Inventaire des bases de données de soins de santé

KCE reports vol. 30B

Federaal Kenniscentrum voor de gezondheidszorg
Centre fédéral d'expertise des soins de santé
2006

Le Centre fédéral d'expertise des soins de santé

Présentation : Le Centre Fédéral d'Expertise des Soins de Santé est un parastatal, créé le 24 décembre 2002 par la loi-programme (articles 262 à 266), sous tutelle du Ministre de la Santé publique et des Affaires sociales, qui est chargé de réaliser des études éclairant la décision politique dans le domaine des soins de santé et de l'assurance maladie.

Conseil d'administration

Membres effectifs : Gillet Pierre (Président), Cuypers Dirk (Vice-Président), Avontroodt Yolande, De Cock Jo (Vice-Président), De Meyere Frank, De Ridder Henri, Gillet Jean-Bernard, Godin Jean-Noël, Goyens Floris, Kesteloot Katrien, Maes Jef, Mertens Pascal, Mertens Raf, Moens Marc, Perl François, Smiets Pierre, Van Massenhove Frank, Vandermeeren Philippe, Verertbruggen Patrick, Vermeyen Karel

Membres suppléants : Annemans Lieven, Boonen Carine, Collin Benoît, Cuypers Rita, Dercq Jean-Paul, Désir Daniel, Lemye Roland, Palsterman Paul, Ponce Annick, Pirlot Viviane, Praet Jean-Claude, Remacle Anne, Schoonjans Chris, Schrooten Renaat, Vanderstappen Anne,

Commissaire du gouvernement : Roger Yves

Direction

Directeur général : Dirk Ramaekers

Directeur général adjoint : Jean-Pierre Closon

Contact

Centre fédéral d'expertise des soins de santé (KCE).
Résidence Palace (10ème étage)
155 Rue de la Loi
B-1040 Bruxelles
Belgium

Tel: +32 [0]2 287 33 88
Fax: +32 [0]2 287 33 85

Email : info@centredexpertise.fgov.be
Web : http://www.centredexpertise.fgov.be

Inventaire des bases de données de soins de santé

KCE reports vol. 30B

STEFAAN VAN DE SANDE
DIRK DE WACHTER
NATHALIE SWARTENBROEKX
JAN PEERS
HANS DEBRUYNE
INGRID MOLDENAERS
BAUDOUIN LEJEUNE
VIRGINIE VAN DAMME
DIRK RAMAEKERS
MARK LEYS

Federaal Kenniscentrum voor de gezondheidszorg
Centre fédéral d'expertise des soins de santé
2006

KCE reports vol. 30B

Titre : Inventaire des bases de données de soins de santé

Auteurs : Stefaan Van de Sande (KCE), Dirk De Wachter (KCE), Nathalie Swartenbroeckx (KCE), Jan Peers (Deloitte), Hans Debruyne (Deloitte), Ingrid Moldenaers (Deloitte), Baudouin Lejeune (Deloitte), Virginie Van Damme (Deloitte), Dirk Ramaekers (KCE), Mark Leys (KCE)

Experts externes : Ingrid Mertens (FOD Volksgezondheid), Dirk Moens (FOD Sociale Zekerheid), Koen Vleminckx (FOD Sociale Zaken), Viviane Van Casteren (WIV), Johan Van der Heyden (WIV) Herman van Oyen (WIV), Patrice Planche (Région Wallonne), Herwig De Kindt (Vlaamse gemeenschap), Gérard Felten (Région Wallonne),Raf Mertens (IMA), Johan Van Overloop (IMA) Hervé Avalosse (IMA), Jouri Guillaume(IMA), Jan De Maeseneer (Huisartsengeneeskunde - UGent), Sara Willems (Huisartsengeneeskunde - UGent), , Liesbeth Van Eycken (Kankerregister), Myriam De Spiegelaere (Observatoire de Bruxelles), Luc Berghmans (Observatoire de Hainaut), Etienne De Clercq (Ecole Santé Public-UCL) Christophe Vanroelen (VUB-MESO), Diana De Graeve (UA), Carine Vandevoorde (KUL), Vincent Lorant (SESA, UCL), Marie Christine Closon (SESA UCL), Ghislaine Adriaensens (Steunpunt bestrijding armoede – Centrum voor gelijke kansen), Jean-Pierre Engels (IMS Health Belgium)

Validateurs : Niko Demeester, Pieter Dewitte, Leïla Bellamammer (Institut National pour la Statistique)
Frank Robben (Banque Carrefour Sécurité Sociale)

Acknowledgement: Des commentaires et suggestions: Muriel Lona (KCE), Carine Vandevoorde (KCE), Stephan Devriese (KCE), Cecile Camberlin (KCE), Dirk Van Den Steen (KCE), Frank Hulstaert (KCE)

Conflict of interest : Aucun conflit déclaré

Disclaimer: Les experts externes et validateurs ont collaboré à la rédaction du rapport scientifique mais ne sont pas responsables des recommandations aux Autorités. Les recommandations aux Autorités ont été rédigées par le Centre d'expertise (KCE).

Mise en Page : Nadia Bonnouh, Dimitri Bogaerts

Bruxelles, mai 2006

Etude : 2004-23
Domaine : Objective Elements-Communication (OEC)
MeSH : Information Storage and Retrieval ; Databases ; Public Health Informatics ; Information Science ; Medical Record Linkage
NLM classification : W 26.55.I4

Langage : français, anglais
Format : Adobe® PDF™ (A4)

Dépôt légal : D/2006/10.273/15

La reproduction partielle de ce document est autorisée à condition que la source soit mentionnée. Ce document est disponible en téléchargement sur le site Web du Centre fédéral d'expertise des soins de santé.

Comment citer ce rapport ?
Van de Sande S, De Wachter D, Swartenbroeckx N, Peers J, Debruyne H, Moldenaers I, Lejeune B, Van Damme V, Ramaekers D, Leys M. Inventaire des bases de données de soins de santé. KCE Reports vol. 30B Bruxelles : Centre fédéral d'expertise des soins de santé (KCE) ; mai 2006.. Ref. D/2006/10.273/15.

Préface

La recherche scientifique en soins de santé destinée à l'aide de la décision se base souvent sur des données enregistrées. Ces données peuvent être recueillies de manière prospective par échantillonnage, ou par des études cliniques. De plus, une masse de données administratives est disponible auprès des assurances maladies, et des données cliniques le sont auprès des hôpitaux. En tenant compte de la protection de la vie privée du patient, ces données, codées et agrégées, peuvent donner lieu à des analyses rétrospectives. Les technologies de l'information et le développement des applications via le réseau à large bande augmentent les possibilités d'exploitation de grandes bases de données pour la recherche en soins de santé.

Sur le plan national comme international, des débats ont lieu concernant l'organisation des bases de données et des données statistiques afin d'apporter une aide à la décision. Ce débat se fait en parallèle avec le développement du dossier électronique du patient et l'échange des données de santé entre les prestataires de soins. La Belgique a le devoir de s'investir dans ce débat. Identifier les besoins d'information des différents acteurs du domaine de la santé est dès lors essentiel.

Pour réaliser sa mission, le KCE dépend fortement de la disponibilité des données en matière de soins de santé. Déjà avant sa création, le paysage belge des données de soins de santé était reconnu pour sa complexité et son cloisonnement, et aucun inventaire des données exploitables pour la recherche n'était disponible. C'est pour ces raisons que le KCE a débuté un projet d'inventaire de données. En attendant cet inventaire, le KCE ne s'est toutefois pas reposé sur ses lauriers. Pensons à de précédents rapports comportant des analyses sur des groupes de patients homogènes au départ de données hospitalières ou, pour les montants de référence, à la collaboration étroite établie avec l'Agence inter-mutualiste et la Cellule technique.

La création de cet inventaire a été un processus laborieux dont le dénouement a été rendu possible grâce au précieux concours des différentes personnes au sein des institutions impliquées et la contribution de plusieurs experts au groupe focus. Sans la contribution de ces personnes, notamment dans l'identification, parfois difficile, de bases de données disponibles au sein de leur propre institution, cet inventaire n'aurait pas été possible.

Un produit est désormais disponible pour nos experts comme pour les diverses équipes de recherche qui collaborent avec le KCE. Cet inventaire sera naturellement un outil de travail pour toutes les organisations et les personnes qui désirent en savoir plus sur les données existantes en soins de santé. Il pourrait également fournir une impulsion au débat nécessaire concernant l'optimalisation des flux des données de soins de santé pour leur exploitation à fins de recherche et d'aide à la décision. Nous espérons que ce rapport, outre son rôle important d'inventorisation, initiera le développement d'une vision stratégique, du point de vue de la société, d'un système d'information en soins de santé.

Jean-Pierre Closon
Directeur général adjoint

Dirk Ramaekers
Directeur général

Résumé

Introduction

Le rapport « inventaire des données » est à situer dans le cadre d'une des missions-clé du Centre Fédéral d'Expertise des Soins de Santé: l'utilisation des données de soins de santé comme instrument pour mener des recherches d'évaluation ayant pour but une aide à la décision politique.

En ce qui concerne la situation belge, le rapport décrit les données générales de soins de santé et les bases de données qui peuvent être utilisées dans le cadre de projets d'équité, un domaine important dans le contexte belge. Une partie spécifique de la recherche a étudié de quelle manière l'Allemagne, l'Ecosse, le Danemark et d'autres organisent leurs banques de données relatives aux soins de santé.

Objectifs

L'objectif de ce projet est de dresser un inventaire des bases de données disponibles qui contiennent des données sur l'organisation des soins de santé. Sur base de cet inventaire, l'accent est mis sur les possibilités et les problèmes qui surviennent actuellement lors de l'utilisation de ces données.

Il s'agit d'un inventaire assez complet, mais l'inventaire ne se veut pas être exhaustif. Cet inventaire doit être considéré comme une première ébauche, qui doit être complétée et actualisée dans le futur.

Méthodologie

Pour la réalisation de l'inventaire, différentes sources ont été utilisées : la littérature, les documents, l'information disponible sur Internet, les contacts individuels avec les gestionnaires de bases de données, les interviews de groupe organisés avec les experts scientifiques et politiques actifs sur le terrain.

Résultats et conclusions

Une première constatation importante et encourageante est que la Belgique dispose de nombreux systèmes d'enregistrements, de bases de données et par conséquent, de données relatives aux soins de santé. De plus, les données sont collectées à un niveau relativement détaillé.

Sur le plan du contenu, certaines lacunes ont été identifiées : absence d'un identificateur unique de la personne qui rend particulièrement difficiles, voire impossibles à réaliser, les analyses longitudinales nécessaires à l'étude du suivi du patient; absence de données disponibles sur les assurances soins de santé non obligatoires; difficultés avec les données de diagnostic et de traitement ; un manque de données sur les soins de santé extramuraux ; des données défectueuses sur la psychiatrie et des données manquantes sur les maisons de repos et les maisons de repos et de soins ; un manque de données sur les technologies utilisées dans les soins de santé et un manque de données sur la consommation de soins non remboursés.

Malgré le grand potentiel de données, la Belgique ne réussit pas bien à rapporter sérieusement et à traduire les classifications et les concepts dans le cadre d'une propre collecte de données qui rende possible de rapporter correctement sur le plan international.

La situation concernant la disponibilité pour la recherche équité est encore loin d'être optimale. De manière générale, on doit travailler avec des données provenant de bases de données administratives, qui se préoccupent moins de la perspective scientifique. Les données sont à peine collectées dans le cadre d'un modèle conceptuel de définitions qui sont pertinentes pour une

recherche d'équité. De plus, les groupes socio-économiques faibles sont sous-représentés dans les enregistrements (administratifs).

En ce qui concerne l'organisation pratique, il y a une dispersion de bases de données et d'institutions. Ce morcellement est à expliquer par la répartition complexe des différents échelons de pouvoir publique. Il est surtout surprenant que les déclarations d'intention et les protocoles pris précédemment afin d'arriver à plus de coordination soient jusqu'à présent restés lettre morte.

De plus, réaliser un inventaire est une mission difficile. Ainsi, retrouver les caractéristiques précises des bases de données pose parfois grand problème. D'une part, il n'est pas toujours facile de savoir qui sont les personnes de contact et où l'information peut être trouvée. D'autre part, il semble que l'information sur les caractéristiques des bases de données ne se trouve pas toujours à portée de main, encore moins qu'elle soit actualisée en permanence.

La qualité des données enregistrées est une condition essentielle pour garantir la pertinence et la crédibilité des analyses. Cependant, il n'est pas question, dans le contexte belge, d'un projet qualitatif (écrit et disponible) axé sur une gestion des données, ni au niveau des bases de données individuelles, ni au niveau du paysage entier des données relatives aux soins de santé.

Pour les utilisateurs secondaires des bases de données, un problème qui se pose régulièrement est l'accessibilité à ces données. Celle-ci est entravée par des procédures lourdes qui prennent beaucoup de temps. L'accessibilité aux bases de données disponibles est influencée par toutes sortes de conditions et des procédures légales, dont la loi sur la protection de la vie privée. Dans la législation actuelle, le « concept d'unité » est manquant, non seulement en relation avec la définition des données (le modèle de données), les modalités pour la collecte, la communication, le stockage et la protection des données, mais aussi et surtout en relation avec l'accessibilité structurée et rapide à ces données pour une exploitation secondaire.

Recommandations

- Le besoin d'une *centralisation de l'information* sur les bases de données, les « métadata » et l'information sur leurs modalités d'exploitation ressort de ce projet. Ce rapport peut constituer une première réponse..
- Dans le contexte de la Belgique, il manque *un modèle conceptuel de données* dans lequel les objectifs sont fixés et les données relatives à la santé et aux soins de santé ont leur place. Sur le plan du contenu d'un modèle conceptuel, la réflexion peut commencer sur le rôle des organisations responsables de la collecte des sets de données, de la coordination des sets de données, des procédures et des méthodes de travail pour exploiter les données. Un modèle conceptuel tient explicitement compte des besoins des utilisateurs primaires et secondaires, en développant la concertation avec le secteur. Le modèle conceptuel doit d'ailleurs prendre en considération le problème spécifique du couplage des bases de données, ce qui peut avoir un impact sur l'architecture des bases de données individuelles. Ce modèle doit aussi intégrer la notion de protection de la vie privée lorsqu'il s'agit de données personnelles. Pour le développement d'un modèle et en ce qui concerne le contenu et la méthode, la Belgique peut se laisser inspirer des pays qui ont déjà réalisé un exercice similaire.
- Dans les différents pays qui travaillent avec un système intégré de données relatives aux soins de santé, *un plan de qualité des données* est considéré comme une partie fondamentale d'une politique axée sur les données. Un plan de qualité tient compte des différents (types d') utilisateurs finaux. Les exemples internationaux nous enseignent qu'un cadre juridique intégré fournit les garanties d'une collecte de données performante et assure le contrôle de la qualité.
- Un *cadre juridique intégré* nécessite un discussion sur les différentes législations existantes en fonction du management des bases de données et une coordination mutuelle de cette législation.
- Le développement d'un modèle de données soins de santé doit être traduit dans un modèle de management et un modèle organisationnel. *Un modèle d'organisation en réseau* avec les fournisseurs actuels, les utilisateurs et les tiers de confiance semble

être une voie indiquée. Un modèle d'organisation en réseau implique que les institutions existantes travaillent ensemble de manière structurelle et organisée. Un rôle de coordination est attribué à une institution centrale.

Scientific summary

TABLE OF CONTENTS

SCIENTIFIC SUMMARY I

INLEIDING 5

HOOFDSTUK 1 : METHODOLOGISCHE ASPECTEN BIJ DE INVENTARISATIE VAN GEZONDHEIDSZORGGEGEVENS 8

1 INTRODUCTIE 9

2 ZOEKTOCHT NAAR BELGISCHE REGISTERS MET GEZONDHEIDSZORGDATA 10

2.1 Zoekstrategie voor de inventarisatie 10

2.2 Opzoeken van wetteksten via de 'Kruispuntbank van de Wetgeving' 11

2.3 Opzoeken van wetenschappelijke literatuur over databankTheorieën en information management systems (IMS) 11

3 SJABLOON VOOR INDIVIDUELE BESCHRIJVING VAN DATABANKEN 12

3.1 Korte omschrijving 12

3.1.1 Typologie 12

3.2 Beheersorganisatie 12

3.2.1 Verantwoordelijken: 12

3.2.2 Financiering en resources: 12

3.3 Achtergrond: doel en gebruik van de databank 12

3.3.1 Grondslag 12

3.3.2 Doel 12

3.3.3 Gebruikers (organisaties) 12

3.3.4 Toepassingen 12

3.3.5 Nieuwe evoluties 13

3.4 Tijdskarakteristieken 13

3.4.1 Periodiciteit van de updates 13

3.4.2 Longitudinaliteit 13

3.4.3 Periodiciteit 13

3.4.4 Data van versieaanpassingen /updates 13

3.5 Inhoud 13

3.5.1 Informatie over de inhoud 13

3.5.2 Evoluties 13

3.5.3 Registratie-items (variabelen) 13

3.6 Dataproces 13

3.6.1 Informatie over het dataproces 14

3.6.2 Gegevensbronnen (organisaties/databanken) 14

3.6.3 Dataproces plan 14

3.6.4 Datakwaliteitsplan 14

3.6.5 Toegankelijkheidsprocedure 14

HOOFDSTUK 2 : INVENTARIS EN DOORLICHTING VAN BELGISCHE GEGEVENSBANKEN MET GEZONDHEIDSZORGGEGEVENS 15

4 INTRODUCTIE 16

5 BELGISCHE INSTELLINGEN EN DATABANKEN MET GEZONDHEIDSZORGGEGEVENS 17

6 ORDENING VAN GEGEVENS OVER GEZONDHEIDSZORG 20

6.1 Gegevens over GEZONDHEIDSzOrgvoorzieningen 20

6.1.1 Gegevens over voorzieningen van intramurale curatieve zorg 20

6.1.2 Gegevens over ziekenhuisfuncties 21

6.1.3 Gegevens over voorzieningen revalidatie en rehabilitatie 21

6.1.4 Gegevens over voorzieningen voor verpleging en verzorging 21

6.1.5 Gegevens over het beschikbare personeel in de gezondheidszorg 21

6.1.6 Gegevens over de beschikbare technologie en uitrusting 22

6.1.7 Gegevens over de infrastructuur 22

6.1.8 Gegevens over zorgprogramma's 22

6.2 Gegevens over het gebruik van gezondheidszorg 22

6.2.1 Medische (publiek terugbetaalde) consumptie 22

6.2.2 Diagnose-behandel kenmerken van gebruikers van gezondheidszorg 23
6.3 Financiële gegevens van zorgvoorzieningen 23
6.4 Uitgaven gezondheidszorg 23
6.4.1 Prijzen en vergoedingen 24
7 KRITISCHE BEOORDELING VAN DE HUIDIGE TOESTAND M.B.T. GEZONDHEIDSZORGGEGEVENS IN BELGIË 25
7.1 Rijkdom van de Belgische gegevens 25
7.2 Versnippering van gegevens 25
7.3 De kwaliteit van de Belgische gegevensbanken 26
7.3.1 Toegankelijkheid en duidelijkheid: gebrek aan (actuele) informatie over de bestaande gegevensbanken 27
7.3.2 Relevantie 28
7.3.3 Accuraatheid 28
7.3.4 Volledigheid 29
7.3.5 Consistentie 30
7.3.6 Coherentie 30
7.3.7 Tijdigheid en punctualiteit 31
7.4 Persoonlijke gezondheids(zorg)gegevens en de identificatieproblematiek 32
7.4.1 De wettelijke bepalingen m.b.t. het inzamelen en verwerken van persoonsgegevens 33
7.4.2 Implementatieproblemen bij koppeling van databanken 34
7.5 Longitudinaliteit 35
7.6 Gegevenshiaten 37
7.6.1 Niet wettelijk verplichte ziekteverzekeringen 37
7.6.2 Verplichte arbeidsongevallenverzekering en het Fonds voor Arbeidsongevallen (FAO) 38
7.6.3 Fonds voor beroepsziekten 38
7.6.4 De Hulp- en Voorzorgkas voor Zeevarenden (HVZ) 39
7.6.5 Diagnose en behandelingsgegevens en gegevens over zorgzwaarte 39
7.6.6 Gegevens over de eerste lijn (huisartsen en thuiszorg) en ambulante zorgverlening buiten de ziekenhuizen (dagcentra) 39
7.6.7 Geen gegevens over zorgtrajecten 40
7.6.8 Gebrekkige gegevens over rust- en verzorgingstehuizen 40
7.6.9 Gebrekkige gegevens over de psychiatrie en geestelijke gezondheidszorg 40
7.6.10 Gegevens over technologie 40
7.6.11 Gegevens over niet-terugbetaalde consumptie 41
7.7 Gezondheidszorgdata en internationale rapportering 41
7.8 Kernboodschappen 42

HOOFDSTUK 3 : INVENTARIS DATABANKEN 'BILLIJKHEID IN DE GEZONDHEIDSZORG' 43
8 INTRODUCTIE 44
9 METHODOLOGIE 45
10 ORDENING VAN DE GEGEVENS 46
10.1 Inhoudelijke aspecten 46
10.1.1 Algemene opmerkingen 46
10.1.2 Gegevens over gezondheidszorg 47
10.2 Meten 48
10.2.1 Hoe meten? 48
10.2.2 Meetniveau 49
10.2.3 Tijdsdimensie 49
10.3 INVENTARIS VAN DATABANKEN i.v.m. BILLIJKHEID EN GEZONDHEIDSZORG) 49
10.3.1 Populatiegegevens en algemene statistieken 49
10.3.2 Gegevens over "gezondheid" 50
10.3.3 Sociaal-economische kenmerken van de populatie 50
10.3.4 Levensomstandigheden en leefstijl 51
10.3.5 In toegang tot gezondheidszorgvoorzieningen (financieel en andere) 51
10.3.6 In het gebruik van gezondheidszorg 52
10.3.7 In de kwaliteit van de geleverde zorg 52
11 DOORLICHTING VAN DE HUIDIGE SITUATIE 53
11.1 Inhoudelijke aspecten 53
11.2 Kwaliteitskenmerken van de databanken 54

12 KERNBOODSCHAPPEN 56

HOOFDSTUK 4 : INTERNATIONAL EXPERIENCES 57
13 INTRODUCTION 58
14 METHODOLOGY 59
15 SCOTLAND 60
15.1 Introduction 60
15.2 Management of OVERALL PROCESS 60
15.2.1 Roles and responsibilities 60
15.2.2 Legal Framework 61
15.2.3 Time Line 61
15.2.4 Decision Model 61
15.2.5 Management aspects 61
15.2.6 Financing 61
15.3 Data collection Process 62
15.3.1 Data Sources, Roles and Responsibilities 62
15.3.2 Data Collection Mechanism 62
15.4 EXTRACTION, TRANSFORMATION AND LOADING OF DATA (ETL) 63
15.5 Quality Controls 63
15.6 Exploitation 63
15.6.1 Primary Exploitation 63
15.6.2 Secondary Exploitation 64
15.6.3 Relations with international organisations such as WHO and OECD 64
15.7 Summary assessment 65
16 GERMANY 66
16.1 Introduction 66
16.2 Management of overall process 66
16.2.1 Roles and responsibilities 66
16.2.2 Legal Framework 67
16.2.3 Time Line 67
16.2.4 Management aspects 68
16.2.5 Financing 68
16.3 Data collection Process 68
16.3.1 Data Sources, Roles and Responsibilities 68
16.3.2 Data Collection Mechanism 69
16.4 EXTRACTION, TRANSFORMATION AND LOADING OF DATA (ETL) 69
16.5 Quality control 70
16.6 EXPLOITATION 70
16.6.1 Primary Exploitation 70
16.6.2 Secondary Exploitation 71
16.6.3 Relations with international organisations such as WHO and OECD 71
16.7 Summary assessment 72
17 DENMARK 73
17.1 INTRODUCTION 73
17.2 Management of OVERALL PROCESS 73
17.2.1 Roles and responsibilities 73
17.2.2 Legal Framework 74
17.2.3 Time Line 75
17.2.4 Management aspects 75
17.2.5 Financing 75
17.3 Data collection Process 76
17.3.1 Data Sources, Roles and Responsibilities 76
17.3.2 Data Collection Mechanism 76
17.4 EXTRACTION, TRANSFORMATION AND LOADING OF DATA (ETL) 77
17.5 Quality control 77
17.6 EXPLOITATION 78
17.6.1 Primary Exploitation 78
17.6.2 Secondary Exploitation 78
17.6.3 Relations with International Organisations such as WHO and OECD 78
17.7 Summary assessment 79
18 CONCLUSIONS 80

HOOFDSTUK 5 : CONCLUSIES EN AANBEVELINGEN 83
19 EEN GEÏNTEGREERD DATAMODEL VOOR GEZONDHEIDSZORGINFORMATIE 84
19.1 Een conceptueel datamodel gezondheidszorg 84
19.2 Participatie van gebruikers, leveranciers en trusted third partIES 84
19.3 Relatie tussen databanken 84
19.4 Financiële middelen 85
19.5 Buitenlandse inspiratie 85
20 EEN NATIONAAL KWALITEITSPLAN VOOR GEZONDHEIDSZORGDATA 86
21 NOOD AAN CENTRALISATIE VAN INFORMATIE OVER DATABANKEN EN METADATA 88
22 EEN GECOORDINEERD PROGRAMMA VOOR DISSEMINATIE VAN BETROUWBARE GEZONDHEIDSZORGGEGEVENS 89
23 HET BELANG VAN EEN UNIEKE GEPSEUDONIMISEERDE PERSOONSIDENTIFICATOR 90
24 EEN GEÏNTEGREERD WETGEVEND KADER 91
25 HET UITTEKENEN VAN EEN NETWERKORGANISATIEMODEL 92
26 REFERENTIES 95
27 BIJLAGEN 98

Inleiding

In de huidige wereld van het beleidsondersteunend onderzoek liggen steeds meer mogelijkheden om op een efficiënte manier gebruik te maken van gegevens uit databanken die om verschillende redenen worden samengesteld. Vooral de ontwikkelingen in de informatietechnologie hebben de mogelijkheden vergroot om gegevens te exploiteren. Bovendien is de evidence-based health care beweging een enorme impuls om de beleidsmakers op alle niveau's te ondersteunen met informatie die is afgeleid uit belangrijke informatiebestanden. Het KCE heeft in deze vraag naar evidence based beleidsondersteuning een opdracht te vervullen.

Dit project kadert in de tweede strategische doelstelling van het KCE: "het verzamelen en ter beschikking stellen van objectieve elementen voor de ondersteuning van analyses en studies in de verschillende domeinen van de gezondheidszorg waarin beleidsbeslissingen dienen genomen te worden". Het onderzoeksproject is een eerste stap om een overzicht te maken en inzicht te verwerven in de huidig beschikbare databanken. Die eerste stap moet aansporen om een model te ontwikkelen ter verbetering van de bruikbaarheid, de toegankelijkheid, de onderlinge samenhang van bestaande data en databanken. De ontwikkeling van dit model vraagt eveneens het opsporen van ontbrekende gegevens of gegevens die van suboptimale kwaliteit zijn.

De resultaten van dit project zullen binnen het KCE in de eerste plaats ondersteuning bieden aan de onderzoeksdomeinen Health Care Services Research (HSR) en socio-economische aspecten van de gezondheidszorg (EPB). Dat neemt natuurlijk niet weg dat de inventaris nuttige informatie kan bevatten die relevant is voor het health technology assessment (HTA) domein. De inventaris dient beleidsmatig ook als opstap voor het ontwikkelen van een doordacht databeleid in de Belgische gezondheidszorg. In tegenstelling tot een aantal buitenlandse pioniers, heeft België nog geen strategische en klare visie ontwikkeld over de ontwikkeling en exploitatie van gegevens over de gezondheidszorg, waardoor het er ook niet in slaagt om op een gepaste en kwaliteitsvolle manier gegevens te rapporteren aan internationale organisaties zoals OECD, WHO, Eurostat, …

Het project heeft tot doel een bruikbaar overzicht maken van de kenmerken van de beschikbare gegevens in België voor beleidsondersteunend gezondheidszorgonderzoek. Het overzicht beschrijft zowel de inhoudelijke mogelijkheden en analyseert elementen uit alle stappen in het gegevensproces: de selectie van gegevens, productie, verwerking, analyse en exploitatie van de gegevens en toegankelijkheid van databanken en data.

Het rapport zelf bestaat uit vier grote delen. In hoofdstuk 2 wordt algemeen beschrijvend een inventaris gemaakt van de databanken die kunnen gebruikt worden voor het beschrijven en analyseren van de gezondheidszorg. Op basis van de inventaris wordt een kritische bespreking gewijd aan de mogelijkheden en problemen die zich vandaag stellen bij het gebruik van de gegevensbanken.

In hoofdstuk 3 wordt een inventaris gemaakt van databanken die toelaten om billijkheidsonderzoek en onderzoek naar sociaal-economische verschillen in de gezondheidszorg uit te voeren.

In hoofdstuk 4 worden voorbeelden beschreven van drie Europese landen die initiatieven hebben genomen om een gezondheidszorginformatiesysteem uit te bouwen. De bedoeling van deze studie is om te onderzoeken welke lessen geleerd kunnen worden uit die voorbeelden om een goed doordacht gezondheidszorginformatiesysteem vorm te geven.

In het laatste hoofdstuk worden conclusies getrokken en beleidsaanbevelingen geformuleerd.

Bij dit rapport hoort een apart, maar zeer uitgebreid supplement, met daarin een meer gedetailleerde beschrijving van een aantal databanken. Dit supplement is vooral bedoeld als ondersteuning voor mensen die concreet aan de slag willen gaan met de betreffende databank.

Zeer belangrijk is te melden dat dit overzicht gemaakt is in de periode april 2005- februari 2006. Dit rapport zien we als de eerste stap in een activiteit die continu onderhouden moet worden, om de informatie over het Belgische gezondheidszorgsysteem actueel te houden. Rekening houdend met de tijdspanne van dit onderzoek en met de tijd en inspanning die het vraagt om informatie te vergaren, pretendeert deze inventaris geen exhaustiviteit. Het overzicht is uitgebreid, maar we kunnen en willen niet garanderen dat alle mogelijke databanken die potentieel kunnen gebruikt worden voor (beleidsondersteunend) onderzoek naar de gezondheidszorg hier zijn opgenomen.

Hoofdstuk 1 : METHODOLOGISCHE ASPECTEN BIJ DE INVENTARISATIE VAN GEZONDHEIDSZORGGEGEVENS

1 INTRODUCTIE

De eerste opdracht van het project was het identificeren, lokaliseren en inventariseren van databanken met gezondheidszorggegevens beschikbaar op overheidsniveau. Het project heeft daarom in december 2004 een eerste verkennende workshop gehouden met vertegenwoordigers van verschillende organen die vertrouwd zijn met de databanken. Hier is een eerste stap gezet tot de inventarisatie

De eerste opdracht van het project was het identificeren, lokaliseren en inventariseren van databanken met gezondheidszorggegevens beschikbaar op overheidsniveau. Het project heeft daarom in december 2004 een eerste verkennende workshop gehouden met vertegenwoordigers van verschillende organen die vertrouwd zijn met de databanken. Hier is een eerste stap gezet tot de inventarisatie.

Bij de inventarisatie werd al snel duidelijk dat, gezien de versnippering op het onderzochte terrein, 100% exhaustiviteit niet mogelijk was. Daarom heeft het KCE, rekening houdende met de praktische randvoorwaarden van dit project, prioriteiten gesteld bij het inventariseren en bespreken van databanken. Prioriteit is gegeven aan (a) databanken met een nationaal belang voor het beschrijven en analyseren van toestanden in de gezondheidszorg; (b) databanken die een beeld kunnen schetsen van het gebruik van de gezondheidszorg; (c) databanken die relevant kunnen zijn om de sociaal-economische situatie van de populatie te beschrijven (d) databanken die vooral via publieke middelen en diensten te beschikking staan.

Een specifiek onderdeel van het project heeft zich toegespitst op de vraag welke databanken beschikbaar zijn voor de analyse van sociaal-economische verschillen en billijkheidsvragen in het gebruik van de gezondheidszorg. Dit is het luik 'Equity'. Specifieke methodologische aspecten van dit luik worden behandeld in hoofdstuk 5.

Hetzelfde geldt voor het luik 'gegevensprocessen in het buitenland', dat in hoofdstuk 4 aan bod komt.

Het aanbod in de gezondheidszorg wordt in dit rapport op een vrij algemene manier geordend. De ordening waarvoor hier is gekozen heeft enkel tot doel om potentiële gebruikers, op een zeer algemene manier een overzicht te bieden over de beschikbare gegevens in de gezondheidszorg. Voor de inventaris is een ordening gebruikt die in sterke mate is geïnspireerd op classificaties van gegevens over gezondheidszorg van de OECD ('System of Health Accounts'), de WHO ('National Health Accounts') en de werkgroep Public Health Statistics (CARE).

In Bijlage 1 is een glossarium opgenomen van de gehanteerde terminologie rond technische aspecten van de databanken. Het glossarium is een belangrijk werkinstrument binnen dit rapport, om de preciese betekenis van gehanteerde concepten te identificeren. De begrippen die in het glossarium zijn opgenomen worden gebruikt in het rapport bij de bespreking van de databanken. Bijlage 2 geeft een tabel met gehanteerde afkortingen in dit rapport.

2 ZOEKTOCHT NAAR BELGISCHE REGISTERS MET GEZONDHEIDSZORGDATA

2.1 ZOEKSTRATEGIE VOOR DE INVENTARISATIE

Een eerste verkenningstocht ging via de typische portalen van de diverse Belgische overheden op het internet (zie bijlage 1.3). Op deze wijze konden de belangrijkste actoren op het vlak van gegevensbanken met gezondheids(zorg)gegevens worden geïdentificeerd en gelokaliseerd. Ze werden opgenomen in de 'roadmap' (zie figuur 1) gehanteerd bij onze zoektocht naar databanken met al dan niet persoonlijke gezondheids(zorg)gegevens.

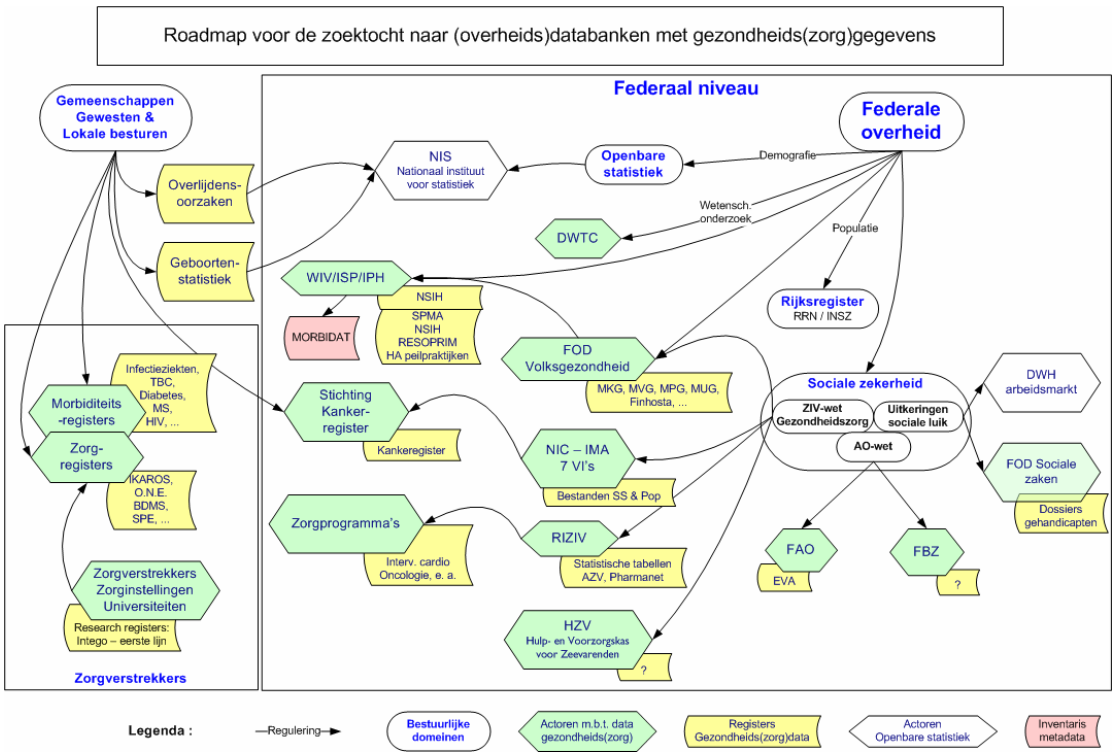
In een tweede, diepergravende internetzoektocht werden de websites van de geïdentificeerde actoren en organisaties bezocht door een ploeg van drie KCE experts. Een grondige exploratie van deze websites liet per organisatie de daar aanwezige gegevensbanken en gegevensstromen te documenteren.

Deze documentatie is aangevuld met bijkomende gegevens verkregen met behulp van een standaardvragenlijst (zie bijlage 1.4) opgestuurd naar geïdentificeerde contactpersonen in deze organisaties.

Per actor of organisatie werd contact opgenomen met de verantwoordelijken van de diverse databanken binnen de organisatie. Dit bleek vooral nuttig bij de grote actoren (FOD Volksgezondheid, RIZIV, IMA, ...) waar een veelheid van gegevensregistraties voorhanden zijn, die niet altijd eenduidig binnen de organisatie gelokaliseerd en gedocumenteerd zijn.

Aan de hand van al deze gegevens werd per specifiek register een detailbeschrijving gemaakt. Hiervoor werd gebruik gemaakt van het sjabloon beschreven in 1.4.

Figuur 1 : Roadmap voor de zoektocht naar (overheids)databanken met gezondheids(zorg)gegevens



2.2 OPZOEKEN VAN WETTEKSTEN VIA DE 'KRUISPUNTBANK VAN DE WETGEVING'

Voor administratieve gegevensinzamelingen door Federale overheidsdiensten bleek vaak dat informatie over metadata, en dan vooral de inhoudelijke beschrijving van de registratie-items en hun wijze en termijnen van inzameling (de zgn. procesmetadata) slechts terug te vinden was in de wetteksten. Deze teksten werden apart opgezocht en bijeengebracht via de 'Kruispuntbank van de wetgeving'

2.3 OPZOEKEN VAN WETENSCHAPPELIJKE LITERATUUR OVER DATABANKTHEORIEËN EN INFORMATION MANAGEMENT SYSTEMS (IMS)

Voor een wetenschappelijke onderbouwing m.b.t. begrippen en methodes, werd, naast het opzoeken via klassieke zoekmotoren zoals PubMed, NLM, Ovid Medline en Embase, websites van een aantal internationale, Europese en nationale statistische referentiecentra bezocht (zie Bijlage 1.5)

3 SJABLOON VOOR INDIVIDUELE BESCHRIJVING VAN DATABANKEN

In de bijlagen worden verschillende databanken met een zekere mate van detail beschreven. Daarvoor is een standaard sjabloon gebruikt zoals hieronder beschreven. Voor een aantal van deze detailrapporten werd hier en daar afgeweken van dit sjabloon, indien toevoeging (of weglating) van specifieke elementen noodzakelijk werd geacht.

3.1 KORTE OMSCHRIJVING

Dit element bevat een korte beschrijving van de gegevensbank.

3.1.1 Typologie

Dit element bevat een aantal sleutelwoorden die de databank typeren. Deze sleutelwoorden worden hernomen in bijlage 6.

3.2 BEHEERSORGANISATIE

Dit element bevat de gegevens van de verantwoordelijke organisatie voor de databank. Ook wordt aangegeven voor welk niveau (via de NUTS-codering) de database gegevens registreert.

3.2.1 Verantwoordelijken:

De verschillende personen die een verantwoordelijkheid over de databank dragen binnen de organisatie, worden hier opgegeven met hun coördinaten en hun rol (situatie 2005).

3.2.2 Financiering en resources:

Dit element verwijst naar de financiële basis van de registratie: wie betaalt, wat is het budget, hoeveel mensen zijn er bij betrokken.

3.3 ACHTERGROND: DOEL EN GEBRUIK VAN DE DATABANK

Dit deel bespreekt de doelstelling(en) en het actuele gebruik van de databank.

3.3.1 Grondslag

De redenen voor de registratie: wettelijke bepalingen, historische achtergrond, wetenschappelijk onderzoek.

3.3.2 Doel

De primaire finaliteit(en) van de databank.

3.3.3 Gebruikers (organisaties)

Gebruikers van de databank zijn die organisaties, anders dan de verantwoordelijke organisatie die gebruik maken van haar gegevens.

3.3.4 Toepassingen

Een lijst van toepassingen die gebruik maken van de gegevens. Eventueel kunnen hier ook positieve en negatieve ervaringen van gebruikers besproken worden.

3.3.5 Nieuwe evoluties

Zijn er nu al gegevens beschikbaar over toekomstige aanpassingen of wijzigingen aan de databank?

