Procesdoc 05 Onderzoek van de gegevens

Inhoud

Omschrijving Proces 1
Randvoorwaarden voor het proces 1
Krachtlijnen van het proces 2
Inpassing /afstemming met andere processen 3
Procesflow 4
Beschrijving Flow 5
Ondersteunende informate ifv personeelsbehoefteplanning 7
Consolidatie behoeften ondersteunende systemen 8

Omschrijving Proces

In dit proces wordt verduidelijkt welke stappen moeten gezet worden om op systematische en betrouwbare wijze de voor het onderzoek noodzakelijke data te verzamelen, te bewerken en te transformeren, te valideren en finaal optimaal ter beschikking te stellen voor een specifiek studieproject.

De omschrijving in deze nota is in de eerste plaats van toepassing op de gekoppelde MKG/MFG gegevens en zal bij het verwerken van andere data progressief worden uitgebreid naar die nieuwe toepassingsgebieden.

Randvoorwaarden voor het proces

M.b.t. onderzoek op persoonsgebonden gegevens

- Zie ook PR12 voor aanvullende omschrijving en bespreking.
- Elke project van studie op persoonsgebonden gegevens (bijv. gekoppelde ziekenhuisgegevens) vereist een voorafgaande machtiging door het Sectorieel Comité (SCKSZ) van de Kruispuntbank van de Sociale Zekerheid krachtens Art. 156, §4 van de Wet van 29 april 1996 houdende sociale bepalingen, gewijzigd door Art. 292, 5° in Hoofdstuk 2 (Oprichting van het Federaal Kenniscentrum voor de Gezondheidszorg), van de Programmawet d.d. 24-12-2002.
- Onder *gekoppelde ziekenhuisgegevens* wordt verstaan de *Minimale Klinische Gegevens* (MKG), door de FOD Volksgezondheid semestrieel ingezameld bij de Belgische Verzorgingsinstellingen, en *de Minimale Financiële Gegevens* (MFG), door het RIZIV semestrieel ingezameld bij de Verzekeringsinstellingen.

- Deze gegevensrecords bevatten persoonsgebonden gegevens, zij het met irreversiebele anonimisatie van de patiëntensleutel, gebaseerd op zijn inschrijvingsnummer bij zijn mutualiteit, en reversiebele codering voor de Belgische verzorgingsinstellingen. Art. 261, 3° en 4°, van diezelfde Programmawet d.d. 24-12-2002 definieert eenduidig de begrippen anonieme gegevens en gecodeerde gegevens.
- Krachtens artikel 156 van de Wet van 29 april 1996 houdende sociale bepalingen, gewijzigd door Art. 292, 4° van hogervermelde Programmawet d.d. 24-12-2002, dient de overdracht van gekoppelde gegevens naar het KCE te gebeuren via de Technische Cel die gemachtigd is deze koppeling uit te voeren en te valideren.
- Deze overdracht vereist de opmaak van een voorafgaandelijke specifieke machtigingsaanvraag bij het Sectorieel Comité van de Kruispuntbank van de Sociale Zekerheid.
- De modaliteiten en inhoudelijke aspecten van deze machtigingsaanvraag worden geregeld in de Wet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens en het Koninklijk Besluit van 13 februari 2001 ter uitvoering van deze wet.
- Het behandelen, uitwisselen en bewaren van gezondheidsgegevens gebeurt onder het toezicht en de verantwoordelijkheid van een geneesheer Toezichtshouder van het KCE; diens identiteit wordt meegedeeld aan het Comité (artikel 26, § 1, Kruispuntbankwet).

