



Bilan et enjeux du partage de données des industriels

Groupe de travail – Partage de données des
industriels



Dans un contexte de **développement exponentiel des capacités de collecte, de stockage et de calcul des données**, l'exploitation des données de santé est incontournable et porteuse de progrès pour les patients et le système de santé dans son ensemble.

Les investissements colossaux dans la donnée par les industriels pharmaceutiques posent la question de leur utilisation, de leur partage et de leur exploitation. Dans ce cadre, un groupe de travail (GT) a été constitué afin de **réfléchir aux enjeux du partage des données générées par les industriels**, de partager les retours d'expérience sur cette pratique et de commencer à **construire une vision commune sur les finalités et les prérequis à ce partage**.

Ce groupe de travail a mobilisé 19 experts représentants de différentes entreprises de santé : AMGEN, ARIIS, ASTRAZENECA, EDWARDS, INCA, IPSEN, MEDTRONIC, PFIZER, RESMED, ROCHE, ROCHE DIA, SANOFI, SERVIER, SNITEM, WLGORE.

Le présent document est le fruit de plusieurs ateliers organisés avec les membres du GT ainsi que d'une enquête « **Etat des lieux du partage des données générés par les industriels** » envoyé aux membres du GT ainsi qu'aux adhérents ARIIS.

Ce travail est issu d'une réflexion commune menée dans le cadre du comité stratégique de filière des industries et technologies de la santé (CSF-ITS). Ces travaux ont été accompagné par Iqvia.

En aucun cas ceci est un engagement des membres de l'ARIIS à partager leurs données.

Table des matières

I.	Qu'est-ce que le partage de données ?.....	2
II.	Qui sont les acteurs de ce partage de données ?.....	3
III.	Pourquoi partager ?.....	4
IV.	Quels modèles de partage ?.....	5
V.	Dans quel cadre partager ?.....	6
VI.	Où en est-on aujourd'hui ?.....	7
VII.	Quels sont les freins au partage et comment les surpasser ?.....	9

I. Qu'est-ce que le partage de données ?

Afin d'établir le bilan et les enjeux du partage de données des industriels, il convient dans un premier temps de **revenir sur la définition des termes qui le compose.**

a. « Le partage »

Le **partage** peut se définir par son niveau d'ouverture, avec aux deux extrêmes le partage fermé et le partage ouvert.

Le *partage fermé* concerne un nombre restreint d'acteurs qui s'accordent afin de partager des données entre eux. Dans ce cadre, ceux qui partagent la donnée contrôlent le pourquoi et le comment et limitent l'usage qui peut en être fait. Ce type de partage est actuellement relativement répandu, par exemple à travers les partenariats entre deux acteurs ou les études coconstruites.

A l'inverse, le *partage ouvert* consiste en un partage avec un accès ouvert à tous. Ceux qui partagent la donnée n'ont pas de regard sur l'usage des données. Cet usage n'est donc pas limité, au-delà de la validité scientifique de la question étudiée et des bonnes pratiques. Le partage ouvert est très peu développé actuellement. La plateforme HDH ou le SNDS sont parmi les seuls vrais exemples de partages ouverts.

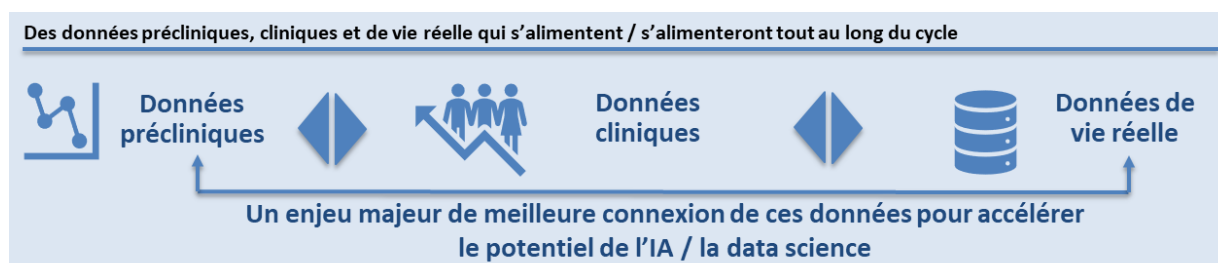
Tout l'enjeu est de trouver le niveau d'ouverture de partage souhaité et réaliste. Les membres du GT ont réfléchi à ce degré d'ouverture et se sont accordés sur le fait que l'enjeu du projet était sur le développement du partage ouvert ou semi-ouvert, encore peu développé aujourd'hui et qui servirait davantage l'intérêt commun.

