

Architectuur Centrale infrastructuur



PARELSNOER INITIATIEF

Versie: 1.1
Status: definitief
Datum: 20 december 2008

Management samenvatting

Het project Parelsnoer beoogt binnen de academische centra een infrastructuur te realiseren voor het verzamelen van klinische data en opzetten van biobanken op interuniversitair niveau. Het Parelsnoer project richt zich in eerste instantie op het faciliteren van combinaties van bijzondere biobanken en patiëntcohorten voor acht ziektebeelden.

Op dit moment zijn veel eisen en wensen voor de inrichting van Parelsnoer nog niet of nauwelijks bekend. Ook is momenteel nog weinig zicht op het gebruik en de omvang van de uit te wisselen informatie. Om toch voortgang te kunnen boeken in het project is een aantal uitgangspunten gekozen:

- Opleveren van de datasets vanuit de instellingen dient met zo min mogelijk inspanning van de kant van de aanleverende instellingen in het zorgdomein plaats te vinden. Dit geldt zowel voor de realisatiefase als voor de operationele situatie.
- De opzet dient zodanig te zijn dat met een beperkte invulling al in 2008 de eerste resultaten kunnen worden getoond. Dit betekent dat wordt uitgegaan van minimale investeringen voor Centrale Infrastructuur (CI), zodat alleen noodzakelijke voorzieningen in de CI kunnen worden opgenomen.
- Uit overwegingen van privacybescherming zijn door de UMC's opgeleverde gegevens niet direct herleidbaar tot personen. Onder bepaalde, nog te definiëren strikte voorwaarden, is deze herleidbaarheid wel mogelijk (omkeerbare pseudonimiteit).
- Bij het verstrekken van een dataset voor onderzoek moet zoveel mogelijk voorkomen worden dat verschillende geautoriseerde datasets die ooit beschikbaar gesteld zijn voor verschillende onderzoeken, gekoppeld dan wel samengevoegd worden tot nieuwe datasets zonder dat daarvoor toestemming is gegeven.
- De Centrale Infrastructuur ondersteunt het opleveren van managementinformatie, zoals bijvoorbeeld het gebruik per PAREL, statistiek van aanleveringen door UMC's etc..

Omdat in 2008 voor Parelsnoer de eerste resultaten moeten kunnen worden getoond wordt voor de startarchitectuur voor Parelsnoer uitgegaan van beperkte centrale voorzieningen gebaseerd op het periodiek "pushen" van datasets. Met deze centrale voorzieningen worden de volgende diensten aan de UMC's geboden:

- Ontvangen, verwerken en beheren van datasets;
- Beschikbaar stellen van datasets aan onderzoekers;
- Een catalogus met informatie over de beschikbare datasets, beeldmateriaal en biomaterialen;
- Voorzieningen voor beveiligde communicatie.
- Management rapportages

De gemaakte architectuurkeuzen zijn:

- De aangesloten instellingen zijn verantwoordelijk voor het periodiek en tijdig beschikbaar stellen van volledige datasets volgens de vastgestelde standaarden en conform een Parelsnoer Informatie Model (PIM).
- De onderzoeker krijgt één dataset aangeleverd. De onderzoeker heeft een eigen onderzoekssysteem en zal alle analyses en bewerkingen van de data in zijn onderzoeksomgeving (geen onderdeel van CI) uitvoeren.
- Bij elke nieuwe aanlevering wordt de oude dataset uit de CI verwijderd en vervangen door de nieuwe dataset.
- De CI heeft geen kennis van het BSN. Het BSN in de datasets van de UMC's wordt gepseudonimiseerd door de UMC's voordat die worden verstuurd naar de CI. Eventueel terugvertalen van de pseudocode naar een BSN vindt plaats door de UMC's. Het UMC dient te voorzien in de mogelijkheid dat een uitgegeven pseudocode kan worden terugvertaald naar het oorspronkelijke BSN.
- Herleidbaarheid van de uitgevoerde handelingen in de CI wordt gerealiseerd door vastlegging van alle interacties (logging).
- De combinatie van beveiligd transport, een besloten CI-omgeving waartoe alleen de CI-beheerder toegang heeft, en een dubbele pseudonimisatieslag is voldoende om te voldoen aan de beveiligingseisen.
- Er wordt gestreefd naar één technische protocol op basis van webtechnologie voor uitwisseling van onderzoeksgegevens. Voor 2008/2009 wordt aangenomen dat tijdelijk meer dan één protocol mogelijk kan zijn, afhankelijk van de mogelijkheden van de UMC's.

