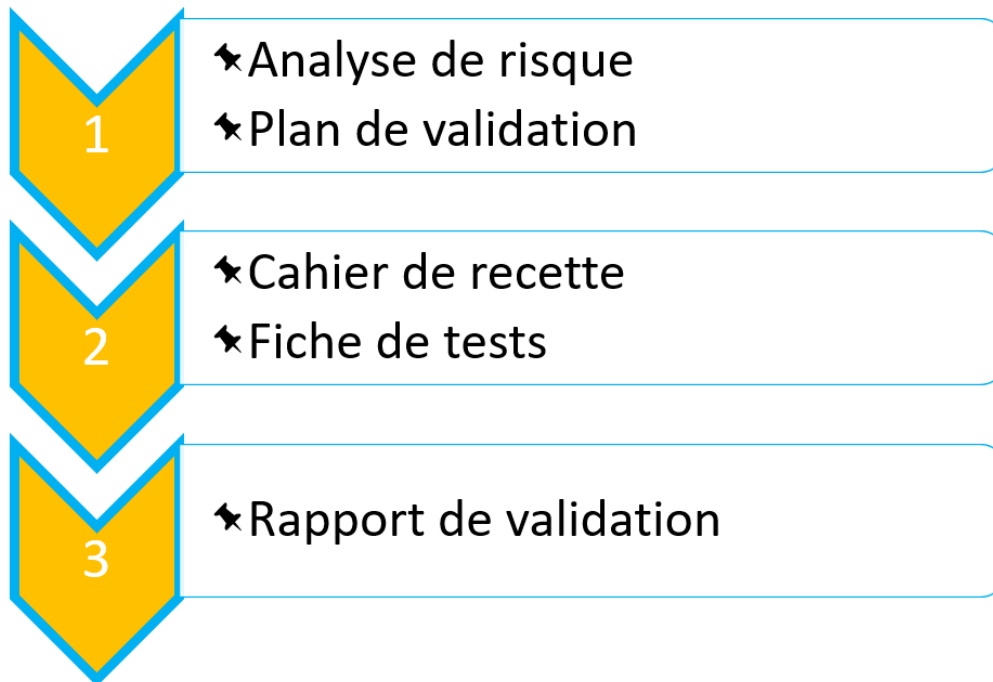


Figure 5 : Démarche générale de la validation présente dans le plan de validation

En parallèle du plan de validation, on effectue, avec une équipe de plusieurs personnes utilisant le logiciel, une analyse de risques de celui-ci, qui va permettre par la suite de déterminer les tests qui seront nécessaires pour la validation. Cette analyse comprend une liste des risques identifiés avec leur niveau de criticité ainsi que les mesures de maitrise du risque mise place, comme on peut le voir dans la figure 6.

Etapes	Activités	n° risque CQ	Risque/événement redouté	P Vraisemblance 1 à 5	G Gravité 1 à 5	C Criticité brut P x G	risque brut	Mesures de maitrise de risques	M MAITRISE 1 à 5	Cr Criticité résidu	risque	Actions à mener	Indicateurs
			Possibilité de créer les dossiers de patients...	2	2	4	risque à surveiller	Monitoring des patients + message d'alerte MACRO	1	4	sous contrôle		
			Echange d'identité entre deux patients	1	4	4	risque à surveiller	Monitoring des patients	1	4	sous contrôle		
			Indication d'une vanne erronée dans la	2	5	10	risque à surveiller	traçabilité avec le formulaire de prescription +	1	10	sous contrôle		
			Perte des données liées à une session ou	1	5	5	risque à surveiller	Sauvegarde journalière	1	5	sous contrôle		
			Report de données pour commencer par une	1	5	5	risque à surveiller	Audit trail du logiciel	1	5	sous contrôle		
			responsable d'exporter les données	1	2	2	risque non critique	Propriétaire remote et administrateur de données	1	2	sous contrôle		
			responsable d'exporter les données pour	1	3	3	risque non critique	Création de profils standards	1	3	sous contrôle		
			Attribution de droits sur une mauvaise	2	3	6	risque à surveiller	Tableau de demande de droits	1	6	sous contrôle		
			Attribution de droits sur un mauvais site	2	3	6	risque à surveiller	Tableau de demande de droits	1	6	sous contrôle		
			Connexion avec le site d'origine	2	5	10	risque à surveiller	Atténuation de non divulgation de mot de passe	2	20	surveillance/vigilance		
			Non respect des réglementations	1	5	5	risque à surveiller	mise en place de procédures conformes aux	1	5	sous contrôle		
			Impossibilité de se connecter au logiciel	2	3	6	risque à surveiller	Procédure dégradée	1	6	sous contrôle		
			Blocage de compte	4	2	8	risque à surveiller	Contact de l'administrateur et/ou back-up	1	8	sous contrôle		
			Verrouillage de CRF	2	2	4	risque à surveiller	Contact de l'administrateur et/ou back-up	1	4	sous contrôle		
			Lenteur du serveur	3	2	6	risque à surveiller	Contact de l'administrateur et/ou back-up	4	24	surveillance/vigilance	Problème remonté à l'éditeur	
			Fermeture spontanée du logiciel	3	2	6	risque à surveiller	Contact de l'administrateur et/ou back-up	4	24	surveillance/vigilance	Problème remonté à l'éditeur	

Figure 6 : Analyse de risque du logiciel MACRO effectué à l'institut Bergonié

Une fois le plan de validation rédigé et l'analyse de risque effectué, il est possible de passer à la mise en place des différents tests et cela commence par la rédaction du cahier de recettes.

### 2.4.2.2. Mise en place des tests

Avant d'effectuer le moindre test il faut tout d'abord rédiger un cahier de recette. Ce cahier comportera l'intégralité des tests à réaliser pour les différents modules utilisés concernant le logiciel MACRO. Celui-ci s'appuie sur le plan de validation, l'analyse de risque, et les différents éléments mis en lumière important pour la validation. Pour écrire le cahier de recette, nous sommes partis des différentes tâches que nous faisons au quotidien et nous avons cherché qu'est-ce qu'il était important de vérifier. Ainsi pour chaque étape où nous utilisons le logiciel nous avons identifié un test, comme on peut le voir dans l'exemple de la figure 7.

#### 5 Actions à réaliser

Le cahier des recettes suivant sera rempli selon les points mis en évidence lors de la rédaction du plan de validation et de l'analyse de risque. Chaque test sera par la suite répertorié dans une fiche test. Le profil test correspond au profil normalement utilisé pour effectuer la tâche décrite dans le test.

LISTE DES DIFFERENTS TESTS A METTRE EN ŒUVRE (cf fiches de test)	
Test	Profil test
Installer MACRO	Informaticien
Création d'un compte (System Management)	Informaticien
Suppression d'un compte (System Management)	Informaticien
Déblocage d'un compte (System Management)	Informaticien
Créer une base de données (System Management)	Informaticien
Créer un nouveau site (System Management)	Informaticien
Créer des profils de droit spécifique : droits en lecture seule, droits de modification (System Management)	Informaticien

*L'impression de ce document est de votre responsabilité, il n'est plus valable s'il fait l'objet d'une révision*

« TUDÉDOC » n° « tsf », version : « n° édition »		Pages 2/3
Affectation de droit à un compte utilisateur : accès à un seul site (System Management)	Informaticien	
Passage d'une étude de DEV en PREPROD et de DEV en PROD (System Management)	Informaticien	
Se connecter au logiciel (Study Definition)	Data-Manager	
Se déconnecter du logiciel (Study Definition)	Data-Manager	
Se connecter à l'instance de notre choix (Study Definition)	Data-Manager	
Se connecter à l'étude de notre choix (Study Definition)	Data-Manager	
Créer un eForm et eForm répété (Study Definition)	Data-Manager	

Figure 7 : Cahier de recette pour le logiciel MACRO

Une fois ce cahier de recette rédigé il était possible de passer à la mise en place des différents tests en passant par la rédaction de la fiche de test associée. Cette fiche de test comprend l'intégralité des tests écrits dans le cahier de recette, avec le résultat qu'a obtenu le test, ainsi que les preuves associées à ce résultat (les preuves sont généralement une capture d'écran de la réalisation de la tâche). Pour le moment les tests de MACRO sont toujours en cours d'exécution et c'est pourquoi il n'est pas possible de montrer la fiche de test pour ce logiciel.

