Assessment and challenges of data sharing by manufacturers

Working Group - Industry Data Sharing

In a context of **exponential development of data collection, storage and calculation capacities**, the exploitation of health data is essential and will bring progress for patients and the health system as a whole.

The colossal investments in data by pharmaceutical companies raise the question of their use, sharing and exploitation. In this context, a working group (WG) has been set up to **reflect on the challenges of sharing data generated by manufacturers**, to share feedback on this practice and to begin to **build a common vision on the purposes and prerequisites for this sharing.**

This working group mobilized 19 experts representing different healthcare companies: AMGEN, ARIIS, ASTRAZENECA, EDWARDS, INCA, IPSEN, MEDTRONIC, PFIZER, RESMED, ROCHE, ROCHE DIA, SANOFI, SERVIER, SNITEM, WLGORE.

This document is the result of several workshops organized with the members of the WG as well as a survey "***State of the art of data sharing generated by industrials***" sent to the members of the WG as well as to ARIIS members.

This work is the result of a joint reflection carried out within the framework of the strategic committee of the health industries and technologies (CSF-ITS). This work was supported by Iqvia.

This is in no way a commitment by ARIIS members to share their data.

Table of contents

I. What is data sharing?

In order to establish the assessment and the stakes of the data sharing of the industrialists, it is advisable at first time to **go back to the definition of the terms that compose it**.

a. "**Sharing**"

Sharing can be defined by its level of openness, with both extremes of closed and open sharing.

Closed sharing concerns a limited number of actors who agree to share data among themselves. In this framework, those who share the data control the why and the how and limit the use that can be made of it. This type of sharing is currently relatively widespread, for example through partnerships between two actors or co-constructed studies.

Conversely, *open sharing* is sharing with open access to all. Those who share the data have no control over the use of the data. This use is therefore not limited, beyond the scientific validity of the question studied and good practices. Open sharing is currently very little developed. The HDH platform or the SNDS are among the only real examples of open sharing.

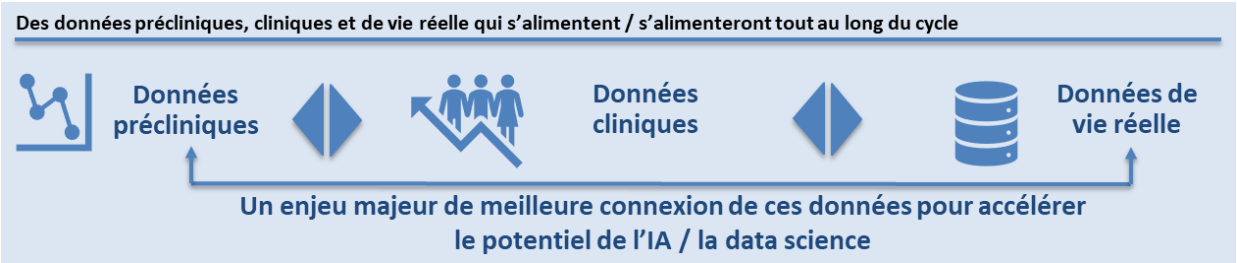
The challenge is to find the desired and realistic level of openness of sharing. The members of the WG reflected on this degree of openness and agreed that the challenge of the project was to develop open or semi-open sharing, which is still underdeveloped today and which would better serve the common interest.

b. "Industry data"

For the purposes of this project, WG members defined *shared/shareable data* as **data generated at the patient level for which the company is a processor, or funds/participates in funding its generation.**

These data follow the life cycle of health products and can therefore be **pre-clinical data, clinical data or real-life data.**

Figure 1: Preclinical, clinical and real-life data feed



Data from clinical and preclinical studies constitute the first "historical" source of exploitable data on pharmaceutical treatments, mainly in the form of randomized controlled trials. Clinical trials remain fundamental both to evaluate the toxicity of a drug (phase I) and to demonstrate the efficacy of a drug or medical device (phases II and III).