Inhoudsopgave

1	Inleiding	6
	1.1 Algemeen	6
	1.2 Terminologie	6
	1.3 Doelstelling van het project Parelsnoer	7
	1.4 Opzet van dit document	7
	1.5 Scope van de architectuur-beschrijving	9
	1.6 Documenthistorie	9
2	Bedrijfsarchitectuur	10
	2.1 Inleiding	10
	2.2 Gebruiksscenario's	11
	2.3 Verantwoordelijkheden	13
	2.4 Eisen vanuit de onderzoeksprocessen	14
	2.5 Eisen aan privacy-bescherming en informatiebeveiliging	16
	2.6 Inrichtingsvarianten	17
	2.7 Het uitwisselingsproces	22
3	Informatie (systeem) architectuur	28
	3.1 Inleiding	28
	3.2 Functionaliteit van de gemeenschappelijke voorzieningen	28
	3.3 Informatiemodellen: semantiek en syntax	28
	3.4 Eisen aan koppelvlakken met instellingssystemen	29
	3.5 Functionaliteit voor beveiliging en beheer.	29
	3.6 Functionaliteit voor communicatie	30
4	Technische architectuur	31
	4.1 Inleiding	31
	4.2 Het servicemodel	31
	4.3 Interfacemodel	32
	4.4 Beveiliging, continuïteit en beheer	33
5	Ontwikkelingsscenario's	34
	5.1 Inleiding	34
	5.2 Rijkere zorg-omgeving	34
	5.3 Rijkere publicatie-omgeving	34
6	Literatuur	35

1 Inleiding

1.1 Algemeen

In dit document wordt de architectuur voor de Centrale Infrastructuur van Parelsnoer beschreven. Dit document beperkt zich tot de hoofdlijnen, dat wil zeggen het beschrijft de belangrijkste uitgangspunten en architectuurprincipes, en is mede bedoeld om de afstemming met en tussen betrokkenen richting te geven en te structureren. In een volgende fase wordt op basis van de uitgangspunten en principes een gedetailleerdere uitwerking gedaan ten behoeve van de realisatie.

Dit document is opgesteld op basis van de uitgangspunten die zijn aangedragen vanuit het project Referentiekaders.

1.2 Terminologie

Algemene biobank	Een biobank die materiaal bevat van een deel van de bevolking en zich niet beperkt tot één ziekte
Analysedomein	Onderzoeksomgeving waarin de ontvangen gegevens worden geanalyseerd
Biobank	Een verzameling van lichaamsmateriaal, met medische, genetische, genealogische en/of andere data daaraan gekoppeld.
Bijzondere biobank	Een cohort of patiëntenbestand met lichaamsmateriaal, vaak hypothesegeïnduceerd.
Catalogus	Een overzicht van kenmerken van beschikbare onderzoeksgegevens, de data-definities en een overzicht van beschikbaar bio-materiaal
Data-elementen	Kenmerken van entiteiten die vastgelegd worden in een dataset. Bijvoorbeeld, het geslacht is een kenmerk van een persoon dat wordt bijgehouden in de verzameling patiënten.
Data dictionary	Een verzameling definities en representaties (datatypes) van data-elementen.
Datamining	Datamining is het op een geautomatiseerde manier patronen en relaties te ontdekken in grote hoeveelheden gegevens. Datamining is gebaseerd op statistiek, machine learning, patroonherkenning, database management.
Dataset	Samenhangende verzameling data-elementen
Datawarehouse	Een datawarehouse is een database waarin data uit verscheidene bronsystemen worden gedupliceerd om in dit gecombineerde gegevenspakhuis patronen te ontdekken en/of analyses uit te voeren.
ETL	Extraction, Transform, and Load. De term staat voor een groep technologieën die veelal gebruikt wordt bij de koppeling tussen bronsystemen en een datawarehouse om de brongegevens in de datawarehouse te laden.

Gepseudonimiseerd	Gepseudonimiseerde gegevens in uitgegeven datasets zijn door onderzoeker niet te herleiden tot de personen waar deze gegevens betrekking op hebben. Een dergelijke herleiding kan alleen plaatsvinden na een te doorlopen procedure
Nationale biobank	Een biobank die beoogt een representatief deel van de bevolking, of de gehele bevolking te beslaan; vaak hypothesegenererend.
Pseudonimisatie	Pseudonimisatie verwijst naar technieken die persoonsidentiteiten vervangen door pseudo-identiteiten (pseudo-ID's), en wel op een zodanige manier dat zij niet in verband kunnen worden gebracht met de overeenkomstige persoonsidentiteit . Bij omkeerbare pseudonimisatie is een procedure beschikbaar om dit verband opnieuw aan te brengen.
Publicatiedomein	Centraal ingerichte omgeving waar de van de UMC's ontvangen gegevens na eventuele bewerking aan onderzoekers in het analysedomein ter beschikking worden gesteld.
Query/reporting	Een opdracht die aan een database wordt gegeven om een bepaalde zoekactie uit te voeren en de zoekresultaten terug rapporteert.
Zorgdomein	Omgeving waar de zorggegevens van patiënten worden vastgelegd

1.3 Doelstelling van het project Parelsnoer

Dit project Parelsnoer beoogt¹ binnen de academische centra een infrastructuur te realiseren voor het verzamelen van klinische data en opzetten van biobanken op interuniversitair niveau. Het Parelsnoer project richt zich in eerste instantie op het faciliteren van combinaties van bijzondere biobanken en patiëntencohorten voor acht ziektebeelden. Om de expertise van de UMC's ten volle te benutten zal de opzet van de infrastructuur en het onderzoek zich richten op topreferente en aan het UMC gelieerde patiëntenpopulaties. De ziektebeelden ('parels') waarvoor in dit project een data- en biobank zal worden gerealiseerd zijn: neurodegeneratieve ziekten, leukemie, reumatoïde artritis, erfelijke darmkanker, diabetes mellitus, nierfalen, cerebro vasculair accident en inflammatoire darmziekten.