3.4 TIJDSKARAKTERISTIEKEN

De tijdsaspecten binnen de registratie en de beschikbaarheid van de gegevens.

3.4.1 Periodiciteit van de updates

De vastgelegde periode tussen verschillende aanpassingen of nieuwe versies van de databank.

3.4.2 Longitudinaliteit

Het al dan niet voldoen aan het opvolgen van entiteiten (patiënten, zorginstellingen, bepaalde uitgaven, …) in de tijd?

3.4.3 Periodiciteit

Over welke periode zijn gegevens beschikbaar in de databank. Zijn er hiaten in de tijd?

3.4.4 Data van versieaanpassingen /updates

Wanneer werden er wijzigingen aan de structuur of ontwerp van de databank doorgevoerd. Dit kan zowel slaan op de introductie van nieuwe variabelen, het (structureel) veranderen van een codeerlijst of een algehele reorganisatie van de gegevens.

3.5 INHOUD

de bespreking van de inhoud van de databank.

3.5.1 Informatie over de inhoud

Welke documenten zijn publiek beschikbaar die de inhoud van de databank beschrijven?

3.5.2 Evoluties

Hier worden de wijzigingen van de inhoud (bvb. nieuwe variabelen) inhoudelijk besproken.

3.5.3 Registratie-items (variabelen)

Dit deel bevat een overzicht van alle variabelen, eventueel gegroepeerd per thema of domein. Van elke variabele wordt indien mogelijk en nodig aangegeven wat haar definitie is, hoe ze gecodeerd wordt, wat de beperkingen zijn (exhaustiviteit), enz…

Wat momenteel niet is opgenomen, is een bespreking van de kwaliteit en de ervaringen van elke individuele variabele. Dit zou een zeer ruime bevraging van de individuele gebruikers vergen.

3.6 DATAPROCES

Een bespreking van de processen om de gegevens te verzamelen, te transformeren en te laden in de database (ETL) met inbegrip van kwaliteitsbevorderende elementen.

3.6.1 Informatie over het dataproces

Welke documenten zijn publiek beschikbaar die de dataverwerking beschrijven om de databank op te stellen en er analyses mee uit te voeren?

3.6.2 Gegevensbronnen (organisaties/databanken)

Wat is de oorsprong van de gegevens? Waar worden ze verzameld?

3.6.3 Dataproces plan

Het plan van de eigenlijke dataverwerking, zowel het ETL deel als het analyse deel.

3.6.4 Datakwaliteitsplan

De verschillende acties die voorzien of geïmplementeerd zijn om de kwaliteit van de gegevens te waarborgen, met specifieke aandacht voor de validatie van de gegevens.

3.6.5 Toegankelijkheidsprocedure

Hoe worden de gegevens toegankelijk gemaakt voor onderzoekers, extern aan de verantwoordelijke organisatie? Wat zijn de wettelijke aspecten? Welke kosten zijn er aan verbonden?

Hoofdstuk 2 : INVENTARIS EN DOORLICHTING VAN BELGISCHE GEGEVENSBANKEN MET GEZONDHEIDSZORGGEGEVENS

4 INTRODUCTIE

In dit hoofdstuk wordt op een algemene manier een inventaris gemaakt van de databanken in België die gebruikt kunnen worden om een inzicht te krijgen in, of een analyse te maken van onderwerpen die van belang zijn voor de organisatie van de gezondheidszorg. We zijn vertrokken vanuit de algemene vraag "in welke mate beschikt België op dit ogenblik over een werkzaam en werkbaar gezondheidszorg informatie systeem (healthcare information system)?".

Zoals in het voorgaande hoofdstuk al aangekaart, is het overzicht vrij uitgebreid, hoewel we niet garanderen dat alle mogelijke databanken, die potentieel kunnen gebruikt worden voor (beleidsondersteunend) onderzoek naar de gezondheidszorg, hier zijn opgenomen. We zijn er echter wel van overtuigd dat de meest belangrijke en vandaag meest courant gebruikte databanken in dit rapport zijn opgenomen.

Het overzicht is opgevat als een stand van zaken 2005. Het zou nuttig zijn de inventaris in de toekomst te actualiseren en aan te vullen waar nodig. Zoals verder in de aanbevelingen nog zal worden bediscussieerd, is er echter behoefte aan een herkenbaar aanspreekpunt dat de inventarisatie en coördinatie van deze informatie op zich neemt.

We bespreken eerst de mogelijkheden die er vandaag bestaan om gegevens over de gezondheidszorg te gebruiken. In bijlage 2.1 is een overzichtstabel te vinden met een beknopte typering van de beschikbare databanken.

In het tweede deel van dit hoofdstuk trachten we tot een soort appreciatie te komen van de huidige situatie. We proberen een aantal troeven en hiaten in het huidige systeem te identificeren. Na een aantal algemene vaststellingen proberen we de kwaliteit van het huidige domein van gezondheidszorgdatabanken, op basis van een aantal eigenschappen te bespreken. Voor de bespreking van kwalitatieve aspecten van de Belgische databanken en gegevens hanteren we dimensies van kwaliteit die onder meer door Eurostat[1], OECD of de Vlaamse APS[2] en van de EC 'Leadership Expert Group on Quality' en 'Statistical Programme Committee'[3] worden gebruikt.

5 BELGISCHE INSTELLINGEN EN DATABANKEN MET GEZONDHEIDSZORGGEGEVENS

In het huidige Belgische landschap beheren verschillende instellingen uiteenlopende databanken die gebruikt kunnen worden voor beleidsondersteunend gezondheidszorgonderzoek. De meeste databanken zijn oorspronkelijk gecreëerd met administratieve doelstellingen voor ogen, en zijn daarom uitgebouwd binnen die administraties die daartoe specifieke bevoegdheden hebben.

Daarnaast zijn in de loop van de tijd andere gegevensregistraties in het leven geroepen, in de schoot van andere organismen. Opvallend is dat de bestaande databanken ook daar aansluiten bij de basisopdrachten van de respectievelijke instellingen.

Onderstaande Tabel geeft weer hoe de publieke dataregistraties verdeeld zijn over verschillende instellingen en in welke mate ze (publiek) toegankelijk zijn.

Table 1 : overzicht databanken

Bevoegdheidsniveau	Administratief niveau	Organisatie(s)	Databank	Publieke toegang
Federaal niveau	Algemene gegevens	Nationaal Insituut voor de statistiek	Bevolkingsstatistieken	Web site
	Rijksregister	FOD binnenlandse zaken	Rijksregisternummer INSZ	-
	Sociale zekerheid	Diverse instellingen sociale zekerheid	Datawarehouse arbeidsmarkt & sociale bescherming	-
		RIZIV	Diverse statistische en boekhoudkundige tabellen Zorgprogramma's (cardiologie, oncologie,… Farmanet MFG (SHA)	- - - -
		Fonds arbeidsongevallen	Gegevensbank AO	-
		Fonds beroepsziekten	Gegevensbank BZ	-
		Hulp en Voorzorgkas voor zeevarenden	Gegevensbank HVZ	-
		7 VI's, NIC en IMA-	Bestand GZ-SS Bestand Populatie	- -
	Volksgezondheid	FOD Volksgezondheid	MKG MVG MPG MUG Finhosta Ziekenhuisstatistieken CIC	- - - - - aggregaten Web site
	Volksgezondheid	Technische cel	MKG-MFG	aggregaten
	Epidemiologie	Wetenschappelijk instituut volksgezondheid	Morbidat Hospitaalinfecties Gezondheidsenquêts Huisartsenpeilpraktijken Resoprim	- - - - -
	Specifieke registratie	Stichting Kankerregister	Kankerregister	aggregaten
Gemeenschappen en Gewesten		Burgerlijke stand	Overlijdensregister Geboortenregister	- -
		Morbiditeitsregisters	Besmettelijke ziekten: TBC register	- -
		Zorgregisters	IKAROS BDMS SPE ONE	- - - -

In een meer gedetailleerde synthesetabel worden de databanken kort getypeerd (bijlage 2.1). In het supplement rapport worden de individuele databanken uitgebreid besproken.

6 ORDENING VAN GEGEVENS OVER GEZONDHEIDSZORG

De meeste dataregistraties die in België op federaal niveau worden georganiseerd, zouden de beleidsmakers moeten ondersteunen in de keuzes die gemaakt worden bij de organisatie van het aanbod. Ze hebben vooral betrekking op gegevens over de intramurale voorzieningen en op registraties van de erkenning van beroepsgroepen en voorzieningen.

Dataregistratie over het gebruik van zorgvoorzieningen is overwegend gekoppeld aan de door de overheid verzekerde zorg.

6.1 GEGEVENS OVER GEZONDHEIDSZORGVOORZIENINGEN

6.1.1 Gegevens over voorzieningen van intramurale curatieve zorg

De meeste gegevens over het aanbod worden in de praktijk verzameld over ziekenhuizen.

Specifieke gegevens over het personeel, over de zware uitrusting van ziekenhuizen en boekhoudkundige gegevens worden geregistreerd in Finhosta. Deze databank bevat ook enkele gegevens over het aantal opnames, overlijdens, verblijfsdagen, etc.

In de ziekenhuisstatistieken zijn veel gegevens terug te vinden over de organisatie, het personeel en de uitrusting van de ziekenhuizen. Zij bevatten gegevens over de verschillende medische diensten die worden aangeboden en over zorgprogramma's (zie verder).

De gegevens over de ziekenhuiserkenningen zijn terug te vinden in een samengestelde databank beheerd door de verschillende gemeenschappen. Ze bevat gegevens over het aantal ziekenhuizen, het aantal erkende bedden, zware uitrusting en over de aanwezige zorgprogramma's en dies meer.

Hoewel ze in strikte betekenis de consumptie van de zorg in de ziekenhuizen registreren kan, via de registratie van de minimale klinische gegevens van de patiënten tijdens hun ziekenhuisverblijven en een aantal administratieve gegevens per verblijf (MKG-RCM), een beeld geschetst worden van het aanbod per ziekenhuis. Voor de psychiatrische behandeling en zorg is een specifieke registratie van verblijven in het leven geroepen (MPG-RPM) waarin gegevens worden geregistreerd over socio-demografische kenmerken van de patiënt, klinische gegevens en gegevens over de behandeling. Gegevens over verstrekte verpleegkundige zorg in de ziekenhuizen worden geregistreerd via de MVG-RIM.

Gegevens over de sociale zekerheidsuitgaven van ziekenhuisverblijven voor alle verplicht verzekerden zijn terug te vinden in de minimale financiële gegevens (MFG-RFM, RIZIV: AZV-SHA).

Via de koppeling van de MFG-MKG is men in staat om voor de algemene, niet psychiatrische ziekenhuizen informatie te verwerven over diagnostisch-therapeutische details van de geregistreerde verblijven en de hieraan verbonden sociale zekerheidsuitgaven.

Binnen de hoger geschetste databanken kan in zekere mate achterhaald worden welke voorzieningen en interventies in daghospitalisatie worden uitgevoerd. De belangrijkste databanken zijn hier de MKG-RCM, de ziekenhuisstatistieken, het

kadaster van de zorginstellingen (het toekomstig centraal instellingen bestand) en voor de psychiatrie, de MPG-RPM.

De meest relevante federale databanken, om de zogenaamde specialistische functies te identificeren, zijn de ziekenhuisstatistieken die gegevens bevatten over het personeel en de jaarlijks activiteiten van de diensten palliatieve zorgen. Ook het CIC bevat gegevens over de palliatieve eenheden.

6.1.2 Gegevens over ziekenhuisfuncties

Informatie over de erkende ziekenhuisfuncties zijn terug te vinden in de databank erkenningen (ziekenhuisapotheken gespecialiseerde urgentiediensten, mobiele urgentiediensten, palliatieve zorgen, dagchirurgiediensten, neonatale diensten, regionale perinatale diensten).

Uit een specifieke databank MUG-SMUR zijn registratiegegevens af te leiden over erkende 'medische urgentie groepen'.

De ziekenhuisstatistieken laten toe om de uitrusting, activiteiten, personeel en organisatie van het programma en bepaalde ziekenhuisfuncties in kaart te brengen.

6.1.3 Gegevens over voorzieningen revalidatie en rehabilitatie

Gegevens over revalidatie en rehabilitatie op federaal niveau zijn in essentie alleen beschikbaar via de databanken die ziekenhuisgegevens registeren, zeker wanneer het de medische en paramedische component van de revalidatie betreft. De meest relevante databanken betreffen hier MKG-RCM, de MVG-RIM en de gekoppelde MKG-MFG.

6.1.4 Gegevens over voorzieningen voor verpleging en verzorging

In de beschikbare federale databanken zijn gegevens terug te vinden over het beschut wonen, de psychiatrische centra (MPG) en de rust- en verzorgingstehuizen.

Voor de rust- en verzorgingstehuizen zijn in de centrale databanken echter nauwelijks gegevens beschikbaar over patiëntenprofielen. Alle beschikbare gegevens moeten gehaald worden uit het zogenaamde centraal instellingenbestand (CIC), dat begin 2006 nog niet ten volle operationeel is.

6.1.5 Gegevens over het beschikbare personeel in de gezondheidszorg

In het algemeen zouden het toekomstig centraal instellingen bestand en centraal beroepenbestand (de zgn. 'kadasterbestanden') de referentiebron moeten worden om gegevens te verzamelen over het beschikbare personeel en de beroepsgroepen. In dit bestand worden gegevens opgenomen over alle relevante beroepen in de gezondheidszorg (artsen en hun specialisme, verschillende categorieën verpleegkundigen, apothekers, kinesisten, etc…)

Specifiek voor de ziekenhuizen is het mogelijk om via de databank Finhosta een zeer gedetailleerd beeld te krijgen van de personeelseffectieven. De databank laat toe om een overzicht te maken van voltijds equivalenten, salarissen en vergoedingen, tewerkstellingsmaatregelen, etc.) voor alle types personeel. De databank ziekenhuisstatistieken laat toe om een beeld te krijgen van de aantallen medisch en paramedisch personeel, de voltijds-equivalenten per instelling, de aanstelling naar diploma (A1-A2) het type functie per type dienst.

De MVG-RIM laten ook toe om gegevens te verzamelen over het verpleegkundig personeel per categorie A1, A2, A3, A4, A5 en het aantal personen en gepresteerde uren.

6.1.6 Gegevens over de beschikbare technologie en uitrusting

Voor het opzoeken van gegevens over uitrusting en technologie (binnen ziekenhuizen) kan gebruik gemaakt worden van verschillende bronnen. De meest gedetailleerde informatie is terug te vinden in de databank ziekenhuisstatistieken (beeldvormingtoestellen en hun karakteristieken, aantal gamma camera's en kenmerken, specifiek uitrusting voor specifieke medisch-technische diensten), maar deze berust op een niet gevalideerde bevraging van de instellingen. De meest betrouwbare informatie is terug te vinden in de databanken erkenningen (aantal PET scans, CT-scans, NMR, aantal operatiekwartieren, verloskwartieren, etc). Ook in Finhosta zijn gegevens opgenomen over NMR, PET-scan, deeltjesversneller.

6.1.7 Gegevens over de infrastructuur

Ook voor de gegevens over de infrastructuur van de ziekenhuizen is informatie te vinden in verschillende databanken.

De databank erkenningen heeft veel gegevens over de beschikbare hospitaalinfrastructuur, het type ziekenhuizen, het aantal bedden, kamertypes, operatiekwartieren etc. Ze bevat ook gegevens over erkende rusthuizen, RVT's, PVT's, het aantal bedden, en het aantal beschutte woningen.

Ook de ziekenhuisstatistieken bevatten informatie over details van ziekenhuisinfrastructuur. Finhosta kan ook een overzicht bieden van het aantal erkende bedden.

Voor de psychiatrie zouden de MPG-RPM databanken gegevens moeten bevatten over de types en aantallen erkende bedden/plaatsen van zowel ziekenhuizen als psychiatrische afdelingen in algemene ziekenhuizen, en gegevens bevatten over beschut wonen en PVT's.

6.1.8 Gegevens over zorgprogramma's

Zorgprogramma's zijn in het leven geroepen om op langere termijn de benadering in termen van ziekenhuisdiensten te vervangen. De zorgprogramma's beogen een reorganisatie van de middelen rond een zorgpakket voor de patiënt. Gegevens over erkende zorgprogramma's zijn terug te vinden via de databank erkenningen: bijv. reproductieve geneeskunde, cardiologische zorgprogramma's, oncologieprogramma's, pediatrie,…)

6.2 GEGEVENS OVER HET GEBRUIK VAN GEZONDHEIDSZORG

6.2.1 Medische (publiek terugbetaalde) consumptie

De gegevens over het gebruik van de gezondheidszorg worden in de federale databanken toegespitst op enerzijds 'verblijven', anderzijds zijn er databanken die de individuele publiek verzekerde zorgconsumptie registreren.

Databanken die 'verblijven' registreren zijn de MKG-RCM, gebaseerd op internationale ICD-9-CM codering en de Belgische RIZIV-nomenclatuur. Deze databank laat ook toe additionele informatie te verwerven over de urgentie van interventies, type anesthésie, ligduur etc. Via de MFG kunnen eventueel ook een aantal consumptiegegevens van ziekenhuizen gereconstrueerd worden.

De medisch verpleegkundige gegevens (MVG-RIM) geven een beeld van het gebruik van verpleegkundige zorgen tijdens een ziekenhuisverblijf, net zoals de MPG-RPM gegevens registreren over de consumptie van zorg en behandeling tijdens een psychiatrisch verblijf.

Voor de verplicht verzekerden kan de zorgconsumptie van medicatie achterhaald worden met de databanken Farmanet. Samen met de databank gezondheidszorgen van het IMA worden de medische consumptie in het kader van de verplichte ziekteverzekering geregistreerd én het aandeel persoonlijke uitgaven voor de patiënt. Het populatiebestand laat toe het ziekteverzekeringstatuut en een aantal sociaal economische kenmerken van de verzekerden te achterhalen.

6.2.2 Diagnose-behandel kenmerken van gebruikers van gezondheidszorg

Gegevens over diagnose en behandelkenmerken van de gebruiker van de gezondheidszorg zijn enkel terug te vinden in de databanken die ziekenhuisverblijven registreren.

De MKG-RCM gegevens registreren voor de verblijven de hoofd- en de nevendiagnose volgden de ICD-9-CM classificatie.

In aparte databanken worden wel specifieke klinische gegevens geregistreerd voor de neonaten (SPE, IKAROS, BDMS).

De minimale verpleegkundige gegevens MVG-RIM registreren (facultatief) de graad van zorgafhankelijkheid per verblijf.

In de minimale psychiatrische gegevens MPG-RPM worden per verblijf de voorlopige diagnose bij opname en de definitieve diagnose bij het verblijf gecodeerd volgens DSM IV en ICD-9-CM. Er zijn ook gegevens opgenomen over psychische symptomen, relationele en sociale problemen, fysieke problemen en ontwikkelingsstoornissen. De niet opgeloste problemen worden ook geregistreerd bij ontslag.

6.3 FINANCIËLE GEGEVENS VAN ZORGVOORZIENINGEN

Financiële gegevens over zorgvoorzieningen zijn in de federale databanken alleen geregistreerd voor de ziekenhuizen. De centrale databank is Finhosta die algemeen en analytisch boekhoudkundige gegevens registreert.

6.4 UITGAVEN GEZONDHEIDSZORG

Onder uitgaven begrijpen we in deze paragraaf de uitgaven uit de collectieve publieke middelen voor de gezondheidszorg. Het gaat hier om gegevens over uitgaven voor zorgverstrekkers (medici, paramedici,…), uitgaven voor medische goederen en uitgaven voor afgebakende groepen van interventies. Specifieke gegevens over de uitgaven die gebeuren in het kader van de conventies zijn ook te achterhalen.

De belangrijkste informatiebronnen om uitgaven in ambulante en intramurale gezondheidszorg te achterhalen zijn de RIZIV databanken (doc C, doc N, doc N^{bis}, doc P, doc H, doc FH, doc PH, doc FM).

Voor de uitgaven voor klassieke hospitalisaties (geen daghospitalisatie) in ziekenhuizen zijn de MFG-gegevens een potentieel relevante databron.

Voor de uitgaven van farmaceutische producten in de ambulante zorg is Farmanet een belangrijke databank.

De IMA databank en gegevens van de individuele V.I.'s vormen een belangrijke aanvullende informatiebron om de uitgaven in bepaalde uitgavencategorieën te beoordelen in functie van het aantal personen dat van bepaalde voorzieningen of diensten gebruik maakt of om de voorschrijvend en uitvoerend verstrekker te kennen.

6.4.1 Prijzen en vergoedingen

In België is de belangrijkste bron van informatie over prijzen en vergoedingen de RIZIV-nomenclatuur.

Voor wat betreft de prijzen voor medische goederen (farmacie, implantaten, hulpmaterialen) moet ook gewerkt worden via de nomenclatuur, maar daarmee is geen precieze inschatting te maken van hun reële kostprijzen. De inschatting van de reële prijzen kan alleen op basis van de assumptie dat het RIZIV een vastgelegd percentage van de goederen terugbetaalt en dat men van daaruit een herberekening kan maken. Aanvullende informatie over de reële prijzen kan achterhaald worden via het Ministerie van Economische Zaken.

7 KRITISCHE BEOORDELING VAN DE HUIDIGE TOESTAND M.B.T. GEZONDHEIDSZORGGEGEVENS IN BELGIË

7.1 RIJKDOM VAN DE BELGISCHE GEGEVENS

Eerst en vooral mag gesteld worden dat België intrinsiek beschikt over een grote hoeveelheid gegevens m.b.t. de gezondheidszorg. Bovendien worden gegevens op een redelijk detailniveau verzameld. Rijkdom en detail zijn vooral te danken aan een aantal structurele elementen:

- het bestaan van een verplichte ziekte- en invaliditeitsverzekering die het overgrote deel van de curatieve gezondheidszorg dekt;
- een uitgebreid tarificatiesysteem voor zorgverstrekkingen met een RIZIV-nomenclatuur die vele medische, zowel als paramedische specialisaties onderscheidt, de meeste hiervan met een hoge graad van detail in aanrekenbare prestatiecodes;
- een terugbetalende overheid die werkt via 7 grote, nationale verzekeringsinstellingen (VI's);
- de in de Wet betreffende de verplichte verzekering voor geneeskundige verzorging en uitkeringen (ZIV-wet) opgelegde inzameling, bij deze VI, van een veelheid van statistische gegevens betreffende uitgaven en uitkeringen, die steeds meer op microniveau aan de overheid overgemaakt worden;
- het bestaan van een verplichte registratie van diagnose- en (para)medische behandelingsgegevens (MKG, MPG, MVG) m.b.t. tot de ziekenhuisverblijven;
- de bij wet geregelde koppeling van de facturatiegegevens (MFG, via de VI's) met de pathologiegegevens (MKG, via de FOD Volksgezondheid) van de ziekenhuisverblijven.

Met uitzondering van het ontbreken van een verplichte zorgregistratie in de ambulante zorg buiten de zorginstellingen (ICPC2-classificatie) en de hiaten van de niet door de ZIV-wet geregelde zorgverlening (zie verder), beschikken we in België op het niveau van de publieke overheid potentieel over een 'goudmijn' van te exploiteren gegevens. Dat neemt echter niet weg dat ook tal van problemen dienen aangepakt te worden om tot een degelijk functionerend en nuttig gezondheidszorginformatiesysteem te komen.

7.2 VERSNIPPERING VAN GEGEVENS

Internationaal (OECD, Eurostat, WHO) worden uitgebreide aanbevelingen geformuleerd over convergentie van gegevensverzameling en internationale samenwerking. Enerzijds lijkt België moeilijkheden te ondervinden om aansluiting te vinden bij de internationale rapportering [4, 5]. Anderzijds lijkt België er niet in te slagen om op nationaal niveau dergelijke doelstellingen te realiseren. De verregaande segmentatie van gezondheidsgegevens in België werd overigens al in vroegere inventarisatieprojecten belicht[6, 7].

In het kader van de bevoegdheidsverdelingen tussen de complexe Belgische staats- en dus administratieve structuren, werd het 'Protocol' van 29 maart 2000' (BS 19/09/2000) betreffende gegevensuitwisselingen tussen de federale

overheid en de overheden bevoegd voor gezondheidszorgbeleid geschreven. Verschillende overwegingen lagen aan de basis van dit protocolakkoord: een gebrek aan efficiëntie en coördinatie bij het verzamelen en verwerken van al dan niet gelijksoortige gegevens, kwaliteitsproblemen in de ruime zin van het woord en de beschikbaarheid van enerzijds de gegevens zelf of anderzijds de resultaten van de gegevensverwerking. Het protocol stelt als uitgangspunt dat voor het voeren van een goed gezondheidsbeleid er nood is aan informatie. Deze informatie komt voort uit de verwerking van verzamelde gegevens door de verschillende overheden bevoegd voor gezondheidszorgbeleid. Het protocolakkoord voorzag in de opmaak van een inventaris van de beschikbare databanken met een verduidelijking van de belangrijkste kenmerken zoals reglementaire basis, doel, inhoud, registratieperioden en -termijnen, enz. Daarnaast was voorzien dat per registratie één contactadres zou zijn waar de huidige stand van zaken in de vorm van schriftelijke informatie kon verkregen worden. In die zin konden we misschien verwachten dat een aantal resultaten van dit protocolakkoord tastbaar zouden zijn in het veld. We hebben echter geen gebruik kunnen maken van concrete producten van dit protocolakkoord. De afspraken gemaakt in het protocolakkoord zijn, na zes jaar, in de praktijk niet gerealiseerd. De complexe bevoegdheidsverdeling tussen de grondwettelijk bepaalde beleidsniveaus (federaal, gemeenschappen en gewesten) in de gezondheidszorg lijkt op dit ogenblik een gecoördineerd databeleid in de weg te staan. De Belgische complexe bevoegdheidsverdeling geeft aanleiding tot nu eens overlappende, vaak verwante, en vooral onderling niet afgestemde registraties, en geeft aanleiding tot specifieke problemen. Twee voorbeelden :

- De erkenning van de ziekenhuizen is een regionale bevoegdheid. Maar de informatie over welke ziekenhuizen erkend zijn, voor hoeveel bedden en voor welke apparatuur is uiteraard van direct belang op het federale niveau, onder meer voor de financiering van de sector.
- De MKG-gegevensregistratie is federaal georganiseerd maar is ook voor de gemeenschappen van belang, onder meer om hun bevoegdheid met betrekking tot de controle van de zorgkwaliteit te kunnen uitoefenen.

De huidige staatsstructuur geeft echter aan geen van deze niveaus de wettelijke bevoegdheid om over deze gegevens te beschikken, tenzij er een samenwerkingsakkoord wordt afgesloten. Het alternatief zou een dubbele dataverzameling zijn voor identieke gegevens, wat omwille van de inefficiëntie ervan en de weerstand ertegen op het terrein, niet gewenst is.

7.3 DE KWALITEIT VAN DE BELGISCHE GEGEVENSBANKEN

Het huidige onderzoek had niet als primaire focus de kwaliteit van elke gegevensbank in detail te onderzoeken. Dit hoorde niet bij de opdracht en het is daarenboven een zeer arbeidsintensieve opdracht, die heel wat maanden in beslag kan nemen. Het is mogelijk dat – ondanks een gebrek aan informatie over kwaliteitsinitiatieven – de kwaliteit van heel wat gegevens toch voldoende is voor wetenschappelijk en beleidsondersteunend onderzoek. Maar bij gebrek aan duidelijke informatie moet dit telkens opnieuw nagegaan worden door de gebruiker, wat zeer inefficiënt is.

Bij onze zoektocht naar informatie blijkt de informatie over een ‘kwaliteitsplan’ of over initiatieven om de kwaliteit te verbeteren of op peil te houden in de meeste gevallen ontoereikend of zelfs ontbrekend.

Kwaliteit is echter een belangrijke voorwaarde om geloofwaardigheid op te bouwen. Studies en onderzoek, die zich op gegevens baseren waarvan de kwaliteit niet gegarandeerd is, zullen gecontesteerd kunnen worden.

De kwaliteitscriteria die we hier bespreken zijn in sterke mate geïnspireerd op criteria die aangehaald worden bij publieke overheden die zich met gegevensregistratie bezighouden (Canadese CIHI[8], Nieuw Zeeland[9, 10], Eurostat[3], Vlaamse statistieken[2], …)

7.3.1 Toegankelijkheid en duidelijkheid: gebrek aan (actuele) informatie over de bestaande gegevensbanken

Informatie vinden over bestaande databanken is problematisch op een aantal niveaus, ondanks het feit dat de initiatieven waar wij ons hebben op geconcentreerd gefinancierd worden met publieke middelen, en bovendien omdat men de exploitatie van de databanken ook als middel ziet om het publieke belang te dienen.

Ten eerste is het zeer moeilijk om een overzicht te vinden van potentieel relevante databanken die in verschillende overheidsinstellingen worden bijgehouden. Bovendien worden tal van gegevens ook verzameld in het kader van zeer uiteenlopende initiatieven en projecten. Voor dit type initiatieven is het nauwelijks mogelijk om op een andere manier dan 'parate terreinkennis', individuele contacten en 'toevalstreffers' enig beeld te krijgen van registraties en databanken die met diverse overheidsgelden zijn gefinancierd. Zelfs binnen eenzelfde overheidsinstelling is men niet altijd in staat een professioneel gedocumenteerde en gestructureerde inventaris van de beschikbare registraties en de bijbehorende metadata aan te bieden.

Onze zoektocht onthulde bovendien een gebrek aan volledige en geactualiseerde documentatie, nochtans een onmisbare voorwaarde voor een efficiënte en correcte verwerking van de gegevens. Zelden kon precieze informatie verstrekt worden over de wettelijk-technische randvoorwaarden voor de verzameling en exploitatie van data.

Ten tweede ontbreekt het aan duidelijkheid. Er duiken moeilijkheden op om de precieze inhoud en kenmerken van een databank te achterhalen. Bij gegevens, waarvan de metadata niet beschikbaar zijn, ontbreken de definities van de variabelen, de gebruikte classificaties, enz. Het gebrek aan duidelijkheid betekent een directe rem op het gebruik van de gegevens en de geloofwaardigheid van onderzoeken die op die gegevens steunen. Wanneer de noodzakelijke metadata van de gegevens (vrij) beschikbaar zijn, kan elke onderzoeker of analist vooraf nagaan of aan de voor hem noodzakelijke kwaliteitscriteria voldaan is. Anderzijds laten deze metadata achteraf, bij het verwerken van de gegevens, toe de beperkingen en premissen waarop de studie steunt beter te omschrijven en documenteren door gebruik te maken van de beschreven definities en procedures die aan de basis lagen van de gegevensbank.

- De grote enquêtes (de Gezondheidsenquête, de PSBH, in mindere mate de sociaal-economische enquête) zijn goede voorbeelden van gegevensverzamelingen met een uitgebreide bibliotheek aan metadata. Van bij de opzet gaat men er van uit dat de gegevens door een brede waaier van onderzoekers zullen gebruikt worden).
- Veel metadata van administratieve databanken zijn slechts beschikbaar via wetteksten, koninklijke besluiten, omzendbrieven of handleidingen. Deze documenten zijn niet alleen dikwijls in

technische termen opgesteld, maar hebben daarenboven vrijwel als enig doel de datatransfer tussen de dataleveranciers en de verantwoordelijke organisatie voor de administratieve databank te organiseren. Ze bevatten dus wel informatie over de variabelen en de deadlines voor de aanlevering ervan, maar veel minder over de structuur van de databank en haar kwaliteitsaspecten.

Toegankelijkheid van de beschikbare data wordt bovendien beïnvloed door uiteenlopende wettelijke voorwaarden en procedures. De belangrijkste is de privacywetgeving. Bovendien wordt de toegang tot andere (private) gegevens vaak gecontroleerd door de eigenaar of een stuurcomité waardoor zowel de data zelf als de informatie over de databank niet altijd even gemakkelijk te verkrijgen zijn.

Tot slot maken de gebrekkige toegankelijkheid van de databanken en de zware procedures het zeer moeilijk om verkennende of pre-analyses te kunnen uitvoeren. Dit soort vooronderzoek is vaak zeer essentieel voor onderzoekers om een inschatting te kunnen maken, in welke mate een vooropgestelde vraagstelling inderdaad onderzocht kan worden, en welke problemen men kan verwachten.

7.3.2 Relevantie

Relevantie verwijst naar de mate waarin de gegevens de doelen dienen waarvoor ze werden verzameld. Daarbij dient een onderscheid gemaakt te worden tussen primair en secundair gebruik. De meeste databanken hebben een specifieke primaire doelstelling. Mogelijk kunnen binnen dat perspectief vragen gesteld worden of individuele variabelen voldoende relevant zijn. Hierover doen we geen uitspraken ten gronde, omdat dit buiten het bereik valt van dit onderzoek. Voor sommige variabelen komt het relevantieprobleem overigens pas sterk tot uiting bij secundair gebruik.

Om relevantie te kunnen beoordelen is het nodig de gebruikers van de gegevens en hun behoeftes te identificeren. Vooral op dit laatste vlak wordt in België niet altijd een systematische methodologie gevolgd. Identificatie van gebruikers en analyse van hun behoeften zijn geen standaardpraktijk voor administratieve databanken, met uitzondering van de datawarehouse arbeidsmarkt. Bij enquêtes (bvb. Gezondheidsenquête, Panel Survey of Belgian Households – PSBH, SHARE – in oprichting) gebeurt dit tegenwoordig meer.

7.3.3 Accuraatheid

In de algemene statistische betekenis gaat het om de vraag: hoe dicht benaderen de gegevens de juiste of werkelijke waarde? De accuraatheid is nauw verwant aan de betrouwbaarheid en de validiteit van de gegevens in de databank.

Validatie is noodzakelijk om de accuraatheid van de gegevens significant te verhogen. Daartoe bestaan verschillende werkwijzen:

- inbouwen van ‘check digits’ bij diverse identificatienummers,
- dubbele data-input,
- secundaire validatie met behulp van andere registraties of andere gegevens,
- nakijken door een tweede persoon,

- ‘computer aided input’: keuze uit een lijst van waarden, consistentie testen tijdens het ingeven,
- feedback naar de dataleveranciers ter controle van de input
- …

Vele registratiesystemen voorzien expliciet validatiestappen, zij het dat ze zich beperken tot eenvoudige controles (nagaan of er correcte waarden zijn ingevuld, in sommige gevallen aangevuld met interne consistentietesten).

- Het controleprogramma voor de MKG gaat na of bepaalde sets van gegevens wel correct zijn door onderlinge toetsing: bvb. een eierstokverwijdering bij mannen kan niet.

Veel minder systematisch wordt gebruik gemaakt van externe validatiestappen, ondermeer ingegeven door de complexiteit en daardoor moeilijke uitvoerbaarheid.

- Het Kankerregister gebruikt bijvoorbeeld Europese gegevens voor een validatie van haar exhaustiviteit. Ook binnen het RIZIV worden bepaalde geaggregeerde resultaten vergeleken met resultaten die verkregen werden op een andere wijze (bvb. vergelijking van de doc. H met de gegevens van doc. C of N).

7.3.4 Volledigheid

Volledigheid van een databank behandelt zowel de vraag naar de aanwezigheid van noodzakelijke variabelen, naar de geregistreerde waarden in de gegevensvelden (‘missing values’), als naar de exhaustiviteit van de databank. Volledigheid is een eigenschap van een individuele databank, die moet beoordeeld worden tegen de achtergrond van primair en secundair gebruik.

Een ander aspect van volledigheid betreft de “dekking” van het gezondheidszorglandschap door de verschillende databanken. Verder in dit hoofdstuk staan we stil bij een aantal hiaten in dat totale landschap.

Gezien we in deze inventaris in de eerste plaats informatie over de metadata van de gegevensbanken hebben verzameld, kunnen we de vraag over de volledigheid van de gegevens niet voor elke databank beantwoorden. Uit de concrete ervaring van experts van het Kenniscentrum en andere onderzoekscentra blijkt dat vele administratieve gegevensbestanden onvolledig zijn: ze bevatten bijvoorbeeld soms lege velden in indexvelden.

De oorzaak hiervoor kan gevonden worden in het soms lange en ingewikkelde dataproces voor de verzameling van de gegevens (bvb. aanmaak van gegevens in het ziekenhuis, verzending naar de verzekeringsinstelling, verwerking en verzending naar het RIZIV) en onvoldoende uitgebouwde of te trage feedbackmechanismen die volledigheid garanderen. Hoe later in de tijd en hoe verder de verwerking (procesmatig) gebeurt van de oorspronkelijke bron, hoe moeilijker het contact en hoe meer moeite het kost om aanvullende gegevens te krijgen.

Bij primaire data-inzamelingen helpt een validatieproces om ontbrekende waarden op te sporen, zodat men eventueel de bron opnieuw kan contacteren of herbevragen.

- Goede voorbeelden zijn onder meer het Kankerregister en de registratie van geboorten bij het Studiecentrum voor Perinatale Epidemiologie. Beide organisaties werken met contactpersonen

in vele ziekenhuizen om rechtstreeks informatie te ontvangen of waar bijkomende informatie kan worden opgevragen.

Een moeilijker probleem betreft de exhaustiviteit. Hiermee bedoelen we dat de gegevens van de volledige populatie zouden moeten opgenomen worden in de databank. Het is een heikel punt dat voor veel dataverzamelingen geldt.

- Voor het Kankerregister is het van belang dat alle gediagnosticeerde kankergevallen ook effectief worden gerapporteerd en opgenomen. Voor de Vlaamse regio is dit doel in belangrijke mate bereikt (d.w.z. binnen de internationale kwaliteitsnormen), terwijl dit nog steeds niet het geval is voor de nationale gegevens.

De enige goede manier om de kwaliteit op dit punt te onderhouden en te verbeteren is via de medewerking van mensen op het terrein. Een wettelijke verplichting tot registratie is op zich een onvoldoende voorwaarde, vooral wanneer er geen reële (positieve of negatieve) gevolgen aan verbonden zijn. Bovendien is de terugkoppeling naar de sector, van wat er met de gegevens gebeurt, te beperkt.

7.3.5 Consistentie

Consistentie houdt in dat de datastructuur en de onderlinge relaties tussen entiteiten van de databank en hun eigenschappen zo niet constant, dan toch verenigbaar worden gehouden (ook in tijd en tussen databanken).

Door de gebrekkige informatie over het technische ontwerp van de databanken is het zeer moeilijk om over dit aspect uitspraken te doen. Slechts in specifieke gevallen is meer informatie hierover beschikbaar.

- Bijvoorbeeld: Het formaat en de gebruikte codering van het IMA populatiebestand is gewijzigd (versie 2.1 tot juli 2005, daarna versie 2.2). Om de integriteit te bewaren werden de gegevens retroactief aangepast.

Er zijn voorbeelden waar de consistentie-eis niet gerealiseerd wordt.

- De MKG-registratie is in 1999 hervormd volgens een volledig nieuw databankontwerp, met gevolgen die vandaag nog wegen. Het nieuwe ontwerp volgt een relationeel model met aantoonbare voordelen voor het gebruik. De gegevens van vóór 1999 bevinden zich echter in een aparte databank en in een ander formaat dan de gegevens vanaf 1999, met alle gevolgen voor de toegankelijkheid en integriteit. Bovendien kent ook het nieuwe ontwerp nog integriteitproblemen. Dit ontwerp bevat volgens de relationele databasetheorie voor elke tabel een index. Een noodzakelijke eigenschap voor indexvelden om een index te kunnen vormen, is dat ze niet leeg zijn. Toch vindt men nog steeds lege indexvelden in de observaties van de MKG-registratie.

7.3.6 Coherentie

Coherentie verwijst naar een eigenschap waarbij gegevens een ordelijke samenhang kennen. Er zijn daarbij vier belangrijke subdimensies te onderscheiden: coherentie binnen databanken (en -sets), tussen databanken (en -sets), in de tijd en in de ruimte.

Coherentie betekent over het algemeen dat de gegevens gebaseerd zijn op gemeenschappelijke concepten, definities, classificaties en ontwerpregels van de gegevens. Indien dat niet kan, moeten op zijn minst de verschillen gekend en gedocumenteerd zijn.

Vooral door de versnippering van gegevensverzameling en gebrek aan coördinatie in België is er sprake van een enorme incoherentie op de vier aangehaalde dimensies.

In de gezondheidszorg zijn vele registraties 'ad hoc' opgezet met een particulier doel, eigen aan de organisatie of dienst die er voor verantwoordelijk is. Vele databanken worden louter pragmatisch samengesteld, op basis van eigen operationaliseringen van doelstellingen en eigen inzichten. Het gevolg is dat heel veel potentieel relevante informatie, die kan ontstaan door koppeling, voor onderzoekers of beleidsmakers blijft liggen, omdat gegevens niet kunnen geïntegreerd worden in functie van specifieke vragen. De huidige praktijk toont aan dat onderzoekers, die informatie uit verschillende bronnen wensen te combineren, verzeild raken in een Babylonische spraakverwarring, waarbij schijnbaar dezelfde variabelen of concepten een andere betekenis hebben in de verschillende registraties of waarbij dezelfde inhoudelijke concepten uiteenlopend worden gecodeerd of geclassificeerd.