b. « Les données des industriels »

Dans le cadre de ce projet, les membres du GT ont défini les *données partagées/partageables* comme étant des **données générées au niveau patient pour lesquelles l'entreprise est responsable de traitement, ou finance/ participe au financement de sa génération.**

Ces données suivent le cycle de vie des produits de santé et peuvent donc être **des données pré-cliniques, des données cliniques ou encore des données de vie réelle.**

Figure 1 : Alimentation des données précliniques, cliniques et de vie réelle



Les *données des études cliniques et précliniques* constituent la première source « historique » de données exploitables autour des traitements pharmaceutiques, principalement sous la forme d'essais randomisés contrôlés. Les essais cliniques restent fondamentaux aussi bien pour évaluer la toxicité d'un médicament (phase I) que pour démontrer l'efficacité d'un médicament ou d'un dispositif médical (phases II et III).

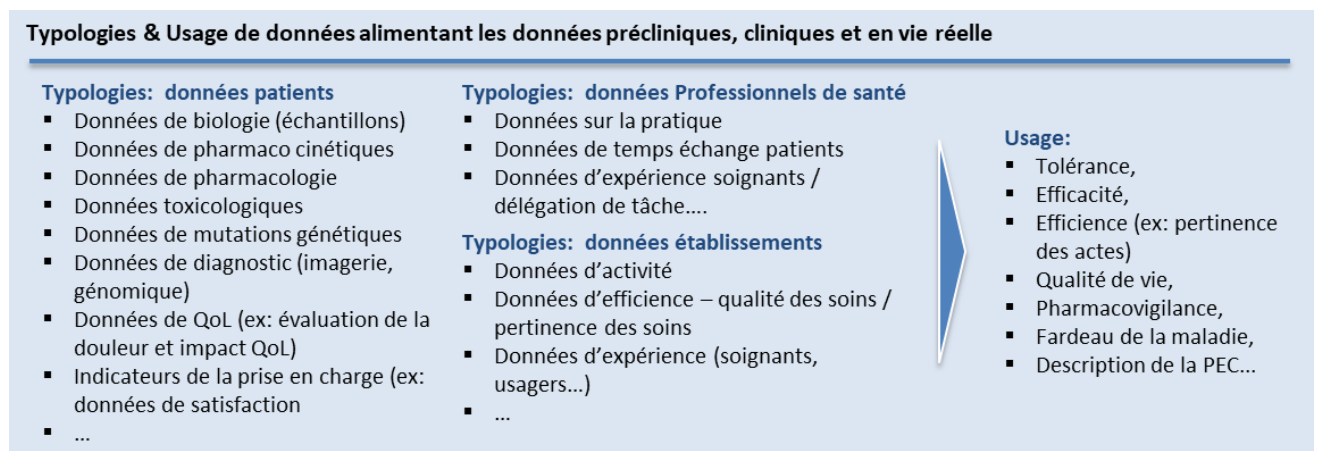
Les *données de vie réelle* décrivent, elles, l'usage d'un traitement dans un environnement non contrôlé de la vie quotidienne du patient tout au long du parcours de soin, à l'opposé des études interventionnelles. Leur analyse lors d'essais cliniques permet de prouver de potentiels bénéfices ou risques liés à des produits médicaux. Elles peuvent être utilisées de différentes manières, couplées à d'autres technologies (IA) et à des analyses statistiques innovantes. Les différentes données de vie réelle produites et collectées sont les suivantes :

- Données d'accès précoce
- Données pour les études observationnelles
- Données pour les études médico-économiques
- Données d'expérience patient - PROMS (Patient-Reported Outcomes Measures) et PREMS (Patient-Reported Experience Measures)
- Données issues de dispositifs médicaux connectés (données de soin)
- Données d'outcome des études précliniques

