Livre blanc

PHOENIX :
Concevoir et déployer **des guichets** pour favoriser et accélérer
LES ACCORDS DE PARTAGE DE DONNÉES DE SANTÉ

MEDICEN PARIS RÉGION
LE PÔLE DE COMPÉTITIVITÉ SANTÉ
DE LA RÉGION ILE-DE-FRANCE

VELTYS
SMART DATA SMART DECISIONS

Sommaire

Edito : PHOENIX, quoi et pourquoi ? 3

Infographie 4

1

1. L'utilisation des données de santé en France : un enjeu majeur entravé par des défis profonds 5

1.1 Les données de santé sont un levier majeur de la transformation des systèmes de santé en France et dans le monde 6

1.2 La France dispose de nombreux atouts et les initiatives visant à favoriser et accélérer le partage se multiplient 8

1.3 On observe encore de nombreux freins à la mise à disposition et à l'exploitation des données de santé en France 10

1.4 Ces difficultés ont été confirmées par le projet PHOENIX et elles témoignent d'un manque de coordination générale autour du partage des données 12

2

2. La mise en place de guichets de données : une solution structurante et modulable pour gagner en efficacité 18

Le projet PHOENIX 19

2.1 Le guichet « *fast-track* » : une solution innovante pour se structurer et accélérer la mise en place d'accords de partage de données 22

2.2 Le guichet « *fast-track* » : une structuration en cinq briques fondamentales déclinable selon les usages souhaités et les bases de données partagées 24

2.3 Le guichet « *fast-track* » : une traduction de la vision, de l'organisation et des contraintes liées au partage de données et propres à chaque acteur 27

3

3. Un déploiement des guichets possible pour tout acteur de l'écosystème en appliquant plusieurs bonnes pratiques 31

3.1 Conduire une réflexion sur l'existant (structure, *process* et verrous) et sur la cible souhaitée 33

3.2 Identifier des solutions opérationnelles, disponibles en interne comme en externe 35

Bonnes pratiques 37

3.3 Mettre en place des indicateurs pour suivre le déploiement et le fonctionnement d'un guichet 46

Notes de fin 50

Remerciements 52

Auteurs 53

Contributeurs 54

Présentation des structures 54

édito

Docteur Christian Deleuze
Président de Medicen Paris Region

PHOENIX, quoi et pourquoi ?

Les données de santé, leur structuration et leur usage sont aujourd'hui au cœur des intérêts de tous les acteurs, qu'ils soient académiques, industriels ou institutionnels. Ces données sont considérées, à juste titre, comme un des leviers majeurs de transformation des systèmes de santé.

De nombreux freins existent pourtant à un bon usage de celles-ci, tels que le manque de lisibilité et visibilité des sources de données existantes ou l'absence de standard permettant de qualifier et valoriser ces données de façon satisfaisante pour les diverses parties. Par ailleurs l'usage de ces données pour la recherche médicale n'est pas encore bien compris par les citoyens, et la collecte au cours du soin ne peut être simplement utilisée à des fins de recherche.

Une vision commune et partagée par tous est indispensable, et il faut donc s'interroger sur les moyens à mettre en place pour accélérer le partage des données de santé, en démontrant l'impact positif pour les citoyens et professionnels de santé.

Le constat de l'intérêt des données de santé pour faire avancer la médecine de demain, et en particulier la personnaliser, est mondial. Certains pays, en Europe notamment, sont en avance dans la collecte et l'utilisation de ces données. La France, longtemps en avance, doit s'organiser pour contribuer à l'essor des innovations basées sur l'utilisation de ces données et ainsi assurer, d'une part une réponse aux besoins spécifiques de sa population, et d'autre part soutenir un tissu industriel et académique riche et prometteur dans ce domaine, avec l'espoir de contribuer à faire émerger un leader Européen dans ce domaine sensible.

Pour cela, la collaboration entre tous les acteurs de la chaine de valeur des données est essentielle afin de construire un écosystème fort autour de standards reconnus et utilisés par tous, et une vision consensuelle tournée vers l'efficience.

Avec le projet PHOENIX, Medicen s'est engagé dans cette voie en proposant une initiative terrain innovante impliquant aussi bien des producteurs que des utilisateurs de la donnée de santé. L'objectif de ce projet est de définir des bonnes pratiques et d'accélérer la mise en place de projets innovants autour des données de santé. L'engagement des partenaires dans ce travail collaboratif, basé sur des échanges transparents et ouverts, est l'une des clefs de la réussite de ce projet.

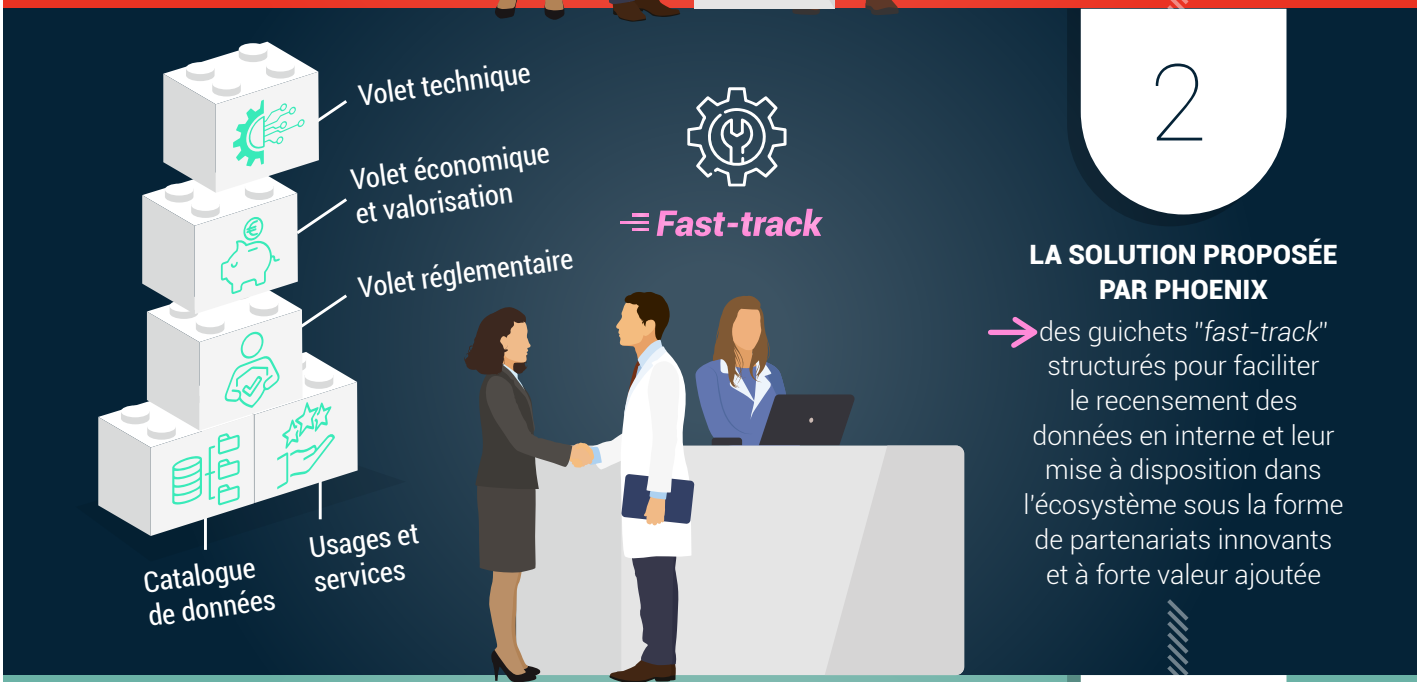
Medicen est fier de vous présenter aujourd'hui les résultats de ces travaux dans ce livre blanc qui se propose d'être un guide et un soutien pour tous les acteurs qui veulent participer à l'effort collectif qui fera des données de santé le moteur d'une médecine personnalisée pour chacun, efficiente et de confiance.

Christian Deleuze

1

L'EXPLOITATION DES DONNÉES DE SANTÉ EN FRANCE

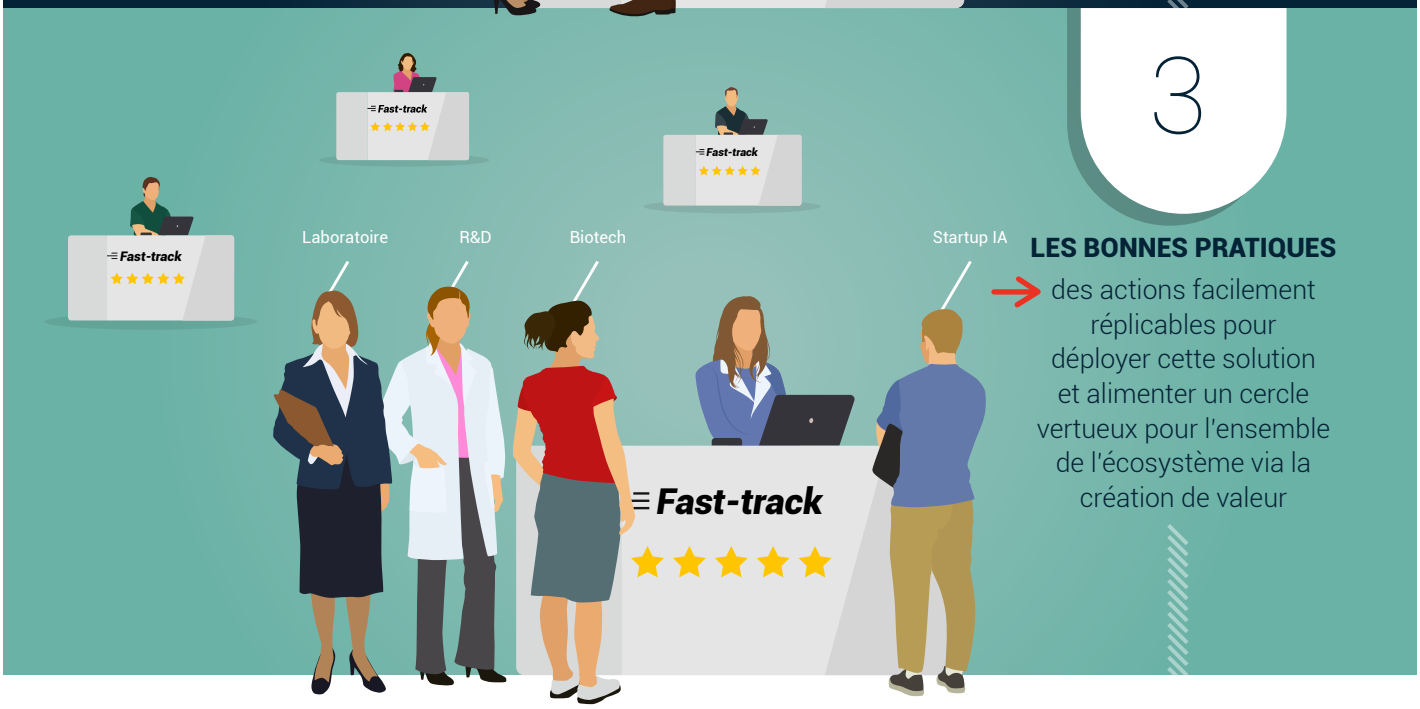
→ des opportunités nombreuses et des freins persistants à différents niveaux



2

LA SOLUTION PROPOSÉE PAR PHOENIX

→ des guichets "*fast-track*" structurés pour faciliter le recensement des données en interne et leur mise à disposition dans l'écosystème sous la forme de partenariats innovants et à forte valeur ajoutée



3

LES BONNES PRATIQUES

→ des actions facilement réplicables pour déployer cette solution et alimenter un cercle vertueux pour l'ensemble de l'écosystème via la création de valeur

1 L'utilisation des données de santé en France :

un enjeu majeur entravé par des défis profonds

1.1

LES DONNÉES DE SANTÉ SONT UN LEVIER MAJEUR DE LA TRANSFORMATION DES SYSTÈMES DE SANTÉ EN FRANCE ET DANS LE MONDE

Tous les acteurs reconnaissent aujourd'hui la valeur clef de la donnée de santé et son potentiel d'exploitation pour transformer le système de santé en profondeur et contribuer plus largement à l'économie que ce soit en termes de croissance, d'emploi ou d'innovation.

L'ensemble des rapports et études produits sur le sujet soulignent ainsi les nombreux impacts positifs d'une exploitation intelligente et massive des données sur le système : acquisition d'une meilleure connaissance du système de soins, identification de facteurs de risque de maladie, aide au diagnostic, au choix et au suivi de l'efficacité des traitements, développement de nouveaux modes de prise en charge, amélioration de l'offre de soins sur le territoire, pharmacovigilance, épidémiologie… À tel point que l'exploitation de données est devenue pour beaucoup un incontournable dans le pilotage médico-économique des établissements de santé, la prise de décision au niveau des politiques publiques ou encore la recherche biomédicale.

Les données de santé sont également au cœur des enjeux des acteurs de la filière santé numérique dont les solutions pourront générer des données massives et diverses, favoriser l'interopérabilité et l'accès sécurisé à ces données ainsi que leur exploitation pour optimiser le parcours de soin pour le patient. Le gouvernement français s'est mobilisé pour favoriser l'émergence de ces solutions innovantes avec le Ségur du numérique en santé embarquant un plan de financement de 2 milliards d'euros ce qui a permis de lancer la stratégie d'accélération « Santé numérique », et de poursuivre une feuille de route visant à accélérer le virage numérique en santé via cinq grandes orientations :

- Renforcer la gouvernance du numérique en santé
- Intensifier la sécurité et l'interopérabilité des données de santé
- Accélérer le déploiement des services numériques socles
- Déployer au niveau national des plateformes numériques de santé
- Soutenir l'innovation et favoriser l'engagement des acteurs •••

+70%

Au niveau européen, la valeur d'usage des données secondaires va passer de 25 Md€ en 2020 à 43 Md€ en 2028

Cela est logique au regard du potentiel de valeur de la santé numérique dont la croissance entre 2019 et 2023 serait de l'ordre de 160 % : de 90 milliards de dollars en 2019 à 235 milliards de dollars au niveau mondial[1]. Les bénéfices de la santé numérique incluent des gains pour les patients, une amélioration de l'offre de soins, des opportunités importantes pour les industriels et une meilleure prise en charge via les assureurs.

A titre d'exemple, un récent rapport de l'Institut Montaigne[2] estime que pour la France, le potentiel de création de valeur dans le domaine de la e-santé est compris entre 16 et 22 Md€ par an répartis de la façon suivante :

- Une plus grande autonomie des patients dans la gestion de leur santé, pour un gain compris entre 3,3 et 4,7 Md€.
- Une meilleure circulation de l'information médicale, notamment grâce à la dématérialisation, pour un gain de 3,4 à 4,7 Md€.
- Le développement de la télémédecine, qui facilite l'accès aux soins, pour un gain de 3,7 à 5,4 Md€.
- Une plus grande efficacité globale du système de soin, pour un gain de 2,4 à 3,4 Md€.
- Des décisions médicales et paramédicales plus sûres et plus fiables, pour un gain de 3,3 à 4,2 Md€.

Si on considère le périmètre plus restreint de l'usage secondaire[3] de la donnée de santé, qui devient un enjeu de plus en plus fort au sein de l'écosystème, là encore les chiffres sont importants :

- La Commission Européenne estime qu'au niveau européen, la valeur d'usage des données secondaires va passer de 25 Md€ en 2020 à 43 Md€ en 2028[4], soit une hausse de 70 %.
- Pour l'industrie pharmaceutique, IQVIA[5] estime que sur la période 2021-2023, les gains liés aux données de vie réelle seront de 4 Md$ pour l'industrie pharmaceutique dans le monde, sur l'ensemble de la chaîne de valeur.

Autre marqueur fort de l'intérêt porté par l'écosystème aux données, de nombreuses licornes[6] se sont développées dans ce domaine (par exemple, *Innovaccer, Komodo Health, Evidation Health, K Health, Clarify Health Solutions, Olive*), et des premiers rachats conséquents ont déjà eu lieu (*Flatiron* a ainsi été racheté par Roche en 2018 pour 1,8 Md€). En particulier, il convient de noter que beaucoup de ces entreprises sont des plateformes, qui structurent les données de leurs clients, offrent des services liés à l'IA, et dans certains cas vendent l'accès à ces données. Le haut niveau de valorisation de ces plateformes montre l'enjeu économique qu'elles représentent.

1) Global Digital Health Outlook, Artificial Intelligence, Care Coordination, Cybersecurity, Data Analytics, Digital Therapeutics, and Telehealth Will Drive the Market, Frost & Sullivan, 2020
2) Institut Montaigne, juin 2020, E-Santé : augmentons la dose
3) i.e. la réutilisation des données dans un usage qui n'est pas celui pour lequel la donnée a été générée initialement
4) Impact Assessment report Accompanying the document "PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space" (1/4)
5) Gores M. and E. Powers, 2021, Capturing Value at Scale: The $4 billion RWE Imperative
6) Les licornes sont des start-ups valorisées à plus d'un milliard de dollars

1.2

LA FRANCE DISPOSE DE NOMBREUX ATOUTS ET LES INITIATIVES VISANT À FAVORISER ET ACCÉLÉRER LE PARTAGE SE MULTIPLIENT

De nombreux projets et initiatives sont menés à différents niveaux (européen, national, par les acteurs de l'écosystème…) pour favoriser l'exploitation des données de santé. La mobilisation d'acteurs de différents profils, privés comme publics (hôpitaux, entreprises privées, nouveaux entrants, institutions…) pour travailler sur ces multiples dimensions est un signe très encourageant quant à la dynamique du secteur et la volonté de traiter ces problématiques.

Au niveau national, l'émergence du *Health Data Hub* en 2018 montre la volonté d'une gouvernance et d'une politique publique forte pour structurer, encadrer et faciliter la mise à disposition de données de santé en France et au niveau européen (par exemple pour répondre aux enjeux législatifs sur l'accès aux données de santé à travers l'Union Européenne[7]).