De wetenschappelijke gegevens die worden gecreëerd met behulp van bovengenoemde infrastructuur zullen leiden tot een betere gezondheid van patiënten met genoemde ziektebeelden, en versterking van de economische positie van de farmaceutische industrie in Nederland. De gecreëerde infrastructuur plaatst Nederland op dit gebied tevens in een Europese koppositie.

1.4 Opzet van dit document

In dit document wordt de globale architectuur van de Centrale Infrastructuur beschreven. De architectuurbeschrijving bestaat uit uitgangspunten en architectuurprincipes. Aangezien op dit moment veel (business) eisen en wensen nog niet bekend zijn, zal, om toch tot een zinvolle uitwerking van de architectuur te komen, een aantal **uitgangspunten** moeten worden gekozen. Deze uitgangspunten zullen in een latere fase

¹ Uit "Werkplan Parelsnoer"

van het project moeten worden gevalideerd. Op basis van deze uitgangspunten worden architectuurprincipes gedefinieerd. Een **architectuurprincipe** is een (beredeneerde) inrichtingskeuze die de ontwerpruimte van een te ontwerpen systeem beperkt.

In sommige gevallen wordt voor de korte termijn (inrichting in 2008/2009) een andere inrichtingskeuze voorzien dan voor de periode daarna. Dit zal dan worden aangegeven met **architectuurprincipe (2008/2009)** respectievelijk **architectuurprincipe (toekomst)**.

1.5 Scope van de architectuur-beschrijving

Een belangrijke voorwaarde voor efficiënte, betrouwbare en veilige gegevensuitwisseling tussen (systemen van) UMC's is de totstandkoming van een aantal lokale UMC-infrastructuren die aansluiten op een gemeenschappelijke, centrale infrastructuur. Om dit te bereiken is een gemeenschappelijk kader noodzakelijk voor alle direct betrokkenen, met name voor ontwikkelaars van systemen en (eventuele) dienstverleners. Dit architectuurdocument levert dit kader door de samenhang van en samenwerking tussen de ICT-voorzieningen te beschrijven.

Afbakening:

Uitgangspunt 1: Administratieve procedures vallen buiten de scope van deze architectuurbeschrijving

1.6 Documenthistorie

Versie	Datum	Toelichting
0.1	17 juni 2007	Vooraf structuurplan
0.2	21 augustus 2007	Interne versie, werkdocument
0.3	31 augustus 2007	Eerste volledige versie
0.4	13 september 2007	Intern commentaar verwerkt
0.5	25 september 2007	Reactie klankbordgroep verwerkt
0.6	11 oktober 2007	Verwerking overig ontvangen commentaar
0.7	6 november 2007	Verwerking commentaar directie PARELSNOER
1.0	8 november 2007	Vorgelegde versie
1.1	30 november	Vastgestelde versie

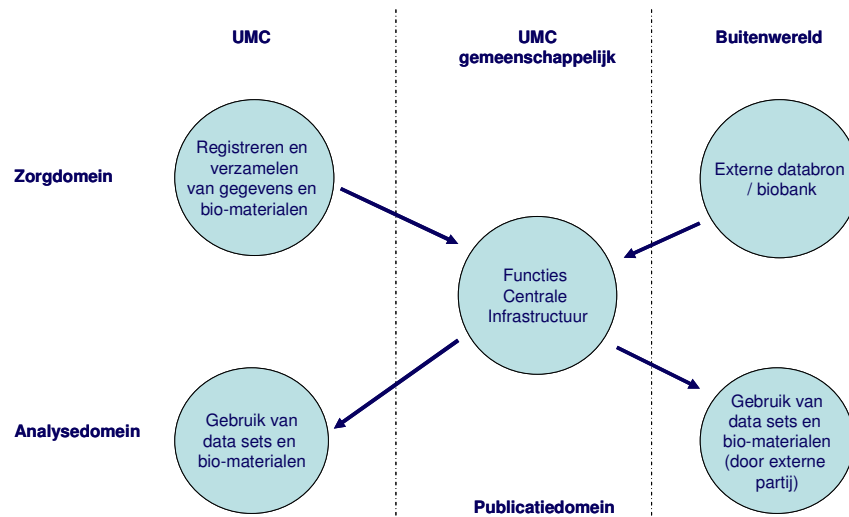
2 Bedrijfsarchitectuur

2.1 Inleiding

In dit document worden alle aspecten beschreven die relevant worden geacht voor de inrichting van de Centrale Infrastructuur (CI). Naast dit document is een Referentiekader gedefinieerd met daarin een meer uitputtende beschrijving van de business georiënteerde onderwerpen.

In het huidige medisch wetenschappelijke onderzoek speelt de relatie tussen fundamentele kenmerken (genen) en processen in het lichaam (eiwitten, metabolieten), én het klinisch beloop een belangrijke rol. In Nederland is, na de integratie van de academische ziekenhuizen en de medische faculteiten tot UMC's, het translationele onderzoek, dat kliniek en laboratorium koppelt, enorm tot bloei gekomen. Kennis over het ontstaan en het variabele beloop van ziekten is in toenemende mate gebaseerd op patiëntcohorten waarvan zowel de klinische data (inclusief digitaal beeldmateriaal) als analyses op lichaamsmateriaal aanwezig zijn. Daarnaast worden behandelingsopties sterk bepaald door waarnemingen in goed gedocumenteerde patiëntengroepen. Een grotere patiëntengroep c.q. patiëntcohort maakt onderzoek beter mogelijk.