Les données alimentant les études précliniques, cliniques ou de vie réelle, et pouvant potentiellement être partagées pour un usage secondaire, se répartissent selon trois typologies de données (cf. figure 2.) :

- Les données patients (ex. données de biologie, d'imagerie),
- Les données de professionnels de santé (ex. données sur la pratique)
- Les données d'établissement (ex. données d'activité)

Figure 2 : Typologie et usage des données



II. Qui sont les acteurs de ce partage de données ?

Les acteurs du partage de la donnée se répartissent selon deux catégories : **les producteurs et les exploitants.**

Les « producteurs » sont les acteurs à l'origine de la donnée produite. Cette production de donnée s'inscrit dans le cadre d'un usage primaire (ex. Réalisation d'un essai clinique). Le producteur d'une base de données est la personne, le plus souvent morale, qui a pris « l'initiative et le risque des investissements »¹ nécessaires à la création du contenu de la base de données.

¹ Source : Article L341-1 du code de la propriété intellectuelle

Le rôle de l'industriel est fonction du modèle opérationnel de partage de données défini. Son rôle peut être les suivants :

- « *Responsable de traitement* » : la personne physique ou morale, l'autorité publique, le service ou un autre organisme détermine les finalités et les moyens du traitement seul.
- « *Co-responsables de traitement* » : la personne physique ou morale, l'autorité publique, le service ou un autre organisme détermine les finalités et les moyens du traitement conjointement avec d'autres.
- « *Sous-traitant* » : personne physique ou morale, l'autorité publique, le service ou un autre organisme qui traite des données à caractère personnel pour le compte du responsable de traitement²

Si les données ont une valeur pour le producteur de la donnée, elles peuvent potentiellement également en avoir pour d'autres acteurs et pour d'autres finalités. Si les données sont partagées par le producteur de la donnée, cela peut permettre à ces autres acteurs d'exploiter la donnée. Ces acteurs sont ainsi les « **exploitants** » de la donnée.

Les producteurs de données peuvent être également des exploitants, pour d'autres sources de données. La figure ci-dessous synthétise les différents producteurs et exploitants des données d'essais cliniques et de vie réelle.

Figure 3 : Producteurs et Exploitants de la donnée

	Données d'essais cliniques		Données de vie réelle	
	Producteur	Exploitant	Producteur	Exploitant
1 Laboratoires pharmaceutiques et fabricants de DM	✓	✓	✓	✓
2 Centres hospitaliers	✓		✓	✓
3 Autorités de santé		✓ (indirect)		✓ (indirect)
4 Payeurs		✓ (indirect)		✓ (indirect)
5 Associations de patients		✓ (indirect)	✓	✓ (indirect)

III. Pourquoi partager ?

L'intérêt du partage réside dans ce que l'on peut faire avec les données récoltées, **au-delà de leur objectif primaire**. Le partage de données ne nécessite pas de transmettre l'intégralité des données car **l'objectif n'est pas de refaire des analyses primaires de ces données**. En effet, la richesse du partage n'est pas dans la réutilisation de la donnée pour refaire l'analyse déjà réalisée (par exemple refaire une safety) mais dans ce que l'on peut faire d'autre avec cette donnée, une utilisation secondaire différente de ce pourquoi la donnée a été initialement produite.

² Source : Legifrance <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000035268202/> Et HDH

Par exemple, un laboratoire qui a généré des données d'imagerie ou biologiques pour mesurer l'effet du traitement dans le cadre d'un essai clinique pourrait partager ses données afin de permettre à un autre acteur d'affiner l'identification de biomarqueurs, de travailler sur de l'IA ou de manière générale de mettre en place des dispositifs intéressants en dehors de l'essai clinique. Dans ce cas ce ne sont pas les données de « l'outcome primaire » qui sont intéressantes mais les autres données générées. Ces données peuvent être riches en informations (données longitudinales, etc...). Selon les membres du groupe de travail, **les laboratoires n'utilisent pas plus de 10% des données générées dans les essais cliniques.**

De la même manière, l'utilisation combinée de plusieurs données de sources différentes, grâce à leur complémentarité, peut permettre de générer de nouveaux résultats, qui n'auront pas nécessairement un lien avec les résultats primaires des différentes sources. La valeur ajoutée du partage peut provenir ainsi d'un croisement entre plusieurs sources de données. Ainsi, par exemple, toute base de données, de vie réelle ou d'essai clinique, qui permet de chaîner les patients avec le SNDS, est un bon sujet de partage. *Par exemple*, pour les données en vie réelle, l'étude des données d'une ATU mises en perspective avec les données du SNDS peut permettre de réaliser une analyse de survie à moyen/long terme.