En parallèle, on observe également à des échelles plus locales des initiatives complémentaires qui visent à faciliter le partage de données dans le cadre de projets collaboratifs publics-privés. La « Clinique des données » portée entre autres par le CHU de Nantes illustre ainsi la volonté des établissements du grand Ouest de structurer à leur échelle la filière data, en permettant l'accélération du nombre de projets de collaboration et de recherche autour de l'exploitation des données. D'autres initiatives regroupant des acteurs privés et publics voient aussi le jour. Par exemple, le projet MELLODY vise à mettre en commun des informations des données issues d'acteurs habituellement concurrents tout en conservant la confidentialité des données afin d'améliorer les performances prédictives pour le développement des médicaments. Le projet CHAIMELEON vise quant à lui à mettre en place un référentiel structuré interopérable à l'échelle de l'UE pour les données d'imagerie médicale contextualisées par des données cliniques issues de plusieurs centres hospitaliers européens. L'objectif est de créer une source ouverte pour l'expérimentation de l'intelligence artificielle dans la gestion de plusieurs cancers. •••

Sebastian Schwarz OWKIN

Data Platform Strategy Lead

Les initiatives autour de la valorisation de la donnée sont nombreuses, y compris dans un contexte européen, et la maturité des acteurs autour de ces questions est grandissante. Cette dynamique est très positive mais des disparités subsistent selon les pays (parfois la valorisation de la donnée et les démarches associées sont mises en place au cas par cas et chronophages). Néanmoins, de plus en plus de producteurs de données de santé intègrent des réseaux comme EHDEN, un réseau Européen consacré à la mise en valeur de la donnée de santé et qui compte plus de 140 partenaires fournisseurs de données, mis en relation avec des PME capables de standardiser et valoriser leurs données de santé. Il est donc important de multiplier les collaborations entre producteurs de données de santé, les PME apportant des solutions innovantes et les acteurs industriels, afin que ces différents acteurs puissent s'unir pour générer ensemble la plus-value la plus importante possible pour le patient en se basant sur la donnée de santé.

7) Lancement d'un projet pilote pour l'Espace européen des données de santé mené par le Health Data Hub https://www.health-data-hub.fr/actualites/lancement-projet-pilote-ehds

Pierre-Antoine Gourraud CHU de Nantes

Professeur des universités et praticien hospitalier en biologie cellulaire

L'activité de la clinique des données a transformé une activité utile mais parfois mal reconnue de soutien épidémiologique à la recherche clinique. Depuis 2015, la Cellule CIC-EC est montée en compétence sur l'exploitation des données massives de par l'exploitation des données issues du SNDS mais également de l'entrepôt de données de santé du CHU de Nantes. Cette évolution a permis de développer le « catalogue » de sources de données exploitables par la Cellule devenue « clinique des données ». La Clinique a ainsi pu affirmer sa position de double ressource potentialisant la communauté des investigateurs du CHU et de Nantes Université en étant classiquement une équipe experte en méthodologie, épidémiologie et biostatistiques des données mais de manière nouvelle en devenant aussi source de données au bénéfice des équipes investigatrices.

Les indicateurs utilisés pour suivre ces effets sont le nombre de projets accompagnés (398 entre 2016 et 2021) et le nombre de publications associant l'équipe (111 entre 2016 et 2021).

C'est dans cette dynamique que s'inscrit le projet PHOENIX[8], qui accompagne les producteurs de données individuellement (chacun sur leur rôle) sur leurs propres difficultés et sur leurs problématiques communes. Le projet vise à lever les verrous dépendant des producteurs de données en mettant en place des guichets de données de santé[9], solution identifiée comme optimale par le groupe de travail du projet PHOENIX car elle vient en complément des solutions existantes et déjà efficaces et met l'accent sur le déploiement d'outils concrets et transverses.

8) La description précise du projet est disponible dans les encadrés pages 19-20 et au début de la deuxième partie.

9) Le guichet est défini dans le glossaire en annexe. Une définition plus exhaustive est l'objet de la deuxième partie de ce document.

1.3

ON OBSERVE ENCORE DE NOMBREUX FREINS À LA MISE À DISPOSITION ET À L'EXPLOITATION DES DONNÉES DE SANTÉ EN FRANCE

La mise à disposition des données de santé devient donc un enjeu de plus en plus crucial au sein de l'écosystème français (comme dans les autres pays comparables) et de nombreux acteurs s'engagent sur ce sujet. Cependant, notre écosystème souffre de nombreux freins qui rendent difficile un partage simplifié et accéléré des données. Ces difficultés sont d'autant plus fortes aujourd'hui qu'une très large majorité de projets dépassent actuellement la seule mise à disposition pour se porter sur des objectifs économiques de création de valeur à partager au sein de l'écosystème.

Les travaux de l'OCDE[10] indiquent que la France se situe plutôt à un haut niveau sur les questions de gouvernance de la donnée, et à un niveau plus intermédiaire sur les questions de maturité et d'usage secondaire des données. D'après l'OCDE, cela s'explique par (i) l'absence de bases de données sur certaines aires thérapeutiques (par exemple le diabète ou les maladies cardio-vasculaires), (ii) le manque d'automatisation dans la collecte des données et (iii) des facteurs plus structurels (temps pour collecter les données, difficultés d'identification unique entre les bases, faible utilisation pour évaluer la performance).

Un nombre de plus en plus conséquent de projets utilisant les données de santé s'appuient sur des outils d'intelligence artificielle pour apporter des solutions efficaces en santé et mieux comprendre certains phénomènes, par exemple avec l'aide au diagnostic, notamment en imagerie. A cet égard, une étude conduite par *AI for Health*[11] (questionnaire adressé à 80 acteurs – hospitaliers, pharmacie, biotech, etc. – et réalisation d'une vingtaine d'entretiens) relève plusieurs freins à l'usage de l'intelligence artificielle en santé en France. La disponibilité et la qualité des données sont perçues comme les premiers obstacles par les répondants, suivies par le manque de budget (notamment dans les institutions publiques comme les établissements de santé et les centres de recherche). L'enjeu de disponibilité / qualité est perçu comme prioritaire dans tous les secteurs. Cela s'est confirmé au cours du projet PHOENIX : l'ensemble des contributeurs ont regretté l'absence de standards de qualité pour évaluer les bases de données. •••

10) J.Oderkirk, 2021, Survey results : National health data infrastructure et governance, OECD Health Working Papers No.127
11) AI for Health Book, 2021

Ces différentes analyses sont également confirmées par les travaux de l'*Open Data Institute (ODI)*[12], qui a conduit une étude qualitative visant à qualifier et caractériser la situation générale des pays européens en termes d'usage secondaire de la donnée de santé. Le rapport se fonde sur une analyse des écosystèmes pays par pays pour classer ces derniers selon plusieurs dimensions (infrastructure, capacités, éthique, innovation, équité, engagement) afin de mesurer leur niveau de maturité générale sur la question. Dans l'évaluation, l'*ODI* pointe du doigt plusieurs freins à cet usage dans le cas français :

Enjeux d'infrastructure

L'*ODI* montre dans son étude que d'un point de vue des infrastructures actuellement mises en place, la France connaît du retard sur certains points, par exemple sur le manque de cadre structuré et mature de collecte et de développement de *PROMs* (*Patient Reported Outcome Measures*) ;

Enjeux d'innovation

Manque de reconnaissance de l'usage secondaire des données de santé dans les politiques gouvernementales ; l'étude indique à la fois qu'en termes d'investissement dans les dossiers patient électroniques (DPI en France : Dossier Patient Informatisé) et qu'en termes de priorisation donnée à l'usage secondaire de la donnée de santé, la France souffre d'un certain retard ;

Enjeux éthiques

(i) Selon l'étude, les cas d'exploitation secondaire des données ne permettent pas suffisamment aujourd'hui d'atteindre une forme d'équité dans le bénéfice dont pourraient profiter les patients, (ii) l'étude indique également que la loi du 2 août 2021 relative à Bioéthique est davantage orientée vers l'exploitation à visée de services de télémédecine que dans un but de recherche scientifique pure, même si elle donne un cadre légal visant à assurer que les exploitants rendent des comptes dans l'usage des données.

Enfin, du point de vue des citoyens[13], le potentiel d'exploitation des données, s'il est identifié et reconnu, soulève également un certain nombre de questions qui sont autant de freins potentiels à leur utilisation par l'écosystème. En premier lieu, la confidentialité et la sécurité des données de santé des patients sont au cœur des préoccupations citoyennes, qui sont par ailleurs renforcées par les cyberattaques et piratages de plus en plus nombreux. Par ailleurs, la notion de consentement à l'utilisation des données occupe également une place importante dans les débats autour de l'exploitation secondaire des données de santé et soulève un enjeu éthique fort doublé d'un enjeu de communication et d'information très important pour les citoyens et très contraignant pour les producteurs de données. Enfin, la question de l'équilibre entre les bénéfices et les risques dans l'usage secondaire des données de santé est prégnante : si les contributions qui visent à améliorer la santé publique, les soins ou encore la recherche sont fortement plébiscitées, l'exploitation de données à caractère commercial suscite une méfiance voire une opposition parfois forte.

Il ressort ainsi des différentes études autour de l'exploitation des données de santé que la France connaît un certain nombre de difficultés structurelles. Ceci étant dit, les nombreuses initiatives mentionnées ci-dessus s'efforcent de lever les verrous et problématiques rencontrées par l'écosystème, tout comme le projet PHOENIX.

12) M. Boyd, M. Zimeta, J. Tennison, M. Alassow – Open Data Institute (ODI) – Secondary use of health data in Europe, 2021

13) J. Maddocks, L. Matthieu, N. Courbon, R. Richards, M. Saelaert, W. Van Hoof – France Assos Santé - Le Débat des Données, une consultation citoyenne en ligne sur la réutilisation des données de santé - 2022 - https://www.france-assos-sante.org/wp-content/uploads/2022/08/TEHDAS-M-8.2-FR.pdf

1.4

CES DIFFICULTÉS ONT ÉTÉ CONFIRMÉES PAR LE PROJET PHOENIX ET ELLES TÉMOIGNENT D'UN MANQUE DE COORDINATION GÉNÉRALE AUTOUR DU PARTAGE DES DONNÉES

La première étape du projet PHOENIX a consisté à dresser le bilan des difficultés rencontrées et des attentes de chaque acteur dans le partage de leurs données de santé. L'un des points forts de cette approche a d'abord été d'assurer le partage de ces informations au sein du groupe de travail. Notre idée était en effet d'identifier les difficultés communes qui constituent les verrous principaux pour l'écosystème et les difficultés internes amenées à être levées avec la mise en place des guichets[14]. Ces échanges se sont également révélés bénéfiques pour les partenaires eux-mêmes en leur permettant de se rendre compte de leur niveau de maturité, de comprendre les contraintes et les efforts des uns et des autres, de créer un lien de confiance et de s'accorder sur des besoins communs à pousser auprès des institutions et de l'écosystème.

Quatre grandes familles de verrous communs à l'ensemble des acteurs émergent :

1. Le manque d'acculturation et d'harmonisation des pratiques

Il s'agit d'une difficulté majeure d'après l'ensemble des acteurs impliqués dans le projet et elle est à la racine de nombreuses difficultés « secondaires » qui sont visibles à tous les niveaux.

L'absence de standards portant sur le format des données et l'absence de pratiques communes sur ce sujet augmente le risque d'incompréhension entre les parties prenantes d'un projet. Le manque d'harmonisation se traduit aussi par des difficultés à assurer et à démontrer la qualité des données et la véracité des informations descriptives associées aux *datasets*. Il en résulte, d'après les acteurs mobilisés, un manque de confiance et des difficultés à contractualiser.

Le manque d'acculturation est un phénomène observé également en interne : la plupart des organisations n'ont ni une connaissance exhaustive ni la maîtrise complète des données dont elles disposent. Pour avancer sur ce point, un embarquement fort des équipes sur ce sujet est nécessaire. Il permet d'assurer une bonne saisie des données – ce qui garantit une partie de la qualité – et la remontée des informations sur les jeux de données existants et exploitables. C'est un point crucial dans les structures décentralisées pour faire remonter les informations. À ce stade, ce point reste une difficulté pour les acteurs.

Sur la réglementation et les procédures à réaliser pour assurer la conformité des projets, le manque d'acculturation se ressent également beaucoup et les trois acteurs mobilisés convergent pour indiquer qu'il s'agit d'un facteur majeur expliquant la lenteur de la contractualisation, voire le renoncement à certains projets. La méconnaissance de la réglementation française et / ou l'incompréhension sur ce sujet des acteurs – qu'ils soient producteurs de données ou porteurs de projets – est encore très marquée, et cela est particulièrement observé chez les acteurs étrangers, même européens. •••

14) Le guichet est défini dans le glossaire en annexe. Une définition plus exhaustive est l'objet de la deuxième partie de ce document.

Marguerite Brac de La Perrière Cabinet LERINS

Avocat associée IT Data et Santé, experte en santé numérique

La réutilisation des données de santé est un sujet qui doit être traité de manière systémique (au niveau de la société), et non au niveau individuel, tenant compte du fait que les patients aussi bien informés qu'ils pourraient l'être ne seront jamais en mesure d'évaluer les garanties et risques, en termes de sécurité et donc de confidentialité, associés à la réutilisation de leurs données. À cet égard, il est regrettable que le consentement des patients à la réutilisation de leurs données figure dans les critères applicables au référencement des services et outils numériques au catalogue de service de l'Espace Numérique de Santé (arrêté du 23-06-2022).

Il convient déjà de distinguer la réutilisation de données de santé anonymisées, de la réutilisation de données de santé pseudonymisées. En dehors des Entrepôts de Données de Santé qui constituent une première brique importante, il convient à mon sens d'appréhender les pratiques sectorielles, de définir des cas d'usage standards et un encadrement spécifique par cas d'usage.

Camille Bachot Roche Pharma

Medical Data Platform Lead

Aujourd'hui, la réglementation française et européenne protège très bien les patients quant à l'utilisation de leurs données de santé, ce qui est une très bonne chose. Ainsi, lorsque l'on souhaite utiliser les données pour une réexploitation secondaire, la mécanique de réinformation ou obtention d'une autorisation à déroger à l'obligation d'informer peut être longue et fastidieuse. Pour autant, les technologies d'anonymisation des données et la mise en place de portail d'information des patients sont des outils qui vont faciliter la réexploitation secondaire des données.

Pauline Touche Hôpital Foch

Responsable adjointe de l'unité data

Le règlement général sur la protection des données (RGPD) impose une information concise, transparente, compréhensible et aisément accessible des personnes concernées. Mais dans la pratique, celle-ci est difficile à mettre en œuvre. Un grand nombre de portails de transparence existent en France, mais aucune initiative nationale n'a vu le jour. Un patient doit donc se connecter à chaque portail pour consulter la liste des études dans lesquels ses données personnelles ont potentiellement été réutilisées et faire valoir ses droits. De mon point de vue, nous avons encore du chemin à parcourir pour que l'information soit réellement transparente ce que je trouve dommage car la très grande majorité des patients sont favorables à la réutilisation de leurs données à des fins de recherche, car ils ont conscience que cela nous permet d'accroître nos connaissances et donc d'améliorer leur prise en charge ou celle de futurs patients.

2. Le « cas par cas », marqueur d'un manque de procédures formalisées et de coordination inter et intra structures

Les trois acteurs participant au projet PHOENIX ont tous décrit un traitement des demandes de partage de données souvent réalisé au cas par cas que ce soit chez les producteurs de données de santé ou chez les porteurs de projet. **Le manque de structuration et de coordination rend nécessaire l'application d'une procédure très spécifique à chaque projet**. Cela constitue un facteur allongeant les durées de traitement et les durées de contractualisation puisqu'il est souvent nécessaire de créer de toutes pièces des solutions sur mesure. Cela crée également un risque de dépendance forte aux membres de l'équipe en charge des partages de données de santé.

Il est souvent nécessaire de créer des solutions sur mesure

Le travail mené au cours du projet a également mis en évidence **l'absence de *process* visant à évaluer le potentiel d'utilisations secondaires** des jeux de données. Ce potentiel est rarement évalué et il semble que très peu de producteurs de données de santé sont assez matures à l'heure actuelle pour essayer de pousser des projets auprès des industriels qui s'appuieraient sur des bases de données déjà créées (pour des besoins internes comme pour des projets embarquant déjà un partage). Ce type d'initiative *techno-push*[15] donnerait pourtant un signal à l'écosystème quant à l'existence des données en question et quant à la maturité et la volonté du producteur de développer ce type de projets en commun.

La réalisation de l'information des patients et du recueil du consentement est également un sujet posant des difficultés aux trois acteurs mobilisés dans le projet PHOENIX. La plupart des procédures existantes à date – au niveau de l'accueil des patients notamment – n'embarquent pas d'outils spécifiques visant à assurer la possibilité ultérieure d'informer les patients. Il s'agit aussi d'un sujet de coordination inter structures notamment dans les cas où les projets mobilisent des données provenant de plusieurs producteurs de données. Cette coordination est aussi essentielle puisque les patients doivent être informés de l'usage spécifique des données qui leur sont relatives.

Un autre facteur limitant pour les partages de données de santé est **la difficulté à assurer la gouvernance de ces données**. Il s'agit ici d'une défiance entre les acteurs qui n'est pas résolue via la mise en place de procédures de coordination entre les acteurs de l'écosystème. Par exemple, il est très **difficile voire impossible à l'heure actuelle de contrôler l'usage des données** fait par un porteur de projet ou de s'assurer qu'il a bien supprimé ces données suite à leur utilisation. Pour les trois partenaires du projet, garder la gouvernance des données est une pratique très importante, à la fois pour des enjeux réglementaires et pour des enjeux stratégiques, et l'absence de pratiques et de coordination en la matière est donc une difficulté forte. •••

Camille Bachot Roche Pharma

Medical Data Platform Lead

La notion de représentativité des données de vie réelle est clé en santé publique ou en épidémiologie. La valeur scientifique est trouvée à partir du moment où les données proviennent de multiples établissements ou sources et offrent une photo régionale, nationale ou même internationale. Les détenteurs de jeux de données ont donc tout intérêt à s'associer pour en augmenter la valeur.

