

## Sommaire

|                                                          |   |
|----------------------------------------------------------|---|
| Description du Processus.....                            | 1 |
| Conditions annexes du processus.....                     | 1 |
| Lignes de force du processus.....                        | 2 |
| Intégration / harmonisation avec d'autres processus..... | 3 |
| Déroulement du processus.....                            | 4 |
| Description du déroulement.....                          | 5 |

## Description du Processus

Ce processus explique les démarches qui doivent être entreprises pour rassembler de manière systématique et fiable les données nécessaires pour la recherche, pour les traiter et les transformer, pour les valider et finalement les mettre à disposition de manière optimale pour un projet d'étude spécifique.

La description figurant dans la présente note est en premier lieu applicable aux données RCM/RFM couplées et, lors du traitement d'autres données, elle sera progressivement élargie à ces nouveaux champs d'application.

## Conditions annexes du processus

### *Concernant l'examen des données à caractère personnel*

- Voir également PR12 pour une discussion et une description complémentaires.
- Chaque projet d'étude sur des données à caractère personnel (par exemple, des données cliniques couplées) requiert une autorisation préalable du Comité sectoriel de la Banque carrefour de la Sécurité Sociale en vertu de l'art. 156, §4 de la Loi du 29 avril 1996 relatives aux dispositions sociales, modifié par l'art. 292, 5° du Chapitre 2 (Création du Centre fédéral d'Expertise des Soins de Santé), de la Loi-programme du 24-12-2002.
- Par *données cliniques couplées*, on entend les *Données cliniques minimales* (RCM), collectées semestriellement par le SPF Santé publique auprès des établissements de soins belges, et les *Données financières minimales* (RFM), recueillies par l'INAMI auprès des Organismes assureurs.
- Ces enregistrements de données comprennent des données à caractère personnel, il est vrai avec anonymisation irréversible de la clé patient sur la base du numéro d'inscription à la mutualité, et un codage réversible pour les institutions de soins belges. L'art. 261, 3° et

4°, de cette même Loi-programme du 24-12-2002 définit de manière univoque les notions de données anonymes et de données codées.

- En vertu de l'article 156 de la Loi du 29 avril 1996 relative aux dispositions sociales, modifié par l'art. 292, 4° de la Loi-programme susmentionnée du 24-12-2002, le transfert de données couplées vers le KCE doit se faire par le biais de la Cellule technique autorisée à effectuer et valider cette liaison.
- Ce transfert requiert l'élaboration d'une demande d'autorisation spécifique préalable adressée au Comité sectoriel de la Banque carrefour de la Sécurité Sociale.
- Les modalités et les aspects relatifs au fond de cette demande d'autorisation sont régis par la Loi du 8 décembre 1992 relative à la protection de la vie privée à l'égard du traitement de données personnelles et l'Arrêté royal du 13 février 2001 portant exécution de cette loi.
- Le traitement, l'échange et la conservation des données relatives à la santé se déroulent sous la surveillance et la responsabilité d'un médecin surveillant du KCE; son identité est communiquée au Comité sectoriel (article 26, § 1, Loi sur la Banque carrefour).

## **Lignes de force du processus**

### **1. *Concernant la création de logfiles et de summary tables with descriptives***

- Tous les traitements et transformations de données doivent être développés, affinés ainsi que documentés de manière systématique dans des "logfiles" qui décrivent de manière uniforme toutes les étapes de traitement et de validation.
- Avec ces logfiles, les summary tables with descriptives constituent un instrument important dans la documentation des étapes consécutives de la transformation/du traitement des données durant la recherche.
- Ce traitement/cette transformation concernent notamment souvent un (re)groupement d'entités (séjours, patients, APrDRG, ...) dans des groupes d'exclusion/inclusion et des sous-groupes spécifiques à l'étude.
- Après chaque (re)groupement, un summary table reprend sous la forme d'un (ou plusieurs) tableau(x) de totalisation par (sous)-groupe et le cas échéant des (sous)-totaux pour les champs numériques pertinents.
- La création de summary tables with descriptives est ainsi un processus itératif obligatoire dans la recherche qui permet au chercheur de résumer les effets de traitement et transformation successifs des données et de les saisir d'un coup d'œil ('fil rouge').
- En ce sens, ils constituent également un outil important pour la validation 3<sup>aire</sup> des données et la validation finale de la recherche (→ PR8 & 9)