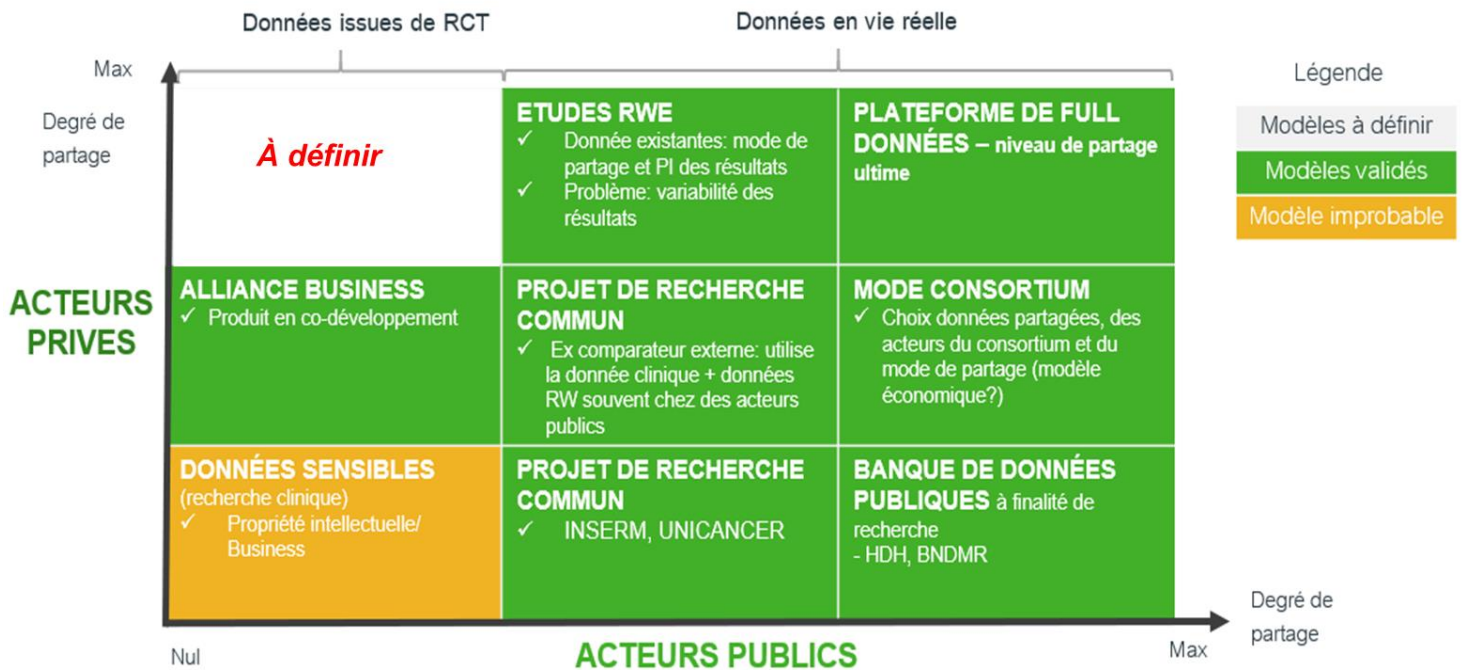
Le partage de données peut également permettre d'accroître la crédibilité des données générées via l'utilisation **d'une seule et même base de données officielle pour réaliser des analyses**. Par exemple, les autorités pourraient travailler sur la même base partagée de données en vie réelle que le laboratoire, tout en s'assurant de la crédibilité via un processus de certification qualité externe, cela permettrait d'éviter des comparaisons de données provenant de sources différentes.