### 2.4.3. Discussion autour de la mise en place de la VSI

MACRO est le premier logiciel pour lequel la VSI a été mise en place. Il constituait ainsi un test pour les différentes procédures élaborées ainsi que pour la suite de la VSI. Les problèmes majeurs rencontrés résidaient dans l'appropriation des procédures fournies par le GIRCI. En effet, celles-ci ne correspondaient pas forcément à ce que l'on attendait à l'institut Bergonié et beaucoup de documents n'ont pas forcément été utilisés. La validation du logiciel MACRO est toujours en cours et il n'est pas possible de comparer avec les autres participants du GIRCI, ceux-ci n'utilisant pas le même logiciel pour leurs essais cliniques. La VSI constitue également une nouveauté pour tous les membres du GIRCI. C'est une notion récente que beaucoup de personnes ne connaissent pas ou peu. Au sein de l'institut la sensibilisation à la VSI est en cours et représente aujourd'hui un enjeu important pour tous les services traitant des données à caractère personnelles. C'est pourquoi nous avons rédigé des diapositives à présenter lors de la revue de direction pour présenter ce que cela représente. Dans un cadre plus générale dans le domaine de la santé, l'importance de cette validation est encore sous-estimée même si celle-ci représente un bon moyen pour respecter les différentes normes et lois en vigueur, comme le RGPD.

L'implication du Data Manager dans la VSI est importante car c'est un élément qui a l'habitude de manipuler les logiciels et de travailler les données personnelles. Il est habitué à respecter des normes que ce soit pour la création des bases de données ou bien la validation des données. Comme les autres métiers, il est habitué à rédiger des documents en suivant un modèle. Tous ces points font qu'il est important d'avoir au sein de l'équipe VSI un tel expert métier. C'est dans ceux-ci que réside la contribution d'un Data Manager dans la VSI.

La mise en place de la VSI est donc assez compliquée de par sa nouveauté et le manque de connaissance autour de ce sujet. Il faut donc s'inspirer de choses existantes mais différentes pour réussir à valider nos logiciels. Aujourd'hui à l'institut Bergonié, même si la VSI a été commencé en interne, tout comme pour certains membres du GIRCI, les différentes équipes réfléchissent à faire appel à des sociétés spécialisées pour l'accompagnement et la revue de procédures, afin de savoir ce qui est important ou non, et également pour la réalisation des tests et la mise en place des différentes validations pour chaque logiciel. L'un des problèmes de faire appel à des sociétés externes est parfois le coup financier de cette opération, principalement pour les structures avec de petits budgets. Il faut réfléchir si investir un tel budget dans cette mission ne permettra un gain de temps et qualité dans la réalisation de la validation. Néanmoins, on peut également se demander en contrepartie si une société externe qui n'utilise pas le logiciel au quotidien ne sera pas moins apte que les experts métiers à faire une telle validation. Il faut donc encore aujourd'hui peser le pour et le contre afin d'évaluer qu'elle est la meilleure option pour continuer cette validation et pour la mener à bien.

## 3. Missions de Data Management à l'institut Bergonié

### 3.1. Missions du Data Manager

Les missions du Data Manager au sein de l'institut Bergonié sont nombreuses et variées. Elles vont de la validation du cahier d'observation à la validation des données, en passant par la création de la base et la rédaction des documents de Data Management.

Au cours de mon alternance, j'ai eu l'occasion de travailler sur plusieurs études cliniques telles que les études BIOPSAR, IBMC ou encore METZOLIMOS. C'est au travers de ces différentes études que je vais décrire les différentes missions qui m'ont été confié à l'institut Bergonié.

#### 3.1.1. Rédaction des documents de Data Management

La première étape avant d'envisager la création des différents documents de Data Management est la lecture du protocole et la validation du CRF. Cette étape permet de se familiariser avec l'étude et de comprendre qu'elles en sont les enjeux. La première étude sur laquelle j'ai pu travailler et rédiger les documents de Data Management est l'étude BIOPSAR. Celle-ci porte sur la performance diagnostique de la biopsie utérine écho-guidée pré chirurgicale dans la prise en charge des tumeurs utérines suspectes de sarcomes. Son objectif est de démontrer que ces biopsies utérines dans le diagnostic des tumeurs utérines suspectes de sarcome présentent une performance acceptable. Ainsi, ma première mission après lecture du protocole était de relire et de valider le CRF fourni par les ARCs. Le but de cette relecture était de vérifier que tout ce qui apparaissait dans le CRF correspondait à ce qui était écrit dans le protocole et qu'il ne manquait pas d'informations. Cela afin de vérifier que les données qui seront utilisées pour les analyses statistiques sont bien présentes. Les commentaires et questions sont reportés dans un fichier excel pour la traçabilité et envoyé par la suite à l'ARC promoteur afin que celui-ci réponde aux différentes interrogations.

La seconde étape, une fois que le CRF est validé, est de commencer à rédiger le DMP ainsi que le DVP. La rédaction du DMP est assez rapide. En effet, une trame de celui-ci est disponible sur la plateforme de partage de documents de l'institut Bergonié. Ainsi, il suffit de l'adapter à l'étude. La rédaction du DVP, quant à elle, est une étape importante dans réalisation de l'étude. C'est dans celui-ci que tous les contrôles de cohérence concernant l'étude vont être définis. Dans le cadre de l'étude BIOPSAR, j'ai pu réfléchir et créer le DVP entièrement. Les contrôles sont divisés en deux types, les contrôles dits automatiques qui apparaissent à la saisie et les contrôles dits manuels qui seront programmés sous le logiciel SAS. Dans ces contrôles manuels, on distingue les contrôles programmés qui sortiront automatiquement les incohérences et les contrôles pas listing qui nécessiteront une vérification humaine. Pour créer ce document, il faut partir du protocole et du CRF et ainsi imaginer toutes les incohérences qui pourraient apparaître lors du remplissage de la base de données. Pour cette étude j'ai pu penser à plusieurs contrôles, comme vérifier la cohérence entre les dates ou bien vérifier si les critères d'inclusion sont bien respectés. Par exemple, on vérifie que les dates de suivis sont bien postérieures aux dates d'inclusion. On peut également mettre des bornes sur certaines variables pour éviter les valeurs aberrantes. Néanmoins, certains contrôles sont difficiles à mettre en place, comme par exemple un critère d'inclusion qui était divisé en 3 parties. Lorsque ce cas se présente on peut faire une vérification par listing, c'est-à-dire qu'on vérifie soit

même si le critère est bien vérifié pour chaque patient en sortant une liste des patients. Un exemple de page du DVP de BIOPSAR est disponible en figure 8, pour permettre de visualiser ce que représente ce document. Une fois tous les contrôles indiqués dans le DVP, celui-ci est validé par l'ARC promoteur, le statisticien, l'investigateur principal et le responsable de l'unité qui peuvent ainsi nous dire si certains contrôles sont à ajouter ou à modifier. Le fait de faire valider le DVP par plusieurs corps de métier permet d'établir un maximum de contrôles grâce à leurs expertises. Les ARCs promoteurs pourront ainsi apporter leur expertise sur l'étude, les médecins leur expertise en médecine, et les statisticiens peuvent vérifier si nos contrôles permettent une bonne vérification de l'éligibilité du patient. Par exemple, il est possible qu'il y ait une borne standard pour la taille des tumeurs à contrôler dans la base qui sera notifiées par l'ARC ou le médecin. Une fois tous les retours effectués le DVP est validé et il est ainsi possible de programmer les contrôles de cohérence sur le logiciel et sous SAS.