Het project Parelsnoer heeft tot doel een infrastructuur te realiseren voor het verzamelen van klinische data en opzetten van biobanken op interuniversitair niveau. Zoals aangegeven in **Figuur 1** zullen registratiesystemen van UMC's in het zorgdomein als bron fungeren voor gegevens en biomaterialen. In een latere fase kunnen ook externe databronnen en/of biobanken zijn betrokken. Uitwisseling naar de onderzoekers in het analysedomein zal plaatsvinden via een aantal gemeenschappelijke voorzieningen, in de figuur aangeduid met "Functies Centrale infrastructuur". De onderzoekers in het analysedomein waar de gegevens aan ter beschikking worden gesteld zijn van de betrokken UMC's en in een latere fase wellicht ook van externe partijen.



Figuur 1: Globale opzet Parelsnoer

2.2 Gebruiksscenario's

2.2.1 Inleiding

Ofschoon nog weinig bekend is over het verwachte gebruik van Parelsnoer, begint zich wel een aantal lijnen af te tekenen. De nu onderkende gebruiksscenario's zijn:

1. Het beschikbaar stellen van een catalogus met informatie over de beschikbare datasets en biomaterialen;
2. Het beschikbaar stellen van (gepseudonimseerde) dataset(s) voor bevraging en analyse;
3. Het beschikbaar stellen van biomateriaal;
4. Het beschikbaar stellen van beeldmateriaal;
5. Het verwijderen van individuele patiëntgegevens uit de in de CI opgeslagen datasets.

2.2.2 Beschikbaar stellen van een catalogus met informatie over de beschikbare datasets en biomaterialen

Voor onderzoekers/coördinatoren zijn per Parel centraal de volgende gegevens in een catalogus beschikbaar:

- definities van data-elementen van de datasets (datadictionary);
- informatie over de beschikbare klinische gegevens (statistische kentallen);
- informatie over het beschikbare biomateriaal (statistische kentallen).
- informatie over het beschikbare beeldmateriaal (statistische kentallen).

2.2.3 Beschikbaar stellen van (gepseudonimiseerde) dataset(s) voor bevraging en analyse

Onderzoekers kunnen na goedkeuring van een onderzoeksvoorstel de beschikking krijgen over datasets voor hun onderzoek. Deze datasets bevatten (gepseudonimiseerde) gegevens over het ziektebeeld van een persoon of groepen van personen.

Uitgangspunt 2: De architectuur voorziet in de mogelijkheid gecombineerde datasets van meerdere onderzoeksgebieden op te leveren, zodat zoekvragen betrekking kunnen hebben op datasets van meerdere onderzoeksgebieden (Parels).

Hoewel op dit moment nog geen Use case is vastgesteld waarbij de behoefte aan deze mogelijkheid aan de orde is, zal hiermee in de architectuur wel rekening worden gehouden. Voor 2008/2009 wordt voorzien dat zoekvragen betrekking hebben op één onderzoeksgebied (Parel).

De omvang van de uit te wisselen datasets is naar verwachting de komende jaren beperkt. Om de complexiteit van de oplossing te beperken wordt vooralsnog niet gekozen voor het incrementeel beschikbaar stellen van de opgetreden wijzigingen in de datasets.

Er wordt er vanuit gegaan dat de onderzoeker voor het adequaat specificeren van zijn onderzoeksvraag gebruik maakt van de catalogus. Met de daarin opgenomen definities van beschikbare informatie kan immers de onderzoeksvraag eenduidig gespecificeerd en vervolgens beantwoord worden.

2.2.4 Beschikbaar stellen van biomateriaal

Op basis van de catalogusgegevens kan een onderzoeker na goedkeuring van een daartoe ingediend verzoek de beschikking krijgen over beschikbaar biomateriaal.

Uitgangspunt 3: Het benodigde logistieke proces om biomateriaal bij de verzoeker (onderzoeker) aan te leveren, wordt vanuit de Centrale Parelsnoerorganisatie gecoördineerd en door CI gefaciliteerd.

Uitgangspunt 4: Alle UMC's stellen actuele gegevens over het aanwezig biomateriaal periodiek als dataset beschikbaar zodat CI een actueel overzicht beschikbaar kan stellen van het beschikbare biomateriaal. Het biomateriaal zelf blijft opgeslagen binnen de UMC's.

2.2.5 Beschikbaar stellen van beeldmateriaal

Op basis van de catalogusgegevens kan een onderzoeker na goedkeuring van een daartoe ingediend verzoek de beschikking krijgen over beschikbaar beeldmateriaal.

Uitgangspunt 5: Er wordt voor gekozen beeldmateriaal niet op te nemen in uit te wisselen datasets tussen UMC's en CI.

Uitgangspunt 6: Alle UMC's stellen actuele gegevens over het aanwezig beeldmateriaal regelmatig als dataset beschikbaar zodat CI een actueel overzicht beschikbaar kan stellen van het beschikbare beeldmateriaal. Het beeldmateriaal zelf blijft opgeslagen binnen de UMC's.