Krachtlijnen van het proces

1. M.b.t. aanmaak van logfiles en summary tables with descriptives

- Alle databewerkingen en -transformaties moeten stelselmatig worden ontwikkeld, verfijnd en ook gedocumenteerd in zgn. logfiles die alle bewerkings- en validatiestappen eenduiding beschrijven.
- Samen met deze logfiles leveren summary tables with descriptives een belangrijk instrument in de dokumentatie van de consecutieve stappen in de gegevensverwerking/transformatie tijdens het onderzoek.
- Deze verwerking/transformatie betreft o.m. vaak een (her)groepering van entiteiten (verblijven, patiënten, APrDRG's,...) in studie-specifieke inclusie/exclusiegroepen en subgroepen.
- Na elke (her)groepering dokumenteert een summary table de effecten ervan onder de vorm van een (of meerdere) totalisatietabel(len) met aantallen per (sub)groep en desgevallend (sub)totalen van relevante numerieke velden.
- Het aanmaken van summary tables with descriptives is aldus een obligaat iteratief proces binnen het onderzoek dat de onderzoeker in staat stelt de effecten van opeenvolgende gegevensverwerkingen en -transformaties de resumeren an aldus te overzien ('rode draad').
- In die zin vormen zij ook een belangrijk hulpmiddel voor de 3aire gegevensvalidatie én de uiteindelijke onderzoeksvalidatie (→ PR8 & 9)

2. M.b.t. de validatie van gegevens

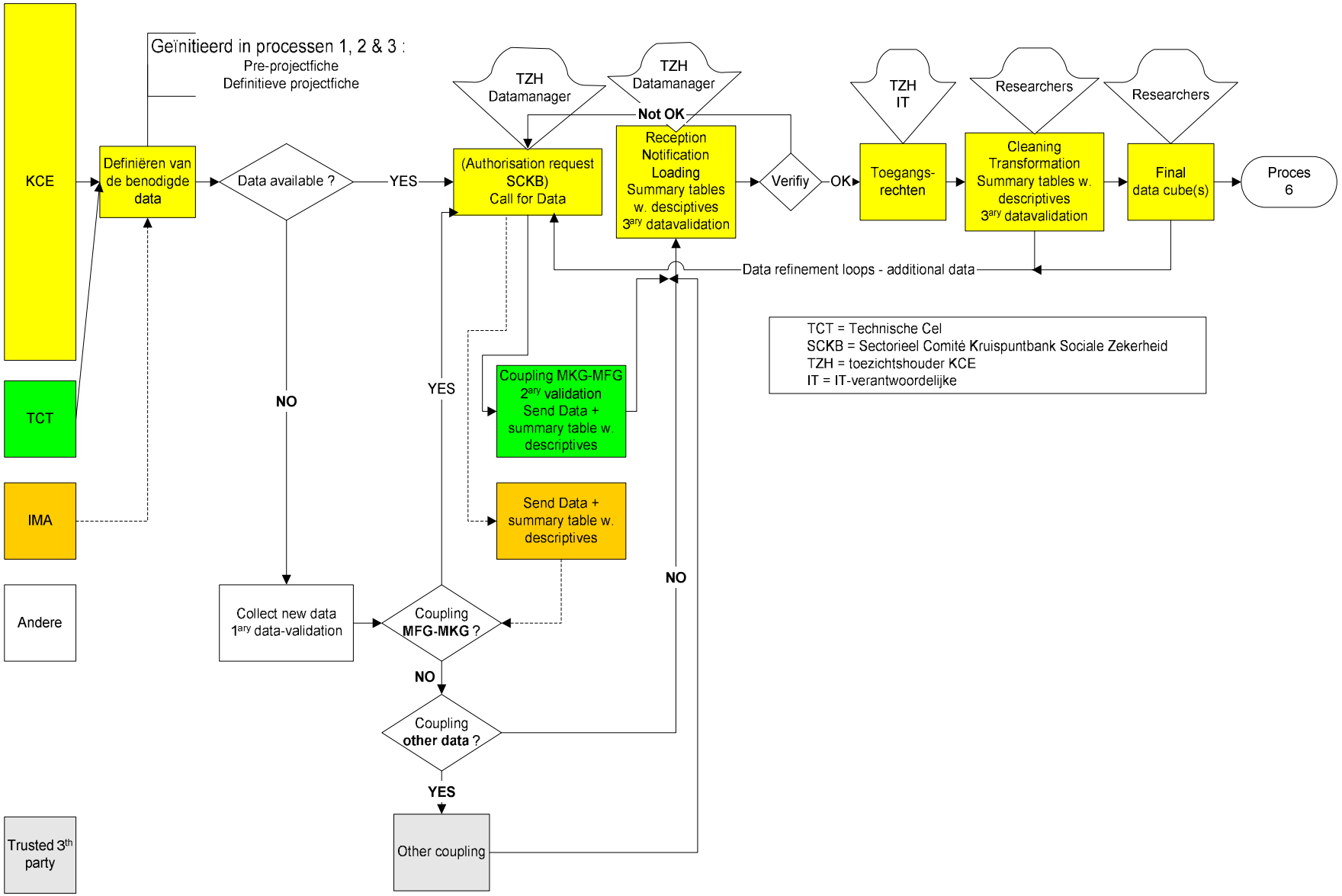
- GEGEVENSVALIDATIE dient strict onderscheiden te worden van ONDERZOEKSVALIDATIE (➔ PR8 & PR9) in die zin dat gegevensvalidatie een iteratief deelproces is gedragen door een attitude van 'constant awareness' omtrent de betrouwbaarheid van geregistreerde gegevens in hoofde van datamanager en elke analyst-onderzoeker betrokken in studieprojecten op gecodeerde datasets.
- Voor een uitgebreide bespreking van gegevensvalidatie verwijzen we naar Procesnota 05.1.
- In de context van studieprojecten van het KCE betreft het vooral de zgn. tertiaire validatie :
    - Object :
        - Essentieel op meerdere 'logisch' bijeenhorende recordsets (= vertikaal), bijv. per patiënt, per pathologie, per ziekenhuis,...
        - Echter ook op globale summary tables with descriptives en/of over meerdere jaren : aantal gekoppelde verblijven per VI, per ziekenhuis,..
    - Actor : KCE
    - Output : creatie van 'ad hoc exclusionflags' bij verwerking/transformatie van de gegevens binnen een specifiek onderzoeks-project (huidig Proces 5)
    - Effect :
        - Short term: exclusie van niet-betrouwbare registraties ('doubtful input data') en dus vermijden van 'contaminatie' van de uiteindelijke studieresultaten
        - Long term: terugkoppeling tot kwaliteitsverbetering in de gegegevensinzameling door de gegevensverstrekker(s) en/of de gegevenskoppeling door de TCT of trusted 3th party