Concernant les contrôles automatiques :

Pour toutes les dates de l'étude: Date > date du jour. MESSAGE: Cette date est postérieure à la date du jour. Merci de vérifier et corriger.

Pour toutes les dates de traitement: Date de début > Date de fin. MESSAGE: La date de fin est antérieure à la date de début. Merci de vérifier et corriger.

Concernant les contrôles manuels :

Aucun contrôle manuel sur les données manquantes ne sera programmé sous SAS (excepté pour les variables relatives aux critères d'éligibilité), une liste des données manquantes sera éditée via MACRO 4.

VISITE	EFORM	QUESTION	INCOHERENCE	MESSAGE	TYPE
INCLUSION INC	Fiche d'inclusion INCFORM	N° Patiente NCENTER, NUMINC	Non unique	Le numéro de la patiente n'est pas unique. Merci de le corriger.	MANUEL
		Date de naissance DT_NAISS	<35 ans	Le critère d'inclusion n°1 n'est pas respecté : la patiente à moins de 35 ans. Merci de corriger ou clarifier.	MANUEL
		Date de signature du consentement DT_CONS	>120 ans <XX/XX/XXXX	Merçi de revoir l'âge de la patiente. La date de signature du consentement est antérieure à la date d'acceptation de l'ouverture réglementaire. Merçi de vérifier et corriger.	MANUEL AUTO
		La patiente est-elle en échec de screening ? FAIL	=Oui & Raison principale de la fin de l'étude ≠ Echec de screening (Fin d'étude) =Oui & Tous les critères d'éligibilité sont respectés = Oui	La patiente est en échec de screening mais la raison de fin d'étude est différente. Merçi de clarifier ou corriger. La patiente est en échec de screening mais tous les critères d'éligibilité sont respectés. Merçi de vérifier, clarifier ou corriger. Listing : Vérifier que les pages 'Critère d'éligibilité' et 'fin d'étude' sont complétées	MANUEL MANUEL
		Date d'échec de screening DT_SCR	<Date de signature du consentement >Date de fin d'étude	La date d'échec de screening est antérieure à la date de signature du consentement. Merçi de corriger ou de clarifier. La date d'échec de screening est postérieure à la date de fin d'étude. Merçi de corriger ou de clarifier.	AUTO MANUEL
		Date d'inclusion DT_INC	<Date de signature du consentement	La date d'inclusion est antérieure à la date de signature du consentement. Merçi de corriger ou de clarifier.	AUTO

Figure 8 : Exemple de première page du DVP de BIOPSAR

Enfin, d'autres documents doivent être initiés lors de la création de la base de données. Le CRF annoté référence le nom de toutes les variables ainsi que leurs codages et paramétrages sous le logiciel. Il sera par la suite utilisé par le statisticien pour programmer ses analyses. La fiche d'identification quant à elle est un document réunissant toutes les informations de base sur l'étude comme la méthodologie de référence utilisée, le type d'étude ou bien encore le nombre de patients attendus. Tous les documents sont validés via des formulaires à faire signer par les responsables de l'étude, comme le promoteur ou l'investigateur principal. La plupart des documents sont édités en amont de la création de la base mais il est également possible de les rédiger en parallèle de cette création.

Ainsi au cours de mes missions j'ai pu rédiger ses documents pour l'étude BIOPSAR, mais j'ai également pu les commencer pour l'étude IBMC portant sur la réduction du risque tabac/alcool chez les femmes traitées pour un cancer du sein avec implication du conjoint lors d'une intervention brève

motivationnelle. Une fois ces documents rédigés il était possible de passer à la création de la base de données.

### **3.1.2. Bases de données et des contrôles de cohérence**

#### *3.1.2.1. Création des bases de données*

Au sein de l'institut Bergonié il existe plusieurs logiciels permettant de concevoir les bases de données relatives aux essais cliniques. Ces logiciels sont respectivement MACRO et RedCap. RedCap, qui signifie Research Electronic Data Capture, est une solution logicielle d'acquisition électronique de données sur navigateur, pour la conception de bases de données de recherche clinique. Il est très utilisé dans le milieu de la recherche universitaire et le consortium RedCap est un réseau collaboratif et international de plus de 2400 partenaires. Bien qu'il soit gratuit pour les partenaires institutionnels, il n'est pas open source et est limité à des fins de recherche non commerciale. [45, 46] Les deux logiciels ayant globalement les mêmes fonctionnalités, MACRO va être utilisé généralement pour les grosses études menées par l'Institut, tandis que RedCap va être dédié aux études plus petites et généralement monocentriques ou bien aux thèses.

La création d'une base de données sous l'un de ces logiciels ne se passe pas comme lorsqu'on crée une base de données avec un logiciel tel qu'Access. En fait, lors de cette création on ne réfléchit pas à la création des différentes tables, à comment les relier les unes aux autres ni aux formes normales. Créer une base de données sous l'un de ces logiciels consiste à mettre en forme des formulaires avec des variables et ce sont ces formulaires qui constitueront par la suite les différentes tables de notre base de données. Ainsi à l'export de la base de données, celles-ci sont plus facile d'utilisation pour les statisticiens.

La première base de données que j'ai pu créer est la base de données concernant l'étude BIOPSAR sous le logiciel MACRO. L'objectif était ainsi de reproduire, à partir du CRF, les formulaires qui serviront à la saisie des ARCs investigateurs. Il faut choisir ainsi les noms des variables et leur format pour que cela soit le plus intuitif et facile d'utilisation possible. Ainsi la base de données doit être une copie conforme du CRF et les formulaires pourraient être un scan des pages de ce CRF. On met également en place des conditions de paramétrage sur certaines variables. Par exemple, on ne fera apparaître la question « Êtes-vous enceinte ou allaitante ? » que si le sexe du patient est féminin. Une fois l'intégralité des variables et conditions ont été créé, la base doit être testée par le Data Manager ainsi que l'ARC promoteur pour vérifier que tout est conforme et que la base est utilisable pour l'étude. Pour cela on va créer de faux patients dans une version test de la base et ainsi s'assurer que toutes les conditions fonctionnent correctement et que toutes les variables sont au bon format.

En parallèle de la création de la base de données les contrôles de cohérence automatiques sont paramétrés dans le logiciel pour assurer la fiabilité et la validité des données saisies.

#### *3.1.2.2. Programmation des contrôles de cohérence*

La programmation des contrôles de cohérence est une étape importante pour la fiabilité des données recueillis lors d'un essai clinique. En effet, ces contrôles permettent de s'assurer qu'il n'y

pas d'erreur lors de la saisie et donc permet de s'assurer que les données sont celles qui ont été renseignées dans le cadre de l'étude. Cette fiabilité permet donc de garantir des résultats statistiques justes et qu'il n'y a pas de problème de qualité et de biais au niveau de la collecte des données.