15) Techno push : processus démarrant suite à une offre du producteur ; par exemple partir d'une base de données existante et la pousser auprès d'industriels pour créer un produit / un projet. Market pull : processus démarrant suite à un besoin du marché ; typiquement le cas d'un industriel approchant un producteur de données pour obtenir un jeu de données spécifique.

3. Des difficultés techniques pour structurer, protéger et partager les données

La troisième famille de difficultés rencontrées par les trois acteurs dans le cadre de partages de données de santé concerne les problématiques techniques, c'est-à-dire les **solutions IT pour collecter, extraire, structurer, mettre en conformité, contrôler et partager** des bases de données. De plus en plus de solutions techniques sont mises au point pour y répondre mais elles doivent encore être adoptées et / ou améliorées pour pouvoir prendre en charge l'ensemble des difficultés auxquelles font face les producteurs de données de santé.

Paul Rinaudo ADLIN Science

CEO & Co-founder

Le partage des données, si elles ne sont pas structurées, n'a qu'une utilité limitée. Avant de penser à la valorisation des données, il faut penser à la structuration et à la qualité. C'est la première question à se poser pour le partage de données.

Une difficulté majeure qui touche notamment une grande partie des hôpitaux est **l'impossibilité d'extraire et de structurer les données issues des logiciels métier**. Le manque d'interopérabilité entre les logiciels, l'accès aux données, et le type de données accessibles (trop souvent composées de commentaires dans des champs textuels à remplir) limitent la capacité des hôpitaux à extraire les données et structurer des bases exploitables.

Antonio Borderia PhD. ADLIN Science

Executive Vice-President Business Development

Au niveau technique un challenge majeur est l'interopérabilité : chaque hôpital et laboratoire recueille des données qui ont des formats différents donc cela implique encore un travail au cas par cas pour connecter les différentes bases issues de sources différentes. Il manque aussi un consensus sur le format à utiliser.

Une solution permettant de mettre en conformité les jeux de données partagés est l'anonymisation des données. Néanmoins à ce stade, **les techniques d'anonymisation font craindre une altération des données, voire des biais**, pouvant nuire aux analyses via le retrait d'information nécessaire à la désidentification des patients : l'agrégation fait perdre le détail des données individuelles, la production de données synthétiques / d'avatars qui reproduisent les schémas observés dans la base d'origine fait perdre la vision sur des caractéristiques peu communes qui pourraient présenter un intérêt. D'autre part ces solutions ne sont pas toujours associées à des brevets ou agréées par la CNIL, ce qui peut réduire la confiance des parties prenantes et limiter leur usage. •••

4. La valorisation des données de santé : une difficulté majeure pour tous les acteurs

Une autre difficulté relevée par les acteurs participant au projet PHOENIX et partagée par l'ensemble de l'écosystème est la difficulté à correctement valoriser les données dans les projets pour lesquels elles sont mises à disposition.

D'une part, la notion de valorisation recouvre plusieurs enjeux tous aussi importants pour les acteurs de l'écosystème : la **valorisation scientifique** de la donnée de santé à travers des brevets ou encore des publications, la **valorisation stratégique** des données qui permet d'élaborer une stratégie d'institution ou de structure à partir de l'exploitation de ces données, la **valorisation marketing** qui vise à utiliser les données pour communiquer sur le travail du producteur de données ou encore la **valorisation économique** et financière liée à la mise à disposition des données partagées.

Les études récentes montrent qu'il n'y a pas de modèle dominant

Sur ce dernier point en particulier, les acteurs font actuellement face à un **manque d'acculturation, d'outillage et plus généralement de cadre commun**, en particulier lorsque la mise à disposition des données se fait dans le contexte de projets collaboratifs au long cours, par exemple sur le développement d'outils d'*IA*. Les enjeux associés au partage de la valeur sont particulièrement complexes dans ce cas précis et les discussions entre partenaires sont bien souvent difficiles à faire aboutir, au risque de voir le projet ne pas se faire. On peut penser par exemple aux enjeux associés au partage de la propriété intellectuelle (PI) et à la détermination de la part de propriété intellectuelle qui est imputable à une base de données et à un algorithme.

Les outils utilisés, les pratiques de chaque acteur et les valorisations chiffrées (ordres de grandeur) sont très peu connus et partagés : l'écosystème est particulièrement opaque sur cet aspect et c'est un facteur important limitant la confiance et ralentissant le lancement des projets. Il y a une forte nécessité à davantage de transparence.

D'autre part, les études récentes[16] confirment tous ces éléments et montrent qu'il n'y a **pas de modèle dominant**. La valorisation se fait bien souvent au cas par cas, ce qui complexifie les projets et rend très difficile l'établissement de standards de marché qui permettraient aux acteurs d'accélérer sur ce volet là encore. Cette difficulté est exacerbée par l'évolution des modèles et des solutions de valorisation qui changent et s'adaptent aux nouveaux usages et nouvelles technologies. •••

Maxime Agostini Sarus Technologies

Président & cofondateur

Les entraves au partage de données proviennent avant tout de la capacité à rassembler de la donnée de qualité et de trouver un cadre réglementaire et contractuel pour le partage. Ce ne sont pas des défis techniques mais des défis organisationnels et juridiques. Des solutions techniques s'attellent à les résorber mais il reste du travail ! Par exemple, l'anonymisation des données facilite l'aspect réglementaire mais soulève des questions théoriques profondes sur le risque de ré-identification. L'apprentissage fédéré est une méthode qui permet de renforcer la sécurité de la donnée mais qui apporte son propre lot de difficultés.

16) Aviesan, Covalliance, Veltys : Etude internationale sur les modèles de valorisation de la donnée de santé, janvier 2020

Linda Nait-Kaoudjt Institut Pasteur

Head of Technology Transfer and Industrial Partnerships

Aujourd'hui les difficultés pour réaliser des partages de données restent multiples : l'obtention de l'autorisation des patients, la structuration d'un catalogue de données avec un format harmonisé des données, la valorisation technique de la donnée et la définition des étapes de montage d'un projet sont des challenges qu'il faut encore relever. Par ailleurs le manque de modèles communs pour la valorisation comme pour la structuration des données est un frein important pour les partages.

Florimond Bourdeaux Veltys

Associé du cabinet en charge des projets santé

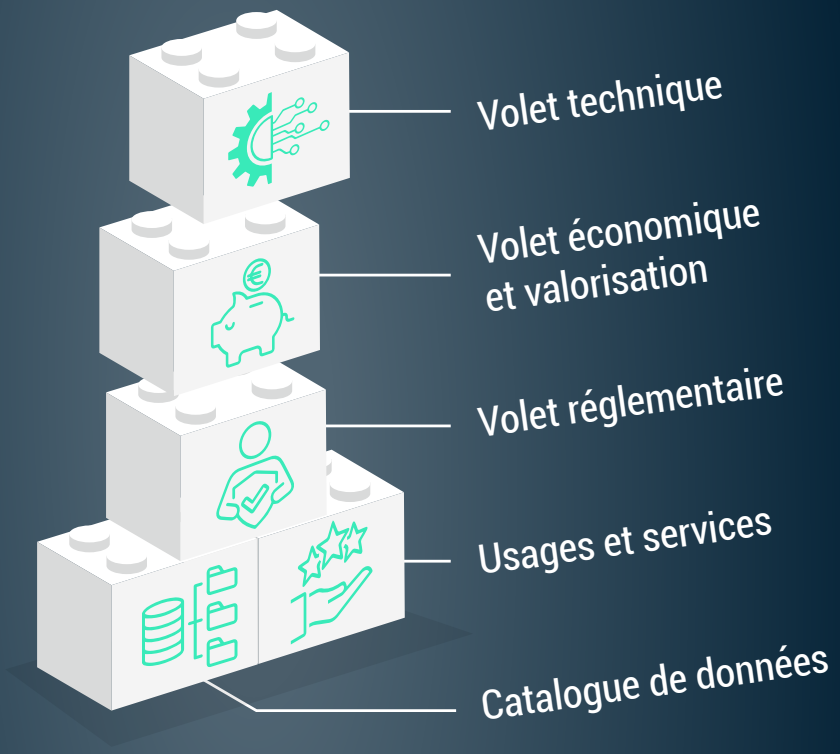
La valorisation des données de santé reste aujourd'hui un défi pour l'ensemble de l'écosystème. Il n'existe pas de solution miracle et chaque projet de collaboration donne lieu à des négociations parfois longues et difficiles entre les potentiels partenaires, négociations qui n'aboutissent pas toujours. Au-delà des outils servant à valoriser les données et répartir la valeur créée, la question de la valorisation montre la difficulté de l'écosystème à trouver un cadre commun suffisamment robuste et clair pour établir une confiance mutuelle entre acteurs.

Face à ces difficultés, le projet PHOENIX a été lancé afin de conduire un travail commun entre plusieurs acteurs de l'écosystème pour pallier les manques précités et proposer un modèle innovant de mise à disposition des données de santé grâce à la structuration et au déploiement de guichets « *fast-track* ». L'objectif était à la fois d'identifier les solutions pour traiter les verrous internes et de s'appuyer sur la réflexion collective pour proposer des solutions à l'écosystème, notamment sur les enjeux d'harmonisation et de standardisation, à travers l'alignement des trois partenaires du projet PHOENIX.

Il s'agit bien d'apporter des réponses concrètes et opérationnelles qui permettent l'engagement des équipes en interne et qui créent les conditions d'échanges entre différents acteurs pour faciliter et accélérer le partage de données de santé via la structuration d'outils et de *process*. Le dispositif mis en place a eu comme force la mobilisation d'acteurs variés au sein d'organisations très différentes. Cela a permis de confronter les expériences pour créer des solutions en interne et s'appuyer sur la multiplicité des acteurs pour alimenter un dialogue vertueux visant à construire des solutions externes, à pousser auprès de l'écosystème.

2 La mise en place de guichets de données :

une solution structurante et modulable
pour gagner en efficacité



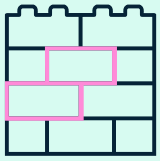
Fast-track

Le projet PHOENIX

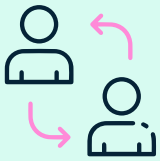
Ce projet a été lancé en 2021 à l'initiative de Medicen dans le cadre du *Medicen Initiatives for Health Data (MIHD)*. L'objectif était de définir les bonnes pratiques pour favoriser les projets collaboratifs autour des données de santé existantes via leur valorisation et l'accélération de la mise en place des projets.

Pour ce faire, nous souhaitions avoir une approche concrète et faire une preuve par l'exemple du potentiel de la structuration du processus de partage avec un guichet « *fast-track* ». La démarche envisagée était la suivante :

▸ Diagnostiquer :
Quels sont les freins au partage de données et comment peut-on accélérer le parcours de mise en place des projets autour de l'exploitation des données ?



▸ Structurer :
Quels sont les leviers pour accélérer les discussions, donner confiance aux porteurs et assurer la réalisation des projets ?



▸ Échanger :
Comment s'assurer que ce travail réponde aux attentes et besoins du plus grand nombre, reflète des bonnes pratiques et soit utile à l'ensemble de l'écosystème ?

Ce projet a réuni un consortium de trois acteurs aux profils différents : un acteur clinique, l'hôpital Foch, un acteur académique, l'institut Pasteur et un acteur industriel, Roche Pharma. Il nous paraissait pertinent de mutualiser les expériences de ces acteurs pour identifier les freins communs, les spécificités de chacun et initier un dialogue ouvert sur leurs différents points de vue. Pour créer une dynamique, organiser le partage d'informations et animer les réflexions individuelles ou communes, un accompagnement conjoint de ces acteurs par Medicen et Veltys a été mis en place. C'est grâce à la confrontation des points de vue entre acteurs de différents types et au sein des services de chacun que des enseignements et des bonnes pratiques profitables aux autres acteurs de l'écosystème ont pu émerger. •••

En parallèle des phases de réflexion et de partage, les acteurs ont progressivement formalisé les objectifs et défini les outils pour mettre en place des guichets répondant à leurs besoins. Ces guichets seront utilisés et évalués dans le cadre d'un Appel à Projets. Pour permettre un déploiement efficace dans les délais, nous avons proposé aux trois partenaires d'utiliser un seul jeu de données ciblant une thématique de leur choix, à partir duquel des porteurs de projets pourront exprimer des besoins précis.

Le projet s'est déroulé en trois phases :

1 *de fin mai 2021 à février 2022*

Cartographie des *process* existants, des *process* et des *KPI* en parallèle de la sélection des *datasets* par les différents acteurs

2 *de mars 2022 à août 2022*

Design des guichets au niveau de leurs éléments constitutifs et mise en place des guichets

3 *à partir d'octobre 2022*

Lancement et suivi des appels à projets

L'objectif de ce projet est de partager les enseignements tirés et les bonnes pratiques identifiées à l'ensemble de l'écosystème et c'est la raison d'être de ce livre blanc : partager le retour d'expérience et les apports de ce travail. Au-delà du partage, nous espérons que les acteurs de l'écosystème s'approprieront et enrichiront les résultats de ce projet avec des initiatives actuelles ou futures. L'objectif à terme est de fédérer un écosystème sur les enjeux de partage de données de santé autour de pratiques communes issues de l'expérience accumulée sur le terrain.

Les enseignements, les bonnes pratiques et également la structure du guichet que nous livrons dans ce livre blanc correspondent aux solutions effectives qui ont été identifiées comme optimales puis déployées progressivement par chacun des membres du groupe de travail.

Camille Bachot Roche Pharma

Medical Data Platform Lead

Il a été extrêmement enrichissant de partager les points de vue d'un établissement de soin, d'un centre de recherche biomédicale et d'un industriel de la pharma sur leur vision respective du partage et de la valorisation des données. Nous avons constaté la nécessité de bien aligner nos visions et de mieux structurer les processus de partage des données.

Linda Nait-Kaoudjt Institut Pasteur

Head of Technology Transfer and Industrial Partnerships

Les échanges au cours du projet PHOENIX nous ont permis de réaliser que les acteurs privés et publics partagent les mêmes problématiques : si la finalité peut être distincte, les questions techniques, réglementaires et de valorisation sont identiques.

D'un point de vue interne, participer au projet nous a permis ou va nous permettre de mettre en avant la conduite de changement à réaliser au niveau de notre structure et d'accélérer notre ambition de réaliser des projets collaboratifs de partage de données. Ce projet est une première étape importante pour le développement de notre guichet unique et nous avons pu y définir une feuille de route pour atteindre cet objectif.

Pauline Touche Hôpital Foch

Responsable adjointe de l'unité data

Le projet PHOENIX coordonné par Medicen avec le soutien du cabinet Veltys, m'a permis d'échanger avec l'Institut Pasteur et le Laboratoire Roche, sur cette thématique. La donnée de santé est très variée, il peut aussi bien s'agir de données issues de recherches impliquant la personne humaine (RIPH), que de données générées dans le cadre du soin, de la recherche fondamentale ou bien générées par le patient lui-même. Le projet PHOENIX a rassemblé des producteurs de données différentes, mais en échangeant nous nous sommes très vite aperçus que nous rencontrions les mêmes difficultés. Pour moi, PHOENIX a été une aventure humaine grâce à laquelle je me suis nourrie de compétences, expériences des différents participants. Ce projet m'a permis de prendre du recul sur notre façon de faire et d'ainsi repenser notre «guichet data» de demain. PHOENIX a été un incroyable accélérateur pour l'amélioration de ce guichet.

2.1

LE GUICHET « FAST-TRACK » : UNE SOLUTION INNOVANTE POUR SE STRUCTURER ET ACCÉLÉRER LA MISE EN PLACE D'ACCORDS DE PARTAGE DE DONNÉES

La solution que nous avons mise en place avec les trois partenaires a été la structuration de *process* en interne au sein d'un guichet de données de santé. Il s'agit en effet d'un facteur déterminant pour favoriser la mise en place d'accords de partage de données, et pour gagner en efficacité entre les premiers contacts et le lancement effectif d'un projet. C'est un moyen optimal d'exploiter les outils existants s'ils sont efficaces et de les inscrire dans un fonctionnement coordonné, construit pour s'appuyer sur ce qui est opérationnel, pour développer des *process* efficaces et pour capitaliser sur l'expérience accumulée en adaptant les pratiques en continu.

Nous pensons qu'il s'agit également d'un outil puissant pour mettre en place des pratiques permettant de pousser des utilisations secondaires de bases de données, en levant les verrous décrits dans la partie 1.4. Comme nous avons pu le constater, cela passe aussi par l'impact en interne. En effet, la mise en place d'un guichet favorise l'acculturation et l'embarquement des acteurs dans la démarche et peut donc assurer un renforcement des liens entre les services / départements et une meilleure compréhension des objectifs.

Un guichet de données de santé se définit comme une structure et une politique de mise à disposition de données de santé mises en place par un acteur producteur de données de santé (hôpital, laboratoire, entreprise…) pour des acteurs externes (publics comme privés). Il peut prendre plusieurs formes mais repose sur la structuration de pratiques et d'outils et sur la coordination d'expertises qui permettent de gagner en efficacité dans la mise en place et la réalisation d'accords de partage de données. •••

Pour cela, nous avons accompagné les acteurs dans la réflexion, la structuration et la mise en place d'un guichet de données de santé, au sein de leur organisation, qui embarque un ou plusieurs mécanismes proposant un parcours par étapes aux porteurs de projet pour :

- **Disposer d'informations** sur les bases de données disponibles
- **Proposer un projet** au producteur de données de santé
- **Contractualiser** en s'appuyant sur une procédure balisée et des outils clairs

Il est nécessaire pour chaque acteur de s'organiser en interne

L'objectif est bien de faciliter et d'accélérer le parcours de mise en place des projets, notamment collaboratifs, autour de l'exploitation des données de santé.

D'après nous, il s'agit également d'une solution qui peut permettre d'assurer et de faciliter l'action des acteurs de l'écosystème, mais aussi des agents en interne amenés à alimenter le guichet autour d'un mécanisme innovant pour faire vivre le partage de données. Cela s'est confirmé au cours du projet avec l'embarquement de participants supplémentaires au sein des équipes de chaque acteur au fil du projet.

