

## **Questionnaire d'auto-évaluation de la gestion de la qualité d'une base de données en santé pour le partage à des fins de recherche**

La qualité des données est essentielle pour la recherche et l'innovation en santé et incontournable pour leur partage. Sous l'impulsion du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI), un groupe de travail<sup>1</sup> s'est mobilisé pour proposer un outil d'auto-évaluation pour les responsables de base de données en santé. L'objectif est d'identifier les points forts et faibles de la gestion de la qualité et du partage. Cet outil n'a pas vocation à interclasser des bases de données. Il doit être utilisé dans cette version sans y apporter de modifications.

---

<sup>1</sup> Romain Casey (OFSEP, Épidémiologiste), Marie Zins et Marcel Goldberg (INSERM, Épidémiologistes, cohorte Constances), Yohann Foucher (MESRI, Chargé de mission), Christophe Bonaldi (Santé Publique France, Statisticien), Emmanuel Bacry (Health Data Hub, Directeur scientifique), Emilie Lanoy (Institut Gustave Roussy, Statisticienne), Paulo Carita (Sanofi, Économie de la santé), Laurie Levy-Bachelot (MSD, Directrice des études en vie réelle), Philippe-Jean Bousquet (INCA, Directeur de la Direction Observation, Science des données, Évaluation).

### **Structuration pour l'ouverture à la recherche**

1. Existe-t-il une publication<sup>2</sup> définissant la base de données et/ou son protocole dans une revue à comité de lecture (critères d'inclusion, variables collectées, etc.) ?

Oui    En cours    Non

Si non, existe-t-il un document accessible avec au minimum les informations suivantes ?  
Population cible, critères d'inclusion, variables collectées, modalité de collecte, contrôle qualité.

Oui                     Non

2. La base de données est-elle partageable<sup>3</sup> avec d'autres utilisateurs que votre équipe ?

Oui                     Non

3. Existe-t-il une procédure formalisée pour accéder à la base de données ?

Oui                     Non

4. Existe-il une instance scientifique pour l'évaluation des projets portant sur la base de données ?

Oui                     Non

5. Une équipe est-elle disponible pour accompagner l'utilisabilité des données (au minimum un data-manager) ?

Oui                     Non

6. Existe-il un dictionnaire des variables (au minimum leurs définitions, formats, et modes de collecte) ?

Oui                     Non

Si oui, les modifications au cours du temps des variables et de leurs codages (*versioning* des modifications) sont-elles intégrées ?

Oui                     Non                     Non applicable (étude transversale)

### **Procédures de mise en qualité des données**

1. Existe-t-il une formation du personnel de collecte ?

Oui                     Non                     Non applicable

2. Existe-t-il un protocole formalisé pour standardiser la collecte<sup>4</sup> ?

Oui                     Non

3. Existe-il des contrôles de cohérence à la collecte ?

Oui                     Non

4. Existe-il des audits de qualité internes ?

Oui                     Non

---

<sup>2</sup> Il peut s'agir d'une publication d'un consortium dans lequel s'inscrit la base de données (exemple d'un registre français s'inscrivant dans une organisation internationale).

<sup>3</sup> Une équipe tierce peut accéder aux données individuelles directement ou indirectement (extraction).

<sup>4</sup> Qu'il s'agisse d'une saisie manuelle ou automatisée depuis une autre source d'information, un document précisant la conformité des données recueillies et les procédures associées de standardisation.

5. Existe-il des audits de qualité externes ?  
 Oui       Non
6. Des classifications standardisées sont-elles utilisées (par exemple : CIM-10, questionnaires validés, etc.) ?  
 Oui       Non       Non applicable
7. Les instruments de mesure sont-ils régulièrement standardisés, calibrés, vérifiés ?  
 Oui       Non       Non applicable
8. Pour les collections biologiques, existe-t-il des consignes de traitement et de stockage des échantillons ?  
 Oui       Non       Non applicable
9. Pour l'imagerie, existe-t-il des consignes de réalisation des clichés ?  
 Oui       Non       Non applicable

### **Tracabilité du processus qualité**

1. Existe-il un cahier des charges de data management décrivant le processus de mise en qualité des données<sup>5</sup> ?  
 Oui       Non
2. Les modifications apportées aux données après la collecte sont-elles traçables ?  
 Oui       Non
3. Peut-on accéder à la fois aux données brutes et retravaillées (par exemple : nettoyage, recodage, etc.) ?  
 Oui       Non       Non applicable
4. Existe-t-il des indicateurs de qualité régulièrement calculés à partir des données collectées ?  
 Oui       Non

### **Normes et réglementation**

1. La base de données est-elle engagée dans un processus normatif (par exemple : BPC, BPE, NF, ISO, etc.) ?  
 Oui       En cours       Non
2. Le régime juridique de la base de données déclarée à la CNIL est :  
 Entrepôt de données<sup>6</sup>

---

<sup>5</sup> Nettoyage des données non qualitatives, suppression des doublons, standardisation des données pour le partage et l'interopérabilité.

<sup>6</sup> Délibération de la CNIL autorisant à collecter des données afin qu'elles puissent être secondairement réutilisées à des fins d'études, de recherches et d'évaluations dans le domaine de la santé.

Recherche, étude ou évaluation ponctuelle<sup>7</sup>

Une demande de changement du statut d'étude à celui d'entrepôt est en cours

3. Est-ce que l'information à l'inclusion des sujets prévoit la réutilisation secondaire de leurs données ?

Oui       Non

Si oui, les canaux d'information<sup>8</sup> des patients sont-ils fonctionnels ?

Oui       Non

### **Évaluation externe, publications et animation**

1. La base de données a-t-elle fait l'objet d'une évaluation par un organisme indépendant (par exemple : ANR, comité d'évaluation des registres, etc.) ?

Oui       Non

2. La liste des projets initiés à partir de la base est-elle publiquement accessible<sup>9</sup> ?

Oui       Non

3. La liste des publications scientifiques issues de la base avec les principaux résultats est-elle publiquement accessible ?

Oui       Non

4. Existe-il une communauté structurée d'utilisateurs de la base (par exemple : club utilisateurs, consortium) ?

Oui       Non

---

<sup>7</sup> Délibération de la CNIL autorisant à collecter des données pour une finalité précise et réponse à une question spécifique et ponctuelle. La durée de la recherche est limitée. Les données sont collectées spécifiquement pour les besoins de la recherche.

<sup>8</sup> Capacité à dispenser une information actualisée aux participants à l'étude (transmission d'une nouvelle notice d'information ou d'un bulletin d'information, site web, etc.).

<sup>9</sup> Action proactive pour mettre à disposition du public les informations actualisées. Le fait de transmettre ces informations uniquement sur sollicitation ne constitue pas un accès public.