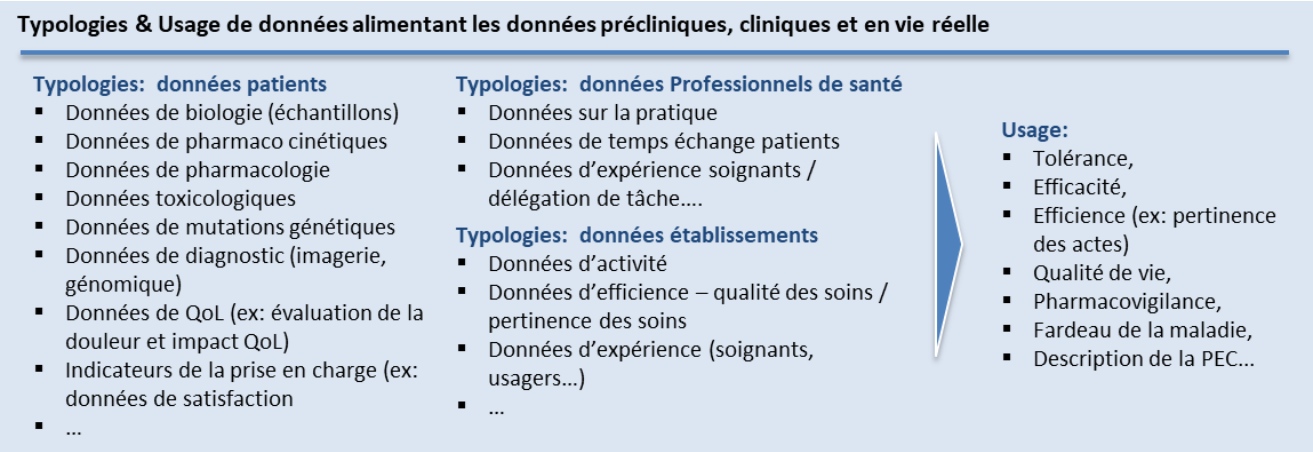
Real-life data describe the use of a treatment in an uncontrolled environment of the patient's daily life throughout the course of treatment, as opposed to interventional studies. Their analysis during clinical trials allows for the proof of potential benefits or risks related to medical products. They can be used in different ways, coupled with other technologies (AI) and innovative statistical analysis. The different real-life data produced and collected are the following:

- Early access data
- Data for observational studies
- Data for medico-economic studies
- Patient Experience D a t a - PROMS (Patient-Reported Outcomes Measures) and PREMS (Patient-Reported Experience Measures)
- Data from connected medical devices (care data)
- Outcome data from preclinical studies

The data feeding into preclinical, clinical or real-life studies, and potentially being shared for secondary use, fall into three data typologies (see Figure 2.):

- Patient data (e.g. biology data, imaging data),
- Health professional data (e.g. practice data)
- Facility data (e.g. activity data)

Figure 2: Typology and use of data



II. Who are the actors of this data sharing?

Data sharing actors fall into two categories: **producers and operators.**

The "producers" are the actors who produce the data. This production of data is part of a primary use (e.g., conducting a clinical trial). The producer of a database is the person, most often a legal entity, who has taken the "initiative and the risk of the investments"[1] necessary to create the content of the database.

[1] Source: Article L341-1 of the Intellectual Property Code

The role of the industrialist depends on the defined data sharing operational model. Its role can be be the following:

- *"Data controller*" means the natural or legal person, t h e public authority, the department or other body determines the purposes and means of the processing alone.
- "*Joint controllers*": the natural or legal person, public authority, department or other body determines the purposes and means of the processing jointly with others.
- "*Processor*" means the natural or legal person, public authority, department or other body that processes personal data on behalf of the controller[2]

If the data has value to the producer of the data, it can potentially also have value to other actors and for other purposes. If the data is shared by the producer of the data, it can allow these other actors to exploit the data. These actors are thus the "**exploiters**" of the data.

Data producers may also be operators for other data sources. The figure below summarizes the different producers and operators of clinical trial and real-life data.

Figure 3: Data Producers and Operators

	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. Why share?

The value of sharing lies in what can be done with the data collected, **beyond its primary purpose**. Sharing data does not require the transmission of the entirety of the data because **the objective is not to redo primary analyses of these data**. Indeed, the richness of sharing is not in the reuse of the data to redo the analysis already done (e.g. redo a safety) but in what else can be done with this data, a secondary use different from what the data was initially produced for.