Pour les producteurs de données, il nous semble essentiel d'**être proactif en interne, dans la structuration des *process*** liés au partage de *datasets*, en amont des demandes de partage par des porteurs de projet ou de propositions de solutions au niveau institutionnel. Le succès et la performance des initiatives impliquant l'utilisation de données de santé dépendent en grande partie de la maturité de l'écosystème, notamment de la préparation des différents acteurs face aux enjeux réglementaires, techniques et économiques. Malgré cela, il est nécessaire pour chaque acteur de s'organiser en interne, quelle que soit la solution envisagée pour le partage.

Pour tous les acteurs il sera utile d'assurer en interne :

- La connaissance et la maîtrise des données existantes au sein de l'infrastructure pour déterminer les possibilités de partage, les opportunités de réutilisations des données pour des projets spécifiques et les *process* et solutions techniques à mettre en place pour optimiser le partage. L'objectif visé est la **préparation des données**.
- La préparation de démarches précises et la création d'outils-type, à décliner différemment selon les situations rencontrées pour accélérer les négociations et la contractualisation ; cela nécessitera une réflexion visant à recenser ces différentes situations et une adaptation en continu pour répondre aux nouveaux cas. L'objectif visé est la **préparation des outils** utilisés pour le partage des données.
- La définition des modalités de partage pour assurer l'aspect fonctionnel : il s'agira de définir le parcours suivi par une demande de partage de données pour optimiser la mobilisation en interne et éviter la réalisation de tâches en doublon. L'objectif visé est la **préparation des *process*** pour les demandeurs de données.

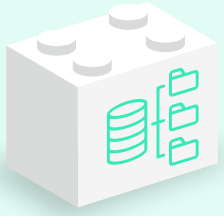
Mener ce type de réflexion et ce travail de structuration permettra d'identifier les difficultés principales rencontrées à date en interne, de définir les bonnes pratiques dans une démarche d'amélioration continue et de les évaluer pour réaliser un pilotage des guichets par l'efficience.

Les grandes initiatives – comme celle poussée par le *Health Data Hub* – sont donc complémentaires de ce type de démarche : elles permettront de créer un environnement plus favorable pour le partage de données. D'autre part les bonnes pratiques et les outils créés de part et d'autre s'alimenteront mutuellement.

2.2

LE GUICHET « FAST-TRACK » : UNE STRUCTURATION EN CINQ BRIQUES FONDAMENTALES DÉCLINABLE SELON LES USAGES SOUHAITÉS ET LES BASES DE DONNÉES PARTAGÉES

Un guichet tel que nous l'avons défini dans le cadre du projet PHOENIX se structure autour de deux points d'entrée :



Le catalogue de données

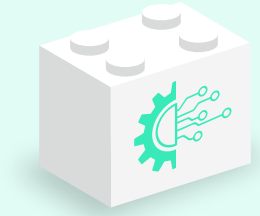
Un répertoire renseignant les sources et les bases de données produites et consolidées en interne par un acteur et potentiellement valorisables dans des partenariats ; il contient idéalement des informations précises sur les jeux de données dont des éléments d'évaluation de leur qualité.

Le recensement des usages associés

La liste la plus exhaustive possible des services, des innovations, des technologies, des types de projet qui peuvent faire l'objet de partenariats autour de l'usage de données de santé avec d'autres acteurs.
Il s'agit ici pour une structure donnée d'être capable d'identifier les types de projet qu'elle va porter.

Pour un guichet donné, **le croisement entre le catalogue de données et les cas d'usages recensés par la structure permet ainsi de définir proprement le périmètre d'action stratégique** sur lequel déployer la mise à disposition des données existantes dans le catalogue. Cela permet en fait de construire une vision stratégique des projets sur lesquels le guichet peut aider à développer des partenariats de manière efficace.

En parallèle de ces deux éléments qui vont permettre de définir la vision cible de la politique de mise à disposition des données (« quelles données pour quels usages ? »), notre travail de structuration des guichets « *fast-track* » s'est porté sur les moyens opérationnels devant servir à mettre à disposition les données (« comment mettre à disposition ? »). Nous avons classé ces moyens en trois volets :



Un volet technique

L'ensemble des outils et services dont un acteur dispose pour mettre à disposition les données contenues dans son catalogue ; cela embarque des notions liées à la constitution et à la structuration des jeux de données, à leur stockage et à leur sécurité, ainsi qu'à leur transfert (ou à la mise en place de l'accès aux bases à un acteur externe).

Un volet réglementaire

L'ensemble des outils et services dont un acteur dispose pour assurer la conformité à la réglementation, notamment au *RGPD,* et pour protéger les données mises à disposition ; cela repose sur des dimensions d'information des patients et de recueil du consentement, de contractualisation, et de contrôle sur l'usage des données réalisées par le porteur de projet.

Un volet valorisation

L'ensemble des outils et services dont un acteur dispose pour valoriser ses apports techniques, scientifiques... dans le cadre d'un partenariat s'appuyant sur des données de santé ; cela nécessite la connaissance des services fournis par le producteur de données pour un projet et la préparation d'offres type embarquant des outils spécifiques pour assurer la valorisation.

Toutes ces dimensions du guichet doivent s'articuler autour d'**un mode de fonctionnement efficace et adapté à l'organisation** du producteur de données.

La sélection et la création d'outils à exploiter dans le cadre du guichet dépendent en grande partie des usages envisagés et des bases de données présentes dans le catalogue. Pour identifier et construire ces ressources, il est possible de prioriser l'une ou l'autre de ces dimensions : définir dans un premier temps les usages souhaités et adapter le contenu du catalogue pour que les demandes d'accès aux données qui y sont recensées correspondent aux ambitions recensées dans les usages. Par exemple, il est possible d'avoir la volonté de favoriser uniquement le développement d'algorithmes d'IA pour la reconnaissance d'images et choisir d'alimenter le catalogue uniquement avec des données d'imagerie.

Une autre approche est de recenser l'ensemble des données ayant un potentiel dans le catalogue et identifier ensuite les usages et les projets qu'il est possible de réaliser à partir de ces données.

Notre expérience sur le projet PHOENIX a montré qu'il est plus aisé de définir les usages à partir

des données disponibles et des projets passés : l'exercice de projection des usages souhaités semblait moins parlant puisque la grande majorité des projets passés avaient été conceptualisés en prenant la base de données comme point de départ pour définir les modalités de partage (techniques, scientifiques, réglementaires et de valorisation). Cet exercice de recensement reste pertinent et il permettrait de mettre en place un « cercle vertueux » en recensant ces usages puis en les confrontant à ce qui a été réalisé a posteriori afin de **gagner en capacité à identifier en amont les usages possibles pour les futurs projets**. Cela peut aussi permettre d'orienter la stratégie de partage de données et de prioriser les usages les plus prometteurs.

Le croisement des bases de données disponibles avec les usages associés fait naturellement émerger une (ou des) offre(s) de services : pour tel type d'usage et tel type de projet, le producteur de données fournit un certain nombre de services à l'exploitant (de l'extraction des données brutes à la collaboration active au projet). Ces services dépendent des volets technique, réglementaire et valorisation puisqu'ils s'appuient sur les outils existant dans chacun de ces éléments. L'offre de services peut donner une orientation au guichet et des informations sur la feuille de route à déployer pour assurer ces services. Elle constitue également un levier important de négociation puisqu'elle aide à rationaliser les pratiques de valorisation (voir en partie 3.2). •••

Camille Bachot Roche Pharma

Medical Data Platform Lead

En tant qu'exploitant de données de santé, le premier besoin couvert par le guichet est le référencement des jeux de données, idéalement avec un niveau de description suffisant pour évaluer si le jeu peut couvrir les besoins du porteur de projet. Lors des ateliers PHOENIX, nous avons bien exploré ce point. Ensuite le guichet pourra faciliter la mise en relation du demandeur et du détenteur des données et accélérer les processus d'évaluation des demandes et de contractualisation.

La mobilisation des équipes intervenant sur le guichet et l'utilisation des outils recensés doit être optimisée : il convient de définir les rôles et les responsabilités de chacun – qui est en charge ? de quoi ? – et définir les étapes pour le fonctionnement du guichet :

▶ **Comment une demande est-elle initiée ? avec quel vecteur ?**
Par exemple dans PHOENIX : via un contact par mail, ou l'envoi d'une demande formalisée avec une fiche en ligne permettant au guichet d'identifier si le projet est d'intérêt vis-à-vis des objectifs de la structure, et s'il est faisable d'un point de vue scientifique.

▶ **Quelles actions cela initie et dans quel ordre ? Quelles sont celles qui doivent être effectuées en parallèle et quelles sont celles qui constituent un prérequis au lancement des phases suivantes ?**
Par exemple dans PHOENIX : lancement d'une évaluation de la conformité juridique réalisée en parallèle d'une étude de la faisabilité scientifique puis début des négociations précédant la mise en place de l'accès aux données. •••

Pour l'ensemble des contributeurs du projet, le bon fonctionnement d'un guichet repose également sur la mise en place de *process* en interne et notamment de l'adhésion des équipes du producteur de données. Cela est nécessaire pour alimenter le catalogue de données. Les besoins et les aspirations de l'ensemble des membres investis dans les sujets liés aux données doivent donc être pris en compte pour faire vivre le guichet et ils doivent être impliqués dans la gouvernance du guichet. Par exemple dans un établissement de santé, les médecins sont à la source de la production des données. Il est alors essentiel de les impliquer dans le guichet. Les médecins sont souvent demandeurs de licences gratuites pour exploiter les logiciels métier et l'utilisation de cet outil de valorisation dans les contrats peut donc être un premier pas vers l'acculturation et mener à terme à une meilleure compréhension de ces sujets et donc à leur implication dans la gouvernance.

La formalisation des guichets autour de ces volets permet également de mieux préparer la stratégie interne d'achat et d'embauche.
Par exemple la définition des standards pour une bonne qualité permet de déterminer comment il faut les structurer en amont. En outre, ce besoin de structuration entrera dans les critères nécessaires que doivent proposer les logiciels métiers.

2.3 LE GUICHET « *FAST-TRACK* » : UNE TRADUCTION DE LA VISION, DE L'ORGANISATION ET DES CONTRAINTES LIÉES AU PARTAGE DE DONNÉES ET PROPRES À CHAQUE ACTEUR

D'une part, le guichet de données de santé conçu dans le cadre du projet PHOENIX est un outil commun dans le sens où il poursuit le même objectif global et se structure autour de cinq éléments articulés conjointement : la catalogue de données, les usages associés et les volets technique, réglementaire et valorisation.

D'autre part, les échanges tenus au cours de ce projet ont montré qu'un certain nombre de solutions et d'ambitions sont partagées dans l'écosystème pour faciliter le partage de données a fortiori dans le cadre d'un guichet :

- **La volonté de mener des projets collaboratifs :** de plus en plus d'institutions souhaitent s'impliquer dans les projets impliquant l'utilisation des données qu'elles collectent, l'objectif étant de maximiser la valeur créée par ces projets et de générer des droits de propriété intellectuelle.

- **Un set d'outils partagés pour assurer la conformité juridique et accélérer les négociations :** l'utilisation d'un site internet pour assurer l'information des patients, la conservation des données dans l'infrastructure SI du producteur pour minimiser le risque d'utilisations hors cadre

des données en cas de transfert des bases (difficile car les porteurs de projet font « pression » pour obtenir l'accès total aux données), l'évaluation objectivée de la qualité des données pour assurer une connaissance commune du potentiel des *datasets* et la connaissance des coûts de fonctionnement pour collecter, structurer, stocker et partager les données.

Cependant les objectifs poursuivis par les acteurs à travers leurs guichets notamment dans leur usage sont différents. Ainsi, les modalités de fonctionnement, les outils utilisés, le niveau de mobilisation interne, les usages envisagés et les données disponibles varient pour chaque producteur de données de santé. **Il n'existe donc pas de modèle unique de guichet**.

Ce résultat s'est confirmé dans le cadre du projet PHOENIX puisque trois modèles de guichet cible ont été construits avec les partenaires du projet :

		Guichet interne à accès décentralisé	Guichet interne exploitant des solutions externes	Guichet externe
	Modèle de guichet	Guichet basé sur le développement en interne d'outils et de *process* de partage autour d'un *datalake* : procédure décentralisée mobilisant plusieurs unités avec une complémentarité d'expertises pour gérer en parallèle les dimensions visant à la mise en place de partenariats.	Guichet structuré autour d'une petite équipe en interne qui gère l'ensemble des dimensions pour mettre en place les partenariats, soutenue par la structure d'entrepôt de données de santé et d'une « boîte à outils » constituée de solutions développées par des prestataires.	Guichet basé sur l'externalisation de la gestion du partage de données : volets technique et réglementaire assurés par un tiers (mise en conformité, transfert du jeu de données, garantie de la qualité, contrôle des traitements réalisés).
Caractéristiques	Catalogue de données	**Très important**, visibilité complète pour alimenter le *datalake*.	**Très important**, visibilité complète des données existantes.	**Non essentiel** : catalogue alimenté par les nouveaux *datasets* structurés.
	Usages	*Market pull* et *techno push*. Objectif de projets collaboratifs.	*Market pull* avec gestion des demandes spécifiques. Objectif de projets collaboratifs.	Mise à disposition de jeux de données, *market pull* sur des données existantes.
	Technique	*Datalake* pour le stockage et la réalisation des traitements de l'exploitant.	À construire, infrastructure *IT* pour accéder et structurer les *datasets*, volet soutenu par des solutions externes.	Externalisé.
	Réglementaire	Juristes tout au long de la chaîne data / projet.	Information patient au plus tôt soutenue par un site internet.	Externalisé / utilisation d'un site internet pour assurer l'information des patients.
	Valorisation.	Modèles à définir / cellule de valorisation déjà en place / volonté d'obtention de droits de propriété intellectuelle / utilisation d'une offre de services.	Objectif d'obtention de droits de PI, outils s'appuyant sur une offre de services et une grille de coûts.	Gratuité et à terme : contrat-type visant surtout à couvrir les coûts *IT*.

Les différences principales entre les modèles de guichets construits dans le cadre du projet PHOENIX sont les suivantes. Il s'agit des dimensions permettant de décliner et de moduler un guichet pour qu'il s'intègre pleinement dans l'organisation interne d'une institution productrice de données de santé :

▶ **L'organisation interne, notamment le niveau d'implication des équipes amenées à interagir avec le guichet à tous les niveaux.**

Il s'agit de définir qui alimente le catalogue de données, qui détermine les usages, qui met en place et opère les solutions techniques visant à extraire, structurer et partager les jeux de données, qui vérifie et met en place les conditions nécessaires à la conformité juridique et qui mène les négociations visant à assurer la valorisation. La distinction majeure observée avec PHOENIX est le niveau de centralisation qui varie entre un cas tout à fait centralisé où une équipe resserrée coordonne tous ces éléments et un cas décentralisé où les producteurs (ou collecteurs) de données interviennent et sont consultés à toutes les étapes de la mise en place d'un accord de partage.

▶ **Le niveau d'externalisation et / ou d'utilisation de solutions externes proposées par des prestataires.**

Il est possible de développer et d'opérer en interne les outils et des services à intégrer sur les volets technique, réglementaire et valorisation. Il est également possible de faire appel à des sociétés externes qui proposent des solutions pour opérer une partie – ou l'ensemble – du guichet, de la structuration de bases de données jusqu'au partage effectif du jeu de données en passant par la prise en charge de la conformité juridique.

▶ **L'organisation et la structure juridique.**

La structure d'entrepôt de données de santé permet de faciliter un certain nombre de démarches réglementaires. La mobilisation de juristes voire leur intégration à l'équipe responsable du guichet et l'acculturation autour des problématiques réglementaires peut également permettre l'accélération de certaines démarches.

▶ **Les outils et les solutions disponibles et leur niveau de formalisation.**

Sur ce point il est possible de construire un set d'outils et de *process* parmi lesquels il sera possible de « piocher » pour adapter le package choisi à chaque type d'acteur et à chaque type de projet, permettant ainsi de proposer toujours une solution modulable. Les éléments sur mesure créés si besoin pour de nouveaux projets viendront compléter ces éléments existants à noter : il s'agit bien ici de s'appuyer sur des outils existants pour rester flexible mais d'éviter le cas par cas sur chaque nouveau projet qui demanderait de créer systématiquement des outils sur mesure.

▶ **Les *process* de fonctionnement.**

Il s'agit de définir le mode d'accès au guichet et le parcours qu'une demande de partage de données va suivre pour aboutir.

Il est possible d'externaliser la contractualisation par un prestataire externe (en conservant en interne un comité qui prendra la décision d'accepter ou non le partage de données) et il est également possible de gérer cet aspect en interne avec une équipe dédiée à contacter via une adresse mail générique ou un site internet en remplissant une fiche type pour décrire le projet. Les étapes suivantes dépendront des services proposés dans le cadre du guichet (exemple : étude de la faisabilité du projet ; voir partie 3.2).

Camille Bachot Roche Pharma

Medical Data Platform Lead

En tant que laboratoire pharmaceutique, Roche exploite des données de santé pour répondre à des questions de santé publique, accompagner les innovations technologiques, affiner la compréhension des maladies... En tant qu'utilisateur régulier de données de santé, Roche apprécie la mise en place de guichets structurés, de contrats de partage. Cela accélère grandement le démarrage des projets.

Sebastian Schwarz OWKIN

Data Platform Strategy Lead

Les guichets construits dans le cadre du projet PHOENIX pourraient constituer un facilitateur de taille dans la construction de projets collaboratifs autour de la valorisation de la donnée de santé, en permettant d'accélérer les processus de négociation et contractualisation qui représentent souvent un frein important à la mise en place de ces projets. OWKIN serait typiquement intéressé par ce guichet !