IV. Quels modèles de partage ?

La possibilité et la volonté de partage dépendent du caractère de la donnée (données d'essais cliniques, données en vie réelle) et des interlocuteurs concernés (acteurs publics, acteurs privés).

Comme illustré dans la figure 2, neuf catégories de partage peuvent être mises en évidence.

Figure 4 : Matrice de partage de données



V. Dans quel cadre partager ?

Le partage de données, **surtout ouvert**, nécessite toutefois la définition de bonnes pratiques.

1. Les partages doivent être encadrés :

- Il faut définir des **scénarios de gouvernance** : Quels sont les scénarios ? Lesquels sont les plus réalistes ? Quels sont les pour/contre ?
- L'engagement ne doit pas être sur celui qui produit la donnée mais **sur celui qui l'utilise**
- Il convient également de définir **les finalités interdites** (comme pour le SNDS) ainsi que des règles d'utilisation dans une charte de bonne pratique afin de limiter les risques liés aux données sensibles (ex. utilisation de données cliniques par le concurrent pour remettre en question un produit)

2. Les modalités de partage peuvent différer selon le type de données partagées.

Ainsi les **données d'essais cliniques** étant plus stratégiques, il convient d'intégrer **des finalités interdites renforcées**. Le partage de ces données implique des contraintes car le laboratoire qui les partage doit garder la main sur la communication de ces données. Le timing de partage après la génération des données est ainsi un élément indispensable à prendre en compte : le partage des données ne peut être fait qu'une fois l'autorisation de mise sur le marché obtenue. **La séparation des données stratégiques et des données ouvertes au partage doit être anticipée** par l'industriel qui produira/partagera la donnée.

Concernant les **données de vie réelle**, ces dernières nécessitent également des règles de bon usage, même si elles seront moins strictes que pour les essais cliniques. Les données observationnelles étant moins sensibles, **le partage de ces données est plus facile à réaliser et le timing de partage est moins stratégique**. Les données partageables sont les caractéristiques de base des patients (permettant de lier plusieurs données entre elles), les produits que le patient a reçus (ex. dans une ATU) et les données de résultats sous réserve de contrôler les finalités (autoriser la transposabilité du SOC mais limiter le head to head).

La méthode utilisée par celui qui veut utiliser la donnée ne doit en revanche pas être limitée ; la méthode n'étant qu'un outil et non une finalité.

Dans un usage secondaire de la donnée, la donnée a déjà été générée donc **les « secondary users » n'auraient pas à supporter le coût de la production initiale de la donnée**. Le coût de production de la donnée étant déjà amorti, le coût le plus important est celui de l'analyse de données faite dans le cadre de ce secondary use. La valorisation financière n'est pas un enjeu pour le partage de la donnée car au-delà du coût de production / de mise à disposition de la donnée, c'est l'usage qui en est fait qui est intéressant, et qui créera de la valeur. C'est donc la « valorisation scientifique » qui sera à considérer.

La qualité de la donnée partagée et le sujet d'étude auquel ce partage doit répondre doivent être vérifiés par un tiers de confiance (comme le CESRESS qui évalue l'intérêt des demandes de travail sur les données du SNDS). Il apparaît nécessaire de définir un processus inter-industriel pour évaluer le risque ou non par rapport aux données demandées, à l'analyse envisagée et ainsi limiter les mauvaises interprétations de données

VI. Où en est-on aujourd'hui ?

a. Constat et exemples de partage

Le premier constat ressorti lors des ateliers et illustré par les réponses au questionnaire « **Etat des lieux du partage des données générés par les industriels** » est la **faiblesse du partage de données actuel**. Ce partage, lorsqu'il est réalisé, se fait encore sur des périmètres très restreints, mettant en évidence notamment **une certaine réticence des industriels à partager leurs données**.