- Een sprekend voorbeeld is de operationalisering van de entiteit 'dienst' in MKG en MFG. In MKG worden de diensten geclassificeerd met een kenletter eventueel aangevuld met een cijfer (A, C, D, I, E, S3,…). In de MFG worden deze gegevens gecodeerd met een tweecijferige code (19, 20,…). Bovendien bestaat er geen één op één relatie tussen de beide codeerlijsten, wat problemen creëert wanneer gegevens gekoppeld en gevalideerd moeten worden.

Maar zelfs binnen één registratie kunnen incoherenties voorkomen, bijvoorbeeld door een wijziging van definities of coderingen in de tijd.

- De codering van het type ziekenhuisverblijf in de MKG vertoont een incoherentie in de tijd. Het type van verblijf had tot 1998 de kenletter 'C' voor een klassieke hospitalisatie. Vanaf 1999 wijzigde dit in de kenletter 'H' voor een hospitalisatie. Vanaf 2003 werd de kenletter 'C' opnieuw ingevoerd, maar ditmaal met de betekenis 'chirurgische dagkliniek'.

Voor onderzoek naar de ontwikkeling van databanken is het noodzakelijk dat de historiek van deze wijzigingen (deel van de metadata) wordt bijgehouden en publiek beschikbaar is. Dit is momenteel niet het geval.

Ook in de verschillende wetgevende initiatieven, die de dataverzameling en exploitatie trachten te kaderen, wordt weinig aandacht besteed aan een geëxpliciteerd en eenduidig kader, niet alleen m.b.t. gegevensdefinitie (het datamodel) en modaliteiten voor gegevensinzameling, -communicatie, -protectie en -opslag, maar ook m.b.t. een gestructureerde en vlotte toegankelijkheid voor secundaire exploitatie door geaccrediteerde wetenschappelijke onderzoeksgroepen. Over het belang van dergelijk kader is in tal van landen geschreven (DBR[11] , Ierland[12] , Zweden[13-16], Canada[17], Manitoba[18-20] en Nieuw Zeeland[9, 10]).

7.3.7 Tijdigheid en punctualiteit

Gegevens moeten, al dan niet in formele afspraken vastgelegd, voldoende snel aangeleverd worden en beschikbaar komen (tijdigheid). Worden afspraken

gemaakt dan kan verwacht worden dat die afspraken stipt worden nageleefd (punctualiteit).

Gegevens die laattijdig beschikbaar komen of niet op de te verwachte datum gepubliceerd worden verliezen namelijk snel aan relevantie, en daardoor aan geloofwaardigheid. Het bereiken van de deadlines is vooral een procedureel gegeven, waarbij er niet alleen aandacht moet zijn voor de doorvoersnelheid van de gegevens, maar ook voor de andere kwaliteitsaspecten (accuraatheid, volledigheid). Tijdige levering van foutieve gegevens heeft even weinig zin als de laattijdige levering van inhoudelijk zeer kwalitatieve gegevens. De doorvoersnelheid is niet alleen van belang voor de eigenlijke gebruikers, maar ook voor de feedback naar de aangevers van de gegevens.

De huidige praktijk, waarbij de feedbacks op basis van de gekoppelde MKG-MFG gegevens twee jaar na de periode waarop ze betrekking hebben (zie addenda) verschijnen, maakt ze deels ongeschikt voor een actuele kijk op de sector, maar ze verliezen ook hun relevantie voor de bijsturing op het terrein.

7.4 PERSOONLIJKE GEZONDHEIDS(ZORG)GEGEVENS EN DE IDENTIFICATIEPROBLEMATIEK

Persoonlijke gezondheids(zorg)gegevens uit het - al dan niet elektronisch - medisch dossier worden frequent extern gecommuniceerd en geregistreerd voor andere, secundaire doelstellingen, zoals :

1. praktijkbeoordeling en kwaliteitsmonitoring van het medisch handelen op lokaal vlak en zelfs ruimer (‘peer review’ bijv.),
2. wettelijk opgelegde registraties in het kader van het gezondheidsbeleid van bevoegde overheidsinstanties,
3. HCSR (health care services research),
4. gezondheidseconomische studies,
5. klinische en/of epidemiologische research, etc …

Typische aspecten van dergelijke secundaire doelstellingen zijn :

1. de gegevens verlaten de context van de strikt confidentiële arts-patiënt relatie en het daaraan verbonden medisch beroepsgeheim om te worden samengebracht in databanken van een veel ruimer domein dan dit van de eigen (para)medische praktijk;
2. voor de meeste van deze secundaire doelstellingen is de werkelijke identiteit van de patiënt op zich niet echt relevant.

Wel is het behoud van de exacte band tussen persoonlijke gezondheids(zorg)gegevens, afkomstig uit verschillende bronnen of tijdsperiodes, essentieel. In die zin wordt internationaal en ook nationaal een debat gevoerd over de unieke persoonsidentificator en de implementatie ervan, in het licht van gegevenscoherentie, longitudinaliteit, koppeling van patiëntgegevens, etc..

Het delicate spanningsveld tussen enerzijds de bescherming van de privacy van de patiënt en anderzijds het belang voor de samenleving van wetenschappelijk en beleidsondersteunend onderzoek kan verzoend worden door het invoeren van een techniek van ‘pseudonimisatie’ van persoonlijke identificatiegegevens of van een nationaal identificatienummer zonder verlies van de uniciteit van deze gegevens. Dergelijke unieke persoonspseudoniemen vormen de gulden middenweg tussen absolute anonimiteit en directe identificatie van fysieke personen. Afhankelijk van de noodwendigheden van de secundaire

doelstelling(en) kunnen twee soorten pseudonimisatiemethoden gehanteerd worden:

1. irreversibele pseudonimisatie, waarbij individuele re-identificatie uitgesloten is, maar het koppelen van gegevens mogelijk blijft,
2. reversibele pseudonimisatie, waarbij deze individuele re-identificatie mogelijk blijft.

Hoewel pseudonimisatie een manier is om directe identificatie tegen te gaan, kan zij het probleem van de indirecte, contextuele herkenning, nooit oplossen. Persoonlijke gezondheids(zorg)gegevens bevatten nl. meerdere specifieke elementen die kenmerkend zijn voor de fysieke of sociale identiteit van de patiënt, zoals leeftijd, geslacht, landcode van de wettige domicilie, nationaliteitsindicator, code van de woonplaats en sociale zekerheidsregime, etc… Daarbij zijn vaak ook andere elementen geregistreerd zoals verblijfsziekenhuis, opname- en ontslagdatum, diagnose- en interventiecodes van de verblijven, met als gevolg dat het risico tot indirecte, "contextuele" herkenning van deze patiënten toeneemt. Om dergelijke contextuele herkenning te vermijden zijn bijkomende maatregelen van gegevensversluiering en 'small cell' exclusies vereist (zie bijlage 7, punt 2).

7.4.1 De wettelijke bepalingen m.b.t. het inzamelen en verwerken van persoonsgegevens

Het inzamelen en verwerken van persoonsgegevens wordt geregeld door verschillende wetten en besluiten waarvan de belangrijkste betrekking hebben op de bescherming van de persoonlijke levenssfeer. Andere leggen de wettelijke basis vast voor de diverse overheidsregistraties, waarbij een onmiskenbare evolutie naar meer en meer inzamelingen op microniveau, met gebruikmaking van een patiëntpseudoniem, op de voorgrond treedt (zie bijlage 2.2).

Het meest opvallende kenmerk van het geheel van de huidige wetgeving is een markante afwezigheid van 'eenheidsconcept' niet alleen m.b.t. gegevensdefinitie (het datamodel), modaliteiten voor gegevensinzameling, -communicatie, -opslag en -protectie, maar ook en vooral m.b.t. een gestructureerde en vlotte toegankelijkheid tot deze gegevens voor secundaire exploitatie door geaccrediteerde wetenschappelijke onderzoeksgroepen. In dit verband is het interessant te wijzen op twee actuele wetsontwerpen, rond persoonlijke gezondheidsgegevens (Be-Health) en dito sociale gegevens (Wetsontwerp van 22-09-2005 tot wijziging van de wet van 4 juli 1962 betreffende de openbare). Beide bieden een uitstekende gelegenheid om de modaliteiten van pseudonimisatie van persoonsidentificatoren het ter beschikking stellen van persoongegevens, evenals de modaliteiten van stockage, uitwisseling en archivering ervan verder te verfijnen. Er ligt een unieke kans tot concretiseren van een 'eenheidsconcept' en meteen ook een kans op een globale coördinatie van alle betreffende wetteksten.

De discussie over bescherming van persoonsgegevens krijgt een extra dimensie bij de koppeling van databanken. Koppeling houdt namelijk een sterke verhoging in van het risico tot indirecte, contextuele herkenning, zeker als aan domeinoverschrijdende koppeling wordt gedacht (individuele gezondheidsgegevens met dito sociale gegevens). Gekoppelde gegevens expliciteren namelijk de informatie kenmerkend voor de fysieke of sociale identiteit van de patiënt, zoals leeftijd, geslacht, landcode van de wettige domicilie, nationaliteitsindicator, code van de woonplaats en sociale zekerheidsregime, etc…), door hieraan nog andere detailinformatie toe te

voegen m.b.t. specifieke gezondheidszorggegevens (gedateerde diagnose- en ingrepencodes, opname- & ontslagdatums, verblijfsziekenhuis, etc..).

Dit is een belangrijke reden waarom elke koppeling van persoonlijke al dan niet gezondheidsgegevens een specifieke autorisatie door het geëigende Sectoraal Comité behoeft. Telkens weer moet hierbij de afweging gemaakt worden tussen enerzijds het recht op privacy van het individu en anderzijds het publiek belang (wetenschappelijk onderzoek en gezondheidsbeleid) wat neerkomt op een ad hoc analyse van 'gevoelige' gegevensvelden en hun eventuele hercodering.

De complexe medische-technische aspecten van machtigingsaanvragen tot verwerking van persoonlijke gezondheidsgegevens hebben er toe geleid dat zulke aanvragen, zowel voor de onderzoekers als het comité dat dergelijke aanvragen behandelt, geen sinecure zijn. Het resultaat hiervan dreigt remmend te werken op het opstarten van nieuwe onderzoeksprojecten. Het ontbreekt voorlopig aan een goed uitgebouwd, gedocumenteerd en door alle partijen welbegrepen globaal kader, dat een goed doordachte typologie schetst van gegevensaanvragen m.b.t. inhoudelijke en procedurele aspecten (gegevensversluiering, 'ad hoc' hercodering en 'small number' restricties) voor onderzoeksdoelstellingen.

Overigens is op het niveau van de onderzoekséquipes niet altijd de verieste competentie aanwezig en ontbreekt degelijke organisatie, met alle gevolgen van dien om op een efficiënte manier gegevens, afgestemd op de onderzoeksbehoeften, te kunnen verkrijgen.

7.4.2 Implementatieproblemen bij koppeling van databanken

De ervaringen met 7 jaar koppeling van MKG- en MFG-gegevens door de daartoe opgerichte technische cel hebben de noodzaak van een gewaarborgd unieke persoonsidentificator treffend in het daglicht gesteld. In het addendumrapport m.b.t. deze gekoppelde gegevensbank wordt hier meer uitgebreid op ingegaan. Meer in het bijzonder wordt gewezen op de feitelijke non-uniciteit van het voor de koppeling gehanteerde patiëntpseudoniem en de consequenties daarvan op de exactheid en volledigheid van de koppeling.

Om informatie te kunnen samenvoegen moet er een overeenkomst ('match') bestaan tussen individuele observaties uit elke bronregistratie. Deze overeenkomst wordt getoetst aan de hand van een inhoudelijke, paarsgewijze vergelijking van een of meerdere velden in de samen te voegen gegevensrecords. Welke velden wordt bepaald door de beoogde entiteit (persoon, ziekenhuis, pathologiegroep, geografische entiteit, etc.) waarvoor men de koppeling wenst uit te voeren.

De belangrijkste en meest betrouwbare 'matching variable' voor de koppeling van persoonlijke gezondheids(zorg)gegevens is een gecertificeerd uniek patiëntpseudoniem. De meeste studies m.b.t. gezondheidsuitgaven en praktijkvariaties, met uitzondering van de macro-economische, vergen een differentiatie (populatiedefiniëring) op grond van de behandelde pathologie als medische entiteit (op basis van APRDRG, diagnose-behandelgroep, type ingreep,…) zowel als uitgavenentiteit (op basis van nomenclatuurcodes). Wil men hierbij alle elementen kunnen vatten (mogelijke causale factoren, co-factoren, zowel als 'confounding factors'), dan is het beschikken over individuele gezondheidsgegevens (microgegevens) noodzakelijk. Gezien deze laatste zo goed als altijd uit verschillende bronnen komen en dus 'gekoppeld' moeten worden, impliceert zulks dat in elke bron een betrouwbaar patiëntpseudoniem moet voorkomen. Dit is een 'conditio sine qua non', maar de domeinspecificiteit van deze pseudoniemen (bijv. het IMA gebruikt een ander patiëntpseudoniem

dan de TCT) heeft dan wel tot gevolg dat beide, in de praktijk verschillende pseudoniemen zullen moeten teruggekoppeld worden naar de gemeenschappelijke persoonsidentificator (i.c. het INSZ) m.a.w. er zullen ad hoc correspondentietabellen nodig zijn.

Wanneer echter verschillende broninzamelingen binnen eenzelfde verantwoordelijke organisatie liggen en alle hetzelfde unieke patiëntpseudoniem hanteren, zijn deze daardoor de facto gekoppeld ("impliciete koppeling"). Een voorbeeld hiervan vinden we bij het IMA waar de 2 basisbestanden, enerzijds het bestand 'facturatiegegevens' en anderzijds het bestand 'populatiegegevens', voldoen aan deze voorwaarde: beide zijn gebaseerd op het INSZ en tweetraps gepseudonimiseerd door de VI (eerste trap) en de 'trusted third party' (TTP) van het IMA (IBM - tweede trap).

In alle andere gevallen moet van de Privacy Commissie een uitdrukkelijke machtiging tot koppeling bekomen worden, die alle aspecten m.b.t. de PrW beslaat, incluis een gedetailleerde beschrijving van de gehanteerde koppelingsprocedure met tussenschakeling van een intermediaire TTP ("expliciete koppeling"). Dit houdt dan tevens de creatie in van een nieuwe patiëntpseudoniem, specifiek voor het door koppeling ontstane nieuwe gegevensbank (zie ook bijlage 7, punten 2 en 4).

Indien de benodigde velden van overeenkomst (koppelingsvelden of 'matching variables') onvolledig aanwezig of niet concordant zijn in de te koppelen databases, moeten ze apart aangevraagd worden door de instelling of organisatie die de koppeling zal uitvoeren. Zo dient de technische cel voor de koppeling van de MKG- en de MFG-verblijven te beschikken over bijkomende correspondentietabellen waarmee informatie, afkomstig uit databases van het RIZIV en die van de FOD Volksgezondheid, kan gelinkt worden (zie het addendumrapport m.b.t. de gekoppelde MKG-MFG databank).

De MKG-MFG koppeling is inderdaad een koppeling op verblijfsniveau, wat betekent dat niet alleen een matching dient te gebeuren op het niveau van het patiëntpseudoniem (patiëntkoppeling), maar ook op het niveau van de verblijven (verblijfskoppeling). Beide brondatabanken hebben inderdaad het ziekenhuisverblijf als index, zij het dat deze index in beide databanken anders samengesteld is. Bij de koppeling op dit laatste niveau zijn daarom secundaire koppelingsvelden vereist, onder meer datum van opname en van ontslag. Voor een exacte koppeling is het noodzakelijk dat deze velden voldoende precies zijn m.a.w. niet vooraf 'versluierd' werden. Zoals besproken werd in het addendumrapport m.b.t. de gekoppelde MKG-MFG databank is dit niet het geval (zie figuur 4 in dit rapport). Beter ware geweest de versluiering op te leggen ná de koppeling, onder vorm van een hercodering van de betreffende datumvelden in de voor studie medegedeelde gegevensextracten. Dit laatste wordt trouwens gepland voor de nieuwe MZG-registratie, die MKG-, MVG, en MUG-registraties bundelt (gepland voor 2007-08).

7.5 LONGITUDINALITEIT

Net zoals er een toenemende vraag is naar het gebruik van gepseudonimiseerde persoonsgegevens, is continuïteit in de tijd zeer relevant voor beleidsondersteunend en wetenschappelijk onderzoek. Het valt op dat er weinig grondige discussies gevoerd worden over de doelstellingen en denkkaders die de ontwerpvoorwaarden voor continue dataregistratie mee helpen inspireren.

Het debat over de longitudinale analyse is trouwens sterk gerelateerd aan het debat over de persoonsidentificator. Voor persoonsgegevens moet bijv. een

unieke identificator gebruikt worden die, wil men een persoon kunnen volgen, onveranderlijk blijft over de elkaar opvolgende registratieperiodes heen.

In meer algemene zin moet, om aan de eis van longitudinaliteit te kunnen voldoen, een tijdsnotie in de databank ingebouwd worden. Ook het databaseontwerp moet onveranderlijk zijn of op zijn minst moeten de wijzigingen gedocumenteerd worden.

Longitudinaliteit is in de conceptualisering van (administratieve) databanken rond gezondheidszorg tot nog toe geen prioritair onderwerp.

- De MKG-registratie laat bijv geen longitudinale analyses toe. De entiteit van deze registratie is het verblijf van een patiënt in het ziekenhuis. Door het ontbreken van een onveranderlijk uniek patiëntpseudoniem (behoudens binnen eenzelfde registratiejaar en in hetzelfde ziekenhuis) wordt, zodra er een opname in of transfer naar een ander ziekenhuis geschiedt, de continuïteit verbroken.
- Continue longitudinale analyses kunnen dan weer wel als gevolg van een 'onvrijwillig neveneffect' van het koppelen van klinische informatie met informatie over uitgaven in de gekoppelde MKG-MFG databank. Via de zogenaamde correspondentietabellen, en het daarin door dubbele hashing-encryptie ontstane gemeenschappelijke patiëntpseudoniem, kan men de identiteit van de patiënt in al zijn verblijven, ongeacht het ziekenhuis, identificeren en longitudinaal exploiteren. (zie bijlage 2.2. en supplement-rapport over de gekoppelde MKG-MFG).
- Het datawarehouse arbeidsmarkt is gebaseerd op een unieke persoonsidentificator (hier het INSZ-nummer) en laat longitudinale analyses toe.
- De bevragingen van het WIV die zich richten op de individuele gezondheidsstatus (Gezondheidsenquête, Huisartsenpeilpraktijken) houden geen patiëntidentificator in, waardoor longitudinale analyses niet mogelijk zijn.

Recentere dataverzamelingen in verband met de verplichte verzekeringen (IMA, Farmanet, Vlaamse Zorgverzekering) hebben veel meer aandacht besteed aan de opvolging van het individu en aan een hogere datumprecisie, waardoor longitudinale analyses mogelijk worden.

- Het initiatief van de permanente steekproef uit de IMA-gegevens moet in de toekomst longitudinale analyses mogelijk maken
- Bevragingen en registers opgezet met wetenschappelijke doelstellingen bevatten bijna altijd een persoonsidentificator die longitudinale opvolging toelaat.

Bijlage 2.3 geeft een (niet exhaustief) overzicht van een aantal databanken en de entiteit waarvoor de eigenschap longitudinaliteit toepasselijk is voor natuurlijke en rechtspersonen.

7.6 GEGEVENSHIATEN

7.6.1 Niet wettelijk verplichte ziekteverzekeringen

Er bestaan geen wettelijke verplichtingen tot registratie van gegevens over de niet wettelijk verplichte verzekeringen voor een tegemoetkoming bij kosten aangegaan voor de gezondheidszorg. Het feit dat er geen wettelijke verplichting op de verzekeraars rust, betekent echter niet dat er geen gegevens bestaan. Dikwijls worden deze gegevens als bedrijfseigendom beschouwd en kunnen pas gebruikt worden na een voorafgaand akkoord met de eigenaar van deze gegevens.

Bij de verzekeraars zijn mogelijk gegevens beschikbaar over vier types verzekeringen, die te klasseren zijn in twee hoofdcategorieën[21] :

- de aanvullende (collectieve) ziekteverzekering voor de leden van ziekenfondsen en voor werknemers voor wie de werkgever een (hospitalisatie)verzekering heeft afgesloten bij een privé-verzekeraar
- de facultatieve ziektekostenverzekering individueel afgesloten bij ofwel een privé-verzekeraar ofwel een ziekenfonds.

Om enig idee te krijgen van de beschikbare gegevens, hebben we de vraag gesteld aan een VI en aan de beroepsvereniging van private verzekeringsondernemingen Assuralia. De beroepsvereniging Assuralia vertegenwoordigt ongeveer 95% van de markt van privé-verzekeringen.

- De meeste leden van deze beroepsvereniging kunnen ad hoc gegevens leveren die een representatief beeld geven van de facultatieve verzekeringen. Deze gegevens omvatten zeker het geslacht (bijvoorbeeld aantal verzekerden per geslacht), de leeftijd en de gemiddelde hospitalisatiekosten per verzekerde. Dit laatste kan verder uitgesplitst worden naar drie verschillende bedragen: de factuur, de terugbetalingen en welk deel de patiënt betaalt indien hij geen facultatieve verzekering zou hebben.
- Voor de aanvullende verzekeringen stellen zich enkele problemen. Enerzijds wordt deze verzekering meestal afgesloten door de werkgever. Hierdoor is er wel informatie over de werkgever beschikbaar, maar nauwelijks over de werknemer. De uitgaven verbonden aan een werkgeverspolis zijn dikwijls gebundeld zonder het detail van de individuele werknemers te kennen. Bij meerdere vestigingsplaatsen wordt de werkgever bovendien enkel geïdentificeerd via het hoofdkantoor. Daarenboven kan er een grote geografische spreiding bestaan in het domicilie van de werknemers. In beide gevallen verhindert dit een geografische analyse van de verdeling qua inschrijving en gebruik van dit type verzekeringen.
- De (geaggregeerde) gegevens komen enkel ter beschikking indien ze commercieel nodig zijn. Er is meestal geen longitudinale collectie voorzien. Het feit dat de ziekteverzekering eerder een klein volume uitmaakt voor de verzekeraars, draagt zeker niet bij tot meer of betere gegevens.

De ziekenfondsen hebben over het algemeen volledigere informatie en zijn meer gehouden aan een omzendbrief van de Controledienst voor de ziekenfondsen

en de landsbonden van ziekenfondsen, die classificatiecodes voorstelt voor diensten vrije en aanvullende verzekering. Op basis van de informatie van één landsbond blijkt dat de gegevens hetzij landelijk verzameld worden, hetzij op het niveau van de locale mutualiteit. Heel veel is hier afhankelijk van het product (de inhoud van de verzekering) en door wie het wordt aangeboden: wordt het product 1) eenvormig door alle mutualiteiten van deze landsbond aangeboden, 2) slechts door enkele mutualiteiten of 3) onder verschillende condities door de verschillende mutualiteiten? Voor producten die eenvormig landelijk worden aangeboden bestaat de meeste kans dat er gegevens over beschikbaar zijn. Voorbeelden zijn:

- de facultatieve verzekering 'kleine risico's' voor zelfstandigen (code 0.00) en de verzekering voor zorgen in het buitenland (code 0.01). Deze verzekeringen sluiten direct aan bij de verplichte verzekering en laten daarom ook een eenvormige en volledige registratie toe.
- Sommige vaccinaties worden terugbetaald door de overheid, maar maken ook deel uit van de aanvullende verzekering (0.09 - diverse). Omwille van deze reden is een nationale registratie voorzien.
- In het geval van aanvullende (codes 2.0x) of facultatieve (codes 2.1x), hospitalisatieverzekeringen en codes 2.2 tot 2.5 is het plaatje minder duidelijk, omdat verschillende producten worden aangeboden op het lokale niveau. Er moet per product nagegaan worden of er gegevens beschikbaar zijn.
- De meeste andere elementen van de aanvullende verzekering (codes 3. tot 8.) worden vooral lokaal georganiseerd en dus zijn gegevens, als ze al bestaan, moeilijk verkrijgbaar.

7.6.2 Verplichte arbeidsongevallenverzekering en het Fonds voor Arbeidsongevallen (FAO)

De arbeidsongevallenwetgeving in België maakt een onderscheid tussen de overheidssector en andere. Beide worden geregeld door afzonderlijke wetten (afgekondigd op 3 juli 1967). Een eerste wet regelt de schadevergoeding voor arbeidsongevallen, voor ongevallen op de weg naar en van het werk en voor beroepsziekten in de overheidssector. De tweede wet verplicht de overige werkgevers een arbeidsongevallenverzekering aan te gaan bij een van de 17 private verzekeringsmaatschappijen, waarvan 6 gemeenschappelijke kassen. Deze 17 verzekeringsondernemingen beheren ± 450.000 polissen en innen daarvoor ongeveer 918 miljoen Euro aan premies. Microdata omtrent financiële tussenkomsten zijn echter nauwelijks te verkrijgen.

Voor gevallen, waar de werkgever in gebreke is gebleven, werd het Fonds voor arbeidsongevallen (FAO) ingesteld. Het Fonds heeft onder meer tot taak een centrale gegevensbank op te richten betreffende de aangegeven arbeidsongevallen en hun regeling. Hiervoor werd recent het EVA-project gestart dat beoogt een volledige reorganisatie uit te bouwen van alle uitwisselingen van gegevens tussen de verzekeringsondernemingen, het Fonds, de Kruispuntbank, het Rijksregister en de andere instellingen van sociale zekerheid. Het gaat om een zuiver administratieve databank, die weinig relevante klinische informatie bevat.

7.6.3 Fonds voor beroepsziekten

Het Fonds voor de Beroepsziekten staat in voor Schadeloosstellingen voor erkende beroepsziekten. Een specifieke taak betreft het uitwerken van regelingen over de aangifte en vaststelling van schade door beroepsziekten en tussenkomsten voor de schadeloosstelling (gerelateerde hospitalisatiekosten, erelonen en farmaceutica, voor zoverre zij niet door de verplichte ziekteverzekering gedekt werden). Kosten die behoren tot het domein van de verplichte ziekteverzekering worden door de 'klassieke' ziekenfondsen terugbetaald. Verder vergoedt het FBZ de uitgaven van sommige kosten die niet gedekt zijn door de ziekteverzekering: vaccins, gammaglobulines, prothesen, orthopedische toestellen, individuele beschermingsmiddelen, huur van sanitair materiaal,…

Het is echter onduidelijk op welk manier de gegevens van dit fonds kunnen gebruikt worden voor gezondheidszorgonderzoek.

7.6.4 De Hulp- en Voorzorgkas voor Zeevarenden (HVZ)

Zeelui (en personen ten laste) tewerkgesteld bij de Belgische koopvaardij én ingeschreven bij de Belgische Pool der Zeelieden zijn verplicht verzekerd bij de openbare Hulp- en Voorzorgkas voor Zeevarenden (HVKZ). Deze Hulpkas is echter niet opgenomen in de cirkel van de 7 verzekeringsinstellingen (de landsbonden) en dus ook niet in het IMA. Microgegevens van deze bijzondere populatie zijn aldus niet geïncorporeerd in de IMA-databanken.

7.6.5 Diagnose en behandelingsgegevens en gegevens over zorgzwaarte

Pathologiegegevens worden vooral gedistilleerd uit de MKG-gegevens en gekoppelde MKG-MFG gegevens. In de MKG is de gebrekkige differentiatie tussen de co-morbiditeit bij opname, en de complicaties die ontstaan tijdens het verblijf een groot probleem. Het gevolg is dat bij analyses nauwelijks aan risk-adjustment kan worden gedaan, wanneer interventies onderling vergeleken moeten worden. Op basis van deze gegevens is het om methodologische redenen nauwelijks mogelijk om in te schatten of comorbiditeit het gevolg is van een verblijf in een ziekenhuis, of het gevolg zijn van initiële comorbiditeiten. In de nieuwe registratie (gepland voor 2007-2008) zou aan dit euvel worden verholpen.

De relatie tussen MKG en pathologie is niet zo vanzelfsprekend als op het eerste gezicht zou kunnen blijken, vooral wanneer een probleemstelling of vraag vertrekt vanuit een diagnose of een behandeling. Wanneer moet gereconstrueerd worden welke test of welke therapie gebruikt wordt voor een bepaalde aandoening, vergt dit aanzienlijk voorbereidend denk- en zoekwerk.

Bovendien kan niet zomaar een relatie gelegd worden tussen pathologie en behandelingen. Specifieke linken tussen prognostische factoren en behandeling zijn niet te leggen en moeten in sterke mate gebeuren op basis van inschattingen: bijv. bepaald medicatiegebruik wordt verondersteld gerelateerd te zijn met een bepaalde aandoening, maar echte zekerheid daarover bestaat niet.

7.6.6 Gegevens over de eerste lijn (huisartsen en thuiszorg) en ambulante zorgverlening buiten de ziekenhuizen (dagcentra)

De huidige registratie van gegevens in België is sterk ziekenhuisgeoriënteerd. Op dit ogenblik liggen nauwelijks gegevens ter beschikking over de eerste lijn en over ambulante vormen van behandeling of verzorging buiten het ziekenhuis. Er zijn kleinere initiatieven (Intego, huisartsenpeilpraktijk, Resoprim,), maar van een gestructureerde en continue registratie in de eerste lijn is geen sprake.

Het ontbreekt in België aan een systematische registratie van zorgverstrekking in de thuiszorg voor paramedische interventies (kinesitherapie, verpleegkunde, ergotherapie, logopedie), met uitzondering van de uitgaven door zorgvoorschrijvers en -verstrekkers (RIZIV) Binnen de grote thuiszorgorganisaties worden wel een aantal activiteiten en kenmerken van het cliënteel geregistreerd, maar bijna uitsluitend met een financieringsdoelstelling voor ogen. De databanken zijn ook niet gecoördineerd. Ze worden bijna uitsluitend gebruikt voor interne managementdoelstellingen of indien specifieke onderzoeksvragen worden gesteld aan de organisaties.

België kent ook geen systematische registratie van de dagcentra. Hoewel deze centra in de strikte betekenis niet specifiek medisch georiënteerd zijn, wordt hun rol in het aanbieden van medische, paramedische en sociale dienstverlening steeds belangrijker. Vooral voor de chronische aandoeningen zullen deze centra een aandeel van gezondheidszorgactiviteiten op zich nemen. In het bijzonder voor specifieke chronische aandoeningen, de revalidatie en de psychiatrie zou dergelijke registratie zeer nuttig zijn. Buiten een aantal sporadische individuele initiatieven over activiteiten en werkwijzen is er over deze dagcentra nauwelijks informatie terug te vinden in systematisch bijgehouden databanken. Een verklaring voor deze hiaten moet allicht gezocht worden in de Belgische bevoegdheidsverdeling.

7.6.7 Geen gegevens over zorgtrajecten

Het is vandaag onmogelijk om zorgtrajecten te reconstrueren. De problematiek houdt rechtstreeks verband met het ontbreken van een patiëntidentificator. Toch wordt het inzicht in de zorgtrajecten van patiënten en de zorgconsumptie in verschillende etappes een belangrijke informatiebron voor de beleidsmakers.

7.6.8 Gebrekkige gegevens over rust- en verzorgingstehuizen

Er worden geen klinische microgegevens geregistreerd in het domein van de ROB, RVT, IBW & CDV, wel aggregaten van aantallen vergoede verpleegdagen op het niveau Katz-schaal.

Over de bestaande registratiesystemen van RVT's is nauwelijks informatie beschikbaar. De beschikbare gegevens zouden moeten toelaten een analyse te maken over het aantal vergoede verblijfsdagen, maar het is zeer onduidelijk wat precies met deze gegevens mogelijk is.

7.6.9 Gebrekkige gegevens over de psychiatrie en geestelijke gezondheidszorg

Er wordt vrij veel aandacht besteed aan de registratie van MPG van de psychiatrische ziekenhuizen en de PAAZ. Het probleem is echter dat niet elke voorziening even goed en evenveel registreert (omdat bepaalde items niet van toepassing zijn), met als gevolg dat niet alle gegevens voor alle leeftijdsgroepen ter beschikking staan. Ondanks dat registratiesysteem is het vandaag zeer moeilijk om een degelijke analyse te maken van de psychiatrie. De huidige MPG-RPM registratie pretendeert een epidemiologische functie te hebben, maar de onderzoekswereld dringt aan op een grondige doorlichting van het registratiesysteem. De oorspronkelijke doelstelling van een soort MKG-RCM voor de psychiatrie lijkt te zijn verdrongen door andere epidemiologische doelstellingen, met alle mogelijke problemen van dien voor de exploitatie.

7.6.10 Gegevens over technologie

Er is op dit ogenblik in België geen databank ter beschikking van gecommercialiseerde medische apparaten of instrumenten. Er bestaat een

Europees initiatief (EUDAMED) waarin de medische hulpmiddelen geregistreerd worden voor ze op de markt komen, maar dit lijkt als databank niet de geschikte informatie te kunnen toeleveren, o.a. omdat een aantal grote landen niet participeren aan deze registratie.

Eenzelfde probleem kan gesignaleerd worden met betrekking tot informatie over moleculaire en genetische testen.

Daardoor wordt, in de Belgische situatie, informatie verzamelen over technologie een zeer arbeidsintensieve bezigheid, die bestaat uit primair onderzoek: surveys, bezoeken van beurzen, contactname met leveranciers, etc. In het kader van het aan belang winnende 'health technology assessment' onderzoek is dit een belangrijk hiaat.

7.6.11 Gegevens over niet-terugbetaalde consumptie

Over het algemeen kunnen we stellen dat onze huidige registratiesystemen en databanken niet toelaten om een beschrijving of analyse te maken van niet door de overheid terugbetaalde gezondheidszorgconsumptie. Uitzondering hierop zijn de databanken van de V.I. en het IMA, waarin het mogelijk is de betaalde ziekenhuissupplementen te achterhalen. Een van de belangrijke hiaten, is de consumptie van niet voorgeschreven geneesmiddelen (zgn. OTC – 'over the counter' medicatie).

7.7 GEZONDHEIDSZORGDATA EN INTERNATIONALE RAPPORTERING

Een van de grootste onduidelijkheden in het Belgische landschap is de wijze waarop aansluiting wordt gezocht bij de internationale organisaties die instaan voor dataregistratie. Ondanks het grote potentieel aan data slaagt België er niet goed in om degelijk te rapporteren en om classificaties en concepten te gebruiken in de eigen dataverzameling, die het mogelijk moeten maken om ook internationaal op een correcte manier te rapporteren. Terwijl internationaal wordt geïnvesteerd in de ontwikkeling van metadatabanken en classificaties voor gezondheidszorggegevens, schijnt er op dit ogenblik geen gestructureerde werkwijze in België te bestaan om deze te implementeren in de dataregistraties op federaal en regionaal niveau. Op dit ogenblik wordt in de schoot van de verantwoordelijke cel van de FOD sociale zekerheid onderzocht hoe de internationale rapportering over de gezondheidszorg strakker georganiseerd kan worden. De Belgische situatie staat in contrast met de oefeningen van verschillende internationale organen (OECD[22], Eurostat[23] en WHO/UN[24]) die onderling een meer op elkaar afgestemde dataregistratie trachten te ontwikkelen. De bevoegdheidsverdelingen in het Belgische landschap, en het gebrek van een geëxpliciteerd mandaat om aansluiting te zoeken bij die internationale ontwikkelingen, is voor internationaal vergelijkend onderzoek (en voor het prestige van ons land) een belemmerende factor. Aansluiting zoeken bij en toeleveren van adequate gezondheidszorggegevens op het internationale niveau zou bovendien een impuls kunnen betekenen voor het ontwikkelen van een gecoördineerd gezondheidszorgdatamodel. België heeft veel data, centraliseert ze en maakt gebruik van informatietechnologie, maar het ontbreekt aan het uittekenen van een visionair en operationeel model dat zowel nationaal als internationaal kwaliteitsvolle informatie zou kunnen toeleveren.

7.8 KERNBOODSCHAPPEN

- Er worden weinig grondige discussies gevoerd over de doelstellingen en de ontwerpvoorwaarden voor databanken gezondheidszorg
- Informatie vinden over bestaande databanken is problematisch. Het ontbreekt aan volledige en geactualiseerde documentatie, nochtans een onmisbare voorwaarde voor een efficiënte en correcte verwerking van de gegevens. Het ontbreekt ook aan duidelijkheid. Er duiken moeilijkheden op om de precieze inhoud en kenmerken van een databank te achterhalen.
- Toegankelijkheid van de beschikbare data voor secundair gebruik wordt beïnvloed zowel door omslachtige wettelijke voorwaarden en procedures, als door een vaak onvoldoende voorbereiding van de onderzoeksequipe.
- Er wordt niet altijd een systematiek of methodologie gevolgd bij administratieve databanken om de relevantie van de gegevens te beoordelen en de behoeften van primaire en secundaire gebruikers te identificeren. Vooral door de versnippering van gegevensverzameling en gebrek aan coördinatie in België is er sprake van incoherentie en inconstistentie binnen databanken (en -sets), tussen databanken (en -sets), in de tijd en in de ruimte.
- Het ontbreekt voorlopig aan een goed uitgebouwd, gedocumenteerd en door alle partijen welbegrepen globaal kader dat een goed doordachte typologie schetst van gegevensaanvragen m.b.t. inhoudelijke en procedurele aspecten (gegevensversluiering, 'ad hoc' hercodering en 'small number' restricties) voor onderzoeksdoelstellingen.
- Het ontbreekt aan gegevens over niet wettelijk verplichte verzekeringen voor een tegemoetkoming bij kosten aangegaan voor de gezondheidszorg.
- Diagnose en behandelingsgegevens zijn niet altijd makkelijk te achterhalen. Pathologiegegevens worden vooral gedistilleerd uit de MKG-gegevens en gekoppelde MKG-MFG gegevens. In de MKG is de gebrekkige differentiatie tussen de co-morbiditeit bij opname, en de complicaties die ontstaan tijdens het verblijf een groot probleem.
- De huidige registratie van gegevens in België is enorm ziekenhuisgeoriënteerd. Op dit ogenblik liggen nauwelijks gegevens ter beschikking voor de eerste lijn en voor ambulante vormen van behandeling of verzorging buiten het ziekenhuis.
- Het is met de huidige gegevensbanken onmogelijk om zorgtrajecten te reconstrueren.
- Er zijn enkel summiere gegevens beschikbaar over RVT's en ROB's.
- De gegevens over de psychiatrie en geestelijke gezondheidszorg kennen kwaliteitsproblemen naar gebruik toe.
- Er is op dit ogenblik in België geen databank ter beschikking van gecommercialiseerde medische apparaten of instrumenten.
- Een van de grootste onduidelijkheden in het Belgische landschap is de wijze waarop aansluiting wordt gezocht bij de internationale organisaties die instaan voor dataregistratie.

Hoofdstuk 3 : INVENTARIS DATABANKEN 'BILLIJKHEID IN DE GEZONDHEIDSZORG'

8 INTRODUCTIE

Een onderdeel van het project "inventaris databanken" heeft zich toegespitst op de vraag welke databanken beschikbaar zijn voor de analyse van billijkheidsvragen in de gezondheidszorg.

In verschillende landen en agenda van internationale organisaties maakt het vraagstuk van billijkheid en sociaal-economische ongelijkheden integraal deel uit van een gezondheids- en gezondheidszorgbeleid. Vele landen (ook België) worstelen met de vraag op welke manier een degelijk gezondheidszorginformatiesysteem kan worden opgebouwd. Verder in dit hoofdstuk gaan we dieper in op een aantal oorzaken voor deze moeilijkheden, die in essentie te maken hebben het feit dat het billijkheidvraagstuk vanuit verschillende perspectieven benaderd kan worden.

We hebben de vraag gesteld aan een groep Belgische experts welke mogelijkheden en moeilijkheden er zich vandaag stellen om billijkheidsvragen te analyseren door gebruik te maken van beschikbare databanken. Het overzicht legt inhoudelijk de nadruk op het gebruik, de toegankelijkheid en de kwaliteit van gezondheidszorg en op de mogelijkheden om verdelingsvraagstukken te bestuderen (dus minder op gezondheidsverschillen). We hebben ons bewust niet toegespitst op de gezondheidsindicatoren, hoewel er moet op gewezen worden dat onderzoek naar billijkheid in de gezondheidszorg de koppeling maakt tussen gezondheidsindicatoren en indicatoren die informatie geven over allocatie van middelen, gebruik van gezondheidszorgvoorzieningen en indicatoren van kwaliteit van zorg.