## 2. *Concernant la validation de données*

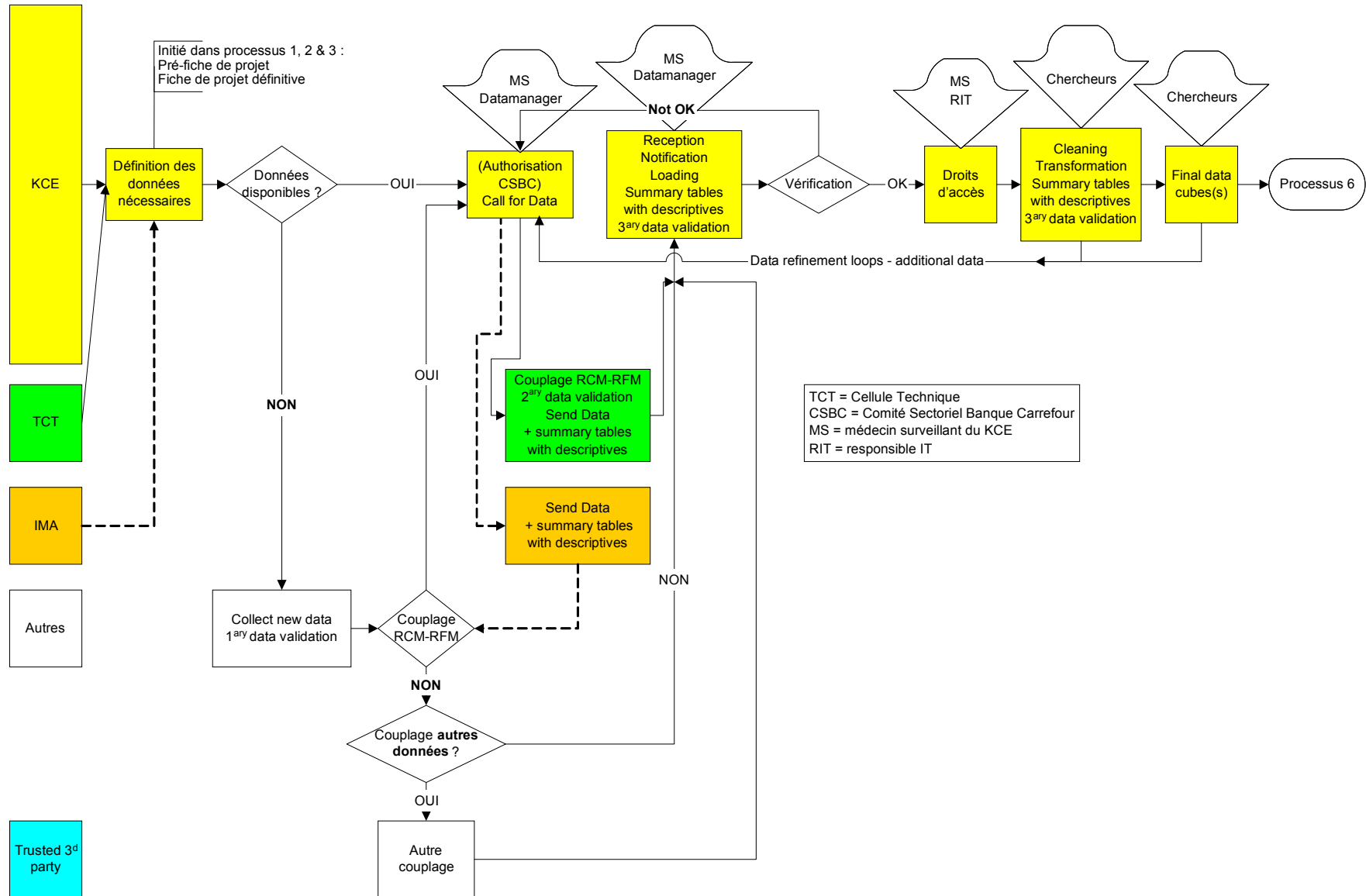
- LA VALIDATION DE DONNÉES doit être distinguée strictement de la VALIDATION DE LA RECHERCHE (→ PR8 & PR9) en ce sens que la validation des données est un sous-processus itératif supporté par une attitude de "constant awareness" relative à la fiabilité des données enregistrées dans le chef du datamanager et de chaque chercheur-analyste concerné par les projets d'études sur des datasets codés.
- La validation des données en tant que processus doit donc être examinée **par phases**. Pour plus de détails, nous vous renvoyons à la Note de processus 05.1 – Validation des données.
- Dans le contexte des projets d'études du KCE, il s'agit surtout de la "validation tertiaire" :
  - Object :
    - Essentiellement sur plusieurs ensembles d'enregistrements correspondants "logiques" (= verticaux), par exemple par patient, par pathologie, par hôpital,...
    - Toutefois aussi sur des summary tables with descriptives globaux et/ou sur plusieurs années : nombre de séjours couplés par organisme assureur, par hôpital,...
  - Actor : KCE
  - Output : création de 'ad hoc exclusionflags' lors du traitement/de la transformation des données au sein d'un projet de recherche spécifique (actuel Processus 5)
  - Effect :
    - Short term: exclusion des enregistrements non fiables ('doubtful input data') afin d'éviter la "contamination" des résultats finaux de l'étude
    - Long term: feed-back en vue de l'amélioration de la qualité dans la collecte de données par le(s) fournisseur(s) de données et/ou la liaison des données par la TCT ou trusted 3<sup>th</sup> party