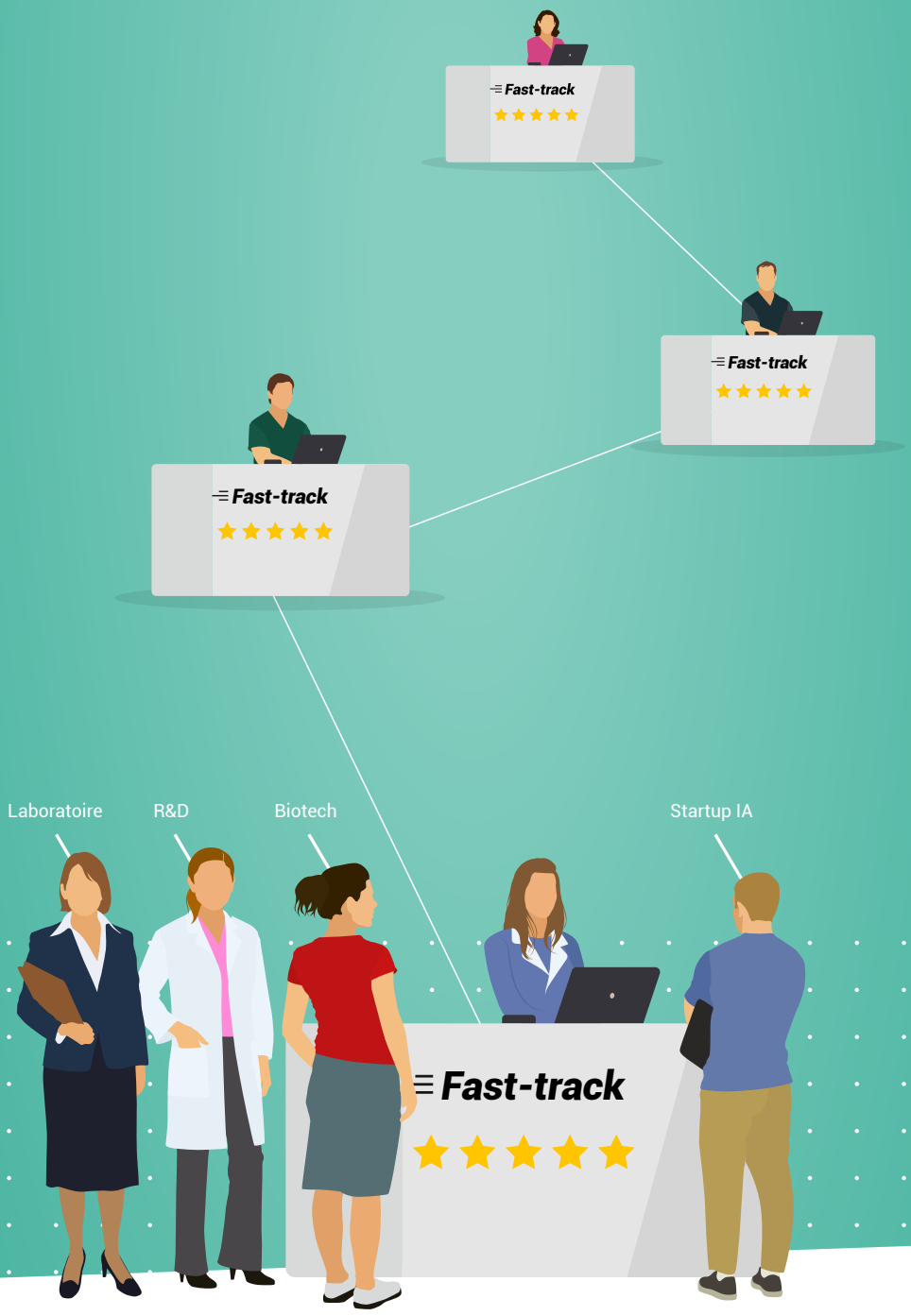
Antonio Borderia PhD. ADLIN Science

Executive Vice-President Business Development

Un atout majeur qu'apportent les guichets est la sensibilisation et l'embarquement des équipes internes à différents niveaux (clinique, juridique, stratégique...). C'est déjà une avancée conséquente puisque cela va accélérer le déploiement des solutions qu'ADLIN peut proposer : cette étape aura déjà été faite via le guichet.

3 Un déploiement des guichets possible pour tout acteur de l'écosystème

en appliquant plusieurs bonnes pratiques



Cette partie restitue les enseignements identifiés au cours du projet. Les bonnes pratiques ont été testées et / ou approuvées par l'ensemble des acteurs du groupe de travail et elles seront évaluées au cours de l'appel à projets puis lors du fonctionnement en routine des guichets une fois complètement déployés. La démarche employée dans le cadre du projet a aussi porté ses fruits : le suivi et l'animation conjointe du dispositif projet par Medicen et Veltys ont permis d'assurer le dynamisme du projet et le partage des réflexions communes ou spécifiques à chaque partenaire. En effet, le mode de travail employé a prouvé qu'il est possible de commencer à déployer des guichets malgré les nombreux verrous identifiés (voir partie 1.4). Les partenaires du projet se sont pleinement engagés et se sont placés dans une posture d'échange et une volonté d'avancer qui a permis d'atteindre ce résultat.

Des temps d'échange collectifs ont d'abord permis aux acteurs de se rencontrer et de mettre en place les liens de confiance nécessaires au partage des expériences de chacun. Nous avons ensuite organisé des ateliers individuels avec chaque acteur pour mener une réflexion spécifique et travailler avec chacun sur son guichet et sa feuille de route. Nous avons aussi échangé au cours de réunions avec l'ensemble du groupe afin de traiter des problématiques communes, parfois en réunissant les partenaires en plénière et parfois en mobilisant des acteurs externes au projet (retour d'expérience de la clinique des données du CHU de Nantes, pitch *start-ups*, échange avec la CNIL, etc.). Ces ateliers étaient essentiels pour partager les réflexions et les travaux menés, identifier les problématiques et les solutions communes, et faire émerger de nouvelles idées via le dialogue entre des acteurs différents qui ne sont pas amenés à travailler régulièrement ensemble autour des mêmes sujets.

La force du projet PHOENIX réside en la réunion de ce consortium d'acteurs pleinement engagés et en une animation intelligente et efficace pour mener une réflexion individuelle avant de créer des solutions collectivement. À ce sujet, il faut noter que l'intérêt des directions des différentes structures a grandi au fil de l'avancement du projet.

Julie Baussand Medicen

Chef de Projets Innovation Santé – Responsable du Medicen Initiatives for Health Data

Les leçons sont nombreuses mais la plus importante est surement la nécessité de travailler collaborativement. L'ensemble des acteurs de la chaine de valeur doivent se mobiliser, bien évidemment en interne mais aussi avec des acteurs externes (potentiels porteurs de projet, CNIL, fournisseurs de technologie, ...). Identifier clairement les objectifs du guichet et associer les divers acteurs dans la définition, la mise en place et le pilotage de celui-ci sera primordial pour initier un tel chantier et s'assurer qu'il corresponde bien à un outil pertinent pour tous. Les sujets à aborder pour constituer un guichet de partage de données sont nombreux et pour certains en constante évolution. Un pilotage en continu des guichets (via des indicateurs de performances (KPI) par exemple) sera nécessaire pour viser l'excellence d'un outil efficient, pertinent et de confiance.

3.1

CONDUIRE UNE RÉFLEXION SUR L'EXISTANT (STRUCTURE, *PROCESS* ET VERROUS) ET SUR LA CIBLE SOUHAITÉE

Le projet PHOENIX a permis de mettre en déploiement trois guichets en moins d'un an pour des acteurs très différents. Avant de passer à l'implémentation de ces solutions, il a été nécessaire de mettre à plat les modes de fonctionnement déjà en place et d'identifier les bonnes pratiques et les difficultés bloquantes et récurrentes. À partir de cette cartographie de l'existant, nous avons défini les modèles cible des guichets ainsi qu'une feuille de route avec les partenaires : notre idée était de déterminer la cible à atteindre à long terme et de prioriser les chantiers les plus importants à court terme. Dans le cadre du projet PHOENIX, nous souhaitions en effet construire les bases de guichets efficaces, mais ils sont amenés à évoluer pour atteindre le modèle-cible. Des *KPI* seront essentiels à définir et à suivre pour assurer un pilotage efficient.

La force du projet PHOENIX pour cette première étape de construction des guichets a été de réunir un consortium d'acteurs avec des profils d'acteurs clinique, académique et industriel. Sur chacun des éléments évoqués ci-dessus, le partage des objectifs, des solutions existantes et des contraintes et difficultés permet de satisfaire en partie le besoin d'acculturation : chacun a pu alors prendre conscience des contraintes des autres acteurs et comprendre qu'il est possible de limiter les retards et la complexité à réaliser des accords de partage de données de santé en travaillant à plusieurs. Différentes solutions ont été partagées notamment en termes d'articulation de *process* entre les acteurs. •••

Pierre-Antoine Gourraud CHU de Nantes

Professeur des universités et praticien hospitalier en biologie cellulaire

Il faut placer la gouvernance de la donnée au centre des pratiques et y apporter les moyens IT, logiciels et humains à la hauteur du potentiel de transformation des données. En effet, la réutilisation des données touche l'intimité des patients comme celles des professionnels de santé (intimité pathologique et intimité des pratiques), ces deux acteurs doivent donc être représentés dans la gouvernance.

Par ailleurs, il y a aujourd'hui une naïveté laissant penser que les données sont immédiatement prêtes à être réutilisées par les data scientists, or elles ne le sont pas, souvent pas du tout, et un gros travail est nécessaire par les data engineers pour structurer et préparer les données.

Au cours de cette étape de réflexion, nous avons donc travaillé sur les points suivants :

1) Cartographie de l'existant

Mettre à plat l'ensemble des pratiques et des outils à disposition pour réaliser des partages de données de santé a été essentiel à la mise en déploiement de guichets. Une fois le modèle-cible construit, cela a permis de mesurer le travail à réaliser pour l'atteindre. Déterminer ce « point de départ » a aussi été utile pour prioriser les chantiers : ce choix doit prendre en compte la maturité à date et l'écart à la cible.

2) Identification des difficultés bloquantes en interne

À travers cette revue des outils et des pratiques existantes, il était utile d'identifier l'ensemble des difficultés qui vont ralentir ou empêcher les partages de données. Il s'agissait ensuite d'identifier celles qui dépendent de l'organisation et / ou des outils en interne, c'est-à-dire celles sur lesquelles il est possible d'agir en autonomie, et celles pour lesquelles des propositions au niveau de l'écosystème peuvent être faites à partir du travail collaboratif avec les trois partenaires.

3) Définition du modèle-cible

Nous avons ensuite défini le modèle-cible de chaque acteur en repartant de l'existant et des difficultés : il a fallu déterminer quel fonctionnement et quelle articulation étaient souhaités pour construire et opérer le guichet. Pour cela, nous avons évalué à quel point les solutions et les *process* déployés étaient satisfaisants à date et quels étaient ceux qu'il convenait de faire évoluer pour atteindre un fonctionnement optimal.

4) Élaboration d'une feuille de route

Une priorisation des différents chantiers a ensuite permis d'établir une feuille de route. Nous avons pris en compte l'écart entre la maturité à date des différents éléments constitutifs des guichets et la cible à atteindre pour prioriser. Plus spécifiquement, un producteur de données pour qui les éléments existants à date dans les volets techniques et réglementaires sont très matures au regard de sa cible devra concentrer ses efforts sur les enjeux liés à valorisation. Une répartition des responsabilités a été réalisée avec les partenaires du projet pour assurer un déploiement efficace des feuilles de route.

Pour suivre ces différentes étapes il nous semblait pertinent de procéder de façon structurée en étudiant un à un chaque élément du guichet avant d'étudier leur articulation globale. Il peut également être intéressant d'évaluer des indicateurs de maturité sur ces différents éléments dont une liste est proposée ci-dessous (partie 3.3).

Grille d'évaluation visant à cartographier les *process* et les outils existants :

Élément constitutif du guichet	Notions à étudier
Catalogue de données	▶ Dans quelle mesure les données disponibles sont-elles connues ? ▶ Quel est le niveau de standardisation / d'harmonisation de vos jeux de données ? ▶ Êtes-vous en mesure d'évaluer la qualité des jeux de données ?
Usages	▶ Avez-vous une bonne connaissance des types de projet et des types d'acteur sur lesquels et avec qui vous pouvez / aimeriez travailler autour de ces sujets ? ▶ Vos partages de données répondent-ils à des demandes extérieures ou embarquent-ils également des cas où vous avez recherché vous-même des acteurs pour exploiter certaines données ? ▶ Êtes-vous en mesure d'identifier des usages secondaires pour les bases de données constituées ?
Volet technique	▶ Comment constituez-vous les bases de données ? ▶ Quelle(s) solution(s) utilisez-vous pour stocker et sécuriser vos données ? ▶ Comment mettez-vous en place l'accès à vos *datasets* ?
Volet réglementaire	▶ Comment est réalisée la mise en conformité des données (anonymisation / données synthétiques / information patient) ▶ Comment se déroule la contractualisation pour les accords de partage de données de santé ? ▶ Êtes-vous en mesure de contrôler les traitements réalisés par les exploitants de données ?
Volet valorisation	▶ Quels outils et solutions utilisez-vous pour valoriser les données ? ▶ Quels sont vos leviers de négociation sur la partie valorisation ?

Paul Rinaudo ADLIN Science

CEO & Co-founder

Il faut comprendre que l'aspect technique n'est pas un problème en tant que tel, il s'agit d'une question d'approche. Il y a des moyens d'utiliser les données en évitant ces problématiques techniques. À ce stade les exploitants souhaitent accéder aux données pour obtenir des résultats agrégés ou pour entraîner un algorithme par exemple alors qu'il serait possible de procéder via des requêtes sur un espace de travail interne au producteur. En identifiant les besoins véritables dans l'exploitation des données, on se rendra compte que la technique est moins importante que l'approche.

Pourtant il existe des solutions mais sauf rares exceptions, les producteurs ne sont pas suffisamment structurés pour aller chercher ces solutions, les intégrer et les rendre accessibles. Par ailleurs, il y a une hétérogénéité forte dans l'écosystème au niveau de la maturité technique et organisationnelle sur ces sujets (les équipes ne sont ni assez formées ni assez structurées pour valoriser les données).

D'autre part, le modèle historique utilisé pour recueillir les données a été l'agglomération des données qui a bien un but statistique et analytique mais qui ne répond pas aux besoins. Or une base de données a de la valeur si elle est contextualisée et il y a donc plus de valeur à structurer plusieurs bases contextualisées au cas par cas plutôt qu'une très grande agglomération. Il ne faut pas rassembler des données sans usage et toujours partir d'une question clé : à quelle question biologique on veut répondre ?

3.2 IDENTIFIER DES SOLUTIONS OPÉRATIONNELLES, DISPONIBLES EN INTERNE COMME EN EXTERNE

Au cours du projet PHOENIX nous avions identifié un panel large de solutions opérationnelles pour chaque élément constitutif du guichet en interne. Ces solutions peuvent embarquer :

- **Des outils techniques, des ressources et des *process* construits en interne**

Chaque producteur de données peut capitaliser sur son expérience et sur les outils dont il dispose déjà pour ajouter des fonctionnalités et déployer des ressources et des *process* visant à structurer le guichet. L'expérience du projet PHOENIX montre qu'un certain nombre de solutions existent déjà dans les différentes structures et qu'une partie du travail de construction des solutions internes consistera en des changements au niveau de l'articulation et/ou du périmètre d'application des solutions existantes.

Au-delà des solutions opérationnelles à mettre en place en interne, il est essentiel d'assurer l'embarquement des équipes amenées à alimenter et opérer le guichet. Cela est primordial notamment pour une organisation décentralisée qui dépend directement des équipes pour alimenter le catalogue de données et déterminer les usages associés. Pour cela, les acteurs ont évoqué la possibilité de prévoir des bulletins de communication pour partager les bonnes pratiques et informer au sujet des résultats d'intérêt des projets impliquant des partages de données de santé. Ce type d'action permettrait aussi de sensibiliser à la réutilisation

des données et à l'identification du potentiel de réutilisation dès le processus de validation d'un projet. Pour soutenir cet objectif, des formations spécifiques – courtes mais permettant un bon partage de l'information – peuvent également être développées et proposées dans les modules d'*onboarding* de nouveaux collaborateurs.

▶ Des solutions proposées par des acteurs externes

Un grand nombre de *start-ups* proposent des solutions et / ou des accompagnements spécifiques pour soutenir le partage de données de santé. Sur la base des freins identifiés dans les sessions de cartographie, une sélection de solutions pouvant répondre à ces besoins a été proposée aux partenaires pour organiser un échange. Ainsi deux sessions de pitch ont été organisées au cours du projet afin d'obtenir une vision large des solutions existantes. La plupart des offres étudiées dans le cadre du projet se répartissent entre des outils de structuration ou de mise en conformité des données et des solutions portant sur une palette plus large du partage de données : de la constitution des bases au suivi du projet en passant par la contractualisation et la mise en place de l'accès au(x) *dataset*(s).

Un enjeu important à prendre en compte en cas d'utilisation de ces solutions est la question de l'intégration de ces solutions dans le guichet : les différentes *start-ups* peuvent avoir un rôle de prestataire ou de véritable partenaire. Ces questions sont importantes notamment pour assurer la gouvernance des données et trouver le bon niveau d'infogérance.