[2] Source : Legifrance https://www.legifrance.gouv.fr/loda/id/JORFTEXT000035268202/ and HDH

For example, a laboratory that has generated imaging or biological data to measure the effect of treatment in a clinical trial could share its data to allow another actor to refine the identification of biomarkers, to work on AI or generally to set up interesting devices outside the clinical trial. In this case, it is not the "primary outcome" data that is interesting but the other data generated. These data can be rich in information (longitudinal data, etc.). According to the members of the working group, **laboratories do not use more than 10% of the data generated in clinical trials.**

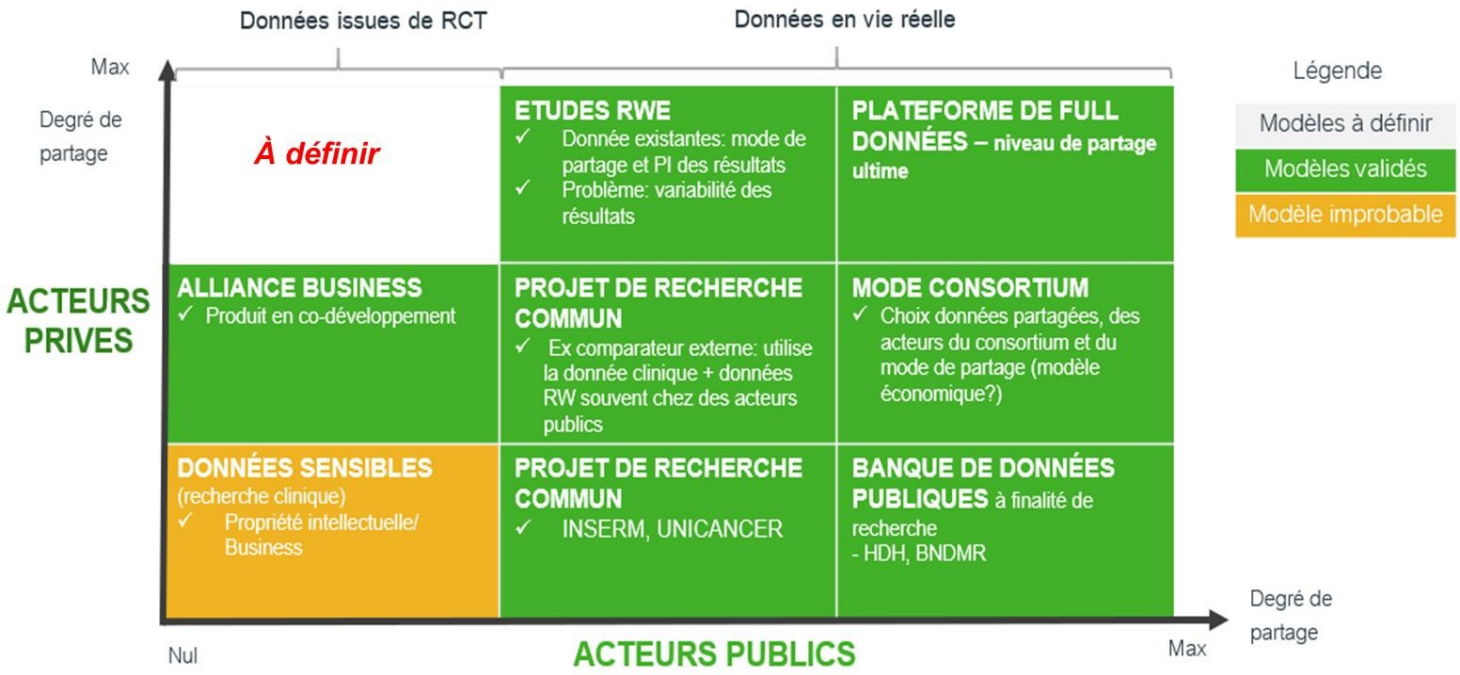
In the same way, the combined use of several data from different sources, thanks to their complementarity, can make it possible to generate new results, which will not necessarily be related to the primary results of the different sources. The added value of sharing can thus come from a crossing between several data sources. Thus, for example, any database, real-life or clinical trial, that allows patients to be linked to the SNDS, is a good subject for sharing. For *example,* for real-life data, the study of ATU data put into perspective with SNDS data can allow for a medium/long-term survival analysis.

Data sharing can also increase the credibility of the data generated through the use **of a single official database for analysis**. For example, authorities could work on the same shared real-life database as the laboratory, while ensuring credibility through an external quality certification process, thus avoiding comparisons of data from different sources.

IV. Which sharing models?

The possibility and willingness to share depends on the nature of the data (test d a t a clinical data, real-life data) and the stakeholders involved (public actors, private actors). As illustrated in Figure 2, nine categories of sharing can be highlighted.

Figure 4: Data Sharing Matrix



V. In what context should we share?

Data sharing, **especially open** data sharing, however, requires the definition of good practices.