Pour programmer les contrôles dans la base de données on part du DVP précédemment créé. Dans ce document les contrôles ont été divisés en deux catégories, les contrôles manuelles et automatiques, voir figure 8 précédemment. Cette division détermine si les contrôles vont être programmé directement sur le logiciel, MACRO ou RedCap, ou bien programmé sous SAS.

La programmation des contrôles automatiques diffère légèrement entre MACRO et RedCap mais reste néanmoins très similaire. La différence majeure réside dans le fait qu'il est possible de faire des fonctions mathématiques légèrement plus poussés sous RedCap. Mais cela n'entraîne pas de distinction majeure entre les deux logiciels. Pour programmer ces contrôles on va donc écrire le message d'alerte préalablement rédiger dans le DVP et implémenter la condition pour laquelle ce message doit apparaître. Par exemple, pour le message « La date de la visite est antérieure à la date de signature du consentement. Merci de corriger ou confirmer. » nous allons implémenter la condition « Date d'inclusion < Date de signature du consentement », voir figure 9 pour cet exemple programmé sur MACRO (on peut également voir un exemple similaire programmé sous RedCap dans la figure 10). Ainsi, lors du remplissage, si la date d'inclusion est antérieure à la date de signature du consentement le message d'alerte apparaîtra et l'investigateur pourra donc vérifier la date et la corriger si nécessaire. En effet, dans certains cas un message peut apparaître mais la valeur, même si elle est aberrante, est correcte et dans ce cas la donnée peut être confirmée. Ce genre de cas apparaît généralement pour les bornes et les intervalles. Dans la plupart des cas les variables numériques sont bornées comme pour le poids que l'on va borner entre 50 et 150 kilogrammes pour un adulte. Cependant, il est possible qu'un adulte fasse un poids en dehors de ces bornes et dans ce cas la donnée peut être confirmée. On va distinguer plusieurs types de contrôles, comme par exemple les contrôles sur des dates qui consistent à comparer deux dates entre elles pour s'assurer de leur chronologie ou bien du respect des délais conformément au protocole, les intervalles pour les variables numériques ou bien encore les contrôles par rapport aux critères d'inclusion lorsque certaines réponses doivent être oui ou non pour que le patient soit éligible.

Validation Type	Validation Term	Validation Message
	Warn if me:value isafter datenow	'La date saisie est supérieure à la date du jour'
	Warn if visit = INC and me:value isbefore INC:INCFORM:DT_CONS	'La date de la visite est antérieure à la date de signature du consentement. Merci de corriger ou confirmer.'
	Warn if visit = 'CPB' and me:value isbefore BIO:BIOPFORM:DT_BIOP	'La date de la consultation post-biopsie est antérieure à la date de biopsie. Merci de clarifier ou corriger.'
	Warn if visit = 'CPC' and me:value isbefore CHIR:CHIRFORM:DT_CHIR	'La date de la visite est antérieure à la date de chirurgie. Merci de clarifier ou corriger.'

Figure 9 : Exemple de contrôle automatique sous MACRO

11	Le critère de non inclusion 5 est coché "Oui" ou est manquant, le participant est donc inéligible. Merci de corriger ou confirmer.	[c5ni] = "" or [c5ni] = "1"		<input type="button" value="Execute"/>	
12	La date de remplissage du questionnaire est antérieure à la date d'inclusion. Merci de vérifier et corriger.	[dt_quest] <> "" and [dt_inc] <> "" and [dt_quest] < [dt_inc]		<input type="button" value="Execute"/>	
13	La date de remplissage du questionnaire d'inclusion est plus de 15 jours après l'inclusion. Merci de vérifier et corriger.	[dt_inc] <> "" and [dt_quest] <> "" and datediff([dt_inc],[dt_quest],"d",true) > 15		<input type="button" value="Execute"/>	

Figure 10 : Exemple de contrôle automatique sous RedCap

Une fois les contrôles automatiques intégrés au logiciel, les contrôles manuels vont être programmés dans SAS. On va également vérifier les données manquantes et faire sortir les différents listings pour les vérifications manuelles. Le modèle type de ce programme SAS est un fichier disponible sur l'espace de stockage partagé de l'institut. Ce modèle permet l'importation des données et la vérification des données manquantes. Ensuite, les autres contrôles sont programmés soit sous forme de macro SAS soit à l'aide de procédure PROC PRINT ou PROC SQL. Les contrôles ainsi programmés vont également être, par exemple, des contrôles sur les dates pour vérifier la chronologie entre deux dates de formulaire différent. Mais ceux-ci peuvent également être pour des calculs de délai entre deux dates. Par exemple, pour l'étude BIOPSAR on vérifie que le délai entre l'inclusion et la biopsie est bien inférieure à 30 jours. On réalise également d'autres calculs mathématiques comme la vérification de l'âge des patients. Une fois le programme SAS terminé celui-ci est testé grâce à des données fictives pour s'assurer que tous les contrôles fonctionnent correctement et que celui-ci pourra être utilisé par la suite pour l'étude.

Ainsi la programmation de ces contrôles de cohérence permet de s'assurer de la qualité des données saisies dans la base de données. Les données saisies font faire l'objet de contrôles automatiques en temps réels qui s'afficheront directement pour les investigateurs, tandis que les contrôles manuelles via programme SAS seront lancés lors des validations des données.

### **3.1.3. Validation des données et autres missions**

#### *3.1.3.1. Validation des données*

La validation des données, tout comme la programmation des contrôles, est également un élément clé pour assurer la fiabilité et la qualité des données. Ces validations sont faites régulièrement au cours des différentes études et constitue un travail important du Data Manager. Au cours de mon alternance, j'ai pu réaliser la validation des données de l'étude METZOLIMOS. Cette étude a pour but d'évaluer la dose recommandée pour la phase 2, ainsi que la dose maximale tolérée, la toxicité et la toxicité dose-limitante de l'association entre plusieurs médicaments pour les patients prétraités pour une tumeur solide avec métastases osseuses ou un ostéosarcome avancé.

La validation des données consiste en une vérification de la qualité des données au cours de l'étude et à la fin de la saisie des données. En effet, en fonction de la durée de l'étude il est possible de faire plusieurs validations. Une validation est effectuée à chaque analyse intermédiaire et lorsque cela est demandé. Pour les études classées RIPH 1, une validation à l'année pendant les inclusions doit être effectuée et constitue un indicateur pour le data management pour la certification ISO de l'institut. Lors de cette validation, une extraction des données est demandée à l'informaticien afin de pouvoir lancer le programme SAS sur les données. Ce programme va ainsi ressortir toutes les incohérences relevées dans la base de données. Une fois tous les contrôles réalisés, on vérifie les différents listings, s'il y en a, afin de ressortir les dernières incohérences.

Lorsque toutes les incohérences ont été identifiées, on peut alors poser des demandes de clarification aux investigateurs afin qu'ils puissent par la suite corriger éventuellement les données. En effet, même si l'erreur paraît évidente il est impossible pour le Data Manager de corriger lui-même les données. Il est simplement possible de demander une clarification mais celle-ci ne doit jamais être orientée la réponse de l'investigateur. On ne peut pas écrire, je pense que c'est une erreur et qu'il faut mettre telle information. Il faut que la demande soit simplement un relevé de



l'incohérence et une demande de vérification avec correction ou confirmation. Il est nécessaire pour l'investigateur de répondre à toutes les demandes et chaque modification est par la suite enregistrée dans l'audit trail. L'audit trail est un suivi des modifications qui enregistre quelle personne a modifié les données, à quel moment et pourquoi, afin d'assurer une traçabilité, la qualité et la sécurité des données. Un fichier d'échange est également créé pour noter les différentes questions ou demandes de confirmation à poser aux ARCs promoteurs pour éviter des demandes de clarification inutiles.