▶ Des solutions ou des besoins devant être pris en charge au niveau institutionnel

Le projet PHOENIX a mis en évidence une convergence d'opinions autour de solutions à développer au niveau institutionnel au cours des échanges réunissant l'ensemble des partenaires du projet. Certains verrous identifiés montrent des besoins communs. La manière d'y répondre est parfois indépendante des acteurs de l'écosystème et doit se faire au niveau institutionnel. À titre d'exemple les structures comme le *HDH* ou la CNIL qui font figure d'instances de premier plan autour de ces sujets doivent permettre de structurer davantage l'écosystème et les pratiques pour faciliter le partage de données. La réflexion commune du consortium d'acteurs mobilisé dans le projet a prouvé le potentiel d'une mobilisation de l'écosystème pour être force de proposition auprès des instances. ●

Linda Nait-Kaoudjt Institut Pasteur

Head of Technology Transfer and Industrial Partnerships

Afin que les projets collaboratifs de données se déroulent plus facilement et rapidement, il sera essentiel que les structures impliquées dans un projet partagent une ambition commune et que chacune ait défini les « bons interlocuteurs », car ce type de projet nécessite des expertises spécifiques. Il serait bon également que des guidelines communes soient établies pour atteindre un certain niveau d'harmonisation, par exemple sur les formats de génération des données ou sur des critères d'évaluation de leur qualité. C'est ainsi, avec la fluidification de l'élaboration d'un projet, que nous pourrons booster la recherche par des projets collaboratifs de data sharing.

Pauline Touche Hôpital Foch

Responsable adjointe de l'unité data

De la part des porteurs de projets, j'attends une meilleure définition de leurs besoins quand ils nous contactent, ainsi que de la patience. La donnée de qualité n'est pas disponible en un claquement de doigt. Il faut réaliser une étude de faisabilité, contractualiser, informer les patients concernés, construire la base de données, contrôler sa qualité, etc.

Catalogue de données

Identification des données disponibles

INTERNE

Alimentation autonome du catalogue par les équipes qui collectent et / ou réunissent les données. C'est un mode de fonctionnement s'inscrivant plutôt dans le cadre d'une organisation décentralisée et qui demande une mobilisation et une implication fortes des équipes pour assurer la richesse du catalogue.

INTERNE

Travail d'alimentation du catalogue au fil de l'eau avec les jeux de données constitués pour les projets internes et ceux répondant aux demandes spécifiques d'exploitants.

INTERNE

Cartographie des bases de données existantes à mener en plusieurs étapes :

1/ Évaluation générale avec des référents de chaque département / service pour obtenir une vision globale puis

2 / Priorisation basée sur ces informations générales pour identifier précisément les données et alimenter le catalogue.

EXTERNE INSTITUTIONNEL

Certains acteurs ont également émis le souhait que les instances puissent soutenir et / ou créer un espace centralisé où les producteurs de données pourraient déposer des informations sur leurs jeux de données (un catalogue de données centralisé par exemple au niveau du *Health Data Hub*).

INTERNE EXTERNE

Prise de références auprès de structures semblables disposant d'un catalogue pour déterminer ce qui pourrait être fait en interne pour le catalogue.

Évaluation de la qualité des données

EXTERNE INSTITUTIONNEL

Utilisation d'outils se basant sur des critères prédéfinis et modulables ; ces critères peuvent être modifiés selon le type de données ou selon le projet : ce type d'outil peut donc à la fois servir à évaluer la qualité a priori des données (en amont d'un projet, dans le catalogue) et à évaluer leur qualité et/ou pertinence pour un projet spécifique (lors des négociations).

EXTERNE

Production de données synthétiques (voir le volet réglementaire).

INTERNE

Construction d'une fiche descriptive type d'un jeu de données qui embarque les métadonnées à intégrer (dictionnaire de variables, statistiques sur certaines variables, etc.).

Audits internes réguliers pour tester les erreurs et les incohérences entre la donnée source et les jeux de données structurés.

Évaluation de la qualité des données par l'exploitant après le partage pour intégrer cette information pour les futurs porteurs de projet.

Catalogue de données

Arthur Souletie Veltys

Associé du cabinet en charge des projets santé

L'évaluation de la qualité d'un dataset constitue un enjeu majeur pour l'écosystème. Déterminer avec précision et justesse le contenu d'un jeu de données, analyser la qualité des informations fournies et attester de leur véracité sont autant d'exigences difficiles à satisfaire et qui peuvent s'avérer bloquantes dans les partenariats. Des approches et des outils se développent pour permettre à la fois aux producteurs de données et à leurs partenaires de réaliser ce travail qui doit servir de garantie et de préalable aux discussions autour des données.

On peut par exemple penser à l'outil VIV-T que Veltys a développé dans le cadre de l'étude conduite pour AVIESAN en 2020 qui permet, suivant un set de critères spécifié, de noter un jeu de données et sa qualité intrinsèque ainsi que sa pertinence relative dans un projet de collaboration donné. Plus largement, des initiatives au sein de l'écosystème voient le jour pour pallier ce manque et appellent une réflexion commune sur la définition de standards de qualité reconnus par tous les acteurs pour faciliter et fluidifier les discussions autour du partage des bases.

Standardisation des bases de données

INTERNE EXTERNE

Utilisation d'un format de données standard type *FHIR* ou *OMOP* ; il semble que le second choix soit de plus en plus usuel dans l'écosystème.

INTERNE

Détermination de critères spécifiques et de *guidelines* pour structurer les données et le cas-échéant assurer une cohérence entre les différents jeux de données du catalogue ; la prise en compte des critères de qualité peut permettre de définir des standards ou des références pour établir ces *guidelines*.

Sebastian Schwarz OWKIN

Data Platform Strategy Lead

Sur la mise en qualité de la donnée pour la rendre exploitable et interopérable, un enjeu important réside dans l'uniformisation de la mise en forme des données. En effet, même si plusieurs entités arrivent à partager leurs données de manière sécurisée, celles-ci seront très difficilement exploitables si elles ne sont pas mises en forme sous un même format. Heureusement, de nombreux langages de données communs (ou Common Data Language, CDM), ont vu le jour ces deux dernières décennies, comme OMOP ou FHIR, et ont réussi à s'imposer partiellement dans l'écosystème de la santé. Les processus de conversion de données brutes en données standardisées sont donc de plus en plus importants dans le cadre des collaborations.

Les futurs enjeux sur le sujet se situent à deux niveaux : tout d'abord réussir à évangéliser ces modèles de données qui ne sont pas encore utilisés universellement et développer les outils techniques qui leur sont associés (incluant la possibilité de réaliser un catalogue de ces données in fine). Deuxièmement, enrichir le spectre des modalités de données couvertes par ces modèles, avec notamment un besoin pressant d'intégrer des données issues de l'imagerie médicale et de la génomique afin de pouvoir baser les collaborations sur des données plus riches.

Usages

Détermination des types de projet

INTERNE

Mobiliser les équipes responsables de la contractualisation, de la mise en place des accords de partage et du suivi des projets pour lister et définir les types de projets auxquels le producteur de données a participé et / ou souhaite participer (ex : académique ou solution industrielle, collaboratif ou « simple » mise à disposition de données, avec de grandes entreprises ou des *start-ups* etc.) .

Réalisation de *market pull / techno push*

INTERNE

Mettre en place une étape d'évaluation du potentiel de réutilisation des bases de données existantes et réaliser une veille en vue de mener des *techno push*.

Utilisation secondaire des bases de données

INTERNE

Réaliser l'information patient et le recueil du consentement au plus tôt pour les nouveaux projets.

Volet technique

Constitution des bases de données

INTERNE

Capitaliser sur le système d'informations existant et monter des projets avec les équipes techniques visant à favoriser la constitution de jeux de données structurés en revoyant des éléments du SI (de la saisie des données au stockage).

EXTERNE

S'appuyer sur des solutions externes pouvant se « brancher » sur le SI pour favoriser l'interopérabilité : de nombreuses solutions existent notamment pour les établissements de santé.

Antonio Borderia PhD. ADLIN Science

Executive Vice-President Business Development

Il y a quelques années, beaucoup d'acteurs intervenaient en fin de chaîne sur les algorithmes ou solutions d'analyse des données. Aujourd'hui, on voit que de plus en plus d'acteurs se sont créés et mobilisés en amont de l'analyse des données pour s'intéresser aux solutions rendant possible cette analyse à la base : anonymisation, enjeux juridiques pour être en mesure de partager les données, sécurité et cryptographie, production de données synthétiques...

Volet technique

Stockage et sécurité des données

INTERNE EXTERNE

Déterminer le type de solution souhaitée : *cloud* vs *on premise*.

S'appuyer sur les structures existantes : entrepôt de données de santé, *datalake* où l'ensemble des données seraient stockées (structurées ou non).

Mise en place de l'accès aux *datasets*

INTERNE

Mettre en place des accès à des espaces de travail numériques sécurisés ; il peut alors être nécessaire de prévoir suffisamment de puissance de calcul.

EXTERNE

Utiliser une plateforme en mesure de proposer un « sas » sécurisé où le producteur peut déposer les données et le porteur de projet les exploiter sans avoir le contrôle total ; plusieurs *start-ups* proposent ce service.

INTERNE EXTERNE

Utiliser une solution évitant un transfert des données vers l'exploitant.

Sebastian Schwarz OWKIN

Data Platform Strategy Lead

Sur les systèmes et usages pour mettre à disposition les données, le partage de données de santé implique un déplacement de la donnée vers un tiers qui la rend vulnérable et risque de compromettre sa confidentialité. L'un des principaux challenges réside donc dans le moyen de sécuriser son transit et son exploitation multipartite, à tous les niveaux. Des investissements conséquents sont nécessaires à la fois dans les systèmes d'informations, mais aussi dans les processus de communication et d'authentification qui peuvent faciliter une collaboration requérant le partage de données de santé. Des approches récentes permettent de minimiser les échanges de données, soit en exploitant les données localement, soit en travaillant sur des données chiffrées.

Ainsi, l'apprentissage fédéré est une méthode qui permet d'entraîner des modèles d'intelligence artificielle de manière décentralisée en faisant transiter le modèle plutôt que la donnée elle-même, ce qui permet de la garder dans un lieu sécurisé. Le modèle peut ainsi apprendre sur différents datasets dans une configuration dite multicentrique, avec des performances similaires à un entraînement "centralisé", et sans jamais exposer la donnée. Le chiffrement homomorphe est une autre méthode, qui consiste à appliquer des calculs directement sur la donnée chiffrée, n'y donnant aucun accès direct.

Maxime Agostini Sarus Technologies

Président & cofondateur

Il y a désormais consensus sur le fait que la donnée médicale peut servir à faire progresser la science plus vite mais qu'il faut pour cela arriver à rassembler plus de données et de meilleure qualité. L'arrivée à maturité de l'écosystème est un processus long qui passe par l'informatisation de la collecte et du stockage de données, la normalisation des informations collectées, la standardisation des formats de fichier et de communication... C'est nécessairement un processus long mais tous les maillons de la chaîne y contribuent et la progression est impressionnante.

Mise en conformité des projets et des données partagées

INTERNE

Mettre en place un site d'information pour les patients sur l'usage de leurs données.

EXTERNE

Utiliser des solutions pour analyser les données sans y avoir accès (envoi d'un algorithme et récupération des résultats).

EXTERNE

Utiliser des données synthétiques plutôt que les données réelles. Un certain nombre de *start-ups* proposent cette solution.

Réaliser l'anonymisation des données ; c'est aussi une solution proposée par plusieurs *start-ups*.

INTERNE

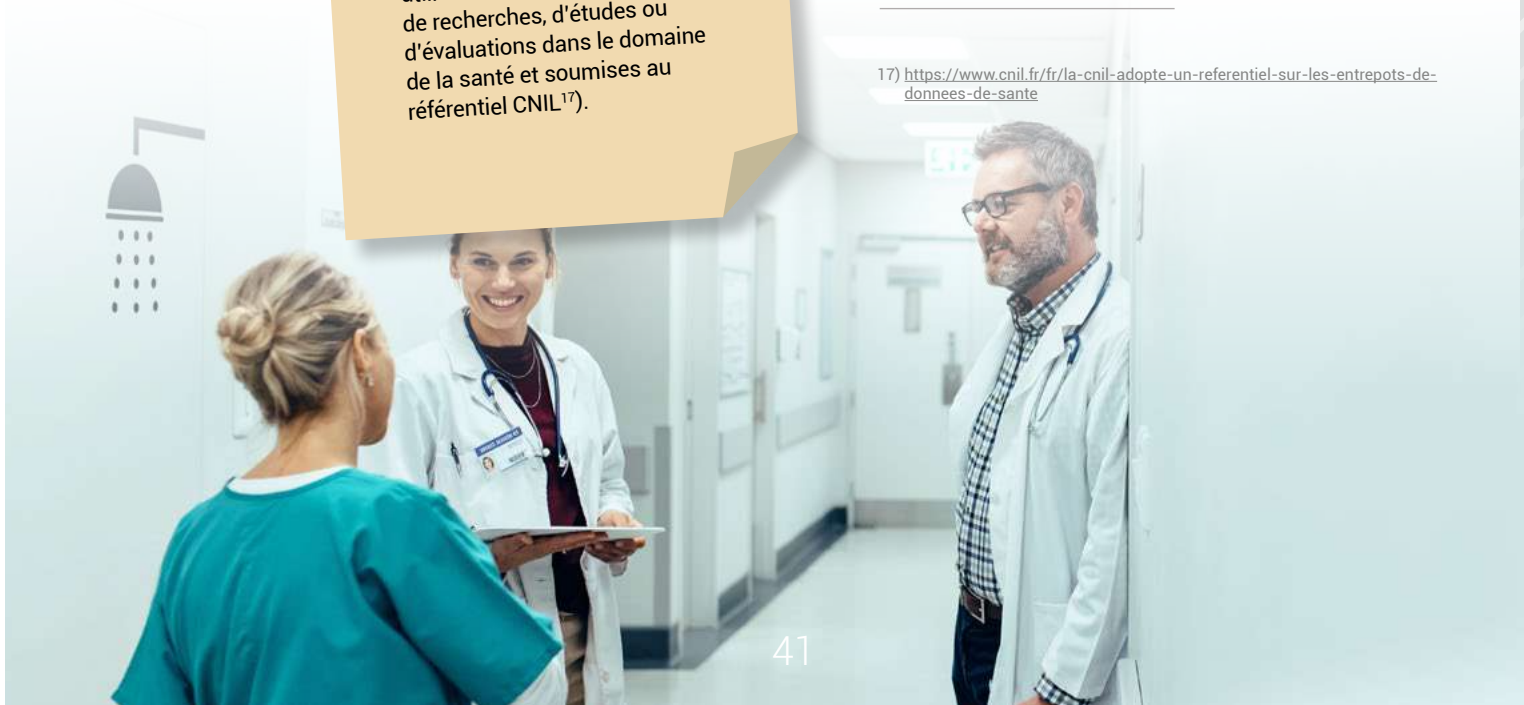
Mettre en place une structure juridique d'entrepôt de données de santé (EDS, entendu ici comme l'ensemble des bases de données destinées à être utilisées notamment à des fins de recherches, d'études ou d'évaluations dans le domaine de la santé et soumises au référentiel CNIL[17]).

Pauline Touche Hôpital Foch

Responsable adjointe de l'unité data

Concernant les start-ups qui travaillent sur la création de données synthétiques, d'avatar, je tiens à les encourager dans leurs développements. Leurs solutions techniques nous permettent d'anonymiser les données de nos patients et d'ainsi mieux les protéger. Cependant il reste du chemin à parcourir, car ces nouvelles catégories de données n'ont pas encore acquis leurs lettres de noblesse dans la communauté scientifique, et les porteurs de projets préfèrent toujours pour le moment de la données «brutes» (non transformées).

17) https://www.cnil.fr/fr/la-cnil-adopte-un-referentiel-sur-les-entrepots-de-donnees-de-sante



Volet réglementaire

Les trois producteurs de données s'accordent également pour considérer que des solutions mises en place au niveau institutionnel seraient également très utiles voire nécessaires pour favoriser le partage de données :

INSTITUTIONNEL

Obtenir des *guidelines* précises et traduites pour permettre aux porteurs de projet (notamment européens) de se mettre en conformité plus simplement.

Créer un site centralisé permettant de gérer une information patient au niveau national (par exemple au niveau du *Health Data Hub*).

Contractualisation

INTERNE

Préparer des contrats type répondant aux croisements entre les différents usages identifiés et les bases de données disponibles dans le catalogue.

INTERNE

Impliquer les juristes aux différentes étapes du parcours d'un porteur de projet au sein du guichet.

Contrôle des traitements réalisés sur les données

INTERNE

Prévoir la possibilité de réaliser ces contrôles dans les contrats.

INTERNE EXTERNE

Utiliser préférentiellement une solution permettant d'auditer l'accès aux données (contrôle des log).

Camille Bachot Roche Pharma

Medical Data Platform Lead

Que chaque acteur de l'écosystème ait une volonté de partager les données qu'il exploite. Un guichet unique, potentiellement géré par un organisme public type HDH ou autre, pourrait aussi constituer une vraie opportunité pour faciliter le partage des données entre tous les acteurs.

Leviers de négociation

Les offres de services construites par chacun des acteurs dans le cadre du projet PHOENIX embarquent toutes des **services inclus par défaut** pour tout type de partage de données (de la mise à disposition de données de santé à un projet collaboratif très ambitieux mobilisant les ressources du producteur de données) et des **services additionnels** qui seront assurés selon le projet et qui dépendent donc du croisement entre les bases de données du catalogue et les usages associés.

Elles se décomposent en quatre grandes catégories :

- Les services visant à l'étude de la faisabilité du projet sur les plans scientifique, technique et réglementaire
- Les services assurant la mise à disposition de jeux de données exploitables
- Les services de description et d'évaluation de la qualité des *datasets*
- Les services de suivi et de soutien à la réalisation des projets

Au sein des partenaires du projet PHOENIX, la formalisation des offres de services a été très efficace d'abord auprès des acteurs eux-mêmes puisque cela leur a permis de prendre conscience de la richesse des services qu'ils offrent ou peuvent offrir à des porteurs de projet. Tous sont également alignés sur le fait que rationnaliser ces services et leurs coûts permettra véritablement d'appuyer le travail mené au cours des négociations en utilisant des éléments objectivés grâce à ces outils.

INTERNE

Formaliser une offre de services modulaire définissant l'ensemble des services proposés via le guichet dans le cadre d'un projet d'exploitation des données de santé.