### **Intégration / harmonisation avec d'autres processus**

La détermination des besoins de données et la demande des données chez le fournisseur de données doivent se faire le plus rapidement possible dans le processus de recherche. Ce processus peut en effet demander un certain temps, surtout si une autorisation du Comité sectoriel est nécessaire. La détermination des besoins de données et la demande se déroulent dès lors en parallèle avec le PR2 (pré-fiche de projet) jusqu'au PR3 (demande d'autorisation STC et fiche de projet définitive), avant la recherche proprement dite.

Le Processus 5 engendre, après la transformation, l'agrégation et la validation des données, une matrice de données ou un dataset "prêt à l'emploi", nécessaire pour le Processus 6 (Analyse des données) et est étroitement lié au Processus 12 (Datamanagement).

## Déroulement du processus



## Description du déroulement

### 1. Définition des données nécessaires

2. Les besoins de données résultent des sujets de recherche et des projets du KCE.
3. Dans le Topic Proposal Form, on demande déjà aux initiateurs des sujets de recherche (→ PR1) d'indiquer quelles sont les données qu'ils estiment indispensables pour étudier les sujets de recherche. L'évaluation des sujets de recherche sera notamment déterminée par une estimation de la faisabilité du rassemblement et du traitement des données indispensables (→ PR1).
4. Dans la spécification du sujet de recherche (= élaboration de la pré-fiche de projet) (→ PR2), on rédige une description des données indispensables (quelles sont les données indispensables, sont-elles disponibles et si oui, où se trouvent-elles, ...).
5. Dans l'élaboration de la fiche de projet finale (→ PR3), la description des données requises est finalisée.
6. La description des besoins de données est une tâche des Chercheurs (Research) et de l'Analyste Datamanagement (RI).