Une fois la validation effectuée et toutes les demandes de clarification fermées par les investigateurs, toutes les données sont enregistrées et il est possible de continuer le bon déroulement de l'étude. Si une incohérence apparaît pour la plupart des patients, celle-ci est remontée à l'ARC promoteur afin de déterminer si ce contrôle est pertinent et pour voir s'il est possible de faire des modifications substantielles du protocole pour que cette erreur ne soit plus considérée comme une incohérence. Cela n'est à ce jour jamais arrivé à l'institut mais reste une possibilité. Lorsque cela arrive le DVP et les contrôles sont ainsi modifiés et une nouvelle version est créée pour la prochaine validation. Lorsqu'il s'agit de la validation finale des données, une fois les demandes de clarification fermées et les différents contrôles nécessaires au lock de la base de données effectués (vérification du codage des événements indésirables, réconciliation de base, plus aucune donnée manquante ...), les données sont extraites et le statisticien peut commencer à travailler sur son rapport.

### *3.1.3.2. Autres missions du Data Manager*

Au cours d'une étude, le Data Manager se voit également confier de nombreuses autres missions. Je n'ai malheureusement pu réaliser aucune de ces différentes missions au cours de mon alternance. Ces missions peuvent être des missions de back-up ou bien des tâches de Data Management classique. En effet, lorsqu'une étude implique une randomisation des patients ou bien la création d'une base de pharmacovigilance pour les événements indésirables graves, il est nécessaire de faire une réconciliation entre les bases dites cliniques et les bases de randomisation ou les bases PV. Cette réconciliation consiste à relier le patient de la base de données clinique à son homonyme dans les autres bases de données afin de pouvoir traiter les données et vérifier que celles-ci sont bien identiques entre les bases.

Il y a d'autres missions qui peuvent être confiées au Data Manager lorsque celles-ci ne peuvent être réalisées par la personne normalement en charge. Par exemple, si le statisticien n'est pas en mesure d'effectuer la randomisation, cette tâche peut être confiée au Data Manager. Certains Data Manager se voient également confier des missions de back-up. Notamment des missions de back-up informatique afin d'assurer, lorsque l'informaticien ne le peut pas, la mise en production des bases de données ou bien encore l'attribution des droits sur les différentes études. Ces missions de back-up permettent d'assurer une permanence. Le Data Manager est également convié aux différentes réunions au tour de la mise en place et du suivi des études afin de donner son avis sur les différents aspects liés au Data Management de l'étude.

Parmi ces autres missions on peut également retrouver son implication dans la VSI au sein de l'institut Bergonié. Cette implication se retrouve notamment dans sa connaissance des différents logiciels, mais aussi dans son habitude à rédiger différents documents et à suivre et gérer les données. Cette implication est explicitée précédemment dans la section 2.4.

## **3.2. Apport personnel de ces différentes missions et discussion autour du Data Management**

### **3.2.1. Apport personnel des différentes missions**

D'un point de vue personnel ces différentes missions m'ont permis d'en apprendre beaucoup sur le métier de Data Manager et sur les différentes notions qui y sont associés. Dans un premier temps j'ai pu être formé aux différentes tâches quotidiennes du Data Manager, en passant tout d'abord par les différentes procédures et documents existants. J'ai pu apprendre à rédiger ces documents et à réfléchir à comment les concevoir. Par exemple, la rédaction du DVP m'a permis de réfléchir à la création des différents contrôles de cohérence ce qui développe les capacités de réflexion et d'analyse. Ensuite, les différents protocoles que j'ai pu lire pour les études sur lesquelles j'ai travaillé m'ont beaucoup appris sur différents domaines de la cancérologie, ce qui m'a apporté de nombreuses connaissances pour ma culture générale et pour la compréhension d'autres protocoles. J'ai pu observer le fonctionnement d'un essai clinique des phases de réflexion jusqu'à la mise en place de celui-ci. J'ai également pu observer le fonctionnement de la structure en général et ainsi comprendre l'organisation de celle-ci et potentiellement des autres structures existantes.

Dans un second temps, ces missions m'ont également permis de développer de nombreuses qualités. Tout d'abord le métier de Data Manager demande de la rigueur et de l'organisation. Il faut ainsi savoir planifier ses tâches et organiser son temps de travail pour respecter les délais. Ensuite, le travail d'équipe est un élément central de ce métier. En effet, nous sommes amenés à travailler avec l'ensemble des corps de métier de l'UREC que ce soit les ARCs ou bien les statisticiens. Mais également les autres Data Manager avec lesquels nous nous entraînons sur chaque étude dès qu'il y a besoin. Cela permet également d'apprendre à se faire comprendre de tous car les procédures sont parfois complexes et il faut savoir les expliquer quand cela s'avère nécessaire.

Enfin la VSI, m'a permis de travailler sur un domaine nouveau. J'ai pu ainsi me rapprocher d'un domaine qui peut faire penser à celui de la recherche. J'ai pu me documenter sur les différentes notions liées à la VSI et ainsi travailler sur la méthodologie que ce soit via la rédaction des différentes procédures ou bien via la transmission des connaissances sur le sujet au reste de l'équipe. La VSI m'a également permis de gérer ce projet pour mon alternance comme les projets qu'on a pu voir au cours de l'année scolaire.

Ainsi, je suis très reconnaissant envers l'ensemble de l'équipe de l'UREC et l'ensemble des Data Manager qui m'ont permis d'apprendre ce métier et de découvrir le Data Management dans le domaine de la santé. Et même si ces compétences s'appliquent au domaine de la santé, il est possible de les extrapoler à d'autres domaines, car la création de base de données et la programmation de contrôle de cohérence pour assurer la fiabilité et la sécurité des données restent des éléments essentiels dans tous métiers de Data Management.

### 3.2.2. Comparaison entre les logiciels de Data Management et les SGBD classiques

Mes différentes missions m'ont également permis de réfléchir sur le Data Management et notamment sur la différence qu'il peut exister entre les logiciels utilisés à l'institut Bergonié et ce qui est vu dans le milieu scolaire. Ainsi, j'ai pu soulever une discussion qui peut sembler intéressante suite à la réalisation de ces missions en tant que Data Manager. Celle-ci réside dans la comparaison entre les logiciels utilisés à l'institut Bergonié pour le Data Management et les Systèmes de Gestion de Base de Données (SGBD) plus classiques que j'ai pu voir au cours de mes études, et plus particulièrement de mon année de Master 2.

L'une des différences majeures entre ces deux systèmes réside dans la façon de concevoir les bases de données. En effet, dans les logiciels de Data Management d'essais cliniques la base de données est pensée comme un formulaire de remplissage pour rentrer les données dans la base. La base est ainsi créée à partir d'un CRF qui doit être suivi à la lettre, et ainsi chaque formulaire constituera une table de notre base de données. En revanche, dans les SGBD traditionnels les bases de données sont pensées pour être optimisées et pour que les requêtes soient faisables le plus facilement possible. La notion de forme normale par exemple n'apparaît pas dans les logiciels de Data Management, ni la notion de table de relation. Ces notions n'apparaissent pas car dans les logiciels de Data Management les données ont pour but d'être stockées puis extraites pour analyse statistique. Une fois que l'étude est terminée, il n'y a plus aucune modification qui va être faite sur la base, sauf cas exceptionnel par exemple lorsqu'un évènement indésirable grave ce déclare. Pour les SGBD, les bases de données ont pour vocation à être toujours en constante évolution, les données doivent être dynamiques et ont pour but d'être réutilisées et parfois modifiées. Ainsi entre les deux systèmes la conception d'une base de données changent énormément, même si dans les deux cas l'objectif reste le même c'est-à-dire, stocker des données afin de les réutiliser dans un but précis. On peut noter par exemple que dans les logiciels de Data Management il n'y a aucune interface pour interagir avec les données via des requêtes SQL, comme on peut en trouver dans les SGBD. On peut dire que les SGBD sont plus permissifs que les logiciels utilisés en Data Management d'essais cliniques.