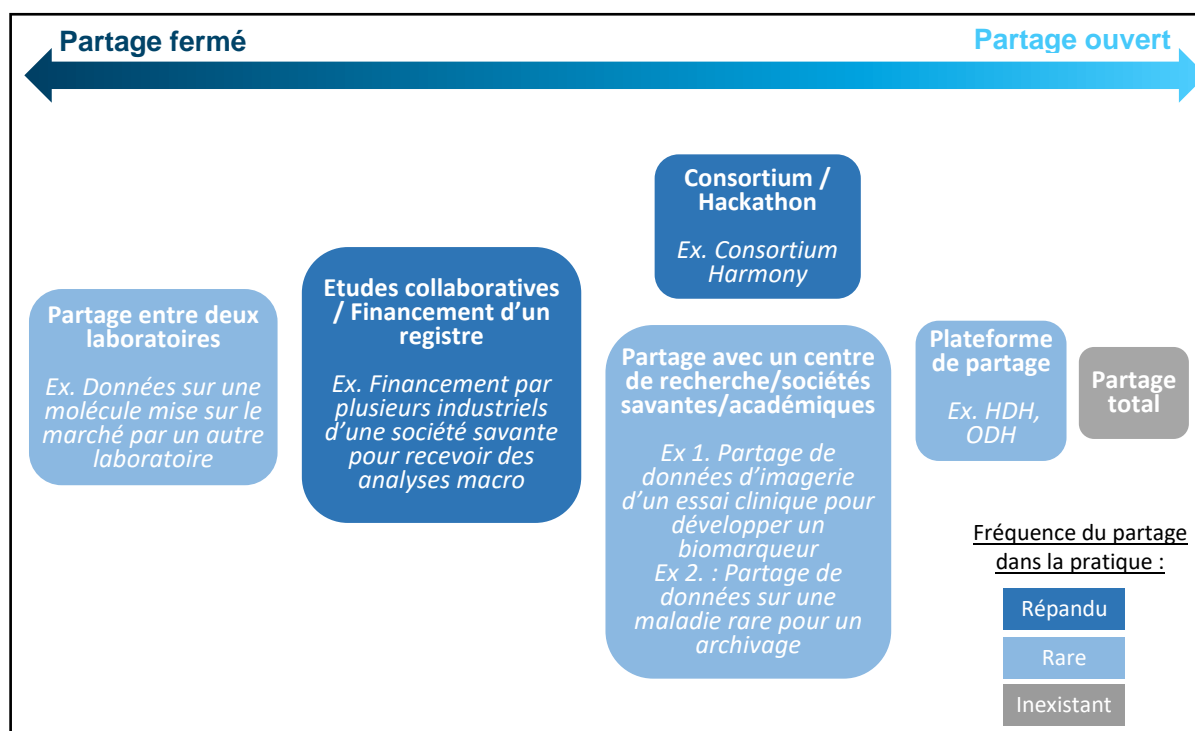
Les principales expériences de partage qui sont ressorties concernent surtout des partages fermés, entre un nombre d'acteurs restreints, et « contrôlés ». Par exemple, certains industriels travaillent avec des académiques, dans le cadre d'études collaboratives où deux pratiques ont été mises en évidence :

- Un ou plusieurs laboratoires financent un « instrument » permettant de réaliser une étude
- Un laboratoire utilise les données d'académiques existantes pour réaliser une étude

Dans les deux cas, **les industriels n'ont pas directement accès aux données brutes**. Les industriels financent tous des registres en échange de rapports consolidés. L'équipe académique ou autre, gérant le registre, fait l'analyse et la partage aux « financeurs », ces derniers n'ont pas accès aux données brutes directement. De la même manière, lors d'un partage de données depuis un laboratoire vers des académiques, ces derniers peuvent avoir accès ou non aux données brutes.

Les partages de données mis en évidence lors des ateliers ont été positionnés selon l'axe du degré d'ouverture du partage (cf. Figure 1).

Figure 5 : Exemples de partages de données selon leur degré d'ouverture et leur fréquence

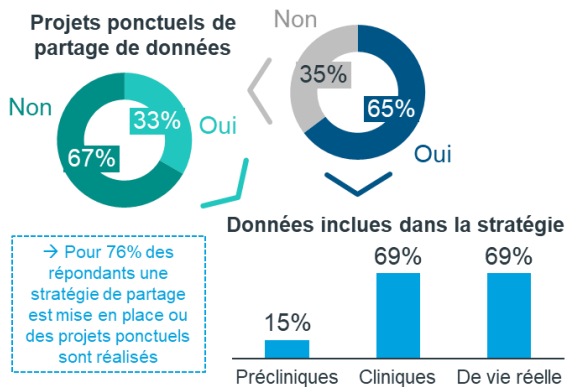


Concernant le type de données partagées, le partage de données de vie réelle est plus développé actuellement que le partage de données d'essais cliniques. Cela est lié au caractère plus stratégique de ces dernières issues du développement d'un produit mais servant également au développement.