9 METHODOLOGIE

De inventaris en appreciatie van de huidige situatie is methodologisch aangepakt via:

1. Documenten- en website analyse.
2. Individuele contacten met verantwoordelijken van databanken relevant voor deze materie (geselecteerd op basis van de input van experts).
3. Focusgroepen met Belgische experts in het onderzoek naar sociaal-economische verschillen in de gezondheid en gezondheidszorg. Ter voorbereiding van de focusgroep is een vragenlijst rondgestuurd (zie bijlage…), waarin experts werden gevraagd te identificeren welke relevante databanken bestaan. Er is ook gevraagd welke behoeften en hiaten er momenteel in onze beschikbare data zitten om zinvolle analyses van het billijkheidsvraagstuk te maken.

Op basis van de input van de experts is een eerste algemene lijst gemaakt van mogelijk interessante databanken. De experts hebben in de focusgroepen commentaren geformuleerd over het belang en de accuraatheid van een aantal databanken. In deze discussie is ook gewezen op het belang van de conceptuele insteek die gehanteerd wordt, omdat deze gevolgen heeft voor het type informatie dat gezocht wordt.

Op basis van die lijst, en rekening houdend met de praktische randvoorwaarden van dit project, heeft het KCE prioriteiten gesteld bij het inventariseren en bespreken van databanken. Tijdens de expertworkshop is er op gewezen dat bij prioriteit een inventaris moet gemaakt worden van databanken die beschikbaar en toegankelijk zijn en waarvan de data bruikbaar zijn voor billijkheids- en onderzoek naar sociaal-economische verschillen in de gezondheidszorg.

Prioriteit is gegeven aan

1. databanken met een nationaal belang voor het beschrijven en analyseren van de toegang toe , het gebruik van en de kwaliteit in de gezondheidszorg;
2. databanken die relevant kunnen zijn om de sociaal-economische situatie van de bevolking te beschrijven (gerelateerd aan gezondheidszorg);
3. databanken die vooral met publieke middelen en diensten gecreëerd zijn en beheerd worden.

Buiten de inventaris die hier wordt voorgesteld, liggen ‘potentieel’ nog tal van gegevens relevant voor onze probleemstelling ter beschikking. Voor tal van gegevens werd het ons niet duidelijk of het gaat om exploiteerbare databanken.

Net zoals gemeld in voorgaande hoofdstukken: dit overzicht pretendeert niet exhaustief te zijn, maar eerder een eerste schets te maken van de mogelijkheden en behoeften. Bovendien willen we er net zoals in voorgaand hoofdstuk op wijzen dat het voor vele databanken soms niet eenvoudig was om precieze informatie over databank of metadata te verkrijgen. Zoals in het voorgaande hoofdstuk willen we er op wijzen dat dit meteen ook problemen in de hand werkt voor de gebruikers van deze databanken.

10 ORDENING VAN DE GEGEVENS

In samenspraak met de experts is een beschrijvend ordeningskader ontwikkeld in functie waarvan we beschikbare gegevens en nuttig databanken kunnen schetsen. Het vormt het raamwerk tegen het welk de huidige (on)mogelijkheden besproken zijn voor sociaal-economisch en billijkheidsonderzoek..

10.1 INHOUDELIJKE ASPECTEN

Het onderzoek naar sociaal-economische verschillen en billijkheidsvragen is enkel zinvol te voeren als vooraf duidelijk wordt gemaakt naar welke vraagstelling men op zoek is. Dergelijk conceptuele verheldering is essentieel indien men ook wil onderzoeken welke data en databanken potentieel relevant kunnen zijn.

10.1.1 Algemene opmerkingen

In de wetenschappelijke literatuur wordt een breed spectrum van concepten en dimensies rond de begrippen sociaal-economische verschillen, billijkheid en toegankelijkheid gehanteerd. Die zijn het gevolg van verschillende inhoudelijke (en operationele) benaderingen. Het is hier niet de bedoeling om al die verschillende benaderingen te schetsen. Sleutelconcepten die gemeenschappelijk aan bod komen zijn in elk geval noden (“need”), toegankelijkheid (“access”), gebruik (“utilisation”) en kwaliteit van zorg (“quality of care”). De discussie over data en databanken is in algemene zin enkel zinvol als voldoende wordt stilgestaan bij het onderscheid tussen:

1. onderwerpen (data) die verband houden met gezondheid;
2. onderwerpen (data) die verband houden met (het gebruik, financiering en kwaliteit van) gezondheidszorg;
3. sociaal-economische kenmerken van de populatie:
    a. In levensstijl
    b. In gezondheidstoestand
    c. In toegang tot gezondheidszorgvoorzieningen (financiële draagkracht, opleiding, cultuur, taal, etc.)
    d. In het gebruik van gezondheidszorg
    e. In de kwaliteit van de geleverde zorg
4. billijkheid in het gebruik van gezondheidszorg” als een normatief concept dat gerelateerd is aan visies over sociale rechtvaardigheid. Er kan een onderscheid gemaakt worden naar :
    a. Billijkheid in gezondheidstoestand
    b. Gelijke voorzieningen
    c. Gelijke toegang tot gezondheidszorg voor gelijke behoeften / gelijke gezondheidstoestand
    d. Gelijk gebruik voor gelijke behoeften
    e. Dezelfde kwaliteit voor iedereen
    f. Financiële bijdrage in verhouding tot beschikbare (financiële) middelen

5. Een gezondheidszorginformatiesysteem zou in staat moeten zijn informatie op te leveren die voor verschillende benaderingen kan gebruikt worden. Informatie is nodig over gezondheid. Data zijn nodig die toegankelijkheidsverschillen (financiële en andere) helpen documenteren, die verschillen in gebruik en financiering beschrijven, en die verschillen in kwaliteit van voorzieningen helpen te identificeren. In ideale omstandigheden zou ook de relatie moeten gelegd worden met informatie over elementen die gezondheidsverschillen en onbillijkheid bepalen die niet kunnen verklaard worden door gezondheidszorg en gezondheidsbeleidmaatregelen (omgevingsfactoren, opleiding, levensomstandigheden,…)
6. In een niet officieel gepubliceerd WHO-werkdocument (Issues in health Information: part 3: integrating Equity into health information systems, April 2005) is een analyse gemaakt van de potentiële moeilijkheden en de noodzakelijke vereisten om een degelijk informatiesysteem uit te bouwen waarmee billijkheidsvragen kunnen worden beantwoord. De opmerkingen die gemaakt wordt in dit rapportstemmen in belangrijke mate overeen met de uitkomsten van de expertworkshop.

10.1.2 Gegevens over gezondheidszorg

10.1.2.1 Financiële bescherming

Er is informatie nodig over de maatregelen die genomen zijn in de financiering van de gezondheidszorg (publieke tussenkomsten, private verzekering en persoonlijk aandeel) en informatie over de proportie van de populatie die (publiek) verzekerd is (inclusief gegevens over de dekking van de verzekering).

10.1.2.2 Gezondheidszorgvoorzieningen

Er moeten gegevens beschikbaar zijn over de allocatie van de middelen besteed aan gezondheidszorg, inclusief de geografische spreiding en de verdeling van middelen over type voorziening, inclusief de gegevens over verspreiding van personeel infrastructuur etc.. Gedetailleerde gegevens over de allocatie aan personeel, uitrusting, technologie, infrastructuur kunnen nodig zijn om vragen met betrekking tot de beschikbaarheid van professionele ondersteuning bij gezondheidsproblemen te kunnen inschatten

10.1.2.3 Gebruik van voorzieningen

Onderzoekers en analisten hebben nood aan gegevens die het gebruik van gezondheidszorgvoorzieningen in kaart brengen, die een aantal barrières in het gebruik helpen te identificeren en die deze gegevens kunnen koppelen aan de kenmerken van socio-economische groepen. Daarbij moet een differentiatie kunnen gemaakt worden tussen verschillende types gezondheidszorgvoorzieningen (preventie zoals vaccinaties), eerste lijn, algemene zorg versus specialistische zorg, maar ook in termen van frequentie en hoeveelheid gebruik van de voorzieningen. Om het gebruik van gezondheidszorg vanuit billijkheidsperspectief te kunnen analyseren is precieze informatie vereist over diagnose en ernst van de aandoening. In ideale omstandigheden zijn precieze gegevens vereist over de noden van de populatie. Om dergelijke vragen te behandelen moet men in de praktijk in staat zijn gegevens uit verschillende databanken te koppelen.

10.1.2.4 Kwaliteit van zorg

Monitoring van kwaliteit is een zeer belangrijk ondersteunend instrument wanneer over billijkheid wordt gesproken. Een zeer specifieke klasse van gegevens is de kwaliteit van zorg; zowel objectieve kwaliteit als de percepties van de gebruikers. Het is nog maar recent dat westerse landen expliciet aan kwaliteitsmonitoring doen via aparte registratiesystemen.

Wanneer geen gebruik wordt gemaakt van gerichte kwaliteitsmonitoring kunnen bepaalde kwaliteitsaspecten eventueel afgeleid kunnen worden uit andere meer algemene databanken: zoals de beschikbaarheid van geschoold personeel, de infrastructuur en het type voorziening, gebruik van het type kamer en type zorgverstrekker, verschillen in gebruik van technologie of zelfs interventietechnieken. Voor specifieke kwaliteitsvragen, is in die omstandigheden een koppeling vereist is van gegevens uit verschillende databanken. Dus ook in dit type onderzoek vraagt de organisatie en het onderliggende model voor de koppeling van databanken een zeer belangrijke reflectie.

10.1.2.5 Impact van ziekte of handicap

Hoewel het in strikte zin buiten het "enge" domein van de gezondheidszorg valt, wordt er ook op gewezen dat informatie over de gevolgen van ziekte of handicap op het leven van mensen, ook een onderdeel vormt van het billijkheidsvraagstuk. Het is in sommige omstandigheden relevant om over informatie te beschikken over de maatregelen waarvan iemand die ziek is, of gehandicapt blijft kan gebruik maken. Dit aspect wordt zeer belangrijk in de context van het toenemende aandeel chronische ziekten, of in de context van langdurige revalidatie. Gegevens over het aandeel ziektekosten in het totale uitgavenbudget van de huishoudens geven een belangrijke indicatie over de impact van ziektekosten op draagkracht en levensstijl. Ook de wijze waarop verschillende groepen toegang hebben tot of gebruik maken van voorzieningen om in geval van ziekte of handicap de personen te ondersteunen moeten gedocumenteerd worden. De relatie leggen met de arbeidsmarktsituatie of de sociale zekerheidssituatie van de persoon die een stoornis heeft, levert relevante onderzoeksinformatie op

10.1.2.6 Algemene informatie over levensomstandigheden

Met het oog op billijkheidsonderzoek zijn ook gegevens vereist over uitgavenpatronen, huishoudbudgetten, woonomstandigheden. Het zijn de meest belangrijke informatiebronnen waaruit indicaties kunnen worden afgeleid over de levensomstandigheden.

10.1.2.7 Persoonsgegevens

Om zinvol billijkheidsonderzoek te kunnen doen is informatie nodig over personen en huishoudens (zie ook paragraaf meetniveau). In functie van deze vraagstelling stelt zich het probleem van de privacy op een zeer dringende wijze. In het voorgaande hoofdstuk is op die problematiek al uitvoerig ingegaan.

10.2 METEN

10.2.1 Hoe meten?

Een wezenlijke vraag die behandeld moet worden heeft te maken "hoe" bepaalde variabelen gemeten zullen worden.

Een klassiek probleem heeft betrekking op het meten van gezondheid en het gebruik van die gegevens in een billijkheidsperspectief: wordt zelfgerapporteerde gezondheid geregistreerd, worden diagnoses genomen, of behandelingen die iemand heeft ondergaan of de soorten zorg die iemand geniet, wordt kwaliteit van leven gemeten, wordt de chroniciteit of de mate van handicap geregistreerd, etc.

Een ander typisch probleem is het meten van sociale positie en sociale stratificatie: In de internationale literatuur wordt sociale stratificatie meestal geoperationaliseerd aan de hand van sociaal-economische status, gender, ethnische afkomst en geografische locatie. Sociaal-economische status wordt meestal geoperationaliseerd als een gecombineerde maat van opleiding, inkomen en beroep. Ook met deze variabelen stellen zich specifieke meetproblemen. Statuslijsten van beroepen zijn vaak niet geactualiseerd, laat staan vergelijkbaar tussen landen. Bovendien is de variabele inkomen een moeilijke (meet men het vermogen? Het equivalent inkomen of geregistreerd inkomen?, op welk moment wordt inkomen gemeten en wordt historiek meegenomen,…).

Geslacht en leeftijd worden standaard geregistreerd, als essentiële variabelen. Toch wordt leeftijd niet altijd op een standaardmanier geregistreerd: soms geboortejaar, soms geboortedatum, soms leeftijd,…).

10.2.2 Meetniveau

In dit type onderzoek zijn gegevens nodig op verschillende meetniveaus. Er is behoefte aan:

1. Individuele gegevens
2. Gegevens over de gezinssituatie / huishouden (gezin verwijst naar verwantschap; huishouden naar samenlevingsvorm)
3. Gegevens over de groep / sociaal netwerk (individuen als leden van een groep met gelijkaardige kenmerken)
4. Gegevens over plaats en ruimte (location)
5. Gegevens over het gezondheidszorgstelsel
6. Gegevens op populatieniveau (exhaustief, soms steekproef,…)

10.2.3 Tijdsdimensie

Bijkomend moet volgens de experts voldoende aandacht besteed worden aan de tijdsdimensie. De meerwaarde van een databank is sterk afhankelijk van de frequentie van dataverzameling, het feit of (continue) longitudinale analyses kunnen gemaakt worden, de standaardisering van de metingen over de tijd, of chroniciteit zorggebruik van patiënten in de loop van de tijd kan opgevolgd worden, etc.

10.3 INVENTARIS VAN DATABANKEN I.V.M. BILLIJKHEID EN GEZONDHEIDSZORG)

10.3.1 Populatiegegevens en algemene statistieken

Algemene statistische informatie is te vinden op:

- voor België op: http://statbel.fgov.be/,
- voor Wallonië op http://statistiques.wallonie.be/

- en voor Vlaanderen op http://aps.vlaanderen.be/index.htm
    - Informatie over de statistische reeksen die aangeboden worden door het Vlaamse "kenniscentrum statistiek" en over de achterliggende bronbestanden is te vinden op http://aps.vlaanderen.be/statistiek/Frameset_over_cijfers.htm. Naast een beschrijving van de data en metadatafiches is ook de kwaliteit besproken volgens de kwaliteitsdimensies van Eurostat.

Demografische informatie (en analyses) is ook te verkrijgen via de databanken die ontwikkeld worden in de steunpunten demografie (VUB en UCL)

10.3.2 Gegevens over "gezondheid"

Een uitgebreid en goed gedocumenteerd overzicht van databanken met gegevens over gezondheid is te vinden op http://www.iph.fgov.be/epidemio/morbidat/. Het heeft weinig zin om deze inventaris te dupliceren. Morbidat maakt een overzicht van databanken bevolkingscijfers, databanken leefstijl, en gegevens over morbiditeit.

De nationale databank mortaliteit laat toe om analyses te maken naar sociale en regionale sterfteverschillen in België. De nationale databank mortaliteit bestaat uit een geanonimiseerde koppeling van de socio-economische gegevens uit de sociaal-economische enquête en de informatie over sterfte uit het Rijksregister.

In functie van specifieke aandoeningen, wordt de epidemiologie en het gebruik van voorzieningen gedocumenteerd via verschillende registers. We verwijzen hier bijv. naar het kankerregister, mucoregister, registers via huisartsenpeilpraktijken,…

Voor gezondheidsgegevens die beschikbaar zijn over de verblijven moet gewerkt worden via de MKG of de MPG, waarin de hoofd- en nevendiagnose wordt geregistreerd en de ernst van de aandoening. In de databank rust- en verzorgingstehuizen zijn scores geregistreerd over de mate van afhankelijkheid.

Informatie over zelfgerapporteerde gezondheid moet achterhaald worden via de gezondheidsenquête, SHARE of PSBH.

Potentieel belangrijke informatiebronnen voor regionale informatie zijn de Observatioria van Brussel en Henegouwen

10.3.3 Sociaal-economische kenmerken van de populatie

Een belangrijke algemene informatiebron voor het bepalen van sociaal-economische positie zijn de fiscale gegevens. In het bestek van deze inventaris zijn we niet op deze databank ingegaan, maar het is een potentieel zeer belangrijke databank die nu in het billijkheidsonderzoek en het onderzoek naar sociaal-economische verschillen zeer weinig wordt gebruikt. Een van de verklaringen moet gezocht in de omslachtige aanvraagprocedure voor de koppelingsprocedure die vereist is.

De sociale kruispuntbank (datawarehouse arbeidsmarkt) is een van de belangrijke bronnen om sociaal-economische gegevens te achterhalen. Het datawarehouse arbeidsmarkt en sociale bescherming koppelt arbeidsmarktgegevens afkomstig van de verschillende instellingen van sociale zekerheid.

De (EU)SILC- enquête peilt naar inkomen en levensomstandigheden. Omdat de bestaande bevraging gebaseerd is op het Rijksregister zijn armoedegroepen echter niet of ondervertegenwoordigd. Op dit ogenblik loopt een initiatief bij

het steunpunt armoede (UA), om deze problematiek in de bestaande gegevensregistraties aan te pakken of te compenseren.

IMA databanken bevatten gegevens over verzekerde statuut, al of niet Maf regeling, en een aantal persoonsgegevens. Opnieuw heeft deze databank alleen gegevens over de sociaal verzekerden. Ze kan dus niet helpen om de situatie te schetsen van de meest precaire maatschappelijke groepen.

De SHARE databank bevat sociaal-economische gegevens voor een steekproef van 50 plussers. Bovendien is de registratie gestandaardiseerd gebeurd in verschillende landen.

De gezondheidsenquête is een belangrijke informatiebron voor sociaal-economische kenmerken van de respondenten. Het gaat echter om een enquête, en geen continue registratie

Als randinformatie vermelden we ook het belang van het rijksregister. Het Rijksregister is een systeem van informatieverwerking dat informatie betreffende de identificatie van natuurlijke personen, op allerlei aspecten omvat. Via het Rijksregister kunnen lijsten getrokken worden, diverse vormen van statistieken alsook documenten die zeer nuttig zijn voor het gewone of buitengewone beheer van de bevolking. Voor het onderzoek in dit veld is vooral het rijksregisternummer een belangrijke variabele die gebruikt wordt bij de koppeling van gegevens.

10.3.4 Levensomstandigheden en leefstijl

De belangrijkste gezondheidsgedragdatabanken zijn geïnventariseerd op de morbidat webstek. Andere gegevensbronnen die indicaties kunnen geven over levensstijl zijn:

In de gezondheidsenquête worden vragen gesteld naar leefgewoonten en naar het gebruik van gezondheidszorgvoorzieningen.

De huishoudbudgetenquête (PSBH) en de EU-SILC behandelen een brede waaier van sociaal-economische en gezinssociologische thema's en laten toe een aantal levensstijlgegevens te identificeren. Het gaat om een panelstudie die dus toelaat een longitudinale opvolging te doen.

De sociaal-economische enquête laat toe om een aantal levensomstandigheden te identificeren.

10.3.5 In toegang tot gezondheidszorgvoorzieningen (financieel en andere)

De analyse van toegankelijkheid van de gezondheidszorg kan vanuit verschillende invalshoeken gebeuren. In die zin is het niet eenvoudig om alle mogelijke relevante aspecten van toegankelijkheid en barrières te gaan bespreken. Er zijn echter gegevensbronnen die verschillende soorten gegevens verzamelen.

Een belangrijke bron van informatie waarin toegankelijkheidsdrempels bevraagd worden is de gezondheidsenquête.

Financiële barrières kunnen afgeleid worden door de PSBH of de SILC en IMA-gegevens.

Geografische spreiding van type voorzieningen zou afgeleid moeten worden van de databanken zoals het CIC het kadaster, etc. Sommige pathologieregisters hebben gegevens waaruit toegankelijkheidsvraagstukken kunnen worden afgeleid.

Uit de kruispuntbank sociale zekerheid kunnen sociale zekerheidsstatuten worden afgeleid.

De databank van de Vlaamse zorgverzekering laat toe om het sociaal-economisch profiel van haar leden op te volgen en te analyseren. Het is mogelijk om een analyse te maken van de gerechtigden naar geslacht, leeftijd, zorgbehoevendheid, etc.

10.3.6 In het gebruik van gezondheidszorg

De databanken die relevant zijn voor het analyseren van het gebruik van de gezondheidszorg zijn dezelfde als diegene die in hoofdstuk 2 vermeld zijn: IMA databank, Farmanet, MKG, MVG, MPG en de databank rusthuizen (op het niveau van verblijven)

Voor het gebruik van perinatale voorzieningen zijn de SPE en BDMS databank een relevante informatiebron, die ook sociaal-economische gegevens registreert.

Het kankerregister zeker, andere pathologieregisters soms, laten toe het gebruik van (specialistische) voorzieningen af te leiden.

De databank Vlaamse zorgverzekering moet voor de gerechtigde leden ook toelaten om een analyse te maken van het gebruik van verzorgende voorzieningen en over de kostprijs van deze voorzieningen.

10.3.7 In de kwaliteit van de geleverde zorg

Geen databanken zijn specifiek georiënteerd op het registreren en het meten van kwaliteit, maar consumptiegegevens laten toe om zekere afleidingen te maken over het type voorziening waarvan gebruik gemaakt wordt, het type behandeling dat de patiënt ondergaat, etc. Gegevens over de organisatie van de voorzieningen bieden een aantal aanknopingspunten .

11 DOORLICHTING VAN DE HUIDIGE SITUATIE

Algemeen is het op dit ogenblik niet vanzelfsprekend om aan de hand van geregistreerde gegevens in België sociaal-economisch en gezondheidszorgonderzoek te doen. Dat betekent niet dat er geen mogelijkheden zijn, maar de situatie staat nog ver van een gewenste.

11.1 INHOUDELIJKE ASPECTEN

Een eerste vaststelling is dat gegevens niet verzameld worden vanuit een conceptueel model, en definities die relevant zijn voor dit type onderzoek. Er moet over het algemeen gewerkt worden met gegevens uit administratieve databanken, die minder bekommerd zijn om het wetenschappelijke perspectief. Het gevolg is dat van bij de start zich het probleem stelt om gegevens te vinden die aansluiten bij een bepaald analyseperspectief. Databanken worden ontwikkeld met eigen conceptualiseringen en operationaliseringen die analyses en interpretaties bemoeilijken. Een vaak vermeld voorbeeld in deze heeft betrekking op het inkomen, waarvoor zeer verschillende operationaliseringen gehanteerd worden. Het gevolg is dat er nauwelijks coherentie bestaat, in het registreren van sociaal-economische informatie. Een ander belangwekkende ontwikkeling voor toekomstig onderzoek is bijvoorbeeld het meten van "scholingsniveau", in een onderwijslandschap dat in Vlaanderen en Wallonië niet altijd op dezelfde manier evolueert. In de paragraaf over het meten zijn een aantal andere aspecten aangekaart.

Het probleem van ondervertegenwoordiging van (vooral lagere sociaal-economische categorieën en kwetsbare groepen) is een onopgeloste kwestie in de huidige registraties of surveys. Die onderregistratie kan het gevolg zijn van het feit dat surveys zich baseren op het rijksregister, of omdat het gekend is dat de non-respons op surveys van kwetsbare groepen bijzonder hoog is. Gezien de intrinsieke opdracht van de verzekeringsinstellingen worden in die databanken ook geen gegevens opgenomen van niet-verzekerden, meestal toch de meest kwetsbare groepen (uitgeslotenen, asielzoekers,...)

Een van de belangrijke vragen om een degelijke interpretatie te maken over het billijke gebruik van gezondheidszorg is de relatie tussen accurate pathologie- of morbiditeitsgegevens, afhankelijkheidsgegevens en gegevens over consumptie.van gezondheidszorg. Op dit ogenblik is het niet altijd mogelijk om systematisch deze relatie te leggen.

Met betrekking tot het gebruik van de gezondheidszorg, waarin objectieve omstandigheden en subjectieve percepties enz een belangrijke rol spelen, zijn enkel in selectieve bronnen gegevens ter beschikking over opvattingen i.v.m. betaalbaarheid van de gezondheidszorg, noch veel informatie over uitstel- en afstelgedrag. de belangrijkste bronnen zijn hiervoor de gezondheidsenquête of de SHARE.

Voor wat betreft het aspect kwaliteit van zorg wordt erop gewezen dat er in België te weinig evaluaties beschikbaar zijn van de geleverde zorg. Er wordt niet bijgehouden "waarom" een interventie gebeurt, wel de frequentie ervan. Een denkoefening over het monitoren van kwaliteit zou ook voor het billijkheidsonderzoek van belang zijn

In hoofdstuk 2 is al uitvoerig ingegaan op het ontbreken van (toegankelijke) gegevens over complementaire en supplementaire verzekeringen. Om een goed zicht te krijgen op ongelijkheden en verschillen m.b.t. tot aspecten van sociale zekerheid is kennis vereist over de aanvullende en facultatieve

verzekeringsstelsels. Er is een gebrekkige (en gecentraliseerde) dataverzameling van de activiteiten van de commerciële verzekeringsmaatschappijen en van de aanvullende en facultatieve diensten van de ziekenfondsen.

Er wordt op gewezen dat het gebruik en exploiteren van gegevens voor (beleids)onderzoek niet enkel mag geïnspireerd worden door te focussen op de gezondheidszorgvoorzieningen. Er is in de eerste plaats nood aan gegevens en databanken over de noden/behoeften van de populatie. De gegevens ontbreken ten dele voor behoefteplanning.

- Om de gezondheidszorgbehoeften te kunnen plannen (bijvoorbeeld vergrijzing: aantal bedden in RVT; chronische ziekten), moet men de zorgnood kunnen opvolgen (monitoring). Daarvoor zijn ook gezondheidsgegevens noodzakelijk.
- Van sommige problemen is weinig informatie beschikbaar: bijvoorbeeld informatie over de extramurale voorzieningen en de persoonlijke consumptie zonder medisch voorschrift. Van dit laatste bestaat er soms wel informatie in private databanken (bijvoorbeeld farmaceutische consumptie). Deze informatie ontbreekt in heel wat landen.

11.2 KWALITEITSKENMERKEN VAN DE DATABANKEN

Zoals in hoofdstuk 2 al is aangekaart, ontbreekt het aan de basisinformatie over de inhoud (metadata) en het gebruik van de databanken. Het is nochtans belangrijk snel een beeld te kunnen krijgen wat de groep of populatie is waarover gegevens worden verzameld, zeker wanneer er inconsistenties opduiken in de gegevens. Een typisch voorbeeld zijn de statistieken van het NIS die alleen geboortecijfers vermelden over personen opgenomen in het bevolkingsregister. Andere registers stellen dat criterium niet, en meten daarom andere geboortecijfers. Een duidelijke en snel beschikbare documentatie over de metadata zou heel wat onderzoek een eind vooruit helpen.

Voor veel databanken is het nauwelijks te achterhalen hoe representatief de verzamelde gegevens zijn. Er stellen zich voor veel databanken dezelfde kwaliteitsproblemen als in hoofdstuk 2 zijn aangekaart.

De actualiteit van enerzijds de verzameling en anderzijds het ter beschikking stellen van de data verdient meer aandacht. Zowel voor het wetenschappelijke werk en beleidsevaluatieonderzoek als voor het adviseren van beleid is dit een zeer heikel punt. De impact van bepaalde sociale correctiemaatregelen kunnen soms niet geëvalueerd worden omdat de reglementering sneller wijzigt dan dat er data ter beschikking komen.

Als algemene opmerking wordt er ook op gewezen dat bij het opzetten van nieuwe registraties, te weinig wordt stilgestaan bij de potentiële relevantie van sociaal-economische informatie. Zo wordt er bijvoorbeeld op gewezen dat de recente registratie van gegevens over borstkankerscreening, een problematiek waarin toch heel wat sociaal-economische aspecten komen kijken, niet eens sociaal-economische gegevens worden geregistreerd. Voor de toekomst is er een conceptuele en operationele reflectie nodig, inclusief het moeilijke probleem van het beschermen van de privacy, hoe een koppeling kan gemaakt worden met geregistreerde sociaal-economische gegevens uit de kruispuntbank.

Er bestaat op dit ogenblik bijzonder weinig coördinatie bij de verzameling van data over gezondheidszorg en sociaal-economische verschillen. Dat brengt zeer specifieke problemen met zich mee wanneer gegevensbanken gekoppeld worden. Het meest fundamentele probleem is het ontbreken van degelijk

geregistreerde gegevens op individueel niveau, en dat binnen alle zorgvoorzieningen. Dergelijke registratie is niet alleen om inhoudelijke redenen van belang, maar zou het een stuk makkelijker maken om in functie van vooropgestelde vragen, verschillende databanken te koppelen. Er worden daarom door alle betrokkenen sterke pleidooien gehouden om persoongegevens systematisch te registreren met een unieke en steeds dezelfde bronidentifier in de verschillende bestanden. Hierbij kan onderzocht worden hoe in het model van de Kruispuntbank Sociale Zekerheid met deze problematiek werd omgegaan. Voor de coördinatie van persoonsgebonden sociale zekerheidsgegevens is hier een oplossing uitgewerkt. Andere voorbeelden zijn te vinden in buitenlandse oplossingen Op dit ogenblik is een ontwerpwet in de maak om dergelijke unieke patiëntenidentificator ook in het Belgische stelsel te creëren.. Een grondige neutrale analyse van de voor- en nadelen van de verschillende aanpakken lijkt op dit ogenblik zeer belangrijk om dit wettelijk initiatief te verfijnen en ervoor te zorgen dat wijzingingen in de reglementering geen bijkomend obstakel vormen om op efficiënte wijze beleidsanalyses uit te voeren.

Voor sociaal-economisch onderzoek is er een groeiende behoefte aan longitudinale data, in het bijzonder om de (verschillen in) zorgtrajecten te kunnen reconstrueren. Op dit vlak loopt er in België geen programma met als doel een model te ontwikkelen hoe dergelijke longitudinale gegevens verzameld en georganiseerd kunnen worden, met uitzondering van de denkoefening over het unieke patiëntennummer. In het kader van het AGORA-programma wordt een voorbereidende studie voorzien voor het complementeren van het gezondheidsinformatiesysteem met gegevens vanuit een dynamisch levensloopperspectief.

Met betrekking tot gegegevens die gekoppeld zijn aan het rijksregisternummer, en in het bijzonder bij de mortaliteitsstatistieken, is de kwaliteit van de geregistreerde sociaal-economische gegevens zo bedroevend dat geen accurate analyses kunnen gemaakt worden. Dit staat in schril contrast met de kwaliteit van de perinatale statistieken, waar heel wat meer mogelijkheden liggen.

Verder gelden gelijkaardige opmerkingen als in voorgaand hoofdstuk. Het vraagt op dit ogenblik enorm veel werk om de concrete inhoud en kenmerken van data binnen de databanken te identificeren, om gegevens uit databanken te koppelen, zowel wat betreft administratieve procedures, als wat betreft de concrete datawarehouse.

Tot slot staan we nog even stil bij de problemen die er vandaag zijn om internationale vergelijkingen te maken op basis van de beschikbare databanken. Zo worden er in nagenoeg alle geïndustrialiseerde landen op regelmatige tijdstippen nationale gezondheidsenquêtes georganiseerd. De gebruikte instrumenten zijn echter zo uiteenlopend, dat vergelijking soms moeilijk, zelfs onmogelijk is. Onder impuls van DG SANCO en EUROSTAT is de ontwikkeling van een standaardinstrument dat in de verschillende lidstaten van de EU zal gehanteerd worden volop in een stroomversnelling geraakt. Op dit vlak speelt het WIV een belangrijke rol.

12 KERNBOODSCHAPPEN

- Gegevens in de beschikbare databanken worden niet verzameld worden vanuit een conceptueel model dat relevant is voor billijkheidsonderzoek. Het gevolg is dat er nauwelijks coherentie bestaat in het registreren van relevante informatie.
- Het probleem van ondervertegenwoordiging van (vooral lagere sociaal-economische categorieën en kwetsbare groepen) is in de huidige registraties of surveys niet opgelost.
- Op dit ogenblik is het niet altijd mogelijk om systematisch de relatie te leggen tussen accurate pathologie- of morbiditeitsgegevens, afhankelijkheidsgegevens en gegevens over consumptie van gezondheidszorg.
- Er zijn in België te weinig gegevens beschikbaar over kwaliteit van zorg (zoals dat ook van belang is voor billijkheidsonderzoek)
- Het ontbreekt aan (toegankelijke) gegevens over complementaire en supplementaire verzekeringen
- Er is behoefte aan gegevens en databanken over de noden/behoeften van de populatie.
- Bij het opzetten van nieuwe registraties, wordt te weinig stilgestaan bij de potentiële relevantie van informatie over sociaal-economische karakteristieken.
- Het gebrek aan coördinatie bij de dataverzameling brengt problemen met zich mee wanneer gegevensbanken gekoppeld worden.
- Een fundamenteel probleem is het ontbreken van degelijk geregistreerde gegevens op individueel niveau, en dat binnen alle zorgvoorzieningen.
- Er is een groeiende behoefte aan longitudinale data, in het bijzonder om de (verschillen in) zorgtrajecten te kunnen reconstrueren.
- De toegankelijkheid en de exploitatie van gegevens uit databanken is zeer omslachtig en remt in zeker zin een snelle exploitatie van de beschikbare gegevens. De informatie over de databanken, data en metadata is overigens in vele gevallen zeer moeilijk te achterhalen.

Hoofdstuk 4 : INTERNATIONAL EXPERIENCES

JAN PEERS
HANS DEBRUYNE
INGRID MOLDENAERS
BAUDOUIN LEJEUNE
VIRGINIE VAN DAMME
MARK LEYS

13 INTRODUCTION

This part of the report focuses on management processes and organisation models of health care data in 3 countries (Scotland, Germany and Denmark). This overview is intended to get a better understanding of choices that can be made in the organisation of health care data warehousing, for the purpose of internal policy and research uses, as well as international reporting to instances as Eurostat, WHO, OECD, etc.

The selection of the countries discussed in this chapter is limited, because of pragmatic, project specific reasons. Three European countries were selected that rather recently made efforts to develop an integrated health information system. The scope of the research was not on understanding all technical details of the datawarehousing, but instead, on trying to understand the organization and management principles used for implementing and developing the health information system.

Further in this report reference will also be made to other interesting countries, having put major efforts in developing health information systems: in particular the Canadian[25];CIHI, 2005 #19;Roos N, 1999 #1;Black C, 1995 #2} and the New Zealand[10] system. For these countries, the information was only coming from document and website analysis, and ad hoc references are made in the report.

14 METHODOLOGY

The data collection was based on both desk research, document analysis and interviews in the respective countries.

The information used during desk research comes mainly from Internet search and to a lesser degree from contacts with local correspondents.

Interview guidelines have been drafted and potential candidates for interviews have been identified. The following subjects were covered in the interview questionnaire:

- Data collection: The process of identifying data sources, defining (using classifications when appropriate) and gathering data about particular issues related to healthcare, taking into account contextual elements and statutory (legal) regulations;
- ETL: Extraction, transformation and loading of collected data, using additional classifications when appropriate;
- Controls: Validation and quality check of collected data;
- Primary exploitation: Processes of primary use (descriptive and analytic) related to the primary purpose of the data;
- Secondary exploitation: Processes of access to and use of databases for other reasons than the primary exploitation of the data, including utilization by and access of the database for external (accredited) users.

In order to prepare the interviews, interviewees were asked, before the meeting, to provide a short overview of the healthcare data collection process in their countries and to draw up a scheme to explain how the process is organised in their country (i.e. organisations involved? contact persons?). The three countries were visited by 2 persons doing the interviews with several key national experts.

15 SCOTLAND

15.1 INTRODUCTION

Scotland is part of the United Kingdom of Great Britain and Northern Ireland (UK), which is made up of four constituent countries: England, Scotland, Wales and Northern Ireland. Until May 1999, Scotland was governed centrally through Westminster. With the establishment of the Scottish Parliament and Welsh Assembly the responsibility for health has been devolved.

Separate NHS organisations operate in each of England, Scotland, Wales and Northern Ireland. These organisations report respectively to the Department of Health, the Scottish Parliament, the Welsh Assembly and the Northern Ireland Office. There are some structural and procedural differences in all four countries but essential services and policy are similar. Each 'home country' has a separate Minister of Health. Currently the National Health Service (NHS) in Scotland is administered by the Department of Health in the Scottish Office.

15.2 MANAGEMENT OF OVERALL PROCESS

15.2.1 Roles and responsibilities

The Scottish situation is characterized by a centralised, standardised and well structured system, which assures the cohesion of information. There are three main actors involved in the whole process of data management from the collection till the secondary exploitation:

1. Minister for Health and Community Care;
2. The Scottish Executive Health Department (SEHD), sets annual national priorities and targets for Scottish healthcare organisations, and decides on information to be collected;
3. NHS National Services Scotland (NHS NSS), in which the business operating unit Information & Statistics Division (ISDScotland), plays a key role at every step of the datamanagement process (data collection, classification, storage and dissemination). ISDScotland works in partnership with a wide range of organisations and individuals to build the national database.

ISDScotland is organised through Groups and Programmes: "Groups" are responsible for developing delivery of information services, including publications, data definitions, information requests, parliamentary questions, and national information management and technology. "Programmes" focus on the key healthcare priority areas for NHSScotland.

Key groups are:

1. The Data Intelligence Group has the remit to identify gaps in current data provision; to undertake collection and provide analysis of healthcare information on behalf of the NHSScotland and Local Authorities; to assure quality of data collected for national information purposes; and to provide coding advice for various coding conventions.
2. The Healthcare Information Group supports NHSScotland and the Scottish Executive Health Department by advising how to use information to ensure efficient and effective delivery of patient care. The group maintains a range of national databases and provides output in the form of routine publications, reports, software packages and information requests relating to patient care.

3. The Epidemiology and Statistics Group ensures that the information is relevant and used effectively in policy making, planning and evaluation of healthcare services.

15.2.2 Legal Framework

Data providing is not compulsory.

Most of the statistical information published by ISDScotland is badged as National Statistics and therefore adhere to the National Statistics Code of Practice and Protocols.

However, not all sources of information and publications from ISDScotland carry the National Statistics badge, for a variety of reasons such as concerns about the reliability and use of secondary data that has been previously published.

All the personal health information supplied is managed and stored by ISDScotland in a secure and confidential manner. The use of data is managed, according to strict rules.

The principle of freedom of information exists but confidentiality is required and assured. Access to personal healthcare information requires special authorisation and is closely monitored by a Privacy Advisory Committee (PAC). This Committee is an independent body, which provides advice on requests for the release of patient identifiable information. In this way, PAC makes sure that any information release is carefully controlled.

15.2.3 Time Line

The data process starts from the collection, which varies from 1 day to 1 month.

Twice a year, results are published with data of the last 3 months.

15.2.4 Decision Model

Annual Targets set by politicians and the SEHD in terms of healthcare objectives are translated in terms of data collection and information. (release each September).

As soon as the national targets are released, Operating Divisions and Community Health Partnerships (CHPs) are asked to put together their proposed input into the Local Health Plans (LHP). A final LHP is agreed between NHS boards and the SEHD during January and February each year, ready for the NHS financial year starting in April.

15.2.5 Management aspects

ISDScotland owns and organises the data warehousing. It sets up data schemes, defining the process and data required, in order to standardise the data collection process. It also manages the process of data use internally or externally.

15.2.6 Financing

Scotland does not have one single reporting system accessible and used by al types of healthcare providers.

For the Primary care sector, a GPASS (General Practitioner Administration System for Scotland) is placed at the disposal of the GPs to send out required information.

ISDScotland estimated that the GPASS has costed £ 3 – 5 millions since the beginning of the project. Recently an evaluation to implement such a system for a single hospital (+- 100 beds) amounted to £10 – 15 millions; NHS (England) valued its project to build such a reporting system in UK to cost £50 billions over the next 25 years.

15.3 DATA COLLECTION PROCESS

15.3.1 Data Sources, Roles and Responsibilities

National Data Schemes stipulate what data are to be collected, how they are collected and where they have to be sent to. ISDScotland has defined and manages more then fifty data schemes to collect healthcare information. Consequently data are provided only once.

15.3.2 Data Collection Mechanism

In the primary care sector, the patients' personal information are recorded into the Community Health Index (CHI) System. The CHI is a computerised index for patients within the NHS in Scotland. It holds name, address including postcode, date of birth, gender and details of registered GP and date joined GP lists. It holds some historical data including previous GP/address/name. Each patient on the CHI has a unique 10 digit CHI number which consists the date of birth and four other numbers. This number is used for communications within the NHS and to track patient medical records.. The information is then kept in the database, accessible for doctors.