Uitwerking van dit gebruiksscenario heeft nog niet plaatsgevonden. Consequenties voor het opslaan, beschikbaar stellen en transporteren van grote datahoeveelheden bij uitwisseling van beeldmateriaal zijn derhalve niet uitgewerk. Bij deze uitwerking zal aandacht moeten worden besteed aan de inrichting van het communicatienetwerk, bijvoorbeeld de inzet van lichtpaden, en de faciliteiten voor verwerking en opslag van grote bestanden, bijvoorbeeld inzet van BIG GRID².

2.2.6 Het verwijderen van patiëntgegevens uit de in de CI opgeslagen datasets.

Het kan voorkomen dat een patiënt zijn toestemming (consent) intrekt om zijn gegevens voor onderzoek te gebruiken. De eisen op dit punt zijn momenteel nog niet helder. Tot hierover helderheid is wordt er van uitgegaan dat voor alle aangeleverde gegevens in de datasets toestemming is voor gebruik in onderzoek.

Uitgangspunt 7: Bij intrekken van de toestemming van de patiënt voor het gebruik van zijn gegevens, mogen deze gegevens niet meer worden gebruikt in nieuwe datasets en wordt door het betreffende UMC direct een nieuwe dataset aan de CI ter beschikking gesteld.

2.3 Verantwoordelijkheden

Uitgangspunt 8: De aangesloten instellingen zijn verantwoordelijk voor het beheren van hun eigen gegevens en de kwaliteit van de gegevens in de aangeleverde dataset.

Architectuurprincipe 1: De aangesloten instellingen zijn verantwoordelijk voor het periodiek en tijdig beschikbaar stellen van volledige datasets volgens de vastgestelde standaarden en conform een Parelsnoer Informatie Model (PIM).

Architectuurprincipe 2: Er is een centrale organisatie die verantwoordelijk is voor:

- Het beheren van centraal vastgelegde datasets en het bewaken de gemaakte afspraken voor aanlevering;
- Het op verzoek van wetenschappelijke commissie(s) beschikbaar stellen van geïntegreerde datasets;
- De coördinatie van het logistieke proces voor het beschikbaar stellen van biomaterialen;
- Het in opdracht koppelen van nieuwe (externe) gegevensbronnen;

² Bovendien moeten nadere afspraken worden gemaakt over het toepassen van standaarden voor beeldopslag in relatie tot de bijbehorende verslagen, zoals bijvoorbeeld DICOM SR. Ook het anonimiseren/pseudonimiseren is daarbij een aandachtspunt. Hiertoe zal een deelproject worden gestart.

- Het aanleveren van management informatie over het gebruik van Parelsnoer
- De ontwikkeling en het onderhoud van centrale standaarden voor:
 - Informatiemodel(len);
 - Codestelsels (vocabularies);
 - Interface-definities en uitwisselingstandaarden.
- Beheer van de centrale infrastructuur.

Uitgangspunt 9: Per PAREL beslist een wetenschappelijke commissie over het ter beschikking stellen van datasets of van biomateriaal, op basis van een door de onderzoeker(s) ingediend onderzoeksvoorstel.

Uitgangspunt 10: De onderzoeker die een dataset, beeldmateriaal of biomateriaal heeft aangevraagd en ontvangen krijgt het gebruiksrecht over deze gegevens, beeldmateriaal of biomateriaal. De gegevens, beeldmaterialen of biomaterialen mogen alleen worden gebruikt binnen het kader van het onderzoek conform het ingediende onderzoeksvoorstel³.

Uitgangspunt 11: Bij het verstrekken van een dataset voor onderzoek moet zoveel mogelijk voorkomen worden dat verschillende geautoriseerde datasets die ooit beschikbaar gesteld zijn voor verschillende onderzoeken, gekoppeld danwel samengevoegd worden tot nieuwe datasets zonder dat daarvoor toestemming is gegeven.

Uitgangspunt 12: De onderzoeker publiceert de onderzoeks(tussen)resultaten en maakt kenbaar waar deze kunnen worden ingezien. Deze verwijzing kan worden opgenomen in de catalogus. Op termijn is het niet uitgesloten dat onderzoeksresultaten op een centrale plek worden opgeslagen en ontsloten

Uitgangspunt 13: De aan de onderzoeker aangeleverde dataset wordt binnen de centrale infrastructuur wordt minimaal gedurende de periode bewaard die de wet voorschrijft.

Uitgangspunt 14: CI ondersteunt de service "onafhankelijke reproduceerbaarheid" gedurende de in uitgangspunt 13 genoemde periode zodat een onderzoeksresultaat desgewenst autonoom herhaald kan worden door een onafhankelijke derde.

2.4 Eisen vanuit de onderzoeksprocessen

Uitgangspunt 15: De gegevens van de UMC's van een Parel worden als één geheel (één dataset) aangeleverd aan CI. Per PAREL wordt een aparte dataset aangeleverd. De CI maakt per Parel een samengestelde dataset die is opgebouwd uit de datasets van de UMC's.

³ De ethische en juridische aspecten worden uitgewerkt in een eind oktober op te leveren Reglement.

Uitgangspunt 16: Overdracht van gegevens is niet tijdkritisch; er hoeft dus geen “real time” overdracht plaats te vinden. Overdracht van de dataset aan de onderzoeker dient na goedkeuring door de wetenschappelijke commissie binnen 1 week te hebben plaatsgevonden.