INTERNE

Construire une grille de coûts associés à chacun des services définis dans l'offre de services : l'objectif est d'obtenir une connaissance fine des coûts de chaque service puisqu'il conviendra a minima de les couvrir avec les différents outils de valorisation disponibles ; il conviendra d'étudier les coûts techniques de structuration et de stockage des données notamment et de recenser les coûts RH par exemple via des feuilles de temps. Par ailleurs, il s'agit d'un facteur de confiance avec les porteurs de projet qui pourront rationnaliser le coût demandé pour accéder aux données.

Volet valorisation

Outils de valorisation

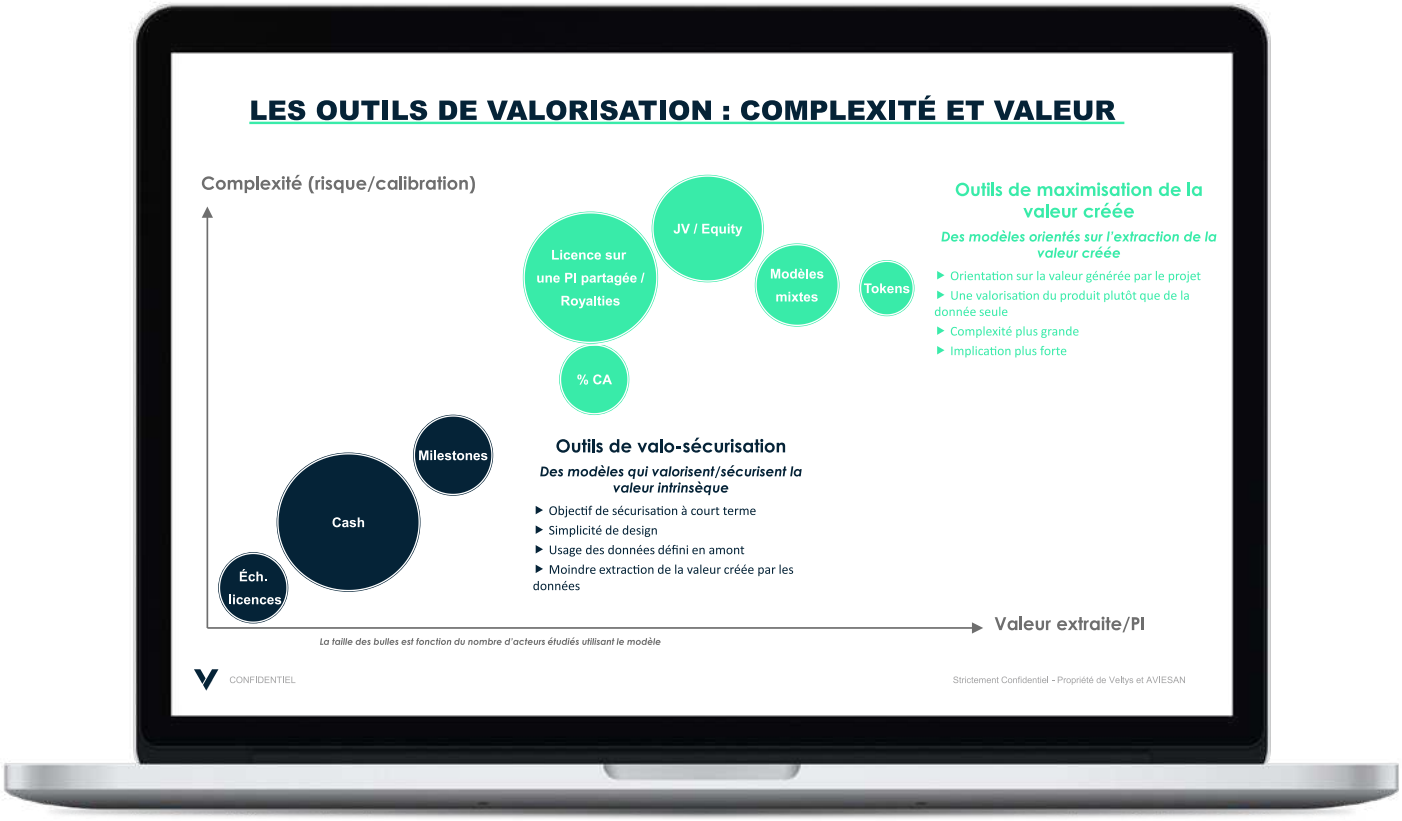
Les données de santé embarquent une **valeur intrinsèque générale** et une **valeur ajoutée spécifique** à chaque cas d'usage. Il est nécessaire de bien faire la distinction pour valoriser les services associés aux données de santé.

La valeur intrinsèque de la donnée de santé est une valeur qui dépend notamment de la qualité de la collecte, de la volumétrie, des traitements ainsi que de tous les services et expertises associés.

La valeur ajoutée spécifique de la donnée de santé correspond à ce que la donnée de santé génère comme valeur dans un projet/partenariat, c'est-à-dire sa contribution à la valeur finale d'un produit ou ce qu'elle a permis de générer.

Les outils de valorisation existants se décomposent en :

- **Des outils de valo-sécurisation :** il s'agit de modèles qui sécurisent la valeur intrinsèque. Ils sont souvent d'une moindre complexité à mettre en place mais ils ne permettent pas d'extraire toute la valeur des données en s'appuyant sur l'expertise du producteur de données de santé.
- **Des outils de maximisation de la valeur créée :** il s'agit de modèles orientés vers l'extraction du potentiel de valeur des *datasets*. Ils sont plus complexes à mettre en place et impliquent un partage du risque entre le producteur et l'exploitant mais ils sont davantage tournés vers des projets collaboratifs ambitieux à forte valeur ajoutée.



Volet valorisation

Outils de valorisation

INTERNE

Déterminer selon le type de projet et les usages, c'est-à-dire avec un croisement entre le catalogue des données et les usages, les différents outils à pousser et à utiliser dans les contrats pour procéder à la valorisation.

INTERNE

Définir les besoins et les ambitions scientifiques du guichet et déterminer les outils à utiliser pour assurer la valorisation scientifique : remerciements dans les publications, co-autorat, etc.

Florimond Bourdeaux Veltys

Associé du cabinet en charge des projets santé

Le travail conduit avec chaque acteur sur la définition et la formalisation d'une offre de services associée au guichet s'est avéré fondamental. Il a notamment permis aux acteurs de prendre conscience de l'ensemble des services qu'ils sont d'ores-et-déjà en capacité de fournir à l'écosystème, au-delà de la seule mise à disposition des jeux de données. Il a également servi de support à une réflexion de fond sur la valorisation non seulement des bases de données de chaque structure mais également de l'ensemble des services qui y sont associés en termes de coûts, par exemple en termes d'infrastructure IT ou de moyens humains mis à disposition dans les partenariats. Ce type de travail est absolument nécessaire aujourd'hui pour redonner confiance aux producteurs de données et faciliter leurs discussions avec les industriels désireux de collaborer avec eux à l'avenir. Il appelle une connaissance intime par les producteurs de leur catalogue de données et plus largement de leur organisation pour permettre d'établir et de proposer des moyens clairs dans les futurs partenariats autour des données.

Cette première brique constitue finalement un prérequis fondamental pour avancer sereinement dans les négociations et envisager une juste valorisation de la contribution des producteurs de données dans les projets, et ce quels que soient les outils utilisés pour valoriser les données et les services associés.

Marguerite Brac de La Perrière Cabinet LERINS

Avocat associée IT Data et Santé, experte en santé numérique

Le partage de la valeur est un sujet qui concerne la réutilisation des données. En gardant en tête que la cession à titre onéreux des données personnelles de santé est interdite ; il convient de valoriser les prestations réalisées sur les données pour permettre leur réutilisation de manière à inciter les producteurs à faire bénéficier des tiers poursuivant une finalité d'intérêt public, dans le respect des droits et libertés des patients.

3.3

METTRE EN PLACE DES INDICATEURS POUR SUIVRE LE DÉPLOIEMENT ET LE FONCTIONNEMENT D'UN GUICHET

Tout au long du projet, la collaboration entre les différents partenaires a permis de déterminer quels éléments doivent constituer un guichet pour qu'il soit efficace. Au cours des phases précédant l'appel à projets, les travaux avaient pour but de préparer chaque guichet, c'est-à-dire de déterminer quels *process*, outils et solutions doivent être utilisés pour chaque dimension du guichet pour réaliser tout type de partage de données, quel que soit le projet. Il s'agissait d'une étape de préparation et d'anticipation. Cela nous a permis d'identifier des indicateurs de maturité sous une forme binaire : existe-t-il un outil / une solution / un *process* pour prendre en charge chaque point lié au partage de données de santé ? La maturité est comprise ici comme une organisation avec des *process* et des outils déployés, en opposition à une situation où aucun outil n'existe et aucun *process* n'est formalisé.

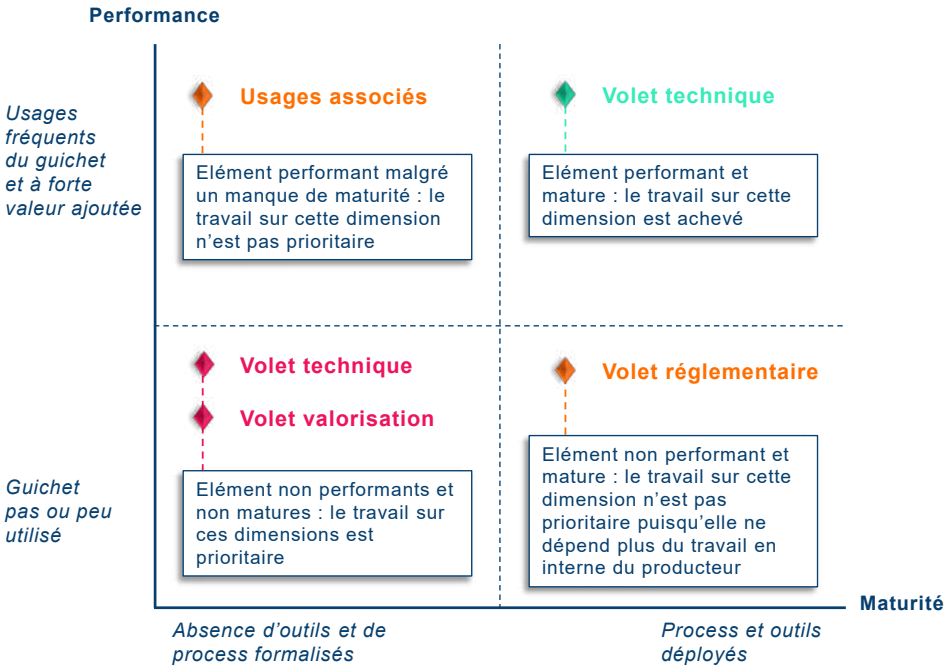
Indicateurs de maturité	
Élément constitutif du guichet	**Indicateurs**
Catalogue de données	▶ Existence d'une fiche standardisée pour décrire un jeu de données ▶ Existence d'une solution d'évaluation de la qualité d'une base de données ▶ Existence d'un catalogue de données
Usages associés	▶ Existence d'une typologie d'acteurs et de projets liés à l'exploitation de données ▶ Existence d'un comité de faisabilité pour évaluer les nouveaux projets
Volet technique	▶ Existence d'une procédure de partage des données conforme et sécurisée ▶ Capacité à anonymiser les données ou à les mettre en conformité (données synthétiques) ▶ Capacité à structurer des bases de données à partir du SI
Volet réglementaire	▶ Existence d'une solution visant à informer les patients et recueillir leur consentement ▶ Existence de contrats type
Volet valorisation	▶ Existence d'une offre de services associée à des outils de valorisation ▶ Connaissance des coûts de fonctionnement (IT et RH)

Au cours de l'appel à projets puis en routine d'opération du guichet, nous pensons qu'il sera important d'évaluer également la performance du guichet. L'indicateur majeur de performance dépend de l'objectif premier de chaque guichet : si l'objectif est la valorisation scientifique, l'indicateur clé est sans doute le nombre de publications ; cet indicateur clé peut aussi être le nombre de projets menés ou le temps moyen de mise en place du partage (à partir des premières discussions entre le producteur et l'exploitant). A nouveau, nous avons défini des exemples d'indicateurs de performance pour chaque élément constitutif du guichet. L'idée est d'étudier la performance du guichet au regard de la fréquence d'usage de cette solution où les procédures mises en place font la preuve d'un bon niveau d'anticipation. •••

Indicateurs de performance	
Élément constitutif du guichet	**Indicateurs**
Fonctionnement général du guichet	▶ Temps moyen pour conclure un contrat à partir des premiers échanges ▶ Nombre de contrats ▶ Nombre de publications ▶ Part des contrats réalisés dans le cadre du guichet ▶ Taux de révision des contrats
Catalogue de données	▶ Proportion de jeux de données partagés qui étaient initialement présents dans le catalogue ▶ Nombre de demandes d'informations supplémentaires sur les données ▶ Nombre de jeux de données dans le catalogue
Usages associés	▶ Proportion de négociations sortant des cadres identifiés (autres types de projet, autres types d'acteur) ▶ Nombre d'utilisations secondaires de bases de données
Volet technique	▶ Temps moyen pour partager le *dataset* ▶ Nombre de cas où un travail technique complémentaire a été nécessaire sur le jeu de données
Volet réglementaire	▶ Proportion de jeux de données conformes (anonymisés ou embarquant déjà le consentement patient) dans le catalogue ▶ Nombre de litiges (patient et/ou porteur de projet)
Volet valorisation	▶ Chiffre d'affaires (moyen) / marge ▶ Nombre de publications scientifiques réalisées grâce au guichet ▶ Nombre de contrats nécessitant des services non présents dans l'offre de services ▶ Nombre de projets ayant amené à un dépassement des coûts budgétés

Le croisement entre les indicateurs de maturité et ceux de performance permettra d'adapter la feuille de route en identifiant sur quels éléments il convient de travailler en priorité pour assurer la performance du guichet. Par exemple, si un élément du guichet est jugé peu mature mais performant il n'apparaît pas prioritaire de travailler spécifiquement sur l'amélioration des *process* et le développement d'outils pour cet élément.

Par ailleurs les *KPI* peuvent constituer un élément de confiance envers les porteurs de projets en rassurant sur le temps de mise à disposition des données et de négociation.



Le guichet :

une solution opérationnelle qui a fait ses preuves et qui alimente des échanges vertueux dans l'écosystème

Après un an de recul, le projet PHOENIX a permis de démontrer l'intérêt majeur pour l'écosystème de se réunir pour des échanges ouverts et transparents autour des enjeux associés à la mise à disposition et la valorisation des données de santé.

Si les freins qui rendent difficiles le partage des données et les projets de partenariats innovants en France sont encore nombreux, des solutions existent pour faciliter la mise à disposition par et pour chaque acteur à travers la structuration de guichets « *fast-tracks* » innovants. Ceux-ci peuvent être déployés rapidement en capitalisant notamment sur les outils et les *process* existants et s'adaptent facilement à la structure, aux spécificités et aux contraintes de chaque acteur. Le projet a également mis au jour de nombreuses bonnes pratiques qui peuvent être répliquées et adaptées et démontrent là-encore que des choses peuvent être faites pour améliorer les *process* de mise à disposition.

En outre, PHOENIX a aussi permis de montrer l'intérêt d'un tel type de dispositif pour favoriser la collaboration entre les acteurs. Au-delà des solutions apportées à chacun pour son propre guichet et ses propres projets, ce genre d'initiative montre l'intérêt de l'écosystème pour davantage de dialogue et de travail en commun afin de construire des solutions partagées qui aideront toutes les parties dans l'exploitation des données. Ces échanges sont essentiels pour créer une relation de confiance entre les acteurs grâce à une meilleure compréhension des contraintes et besoins des uns et des autres. En ce sens, PHOENIX invite à mener une réflexion à large échelle et un chantier plus global qui impliquerait tous les acteurs afin de transformer l'essai et faciliter encore plus la mise à disposition des données grâce à la co-construction de standards communs.

C'est là l'ambition de ce livre blanc qui doit encourager l'écosystème à toujours plus collaborer et partager.

Julie Baussand Medicen

Chef de Projets Innovation Santé – Responsable du Medicen Initiatives for Health Data

Tout d'abord, il faudra que ce travail serve pour amorcer les futurs guichets des acteurs qui souhaitent se structurer. Les souhaits de se structurer sur le territoire sont nombreux et ce travail pourra avoir un vrai impact pour amorcer ces guichets. Bien sûr chacun devra adapter les bonnes pratiques et recommandations à ses propres objectifs mais cela constituera un support essentiel. D'autres enseignements seront probablement à partager d'autres réplications de l'approche, et la continuité de ce partage de bonnes pratiques issues d'autres initiatives permettrait de constituer une référence acceptée par le plus grand nombre puisque issue de l'écosystème et de son expérience. Ensuite, les questions de mutualisation des données et de partage de données avec l'étranger, seront probablement les prochains gros chantiers à s'appuyer sur ces guichets. Les acteurs souhaitant adresser ces questions pourront s'inspirer de l'approche collaborative et pragmatique du projet PHOENIX.

Paul Rinaudo ADLIN Science

CEO & Co-founder

En France les acteurs n'arrivent pas à collaborer : ils se voient comme concurrents avant de se voir comme partenaires. Il y a une progression sur ces points notamment via le travail fondamental mené par Medicen. Cela mène à des rapprochements et à des partenariats entre start-ups. Néanmoins il y a surtout des partenariats entre même typologie d'acteurs (entre académiques, entre start-ups, entre laboratoires pharmaceutiques...) mais pas ou peu de partenariats structurants entre plusieurs types d'acteurs à cause du manque de confiance. Le projet PHOENIX vient contribuer à ce type d'initiative et c'est la bonne approche.

Camille Bachot Roche Pharma

Medical Data Platform Lead

Roche a déjà conclu et continuera à conclure de nombreux partenariats avec des partenaires publics et institutionnels. La collecte des données coûte extrêmement cher. Tous les acteurs de l'écosystème ont donc intérêt à exploiter les jeux de données déjà constitués ce qui permet de maintenir la soutenabilité de la recherche épidémiologique ou la santé publique.