In the secondary and tertiary care sector, whenever a patient is admitted in a healthcare establishment, information (such as accident and emergency treatments, inpatient and / or day cases admissions, out patient attendances at hospitals and clinics, and day hospital attendances) are recorded. Each episode of care has to be recorded in the SCI (Scottish Care Information). The SCI is an integrated programme of Information Management & Technology aimed at delivering NHSScotland-owned products and standards to support clinical communication and Electronic Patient Record (EPR) and Electronic Health Record (EHR) development

Data are transmitted without any legal constraints to ISDScotland.

1. In the primary care sector the PTI (Practice Team Information) collect data on patients care on a sample of currently 62 Scottish general practices, The PTI is using the GPASS computing system, submitting information regarding face-to-face consultations. The data collected are then extrapolated by ISDScotland.
2. The secondary and tertiary care sectors provide ISDScotland data about the utilisation of healthcare resources. These data are collected by healthcare staff without specific funding. Nevertheless training and guidance is provided by ISDScotland.

Data collection is mainly electronic. The information collected includes the CHI- number which is a computerised index for patients within the NHS in Scotland. However data providers are free to use the mean of their choice to report. As a consequence, ISDScotland expectionally receives data on paper.

1. In the primary care sector, GPs are free to use any type of reporting system. ISDScotland provides the GPs with the free software called GPASS which is a completely integrated Windows-based system for supporting general practitioners (GPs) in their practice tasks.

   The payment of the primary care providers is related to the data they provide to ISDScotland via the Quality & Outcomes Framework (QOF). The QOF is using evidence-based quality indicators. Participation is voluntary, but almost all practices in Scotland take part in the QOF.

2. In the secondary and tertiary sector, ISDScotland does not provide any reporting software. These care providers are free to use the mean of their choice to report their healthcare data to ISDScotland.

   The secondary and tertiary sector data providers do not have any incentives for sending their data. Payment of hospitals is not linked to their reporting. The transmission of information is not a criterion for accreditation.

15.4 EXTRACTION, TRANSFORMATION AND LOADING OF DATA (ETL)

ISDScotland is the only actor involved in the ETL of data. ISDScotland is responsible for the data warehousing. However, ISDScotland works in partnership with a wide range of other national organisations to build the national database, using national datasets to support the collection and management of information

For "record linkage" (i.e. a pathology linked to a patient), the NHS charges because this service is often used for research's purposes by scientists, and needs ad hoc interventions. For confidentiality reasons, names are replaced by digits.

15.5 QUALITY CONTROLS

Healthcare Professionals validate on logic their own data before sending them to ISDScotland. ISDScotland has its own quality assurance process via:

1. Audits: based on a sample of 1 %, checked by auditors in person.
2. Training: to support data quality, ISD provides advice, training and guidance on definitions and standards for data gathering, clinical coding, data validation, accreditation and quality assurance.
3. Collaboration with colleagues throughout Scotland is essential to every aspect of managing the quality and completeness of national data.

The majority of statistics released by ISDScotland is published under National Statistics. National Statistics is a quality marker applied to certain of the UK's official statistics, guaranteeing that criteria were met.

15.6 EXPLOITATION

15.6.1 Primary Exploitation

NHS naturally becomes the owner of data it receives from healthcare professionals. ISDScotland, on behalf of NHSScotland, manages the National Database. The exploitation of the data are primarily used to inform policy makers to improve Scotland's patient care.

The NHS National Services Scotland (NHS NSS) handles patient data for administrative, research related and medical purposes, which contribute to

improving people's healthcare. This information is used for planning and managing healthcare (analysis of the provision and the effectiveness of healthcare services in Scotland).

PTI data from the primary care sector can be used to estimate incidence and prevalence of any condition presenting to general practice. Results can be used to show trends over time; to determine the most common conditions/diseases; and can be shown by gender, age and deprivation.

Information sent to ISDScotland ends up in a number of different reports and systems which allows healthcare stakeholders to easily compare one area to another for different aspects of care.

15.6.2 Secondary Exploitation

Researchers and external people can use data for various purposes such as management, audit, planning, and research.

Data can be obtained on request. However, a set of publications are available directly on paper or via other means such as websites (i.e. SDScotland website Scottish Health Statistics or Scotland's Health on the Web - SHOW).

Depending on topics, publications are either quarterly or bi-annually or annually released.

15.6.3 Relations with international organisations such as WHO and OECD

For the international relations, Scotland is under the UK reporting. The Office of National Statistics (England) uses the EU and UK classification in order to record national number of diseases.

15.7 SUMMARY ASSESSMENT

Strenghts

- Centralised, standardised and well structured system, which assures the cohesion of information.
- Long tradition of collecting healthcare data: schemes developed by the NHSScotland give a clear framework of the type and timing of data to be collected.
- Representative groups of GP's are contracted to collect and transmit data about GP's practices within the framework of a paid for ad hoc system.
- Use of a comprehensive framework (QOF) to promote exchange of info and to improve quality of care.
- National Statistics label.
- Quality controls are enforced by systematic auditing on top of logic check.

Weaknesses

- Lack of flexibility due to difficulties to change the traditional way of working.
- Feedback to data providers is not resulting in a strong incentive.
- Limited budget.
- GPASS is expensive.
- Unique Patient Identifier under development.

16 GERMANY

16.1 INTRODUCTION

Germany is a federal republic consisting of 16 states ("Länder"). Decision-making power in health care is shared between the Länder, the federal government and legitimized civil society organizations.

1. At the national level, the Federal Assembly, the Federal Council and the Federal Ministry for Health and Social Security are responsible for passing health reforms concerning statutory insurance.
2. The Länder are responsible for planning inpatient capacities and nancing investments in hospitals, nursing homes and institutions for social care. In addition, they supervise corporatist actors and pharmaceutical manufacturers in their constituency.
3. The corporatist level is represented by the non-pro t, quasi-public sickness funds and their associations and associations of statutory health insurance (SHI)-affiliated physicians' and dentists' on the provider side.
4. The sickness funds are the collectors, purchasers and payers of statutory health and long-term care insurance.
5. Physicians treating SHI-insured patients are organized in regional physicians' associations, based on obligatory membership and democratically elected representation.

16.2 MANAGEMENT OF OVERALL PROCESS

16.2.1 Roles and responsibilities

The German health care datamanagement is characterized by a highly decentralised system at the regional level and is accomplished in cooperation with numerous other stakeholders having experience and expertise in a specific field.

The Federation and the Länder share the statistical tasks to be performed in line with the federal structure of the Federal Republic of Germany. At the Federal level, two main actors, the Robert Koch Institute (RKI) and the Federal Statistical Office, collect healthcare data in the following areas: services and utilisation of the healthcare sector, resources of healthcare and cost and funding of the healthcare sector. The Robert Koch Institute, the Federal Statistical Office and the Federal Ministry of Health and Social Security are jointly responsible for monitoring the healthcare sector at the federal level. They operate under the umbrella of the Federal Ministry for Health and Social Security (BMGS), which shapes the health policy and ensures the functioning of the healthcare system. BMGS has the administrative responsibility for healthcare monitoring. Proposals of BMGS have a major impact on decisions on the kind of information to collect. The discussion process is based on laws, voted by the Government but enforced by the BMGS for health matters.

1. The Robert Koch Institute (RKI), based in Berlin, serves the Federal Ministry of Health. The RKI is responsible for coordinating and carrying out the Federal Health Reporting service (Gesundheitsberichterstattung des Bundes, GBE). The RKI collaborates with the Federal Statistical Office and other Federal authorities to maintain and update registries and other data at the federal level.

2. The Federal Statistical Office (FSO) – "Statistisches Bundesamt", based in Bonn, is a central agency responsible for:
   a. the methodological and technical preparation of federal statistics; It is responsible for the operation of the information system, evaluates only available data and provides data in an aggregated and standardised form.
   b. the coordination of the complex cooperation processes in the various fields of statistics, ensuring that federal statistics are produced in the Länder without overlaps, according to the agreed uniform concepts, and in time; the compilation and dissemination of federal data;
   c. the cooperation with international and supranational institutions (European Union, United Nations, OECD,) and relations with statistical institutes abroad.
   d. information dissemination.
3. 16 regional statistical offices of the Länder are responsible for uniform and timely data collection and processing and develop and disseminate Länder statistics.

16.2.2 Legal Framework

Laws about data gathering have to pass both chambers of Parliament, authorizing and obliging the statistical offices are to implement the survey. with the statistical offices of the Länder, existing in all fields of statistics

A specific law on federal statistics defines the tasks of the Federal Statistical Office as the central institution and the tasks of the regional statistical offices of the Länder.

Objectivity, neutrality and scientific independence of the statistical offices at federal and Länder level are laid down in the statistics law. Detailed rules for statistical confidentiality and protection of privacy have to be followed. The Federal Statistical Office is entrusted with the functions of planning, conception and coordination as well as the aggregation of all Länder results. The central institution must ensure that user needs are comprehensively taken into account and results of individual statistics, produced by different parties are comparable creating the possibility to aggregate and benchmark data.

The statistical offices of the Länder perform data collection and data processing on the level of the Länder. They operate closer to the respondents and have better knowledge of the relevant regional particularities. Access to administrative data is usually easier because of the federal structure of Germany. Exceptions to decentralised production of data are possible if specified in a specific law (the roles of the institutions involved are defined for the special circumstances for which statistics are collected).

A legal framework has been developed to coordinate state level statistical initiatives into a federal reporting.

Statistical activities – with very few exceptions – must be ordered by law. Knowing that legislation is an extremely time-consuming process, German statistics show a clear lack of flexibility also on the output side.

16.2.3 Time Line

There is no strict timing for the provision of data to the federal level. Routine collection of data varies from several days to several weeks and up to several

months. Due to the German specificities, planning and coordination of work and the processes of decision-making are complex and time-consuming.

16.2.4 Management aspects

No specific tools are available for data providers, leading to a situation were data are provided in various forms.

Data are stored publicly and free of charge on the Federal Statistical Office website. There are two other routes for the dissemination of healthcare reporting at the federal level: publication via print media (www.rki.de) or the Information System of the Federal Health Monitoring (GBE) (www.gbe-bund.de). Ministries can access objective information at the Federal Statistical Office and via the GBE service. These data are used for policy-making and policy-support

At the regional level, the results are published on individual regional statistics websites. A common webbased central access point (GENESIS - www.statistikportal.de), a joint initiative of the FSO and the 16 regional statistics offices, has been implemented. Its purpose is to provide basic statistical information and to facilitate comparison between healthcare information across Länder.

16.2.5 Financing

The funding of official statistics in Germany is similar to the funding of the whole administration. Every level pays for the tasks it is entrusted with. Decisions about the amount of funds available are taken only by the level performing the relevant tasks.

Healthcare reporting is commissioned and financed by the Federal Ministry of Health and Social Security, which also finances in collaboration with the Federal Statistical Office- the data management process of the GBE information system.

16.3 DATA COLLECTION PROCESS

16.3.1 Data Sources, Roles and Responsibilities

The GBE, jointly managed by the RKI and the FSO, aggregates data collected from the multitude of institutions in the healthcare sector. The GBE occasionally launches its own healthcare surveys. Data are collected in the following healthcare fields: resources of the healthcare system; benefits and utilization of the healthcare system; and expenditure on, costs and financing of the healthcare system.

In addition to official statistics from the FSO, data are obtained through surveys, registries, or in the course of medical treatment.

Major data sources are the health insurance agencies, the special social insurances and the hospital statistics. 118 data sources are collected on a yearly basis for the purposes of the GBE. 48 sources belong to the field of responsibility of the Federal Statistical Office and the statistical offices of the Länder; 70 sources come from other institutions outside the statistical offices.

For clinical health research purposes and health economics studies the 'Telematikplattform für Medizinische Forschungsnetze' (TMF), together with the Institut für Medizinische Statistik und Dokumentation, have developed and implemnted a generic model ('Datenschutzkonzept') in which all aspects of pseudonimisation and protection of the privacy have been integrated [26, 27, 11, 28, 29].

16.3.2 Data Collection Mechanism

Uniform and timely data collection and processing up to the level of Länder results is generally the task of the 16 statistical offices of the Länder.

A close cooperation is needed between the FSO and the 16 statistical offices of the Länder in the field of data collection. It is one of the main tasks of the Federal Statistical Office to create a uniform and harmonised methodology for all federal statistics.

During the preparatory stage, a large number of questions concerning the objectives, the concept and the implementation procedures of the survey have to be settled. The federal ministry in charge of the respective subject matter, will consult the Federal Statistical Office early in the process to receive advice on the conceptual arrangements of the survey.

The methodological and technical details of the implementation, which are not defined by law, are discussed and agreed upon at joint expert meetings. From the very beginning in the decision making process, methodological and technical preparations aim at securing uniform federal data. All statistical components need to fit in one consistent reporting system by using corresponding questions and definitions in the various surveys, as well as drawing upon comparable classifications etc.

The FSO has no imposing power. Issues of fundamental importance and conflicts which couldn't be solved on the working level are discussed at the regular meetings of the heads of the statistical offices. Since the Länder are generally interested in having nationally uniform results at their disposal for their own purposes, all parties involved are motivated to agree on uniform solutions. However, sometimes more consensual decisions are taken in stead of methodological and technical sound solutions

1. In the primary care sector the reporting system is linked to the GPs' payment-system. GPs report their data to the doctors associations: Revenues are distributed from the Statutory Health Insurance (SHI) sickness funds to GPs through these GP-associations.. In order to receive money to pay the GPs, the medical associations give aggregated data to the Insurance Funds who forward these to the Federal Statistical Office. Personal information is not available for the Federal Statistical Office since data have been aggregated on the Funds' level.

   A law is in prepartion to oblige doctors' associations and the statutory health funds to report data for policy-support. This debat is linked to a discussion of instauring a UPI.

2. In the secondary and the tertiary sector, healthcare providers have to report by law to the regional statistic office (Länder) and to the INEK, the German Institute for the Prospective Hospital Payment System. INEK collects data for reimbursement purposes based on DRG's.

16.4 EXTRACTION, TRANSFORMATION AND LOADING OF DATA (ETL)

German healthcare data providers are free to choose their way of reporting data. There is no software provided to the healthcare professionals in order to transmit the data, resulting in a wide range of different formats. The data are sent either manually (on paper) or electronically (Microsoft Excel or Access files).

The organizational responsibility for healthcare reporting at federal level is assumed by the RKI. To accomplish their tasks, the RKI cooperates closely with the FSO, taking into account the interest and impulses given by the various institutions of the German healthcare system and anticipating to the information needs of the users of healthcare reporting data.

At the federal level, there is no need for anonymisation because data are already aggregated and anonymous.

The medical associations do not provide patients related data to the SHI sickness funds. Statistical offices only dispose of aggregated data – rates or percentages of persons but no individual data.

16.5 QUALITY CONTROL

Auditing of the data providers is not foreseen.

At the regional level, a logic quality check is performed on the healthcare data by each regional statistical office.. The fact that data are provided in different forms and electronic formats represents a heavy task for the Federal Statistical Office which has to compile them.

The quality information process is supported by the GBE, which focuses on different data sources. There is feedback provided by data delivering institutions during the different stages of the process (acquisition, transformation, publication). Before publication, data delivering institutions revise methods and documentation sources.

16.6 EXPLOITATION

16.6.1 Primary Exploitation

The exploitation of data on the federal level has two prominent objectives:

1. Information on and identification of actions to be taken within the healthcare system;
2. Monitoring, reporting of healthcare situation and identification of discrepancies between services needed and offered.

Healthcare information database are owned and exploited by :

1. Federal Statistical Office:
2. 16 regional statistical offices of the Länder
3. the Common New Statistical Information System (Gemeinsames Neues Statistisches Informations-System - GENESIS), the FSO and the statistical offices of the Länder have set up new information databases in order to make the wide range of official statistics available via the Internet in an easy-to-use manner; Its purpose is to provide central access to basic statistical information and to facilitate comparison between such information across Länder. The information offered through the portal consists of a multitude of retrievable data tables containing information on the Federal Republic of Germany and its Länder as well as online access to statistical databases
4. RKI collects data and disseminates via the Information System of the Federal Health Monitoring (IS-GBE); The GBE provides a fact-based foundation for political decisions. It serves to monitor the success of measures taken and aids the development and evaluation of targets within the healthcare sector. The IS-GBE contains a broad range of information: in addition to text extracts, tables and graphs from the healthcare report, it also offers flexible,

user-controlled indicator tables, descriptions of data sources, code definitions and further background information. IS-GBE uses a relational database to store and retrieve numerical and non-numerical information and a multidimensional database system to present numerical data as arrays or data cubes. The system is accessible via web servers.

5. The Institute for Medical Statistics (IMS) owns health care data not provided to IS-GBE,
6. INEK collects data for reimbursement purposes and makes them available to IS-GBE.

16.6.2 Secondary Exploitation

The data are publicly available and widely used by healthcare stakeholders, politicians, journalists, scientists, students and citizens in general.

16.6.3 Relations with international organisations such as WHO and OECD

The health division of the Federal Statistical Office is in charge of collecting information for international level purposes. Germany delivers duly filled out questionnaires in response to the requests of these international institutions.

16.7 SUMMARY ASSESSMENT

Strenghts

- Data free of charge and easily accessible: Data from the FSO are publicly available online and free of charge.
- Entire independence of the statistical offices of the Länder from the Federal Statistical Office in terms of organisation and funding.
- Credibility of every single element of the system: such credibility is derived from the professionalism and scientific independence in data production, from the objectivity and neutrality of the institutions, in particular with regard to data dissemination, and from safeguarding the interests of the respondents, both regarding the limitation of burdens placed on them and the strict adherence to the principle of statistical confidentiality.
- Compared with a centralised system, decentralisation has at least in large countries some clear advantages: the system ensures close contact between the Länder statistical offices and the respondents, utilizing the statistical offices' detailed knowledge of regional particularities.

Weaknesses

- Due to the decentralised system the planning and coordination of work and the processes of decision -making are complex and time-consuming. Delays may occur in providing federal results, the reason being that statistics produced on a decentralised basis cannot be compiled and published until all or nearly all statistical offices of the Länder have supplied their data.
- The FSO has no decision-making power in methodological issues. Decisions on methodological issues can be taken only in consensus. This process is time consuming and compromises reached at the end of the process are not always the best solutions.
- Uniform results at federal level can only be guaranteed as they are ordered by law, regulating a large number of details of the statistics to be produced.
- Little feedback to data providers.
- Huge collecting effort: data are collected from many sources and in many forms.
- Information overload: there is a lot of aggregated data, difficult to use for policy-making.
- Lack of global vision / mission / objectives. Data collection needs to be more based on exploitation purposes.

17 DENMARK

17.1 INTRODUCTION

Administratively, Denmark is divided into 14 counties, 275 municipalities and the metropolitan areas of Copenhagen and Frederiksberg, which have both county and municipality status.

1. Responsibilities of the 14 counties are: secondary schools, roads and environmental issues; healthcare expenditure accounts for approximately 70% of the budget; ownership and running of hospitals and prenatal care centres; most county councils have set up committees on health and social affairs and hospital committees to oversee their healthcare responsibilities; finance general practitioners, specialists, physiotherapists, dentists and pharmacists through the National Health Security System (NHSS)
2. The 275 municipalities provide services such as: nursing homes, home nurses, health visitors, municipal dentists and school health services. These activities are financed by taxes, with funds distributed through global budgets, and carried out by salaried health professionals.

The defining feature of the Danish healthcare system is decentralized responsibility for primary and secondary healthcare. Most decisions regarding the form and content of healthcare activity are taken at county and municipal level. However, there are important channels, for negotiation and coordination between the state, counties, municipalities, and the political focus on controlling healthcare costs has encouraged a trend towards more formal cooperation.

17.2 MANAGEMENT OF OVERALL PROCESS

17.2.1 Roles and responsibilities

The Danish healthcare system evolves around a centralised system supporting decentralised operations. The uniformity of the health care system is a favourable framework for standardisation and centralisation of the information process, assuring cohesion in information.

The main actors involved in the whole process of data management from the collection till the secondary exploitation are:

1. Ministry of the Interior and Health;
2. National Board of Health (NBH) intervenes in the data collection process, classification, storage and dissemination;
3. Sector research institutions, e.g. National Institute of Public Health;
4. Statistics Denmark, which is in charge of all non-health related data;
5. Danish counties, which are responsible for the decentralized collection and distribution of health statistics.

The National Board of Health is the central authority for activities in health statistics in Denmark It is an autonomous governmental agency depending on the Ministry of Health. NBH's mission is: to monitor healthcare and follow developments within health issues; to advice the Ministry of Health and other public authorities; to supervise healthcare professionals. The NBH gathers information on healthcare deliveries and incidence of diseases in registries. The registries are foremost used for preparing health policies, administration, planning and research, as well as for communicating health related statistics. 22

specialist advisers assist the NBH in monitoring developments within their own area of expertise

Documentation tasks within the NBH are done by 3 divisions: Health Statistics, Health Informatics and Health Economics:

1. The Health Statistics department collects, analyses, stores and disseminates relevant information. This involves running and developing a number of central registries.
2. Health Informatics is charged with the creation of a common foundation for the application and development of IT within health service, in order to ensure the communication of information.
3. Health Economics is responsible for developing, maintaining and applying management and planning methodologies and tools within the health service. It uses the DRG-method.

17.2.2 Legal Framework

Data providing is compulsory in Denmark. The NBH has the final word but is obliged to consult the "Coordinating Committee for registering patient-data including the Ministry of Health and the Interior, the NBH and representatives of counties. This Committee gives advice to the NBH.

The Danish data collection system is mainly based upon the central law, stipulating that the National Board of Health has the obligation to "… follow the development in the health status of the population. Consequently, the NBH can ask –free of charge- all data subject to improve the healthcare system. The law provides the possibility to penalize institutions that would not release their data. However, so far, no cases have been reported. The penalties would mainly consist of fines.

The NBH can request without any charges all the necessary data that help to improve the healthcare system.

Furthermore, the NBH has an obligation to answer questions from citizens and to guide them in their choice related to health concerns.

Linking different registers by CPR (Social Security Number) is allowed only for certain administrative and research purposes

Specific regulations do exist related to the "national patient register" and the "causes of death register".

EU regulation on data protection is strictly followed. Access to data varies:

1. Within the central level of the NBH, everybody has access to all kinds of anonymous data without any legal restriction. However, only a small team of 4-5 persons has access to personal data related to the CPR number;
2. Local divisions of the NBH do not have direct access to the data. However, they can apply to obtain access;
3. Public institutions and private persons/companies can ask to access these data for free (except for tailor-made data requests). A department composed of scientists will investigate each request separately in order to evaluate the relevance of the request (for instance scientific purposes vs commercial). In order to obtain non-aggregated data, both the Danish Data Protection Agency and the National Board of Health need to approve the given project. The data will be delivered by charge.

17.2.3 Time Line

The data development process from data collection to secondary use varies from 1 day to 11 months. The goal is to achieve a gap of maximum six month before secondary use.

17.2.4 Management aspects

The NBH owns and organises the data warehousing. The strategic decisions and basic structure of the reported data are taken by the NBH in cooperation with the municipal and county level

Health care providers are supported by the NBH through an ERS system permitting them to provide and transmit data in a standardized way without extra costs.

The public can access information through the websites Patientinfo.dk and Sundhedsdata.dk. This way the NBH guarantees direct interaction between the registers and the patients.

17.2.5 Financing

The data development process is funded by the Danish government. The Ministry of the Interior and Health yearly allocates a budget for the NBH. This budget is partly used for the regular maintenance of the registries.

For specific initiatives the NBH can apply for additional money, which is not included in the initial budget, to fund expensive projects (e.g. large software updates).

The costs of development and implementation of the ERS in terms of human and financial resources are very reasonable. Only two persons are working on the ERS. This project has been costing so far 4 millions of Danish Krone. As the ERS is an on-going process, other costs will have to be add in the future.

1. In the primary care sector, the National Board of Health gives the facility to the GPs to deliver the data electronically. The GPs would rather have access to free alternatives. An incentive exists for primary care. Their payment is based upon the number of visits reported. More than 80% of GPs uses a computer and up to 100% of those have access to internet. 100% of the data are entered into the system.
2. In the secondary and the tertiary sector, healthcare establishments often choose the full integration into local systems. Although the financing is independent of the number of registrations, healthcare establishments need to report a certain number of registrations to be accredited. Consequently, healthcare establishments are financed according to the patient-mix as shown by the registration of the pathologies.

17.3 DATA COLLECTION PROCESS

17.3.1 Data Sources, Roles and Responsibilities

Denmark has a tradition of decentralized coding (since 1970).

1. In the primary care sector: the general practitioners provide – on a daily basis – all information about patient visits directly to the National Board of Health.
2. In the secondary and tertiary care sector: the healthcare establishments provide on a daily basis all information about patient visits and treatments to the National Board of Health.
3. At the municipal level, information about services such as: nursing homes, home nurses, health visitors, municipal dentists and school health services are released to the National Board of Health.

17.3.2 Data Collection Mechanism

Prior to 2003, data were manually collected. All healthcare providers had to complete paper forms and send back to the NBH. At that time, the NBH system had no central strategy (no central overview, access...), many separate small systems for reporting and for sending the data, and no economy of scale (lots of services, high maintenance costs…).

Healthcare data are collected via the SEI, which is the Danish abbreviation of ERS (Electronic Reporting System).

The project of "electronic reporting and automated coding of data" was initiated in 2003 and progressively replacing the manual coding. The ERS is a tool to transfer data electronically such as admission, diagnosis and treatment from data providers to the NBH.

There are four different technical infrastructures for reporting healthcare data to the National Board of Health. The two first ones do not request any local system development. The two last ones request local system development:

1. The ERS-solution is an "all-in-a-box" system for reporting the data. The ERS is used for filling out the form, signing off the document and sending it back to the NBH. This is the most common used since it is free, easy to access and it does not need a permanent internet connection.
2. Client/server solution needs a browser and internet-access to connect to the register. This access is free of charge. The Citrix setup enables web access through a traditional browser with a Citrix plug-in. Compared to the ERS-solution, there is no need for local installation. The disadvantage of the Citrix solution is the need to be constantly on-line when filling out the forms. Furthermore, the data are placed on a NBH server and consequently they are not directly accessible for the local systems. The NBH leaves it to the reporters to choose the optimal solution.
3. Import/export facilities to and from local systems and ERS. The Import/export facility is integrated in the ERS and therefore free of charge.
4. Full integration into local systems. As this solution is not free of charge, it is not yet widely used. Primary sector stakeholders would rather have access to free alternatives. On the other hand, healthcare establishments do use it. To ensure data consistency, the most efficient solution is to integrate the

reporting directly into the systems used for general patient administration. This is expected to happen with the coming Electronic Patient Records.

The Danish reporting system assures quick, reliable and complete feed back. The timeframe has evolved from a yearly data collection in 1977 over a monthly data collection mid-nineties, to a current daily electronic collection. The output received is greater than the input provided, what motivates data providers to play an active role and participate in surveys.

17.4 EXTRACTION, TRANSFORMATION AND LOADING OF DATA (ETL)

VPN (Virtual Private Network) connections are used for reporting data in a closed system by regional authorities - either via FTP (File Transfer Protocol) for bulk data, or via HTML (Web pages) for single reports.

All this information is stored in a data warehouse, which is the central data storage unit. All data ends up here eventually, with CPR, but can also be seen without CPR by logical column separation, a standard database technology. The data here are accessed by in-house statisticians (with CPR), as well as certain external organizations (without CPR) by secure connection.

All the data stored in the data warehouse are made available publicly accessible via a web server. NBH has not yet consolidated to a single output system for all users. Therefore PIF (outgoing data), which is just one of many, should in the long run be replaced by a general system. Data are gathered in several "registers". Each register covers not only the transformation of data into useful information, but is also responsible for the management of transit of the secondary exploitation

Registers are organised on basis of acronyms of 3 letters (for instance, CAR is set for Cancer Register; IVF In Vitro Fertilisation). Everybody can log into the database and find its own data thanks to digital signature. On the web page, when clicking on acronyms, diseases are linked as well as the variable and examples. All registrations are done on individual patient level and are linked to a unique number, the so-called CPR (Social Security Number)… The information stored under the unique key is identification data and other relevant data as professional and family situation, size of house, diseases… Regulated by law, this 10-digits CPR allows a unique identification and a key for combining information from different registers.

These registers are controlled by ethical committees and by the data inspection.

17.5 QUALITY CONTROL

The NBH takes care of the quality control process. There are different ways to ensure quality of data.

1. Training: coding is an integrated part of the education of doctors and medical secretaries. The final responsibility (quality, correctness of data) remains at doctor level.
2. Information and documentation: the NBH distributes information and documentation on what and how to register.
3. Quality checks: a series of quality checks are executed on different levels and via different mechanisms. For example, concerning data validation for public hospitals, all data sent to the NBH server are automatically checked on their logic by the Board of Health Databases. Up to 98-99 % of data are technically correct (i.e. incompatibility of data). These data are transferred

to the specific register. Incorrect data are sent back to the hospital within one month after submission of the data.

4. Feedback by the NBH towards hospitals: the NBH must provide hospitals with feedback by means of a monthly publication of statistics. These publications allow data providers to compare and to improve their data and their care. This mechanism increases the pressure to provide correct data on time.

The healthcare providers are legally responsible for the validity of the information. This is enhanced by the fact that they receive complete feedback after 2 to 3 months.

17.6 EXPLOITATION

17.6.1 Primary Exploitation

The legal owner and primary user is the NBH. Information is used by the following departments: public health, planning, quality, monitoring and supervision.

They provide these data to secondary users into applicable useful information on: health monitoring and prevention, health planning, supervision of health personnel and health system, quality assessment (i.e. indicators and standards), administration and economy (indeed case mix systems (DRG) depend on the hospital, the pathology and cost per pathology).

17.6.2 Secondary Exploitation

Health statistics are widely used by the following categories of users:

1. Administrators in the healthcare sector have an extensive demand for statistics and records; this information will serve to evaluate the right amount of money each hospital will need.
2. Scientists will be asking for statistics and records for research.
3. Journalists need statistics in order to strengthen their daily articles.
4. Students can get easy access to specific statistics with the purpose of preparing a thesis.
5. Citizens: as better communication will improve the quality of healthcare, NBH is patient-oriented, patients are thus empowered. The Patient Information System (www.patientinfo.sst.dk; www.sundhed.dk) allows the public to get information through a Patient Information Portal, assisting patients in making a choice of healthcare provider and providing info related to waiting times, activity statistics and quality indicators.

The dialogue is informal with user community through e-mails and meetings. Services provided by the NBH are freely available, except for requests demanding lots of work (i.e. developing new systems) or for requests outside the ordinary strategy, such as research, industry…

17.6.3 Relations with International Organisations such as WHO and OECD

Denmark reports in a systematic way to the annual request of these international organisations. The responsible actors to report are: the National Board of Health (which sends non-financial data such as resources, utilisation, outcomes); the Ministry of the Interior and Health (which is in charge of giving the SHA (financial data) expenditure statistics to Eurostat and OECD); and the

Statistics Denmark (which gives mortality rates to New Cronos, a Eurostat database for some health indicators).

17.7 SUMMARY ASSESSMENT

Strenghts

- NBH is one structured system recognised by everyone as working efficiently. This homogeneous system collects information only once ("same question answered only once"), which increases its efficiency.
- By law, any person can request insight in the content of the Danish Healthcare Registers.
- The patient can play an active role by asking the Specialist Advisers how to optimize its care. The patient can interact with the registers through the information systems, Patientinfo.dk and Sundhedsdata.dk.
- Thanks to the consultation of all healthcare stakeholders, a high degree of implementation can be reached.
- Thanks to regular collection and feedback, the healthcare system improves.
- Advantages of the encoding system are: data quality assurance (feedback to providers); coding consistency – for Danish and international analysing purposes (using a specified validated form); speed up of the process of updating the Registers; facilitating of the reporting; reduction of the logical errors in the reporting phase.
- The care deliverers are provided with a basic system (ERS) permitting them to provide and transmit data in a standardized way without extra costs.
- All actors (public, care providers, patients, politicians, decision makers, scientists,…) have access to the same information.
- The use of a unique patient identifier simplifies obtaining coherent data and avoids duplication.
- The NBH gives quick and reliable feedback to the transmitters of data.

Weaknesses

- The NBH faced difficulties in changing the system. However, after several demonstrations and assiduity, the benefits were clear and the system was finally adopted.
- Slow implementation of ERS (Electronic Reporting System); limited fund to provide the healthcare professionals with the latest tools.

18 CONCLUSIONS

The practice of healthcare reporting in the countries studies clearly varies. The development of a health care information system is very much related to the particularities of the health care systems. A centralised and standardised system, assuring the cohesion of information, is in place in Scotland and in Denmark. A more decentralized system is in use in Germany.

1. The rather newly established Danish system is very structured and the more advanced. The success of the Danish system is closely linked to the uniformity of the Danish healthcare system itself. This uniformity is a favourable framework for standardisation and centralisation of the information process, assuring cohesion in information. On the other hand, this centralized system is suffering also of lack of flexibility (cultural resistances to changes) and of the cost of the establishment of a fully centralised data reporting and data warehousing system.
2. The Scottish healthcare reporting system which is also characterized by its centralisation and its standardisation, and leading to particular advantages in the datastreamlining. Comparable to the Danish system high efforts and resources have been put to set up a particular health care information unit
3. The more decentralised model in Germany is in place for more than 50 years. The model ensures that statistics are produced in close contact of the Länder statistical offices with the respondents, utilizing the statistical offices detailed knowledge of regional particularities. Moreover, this organisation guarantees that the Länder specific data requirements are sufficiently taken into account and that a comprehensive programme of regional statistics will be built up and maintained.

   But the decentralised structure has its shortcomings. In particular, the planning and coordination of work and the processes of decision-making within the system are complex and time-consuming. Moreover, delays may occur in providing federal results, the reason being that statistics produced on a decentralised basis cannot be compiled and published until all or nearly all statistical offices of the Länder have supplied their data. The lack of mechanisms for solving conflicts certainly is one of the weak points in the German system. Moreover, uniform results for the Federation can be guaranteed only by the fact that statistics are ordered by law and, more specifically, by laws regulating a large number of details of the statistics to be produced. Moreover, in the particular case of health care data, a lot of initiatives are taken by professional groups and insurance companies, who develop and exploit their own health care data bases

Some conclusions can be drawn on these experiences:

1. A Healthcare system with clear defined roles and responsibilities for all stakeholders (well structured relations – uniformity) provides the context to develop a useful healthcare information system.
2. A highly centralised data system allows better use and coherence of information as well as better protection of data (privacy regulation). Countries as Denmark and Scotland, with a centralised system of data collection, data definition, and healthcare reporting, have noticeable advantages in gathering consistent and homogeneous data.
3. A Clear framework for data collection, and ETL is needed during the whole information chain, from collection to exploitation, standardisation, clear

definitions, clear roles and responsibilities are crucial. Therefore, a clear framework defining the process and data required may help to standardise the data collection process. In a well-structured system, healthcare professionals are guided and know where and when to send which data. This could be done following the example of the Scottish 'data scheme', which clearly defines the kind of data required.

4. The datamanagement process has to provide systematic feedback to data providers. Data providers need to receive rapidly reliable and qualitative feedback in order to allow them to take actions and manage their processes. Limited time between the data collection and feedback process will improve the social control on data. Denmark seems to reach a high level of collaboration and optimal utilisation through quick feedback.
5. The Data quality depends on a set of different elements such as:
    a. A centralised reporting system is a backbone for a modern healthcare system. Standardization is necessary to reach a high level of electronic exchange of information. (Denmark ERS advanced stage – Scotland GPASS provided to primary care stakeholders).
    b. A central database assures the coherence in the use of data and thus quality.
    c. Reliability and correctness of data is the result of a multitude of interventions such as training for healthcare providers, audits, identification of extreme values in the data provided, logical checks on data, benchmarking reports …
    d. The development of a national label in order to attest the quality of data can be considered as a quality standard.
6. One of the main aspects of a user friendly data system is to avoid duplication of input. Once information is registered in the system, this information needs to be automatically available for all users. The use of existing data sources from other departments is crucial to avoid a bureaucratic burden; A high degree of interactivity between the main actors and the central collection point of healthcare data, should avoid multiplication of data transmission. Information has to be registered and transmitted only once in the system.
7. The use of a unique identification number reduces the risk of mistakes in patient information and avoids duplication of data input. It allows centralisation of the patient's information as well as linkage of databases. Denmark is a good example. In Scotland, the unique patient identifier is under development while in Germany, data are aggregated at the federal level.
8. The data system and the related reporting system must take into account all legislation regarding privacy and confidentiality of data. As far as confidentiality is concerned, it is very important to limit access to individual data.
9. Accessibility of data should be guaranteed and well organised at the different health care levels. Using a unique identification number for patients and high security standards are key elements. All actors (public, care providers, patients, politicians, decision makers, scientists,…) need to have access to the same aggregated and anonymous information, using the same sources. This will improve the quality of information and motivate persons and institutions to play an active role in the data collection process.

10. A legal framework has to support the role and responsibility of every player in the process, to enhance reliability of the data and clearly define the time frames to be respected. Compliance with these criteria could be used for accreditation of healthcare providers and other stakeholders.
11. Financing linked to information. Linking the information collection to the payment of healthcare professionals increases the participation rates in healthcare reporting. Nevertheless, financial stimuli is not enough, there need to be clear quality assurances plans.

Hoofdstuk 5 : CONCLUSIES EN AANBEVELINGEN

19 EEN GEÏNTEGREERD DATAMODEL VOOR GEZONDHEIDSZORGINFORMATIE

19.1 EEN CONCEPTUEEL DATAMODEL GEZONDHEIDSZORG

In de Belgische situatie ontbreekt het aan een conceptueel datamodel waarin doelstellingen worden vastgelegd en gezondheids- en gezondheidszorggegevens een plaats krijgen. De huidige gegevens en gegevensstromen zijn historisch ontstaan, meestal uit financieringsdoelstellingen, en bestaan vaak los van elkaar. Enkele uitzonderingen daarop zijn de gekoppelde MKG-MFG databank en recent - in het kader van beleidsondersteunend onderzoek - de MKG-gegevens gekoppeld aan longitudinale mutualiteitsgegevens via het IMA. Een conceptueel datamodel gezondheidszorg vraagt een grondige reflectie over de inhoudelijke doelstellingen en de informatiebehoeften om beleidsmakers te ondersteunen en om maatschappelijk relevante vragen te behandelen. Vanuit een inhoudelijk conceptueel model kan de reflectie starten over de rol van de verantwoordelijke organisaties voor de verzameling van de datasets, coördinatie van datasets, procedures en werkwijzen om data te exploiteren.

Een datamodel moet naar inhoud voldoende breed geconceptualiseerd worden (zo exhaustief mogelijk, inclusief het billijkheidsvraagstuk), met vooral beleidsondersteunende en onderzoeksdoelstellingen voor ogen. Het model moet primair en secundair gebruik verzoenen. Het model dient eveneens als leidraad bij de opzet van nieuwe registraties en de aanpassing van bestaande dataverzamelingen. Het moet dynamisch opgezet worden en generiek genoeg zijn om er in de toekomst de informatie over nieuwe datasets aan toe te kunnen voegen.

19.2 PARTICIPATIE VAN GEBRUIKERS, LEVERANCIERS EN TRUSTED THIRD PARTIES

Er dient een grondige analyse gemaakt te worden van de specifiek primaire en secundaire informatiebehoeften voor nationale beleidsdoelen, administratieve doelen, onderzoeksdoelen en internationale vragen. Een conceptueel model moet expliciet de behoeften van de primaire en secundaire gebruikers, leveranciers van gegevens en trusted third parties inbouwen. Het model wordt in overleg met de sector en met kennis van zaken ontwikkeld. Hoewel het te ontwikkelen model van een klare visie moet vertrekken, heeft het naar de realisatie weinig zin om geen rekening te houden met bestaande rollen en verantwoordelijkheden zoals die in de huidige situatie bestaan. De participatie van de verschillende "stakeholders" in het proces is een essentiële stap om tot een in de praktijk realiseerbare werkwijze te komen.

19.3 RELATIE TUSSEN DATABANKEN

Het ontwikkelen van een conceptueel datamodel vraagt een benadering die verder reikt dan een reflectie over individuele databanken. Er is behoefte aan een geïntegreerd databankenmodel. De relatie en samenhang (coördinatie en integratie) tussen verschillende databanken en datasets moet eveneens bestudeerd worden.

Het conceptueel model moet bovendien oog hebben voor de specifieke problematiek van de koppeling van (gegevens uit) de databanken. Ook hier moet vanuit een visie gewerkt worden die mede een impact kan hebben op de architectuur van de individuele databanken.

19.4 FINANCIËLE MIDDELEN

In functie van de afbakening van die rollen en verantwoordelijkheden zijn voldoende (financiële) middelen vereist. Deze oefening vraagt in de eerste plaats een grondige assessment van het gebruik van de middelen die vandaag al besteed worden aan dataregistratie en -analyse. Op basis van die analyse kan geprogrammeerd worden welke aanvullende middelen vereist zijn om tot de uitbouw een zeer performant gezondheidszorginformatiesysteem te komen, zoals een aantal initiatieven in het buitenland illustreren.