1. Sharing must be supervised:
   - **Governance scenarios** must be defined: What are the scenarios? Which ones are the most realistic? What are the pros/cons?
   - The commitment should not be on the one who produces the data but **on the one who uses it**
   - It is also necessary to define **the prohibited purposes** (as for the SNDS) as well as the rules of use in a charter of good practice in order to limit the risks linked to sensitive data (e.g. use of clinical data by the competitor to question a product)

2. Sharing arrangements may differ depending on the type of data being shared.

Thus, since ***clinical trial data*** is more strategic, it is appropriate to incorporate **reinforced prohibited purposes**. The sharing of these data implies constraints because the laboratory that shares them must keep control of the communication of these data. The timing of sharing after data generation is thus an essential element to take into account: data sharing can only be done once the marketing authorization has been obtained**. The separation of strategic data and data open to sharing must be anticipated** by the manufacturer who will generate/share the data.

Real-life data also require rules of good use, even if they are less strict than for clinical trials. As observational data are less sensitive**, sharing them is easier and the timing of sharing is less strategic**. The data that can be shared are the basic characteristics of the patients (allowing several data to be linked together), the products that the patient has received (e.g. in a UTA) and the results data subject to controlling the purposes (allowing the transposability of the SOC but limiting the head to head).

The method used by the person who wants to use the data should not be limited; the method being only a tool and not an end.

In a secondary use of data, the data has already been generated so **the secondary users would not have to bear the cost of the initial production of the data**. Since the cost of producing the data has already been amortized, the most important cost is the cost of analyzing the data in the context of this secondary use. Financial valuation is not an issue for data sharing because beyond the cost of producing / making the data available, it is the use that is made of it that is interesting, and that will create value. It is therefore the "scientific valorization" that should be considered.

The quality of the shared data and the subject of the study to which this sharing must respond must be verified by a trusted third party (such as CESRESS, which evaluates the interest of work requests on SNDS data). It appears necessary to define an inter-industry process to evaluate the risk or not in relation to the data requested, to the analysis envisaged and thus to limit misinterpretations of data

VI. Where do we stand today?

a. Findings and examples of sharing

The first observation made during the workshops and illustrated by the answers to the questionnaire "***State of the art of data sharing generated by industrialists***" is the **weakness of current data sharing**. This sharing, when it is carried out, is still done on very restricted perimeters, highlighting in particular **a certain reluctance on the part of manufacturers to share their data.**

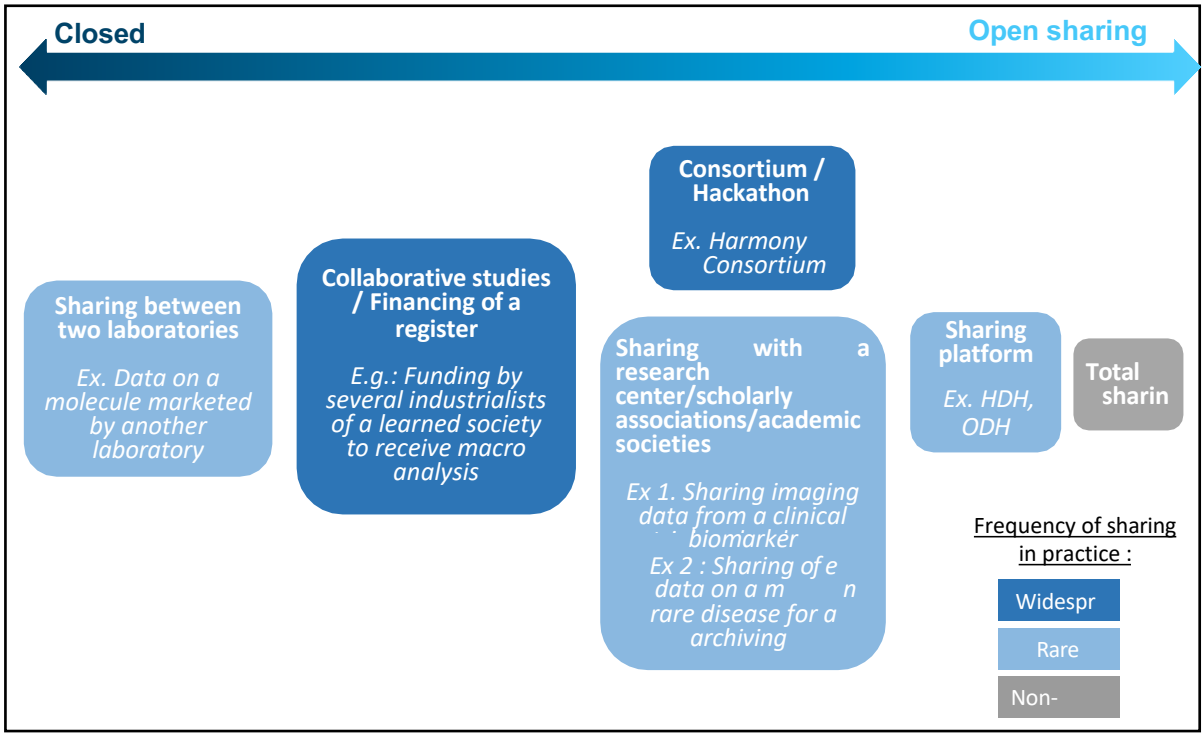
The main experiences of sharing that have emerged concern mainly closed sharing, between a limited number of actors, and "controlled". For example, some industrialists work with academics in the framework of collaborative studies where two practices were highlighted:

- One or more laboratories fund an "instrument" to conduct a study
- A laboratory uses existing academic data to conduct a study

In both cases, **industry does not have direct access to the raw data**. Industry all fund registries in exchange for consolidated reports. The academic or other team managing the registry does the analysis and shares it with the "funders", who do not have direct access to the raw data. Similarly, when sharing data from a laboratory to academics, the latter may or may not have access to the raw data.

The data shares highlighted in the workshops were positioned along the axis of the degree of openness of sharing (see Figure 1).

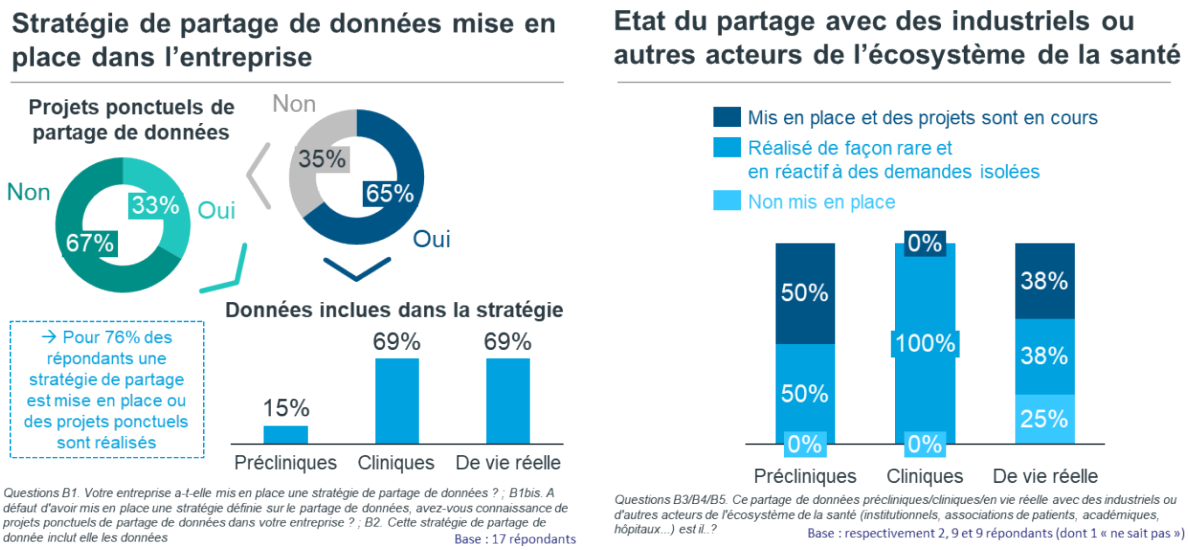
Figure 5: Examples of data sharing by degree of openness and frequency



Regarding the type of data shared, real-life data sharing is currently more developed than clinical trial data sharing. This is linked to the more strategic nature of the latter, which is derived from the development of a product but is also used for development.

In the survey "*State of the art of data sharing generated by manufacturers*", which was conducted among 17 manufacturers, 76% of respondents indicated that a data sharing strategy was in place in their company or that they had specific data sharing projects. These strategies concern both clinical and real-life data. However, if we look at the status of data sharing, it is **only done rarely and in response to isolated requests for clinical data**. **Sharing of real-life data is a little more advanced**, with more than a third of respondents indicating that sharing is in place and that projects are underway.