Uitgangspunt 17: De opzet van Parelsnoer dient schaalbaar te zijn; dit betekent dat:

1. nieuwe Parels kunnen worden toegevoegd;
2. nieuwe externe (commerciële) bronnen kunnen worden toegevoegd;
3. de dataset van een Parel kan worden uitgebreid.

Uitgangspunt 18: De catalogusgegevens dienen continu 7 * 24 online beschikbaar te zijn.

Uitgangspunt 19: Het beschikbaarheidspercentage van de Centrale Infrastructuur (CI) is minimaal 95%.

Uitgangspunt 20: Aangeleverde en uitgegeven datasets worden gekenmerkt door een uniek versienummer. Op de individuele patiëntgegevens vindt geen versiebeheer plaats.

Uitgangspunt 21: Aanvullende vragen van een onderzoeker op reeds aangeleverde datasets hebben nooit betrekking op de eerder aangeleverde datasets, maar altijd op de meest actuele dataset.

Uitgangspunt 22: Indien een onderzoeker over meer informatie wil beschikken die wel binnen CI beschikbaar, maar waar geen toestemming voor is verleend, dan dient een nieuwe formele aanvraag te worden ingediend.

Uitgangspunt 23: Inhoudelijke updates van een dataset worden niet pro-actief doorgeleverd aan een onderzoeker anders dan het beschikbaar stellen van een geheel nieuwe dataset.

Uitgangspunt 24 (toekomst): Indien een onderzoeker van een of meerdere UMC's meer informatie omtrent een patiënt dan wel geregistreerde data wenst, dan kan hiervoor een aanvraagprocedure voor geïnitieerd worden.

Uitgangspunt 25: (toekomst): Op aanvraag kan een onderzoeker ten behoeve van longitudinaal onderzoek een nieuwe dataset aangeleverd krijgen met daarin dezelfde patiënten.

Uitgangspunt 26 (toekomst): Op termijn dient een koppeling met bestaande biobanken buiten de UMC's te kunnen worden gerealiseerd.

Uitgangspunt 27 (toekomst): Op termijn dienen ook nader vast te stellen derde partijen te kunnen worden aangesloten op de CI voor het ontvangen van aangevraagde datasets.

2.5 Eisen aan privacy-bescherming en informatiebeveiliging⁴

Uitgangspunt 28: Uit overwegingen van privacybescherming zijn door de UMC's opgeleverde gegevens niet direct herleidbaar tot personen. Onder bepaalde, nog te definiëren strikte voorwaarden, is deze herleidbaarheid wel mogelijk (omkeerbare pseudonimiteit).

Architectuurprincipe 3: Data-integratie van datasets van verschillende PARELS en/of verschillende instituten en herleidbaarheid bij omkeerbare pseudonisatie vereisen een eenduidige koppeling van gegevens aan personen. Gegeven het toenemend – en in het zorgdomein zelfs verplicht – gebruik van BSN, vormt het BSN de beste kandidaat voor eenduidige koppeling van gegevens aan personen.

Architectuurprincipe 4: Het BSN in de datasets van de UMC's wordt gepseudonimiseerd door de UMC's voordat die worden verstuurd naar de CI.

Dit architectuurprincipe betekent niet noodzakelijkerwijs dat het BSN door de gehele organisatie en in alle systemen van een UMC moet zijn ingevoerd. Een UMC kan ook er voor kiezen om pas in het proces van samenstellen van een dataset voor aanlevering aan CI het BSN op te zoeken.

Uitgangspunt 29: Er is een juridisch kader voor het gebruik van het BSN binnen het zorgdomein.

Architectuurprincipe 5: In de door de UMC's aangeleverde datasets aan de CI geschiedt het relateren van gegevens aan personen aan de hand van een pseudocode die is afgeleid van het BSN. Eventueel andere persoonsidentificerende gegevens worden verwijderd. Vanwege de vereiste omkeerbare pseudonimiteit is de pseudocode conform een nog vast te stellen procedure herleidbaar naar het oorspronkelijke identificatienummer (BSN).

De pseudocode wordt bepaald op basis van een gestandaardiseerd algoritme die voor alle aanleverende instellingen (UMC's) hetzelfde is. Daardoor zullen verschillende databronnen voor hetzelfde BSN dezelfde pseudocode berekenen en is het mogelijk om binnen de CI gegevens van dezelfde persoon te koppelen op basis van die pseudocode.

Architectuurprincipe 6: Het algoritme om de pseudocode te berekenen is een cryptografisch one-way algoritme. Dit algoritme wordt gestandaardiseerd en dient door alle aanleverende instellingen toegepast te worden.

Architectuurprincipe 7: De CI heeft geen kennis van het BSN. Eventueel terugvertalen van de pseudocode naar een BSN vindt plaats door de UMC's. Het UMC dient te voorzien

⁴ Deze paragraaf is via Kees Louwerse afgestemd met de SIG IB (Special Interest Groep InformatieBeveiliging) van de UMC's

in de mogelijkheid dat een uitgegeven pseudocode kan worden terugvertaald naar het oorspronkelijke BSN.

Architectuurprincipe 8: De pseudonimisatiegegevens worden samen met de overige logginggegevens meerjarig bewaard.

Architectuurprincipe 9: Herleidbaarheid van de uitgevoerde handelingen in de CI wordt gerealiseerd door vastlegging van alle interacties (logging).