### 7. Rassemblement de nouvelles données

8. S'il ressort de ce qui précède que de nouvelles données sont nécessaires, le rassemblement de ces données sera prévu par le datamanager en concertation avec les chercheurs et le manager du projet. Cela commence dès l'élaboration de la pré-fiche de projet (→ PR2) pour être finalisé dans la fiche de projet définitive (→ PR3).

### 9. Demande d'autorisation

10. Dès que les besoins de données seront clairs (PR1 of PR2), le datamanager devra déterminer si une autorisation du Comité sectoriel est indispensable (→ PR12) lorsque les données requises concernent des **données cliniques couplées**. Dans ce cas, la description des besoins de données se fera en concertation avec le Contrôleur de la Cellule technique et ce sera concrétisé dans la demande d'autorisation requise auprès du CSBC ('Pre-authorisation Study').

### 11. Demande de données auprès du fournisseur de données

12. Dans cette étape, le KCE demande les données au fournisseur de données ('Call for data').
13. S'il s'agit de *données cliniques couplées*, la demande est envoyée à la *Cellule technique* accompagnée de l'autorisation du Comité sectoriel. (→ PR12).

### 14. Fourniture de données par le fournisseur de données

15. Après validation, les fournisseurs de données mettront les données à disposition sous différents formats (sur serveur dédié, via cd-rom, fichier plat,...).
16. On y ajoute un ou plusieurs summary table(s) with descriptives établis par le fournisseur de données.

### **17. Réception de données**

18. Les données sont reçues par le KCE (→ PR12)

19. Les données sont stockées et adaptées par le datamanager afin d'être utilisables pour la transformation en fonction de la recherche spécifique.

20. Le datamanager envoie une 'Data Received Notification' au(x) fournisseur(s) de données.

21. Il élabore un premier (une première série de) summary table(s) with descriptives : cf. ligne de force 1 ci-dessus.

22. On part du principe que les fournisseurs de données ont appliqué la validation 1<sup>aire</sup> ou 2<sup>aire</sup> nécessaires (en cas de liaison par la TCT) sur leurs datasets. Le datamanager est chargé de vérifier cette étape et le cas échéant de débiter la validation 3<sup>aire</sup>. S'il constate des problèmes, il prend contact avec le(s) fournisseur(s) de données pour rectifier les données fournies.

### **23. Détermination des droits d'accès**

24. Dans chaque projet, on définit concrètement les droits d'accès aux données dans le système pour les différentes personnes concernées par la recherche. Cela se fait sur l'instruction du Médecin surveillant à l'attention du responsable IT qui assure la programmation des droits d'accès dans le système de données (→ PR12).

### **25. Data Cleaning, transformation, summary table(s) with descriptives supplémentaires & validation tertiaire**

26. In globo, il s'agit de:

- L'exclusion des enregistrements inconsistants ou non-pertinents (cleaning)
- L'agrégation des enregistrements spécifique à l'étude, avec ou sans introduction de variables spécifiques à l'étude (transformation)
- La suite de la création de tous les summary tables with descriptives supplémentaires requis
- La poursuite de la validation de données 3<sup>aire</sup>, finalisée dans PR 6

27. Tous ces traitements et transformations de données devront être développés, affinés et documentés de manière systématique ("logfiles"). Ces tâches sont effectuées par l'analyste-chercheur, assisté ou non du datamanager.

### **28. Data Refinement Loops**

29. Au fil de l'avancement de la recherche (PR5) et de l'analyse des données (→ PR6), un besoin de données supplémentaires peut être constaté. Cela engendre des data refinement loops. Le data refinement peut se faire au sein des datasets déjà obtenus existants, ou peut nécessiter la demande de données supplémentaires (avec la nécessité éventuelle de déposer une nouvelle demande d'autorisation).

### **30. Data cubes**

31. L'output final du processus 5 comprend:

- un ou plusieurs data cubes ad hoc et "prêts à l'emploi"
- Le(s) logfile(s) de toutes les étapes du traitement/de la transformation des données
- la série complète de summary tables with descriptives