Figure 6: Data Sharing Practice Strategy and Sharing Status



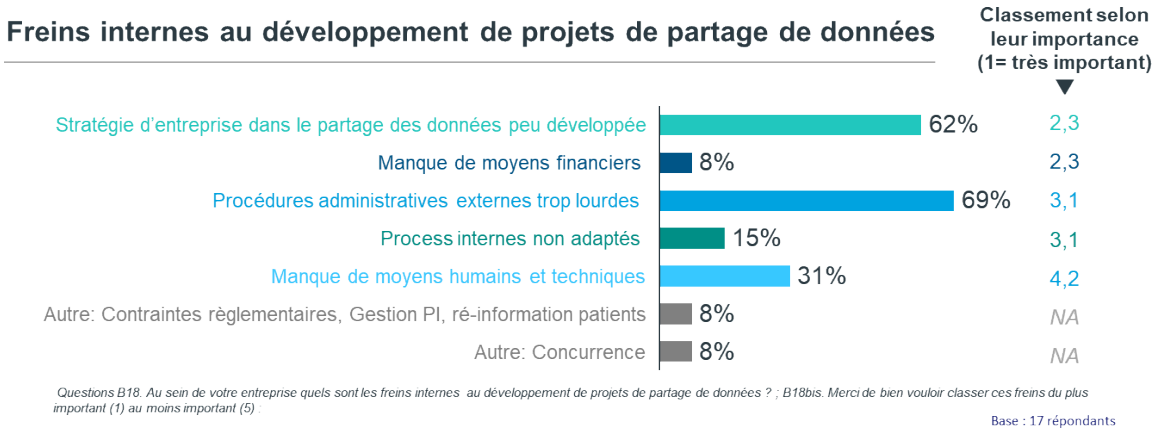
Concerning the type of sharing carried out, the survey shows that it **is the relatively closed sharing that is most practiced by manufacturers**. **For preclinical and clinical data, 100% of the sharing carried out is closed sharing**, whether it is multi-actor or bilateral.

VII. What are the obstacles to sharing and how can they be overcome?

The current limited experience with data sharing is related to various disincentives still present for data sharing.

Our survey "*State of the art of data sharing generated by manufacturers*" has highlighted the most common and most important obstacles for manufacturers today. **Cumbersome external administrative procedures** are the most common impediment today (69% of respondents), followed **by an underdeveloped corporate strategy for data sharing** (62% of respondents). In terms of importance, manufacturers rank this last obstacle as the most important, on the same level as the lack of financial means, but which is less widespread (8% of respondents).

Figure 7: Internal barriers to the development of data sharing projects



Faced with the various obstacles that manufacturers may encounter in sharing their data, **various key prerequisites have been highlighted** along two axes: the prerequisites of the manufacturer and the prerequisites of the ecosystem.

Barriers to data sharing	Industry requirements for data sharing
Weakness of the strategy in the sharing of data	Organize an **internal administrative framework conducive to** sharing and **dedicated internal financial and human resources** to : - Manage the administrative part which can be heavy and time consuming - Evaluate the opportunities (partnerships...), the risks, define the framework, - Review sharing requests - Extract data in the right format, in the right way, with the right quality controls - Establish contracts with devices allowing sharing - Dedicate funding for the entire process
Lack of visibility of available data by other actors	Enhance the value of the data **through a catalog of data made available to all** Ex. Aviesan portal
Presence of regulatory constraints on the reuse of data for a purpose other than the primary purpose	Anticipate upstream and **prepare the regulatory framework by informing the patient about the reuse of the data.** *Ex. Patient information portal with the information required to share data: the scientific process, patient rights, possible reuses of data, etc.*

Barriers to data sharing	Ecosystem requirements for data sharing
Risk of sharing and lack of trust	Ensure **reciprocity and equity of public actors and private sector** in the operationalization of the sharing
Lack of knowledge of the French ecosystem in the case of sharing from a global to a local level actor	Establish continuity between the local and the global to avoid duplication of governance processes with the CNIL, public or private actors and internal resources
Lack of a stable legal framework on data sharing	Establish **a fixed and permanent legal framework ("safety net") for industry**, in addition to the guidelines, in order to avoid laboratories that are willing to share their data having to make changes to the Following changes in recommendations.

Participants

First name	Name	Organization
Gaelle	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	of Léotoing	**WLGORE**