19.5 BUITENLANDSE INSPIRATIE

Voor het ontwikkelen van een model kan België zich zowel qua inhoud als qua methode laten inspireren door landen die een gelijkaardige oefening hebben gemaakt (zoals Canada[18, 20, 8, 25], Zweden[13, 30, 31, 16, 32, 33], Nieuw Zeeland[9, 10], Schotland, Denemarken,…) Die recente initiatieven hebben zelf goed gedocumenteerd hoe zij te werk zijn gegaan.

20 EEN NATIONAAL KWALITEITSPLAN VOOR GEZONDHEIDSZORGDATA

In verschillende landen die werken met een geïntegreerd systeem van gezondheidszorggegevens wordt een degelijk kwaliteitsplan als een fundamenteel onderdeel van een adequaat databeleid gezien. Op dit ogenblik is er in België geen sprake van een geïntegreerde veralgemeende datakwaliteitsaanpak in alle instellingen. De kwaliteit van de gegevensbanken gezondheidszorg wordt bijna uitsluitend gedragen door persoonlijke inzet van de verantwoordelijken voor de databank. Er bestaat, vooral bij administratieve databanken, weinig informatie over de wijze waarop de kwaliteitsbewaking systematisch is georganiseerd en aangepakt. Er wordt te weinig aandacht besteed aan de uitbouw van organisatorische structuren om het uitwerken van een datakwaliteitsplan te garanderen.

Een datakwaliteitsplan vertrekt het beste vanuit een geëxpliciteerd referentiekader. Eppler en Wittig[34] stellen dat volgende aspecten zeker in het achterhoofd moeten worden gehouden :

- Een referentiekader voor kwaliteitsbewaking moet generiek zijn, niet specifiek voor een bepaalde toepassing.
- Een referentiekader documenteert de onderliggende samenhang van de verschillende vooropgestelde kwaliteitscriteria.
- Een referentiekader overstijgt een eenvoudige lijst van "kwaliteitscriteria" en gaat uit van specifieke probleemvelden en indicatoren.
- Een referentiekader is zowel theoretisch als praktisch relevant.

Een 'Data Quality Framework' zoals is uitgewerkt door het Canadese CIHI of de Nieuw-Zeelandse gezondheidszorg[10] levert interessante voorbeelden hoe zo'n raamwerk concreet kan ontwikkeld worden. Het Canadian Data Quality Framework[8] stelt methoden voor om de datakwaliteit te beoordelen, legt standaardeisen op om documentatie over de data en databanken samen te stellen en verplicht tot consistente werkprocessen. Een 'assessment tool' is ontwikkeld dat de gebruikers in staat moet stellen om de beperkingen en sterktepunten van een databank te identificeren. Op basis van een identificatie van de zwaktes worden verbetervoorstellen geformuleerd. Het instrument maakt gebruik van 5 assessment dimensies (accuraatheid, tijdigheid, vergelijkbaarheid, bruikbaarheid en relevantie) en 58 criteria. Voor elke dimensie worden meer specifieke kenmerken geïdentificeerd, en voor elk van de kenmerken worden verschillende criteria vastgelegd waarop aspecten van datakwaliteit kunnen beoordeeld worden .

Een datakwaliteitsplan houdt dus ook rekening met de verschillende (types) eindgebruikers. Participatie van eindgebruikers is belangrijk om te kunnen bepalen welk niveau van kwaliteit uiteindelijk verwacht wordt van data. Een van de eerste stappen bij het opmaken van een kwaliteitsplan is het identificeren van het gebruik en de gebruikers van data en databanken en het vastleggen van wat het toekomstig gebruik van data zou kunnen zijn.

De kwaliteit van de gegevens is in sterke mate afhankelijk van de gevolgde procedures. Al van bij de primaire datacollectie zijn aanbevelingen vereist om de kwaliteit van de gegevens te optimaliseren: juiste syntax en formaat, validatie van input, uitvoeren van "integrity checks", gevalideerde conversie van data, etc. Naarmate men opschuift in de keten van de gegevensprocessen wordt het niet

alleen moeilijker om de validatie van deze primaire gegevens uit te voeren, er zijn ook meer middelen nodig voor het verbeteren van de gegevens. Vanuit dit oogpunt is het essentieel dat, in de Belgische context, veel sneller feedback wordt gegeven aan de instellingen en zorgvertrekkers die de feitelijke basisregistraties verzorgen.

21 NOOD AAN CENTRALISATIE VAN INFORMATIE OVER DATABANKEN EN METADATA

Internationale organisaties (WHO, OECD, EuroSTAT) onderkennen dat de beschikbaarheid van metadata een noodzakelijke voorwaarde is voor de uitbouw van een datamodel en voor een betrouwbare en efficiënte exploitatie[23, 24, 31, 33]. Bij de ontwikkeling van een gecoördineerd en beleidsondersteunend 'healthcare information system', moet ervoor gezorgd worden dat zowel de informatie over de bestaande en beschikbare databanken, als informatie over data en metadata, als informatie over de modaliteiten voor de exploitatie van de data, gecentraliseerd wordt. Het in dit rapport geciteerde protocolakkoord rond de coördinatie van gezondheidszorg en gezondheidsgegevens zou nieuw leven moeten worden ingeblazen en gekaderd worden in een conceptueel data model.

België schiet vandaag tekort in rapportering naar internationale organisaties Een centrale coördinatie van alle gegevens en een expliciete opdracht, om via die centralisatie ook de internationale organisaties informatie te verstrekken, zal een nuttig werkinstrument zijn.

Er is een debat vereist over de wijze waarop de centralisatie van informatie over met publieke middelen gefinancierde registraties eventueel ondersteund kan worden door wettelijk opgelegde verplichtingen. Onderzocht moet worden hoe iedereen, die een register wenst op te zetten met publieke middelen, verplicht kan worden om gegevens over de databank en de metadata beschikbaar te stellen en actueel te houden, met inbegrip van een rapport over de historiek van een databank.

22 EEN GECOORDINEERD PROGRAMMA VOOR DISSEMINATIE VAN BETROUWBARE GEZONDHEIDSZORGGEGEVENS

België beschikt over een niet onbelangrijk potentieel van gezondheids(zorg)gegevens, maar ontbeert een kader waarbinnen deze gegevens, zowel voor beleidsdoeleinden als voor onderzoeksdoeleinden én voor internationale rapportering, veel meer en efficiënter geëxploiteerd kunnen worden. Sleutelvoorwaarde hiertoe is het bouwen aan een vlotte toegankelijkheid tot en disseminatie van betrouwbare gegevens.

Datakwaliteit is een essentiële voorwaarde om geloofwaardigheid op te bouwen en zo het vertrouwen te winnen van de verschillende type gebruikers van data. Gebruikers van met publieke gelden gerealiseerde databanken verwachten terecht relevante, volledige, betrouwbare en 'up to date' gegevens. Overigens zal, naarmate het gebruik van gegevens uit publieke databanken toeneemt, ook een debat moeten gevoerd worden over de verantwoordelijkheid en aansprakelijkheid in gevallen waar analyses het resultaat blijken te zijn van foute of niet kwaliteitsvolle gegevens.

Bij publiek gefinancierde databanken moet werk gemaakt worden van een professionele omkadering. Meer in het bijzonder moeten waarborgen van wetenschappelijke onafhankelijkheid, objectiviteit en neutraliteit m.b.t. disseminatie van data worden ingebouwd. Bij het ter beschikking stellen van data moet er verder voor gewaakt worden de belangen van zowel leveranciers en gebruikers van data te vrijwaren en tezelfdertijd het respecteren van de confidentialiteit van persoonsgegevens te garanderen.

23 HET BELANG VAN EEN UNIEKE GEPSEUDONIMISEERDE PERSOONSIDENTIFICATOR

De 'good practice' voorbeelden uit de ons omringende landen illustreren het belang van een unieke persoonsidentificator (UPI) met betrekking tot gezondheids(zorg)gegevens. Zowel primaire als secundaire gebruikers van gegevens hebben uit inhoudelijke en methodologische overwegingen nood aan een UPI. Op Europees niveau wordt verder zeer veel belang gehecht aan de 'electronic healthcare card' om redenen van mobiliteit van patiënten (vrij verkeer van EU-onderdanen), om redenen van administratieve vereenvoudiging en omwille van het koppelen van gegevens voor 'in depth' exploitaties. Ook in dit verband wordt gedacht aan een UPI.

De belangrijkste en meest betrouwbare 'matching variable' voor de koppeling van persoonlijke gezondheids(zorg)gegevens is een gecertificeerd unieke persoonsidentificator (UPI), die om redenen van confidentialiteit gepseudonimiseerd moet zijn (UPP).

Een UPI-UPP is een belangrijk middel om de koppeling van databanken exact uit te voeren en dus meer garanties te bieden voor de exploitatie van valide en betrouwbare gegevens. Om potentiële fouten hierbij te vermijden is een georganiseerde en doordachte werkwijze vereist.

De discussie over het gebruik van UPI-UPP moet gepaard gaan met maatregelen voor de bescherming van de persoonlijke levenssfeer in het bijzonder met betrekking tot gezondheidsgegevens, evenwel zonder dat deze beschermingsmaatregelen de exploitatie van databanken zou blokkeren. Dit is een delicate oefening is en moet met de nodige aandacht worden aangekaart. In bijlage 7 is een aanzet voor deze discussie uitgewerkt).

24 EEN GEÏNTEGREERD WETGEVEND KADER

In de bestudeerde landen (Denemarken, Schotland, Duitsland) gaat de uitbouw van een datawarehousemodel gepaard met een grondige reflectie over en het uitwerken van juridische randvoorwaarden. Ook in andere landen (Canada, Nieuw Zeeland, Ierland, Zweden, …) wordt gewezen op het belang van een wel doordacht en op dataverzameling en -exploitatie afgestemd juridisch kader.

België heeft ook tal van data-initiatieven op stapel staan, maar dreigt op dit ogenblik te weinig aandacht te besteden aan de coördinatie van verschillende initiatieven. We geven een aantal voorbeelden:

- Sectoraal comité gezondheids(zorg)gegevens (SCGG)
- Be-Health platform
- Electronic Health Record (EHR)
- Globaal Medisch Dossier (GMD)

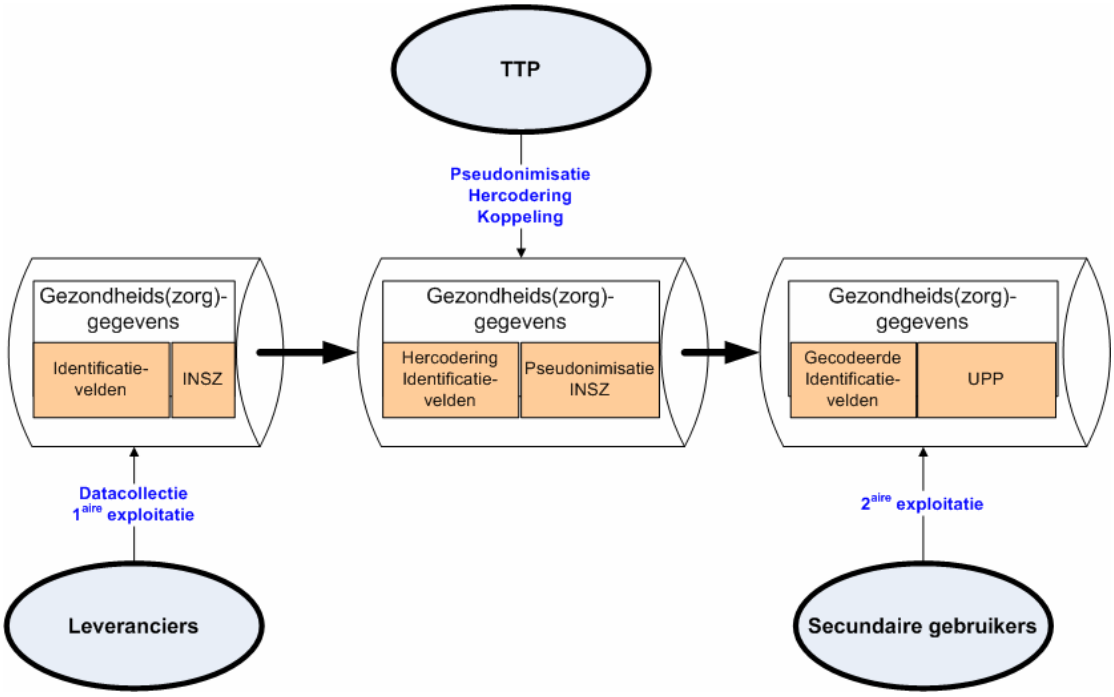
De indruk in het werkveld is dat de respectievelijke initiatieven onvoldoende gecoördineerd verlopen. Een geïntegreerd juridisch kader vraagt een discussie over verschillende wetgevingen in functie van databankmanagement. Naast het integreren van het wetgevend kader is er ook behoefte aan een geglobaliseerd uitvoerings-KB (zie bijlage 8).De buitenlandse voorbeelden leren dat een dergelijk kader garanties geeft om aan performante dataverzameling en efficiënte kwaliteitsbewaking te doen.

Voor de kwaliteitsbewaking en het opzetten van een juridisch kader kan onder andere inspiratie gevonden worden in het vandaag voorliggende wetsontwerp ter wijziging van de wet betreffende de openbare statistiek[35] (een belangrijke toepassing van overheidsdatabanken) dat streeft naar een oplossing voor de certificatie van statistieken en ook pleit voor een orgaan dat zal instaan voor de coördinatie.

25 HET UITTEKENEN VAN EEN NETWERKORGANISATIEMODEL

De ontwikkeling van een gezondheidszorgdatamodel moet vertaald worden in een management- en organisatiemodel. Het heeft geen enkele zin enkel wettelijke protocols uit te schrijven en voorwaarden vast te leggen, als ook geen organisatie- en managementstructuur wordt uitgebouwd, die zorgt voor een coördinatie, voor permanent overleg en participatie, en voor het opvolgen van afspraken tussen leveranciers van gegevens, primaire en secundaire gebruikers en een TTP. De kracht van een degelijk gezondheidszorginformatiesysteem ligt in de mogelijkheden die worden voorzien om gegevens uit uiteenlopende databanken te koppelen, in de plaats van een groot aantal individuele registraties op te zetten, met enerzijds veel redundante (en tijdrovende) informatieverzameling en anderzijds met heel wat hiaten. In ons huidige systeem is structureel geen enkel organisatorisch mechanisme aanwezig om de potentiële rijkdom aan gegevens rond gezondheidszorg op elkaar af te stemmen en te integreren.

Het uittekenen van het netwerkorganisatiemodel vereist voldoende aandacht voor de 'rollen' en aanwezige competenties van verschillende betrokken actoren als leveranciers, gebruikers van data en trusted third party. Deze 'driehoek' van relaties en rollen vraagt om mee opgenomen te worden in het organisatieconcept (zie figuur).



Het zou weinig efficiënt zijn om, louter op basis van een puur theoretisch concept, verantwoordelijkheden te definiëren, rollen vast te leggen en 'ownership' te bepalen. Gezien de huidige initiatieven en de specifieke bevoegdheidsverdelingen tussen de verschillende overheden is de optie van één organisatie, die de opdracht zou krijgen om alle datataken in de praktijk te gaan overnemen en implementeren, zeker niet wenselijk. Er is in België heel wat operationele competentie aanwezig.

Een netwerkorganisatiemodel houdt in dat bestaande instellingen, op een structurele en gecoördineerde manier, samen aan een gedeelde agenda werken.

Op dit ogenblik werken teveel instanties in zichzelf gekeerd en opereren ze in functie van hun eigen organisatiedoelstellingen. Er is geen enkel orgaan dat aanspoort om de eigen organisatiedoelstellingen en werkwijzen te overstijgen. Allicht ligt hier een verklaring waarom België het zo moeilijk heeft om degelijk internationaal te rapporteren.

In een netwerkorganisatiemodel wordt een samenwerking uitgeklaard tussen verschillende actoren en wordt een coördinerende rol toebedeeld aan een centrale instantie. Met het oog op een optimale werking en het vermijden van belangenconflicten, wordt dergelijke instantie het beste gecreëerd buiten het kader van de operationele taken van individuele ministeries, instellingen of organen.

Een centrale coördinatie-instantie krijgt de bevoegdheid om, in samenspraak met alle actoren, een conceptueel datamodel uit te tekenen en verder te ontwikkelen. De communicatielijnen met en tussen de respectieve instellingen worden daarom zeer expliciet uitgetekend. Degelijke instantie vraagt ook om een expliciet mandaat om in het verlengde van deze afspraken richtlijnen uit te vaardigen en af te dwingen, evenals certificaties voor kwaliteitsvol gegevensbeheer te verlenen. Daarenboven moet ze op een directe manier kunnen beschikken over de verzamelde data (en de metadata) van de moederorganisaties. Ze heeft ook een servicefunctie te vervullen naar de gebruikers en de overheid toe.

Voor de concrete coördinatie zijn verschillende organisatiemodellen mogelijk. De meeste buitenlandse voorbeelden die recent hervormingen hebben doorgevoerd kiezen resoluut voor een centraal en duidelijk identificeerbaar autonoom orgaan met precies omschreven opdrachten. België kan voor het organisatiemodel verder een analyse maken van ervaringen opgedaan met de opbouw van de Kruispuntbank sociale zekerheid. Die organisatie is tegelijk de interface voor overleg met secundaire gebruikers en heeft bij het ontwerp rekening gehouden met secundaire doelstellingen in het databaseontwerp. Er zou kunnen geopteerd worden voor een model kruispuntbank gezondheids(zorg)gegevens (KBGG), als draaischijf voor gegevensbanken met persoonlijke gezondheids(zorg)gegevens.

Gebaseerd op een aantal buitenlandse voorbeelden zou een centrale coördinatie- en certificatie-instantie (CCI) volgende wettelijk verankerde opdrachten moeten opnemen:

- Informatie verzamelen, analyseren en ter beschikking stellen over beschikbare databanken en registraties gezondheidszorg;
- Het uittekenen en coördineren van het conceptueel datamodel gezondheidszorg en het mandaat krijgen om, in samenspraak met de sector, alle dataregistraties op te bouwen conform dat conceptueel model;
- Informatie verzamelen en coördineren over de internationale ontwikkelingen en rapportagevereisten en het conceptueel datamodel daarop afstemmen;
- Het certifiëren van databanken en het accrediteren van gebruikers van databanken die zich aan de standaarden van goed gebruik van databanken moeten houden;
- Een competentiecentrum worden dat individuele organisaties of actoren ondersteunt bij de uitbouw of aanpassing van hun databanken;

- Ondersteuning bieden bij de ontwikkeling en implementatie van datakwaliteitsplannen en daarvoor methodologieën aan de betrokken actoren ter beschikking stellen;
- Het ondersteunen, coachen en coördineren van aanvragen waarbij een koppeling van gegevensbanken nodig is;
- Een strategie ontwikkelen m.b.t. de 'buy in' van voldoende computertechnische, specifieke juridische en medische competentie ('human capital').

Specifiek met betrekking tot persoonsgegevens zou een CCI volgende verantwoordelijkheden moeten opnemen:

- beëdigd leverancier zijn van de UPP's en hun eventuele correspondentietabellen;
- bewaarder zijn van de UPP's en hun eventuele correspondentietabellen (voor terugkeer naar de bron-UPI of enig andere gehercodeerde rechtspersoon);
- fungeren als intermediaire trusted third party (TPP) bij gemachtigde onderlinge koppeling en hercodering van gezondheidszorggegevens op basis van een UPP;
- instaan voor de datawarehousing van aldus ontstane nieuwe gegevensbanken.

26 REFERENTIES

1. Quality in Statistics. Standard quality indicators. 7th meeting of 'Quality in Statistics'. 2005.
2. APS. Kwaliteitszorg Statistisch productieproces – Aanbevelingen. http://aps.vlaanderen.be/statistiek/publicaties/pdf/kwaliteit/. 2003.
3. Eurostat. European statistics code of practice, Part on statistical output. 2005.
4. Eurostat. Results of Health Account Data in Europe - Final Report. 2004.
5. OECD Stats. Belgium - country statistical profiles 2006. 2006.
6. Bossuyt N TJ, Van Casteren V, Van Oyen H. Belgium: available health information. IPH. 2002.
7. Closon MC PJ. Statistische behoeften op het vlak van gezondheid en gezondheidszorg in België. CIES - UCL, HIVA - KUL. http://www.belspo.be/belspo/agora/doc/gezondheid.pdf . 2003.
8. CIHI. The CIHI Data Quality Framework. http://www.cihi.ca/cihiweb/en/downloads/Data_Quality_Framework_2004_e.pdf. 2005.
9. The WAVE Advisory Board to the Director-General of Health. From Strategy to Reality — WAVE (Working to Add Value through E-information). 2001.
10. Gauld R. One step forward, one step back? Restructuring, evolving policy, and information management and technology in the New Zealand health sector. Government Information Quarterly. 2004; 21:125–42.
11. Reng C, Debold P, Adelhard K, Pommerening K. Akzeptiertes Datenschutzkonzept. Deutsches Ärzteblatt. 2003; 33.
12. Roinn A, Leaneí S. Health Information – a national strategy. Dept. of Health and Children. 2004.
13. Sundgren B. Conceptual Design of Data Bases and Information Systems. Statistics Sweden. 1984.
14. Sundgren B. Conceptual Modelling as an Instrument for Formal Specification of Statistical Information Systems. ISI 47th Session. 1989.
15. Malmborg E, Sundgren B. Integration of statistical information systems - theory and practice Proceedings of the Seventh International Conference on Scientific and Statistical Database Management. University of Virginia. 1994.
16. Sundgren B. Documentation and Quality in Official Statistics. EU conference on Quality in Statistics. Stockholm, Sweden. 2001.
17. Upshur R, Goel V. The health care information directive. BMC Medical Informatics and Decision Making. 2001.
18. Black C, Burchill C A, L R. The Population Health Information System: data analysis and software. Med Care. 1995; 33(12 Suppl): 127-31.
19. Black C. Population-based information systems: how they can provide tools for decision-making and monitoring. http://www.aes.es/charlyn_black_18.pdf. 1998.
20. Roos N, Black C, Roos LL, Frohlich N, DeCoster C, Mustard C, et al. Managing health services: how the Population Health Information System (POPULIS) works for policymakers. Med Care. 1999; 37(6 Suppl): 27-41.

21. Berghman J. Eindrapport van het AGORA onderzoeksproject AG/FF/084: Aanvullende sociale voorzieningen in de tweede en derde pijler. http://www.belspo.be/belspo_off/home/publ/pub_ostc/agora/ragff084_nl.pdf 2005.

22. OECD. Quality framework for OECD statistics. 2002.

23. EuroStat. SDDS metadataserver. http://europa.eu.int/estatref/info/sdds/en/health/health_index.htm. 2004.

24. UN/ECE. Guidelines for the modeling of statistical data and metadata. Conference of European Statisticians. 1995; Geneva, 1995.

25. CIHI. CIHI - Taking Health Information Further. http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=profile_e . 2006.

26. Pommerening K. Medical Requirements for Data Protection. Institut für Medizinische Statistik und Dokumentation. http://www.staff.uni-mainz.de/pommeren/publInf.html. 1994.

27. Pommerening K. Pseudonyme - ein Kompromiß zwischen Anonymisierung und Personenbezug. In H. J. Trampisch and S. Lange, editors, Medizinische Forschung - ¨ Arztliches Handeln – 40. Jahrestagung der GMDS. 1995:329 – 33.

28. Pommerening K. Secondary Use of the EHR via Pseudonymisation. Medical Care Compunetics 1. 2004.

29. Pommerening K. Pseudonymisierung in der medizinischen Forschung - das generische TMF-Datenschutzkonzept. http://www.egms.de/pdf/journals/mibe/2005-1/mibe000017.pdf . 2005; 1(3): Doc 17.

30. Sundgren B. Metadata: A Quality Element in Official Statistics. Conference of the American Statistical Association. 1996; Chicago, U.S.A., 1996.

31. Sundgren B. The Swedish Statistical Metadata System. Eurostat conference. 2000; Luxembourg, 2000.

32. Sundgren B. Strategies for Development and Implementation of Statistical Metadata Systems. Proceedings of the ISI Conference. 2003; Berlin, 2003.

33. Sundgren B. Developing and implementing statistical metadata systems. EU project MetaNet. 2003.

34. Eppler MJ, Wittig D. Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years. Net Academy. 2000.

35. Belgian Chamber. Wetsontwerp tot wijziging van de wet van 4 juli 1962 betreffende de openbare statistiek. 2006.

36. Menezes A, Van Oorschot P, Vanstone S. Handbook of Applied Cryptography – Chapter 9 - Hash Functions and Data Integrity. CRC Press. 1996.

37. NHS. Draft Guidance on Handling Small Numbers (v1.14). National Services Scotland. 2005.

38. Washington State Department of Health. Guidelines for Working With Small Numbers. 2006.

39. Office for National Statistics ONS. Definitions Census 2001. 2004:20.

40. Codd E. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM. 1970; 13(6):377-87.

41. Even A, Shankaranarayanan G. The metadata enigma. Communications of the ACM. 2006; 48(2).

42. Newcombe H. Handbook of Record Linkage Methods for Health and Statistical Studies, Administration and Business. U.K. Oxford University Press. 1988:103-6.

43. Winkler W. The state of record linkage and current research problems, Statistical Research Division, Statistical Research Report Series No. RR99/04. Bureau of the Census. 1999.

44. Wajda A, Roos LL, Layafski M, Singleton JA. Record linkage strategies: Part II. Portable software and deterministic matching. Methods Inf Med. 2001; 30:210-4.

45. Faldum A, Pommerening K. Ein Algorithmus zur Erzeugung von pseudonymen Identifikatoren. 2001.

46. Faldum A, Pommerening K. An optimal code for patient identifiers. Computer Methods and Programs in Biomedicine. 2005; 79(81-88).

27 BIJLAGEN

BIJLAGE 1: GLOSSARIUM

1.1. CONCEPTEN EN DEFINITIES MET BETREKKING TOT GEGEVENS

Bestaande en nieuwe concepten en definities m.b.t. gegevens en gegevensbanken worden hieronder gebundeld.

1.1.1. M.b.t. intrinsieke kenmerken van de gegevens

- Onder "persoonsgegevens" wordt in de Wet van 08-12-1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens (PRW) verstaan: iedere informatie betreffende een geïdentificeerde of identificeerbare natuurlijke persoon.
    - Als *"identificeerbaar"* wordt beschouwd een persoon die direct of indirect kan worden herkend, met name aan de hand van een identificatienummer of van één of meer specifieke elementen die kenmerkend zijn voor zijn of haar fysieke, fysiologische, psychische, economische, culturele of sociale identiteit.
    - Een *"natuurlijke persoon"* is de juridische term voor een, al dan niet levende, fysieke entiteit.
    - Een *"rechtspersoon"* is de juridische term voor een niet noodzakelijk fysieke entiteit die kan optreden als een persoon in het rechtsverkeer.
- Onder "uniek identificatienummer" of "unieke persoonsidentificator" (UPI) wordt verstaan een code bestaande uit een combinatie van letter- en/of cijfertekens die een eenduidige identificatie van een fysieke persoon mogelijk maakt en gebruikt wordt in een strikt afgebakend domein van gegevensinzameling. Idealiter bevat zo'n code geen enkele persoonlijke informatie of werd deze "gepseudonimiseerd" (zie verder). In de context van gezondheidsgegevens is de UPI feitelijk het inschrijvingsnummer bij de sociale zekerheid (INSZ), dat dan wel persoonlijkje informatie bevat, met name de geboortedatum en geslacht van de betrokkene.
- Onder "gezondheidsgegevens" worden verstaan alle gegevens van persoonlijke aard waaruit informatie kan worden afgeleid omtrent de vroegere, huidige of toekomstige [a] fysieke of psychische gezondheidstoestand van de natuurlijke persoon die is of kan worden geïdentificeerd.
- "Sociale gegevens van persoonlijke aard": alle gegevens met betrekking tot de sociale zekerheidsstatus van een natuurlijke persoon die is of kan worden geïdentificeerd.

[a] Bedoeld wordt hier de prognostische gegevens.

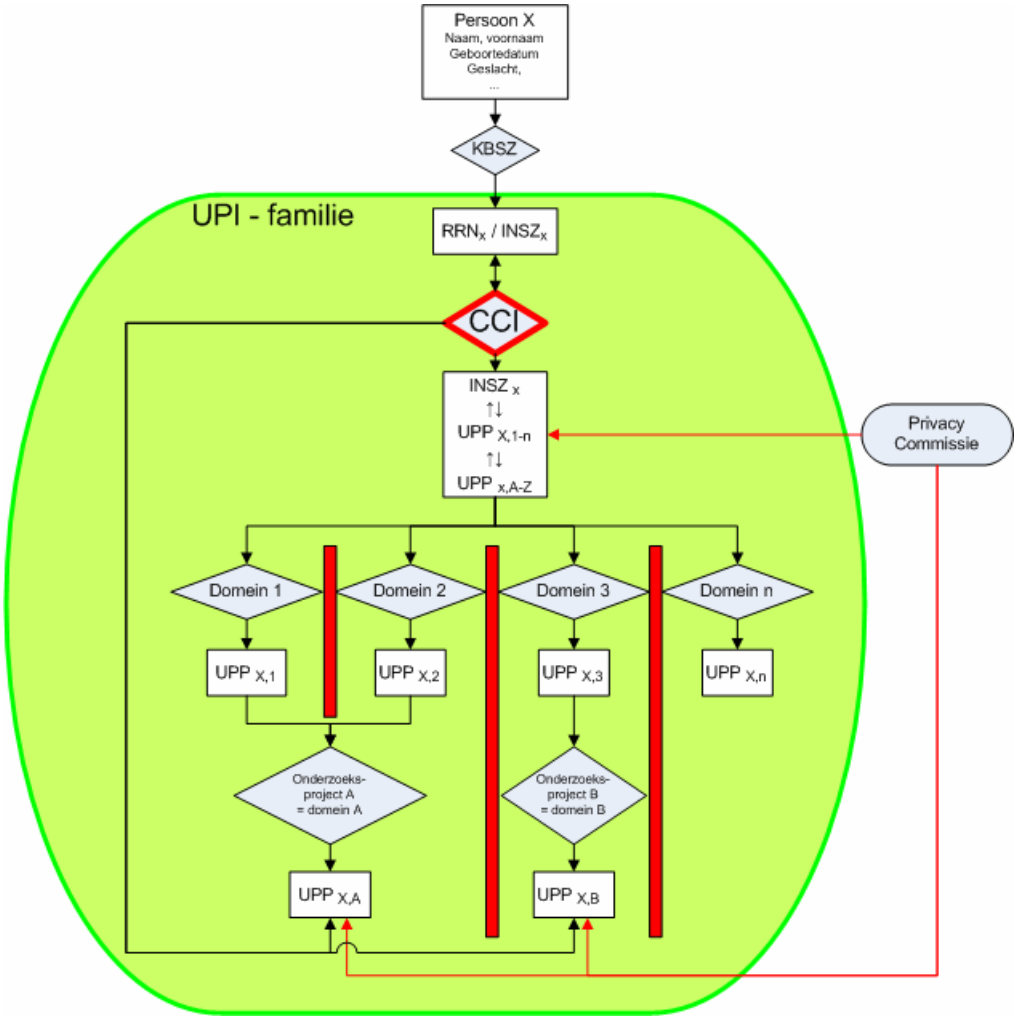
- "Identificatiegevoelige gegevens" zijn gegevens die, op zichzelf of in hun onderlinge samenhang, een bijzonder risico inhouden m.b.t. potentiële indirecte herkenning van hetzij fysieke personen (i.c. patiënten of zorgverstrekkers), hetzij rechtspersonen (i.c. verzorgingsinstellingen).
- "Operationele gegevens" zijn algemeen gesproken gegevens die de productiesystemen van een bedrijf of organisatie ondersteunen. Binnen de federale gezondheidszorg gaat het hier over bijvoorbeeld de financiële gegevens (MFG), de medische kenmerken (gecodeerde diagnoses en behandelingen) van de ziekenhuisverblijven (MKG), …

1.1.2. M.b.t. verwerking en transformatie van gegevens

- Onder "verwerking" wordt elke bewerking of elk geheel van bewerkingen al dan niet uitgevoerd met behulp van geautomatiseerde procédés. In de PRW wordt verweking van persoonsgegevens zeer breed gedefinieerd: het verzamelen, vastleggen, ordenen, bewaren, bijwerken, wijzigen, opvragen, raadplegen, gebruiken, verstrekken door middel van doorzending, verspreiden of op enigerlei andere wijze ter beschikking stellen, samenbrengen, met elkaar in verband brengen, alsmede het afschermen, uitwissen of vernietigen van persoonsgegevens.
- Onder "data-analyse" verstaat men het geheel van methoden toegepast om diverse observaties te beschrijven en hierin patronen te herkennen, verklaringen te vinden en hypothesen te toetsen.
- "Anonieme gegevens": gegevens die niet met een geïdentificeerd of identificeerbaar persoon of organisatie in verband kunnen worden gebracht (Wet van 12/08/2000 houdende sociale, budgettaire en andere bepalingen en Programmawet van 24/12/2002).
- "Gecodeerde persoonsgegevens" zijn gegevens die slechts door middel van een code of combinatie van codes in verband kunnen worden gebracht met een geïdentificeerd of identificeerbaar persoon of specifieke kenmerken ervan. De Europese wetgeving (EU Commission Regulation No. 831/2002 – 17/05/2002) spreekt hier van "confidentiële gegevens": gegevens die slechts een indirecte identificatie toelaten. "Geanonimiseerde microgegevens" worden hier gedefinieerd als individuele statistische observaties die aangepast werden om, volgens de huidige beste praktijk, het risico op identificatie te minimaliseren van de natuurlijke of rechtspersonen naar wie zij verwijzen.
- "Coderen" is het proces waardoor informatie over een gegeven of eenheid vertaald wordt in een code. Het doel van coderen is onder meer efficiëntere dataopslag, encryptie of versluiering, beheer of analyse.
- "Hercodering" is het transformeren van een code of set van codes in een andere code, (met behoud van een controleerbare omkeerbaarheid). In een meer restrictieve context wordt hier bedoeld: de omzetting van een algemeen gebruikte code in een

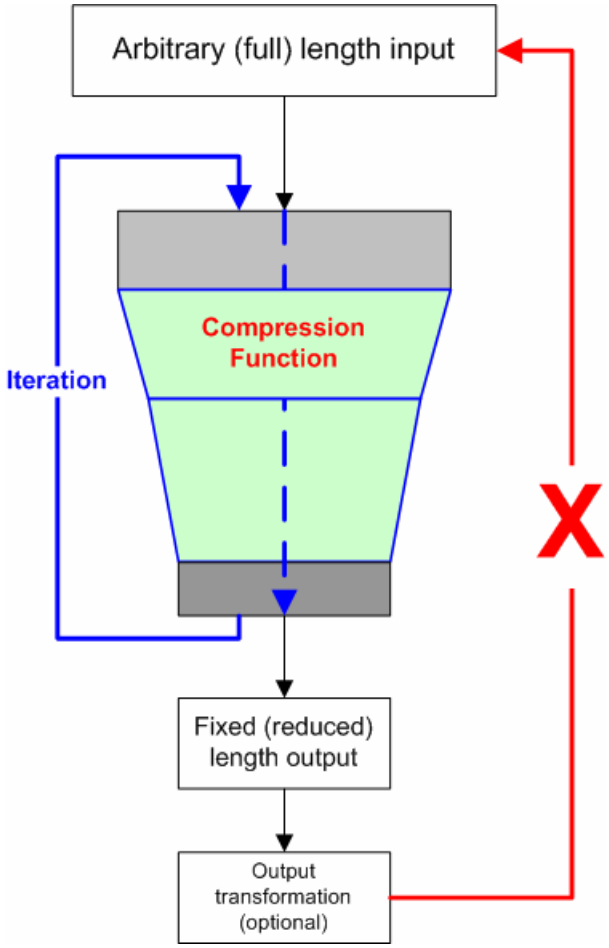
andere, projectspecifieke code. Voorbeeld: omzetting van het publiek toegankelijke erkenningsnummer van de ziekenhuizen tot een niet direct herkenbaar, zgn. 'anoniem' ziekenhuisnummer.

- "Pseudonimisatie" is een specifieke vorm van hercoderen, nl. het transformeren van een uniek maar publiek identificatienummer van een persoon bv. het rijksregister nummer (RRN) of het inschrijvingsnummer bij de sociale zekerheid (INSZ) of een set van specifieke elementen die kenmerkend zijn voor de fysieke of sociale identiteit van deze persoon (zoals bv. leeftijd, geslacht, nationaliteit, wettige domicilie, woonplaats, sociale zekerheidsregime, etc…) door een al dan niet numerieke code, met het doel directe identificatie onmogelijk te maken maar toch de uniciteit van de persoon te behouden. Afhankelijk van de finaliteit kan deze transformatie irreversibel, dan wel reversibel zijn (bv. i.v.m. longitudinaliteit, koppeling,…). Pseudonimisatie sluit weliswaar directe persoonsidentifiactie uit, maar kan op zichzelf nooit het probleem van de indirecte, contextuele herkenning oplossen (zie bijlage 8).
- Een gepseudonimiseerde UPI wordt "uniek patiëntpseudoniem" (UPP) genoemd.
- Een "UPI-familie" omvat het geheel van alle domeinspecifieke UPP's behorende tot eenzelfde UPI (figuur 1). Voor een meer conceptuele bespreking verwijzen we naar de bijlage 7.



- **Figuur 1 : het concept 'UPI-familie'**

- "Hashing" is een gegevensbewerking in de informatica die een gegeven met een willekeurige lengte omzet in een binaire reeks met vaste lengte[36] (figuur 2). Een goede hashfunctie is er een die weinig botsingen veroorzaakt in het gegevensdomein waarmee ze werkt, dit wil zeggen dat er weinig kans is dat twee verschillende invoerwaarden dezelfde uitvoer geven ('non collision'). Daarnaast moet het quasi onmogelijk zijn de oorspronkelijke waarde terug te vinden en dus op basis van de output de originele input af te leiden (one-way transformation). Het resultaat van de hashing, ook wel "hashingstring" genoemd, zorgt ervoor dat de informatie beschermd wordt tegen frauduleus 'geknoei' met gegevens.



- **Figuur 2 : principe van hashing - naar Menezes**

- "Encryptie" is het proces waardoor informatie dermate getransformeerd (versleuteld) wordt dat ze volkomen onleesbaar wordt zonder speciale kennis van de gebruikte werkwijze. Bij dit proces wordt evenwel de totaliteit van de broninformatie niet gereduceerd (wat wel het geval is bij hashing). Oorspronkelijk bestond encryptie uit het hercoderen van de informatie met een geheime codeerlijst[b]. Later kwamen algoritmes in zwang die toelieten het encryptieproces te automatiseren. Het algoritme omhelst een reeks van transformatiestappen die in een vooraf vastgelegde procedure op de broninformatie toegepast worden en gestuurd worden door een specifieke "encryptiesleutel". Deze is geheim voor niet ingewijden. Pas na toepassen van de omgekeerde procedure ("decryptie") wordt de informatie in zijn totaliteit weer 'beschikbaar' voor de gebruiker die over de sleutel beschikt.

- Er bestaan twee klassen van algoritmes: de symmetrische waarbij dezelfde sleutel gebruikt wordt voor encryptie en decryptie en de asymmetrische waarbij een verschillende (publieke en private) sleutel gebruikt worden. De encryptiesleutels en het algoritme dienen dusdanig te worden gekozen dat het 'breken' ervan praktisch onmogelijk is, rekening houdend met de technische mogelijkheden van de bestaande informaticasystemen. Encrypteringsalgoritmes dienen dan ook constant mee te evolueren

[b] Zie: Barbara W. Tuchman. The Zimmerman telegram. ISBN: 0026203200.

met de evolutie van de computerrekenkracht en de cryptografische wetenschap.

- "Gegevensanonimisatie" is het geheel van processen waardoor gegevens dermate getransformeerd worden dat ze niet meer met een geïdentificeerd of identificeerbaar fysieke persoon of rechtspersoon in verband kunnen worden gebracht. Om dit te bereiken worden een of meerdere van de volgende methodes, al dan niet in combinatie of sequentieel, toegepast op de identificatiegevoelige gegevensvelden :
    - codering,
    - hercodering,
    - pseudonimisatie,
    - versluiering.
- "Gegevensversluiering" is het geheel van technieken die leiden tot een reductie van de informatie in individuele gegevensvelden met de bedoeling het identificatierisico aan de hand van gevoelige gegevens te beperken tot het niveau van anonieme gegevens of, bij onmogelijkheid hiertoe, tot een aanvaardbaar, door de Privacy Commissie vooraf opgelegd niveau. Hierbij mag geenszins de finaliteit van de gegevensverwerking in het gedrang komen. Versluiering kan bekomen worden door:
    - gegevensreductie door de precisie te verminderen, bijv. datumwaarden enkel in YYYY-MM formaat i.p.v. YYYY-MM-DD, ICD-9-CM codes in "three digit format" i.p.v de volledige codes in 4 of 5 cijfers,
    - gegevenseliminatie door verwijdering of niet mededeling van sommige gegevensvelden,
    - gegevensaggregatie van het niveau van individuele entiteiten tot dit van groepsentiteiten,
    - gecontroleerde gegevenswijziging ("barnardisation" of "random perturbation") in groepsentiteiten met kleine aantallen[37, 38],
    - gecontroleerde "record swapping" in kleine groepsentiteiten[39].