Inpassing /afstemming met andere processen

De bepaling van de databehoeften en de aanvraag van de data bij de dataverstrekker dient zo snel mogelijk in het onderzoeksproces plaats te vinden. Dit proces kan immers enige tijd vergen, zeker als een machtiging van het Comité noodzakelijk is. De bepaling van de databehoeften en de aanvraag vindt dan ook parallel plaats met PR2 (pre-projectfiche) tot PR3 (machtigingsaanvraag STC en definitieve projectfiche), voorafgaand aan het feitelijke onderzoek.

Proces 5 leidt, na ad hoc transformatie, aggregatie en gegevensvalidatie, tot een 'klaar voor gebruik' gegevensset of gegevensmatrix, nodig voor Proces 6 (Analyse Gegevens) en is nauw verbonden met proces 12 (Datamanagement).

Procesflow


KCE
TCT
IMA
Andere
Trusted 3th party
Geïnitieerd in processen 1, 2 & 3 :
Pre-projectfiche
Definitieve projectfiche
Definiëren van de benodigde data
Data available ?
YES
NO
TZH Datamanager
(Authorisation request SCKB)
Call for Data
TZH Datamanager
Not OK
Reception
Notification
Loading
Summary tables w. desciptives
3ary datavalidation
Verifiy
OK
TZH IT
Toegangs-rechten
Researchers
Cleaning
Transformation
Summary tables w. descriptives
3ary datavalidation
Researchers
Final data cube(s)
Proces 6
Data refinement loops - additional data
TCT = Technische Cel
SCKB = Sectorieel Comité Kruispuntbank Sociale Zekerheid
TZH = toezichtshouder KCE
IT = IT-verantwoordelijke
Coupling MKG-MFG
2ary validation
Send Data + summary table w. descriptives
Send Data + summary table w. descriptives
YES
NO
Collect new data
1ary data-validation
Coupling MFG-MKG ?
NO
Coupling other data ?
YES
Other coupling