Dans l'enquête « *Etat des lieux du partage des données générés par les industriels* » qui a été menée auprès de 17 industriels, 76% des répondants ont indiqué qu'une stratégie de partage de données était mise en place dans leur entreprise ou qu'ils avaient des projets ponctuels de partage de données. Ces stratégies concernent autant les données cliniques que les données de vie réelle. Néanmoins, si l'on regarde l'état du partage, ce dernier n'est **réalisé que de façon rare et en réactif à des demandes isolées pour les données cliniques**. Le **partage de données de vie réelle est quant à lui un peu plus avancé** avec plus d'un tiers des répondants pour lesquels le partage est mis en place et que des projets sont en cours.

Figure 6 : Stratégie de pratique de partage de données et état du partage

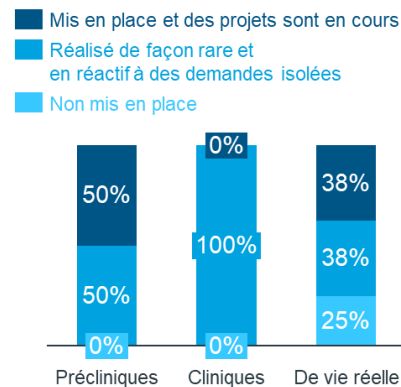
Stratégie de partage de données mise en place dans l'entreprise



Questions B1. Votre entreprise a-t-elle mis en place une stratégie de partage de données ? ; B1bis. A défaut d'avoir mis en place une stratégie définie sur le partage de données, avez-vous connaissance de projets ponctuels de partage de données dans votre entreprise ? ; B2. Cette stratégie de partage de données inclut-elle les données

Base : 17 répondants

Etat du partage avec des industriels ou autres acteurs de l'écosystème de la santé



Questions B3/B4/B5. Ce partage de données précliniques/cliniques/en vie réelle avec des industriels ou d'autres acteurs de l'écosystème de la santé (institutionnels, associations de patients, académiques, hôpitaux...) est-il ?

Base : respectivement 2, 9 et 9 répondants (dont 1 « ne sait pas »)

Concernant le type de partage réalisé, l'enquête montre que **c'est le partage relativement fermé qui est le plus pratiqué par les industriels. Pour les données précliniques et cliniques, 100% des partages réalisés sont des partages fermés**, qu'ils soient multi acteurs ou bilatéraux.

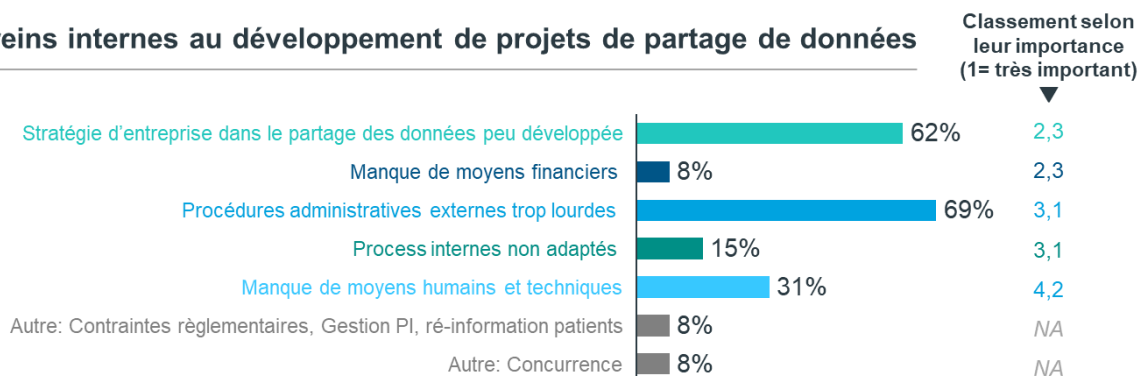
VII. Quels sont les freins au partage et comment les surpasser ?