Une seconde différence importante entre ces deux systèmes est également dans l'importation des données dans les bases. Pour les SGBD classiques, la base de données va généralement être alimentée par des fichiers externes. Les données peuvent être constamment modifiées et mise à jour si nécessaire et parfois même supprimées. Dans un logiciel de Data Management d'essais cliniques, une partie de ce logiciel est prévue pour le remplissage de la base. Ainsi chaque patient va être rentré individuellement dans la base à partir des formulaires. Dans ces logiciels, on ne supprime jamais les données d'un patient et chaque modification est tracée (via l'audit trail). Les données des patients ne sont pas voué à évoluer contrairement aux données dynamiques enregistrées dans les SGBD classiques. Les données des patients représentent une image à un moment précis et c'est pourquoi il n'y a pas de notion de modification à moins qu'une erreur de saisie ait été effectuée.

Il est intéressant de voir qu'un domaine comme le Data Management peut être vu de façon très différente en fonction du milieu dans lequel il est utilisé. En effet, dans le domaine de la santé et

plus particulièrement des essais cliniques, les bases de données ont une fonction de stockage afin de représenter la réalité à un moment donnée dans le but d'effectuer une analyse statistique sur un sujet précis. Dans des domaines tel que le commerce, les bases de données vont un voir un but de gestion des clients, des commandes et des stocks par exemple. Les bases vont ainsi être en constante évolution avec possiblement des changements dans les données des clients, l'ajout de nouvelles commandes au fil du temps et l'évolution des stocks. Les données sont dynamiques et reflètent la réalité en temps réel. C'est le domaine qui va donc déterminer les besoins auxquels le logiciel de Data Management va devoir répondre. C'est pour ça qu'on observe tant de différence entre les logiciels de Data Management d'essais cliniques et les SGBD plus classiques.

## **4. Discussion, limites et conclusion**

### **4.1. Discussion et limites**

#### **4.1.1. Discussion générale**

Mon travail au cours de cette année d'alternance m'a permis d'apprendre beaucoup sur le métier de Data Manager. J'ai pu ainsi constater les différences qui peuvent exister entre le Data Management tel que nous l'apprenons en cours et le Data Management tel qu'il se passe réellement au sein d'un institut gérant des essais cliniques. On peut ainsi constater une grande différence entre les cours et la pratique, que ce soit dans les logiciels utilisés ou bien dans les méthodes de création des bases de données. J'ai eu la chance de bénéficier d'une excellente formation au métier de Data Manager au sein de l'institut Bergonié. Cette formation m'a permis d'apprendre les bases du métier et m'a apporter de nombreuses connaissances autour du domaine des essais cliniques.

En ce qui concerne la VSI, la nouveauté du domaine constitue une limite importante de ce projet. En effet, la compréhension de toutes les notions liées à cette VSI était parfois complexe et impliquait de se documenter autour de nombreux sujets divers. Que ce soit pour comprendre ce qui était attendu ou bien pour mettre en œuvre la VSI, en faisant par exemple l'analyse de risque. Je n'avais également jamais eu de formation méthodologique à la création de procédures et de documents, j'ai ainsi dû l'apprendre au cours de mon alternance en parallèle de mes autres tâches. Une autre limite liée à la VSI réside dans la durée de mise en place de celle-ci. De par sa nouveauté il est difficile d'estimer combien de temps est nécessaire pour valider un système informatisé. Il y a également un grand besoin de formation autour du sujet pour l'ensemble des services pour une implication plus globale.

#### **4.1.2. Impact de l'épidémie de Covid-19 sur mon alternance**

L'épidémie de Covid-19 a également eu un impact sur mon alternance. Ainsi j'ai été en télétravail dès le début du confinement et ce pendant presque 4 mois. J'ai malgré tout eu la chance de bénéficier d'un excellent suivi tout au long du confinement. J'avais accès à un VPN me permettant de travailler à distance. J'ai donc pu assurer mes différentes tâches de Data Manager, tout en continuant à travailler sur les différents aspects de la VSI. Néanmoins, avec l'épidémie les études étaient pour la plupart ralenties et l'activité réduite. Une des difficultés majeures du confinement résidait dans le manque de place pour travailler à mon domicile, en effet nous vivions à deux dans un petit appartement. Ceci a impacté ma productivité et rendait parfois difficile le travail. Je m'estime cependant chanceux d'avoir pu être si bien suivi et d'avoir pu continuer à travailler malgré les difficultés liées à cet épidémie.

## 4.2. Conclusion personnelle

L'implication du Data Manager dans la VSI se retrouve dans de nombreux points. De par sa capacité à rédiger des documents et sa rigueur acquise grâce à ses tâches quotidiennes, le point de vue du Data Manager est important pour mettre en place une procédure complexe comme la VSI. Les recherches autour du sujet, la compréhension des différents documents fournis par le GIRCI, mais également la mise en place des tests sont tant d'aspect sur lesquels le Data Manager a pu s'avérer utile. Le métier de Data Manager est un métier en lien avec l'informatique ce qui peut aussi faciliter l'appréhension d'une mission tel que la VSI. Ainsi, le premier logiciel pour lequel la VSI a été mise en place à l'institut Bergonié est un logiciel de Data Management. Et même si celle-ci est toujours en cours et qu'il y a encore beaucoup à faire, la VSI est en bonne voie d'être réalisée au sein de la structure. Il faut encore réfléchir sur plusieurs sujets comme celui de faire appel à un prestataire externe ou bien encore la formation des autres corps de métiers et du personnel à cette VSI. Néanmoins tout le travail accompli jusqu'ici permet de poser les bases et d'avoir toutes les clés en main pour mettre en œuvre la VSI et instaurer ces différentes réflexions.

Pour conclure, on peut dire que cette alternance m'a beaucoup apportée sur le plan personnel et professionnel. J'ai ainsi pu découvrir le métier de Data Manager au sein de l'institut Bergonié. J'ai pu apprendre en partie les ficelles du métier, que soit via la création des bases de données ou via la rédaction des documents de Data Management. Je pense ainsi avoir trouvé un métier qui me plait et vers lequel j'ai envie de continuer à m'investir. L'alternance et le télétravail m'ont permis de développer de nombreuses qualités, notamment en matière d'organisation et de rigueur principalement. La VSI d'autre part m'a permis de rester encore d'une certaine façon dans le domaine de la recherche, me rappelant ainsi le stage que j'ai pu faire à l'INSERM lors de mon année de Master 1. En effet, j'ai dû beaucoup me documenter et cela m'a permis de faire la transition entre les cours et le milieu professionnel. Grâce à cette alternance j'ai la chance d'avoir trouvé un CDI dans une autre structure que l'institut Bergonié où je vais continuer à faire du Data Management mais dans un autre domaine. Je vais donc exercer le métier de Data Manager, je l'espère pour longtemps, et ce grâce à cette alternance et à ma formation.