NOTES DE FIN

Glossaire

Guichet de données de santé : structure et politique de mise à disposition de données de santé mises en place par un acteur producteur de données de santé (hôpital, laboratoire, entreprise…) pour des acteurs externes (publics comme privés). Il peut prendre plusieurs formes mais repose sur la structuration de pratiques et d'outils et sur la coordination d'expertises qui permettent de gagner en efficacité dans la mise en place et la réalisation d'accords de partage de données. La partie 2 de ce document porte sur la définition précise et exhaustive d'un guichet.

Techno push : processus démarrant suite à une offre du producteur ; par exemple partir d'une base de données existante et la pousser auprès d'industriels pour créer un produit / un projet.

Market pull : processus démarrant suite à un besoin du marché ; typiquement le cas d'un industriel approchant un producteur de données pour obtenir un jeu de données spécifique.

Précision sur la terminologie

Dans le cadre de ce livre blanc, les échanges ou partages de données ne sont pas distingués et s'entendent comme tout type de procédure donnant un accès aux données du producteur plus ou moins important à un exploitant ; ces termes ne correspondent pas à la terminologie juridique.

De la même façon, il peut arriver que les données de santé soient mentionnées avec un possessif (« les données des patients », « leurs données »), nous avons bien conscience qu'il n'y a pas de notion de propriété valide pour les données de santé.

Nous rappelons ici que la cession à titre onéreux des données personnelles de santé est interdite : il n'est pas légal de valoriser financièrement les données elles-mêmes. Les mentions de « valorisation des données de santé » dans ce livre blanc portent sur les valorisations annexes (scientifique notamment) et, par abus de langage parfois, sur la valorisation des services et des prestations fournies par les producteurs de données.

Abréviations

HDH : Health Data Hub

CNIL : Commission nationale de l'informatique et des libertés

PI : propriété intellectuelle

ODS : offre de services

IT : information technology

SI : système d'information

IA : intelligence artificielle

ODI : Open Data Institute

OCDE : Organisation de coopération et de développement économiques

PROM : Patient reported outcome measure

RGPD : règlement général sur la protection des données

M€, Md€ : million(s) d'euros, milliard(s) d'euros

UE : Union Européenne

KPI : Key Performance Indicator

REMERCIEMENTS

Tout d'abord, Medicen et Veltys tiennent à remercier les 3 partenaires : **l'Hôpital Foch, l'Institut Pasteur et Roche Pharma** pour leur engagement dans ce projet qui a permis d'apporter une vision sur des problèmes concrets d'acteurs qui structurent leur guichet. En particulier, nous remercions pour leur disponibilité, expertise et transparence durant les ateliers et pour la rédaction de ce livre :

- **Pauline Touche**, *Responsable Adjointe de l'Unité Data* – Hôpital Foch
- **Perrine Crequit**, *Pneumologue, MD PhD et responsable de l'unité Data* – Hôpital Foch
- **Amélie Chioccarello**, *Directrice adjointe en charge du Juridique et des Partenariats, DRCI, déléguée à la protection des données* – Hôpital Foch
- **Ségolène Salomez**, *Déléguée adjointe à la protection des données, Juriste en protection des données personnelles* – Hôpital Foch
- **Linda Nait-Kaoudjt**, *Head of Technology Transfer and Industrial Partnership* – Institut Pasteur
- **Claire Héritier**, *Global Partnership Contract Leader* – Institut Pasteur
- **Cloé Giquel**, *Juriste conformité santé publique & recherche* – Institut Pasteur
- **Anne-Caroline Deletoille**, *Responsable de la Plateforme de Data Management* – Institut Pasteur
- **Camille Bachot**, *Medical Data Platform Lead* – Roche Pharma
- **Valérie Machuron**, *External Data Acquisition Manager* – Roche Pharma
- **David Pau**, *BioStatistics & Data Science Lead* – Roche Pharma

Nous souhaitons également remercier l'ensemble des structures et personnes ayant accepté de partager leur expertise lors des ateliers thématiques et ayant ainsi contribué à l'avancée des réflexions :

- **Pierre-Antoine Gourraud**, *Professeur des universités et praticien hospitalier en biologie cellulaire* – CHU de Nantes
- **Marguerite Brac de La Perrière**, *Avocat associée IT Data et Santé, experte en santé numérique* – cabinet LERINS
- **Aurore Gaignon**, *Juriste service Santé* – CNIL
- **Hélène Guimiot-Brehaud**, *Chef du service santé* – CNIL
- **Paul Rinaudo**, *CEO & Cofounder* – ADLIN Science
- **Antonio Borderia PhD.**, *Executive Vice-President Business Development* – ADLIN Science
- **Maxime Agostini**, *Président et cofondateur* – Sarus Technologies
- **Sandrine Murcia**, *CEO* – Cosmian
- **Sebastian Schwarz**, *Data Platform Strategy Lead* – OWKIN
- **Jocelyn Dachary**, *SVP IT & Data Solutions* – OWKIN
- **Emeric Lemaire**, *Co-founder & Co-CEO* – Arkhn
- **Corneliu Malciu**, *Co-founder & Co-CEO* – Arkhn
- **Nicolas Glatt**, *CEO* – Clinityx
- **Seamus Larroque**, *Co-Founder & CEO* – Illiomad
- **Pierre Malvoisin**, *Co-Founder & COO* – Illiomad

AUTEURS

Jessica Leygues

Déléguée Générale Medicen

Ingénieure *AgroParisTech*, Jessica Leygues a débuté sa carrière au sein du cabinet de conseil *Alcimed*. Pendant huit ans, elle y a pris en charge le développement de nouveaux marchés en santé (biotech/*medtech*), pour les créateurs d'entreprises, *start-up*, PME et grands groupes, ainsi que pour la sphère publique et para-publique de l'innovation. Jessica Leygues a ensuite rejoint le laboratoire pharmaceutique Celgene pour prendre en charge le pôle des partenariats innovants, visant à développer le modèle de la valeur en santé.

Julien Ettersperger

Directeur de l'innovation Medicen

Julien Ettersperger est titulaire d'un doctorat en immuno-oncologie de l'université Paris Cité, qu'il a réalisé dans l'équipe du Dr Nadine Cerf-Bensussan, à l'Institut Imagine à Paris. Il a ensuite complété sa formation avec un MS en bio-entrepreneuriat de l'École polytechnique et d'HEC Paris en 2017. Il a notamment travaillé sur le projet de création de la société *Smart Immune*, dont il a par la suite occupé le poste de responsable du *business development*, avant de rejoindre Medicen en mars 2019.

Julie Baussand

Responsable du Medicen Initiatives for Health Data (MIHD)

Titulaire d'un doctorat en bio-informatique de l'université Paris VI, Julie Baussand a consacré la première partie de sa carrière à la recherche médicale, dans des laboratoires INSERM à Paris, puis au *Medical Research Council* à Londres. Après cela, elle a rejoint Dassault Systèmes pour participer aux développements d'applications de modélisation et de simulation dédiées au développement de nouvelles molécules thérapeutiques. Julie Baussand a rejoint Medicen en janvier 2020 comme responsable du secteur Imagerie Médicale, avant de mettre en place en 2021 le Medicen Initatives for Health Data (MIHD), l'axe dédié aux données de santé et Intelligence artificielle de Medicen.

Florimond Bourdeaux

Associé du cabinet Veltys en charge des projets santé

Diplômé de l'École Normale Supérieure de Cachan et de l'École d'Économie de Paris, Florimond Bourdeaux a rejoint Veltys dès sa création en 2013. Responsable de la *practice* santé, il conduit depuis près de 10 ans de nombreuses missions auprès de l'écosystème (ministère de la Santé, ARS, établissements, laboratoires pharmaceutiques, *start-ups*, consortiums publics, etc.) et possède une connaissance fine de ses enjeux data. Il a notamment développé une expertise très riche en matière de valorisation de données et accompagne l'écosystème dans l'étude, la conception et le déploiement de stratégies de valorisation data innovantes. Il a notamment conduit l'étude AVIESAN (2020).

Arthur Souletie

Associé du cabinet Veltys en charge des projets santé

Titulaire d'un doctorat en économie et ancien élève de l'École Normale Supérieure de Cachan, Arthur Souletie a démarré sa carrière à la Direction Générale du Trésor, au Bureau « transport et Énergie », où il effectuait des missions de conseil et d'évaluation de politiques publiques. Il a ensuite rejoint le cabinet Veltys pour co-diriger la *practice* santé au sein de laquelle il assure le pilotage des programmes de santé nationaux développés par le cabinet pour accompagner le système de santé. Il conduit également les missions d'évaluation de politiques publiques de santé pour le Ministère (dispositif FIDES Séjours, dispositif des hébergements temporaires non médicalisés (HTNM)).

Antoine Robins

Consultant data-scientist expert santé et data

Diplômé de l'ENSAE et de l'ESSEC, Antoine Robins est consultant au sein de la *practice* santé depuis son arrivée chez Veltys. Il y a développé une expertise sur l'exploitation intelligente des données afin d'accompagner ses clients dans l'évaluation et l'amélioration de leur organisation interne et dans l'évaluation de politiques de santé. Il a également contribué aux missions de valorisation data du cabinet auprès de l'écosystème. Lors de ses précédentes expériences, il a travaillé sur des sujets d'économie de la concurrence dans un cabinet de conseil et sur des articles de recherche dans un laboratoire de biostatistiques.

CONTRIBUTEURS

Pauline Touche,
Responsable Adjointe de l'Unité Data
Hôpital Foch

Linda Nait-Kaoudjt,
Head of Technology Transfer and Industrial Partnership
Institut Pasteur

Camille Bachot,
Medical Data Platform Lead
Roche Pharma

Perrine Crequit,
Pneumologue, MD PhD et responsable de l'unité Data
Hôpital Foch

Claire Héritier,
Global Partnership Contract Leader
Institut Pasteur

Valérie Machuron,
External Data Acquisition Manager
Roche Pharma

Amélie Chioccarello,
Directrice adjointe en charge du Juridique et des Partenariats, DRCI, déléguée à la protection des données
Hôpital Foch

Cloé Giquel,
Juriste conformité santé publique & recherche
Institut Pasteur

David Pau,
BioStatistics & Data Science Lead
Roche Pharma

Ségolène Salomez,
Déléguée adjointe à la protection des données, Juriste en protection des données personnelles
Hôpital Foch

Anne-Caroline Deletoille,
Responsable de la Plateforme de Data Management
Institut Pasteur

PRÉSENTATION DES STRUCTURES

Créé en 2005, **MEDICEN PARIS REGION est le pôle de compétitivité santé de la Région Île-de-France**. Son réseau unique fédère plus de 510 acteurs de l'innovation en santé parmi lesquels 430 startups et PME de la *healthtech*, les industriels de la santé, les principaux instituts de recherche nationaux et établissements de soins du territoire.

En rassemblant acteurs privés et publics autour des enjeux d'innovation pour développer les solutions thérapeutiques et diagnostiques de demain, Medicen est le tiers de confiance de la filière, au service du développement et de l'aboutissement des projets.

Afin d'œuvrer à la croissance du secteur, à la mise sur le marché de produits et services de santé innovants ainsi qu'à la création d'emplois, le pôle dispose de trois leviers d'action :

Favoriser l'émergence de projets de recherche et d'innovations notamment par l'animation de l'écosystème

Orienter les porteurs de projets vers les bons guichets de financements publics ou privés

Accompagner la croissance des startups et PME en France et à l'international

Medicen, pôle de compétitivité mondial, est un acteur de terrain avec plus de 15 ans d'existence. Avec des actions concrètes sur un large scope, innovation santé, développement des entreprises, Europe et international, ses équipes et ses experts assurent un accompagnement adapté au profil de chaque adhérent tout au long d'un projet et ce, quel soit son niveau de maturité.

PRÉSENTATION DES STRUCTURES

VELTYS est un cabinet de conseil en stratégie, économie et data science créé en 2013 par des chercheurs pour fournir aux acteurs privés et publics un accompagnement stratégique et opérationnel sur-mesure pour toutes les problématiques complexes autour des questions économiques & data.

Le cabinet dispose d'une expertise scientifique à la pointe fondée sur des liens forts avec la recherche en économie, politiques publiques, statistiques et *machine learning* et une équipe de consultants-chercheurs issus des grandes écoles où ils poursuivent des missions d'enseignement et de recherche. Les interventions du cabinet reposent notamment sur une maîtrise de tous les maillons de la chaîne de valeur data, de la collecte et du traitement des données à la construction de recommandations stratégiques et d'actions concrètes nourries par l'analyse de données.

Partenaire de référence du secteur de la santé, Veltys intervient auprès de ses principaux acteurs depuis près de 10 ans autour de trois missions : (i) l'accompagnement data et l'aide à l'exploitation de données du système de santé (établissements, ARS, DGOS), (ii) le conseil économique et data auprès des acteurs privés dont les laboratoires pharmaceutiques, et (iii) la fourniture d'une expertise économique et stratégique sur les enjeux associés à la valorisation des données de santé. C'est dans ce contexte que Veltys a notamment conduit les travaux réalisés en 2020 pour AVIESAN, expérience sur laquelle le cabinet a pu capitaliser dans le cadre du projet PHOENIX.

INSTITUT PASTEUR

L'Institut Pasteur est une fondation à but non lucratif dont la mission est de contribuer à la prévention et au traitement des maladies, en priorité infectieuses, par la recherche, l'innovation, l'enseignement, et des actions de santé publique. L'Institut Pasteur développe de nombreux grands projets internationaux en partenariat avec les grandes instances scientifiques internationales comme l'Organisation mondiale de la santé, et de nombreuses organisations, fondations, instituts de recherche, universités ou autres acteurs privés du monde entier. Par ailleurs, l'Institut Pasteur est membre du Pasteur Network, réseau de coopération internationale en santé publique, enseignement et recherche, qui regroupe 33 membres dans le monde.

HÔPITAL FOCH

L'hôpital Foch est le plus grand hôpital privé universitaire de France avec un effectif de 2 300 collaborateurs dont plus de 300 médecins, 611 lits installés, 260 000 consultations hors maternité et urgences, et plus de 60 000 hospitalisations par an. Ses prises en charge pluridisciplinaires de haut niveau dans la quasi-totalité du champ médical et chirurgical de l'adulte, sa forte implication dans l'enseignement, la formation et la recherche, son plateau médicotechnique de pointe, sa tradition d'accueil en font l'un des hôpitaux privés à but non lucratif les plus performants de France.

ROCHE

Fondée en 1896 à Bâle, Suisse, Roche compte parmi les tout premiers fabricants industriels de médicaments de marque. Elle s'est hissée au rang de plus grande entreprise de biotechnologie au monde et est par ailleurs le numéro un mondial du diagnostic in vitro. Visant l'excellence scientifique, Roche a pour ambition de découvrir et de développer des médicaments et des diagnostics permettant d'améliorer la qualité de vie des patients et de sauver des vies dans le monde entier. Elle est à l'avant-garde de la médecine personnalisée et entend renforcer sa contribution en continuant à faire évoluer la prise en charge des patients. Afin d'apporter à chacun les meilleurs soins, Roche coopère avec de nombreux partenaires et conjugue ses compétences au sein des divisions Diagnostics et Pharma avec des données issues de la pratique clinique.

Résumé

Le projet PHOENIX a été initié en mai 2021 par Medicen via le MIHD (*Medicen Initiatives for Health Data*) pour répondre aux difficultés de partage des données de santé dans l'ensemble de l'écosystème. Pour cela, le projet PHOENIX a réuni un consortium de trois acteurs emblématiques : l'Institut Pasteur, l'Hôpital Foch et Roche Pharma, complémentaires sur la chaine de valeurs des données de santé. Il a pour vocation d'aider l'écosystème à favoriser et accélérer les accords de partage de données de santé via la mise en place de guichets "*fast-track*" innovants. Avec l'aide du cabinet Veltys, Medicen a accompagné pendant plus d'un an ces trois acteurs dans la conception et le déploiement de leurs guichets afin de les aider à améliorer leur politique et leurs pratiques de mise à disposition de données de santé vis-à-vis de l'écosystème.

Au cours d'ateliers collectifs et individuels, Medicen, Veltys et les trois acteurs ont ainsi conduit une réflexion commune sur les enjeux et les difficultés liés au partage de données afin de coconstruire des solutions qui pourraient profiter à l'ensemble de l'écosystème.

Le présent livre blanc vise à diffuser au maximum les enseignements qui découlent de ces travaux pour favoriser et accélérer les accords de partage des données de santé et ainsi optimiser leur utilisation pour accélérer l'innovation en santé. Le livre blanc rend ainsi compte des motivations du projet, de la démarche adoptée et du dispositif innovant proposé par PHOENIX. Il présente également l'ensemble des bonnes pratiques identifiées au cours du projet pour faciliter les accords de partage de données de santé.

Ce livre blanc est une première étape du projet, qui sera suivie par le lancement d'un appel à projets pour mettre en pratique ces enseignements. Il pourra donc être enrichi des enseignements tirés par les trois acteurs dans le cadre de cet appel à projets.

Julie Baussand Medicen

Chef de Projets Innovation Santé Responsable du Medicen Initiatives for Health Data

Le projet PHOENIX est né des retours terrains d'acteurs producteurs ou exploitants de données de santé qui font face à des difficultés à différents niveaux pour partager ces données. Cette situation représente un frein pour la mise en place de projets de recherche collaboratifs et génère un contexte de défiance entre les acteurs.

La constitution d'un petit groupe de travail entre acteurs, pour certains pouvant être aussi bien dans une position de producteur que d'exploitant, dans un contexte d'échange ouvert sur les difficultés du partage car sans enjeux contractuels entre eux, a permis d'une part de réaliser la volonté commune à tous d'avancer sur ces questions pour faciliter le partage de données, et d'autre part de prendre conscience des contraintes auxquelles doivent faire face les uns et les autres pour mettre en œuvre ce partage malgré les bonnes volontés.

En d'autres termes, ce travail a permis de mettre en place un contexte de confiance nécessaire pour aborder concrètement et en toute transparence les fondations des guichets.