Beschrijving Flow

1. **Definiëren van de benodigde data**
2. De databehoeften komen voort uit de onderzoeksonderwerpen en projecten van het KCE.
3. De indieners van onderzoeksonderwerpen (➔ PR1) wordt in de Topic Proposal Form reeds gevraagd aan te geven welke data zij denken dat noodzakelijk zijn om het onderzoeksonderwerp te onderzoeken. De beoordeling van de onderzoeksonderwerpen zal mee bepaald worden door een inschatting van de haalbaarheid van de noodzakelijke datavergaring en verwerking (➔ PR1).
4. Bij de specificatie van het onderzoeksonderwerp (=opmaak van de pre-projectfiche (➔ PR2) wordt een omschrijving van de noodzakelijke data opgesteld (welke data zijn noodzakelijk, zijn ze beschikbaar en, zo ja, waar zijn ze, …).
5. Bij de opmaak van de finale projectfiche (➔ PR3) wordt de omschrijving van de vereiste data gefinaliseerd.
6. De omschrijving van de databehoeften is een taak van de Onderzoekers (Research) samen met de Analist Datamanagement (RI).
7. **Inzameling van nieuwe data**
8. Indien uit het voorgaande blijkt dat nieuwe data nodig zijn, zal in de inzameling hiervan voorzien worden door de datamanager in samenspraak met de onderzoekers en de projectmamanger. Dit begint reeds bij de opmaak van de pre-projectfiche (➔ PR2) en wordt gefinaliseerd in de definitieve projectfiche (➔ PR3).
9. **Machtigingsaanvraagsaanvraag**
10. De datamanager zal, van zodra de databehoeften duidelijk zijn (PR1 of PR2), moeten bepalen indien een machtiging van het Sectorieel Comité noodzakelijk is (➔ PR12) i.c. wanneer de vereiste data **gekoppelde ziekenhuisgegevens** betreffen. In dit geval zal de omschrijving van de databehoeften gebeuren in overleg met de Technische Cel en geconcretiseerd worden in de vereiste machtigingsaanvraag bij het STC ('Pre-authorisation Study').
11. **Aanvraag data bij dataverstrekker**
12. Binnen deze stap vraagt het KCE de data aan bij de dataverstrekker ('Call for data').
13. Indien het gaat om *gekoppelde ziekenhuisgegevens* wordt de aanvraag vergezeld van de machtiging van het sectorieel Comité verzonden naar de *Technische Cel*. (➔ PR12).
14. **Verstrekking data door dataverstrekker**
15. De dataverstrekkers zullen de data, na validatie, ter beschikking stellen in verschillende formats (op dedicated server, via cd-rom, flat file,…).
16. Hieraan wordt een of meerdere summary table(s) with descriptives opgesteld door de dataverstrekker toegevoegd.

17. Ontvangst data

18. Data worden door KCE ontvangen (➔ PR12)

19. De data worden door de datamanager dusdanig gestockeerd en vormgegeven dat ze bruikbaar worden voor transformatie in functie van het specifieke onderzoek.

20. De datamanager stuurt een 'Data Received Notification' aan de dataverstrekker(s).

21. Hij maakt een eerste (reeks) summary table(s) with descriptives : zie krachtlijn 1 hoger.
22. Er wordt van uitgegaan dat de dataverstrekkers de nodige 1aire of 2aire gegevensvalidatie (in geval van koppeling door de TCT) hebben doorgevoerd op hun datasets. Verificatie hiervan en desgevallend eerste aanzet tot 3aire validatie gebeurd door de datamanager en bij vaststelling van problemen wordt contact genomen met de dataverstrekker(s) voor rectificatie van de dataverstrekking.