## Bibliographie

- [1] Unicancer. (page consultée le 15/04/2020). Qui sommes-nous, [en ligne]. <http://www.unicancer.fr/unicancer/le-groupe>
- [2] Hopital.fr. (page consultée le 15/04/2020). Centre Régional de Lutte Contre le Cancer Institut Bergonié (Bordeaux), [en ligne]. <https://www.hopital.fr/annuaire-etablissement/centre-regional-de-lutte-contre-le-cancer-institut-bergonie-bordeaux,6531>
- [3] Bergonié. (page consultée le 15/04/2020). Histoire, [en ligne]. <https://www.bergonie.fr/histoire/>
- [4] European Scientist. (page consultée le 15/04/2020). L'autre invention de Marie Curie : « Le modèle Curie », [en ligne]. <https://www.europeanscientist.com/fr/opinion/lautre-invention-de-marie-curie-le-modele-curie/>
- [5] Unicancer. (page consultée le 15/04/2020). Bordeaux, Institut Bergonié, [en ligne]. <http://www.unicancer.fr/centre/institut-bergonie-bordeaux>
- [6] UREC-DCO-0001 : Organisation générale et gouvernance de la Recherche clinique à l'institut Bergonié (Document interne à Bergonié)
- [7] Recherche clinique Paris Centre. (page consultée le 15/04/2020). Les différents intervenants dans la recherche clinique et Relation entre ces différents acteurs, [en ligne]. [https://www.recherchecliniquepariscentre.fr/wp-content/uploads/2017/11/Cours-Communs-Les-intervenants-de-la-RC\\_03-11-2017\\_SMF.pdf](https://www.recherchecliniquepariscentre.fr/wp-content/uploads/2017/11/Cours-Communs-Les-intervenants-de-la-RC_03-11-2017_SMF.pdf)
- [8] UREC-FITC-0001 : Organisation opérationnelle de la recherche clinique (Document interne à Bergonié)
- [9] Bergonié. (page consultée le 15/04/2020). L'ensemble de la recherche clinique du réseau unicancer certifié ISO 9001, [en ligne]. <https://www.bergonie.fr/lensemble-de-recherche-clinique-reseau-unicancer-certifie-iso-9001/>
- [10] GIRCI-SOHO. (page consultée le 15/04/2020). Historiques et Missions des GIRCI, [en ligne]. <http://www.girci-soho.fr/content/historique-et-missions-des-girci>
- [11] Bergonié. (page consultée le 15/04/2020). Organisation recherche, [en ligne]. <https://www.bergonie.fr/fonctionnement-et-organisation/organisation-recherche/>
- [12] ANSM. (page consultée le 15/04/2020). Qu'est-ce qu'un essai clinique ?, [en ligne]. [https://www.ansm.sante.fr/Activites/Essais-cliniques/Qu-est-ce-qu-un-essai-clinique/\(offset\)/4](https://www.ansm.sante.fr/Activites/Essais-cliniques/Qu-est-ce-qu-un-essai-clinique/(offset)/4)
- [13] Institut National du Cancer. (page consultée le 15/04/2020). Qu'est-ce qu'un essai clinique ?, [en ligne]. <https://www.e-cancer.fr/Comprendre-prevenir-depister/Comprendre-la-recherche/A-quoi-sert-la-recherche/Les-essais-cliniques>
- [14] Institut Curie. (page consultée le 15/04/2020). Qu'est-ce qu'un essai clinique ?, [en ligne]. <https://curie.fr/page/quest-ce-quun-essai-clinique>

[15] La ligue contre le cancer. (page consultée le 15/04/2020). Les essais cliniques et le comité de patients. Les différentes phases des essais cliniques, [en ligne]. [https://www.ligue-cancer.net/article/37840\\_les-differentes-phases-des-essais-cliniques](https://www.ligue-cancer.net/article/37840_les-differentes-phases-des-essais-cliniques)

[16] ANSM. (page consultée le 15/04/2020). Essais cliniques RIPH2 et RIPH3, [en ligne]. [https://www.ansm.sante.fr/Activites/Essais-cliniques/Essais-cliniques-RIPH2-et-RIPH3/\(offset\)/9](https://www.ansm.sante.fr/Activites/Essais-cliniques/Essais-cliniques-RIPH2-et-RIPH3/(offset)/9)

[17] Les entreprises du médicament. (page consultée le 15/04/2020). Les essais cliniques en 12 questions, [en ligne]. <https://www.leem.org/sites/default/files/questions%20sur%20les%20essais%20cliniquesVF.pdf>

[18] Pharmacomédicale.org. (page consultée le 15/04/2020). Développement et suivi des médicaments. Aspects législatifs des essais cliniques, [en ligne]. <https://pharmacomedicale.org/pharmacologie/developpement-et-suivi-des-medicaments/26-aspects-legislatifs-des-essais-cliniques>

[19] Les entreprises du médicament. (page consultée le 15/04/2020). Les essais cliniques en 10 questions, [en ligne]. [https://www.leem.org/sites/default/files/2018-12/10questions%20sur%20les%20essais%20cliniques\\_27nov2018%20revu%20CPH.pdf](https://www.leem.org/sites/default/files/2018-12/10questions%20sur%20les%20essais%20cliniques_27nov2018%20revu%20CPH.pdf)

[20] ANSM. (page consultée le 15/04/2020). Textes en vigueur, [en ligne] [https://www.ansm.sante.fr/Activites/Essais-cliniques/Reglementation/\(offset\)/8](https://www.ansm.sante.fr/Activites/Essais-cliniques/Reglementation/(offset)/8)

[21] (page consultée le 15/04/2020). Les bonnes pratiques cliniques : avis aux promoteurs et aux investigateurs, [en ligne]. [https://www.inserm.fr/sites/default/files/2017-09/ANSM\\_BonnesPratiquesCliniques\\_RecherchePersonneHumaine.pdf](https://www.inserm.fr/sites/default/files/2017-09/ANSM_BonnesPratiquesCliniques_RecherchePersonneHumaine.pdf)

[22] Oracle. (page consultée le 15/04/2020). Principes fondamentaux du Data Management, [en ligne]. <https://www.oracle.com/fr/database/definition-data-management.html>

[23] Paris School of Business. (page consultée le 15/04/2020). Le data management, c'est quoi ?, [en ligne]. <https://www.psb.edu.paris/fr/actus/data-management-defenition>

[24] UREC-PROC-0007 : Définition de la gestion des données (Data Management) (Procédure interne à Bergonié)

[25] UREC-FIFO-0003 : Data Manager, gestionnaire de données pour des études et essais, pôle Promotion (Document interne à Bergonié)

[26] UREC-PROC-0042 : Elaboration d'un e-CRF (Document interne à Bergonié)

[27] UREC-PROC-0041 : Plan de Data Management (Document interne à Bergonié)

[28] UREC-FITC-0004 : Plan de validation des données (DVP) (Document interne à Bergonié)

[29] UREC-PROC-0011 : Définition des contrôles informatisés des données (Document interne à Bergonié)

[30] Wikipédia. (page consulté le 15/04/2020). Validation, [en ligne]. <https://fr.wikipedia.org/wiki/Validation>