Zoals aangegeven in uitgangspunt 10 moet bij het verstrekken van een dataset voor onderzoek zoveel mogelijk voorkomen worden dat verschillende geautoriseerde datasets die ooit beschikbaar gesteld zijn voor verschillende onderzoeken, gekoppeld danwel samengevoegd worden tot nieuwe datasets zonder dat daarvoor toestemming is gegeven.

Architectuurprincipe 10: De dataset die beschikbaar wordt gesteld aan een onderzoeker bevat een ander identificerend gegeven dan de pseudocode die door de UMC's wordt aangeleverd. Dit gegeven is verschillend voor elke dataset die beschikbaar gesteld wordt aan een onderzoeker. Hiervoor wordt een cryptografisch one-way algoritme gebruikt.

Hierbij geldt dat dezelfde patiënt in verschillende afgegeven datasets aan een onderzoeker een verschillend identificerend gegeven heeft waardoor onderlinge koppeling/samenvoeging niet zonder meer mogelijk is. Tevens wordt hiermee een extra maatregel ter bescherming van de privacy gerealiseerd.

Architectuurprincipe 11: Per afgegeven dataset van de CI aan een onderzoeker wordt door CI de gegevens bijgehouden waarmee het identificerend gegeven van de dataset gerelateerd kan worden aan de pseudocode in de CI.

In feite ontstaat in de hele keten UMC-CI-onderzoeker een dubbele pseudocode:

- BSN → pseudocode CI
- pseudocode CI → pseudocode in de afgegeven dataset aan onderzoeker

Bij het inschatten van de risico's m.b.t. de privacybescherming dient de gehele keten in beschouwing te worden genomen. Hoewel de eerste stap van pseudonimisatie vanwege de dataintegratie-eisen in de CI een beperkte bescherming oplevert, is de combinatie van beveiligd transport, een besloten CI-omgeving waartoe alleen de CI-beheerder toegang heeft, en een dubbele pseudonimisatieslag voldoende veilig.

2.6 Inrichtingsvarianten

Uitgangspunt 30: Opleveren van de datasets vanuit de instellingen dient met zo min mogelijk inspanning van de kant van de aanleverende instellingen in het zorgdomein plaats te vinden. Dit geldt zowel voor de realisatiefase als voor de operationele situatie.

Uitgangspunt 31: De opzet dient zodanig te zijn dat met een beperkte invulling al in 2008 de eerste resultaten kunnen worden getoond.

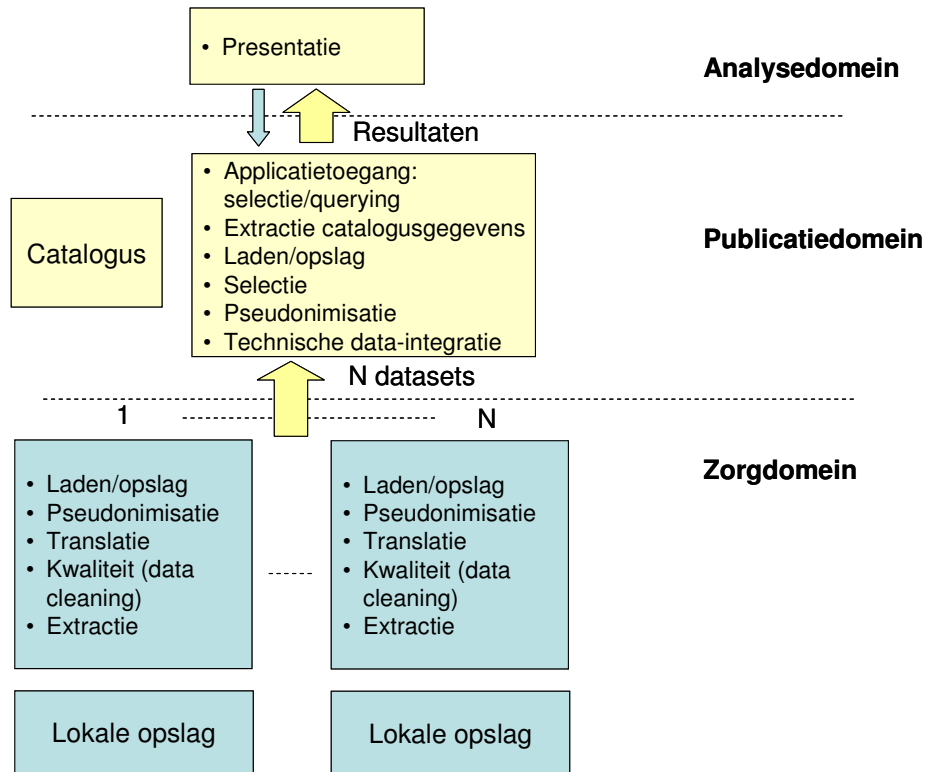
Op hoofdlijnen zijn de volgende implementatievarianten mogelijk:

1. Uitgebreide centrale voorziening: er komt een gemeenschappelijke voorziening die de gegevens verzamelt en bewaart (datawarehouse) met een uitgebreide functionaliteit om te worden bevraagd door de onderzoeker (querying).
2. Beperkte centrale voorziening: er komt een gemeenschappelijke voorziening die de gegevens verzamelt en die na integratie en verdere pseudonimisatie voor de doorgifte naar de onderzoeker zorg draagt.
3. Federatief model waarbij de gezamenlijke voorzieningen bij de UMC's zich gedragen als één "on line" virtueel datawarehouse met uitgebreide functionaliteit.