1.1.3. M.b.t. het niveau van de gegevens

Micro gegevens

Gegevens op het laagste operationele niveau. Hiertoe behoren de facto persoonsgebonden gegevens van patiënten of van individuele verstrekkers (natuurlijke personen).

Meso gegevens

Gegevens op een tussenniveau (bijv. groepen, afdelingen, ziekenhuizen). Deze gegevens verwijzen naar rechtspersonen of groepen van natuurlijke personen.

Macro gegevens

Gegevens gegroepeerd op een algemeen, hoger niveau (bijv. regio's, nationale gegevens, Europese gegevens,…)

Geaggregeerde gegevens

Geaggregeerde gegevens zijn het resultaat van samenvoeging of combinatie van gegevens (van een lager niveau)

Niet-geaggregeerde gegevens

Met niet-geaggregeerde gegevens worden bedoeld gegevens op hun oorspronkelijk niveau van inzameling.

1.2. CONCEPTEN EN DEFINITIES MET BETREKKING TOT GEGEVENSBESTANDEN OF –BANKEN

In deze sectie worden de elementen gedefinieerd die determinerend zijn voor de structurele samenhang van gegevens in het geheel van een gegevensbestand of een gegevensbank. Verder worden een aantal essentiële kernbegrippen m.b.t. deze laatste uitgewerkt.

Databank / gegevensbank / database / dataregister

Een databank is een verzameling van onderling samenhangende gegevens, georganiseerd in een logische structuur (het datamodel), zodat de toegang optimaal is voor de inzameling, het lezen, de wijziging en de verwijdering van deze gegevens. Een databank wordt beheerd en opgeslagen als een eenheid.

Relationele database

Een database die gestructureerd is volgens het relationele model[40]. De gegevens zijn geordend in verschillende tabellen (relaties) die middels gemeenschappelijke indices met elkaar in verband worden gebracht.

De relationele database is het bouwblok van het hart van moderne informatiesystemen: het datawarehouse. Ze wordt beheerd met een relational database management system (RDBMS), dat helpt om na te gaan of aan de voorwaarden van het relationele model is voldaan: uniciteit, coherentie en integriteit van de gegevens van de gedefinieerde relaties. Ontwerpers van RDBMS bieden een implementatie aan van de databasetaal SQL (structured query language) voor bewerkingen met de relationele database.

Dataset / gegevensbestand

Een dataset is een georganiseerde verzameling van samenhangende gegevens. Een dataset bestaat typisch uit records die op hun beurt bestaan uit de individuele waarden voor de set van variabelen (velden) in de dataset. Een dataset (ook wel tabel genoemd) kan primair of secundair zijn als resultaat van een bevraging op één enkele of een combinatie van primaire datasets.

Art. 1, §3 van de PRW definieert een "bestand" als "elk gestructureerd geheel van persoonsgegevens die volgens bepaalde criteria toegankelijk zijn, ongeacht of dit geheel gecentraliseerd dan wel gedecentraliseerd is of verspreid op een functioneel of geografisch bepaalde wijze".

Gegevensrecord of gegevensrij

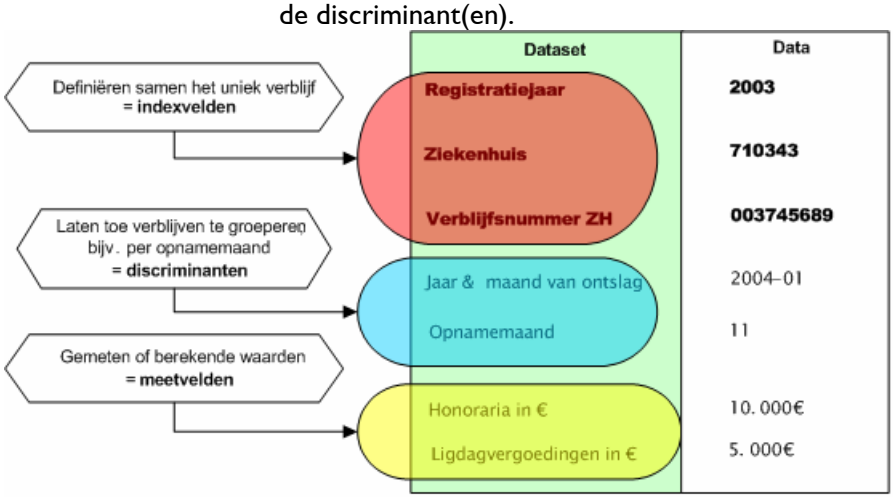
Het geheel van gegevens in een dataset die verwijst naar een entiteit. In een relationele database is gegevensrij de uitdrukking van de relatie tussen de variabelen in de dataset, voor een zelfde entiteit.

Gegevensveld (of variabele)

Een specifiek bedoeld gegevenselement als onderdeel van het gegevensbestand, rechtstreeks geregistreerd of door bewerking tot stand gekomen (figuur 3).

Types van gegevensvelden zijn:

- '*Indexveld(en)*' : één of meerdere gegevensvelden die samen de *index* vormen die de uniciteit van een record in de (oorspronkelijke) dataset bepaalt. Indexvelden zijn kenmerkend voor relationele databases.
- *'Discriminant'* : gegevensveld in een dataset dat gebruikt wordt om observaties onderling te onderscheiden. Alle indexvelden zijn discriminanten.
- Een *primaire of hoofddiscriminant* is die discriminant waarvoor in eerste instantie de informatie wordt opgeslagen in meetvelden van de dataset. Gewoonlijk is er maar één, wat niet uitsluit dat er in complexe databases meerdere hoofddiscriminanten kunnen opgenomen zijn.
- De set van *secundaire of nevendiscriminanten* geeft aan op welk niveau van toewijzing de informatie gemeten werd. Secundaire discriminanten laten dan ook toe de informatie naderhand te (her)groeperen. De primaire finaliteit van de registratie bepaalt of een veld al dan niet 'discriminatoir' is en of het een hoofd- dan wel nevendiscriminant is.
- *'Meetveld'* : veld dat een kwalitatief of kwantitatief gemeten of een berekende grootheid bevat in relatie tot de discriminant(en).



- **Figuur 3 : voorbeeld van indexvelden, discriminanten en meetvelden**

Primaire gegevensbanken

"Eerste orde" gegevensbanken die gecreëerd werden voor de inzameling en bundeling van gegevens met een specifieke administratieve of wetenschappelijke finaliteit.

Secundaire gegevensbanken

Een zelfstandige "afgeleide" databank, waarvan de gegevens geconstrueerd zijn door een selectie, hercodering, wijziging van gegevens, de samenvoeging van gelijksoortige data of de koppeling van samenhorende data afkomstig van een of meerdere primaire databanken.

Gekoppelde gegevensbanken

Een secundaire databank die tot eenzelfde entiteit behorende informatie uit verschillende primaire en/of secundaire gegevensbanken samenbrengt en integreert en op deze wijze de informatie die inherent aanwezig was in de verschillende databanken expliciteert.

Correspondentietabellen

Specifieke secundaire gegevensbanken die apart geconstrueerd, bewaard en beheerd worden met het doel de latere conversie van gehercodeerde of gepseudonimiseerde gegevensvelden terug om te zetten naar de oorspronkelijke codering (decodering of depseudonimisering); ook wel "conversietabellen" of 'codeer-decodeerlijsten" genoemd. De opmaak ervan is een exclusief voorrecht van de intermediaire organisatie (TTP).

Referentielijsten (codeerlijsten of "look-up tables)

Gestructureerde lijst met codes voor registratie-elementen in databanken m.b.t. de gezondheidszorg, samengesteld met als doel informatie te ordenen en naderhand te kunnen verwerken (bijv. nomenclatuur, ICD-9-CM classifcatie, APR-DRG lijst, RIZIV-nomenclatuur, lijst van farmaceutische specialiteiten, ATC-codering, ziekenhuiserkenningsnummers, etc.).

Metadata

Metadata zijn gegevens (data) die de karakteristieken van andere gegevens beschrijven, ofwel 'data over data'. Meer algemeen spreekt men ook van meta-informatie. Meta-informatie is informatie die er toe bijdraagt dat gegevens als informatie kunnen geïnterpreteerd worden. Naar functionaliteit worden hierbij 6 types van metadata onderscheiden[41]:

- Gegevens m.b.t. de hardware- en softwareconfiguratie gebruikt voor de gegevensverwerking *('infrastructure metadata')*. Deze informatie is vooral belangrijk voor 'system maintance'.
- Documentatie van de conceptuele, logische én fysieke structuur van de database, de definities van en de onderlinge verbanden tussen de variabelen evenals de relaties tussen datasets binnen de database *('model metadata')*.
- De beschrijving van de gebruikte methodologie van datacollectie, datavalidatie en ETL ('extraction,

transformation & loading' – zie verder) van data, evenals het eigenaarschap van databanken *('process metadata')* .

- Informatie over kwaliteitaspecten van de data *('quality metadata')*.
- Gegevens over toegankelijkheid en dissiminatie van gegevens met informatie over aard en inhoud van de geleverde gegevensextracten *('interface metadata')*.

- Tot slot ook het bijhouden van een register van gemachtigde gebruikers van de data met informatie over het opzet en de finaliteit van het onderzoek én de toegestane looptijd ervan. Kortom: wie heeft toegang tot welke gegevens, onder welke vorm, hoelang en om welke (gewettigde) reden(en)? Al deze elementen vormen de *'administration metadata'*.

Met uitzondering van de eerste catagorie, zijn dit grosso modo de elementen verzameld in de diverse annexen van dit rapport.

Longitudinaliteit

Verwijst naar de temporele eigenschap van de dataverzameling waarbij observaties behorende bij eenzelfde entiteit, maar geregistreerd op verschillende tijdstippen, intrinsiek met elkaar gekoppeld kunnen worden. Bij longitunale databanken kunnen karakteristieken van entiteiten in de loop van de tijd gevolgd worden.

1.3. CONCEPTEN EN DEFINITIES MET BETREKKING TOT GEGEVENSPROCESSEN

Het samenstellen van een databank omvat een aantal stadia of processen. Om goed te begrijpen hoe een databank gecreëerd wordt en welke mogelijkheden er zijn om een databank te exploiteren voor onderzoeksdoeleinden, is het nuttig verschillende stadia van het "gegevensproces" analytisch te onderscheiden. Op deze manier wordt het ook mogelijk een "assessment" te maken die verwijst naar de verschillende stadia van het gegevensproces.

1.3.1. Verzameling

Het proces van identificatie van databronnen, het definiëren (eventueel met behulp van classificaties) van variabelen en het inzamelen van data die verband houden met de gezondheidszorg.

1.3.2. Extraction, Transformation en Loading (ETL) van data

Het geautomatiseerde proces van het onttrekken (E) van ruwe data, transformeren (T) en gestandaardiseerd inladen (L) van de gegevens in een databank. De extractie dient om de gegevens te lezen uit bronbestanden in diverse formaten of uit andere gegevensbanken. Transformatie zorgt voor het omvormen van deze gegevens in het juiste formaat (tekst, numeriek, aantal cijfers of tekens, …), voor de (her)codering van de gegevens in de gewenste classificatie of codering, of voor de aanmaak van de nodige (secundaire) variabelen uit de geëxtraheerde gegevens. De laatste stap is het inladen van deze getransformeerde gegevens op de juiste plaats in de logische structuur van de databank.

1.3.3. Validatie en controle

Het verifiëren van de inhoud en of het formaat van de gegevensinput overeenstemt met de definities van het databankontwerp.

1.3.4. Exploitatie

Primaire exploitatie

Proces waarin de gegevens gebruikt worden in functie van de primaire doelstelling van de databank.

Secundaire exploitatie

Proces waarin de gegevens gebruikt worden in functie van afgeleide doelstellingen die niet noodzakelijk direct verband houden met de primaire doelstellingen van de databank. Hierbij is de wijze waarop externe (geaccrediteerde) gebruikers toegang krijgen tot en gebruik kunnen maken van de gegevens een belangrijk aandachtspunt.

1.3.5. Koppeling van gegevensbanken

Bepaalde zinvolle gegevens voor onderzoeksvragen zijn het resultaat van het koppelen van data uit verschillende databanken. Daarom wordt in dit rapport ook stilgestaan bij de organisatorische mogelijkheden en problemen van het koppelen van gegevensbanken.

- In de meest algemene zin is *"koppeling"* het samenvoegen van records van twee of meer registraties waarvan men aanneemt dat ze naar dezelfde entiteit verwijzen (een entiteit is afhankelijk van de context, maar kan zowel verwijzen naar een natuurlijke persoon, (patiënt of zorgverlener) als naar een rechtspersoon, (ziekenhuis, …). Het doel van koppeling kan ook zijn om duplicaten uit een of meerdere databanken te verwijderen[42, 43]. In dit geval zijn alle velden van de records inhoudelijk van gelijke waarde.
- Om records te kunnen samenvoegen, moeten er overeenkomsten bestaan tussen één of meerdere velden uit deze records. Blijken de records bij paarsgewijze vergelijking van deze velden tot dezelfde entiteit te behoren, dan spreekt men in het Engels van een *"match"*.
- Deterministische overeenkomst (*"deterministic matching"*) verwijst naar de directe methode waarin een set van individuele velden paarsgewijs worden vergeleken om te bepalen of men te doen heeft met een koppeling ("link") of niet ("non-link")[44]. Een concreet voorbeeld zou de koppeling zijn op basis van een unieke patiëntidentificator (UPI), die in beide registraties, elk met hun eigen UPP voorkomt. Behoudens incorrecte inputgegevens bij de creatie van de UPP, zullen alle "links" ook werkelijk naar dezelfde entiteit verwijzen.
- Probabilistische overeenkomst ("probabilistic matching") verwijst naar een methodiek die gebruik maakt van conditionele kansrekening om de waarschijnlijkheid te begroten dat een recordpaar naar dezelfde entiteit verwijst of juist niet. De uitkomst van deze methodiek is gebaseerd op gewichten die bepaald worden uit de vergelijking van de koppelingsvelden. Het resultaat is een

waarde die de kans uitdrukt dat de koppeling naar eenzelfde entiteit verwijst.

1.3.6. Datawarehousing & datawarehouse

Datawarehousing is het geheel van processen leidend tot de aanmaak, onderhoud en exploitatie van een datawarehouse.

Een datawarehouse is een omvangrijke secundaire databank met informatie over de bedrijfsverrichtingen en activiteiten. De brongegevens worden op regelmatige tijdstippen opgevraagd in heterogene databanken en na ETL-stappen en koppeling samengebracht in het datawarehouse. Het datawarehouse zorgt voor efficiënte data-analyse en rapportering, dikwijls uitgewerkt tot een Management Information System. Het is dus een werktuig om gegevens te transformeren tot (al dan niet geaggregeerde) informatie volgens relevante dimensies voor beleidsondersteuning, onderzoek, planning, controle en algemeen beheer van het bedrijf.

In de context van voorliggend rapport kan 'bedrijf' verruimd worden tot de gehele sector van de gezondheidszorg of zelfs van de sociale zekerheid in zijn geheel.

BIJLAGE 2: LIJST MET AFKORTINGEN

AG	Algemeen geneeskundige
APS	Administratie Planning & Statistiek - Vlaamse Gemeenschap
AR	Auditoriaatsrapport van het Sectoraal Comité Sociale Zekerheid
AVZ	Anonieme Ziekenhuis Verblijven = SHA (Fr) Synoniem : MFG = Minimale Financiële gegevens
AZ	Algemeen ziekenhuis = niet-psychiatrisch / acuut of chronisch
BCFI	Belgisch Centrum voor Farmacotherapeutische Informatie
BDMS	Banque de Données Médico-Sociales
BE	NUTS-niveau België
BE1	NUTS-niveau Brussel Hoofstedelijk Gewest
BE2	NUTS-niveau Vlaanderen
BE3	NUTS-niveau Wallonië
BGDA	Brusselse Gewestelijke Dienst voor Arbeidsbemiddeling (= ORBEM)
BisR	Bis-Register of Register van de KSZ.
BMGS	Bundesministerium für Gesundheit - Germany
BZ	(FOD) Binnenlandse zaken
CBIP	Centre Belge d'Information Pharmacothérapeutique
CCI	Coördinatie & Certificatie Instantie
CDV	Centrum voor dagverzorging
CHI	Community Health Index
CHPs	Community Health Partnerships (CHPs) – Scotland
CIC	Centraal Instellingen bestand – fichier des Institutions Centralisé (FOD VG – SPF SP)
CIHI	Canadian Institute for Health Information
CIRCA	Comunication & Information Resource Centre Administrator - Forum for information on Statistical Metadata
CIV	Centrum voor Informatie Verwerking van de FOD Volksgezondheid
CNK	Uniek nummer voor farmaceutica, tot op niveau verpakking
CPR	Centralt PersonRegister - Danmark
CSJ	Centre de soins de jour
CSSS	Comité Sectoriel Sécurité Sociale
D&D	Vergoede deliverables & disposables
D&T	Diagnose & Therapie specificaties
DFMA	Multifunctionele aangifte (KBSZ)
DG SANCO	Directorate General for Health and Consumer Affairs
DGV	Dienst voor Geneeskundige Verzorging - RIZIV
DSM	Diagnostic and Statistical Manual of Mental Disorders

DWHA	DataWarehouse Arbeidsmarkt
DWTC	Federale diensten voor wetenschappelijke, technische en culturele aangelegenheden
ECHP	European Community Household Panel
EHR	Electronic Health Record = electronisch GMD
EPB	Equity & Patient Behaviour
EPR	Electronic Patient Record
ERS	Electronic Reporting System
ETL	Extraction, Transformation and Loading (of data)
EVA	Elektronische Verbinding tussen de verzekeringsondernemingen en het FAO
EZ	(FOD) Economische zaken
FAO	Fonds voor Arbeidsongevallen
FBZ	Fonds voor Beroepsziekten
FOD	Federale OverheidsDienst
FOREM	Office Communautaire et Régional de la Formation Professionnelle et de l' Emploi
FSO	Federal Statistical Office - Germany
FTE	Full time equivalent
G(Z)G	Gezondheids(zorg)gegevens
GBE	Gesundheitsberichterstattung des Bundes - Germany
GENESIS	Gemeinsamen Neuen Statistischen Informations-System - Germany
GMD	Globaal medisch dossier
GP	General Practitioner
GPASS	General Practitioner Administration System for Scotland
H&E	2x hashing-encryptie = irreversibel
HC	Health Care = gezondheidszorg
HCD	Healthcare data = gezondheids(zorg)gegevens
HC-DWH	HealthCare DataWareHouse
HC-mDWH	HealthCare metaDataWareHouse
HCSR	Health Care Services Research
HIS	Healthcare Information System = informatiesysteem m.b.t. G(Z)G
HL7	Health Level 7 = internationaal communicatieprocol voor transmissie G(Z)G
HP	Hôpitaux psychiatriques
HSR	Health (Care) Services Research
HTA	Health Technology Assessment
HZV	Hulp- en Zorgkas voor zeeVarenden
IBW	Initiatief beschut wonen = IHP

ICD-9-CM	International Classification of Diseases, 9th revision, Clinical Modification (WHO)
ICPC2	International Classification of Primary Care, 2nd edition
IHP	Initiative habitation protégéé = IBW
IKAROS	geIntegreerd Kind Activiteiten Regio OndersteuningsSysteem
IMA	Intermutualistisch Agentschap
IMS	Information Management System(s)
INEK	Institut für das Entgeltsystem im Krankenhaus - Germany
In-patient	Gehospitaliseerde patiënt
IPH	Institute for Public Health
ISDS	Information & Statistics Division Scotland
IS-GBE	Informationssystem der Gesundheitsberichterstattung des Bundes - Germany
ISP	Information Security Policy = globaal plan m.b.t. beveiliging van bedrijfsinformatie
ISP	Institut de Santé Publique
IWEPS	Institut wallon de l'évaluation, de la prospective et de la statistique
Katz	Afhankelijkheidscategorie volgens de schaal van Katz
KBO	Kruispuntbank Ondernemingen
KSZ	Kruispuntbank van de Sociale Zekerheid
KUL	Katholieke Universiteit Leuven
LHP	Local Health Plans - Scotland
MAST	Medical Antishock Trousers
MFG	Minimale financiële gegevens = AZV-SHA
MG	Médecin généraliste
MM	Maand in 2 cijfers, mois en deux positions
MRPA	Maison de repos personnes agées
MS	Multiple Sclerose
MSP	Maison de soins psychiatriques
NACE	Nomenclature générale des Activités économiques dans les Communautés Européennes
NAPHSIS	National Association for Public Health Statistics and Information Systems
NBH	National Board of Health - Danmark
NBP	Niet bruggepensioneerden
N-groep	Aggregatieniveaus van codenummers, die overeenstemmen met de onderverdelingen van de nomenclatuur, zoals bijvoorbeeld de tandheelkunde, anesthesiologie, medische beeldvorming, enz.
NHA	Nationale Health Accounts
NHS	National Health Service (UK)
NIC	Nationaal Intermutualistisch College
NIS	Nationaal Instituut voor Statistiek; thans Algemene Directie Statistiek en

	Economische Informatie van de FOD Economie
NISS	Numéro d'inscription à la sécurité sociale
NISZ	Nationaal inschrijvingsnummer sociale zekerheid
NLM	National Library of Medicine (USA)
NMR	NucleoMagnetische Resonantie
NNR	Nationaal Nummer Rijksregister
NRN	Numéro au régistre national
NSS	National Services Scotland
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Co-operation and Development
ONE	Office de la naissance et de l'enface
ONS	Office of National Statistics - UK
OTC	'Over the counter' = niet voorgeschreven medicatie
OTD	Office de tarification – tarifcatiedienst (Pharmanet)
Out-patient	Eendagsverblijf of ambulante patiënt
PAAZ	Psychiatrische afdeling in een algemeen ziekenhuis
PAC	Privacy Advisory Committee (Scotland)
PET	Positron Emissie Tomografie
POD	Programmatorische OverheidsDienst
POD MI	Programmatorische OverheidsDienst Maatschappelijke Integratie
PPP	Part Payed by Patient (out of pocket) = persoonlijk aandeel patiënt
PRW	Privacy Wet
PSBH	Panel Study of Begian Households
PTD	Pharmanet tarificatiedienst
PTI	Practice Team Information – Scotland
PTL	Personen ten laste
PTO	Pharmanet Tarification Offices
PUG	Primair uitkeringsgerechtigde
PVT	Psychiatrisch verzorgingstehuis
RDBMS	Relationeel Database Management Systeem
RevC	Revalidatiecentra
RIZIV	RijksInstituut voor Ziekte- en InvaliditeitsVerzekering
RKI	Robert Koch Institute – Germany
RKW	Rijksdienst voor Kinderbijslag van Werknemers
ROB	Rustoord bejaarden
RR	RijksRegister
RRN	Rijksregisternummer
RSVZ	Rijksinstituut voor de Sociale Verzekeringen der Zelfstandigen

RSZ	Rijksdienst voor Sociale Zekerheid
RSZPPO	Rijksdienst voor Sociale Zekerheid van Provinciale en Plaatselijke Overheidsdiensten
RT	Recordtype
RUG	Rijksuniversiteit Gent
RVA	Rijksdienst voor Arbeidsvoorziening
RVP	Rijksdienst voor Pensioenen
RVT	Rust- en verzorgingstehuis
SBA	Statistisches Bundesamt - Germany
SCGG	Sectoraal Comité Gezondheids(zorg)gegevens
SCI	Scottish Care Information
SCSS	Sectoral Comittee Social Security
SCSZ	Sectoraal Comité Sociale Zekerheid (SCSZ)
SEDH	Scottish Executive Health Department
SEE	Sociaal-Economische Enquête (de 'volkstelling')
SES	Socio-economische status
SFF	Sickness Fund Federation = ziekenfondsfederatie = landsbond = VI
SHA	Séjours Hospitaliers Anonymes = AZV Synoniem : RFM = Résumé Financier Minimum
SHA	System of Health Accounts (OECD)
SHARE	Survey on Health Ageing Retirement Europe
SHI	Statutory Health Insurance
SHOW	Scotland's Health on the Web
SILC	Survey on Income and Living Conditions
SPE	Studiecentrum voor Perinatale Epidemiologie
SPF	Service Public Fédéral
SPHG	Services psychiatriques en hôpitaux généraux
SPP	Service Public de Programmation
SZR	Sociale ZekerheidsRegime van de gerechtigde (in België) = codes gerechtigde 1 & 2
SZS	Sociale zekerheidsstatus (= meer gedetailleerd dan SSR)
TMF	Telematikplattform für Medizinische Forschungsnetze
UA	Universiteit Antwerpen
UCL	Université Catholique de Louvain
ULB	Université Libre de Bruxelles
ULg	Université de Liège
VDAB	Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding
VI	Verzekeringsinstelling = landsbond van mutualiteiten
VTR	Verhoogde tegemoetkoming rechthebbenden

WHO	World Health Organization
WIGW	Weduwen, Invaliden, Gepensioneerden en Wezen
WIV	Wetenschappelijk Instituut voor Volkdgezondheid
YYYY	Jaartal in 4 cijfers
YYYYS	Semester in het jaar YYYY
YYYYT	Trimester in jaar YYYY
ZIV	Ziekte- en Invalididteitsverzekering

BIJLAGE 3: LIJST VAN BEZOCHTE BELGISCHE INTERNET-PORTALEN

http://www.belgium.be/eportal/

https://www.socialsecurity.be (FOD Sociale zekerheid)

http://health.fgov.be (FOD Volksgezondheid)

http://ksz.fgov.be (Kruispuntbank Sociale Zekerheid)

http://www.rijksregister.fgov.be (Rijksregister)

http://statbel.fgov.be/ (Nationaal instituut voor de statistiek)

http://www.belspo.be (Federaal Wetenschapsbeleid)

http://www.iph.fgov.be (WIV/ISP/IPH - afdeling Epidemilogie)

http://www.vlaanderen.be (welzijn en gezondheid)

http://www.wallonie.be (Portail de la Région wallonne)

http://guide.wallonie.be/ (Action sociale et santé)

http://www.cfwb.be (Communauté Française)

http://www.brussel.irisnet.be (Hoofdstedelijk Gewest Brussel)

http://www.ccf.irisnet.be (COCOF)

http://www.vgc.be (Vlaamse Gemeenschapscommissie Brussel)

http://www.belgielex.be (Kruispuntbank Wetgeving).

BIJLAGE 4: VERKENNENDE VRAGENLIJST

1. Bestaan er in uw organisatie gegevensbanken met gezondheids- en/of gezondheidszorggegevens? Zo ja, waarom worden deze gegegevens ingezameld en wat is hun doel (1aire finaliteit)?
2. Welke documenten zijn er beschikbaar die de inhoud nauwkeurig beschrijven?
    a. Waar of bij wie zijn deze documenten beschikbaar? Graag de contactgegevens of URL.
    b. Indien er geen documenten beschikbaar zijn, gelieve dan kort de inhoud van de ingezamelde gegevens te beschrijven.
3. Wat zijn de bronnen van de gegevens die verzameld worden (organisaties, zorgverstrekkers, 'upstream' databanken,...)?
4. Welke documenten zijn er beschikbaar die het dataproces beschrijven (van datacollectie tot exploitatie)?
    a. Waar of bij wie zijn deze documenten beschikbaar? Graag de contactgegevens of URL.
    b. Indien er geen documenten zijn, geef dan een korte beschrijving van het dataproces.
5. Wat wordt gedaan om de datakwaliteit te meten en op te volgen (controle, validatie,...)?

BIJLAGE 5: URL'S VAN BEZOCHTE STATISTISCHE REFERENTIECENTRA

STATISTISCHE REFERENTIECENTRA :

http://unstats.un.org/unsd/methods.htm (United Nations Statistics)

http://www.who.int/healthinfo/statistics (WHO)

http://www.oecd.org/statsportal (OECD[c]- Statistical Information System - SIS)

http://epp.eurostat.cec.eu.int/portal (Eurostat)

http://europa.eu.int/comm/eurostat/ramon/index.cfm (Eurostat's RAMON [d] metadata server)

http://forum.europa.eu.int (EU CIRCA[e] workspace)

http://www.statcan.ca/english/concepts (Statistics Canada)

http://secure.cihi.ca/cihiweb (Canadian Institute for Health Information – CIHI)

http://www.hhs.gov/healthit/ (Office of the National Coordinator for Health Information Technology – ONC, USA)

http://www.cdc.gov/nchs (CDC, National Center for Health Statistics, USA)

http://www.doh.wa.gov/Data/Guidelines (Washington State Dep. of Health – Health data guidelines)

http://www.va.gov/OIT (Office of Information & Technology - Veterans Affairs, USA)

http://www.cihi.ca The Canadian Institute for Health Information

http://www.isdscotland.org/isd (Scottish Health Statistics[f])

http://www.naphsis.org (NAPHSIS[g])

http://aps.vlaanderen.be/ (Administratie Planning & Statistiek - Ministerie Vlaamse Gemeenschap)

LITERATUUR OVER DATAMANAGEMENT:

http://www.scienceresearch.com/search/index.php (ScienceResearch.com™ portal)

http://www.scirus.com/srsapp/ (Science-specific search engine)

http://www.researchindex.com/ (topics: data quality, data warehousing systems, etc…)

http://www.scicentral.com/ (Gateway to the scientific research news sources)

http://www.topicarticles.com/data-management/ (Top Articles Data Management)

---

[c] Organisation for Economic Co-operation and Development
[d] Vroeger CODED
[e] Comunication & Information Resource Centre Administrator - Forum for information on Statistical Metadata.
[f] National Health Service Scotland
[g] National Association for Public Health Statistics and Information Systems, USA

CRYPTOGRAPHY & PRIVACY:

http://www.epic.org/ (Electronic Privacy Information Center, USA)

http://www.cyphernet.org/info/info.php (Cyphernet)

http://www.staff.uni-mainz.de/pommeren/ (Institut für Medizinische Biometrie, Epidemiologie und Informatik – IMBEI, DBR)

BIJLAGE 6 : OVERZICHTSTABEL INVENTARIS DATABANKEN

Bron	Database	Classificatie		Primaire finaliteit		Bereik			Tijdsaspecten		
		Hoofddomein	Subdomein(en)	Kerngegevens	Hoofddiscriminant(en)	Nuts	Doelpopulatie	Zorglijnen	Start	Continuïteit Periodiciteit	Stop
FOD VG	MKG-RCM	Health Care Consumption	Utilization, D&T specifications	Klinisch	Niet-psychiatrische ZH	BE	alle verblijven	intramuraal	1995	continu jaarlijks	-
FOD VG	MVG-RIM	Health Care Consumption	Utilization, D&T specifications	Verpleegkundig	Niet-psychiatrische ZH	BE	alle niet-psychiatrische verblijven	intramuraal	1988	discontinu 20 dd. per jaar	-
FOD VG	MPG-RPM	Health Care Consumption	Utilization, D&T specifications	Klinisch-psychiatrisch	PZ, PAAZ	BE	alle psychiatrische verblijven	intramuraal	1996	continu jaarlijks	-
FOD VG	MUG-SMUR	Health Care Consumption	Utilization, D&T specifications	Klinisch Administratief	MUG-interventies	BE	alle MUG-interventies	intra- & extra-muraal	okt. 2003	continu semesterieel	-
FOD VG	FINHOSTA	Health Care Production Health Care Outcome	Productivity Human resources	Boekhouding Jaarrekening Personeelsdata	Zorginstellingen	BE	alle ziekenhuizen	N.v.t.	1988	continu jaarlijks	-
FOD VG	Jaarlijkse ziekenhuisstatistieken	Health Care Outcome Health Care Production	Productivity Human resources	Infrastructuur & uitrusting Organisatie & informatica Personeelsdata	Zorginstellingen	BE	alle ziekenhuizen	N.v.t.	2000	continu jaarlijks	-
FOD VG	Centraal instellingen bestand CIC	Health Care Production	Resources Technology , Infrastructure Human capital (P&O), Specific healthcare programs	Diverse erkenningen & zorgactiviteiten Contactinfo & organisatie	Zorginstellingen	BE	volledig	N.v.t.	2006?	continu jaarlijks	-
FOD VG	Centraal bestand gezondheidsberoepen	Health Care Production	Resources Human capital	Beroepspersonalia	Zorgverstrekkers	BE	erkende beoefenaars	N.v.t.	2006?	continu 'real time'	-
RIZIV	Doc. C	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	Boekhoudcodes RIZIV	BE	alle verplicht verzekerden	alle zorglijnen	1964	continu jaarlijks	2006

Bron	Database	Classificatie		Primaire finaliteit		Bereik			Tijdsaspecten		
		Hoofddomein	Subdomein(en)	Kerngegevens	Hoofddiscriminant(en)	Nuts	Doelpopulatie	Zorglijnen	Start	Continuïteit Periodiciteit	Stop
RIZIV	Doc. N	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	RIZIV-nomenclatuur	BE	alle verplicht verzekerden	alle zorglijnen	1964	continu jaarlijks	2006
RIZIV	Doc. N^{bis}	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	RIZIV-nomenclatuur Status sociale zekerheid	BE	alle verplicht verzekerden	alle zorglijnen	2006	continu jaarlijks	-
RIZIV	Doc. P	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	Zorgverleners RIZIV-nomenclatuur	BE	alle verplicht verzekerden	alle zorglijnen	1964	continu semesterieel	-
RIZIV	Doc. H	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	Zorginstellingen Verpleegdagcodes	BE	alle verplicht verzekerden	intramuraal	1994	continu semesterieel	-
RIZIV	Doc. FH	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	Ziekenhuizen Forfaitcodes	BE	alle verplicht verzekerden	intramuraal extramuraal	1994	continu semesterieel	-
RIZIV	Doc. PH	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering Persoonlijk aandeel patiënt	Ziekenhuizen Farmaceutica	BE	alle verplicht verzekerden	intramuraal	1994	continu semesterieel	-
RIZIV	Doc. FM	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	Revalidatiecentra, ROB, RVT, PVT & IBW Forfaitaire vergoedingen	BE	alle verplicht verzekerden	intramuraal transmuraal (RevC)	1994	continu semesterieel	juli 2002
RIZIV	Statistieken ROB-RVT-CDV	Health Care Production Health Care Consumption	Utilization Human resources	Aantal vergoede verblijfsdagen Personeelsdata	ROB, RVT, CDV verpleegdagen Personeel	BE	alle verplicht verzekerden	intramuraal transmuraal (CDV)	2004	continu trimesterieel	-
RIZIV	MFG-RFM AZV-SHA	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering	Ziekenhuisverblijven	BE	alle verplicht verzekerden	intramuraal	1995	continu jaarlijks	-
RIZIV	STATMD	Health Care Consumption	Production	Facturatie-bestanden	Zorginstellingen VI's	BE	alle verplicht verzekerden	intramuraal	1994	continu maandelijks	-
RIZIV	PAT	Health Care Consumption	Utilization Practice volume	Aantal patiënten (patiënteel)	Algemeen geneeskundigen	BE	alle verplicht verzekerden	extramuraal	1994	continu semesterieel	-

Bron	Database	Classificatie		Primaire finaliteit		Bereik			Tijdsaspecten		
		Hoofddomein	Subdomein(en)	Kerngegevens	Hoofddiscriminant(en)	Nuts	Doelpopulatie	Zorglijnen	Start	Continuïteit Periodiciteit	Stop
RIZIV	Pharmanet 1e generatie	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering Persoonlijk aandeel patiënt	Farmaceutica Voorschrijver	BE	alle verplicht verzekerden	extramuraal	1996	continu semesterieel	2004
RIZIV	Pharmanet 2e generatie bestand PH_PRS	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering Persoonlijk aandeel patiënt	Farmaceutica Voorschrijver Verzekerden	BE	alle verplicht verzekerden	extramuraal	2004	continu semesterieel	-
RIZIV	Pharmanet 2e generatie bestand PH_POP	Population data	Socio-economic status	Sociale zekerheidsregime	Verzekerden	BE	alle verplicht verzekerden	extramuraal	2004	continu semesterieel	-
TCT	MKG-MFG	Health Care Consumption	Utilization, D&T specifications Expenditure	APRDRG's Uitgaven verplichte ziekteverzekering	Verblijven	BE	volledig	intramuraal	1996	continu jaarlijks	-
IMA (7 VI's)	IMA Gezondheidszorgen	Health Care Consumption	Expenditure	Uitgaven verplichte ziekteverzekering Persoonlijk aandeel patiënt	Verzekerden Voorschrijver-verstrekker	BE	alle verplicht verzekerden	alle zorglijnen	2002	continu jaarlijks	-
IMA (7 VI's)	IMA Populatiebestand	Population data	Socio-economic status (SES)	SZ status & rechten	Verzekerden	BE		N.v.t.	2002	continu semesterieel	-
Fod BZ	Rijksregister RR	Population data	Demographics	RRN civiele persoonsgegevens	Inwoners België	BE	volledig	N.v.t.	1983	continu dagelijks	-
KBSZ	Rijksregister-bis INSZ	Population data	Demographics	Administratief	Ingeschrevenen SZ	BE	volledig	N.v.t.	1990	continu wekelijks	-
KBSZ	Datawarehouse arbeidsmarkt DWA	Population data	Socio-economic status Social security status	Arbeids- & sociale zekerheidsstatuut	Ingeschrevenen SZ (INSZ)	BE	volledig	N.v.t.	1997	continu jaarlijks	-
WIV-ISP	Huisartsenpeilpraktijken	Health status Health Care	Morbidity, epidemiology Utilization, outcome	Klinisch-epidemiologisch	Pathologie (thema)	BE	sample	Extramuraal	1979	continu één jaar	-
WIV-ISP	Gezondheidsenquête Health Interview Survey	Health status Healt care	Morbidities, lifestyles	Population Health survey	Inwoners België	BE	sample	N.v.t.	1997	discontinu 4-jaarlijks	-

Bron	Database	Classificatie		Primaire finaliteit		Bereik			Tijdsaspecten		
		Hoofddomein	Subdomein(en)	Kerngegevens	Hoofddiscriminant(en)	Nuts	Doelpopulatie	Zorglijnen	Start	Continuïteit Periodiciteit	Stop
		consumption Populaton data	Utilization, expenditure Socio-economic status, income								
NIS	Federale overlijdensregister	Health status	Mortality causes	Administratief Doodsoorzaken	Inwoners België	BE	volledig	N.v.t.	(1948) 1991	continu jaarlijks 1997 laatste jaar	-
NIS	Sociaal Economische Survey	Health Population data	Heath determinants, environment Socio-economic status, income	'Volkstelling'	Inwoners België	BE	volledig	N.v.t.	(1846) 1991	Discontinu ±10-jaarlijks	-
NIS	Panel Study on Belgian Households PSBH	Health status Population data	Lifestyles & environment Socio-economic status	Socio-economisch	Gezinnen	BE	sample	N.v.t.	1992	discontinu jaarlijks	2002
NIS	Huishoudbudget Onderzoek HBO	Health Care Consumption Population data	Expenditure Socio-economic status	Socio-economisch	Gezinnen	BE	sample	N.v.t.	1995	discontinu jaarlijks	-
Eurostat	EU-Survey on Income & Living Conditions SILC	Health determinants Population data	Lifestyles & environment Socio-economic status	Socio-economisch	Gezinnen	EU	sample	N.v.t.	2005	discontinu jaarlijks	-
Eurostat	Survey on Health Ageing & Retirement Europe SHARE	Health Population data	Disabilities, lifestyle Socio-economic status	Socio-economisch	Personen > 50 jr. + partner	EU	sample	N.v.t.	2004	continu jaarlijks	-
APS	Vlaamse overlijdensregister	Health status	Mortality causes	Administratief Doodsoorzaken	Inwoners Vlaanderen, Brussel Hoofdstedelijk gewest	BE1 BE2	volledig	N.v.t.	1998	continu jaarlijks 2003 laatste jaar	-
IWEPS	Waalse overlijdensregister	Health status	Mortality causes	Administratief Doodsoorzaken	Inwoners Waalse gewest	BE3	volledig	N.v.t.	1998	continu jaarlijks 1999 laatste	-

Bron	Database	Classificatie		Primaire finaliteit		Bereik			Tijdsaspecten		
		Hoofddomein	Subdomein(en)	Kerngegevens	Hoofddiscriminant(en)	Nuts	Doelpopulatie	Zorglijnen	Start	Continuïteit Periodiciteit	Stop
										jaar	
SKR	Kankerregister	Health Health Care	Morbidity Utilization, D&T specifications	Personalia Klinische gegevens D&T	Kankerpatiënt	BE	onvolledig	Intramuraal	(1983) 1997	continu jaarlijks	-
Kind & gezin	IKAROS	Health status Health Care Consumption Population data	D&T specifications Utilization Demographics, SES	D&T Epidemiologisch Administratief	Kinderen tot 3 jaar	BE2	onvolledig	alle zorglijnen	1999	continu 'real time'	-
ONE	Banque de Données Médico-Sociales BDMS	Health status Health Care Consumption Population data	D&T specifications Utilization Demographics, SES	Administratief Epidemiologisch	Zwangeren, neonati, KK tot 3 jr.	BE1 BE3	onvolledig	alle zorglijnen	1994	continu jaarlijks	-
SPE	Studiecentrum Perinatale Epidemiologie	Health status Health Care Consumption Population data	D&T specifications Utilization Demographics, SES	Obstetrische en neonatale gegevens moeder & kind	Zwangeren & neonati	BE1 BE2	onvolledig	alle zorglijnen	(1987) 1996	continu jaarlijks	-
VRGT + FARES	Tuberculoseregister	Health Health Care	Morbidity Utilization, D&T specifications	Personalia Klinische gegevens D&T	TBC-patiënten	BE	volledig	N.v.t.	1995	continu jaarlijks	-
Vlaams zorgfonds	Vlaamse Zorgverzekering	Health Care	Utilization	Administratief Rechten & gebruik	Vlaamse gemeenschap	BE1 BE2	onvolledig volledig	extramuraal transmuraal	2001	continu jaarlijks	-
Diabetes fonds	Diabetesregister BDR	Health Health Care	Morbidity Utilization, D&T specifications	Personalia Klinische gegevens D&T	Diabetespatiënt Eerstegraadsverwanten <40 jr.	BE	onvolledig	N.v.t.	1989	continu jaarlijks	-
BVSM-ABLM	Mucoviscidoseregister	Health Health Care	Morbidity Utilization, D&T specifications	Personalia Klinische gegevens D&T	Mucoviscidose patiënten	BE	onvolledig	N.v.t.	1998	continu jaarlijks	-

BIJLAGE 7: CONCEPTEN M.B.T. UPP'S EN HUN OPERATIONALISATIE

7.1 VOORWAARDEN VOOR EEN 'BEST PRACTICE' UPP

Een 'best practice' UPP zou aan volgende voorwaarden moeten voldoen:

1. een gegarandeerd unieke persoonsidentificatie mogelijk maken, m.a.w. vertrekken van een publiek en uniek persoonlijk nummer (UPI) zoals het RRN of het INSZ;
2. domeinspecifiek c.q. projectspecifiek zijn, m.a.w. slechts gebruikt worden in een strikt afgebakend domein van gegevensverwerking met respect voor longitudinaliteit en desgewenst terugkoppelbaarheid; het begrip domein wordt hierbij contextueel én strategisch bepaald (zie punt 7.2)
3. volledig gepseudonimiseerd[27, 36, 28, 29] zijn (UPP), m.a.w. mag geen enkele persoonlijke informatie bevatten waaruit de identiteit van de betrokkene zou kunnen afgeleid worden. 'Hashing' alleen voldoet niet, omdat de hashingstrings eventueel kunnen gereconstrueerd worden door ieder die over de basisnummers (RRN of INSZ, beide publiek domein) beschikt én over het gehanteerde hashingalgoritme (bijv. RIPEMD-160), dat vrij gemakkelijk terug te vinden is;
4. bovendien zo geëncrypteerd zijn dat hij bestendig is tegen niet gerechtigd 'kraken' van de code met behulp van de bestaande informaticahulpmiddelen;
5. afhankelijk van de operationele doelstellingen van het registratiedomein moet een UPP koppelbaar zijn met de UPP van andere registraties. Dit impliceert een strikt voorwaardelijke terugkoppelbaarheid van de UPP door exclusieve en beveiligde bewaring, door een bevoegde centrale instantie, van conversietabellen die een terugkeer naar het onder punt 1. vermelde persoonsnummer (UPI) mogelijk maakt;
6. in zijn eindvorm voldoende eenvoud van output vertonen zodat fouten te wijten aan de complexiteit van de resulterende tekenstring geminimaliseerd; deze laatste mag m.a.w. niet te lang, noch te kort bepaald worden;
7. ruimte laten voor foutdetectie met behulp van ingesloten 'check digits'[h] , wat mogelijkheden biedt tot directe validatie ('integrity check') en zelfs in zekere mate ook tot foutcorrectie[45, 46] ;
8. afgeleverd worden door een bevoegde en onafhankelijke 'trusted third party' (TTP), als intermediare organisatie tussen de leveranciers van de brondata en de uiteindelijke verwerkers van de gekoppelde gegevens ;
9. evolutief zijn op het vlak van procedures, hashing- en encryptiestandaarden in functie van technologische vooruitgang maar met de garantie dat longitudinale analyses mogelijk moeten blijven.