- [31] Wikipédia. (page consulté le 15/04/2020). Analyse des risques, [en ligne]. [https://fr.wikipedia.org/wiki/Analyse\\_des\\_risques](https://fr.wikipedia.org/wiki/Analyse_des_risques)
- [32] Piloter.org. (page consultée le 15/04/2020). Méthode d'analyse des risques, [en ligne]. <https://www.piloter.org/asoncompte/outils-entreprise/methode-analyse-des-risques.htm>
- [33] (page consultée le 15/04/2020). Le processus d'analyse et d'évaluation des risques, [en ligne]. [http://gpp.oig.qc.ca/le\\_processus\\_d\\_analyse\\_et\\_d\\_evaluation\\_des\\_risques.htm](http://gpp.oig.qc.ca/le_processus_d_analyse_et_d_evaluation_des_risques.htm)
- [34] Yan C, Chen F, Xia JL, Zheng QS, Liu D. Computerized system validation of clinical researches. Yao Xue Xue Bao, 2015 Nov ;50(11) : 1380-7.  
<https://www.ncbi.nlm.nih.gov/pubmed/26911028>
- [35] Deltenre Consulting Sprl. (page consultée le 15/04/2020). Qu'est-ce que la validation de systèmes informatisés ?, [en ligne]. <https://deltenre-consulting.com/index.php/2018/05/02/quest-ce-que-la-validation-de-systemes-informatises/>
- [36] GetReskilled. (page consultée le 15/04/2020). What is Computer System Validation (CSV) ?, [en ligne]. <https://www.getreskilled.com/what-is-computer-systems-validation-csv/>
- [37] CNIL. (page consultée le 15/04/2020). RGPD : de quoi parle-t-on ?, [en ligne]. <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>
- [38] Marovic B, Curcin V. Impact of the European General Data Protection Regulation (GDPR) on Health Data Management in a European Union Candidate Country : A case Study of Serbia. JMIR Med Inform. 2020 Apr 17 ;8(4) :e14604.  
<https://pubmed.ncbi.nlm.nih.gov/32301736/>
- [39] CNIL. (page consultée le 15/04/2020). RGPD : se préparer en 6 étapes, [en ligne]. <https://www.cnil.fr/fr/principes-cles/rgpd-se-preparer-en-6-etapes>
- [40] Shin Y, Kim KW, Lee AJ, Sung YS, Ahn S, Koo JH, Choi CG, Ko Y, Kim HS, Park SH. A good Practice-Compliant Clinical Trial Imaging Management System for Multicenter Clinical Trials : Development and Validation Study. JMIR Med Inform. 2019 AUG 30 ;7(3) :e14310.  
<https://www.ncbi.nlm.nih.gov/pubmed/31471962>
- [41] ISPE. (page consultée le 15/04/2020). GAMP 5 Guide : Compliant GxP Computerized Systems, [en ligne]. <https://ispe.org/publications/guidance-documents/gamp-5>
- [42] CVO-Europe. (page consultée le 15/04/2020). GAMP 5 : Utilisation pragmatique dans la mise en conformité réglementaire, [en ligne]. <https://www.cvo-europe.com/formation-cvo-europe/gamp-5-utilisation-pragmatique-dans-la-mise-en-conformite-reglementaire>
- [43] Labware. (page consultée le 15/04/2020). Classification des systèmes informatisés selon les besoins, [en ligne]. [https://a3p.org/wp-content/uploads/2010/09/article\\_scientifique\\_vague30\\_0pdf\\_articles\\_30pdf4.pdf](https://a3p.org/wp-content/uploads/2010/09/article_scientifique_vague30_0pdf_articles_30pdf4.pdf)

[44] Elsevier. (page consultée le 25/06/2020). MACRO Ensemble de données avancées pour la recherche clinique, [en ligne]. <https://www.elsevier.com/fr-fr/solutions/macro>

[45] REDCap. (page consultée le 25/06/2020). About, [en ligne]. <https://projectredcap.org/about/>

[46] Wikipédia. (page consultée le 25/06/2020). REDCap (Research Electronic Data Capture), [en ligne]. [https://fr.wikipedia.org/wiki/REDCap\\_\(Research\\_Electronic\\_Data\\_Capture\)](https://fr.wikipedia.org/wiki/REDCap_(Research_Electronic_Data_Capture))

## **Table des annexes**

Annexe 1 : UREC-DCO-007 : Organigramme de la recherche clinique (Document interne à Bergonié)

Annexe 2 : UREC-DCO-0058 : Politique de VSI (Document interne à Bergonié)





### Politique de validation des systèmes informatisés (VSI) utilisés dans le cadre des activités de recherche clinique de l'Institut Bergonié

L'Institut Bergonié utilise pour ses activités de recherche clinique des systèmes informatisés dans le cadre des projets impliquant la personne humaine. Dans ce contexte, la direction de l'Institut Bergonié approuve la mise en place d'une démarche de validation et de documentation de ces systèmes afin de garantir l'intégrité des données traitées.

Cette validation est en conformité avec les exigences réglementaires applicables (telles que les BPC, le RGPD, CFR 21 part 11) et selon les guides et recommandations méthodologiques (GAMP 5, PIC/S).

Pour cela, l'Institut Bergonié s'engage à :

- Valider tout nouveau système informatisé prospectivement avant sa mise en application (si cela est applicable) ;
- Valider toute mise à jour d'un système informatisé déjà en production ;
- Décrire, planifier, exécuter, et documenter les opérations de validation au moyen de personnel qualifié et formé ;
- Mettre en place des audits de la validation des systèmes informatisés au moyen de personnel habilité et compétent ;
- Former le personnel aux aspects inhérents à la validation des systèmes informatisés ;
- Mener des revues périodiques des systèmes afin de s'assurer de leur bonne adéquation par rapport aux besoins.

L'analyse des besoins et la coordination des différentes validations sont pilotées par la Direction de la Recherche et l'Innovation, en lien avec les fournisseurs, les utilisateurs ou les experts métiers des logiciels et la Direction du Système d'Information lorsqu'elle est impliquée.

Dans ce contexte, l'Institut Bergonié s'associe au GIRCI SOHO pour mener à bien l'ensemble des étapes nécessaires à la validation des systèmes informatisés et impliquer l'ensemble des personnels concernés dans cette démarche.

Pour l'établissement  
Directeur(s) de la DRCI

Date : 13/09/2019

Signature

UREC-DCO-0058\_V001

## Résumé

Ce rapport d'alternance traite de l'implication d'un Data Manager dans la mise en place de la Validation du Système Informatisé (VSI) de l'institut Bergonié. Dans un premier temps, ce rapport contextualise cette mission avec une description du métier de Data Manager et une explication appuyer sur plusieurs recherches de ce qu'est la VSI. Dans un second temps, ce rapport traite des différentes missions menées durant mon alternance à l'institut Bergonié en tant que Data Manager. Enfin ce rapport discute des différences qu'il peut y avoir entre le Data Management d'essais cliniques et les SGBD classiques, tout en abordant les conclusions liées à la VSI. La mise en place de celle-ci est une étape longue et complexe qui demande beaucoup de réflexion. Elle implique la création d'une équipe dédiée avec plusieurs experts métiers afin de mettre en place les meilleures procédures possibles applicables à tous les services. Pour conclure, ce rapport parle des différents apports personnels de cette alternance ainsi que de l'impact de l'épidémie de Covid-19 sur celle-ci.

**Mots-clés** : VSI, Validation des Systèmes Informatisés, Data Management, Data Manager

## Abstract

This work-study report deals with the involvement of a Data Manager in the implementation of the Computerized System Validation (CSV) of the Bergonié Institute. Firstly, this report contextualizes this mission with a description of the Data Manager's job and an explanation based on several researches of what the CSV is. Secondly, this report deals with the different missions carried out during my work-study at the Bergonié Institute as a Data Manager. And finally, this report discusses the differences that can exist between clinical trial data management and traditional DBMS, while addressing the conclusions related to CSV. The implementation of CSV is a long and complex step that requires a lot of thought. It involves the creation of a dedicated team with several experts in order to implement the best possible procedures applicable to all departments. To conclude, this report discusses the various personal contributions of this work-study program as well as the impact of the Covid-19 epidemic on it.

**Keywords** : CSV, Computerized System Validation, Data Management, Data Manager