La faible expérience de partage de données qui existe actuellement est liée à différents freins encore présents pour le partage des données.

Notre enquête « *Etat des lieux du partage des données générés par les industriels* » a mis en évidence les freins les plus courants et les plus importants pour les industriels aujourd'hui. Les **procédures administratives externes trop lourdes** sont les freins les plus répandus actuellement (69 % des répondants), suivi de **la stratégie d'entreprise dans le partage des données peu développée** (62% des répondants). En matière d'importance, les industriels classent ce dernier frein comme étant le plus important, au même niveau que le manque de moyens financiers mais qui lui est moins répandu (8% des répondants).

Figure 7 : Freins internes au développement de projets de partage de données

Freins internes au développement de projets de partage de données



Questions B18. Au sein de votre entreprise quels sont les freins internes au développement de projets de partage de données ? ; B18bis. Merci de bien vouloir classer ces freins du plus important (1) au moins important (5) :

Base : 17 répondants

Face aux différents freins que peuvent rencontrer les industriels pour partager leurs données, **différents prérequis clé ont été mis en évidence** selon deux axes : **les prérequis de l'industriel** et **les prérequis de l'écosystème**.

Freins au partage de données	Prérequis de l'industriel pour le partage de données
Faiblesse de la stratégie d'entreprise dans le partage de données	Organiser un cadre administratif interne favorable au partage et des moyens financiers et humains internes dédiés pour : <ul style="list-style-type: none"> - Gérer la partie administrative qui peut être lourde et chronophage - Evaluer le ou les opportunités (partenariats...), les risques, définir le cadre, - Revoir les demandes de partage - Extraire les données dans le bon format, de la bonne façon, avec les bons contrôles de qualité - Etablir les contrats avec les dispositifs permettant le partage - Dédier un financement pour tout le process
Manque de visibilité des données disponibles par les autres acteurs	Mettre en valeur les données à travers un catalogue de données mis à disposition de tous Ex. Portail Aviesan
Présence de contraintes réglementaires pour la réutilisation de données dans un objectif autre que celui principal	Anticiper en amont et préparer le cadre réglementaire en informant le patient sur la réutilisation de la donnée. <i>Ex. Portail d'information patient avec les informations requises pour pouvoir partager les données : la démarche scientifique, les droits des patients, les réutilisations possibles des données...</i>

Freins au partage de données	Prérequis de l'écosystème pour le partage de données
Risque du partage et manque de confiance	S'assurer de la réciprocité et de l'équité des acteurs publics et privés dans l'opérationnalisation du partage
Méconnaissance de l'écosystème français dans le cas d'un partage d'un global vers le local d'un autre acteur	Instaurer une continuité entre le local et le global pour éviter la duplication de process de gouvernance avec la CNIL, les acteurs publics ou privés et les ressources en interne
Absence d'un cadre légal stable sur le partage de données	Mettre en place un cadre légal fixe et pérenne (« filet de sécurité ») pour les industriels , en plus des guidelines, afin d'éviter à des laboratoires se mettant en disposition de partager ses données de devoir refaire des changements à la suite de modifications de recommandations.

Participants

Prénom	Nom	Organisme
Gaëlle	Desamericq	AMGEN
Hanna	Mhamdi	AMGEN
Antoine	Perpoil	AMGEN
Anne	Haziza	ARIIS
Muriel	Licour	AZ
Christophe	Roussel	EDWARDS
Philippe Jean	Bousquet	FIAC/INCa
Maria	del Pilar Schneider	IPSEN
Cécile	Fouret	MEDTRONIC
Pierre-Alexandre	Squara	PFIZER
Sylvie	Troy	PFIZER
Anne	Josseran	RESMED
Remy	Choquet	ROCHE
Laura	Oprea	ROCHE DIA
Aurélie	Lavorel	SNITEM
Camille	Nguyen Van	SNITEM
Lucie	de Léotoing	WLGORE