[h] Cf. het controlegetal van het RRN.

7.2 DOMEIN- EN PROJECTSPECIFICITEIT

Strategisch zijn twee dimensies ('assen') cruciaal bij het invoeren van UPP's voor databanken met persoonlijke gezondheids(zorg)gegevens: enerzijds het exploitatieniveau en anderzijds de domein- en projectspecificiteit.

1. Exploitatieniveau :

   Primaire exploitatie verwijst naar het operationele, administratieve gebruik van databanken. Secundaire exploitatie verwijst naar afgeleide beleidsondersteunende analyses en wetenschappelijk onderzoek. Meestal gaat het hier om zowel in gegevensomvang als in tijd duidelijk afgebakende onderzoeksprojecten, elk met hun eigen, specifieke finaliteit(en).

2. Domein- en projectspecificiteit :

   Domeinspecificiteit is een noodzakelijke dimensie omdat zij in eerste instantie te maken heeft met het inschatten van de risico's op het doorbreken van de privacy van het individu én het medisch beroepsgeheim. Hierbij gaat het niet zozeer om het gevaar voor directe identificatie (die door de pseudonimisatie uitgesloten is), maar vooral om de risico's van indirecte, contextuele herkenning. Hoe meer persoonlijke gezondheids(zorg)gegevens samengebracht worden in omvangrijker databanken, m.a.w. hoe meer individuele detailgegevens bijeengebracht worden, hoe groter deze risico's worden. De enige oplossing voor dit fundamentele probleem is een rigoureuze afbakening van domeinen m.b.t. persoonlijke gegevensbanken, die elk hun eigen, van overheidswege opgelegde UPP moeten krijgen.

   Voor de afbakening van domeinen m.b.t. gezondheids(zorg)gegevens kan een functionele classificatie gebruikt worden, waarbij onderscheid wordt gemaakt tussen 4 types van domeinen: lokale domeinen, samenwerkingsdomeinen, koepeldomeinen en publieke domeinen.

   a. *Lokale domeinen* worden gedefinieerd per afzonderlijke operationele entiteit op het terrein van gezondheids- of welzijnszorg, bijv. een zorginstelling, een medische- of paramedische praktijk (solo- of groepspraktijk), een lokaal OCMW-bestuur, etc. Elk van deze lokale domeinen heeft het recht in volle vrijheid zijn eigen UPI of UPP te gebruiken, ook voor secundaire exploitaties in eigen beheer. Als enige voorwaarde moet hier gesteld worden dat de gebruikte UPI/UPP zou beantwoorden aan opgelegde kwaliteitstandaarden.

   b. *Samenwerkingsdomeinen* betreffen situaties waarbij meerdere afzonderlijke zorgverstrekkers beslissen operationeel samen te werken, hetzij lokaal, hetzij regionaal; bijv. een wachtdienst van regionale huisartsen, die een gemeenschappelijk wachtdienstregister aanhouden; samenwerking tussen ziekenhuizen met het oog op een primaire exploitatie van hun gegevens. Zodra hierbij persoonlijke gezondheids(zorg)gegevens worden samengevoegd, wat in se het informed consent van de patiënt in zijn oorspronkelijke lokale domein doorbreekt (arts, ziekenhuis,…), moet hier de voorwaarde gelden van een officiëel geregistreerde UPP, die dan voor primaire en secundaire exploitaties in eigen beheer mag gebruikt worden. Worden echter gegevens voor secundaire exploitatie overgemaakt aan externe onderzoeksgroepen, dan dienen deze tot het publieke domein gerekend te worden en behoeven ze een nieuwe, eigen UPP (zie verder: projectspecificieit).

c. *Koepeldomeinen* betreffen supraregionale organisaties van ziekenhuizen behorend tot eenzelfde 'zuil' (VVI, VOZ,…) of de landsbonden van de ziekenfondsen (VI's), die elk een apart koepeldomein beslaan, of het IMA dat de 7 VI's overkoepelt. M.b.t. UPP's in primaire en secundaire exploitatie gelden hier dezelfde basisregels als uitgewerkt onder b). De enige reden tot het maken van het onderscheid betreft de schaalvergroting die koepeldomeinen met zich meebrengen, wat o.i. reden genoeg is om ze semantisch te onderscheiden van samenwerkingsdomeinen.

d. *Publieke domeinen* tot slot kaderen zich in de klassieke openbare gezondheidszorgsector, geregeld en gecontroleerd door de bevoegde federale en/of ondergeschikte overheden. Alle secundaire exploitaties, aangevraagd door lokale domeinen, samenwerkingsdomeinen of koepeldomeinen dienen tot het publieke domein gerekend te worden, zodra deze exploitaties de transmissie van persoonlijke gezondheids(zorg)gegevens naar externe onderzoeksgroepen inhouden.

Voor dit secundaire exploitatieniveau moet het begrip 'domeinspecificiteit' vernauwd worden tot '*projectspecificiteit*'. Voor secundaire exploitaties geldt immers de door de PrW opgelegde eis tot restrictie van de benodigde gegevens tot deze relevant voor en proportioneel met de gepredefinieerde finaliteit van het project. Daarenboven moeten de geëxtraheerde gegevens geïsoleerd worden van de 'moederdatabank', zodat niet gerechtigde samenvoeging met andere extracties hieruit uitgesloten wordt en een impliciete koppeling 'sui generis' onmogelijk wordt.

Rekening houdend met het feit dat misbruik, zelfs institutioneel, altijd mogelijk is, moet een strategische afbakening van domeinen m.b.t. databanken met persoonsgegevens, elk met hun eigen, van overheidswege opgelegde UPP, centraal staan in het conceptueel model. Gezond verstand en bestuurlijke wijsheid moeten hier aanzetten tot een voorzichtigheidshalve afzweren van het idee één enkele UPP voor alle gecentraliseerde registraties van gezondheids(zorg)gegevens in te voeren.

Daarenboven moet het oog van de regulerende overheid ook gericht blijven op nakende evoluties in de geneeskunde (zoals de moleculair-genetische diagnostiek). De stap naar digitalisatie en opslag van individuele genetische informatie zal vlug gezet worden. De vraag rijst of het verstandig is degelijke gegevens toe te laten tot een allesomvattend domein van persoonlijke gezondheidszorggegevens (het 'electronic healthcare cardnumber', zoals voorzien in het wetsontwerp m.b.t. het Be-Health platform en de medische telematica). Degelijke extra beveiliging, zoals een stricte domeinafbakening (aparte domeinspecificiteit) voor dergelijke gegevensbanken lijkt hier op zijn plaats, aangevuld met een expliciete informed consent m.b.t. toegang tot deze gegevens voor onderzoeksdoeleinden.

Aan de hand van de hier besproken 'strategische assen' en rekening houdend met het noodzakelijke behoud van exploiteerbaarheid van de uiteindelijke gegevensverzamelingen, kunnen aldus functionele domeinafbakeningen gemaakt worden, elk met hun aangepaste UPP-vereisten (zie tabel).

<table>
<tr><th colspan="3">Afbakening van domeinniveaus bij verwerking van persoonlijke gezondheidsgegevens en de daaraan verbonden UPI / UPP vereisten</th></tr>
<tr><th></th><th>1air domein = primaire exploitatie</th><th>2air domein = secundaire exploitatie</th></tr>
<tr><td rowspan="2">Lokaal domein :
zorginstelling,
artsenpraktijk,
OCMW bestuur</td><td colspan="2">Eigen UPI*</td></tr>
<tr><td colspan="2">Voorbeelden:
- (Operationele) databank patiënteel individueel ZH
- (Operationele) databank patiënteel solo- of groepspraktijk
- Lokaal research project in eigen beheer</td></tr>
<tr><td rowspan="2">Samenwerkingsdomein :
regionale groepering van ZH,
regionale HA wachtdienst</td><td colspan="2">Domeinspecifieke UPP**</td></tr>
<tr><td>Voorbeelden:
- (Operationele) databank v.e. samenwerkingsverband v.e. aantal afzonderlijke ZH
- (Operationele) databank v.e. huis-artsenwachtdienst</td><td>Voorbeelden:
- Studieprojecten in eigen beheer op databankextracties uit de samengestelde databank van meerdere afzonderlijke ZH
- Researchproject in eigen beheer op gegevens v.e. samengestelde databank van meerdere afzonderlijke ZH</td></tr>
<tr><td rowspan="2">Koepeldomein :
VVI, VOZ, …
landsbonden VI's</td><td colspan="2">Domeinspecifieke UPP</td></tr>
<tr><td>Voorbeelden:
- Operationele databank v.e. koepelorganisatie van ZH
- Operationele databank van een koepel-VI (landsbonden)</td><td>Voorbeelden:
- Studieprojecten in eigen beheer op databankextracties uit de databank v.e. koepelorganisatie ZH
- Studieprojecten in eigen beheer op databankextracties uit de databank v.e. koepel-VI
- Studieprojecten in eigen beheer op databankextracties van het IMA</td></tr>
<tr><td rowspan="2">Publieke domein :
Gereguleerd door de overheid, federaal of communautair</td><td>Domeinspecifieke UPP</td><td>Projectspecifieke UPP</td></tr>
<tr><td>Voorbeelden:
- (Administratieve) databanken van de Federale overheid
- (Administratieve) databanken v.e.Communautaire overheid
- (Administratieve) databanken v.h. RIZIV
- Nationale morbiditeitregisters
- Nationaal kankerregister
- Nationaal TBC-register
- Nationale registers van zorgprogramma's
- Netwerk EHR (Elektronisch Medisch Dossier)
- 'DNA-databank' (in de toekomst)</td><td>Voorbeelden:
- Beleidsvoorbereidende of wetenschappelijke studieprojecten op databankextracties uit het publieke domein, een koepeldomein of een samenwerkingsdomein
- Beleidsvoorbereidende of wetenschappelijke studieprojecten op extracties van gekoppelde databanken uit het publieke domein, een koepel-domein of een samenwerkingsdomein</td></tr>
<tr><td colspan="3">* Eigen UPI/UPP's vereisen geen TTP-tussenkomst; voor domein- en projectspecifieke UPP's is dit wel het geval.
** Verschillende UPP's per domein en per project.</td></tr>
</table>

Voor de 1aire en 2aire exploitatieniveaus, met uitzondering van die op strict lokale domeinen, moeten andere accenten gelegd worden m.b.t. registratie en transmissie en de voorwaarden waaraan de exploitatie minimaal moet voldoen (zie tabel).

Basisvoorwaarden voor transmissie & exploitatie van persoonlijke gezondheids(zorg)gegevens	
Primaire exploitatie	**Secundaire exploitatie**
Wettelijke basis na advies van de PrC	Machtiging Sectoraal Comité
Domeinspecifieke UPP	**Projectspecifieke UPP**
Welomschreven finaliteit & proportionele gegevenstransmissie ('no more, no less')	
Gecertifieerd datamodel & DWH	
Datavalidatieplan & datakwaliteitsplan	
Toegankelijke metadata & m-DWH	
Koppeling door 'deterministic matching'	
Maximale dataprecisie	Ad hoc hercoderen / versluieren van 'gevoelige' gegevensvelden; 'small number' restricties
Gecertifieerd datamodel	Geaccrediteerd research team
Aandacht voor longitudinaliteit	In tijd beperkte toegang tot de gegevens
Aandacht vr. noden 2^{aire} exploitatie	Destructie van alle persoonlijke gegevens op het einde van het project

7.3 OPERATIONALISATIE

Op operationeel vlak vereist de aflevering van UPP's de tussenkomst van een bevoegde en onafhankelijke 'trusted third party' (TTP) als intermediair organisme, zeker als het gaat om koppeling van gegevens. De data-technische complexiteit van UPP-afleveringen en gegevenskoppelingen sluit uit dat beide bevoegdheden zouden toegekend worden aan de Privacy Commissie zelf.

BIJLAGE 8 : WETTELIJKE ASPECTEN

8.1 PERSOONSGEGEVENS EN DE WETTELIJKE BASIS VAN DE VERSCHILLENDE OVERHEIDSREGISTRATIES

Basisconcepten en definities m.b.t. het verwerken van persoonsgegevens wordt in eerste instantie gedicteerd door de Wet van 8 december 1992 tot bescherming persoonlijke levenssfeer (PRW), gewijzigd door de Wet van 11 december 1998 tot omzetting van de richtlijn van 24 oktober 1995 van het Europese Parlement en de Raad betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens. Het KB van 13/02/2001 brengt de uitvoeringsbesluiten op de PRW.

Historisch gesproken nochtans wordt de bekommernis voor een dergelijke afscherming al gevat in de Wet van 4 juli 1962 betreffende de openbare statistiek en de volkstellingen, gewijzigd door de Wet van 1 augustus 1985. Hier wordt het concept van *"globale en naamloze statistieken"* ingevoerd (Art. 2a en 6): "*Deze statistieken mogen gepubliceerd of aan derden medegedeeld worden, behalve indien, ingevolge het beperkt aantal aangevers, de onthulling van individuele toestanden mogelijk is*" (principe van de *'small numbers'*). En verder: "*In dit geval mogen zij niet gepubliceerd of aan een derde medegedeeld worden dan met de voorafgaande toestemming van de aangever of van de belanghebbende getelde*" (principe van de 'informed consent'). "*Bij ontbreken van dergelijke toestemming mag het Nationaal Instituut voor de Statistiek (NIS) bedoelde statistieken evenwel vertrouwelijk mededelen aan de belanghebbende ministeriële departementen, Rijksdiensten of diensten van een Gemeenschap, met uitsluiting van de fiscale besturen. In geen geval mogen de statistische onderzoekingen en studies van het NIS betrekking hebben op de persoonlijke levenssfeer, onder meer op het seksueel gedrag, de overtuiging of activiteit op politiek, levensbeschouwelijk of godsdienstig gebied, het ras of de etnische afstamming*" (concept van de 'gevoelige persoonsgegevens').

Anderzijds werden in Art. 206 van de Wet betreffende de verplichte verzekering voor geneeskundige verzorging en uitkeringen, gecoördineerd op 14 juli 1994[i], en het Koninklijk besluit van 3 juli 1996 (+ de bijwerkingen ervan) tot uitvoering van deze wet, aan de verzekeringsinstellingen de verplichting opgelegd tot overmaken van zgn. "statistische tabellen" m.b.t. de geboekte uitgaven van deze verplichte verzekering (artikels 347 tot 353). De inhoudelijke aspecten van de registraties en de termijnen waarbinnen ze moeten gebeuren worden bij Koninklijk besluit opgelegd. Sommige van deze tabellen bevatten identificeerbare persoonsgegevens m.b.t. tot de zorgvoorschrijver resp. zorgverstrekker (bijv. Doc P, Pharmanet,…). Andere bevatten weliswaar gepseudonimiseerde patiëntidentificaties (bijv. Pharmanet – uniek spoor sinds 2004).

Verder regelt het Koninklijk besluit van 6 december 1994 de inhoudelijke aspecten en modaliteiten van de verplichte mededeling door de niet-psychiatrische ziekenhuizen van de Minimale Klinische Gegegevens (MKG) en de Minimale Verpleegkundige Gegevens (MVG) aan de Minister die de Volksgezondheid onder zijn bevoegdheid heeft. Hierdoor wordt een systematische registratie van persoonlijke gezondheidsgegevens ingevoerd, zij het beperkt tot de ziekenhuisverblijven en zonder unieke patiëntidentificator: het ziekenhuis bepaalt zelf zijn patiëntnummer en is alleen gehouden ditzelfde nummer aan te houden binnen eenzelfde registratiejaar of voor langdurige verblijven.

De MPG-registratie wordt geregeld door de KB's van 25/02/1996, 20/09/1998 en 01/10/2002 (en vooral hun bijlagen voor wat betreft de precisering van de registratie-items). Het eerste van deze KB's werd trouwens op 26-01-2000 geannuleerd door een arrest van de Raad van State op grond van een schending van het anonimiteitsprincipe opgelegd in artikel 86 van de

[i] In de huidige context een aanvulling op de Wet van 9 augustus 1963 tot instelling en organisatie van een regeling voor verplichte verzekering voor geneeskundige verzorging en uitkeringen.

wet op de ziekenhuizen, gecoördineerd op 07/08/1987. Artikel 86 werd inmiddels aangepast door het art. 125 van de Wet van 12/08/2000 houdende sociale, budgettaire en andere bepalingen: *"… bedoelde gegevens die verband houden met de medische activiteiten mogen geen gegevens bevatten die de natuurlijke persoon waarop ze betrekking hebben rechtstreeks identificeren. Er mogen geen handelingen worden verricht die erop gericht zijn om deze gegevens in verband te brengen met de geïdentificeerde natuurlijke persoon waarop ze betrekking hebben, tenzij deze nodig zijn om de door de FOD VG aangewezen ambtenaren of aangestelden de waarachtigheid van de medegedeelde gegevens te laten nagaan.*"

Voor de in 2003 ingevoerde MUG-SMUR registratie en de eraan gekoppelde export van (gecodeerde) persoonlijke gezondheidsgegevens naar de FOD VG bestaat er weliswaar een wettelijke geregelde financiering (Koninklijk besluit van 29/03/2003 tot wijziging van het koninklijk besluit van 25 april 2002 betreffende de vaststelling en de vereffening van het budget van financiële middelen van de ziekenhuizen), maar er is geen detailbeschrijving van de aan de overheid mede te delen registratie-items opgenomen.

8.2. KOPPELING VAN PERSOONSGEGEVENS

De gekoppelde MKG-MFG gegevens zijn een voorbeeld bij uitstek van gegevensbanken die persoonsgegevens van patiënten en zorgverstrekkers bevatten, in de zin van Art. 1, §1 van de PRW. Deze wet maakt een duidelijk onderscheid tussen "natuurlijke personen", op wie ze van toepassing is, en "rechtspersonen", op wie ze niet (rechtstreeks) van toepassing is.

Het basisimperatief voor de MKG-MFG koppeling is dat het inschrijvingsnummer van de patiënt bij zijn VI (of, vanaf 2006, het INSZ) d.m.v. een procedure van *dubbele hashing-encryptie* irreversibel wordt getransformeerd tot een pseudoniem van de rechthebbende, dat door latere verwerkers van de gegevens niet meer kan teruggekoppeld worden tot de fysieke persoon. Hierbij is de tussenkomst van een onafhankelijke, intermediaire organisatie (een zgn. 'trusted third party' – TTP, in casu de TCT) vereist, als noodzakelijke tussenpersoon tussen de leveranciers van de brondata en de eindgebruiker(s). De TTP staat nl. in voor de 2^{e} hashing-encryptie van het pseudoniem. Technisch gezien en op zichzelf beschouwd kan dit pseudoniem m.a.w. bestempeld worden als een geanonimiseerd gegeven (geanonimiseerde variabele).

De gekoppelde gegevens in hun geheel bevatten echter ook meerdere specifieke elementen die kenmerkend zijn voor de fysieke of sociale identiteit van de patiënt, zoals leeftijd, geslacht, landcode van de wettige domicilie, nationaliteitsindicator, code van de woonplaats en sociale zekerheidsregime, etc… waardoor het geheel bezwaarlijk kan vallen onder de definitie van *"anonieme gegevens*", maar veeleer onder de definitie van *"gecodeerde gegevens"*. Samen met elementen als opname- en ontslagdatum evenals het ziekenhuis van het verblijf met de diagnose- en interventiecodes, houdt dit een risico in tot *indirecte, "contextuele" herkenning* voor een aantal van deze patiënten. Dit risico is des te groter naarmate deze elementen samen een "singulariteit" afbakenen. Dit is een klassiek probleem bij statistisch onderzoek: outliers aan beide zijden van de distributiecurve verdwijnen niet in de anonimiteit van de massa, gezien ze de facto marginale gevallen zijn[37, 38].

Om die reden werden op voorschrift van de Commissie voor de bescherming van de persoonlijke levenssfeer een aantal beperkingen opgelegd m.b.t. tot het toegestane detail (precisie) van de individuele MKG- resp. MFG-gegevens. Het doel hiervan is de indirecte herkenning te bemoeilijken door introductie van een zekere graad van "versluiering" ("fuzziness"). Deze beperkingen omvatten:

- geen volle geboortedatum maar enkel geboortejaar van patiënt in de MKG en enkel leeftijdscategorie in de MFG;
- geen volledig adres maar enkel postcode / NIS-code van de woonplaats van patiënt en land van domicilie in de MKG;
- geen volle nationaliteitscode maar enkel een nationaliteitsindicator (Belg, EU, niet-EU) in de MKG;
- geen volledige opname- en ontslagdatums, maar enkel jaar en maand voor MKG en MFG; in de MKG-gegevens werd echter ook nog de notie "weekdagnummer" (1 tot 7) toegestaan.

Jammer genoeg werden dezelfde principes gehanteerd m.b.t. tot de correspondentietabellen, wat erop neerkomt dat de gewilde versluiering t.a.v. de latere verwerkers van de gegevens problemen gecreëerd heeft op het vlak van de exacte koppeling van de MKG en de MFG verblijven (zie bespreking van het data process plan van de MKG-MFG in addenda van onderhavig rapport).

Verder is het belangrijk te onderstrepen dat de WVP in art. 7, §2, k) een verwerking van identificeerbare persoonsgegevens toestaat, mits "de verwerking noodzakelijk is voor het wetenschappelijk onderzoek en verricht wordt onder de voorwaarden vastgesteld door de Koning bij een in Ministerraad overlegd besluit, na advies van de Commissie voor de bescherming van de persoonlijke levenssfeer".

8.3. IDENTIFICATIE VAN VERZORGINGSINSTELLINGEN EN ZORGVERSTREKKERS

Ziekenhuizen zijn juridisch gesproken rechtspersonen zodat hun al dan niet directe identificatie sensu stricto niet onderworpen is aan de bepalingen van de Wet van 08-12-1992.

De risico's op indirecte, "contextuele" herkenning van fysieke personen, in casu niet alleen de ziekenhuispatiënt maar ook de zorgverstrekker in dit ziekenhuis, leiden echter tot de consequentie dat toch aandacht moet besteed worden aan een hercodering van deze ziekenhuizen t.a.v. de latere verwerkers van de gegevens. Vooral voor de outliers naar onder toe (kleine instellingen), zou er inderdaad wel een probleem kunnen ontstaan indien over een op een of andere manier a posteriori geïdentificeerd ziekenhuis resultaten worden gevonden met betrekking tot bepaalde kenmerken aangaande de onderzochte klinische praktijk, terwijl slechts een zeer beperkt aantal zorgverstrekkers in dat ziekenhuis dergelijke praktijken uitvoert. In dat geval zal iemand, die persoonlijk over voldoende kennis beschikt over dat ziekenhuis, uit de studieresultaten gegevens kunnen afleiden aangaande deze zorgverstrekker. Meer nog, indien die klinische praktijk in het geïdentificeerde ziekenhuis slechts een enkele keer zou zijn toegepast en indien iemand over voldoende kennis beschikt om te weten welke patiënt die specifieke klinische ingreep in dat ziekenhuis heeft ondergaan, dan is zelfs de identificatie van een patiënt niet uit te sluiten met daarenboven het bezwarende feit dat het dan nog veelal gaat om gezondheidsgegevens. Deze laatste zijn in de Wet van 08-12-1992 aan bijzondere beperkingen onderworpen.

Deze problematiek van indirecte herkenning heeft ertoe geleid dat, m.b.t. de overdracht van de gekoppelde MKG-MFG gegevens door de TCT voor verwerking in het kader van afgebakende onderzoeksprojecten, de verplichtig tot systematische, zij het vaak omkeerbare

'ad hoc' hercodering van de verzorgingsinstellingen werd opgelegd. De omkeerbaarheid is echter aan restrictieregels onderworpen (zie verder punt 8.6.).

8.4 HET PRINCIPE VAN DE 'AD HOC' HERCODERING VAN 'IDENTIFICATIEGEVOELIGE GEGEVENSVELDEN'

In de gekoppelde MKG-MFG gegevens, maar ook in alle andere registraties van persoonlijke gezondheids(zorg)gegevens (MPG, MVG, IMA-databeses, Pharmanet, MUG-registratie) komen volgende identificatiegevoelige gegevensvelden voor:

A. patiëntspecifieke velden:

- het dubbel gehashed & geëncrypteerd pseudoniem van de patiënt in MKG-MFG, IMA- en Farmanet- registraties;
- het patiëntnummer bij de VI in de MFG;

B. ziekenhuisspecifieke velden:

- het patiëntnummer in het ziekenhuis (MKG, MPG, MVG, MUG);
- het verblijfsnummer in het ziekenhuis (MKG, MPG, MVG, MUG);
- het erkenningsnummer van het ziekenhuis (MFG, MPG, MVG, MUG, IMA, …);
- het CIV-nummer van het ziekenhuis (MKG, MPG, MVG, MUG);
- Voor de mededeling van persoonlijke gezondheids(zorg)gegevens aan onderzoeksgroepen geldt de verplichting tot een 'ad hoc' hercodering van deze velden. De bedoeling hiervan is het isoleren van de in het kader van een welbepaalde, gemachtigde studie medegedeelde gegevens, opdat ze niet zouden kunnen samengebracht worden met gelijksoortige gegevens medegedeeld in het kader van enige andere studie. Hiermede wordt dan voor een deel tegemoet gekomen aan de vereiste dat de medegedeelde gegevens enkel zouden gebruikt worden voor de oorspronkelijke en gemachtigde finaliteit van de betreffende studie (zie eerste schema verder: 'restricted to original purpose'). Dit 'restrictieprincipe' geldt trouwens voor alle gegevensopvragingen van persoons- en gezondheidsgegevens, wat dus impliciet een extrapolatie van de verplichting tot 'ad hoc' hercodering inhoudt (behoudens specifiek gemachtigde uitzonderingen).

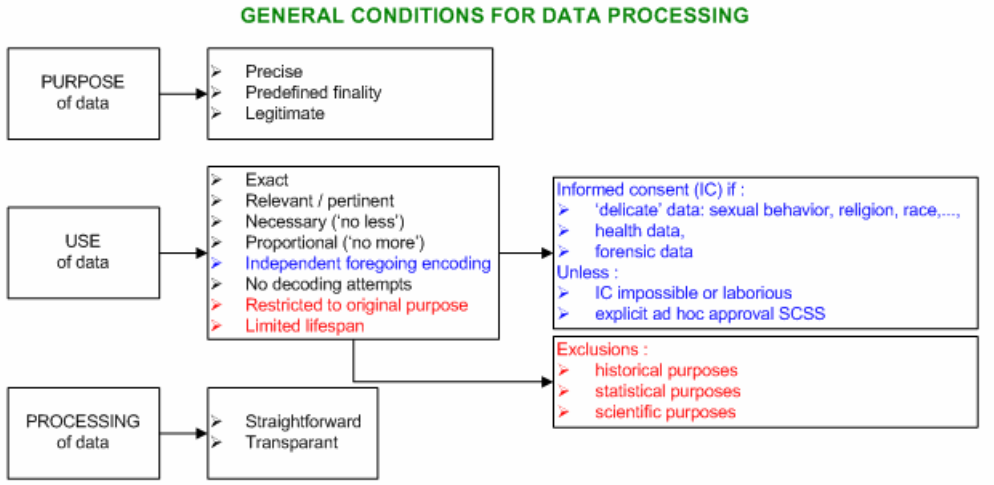
Theoretisch bestaat immers de mogelijkheid dat eenzelfde onderzoeksgroep, betrokken in 2 volkomen onderscheiden studieprojecten - bonafide of malafide - de bekomen gegevens met gebruikmaking van de hierboven opgesomde gegevensvelden samenbrengt in een groter geheel, wat volkomen in strijd is met het restrictieprincipe. In een 'worst case' scenario zou men zelfs, door herhaling van deze handelswijze, een quasi volledige database kunnen samenstellen.

8.5 TOEGANG TOT GEGEVENS

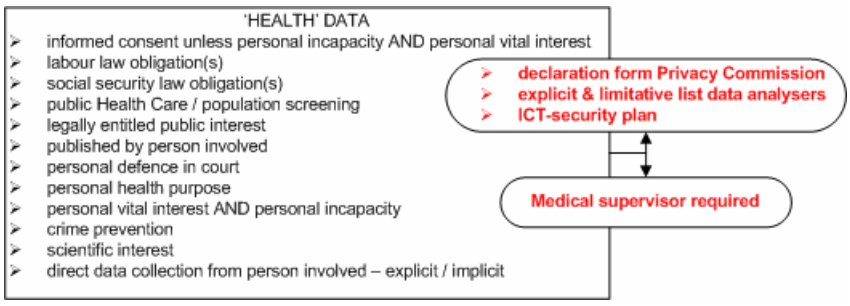
De toegang tot persoonsgegevens wordt strikt geregeld door de PRW : elke verwerking van persoonsgegevens is onderworpen aan strenge wettelijke beperkingen. Deze laatste zijn nog stringenter als het gaat om gezondheidsgegevens (zie fig. 2 & 3).

Samengevat is de toegang tot persoonsgegegevens m.b.t. de gezondheid onderworpen aan volgende voorwaarden:

- Informed consent van de patiënt (expliciet);
- Bij ontstentenis hiervan een specifieke machtiging door de Privacy Commissie bij monde van een daartoe opgericht Sectoraal Comité Sociale Zekerheid[j]
- Een officiële, gedetailleerde aangifte van de verwerking bij de Privacy Commissie, i.v.m. openbaarheid van de gegevensverwerking (PRW);
- Het overmaken van een expliciete en limitatieve lijst van de personen die de gegevensverwerking (zullen) uitvoeren;
- De aanstelling van een beoefenaar in de gezondheidszorg aan onder wiens toezicht en verantwoordelijkheid de verwerking en de analyse van persoonsgegevens betreffende de gezondheid gebeurt;
- Een goedgekeurd veiligheidsplan, zowel qua materiële beveiliging als qua informatie beveiliging en begeleid door een aangestelde externe veiligheidsconsulent.



Figuur 2: algemene voorwaarden voor (geheel of gedeeltelijk geautomatiseerde) verwerking van persoonsgegevens



Figuur 3: bijzondere voorwaarden voor (geheel of gedeeltelijk geautomatiseerde) verwerking van persoonlijke gezondheidsgegegevens

[j] naast andere sectorale comités: sectoraal comité voor het Rijksregister voor de toegang tot en het gebruik van het identificatienummer van dit register, sectoraal comité voor de Federale Overheid voor machtigingsaanvragen van een Federale Overheidsdienst voor toegang tot gegevens die door een andere Federale Overheidsdienst werden verzameld en sectoraal comité voor de Kruispuntbank van Ondernemingen voor machtigingen m.b.t. de toegang tot of de mededeling van de informatiegegevens van de Kruispuntbank van Ondernemingen.

8.6 MEDEDELING VAN GEKOPPELDE ZIEKENHUISGEGEVENS DOOR DE TECHNISCHE CEL

Op het vlak van mededeling van de gekoppelde gegevens door de technische cel, geldt als algemeen principe dat hierbij niet alleen de fysieke, maar ook de rechtspersonen 'geanonimiseerd' worden (Art. 156, §1 en §3 van de wet van 29-04-1996). Het KB van 18/10/2001, tot uitvoering van artikel 156, § 3 van dezelfde wet, gaat hier dieper op in en maakt een duidelijk onderscheid tussen:

- enerzijds de Federale Overheidsdienst Volksgezondheid en het Rijksinstituut voor Ziekte- en Invaliditeitsverzekering, die beide van rechtswege toegang krijgen tot gegevens waarbij rechtspersonen zijn geïdentificeerd en dit op hun verzoek, maar uitsluitend voor zover dit noodzakelijk is voor de uitvoering van hun respectievelijke wettelijke opdrachten;
- anderzijds de leden van de beheers- of adviesorganen van de Federale Overheidsdienst Volksgezondheid of het RIZIV en alle andere personen, die om deze gegevensoverdracht verzoeken. Het meedelen van gegevens met identificatie van de rechtspersoon kan hier enkel gebeuren indien deze identificatie onontbeerlijk is voor het uitvoeren van de opdracht.

Onder "gegevens noodzakelijk voor de uitvoering van de opdrachten" wordt verstaan: alle voor de kennis, de conceptie en het beheer van het ziekenhuisbeleid onmisbare gegevens. Hiertoe behoren de gegevens aangewend met beleidsvoorbereidend oogmerk en voor doeleinden van wetenschappelijk onderzoek of meer specifiek:

- de gegevens die toelaten verbanden te leggen tussen de uitgaven van de verzekering voor geneeskundige verzorging en de behandelde pathologie;
- de gegevens die toelaten financieringsregels, erkennings- en programmatienormen uit te werken;
- de gegevens die een permanente evaluatie van de medische praktijk toelaten.

This page is left intentionally blank.

Dépôt légal : D/2006/10.273/15

KCE reports

1. Efficacité et rentabilité des thérapies de sevrage tabagique. D/2004/10.273/2.
2. Etude relative aux coûts potentiels liés à une éventuelle modification des règles du droit de la responsabilité médicale (Phase 1). D/2004/10.273/4.
3. Utilisation des antibiotiques en milieu hospitalier dans le cas de la pyélonéphrite aiguë. D/2004/10.273/6.
4. Leucoréduction. Une mesure envisageable dans le cadre de la politique nationale de sécurité des transfusions sanguines. D/2004/10.273/8.
5. Evaluation des risques préopératoires. D/2004/10.273/10.
6. Validation du rapport de la Commission d'examen du sous financement des hôpitaux. D/2004/10.273/12.
7. Recommandation nationale relative aux soins prénatals: Une base pour un itinéraire clinique de suivi de grossesses. D/2004/10.273/14.
8. Systèmes de financement des médicaments hospitaliers: étude descriptive de certains pays européens et du Canada. D/2004/10.273/16.
9. Feedback: évaluation de l'impact et des barrières à l'implémentation – Rapport de recherche: partie 1. D/2005/10.273/02.
10. Le coût des prothèses dentaires. D/2005/10.273/04.
11. Dépistage du cancer du sein. D/2005/10.273/06.
12. Etude d'une méthode de financement alternative pour le sang et les dérivés sanguins labiles dans les hôpitaux. D/2005/10.273/08.
13. Traitement endovasculaire de la sténose carotidienne. D/2005/10.273/10.
14. Variations des pratiques médicales hospitalières en cas d'infarctus aigu du myocarde en Belgique. D/2005/10.273/12
15. Evolution des dépenses de santé. D/2005/10.273/14.
16. Etude relative aux coûts potentiels liés à une éventuelle modification des règles du droit de la responsabilité médicale. Phase II : développement d'un modèle actuariel et premières estimations. D/2005/10.273/16.
17. Evaluation des montants de référence. D/2005/10.273/18.
18. Utilisation des itinéraires cliniques et guides de bonne pratique afin de déterminer de manière prospective les honoraires des médecins hospitaliers: plus facile à dire qu'à faire.. D/2005/10.273/20
19. Evaluation de l'impact d'une contribution personnelle forfaitaire sur le recours au service d'urgences. D/2005/10.273/22.
20. HTA Diagnostic Moléculaire en Belgique. D/2005/10.273/24, D/2005/10.273/26.
21. HTA Matériel de Stomie en Belgique. D/2005/10.273.28.
22. HTA Tomographie par Emission de Positrons en Belgique. D/2005/10.273/30.
23. HTA Le traitement électif endovasculaire de l'anévrysme de l'aorte abdominale (AAA). D/2005/10.273.33.
24. L'emploi des peptides natriurétiques dans l'approche diagnostique des patients présentant une suspicion de décompensation cardiaque. D/2005/10.273.35
25. Endoscopie par capsule. D2006/10.273.02.
26. Aspects médico-légaux des recommandations de bonne pratique médicale. D2006/10.273/06.
27. Qualité et organisation des soins du diabète de type 2. D2006/10.273/08.
28. Recommandations provisoires pour les évaluations pharmacoéconomiques en Belgique. D2006/10.273/11.
29. Recommandations nationales Collège d'oncologie : A. cadre général pour un manuel d'oncologie B. base scientifique pour itinéraires cliniques de diagnostic et traitement, cancer colorectal et cancer du testicule. D2006/10.273/13.
30. Inventaire des bases de données de soins de santé. D2006/10.273/15.