23. Bepaling toegangsrechten

24. Binnen elk project wordt concreet de toegangsrechten tot de data in het systeem vastgelegd voor de verschillende betrokkenen bij het onderzoek. Dit gebeurt op instructie van de Geneesheer Toezichthouder t.a.v. de IT-verantwoordelijke die instaat voor het programmeren van deze toegangsrechten in het datasysteen (➔ PR12).

25. Data Cleaning, transformatie, bijkomende summary table(s) with descriptives & 3ary validatie

26. In globo komt dit neer op:

    - exclusie van inconsistente of niet-relevante records (cleaning)
    - studiespecifieke aggregatie van records al dan niet met introductie van studie-specifieke variabelen (transformatie)
    - verdere aanmaak van alle vereiste bijkomende summary tables with descriptives
    - voorzetting van de 3aire gegevensvalidatie, die gefinaliseerd wordt in PR 6

27. Al deze databewerkingen en -transformaties zullen stelselmatig moeten worden ontwikkeld, verfijnd en ook gedocumenteerd (zgn. logfiles). Deze taken worden uitgevoerd door de analist onderzoeker al dan niet bijgestaan door de datamanager.

28. Data Refinement Loops

29. Naarmate het onderzoek (PR5) en de data-analyse (→ PR6) vordert kan er een behoefte aan bijkomende data vastgesteld worden. Dit geeft aanleiding tot data refinement loops. De data refinement kan gebeuren binnen de bestaande, reeds verkregen datasets, of kan vereisen dat bijkomende data worden opgevraagd (waarbij eventueel ook nieuwe machtigingsaanvraag moet worden ingediend).

30. Data cubes

31. De finale output van de proces 5 omvat :

    - één of meerdere ad hoc en 'klaar voor gebruik' data cubes
    - de logfile(s) van alle stappen in de gegevensverwerking/transformatie
    - de volledige reeks van summary tables with descriptives

Ondersteunende informate ifv personeelsbehoefteplanning

Hieronder zijn een aantal rollen uitgewerkt die een taak hebben binnen het proces 5.
Bepaalde rollen zijn combineerbaar met elkaar, en kunnen door eenzelfde persoon uitgevoerd worden. De uitwerking hiervan gebeurt in personeelsbehoefteplan (zie ook Proces 12 Datamanagement).

Wettelijk Vastgelegde Rol: Toezichthoudend Geneesheer

- Toezicht en verantwoordelijkheid over behandelen, uitwisselen en bewaren van gezondheidsgegevens gebeurt onder het toezicht en de verantwoordelijkheid van een geneesheer
- bij naam aanwijzen van de personen die de gezondheidsgegevens mogen opslaan, raadplegen, wijzigen, verwerken of vernietigen;

Wettelijk Vastgelegde Rol: Veiligheidsconsulent

- De veiligheidsconsulent staat in voor de veiligheid van de persoonsgegevens die door zijn instelling worden verwerkt of uitgewisseld; hij verstrekt deskundige adviezen aan de dagelijkse bestuurder van de instelling en voert de door deze laatste aan hem toevertrouwde opdrachten uit (artikel 25 Kruispuntbankwet).

Rol: Functionele en technische ondersteuning

- ondersteunen onderzoekers-analisten bij vertaling van onderzoeksvraag naar behoeften inzake data en data-analyse
- loading van de data
- ondersteuning van de data-analisten bij de transformation van de data (hanteren ETL-log, metadata,…)
- ondersteuning van de data-analisten bij de data-analyse (hanteren tools, programmering,…)
- mee instaan voor gestructureerde feedback aan de dataverstrekkers (secondary and tertiary validation feedback loops)

Rol: ICT-beheerder

- Beheer Software (tools expert, versiebeheer,…)
- Beheer Hardware (database administrator, onderhoud, backup, security,…)

Rol: Onderzoekers/ Data-analisten (Research)

- Uitwerken onderzoeksvraag en mee databehoeften aangeven
- Instaan voor de transformatie van de data (eventueel ondersteund door datamanager)
- Instaan voor correcte ETL-logging
- Data-analyse met reporting- en analysetools

Consolidatie behoeften ondersteunende systemen

Zie Proces 12 Datamanagement.