Bij alle categorieën zijn weer verschillende subvarianten mogelijk. Eerst wordt op hoofdlijnen een keuze gemaakt tussen de hoofdcategorieën, waarna eventuele subvarianten aan de orde kunnen komen.

2.6.1 Uitgebreide centrale voorziening

Om de UMC's zo veel mogelijk te ontlasten kan worden gekozen voor een uitgebreide centrale voorziening met een data warehouse en applicatietoegang met analyse/querying functionaliteit. Dit betekent dat de aanlevering vanuit het zorgdomein, de inrichting van de onderzoeksomgeving in het analysedomein binnen de UMC's eenvoudiger kan worden ingericht. In **Figuur 2** is deze situatie weergegeven. Zoals aangegeven bevat de centrale omgeving, in de figuur aangeduid met de term Publicatiedomein, uitgebreide functies voor o.a data-integratie, eventueel aanvullende pseudonimisatie en applicatietoegang.



Figuur 2: Opzet met een uitgebreide centrale voorziening

Bij de uitgebreide centrale voorziening zal vooral afstemming plaats moeten vinden met de onderzoekers in het analysedomein over de te bieden functionaliteit van de Centrale Infrastructuur. Nieuwe onderzoeksvragen vereisen mogelijk aanpassingen aan datamodel/datamarts om bevestigingen te kunnen optimaliseren. De autorisatie-functie is bij een dergelijk "on line" bevestiging naar verwachting complex. Er zullen centraal uitgebreide voorzieningen moeten worden gerealiseerd.

Een dergelijke invulling van de centrale infrastructuur vereist naar verwachting een meerjarige inspanning alvorens tot resultaten te kunnen komen.

2.6.2 Beperkte centrale voorziening

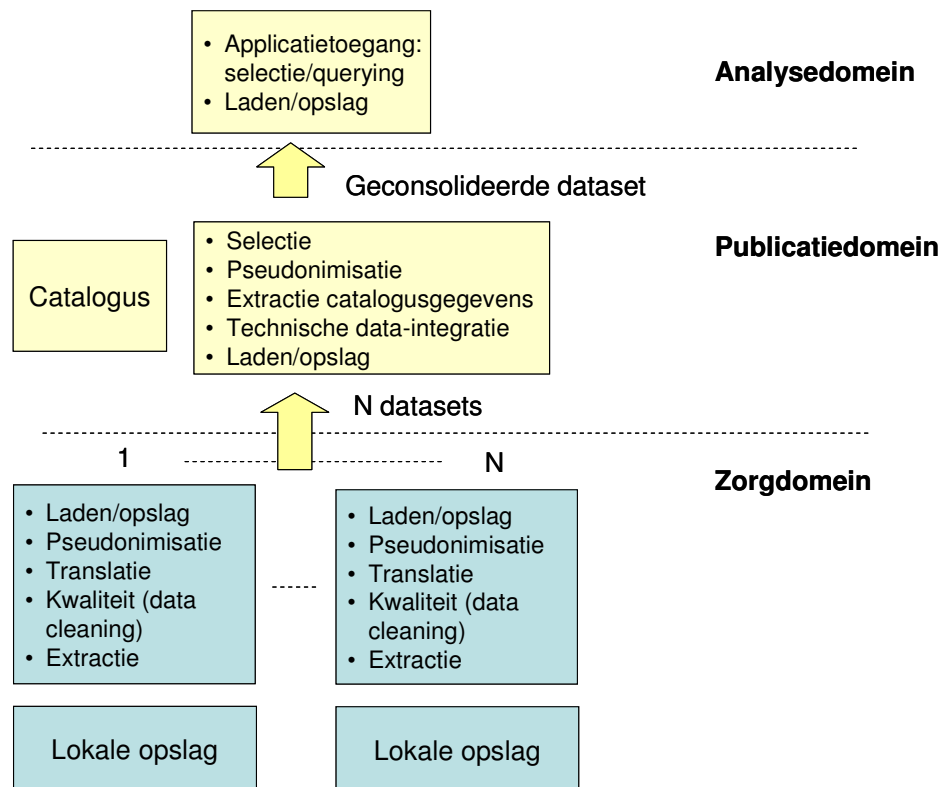
Om de complexiteit van het project te verminderen en toch een ontlasting voor de UMC's te realiseren

kan worden volstaan met een minder uitgebreide centrale voorziening. Een opzet van een dergelijke

invulling is aangegeven in **Figuur 3**.

Zoals aangegeven bevat de centrale omgeving naast de functies voor o.a data-integratie, eventueel aanvullende pseudonimisatie slechts een beperkte functionaliteit. Vanwege de eenvoudiger invulling wordt het mogelijk om snel resultaten te boeken en de toegevoegde waarde zichtbaar en inzichtelijk te maken. Met name voor het scenario 2008/2009 is van belang dat het een beperkt ontwikkeltraject vergt, waardoor realisatie

in 2008/2009 haalbaar wordt. Bovendien is het een minder kostenintensieve oplossing hetgeen van belang is gezien het gelimiteerde budget.

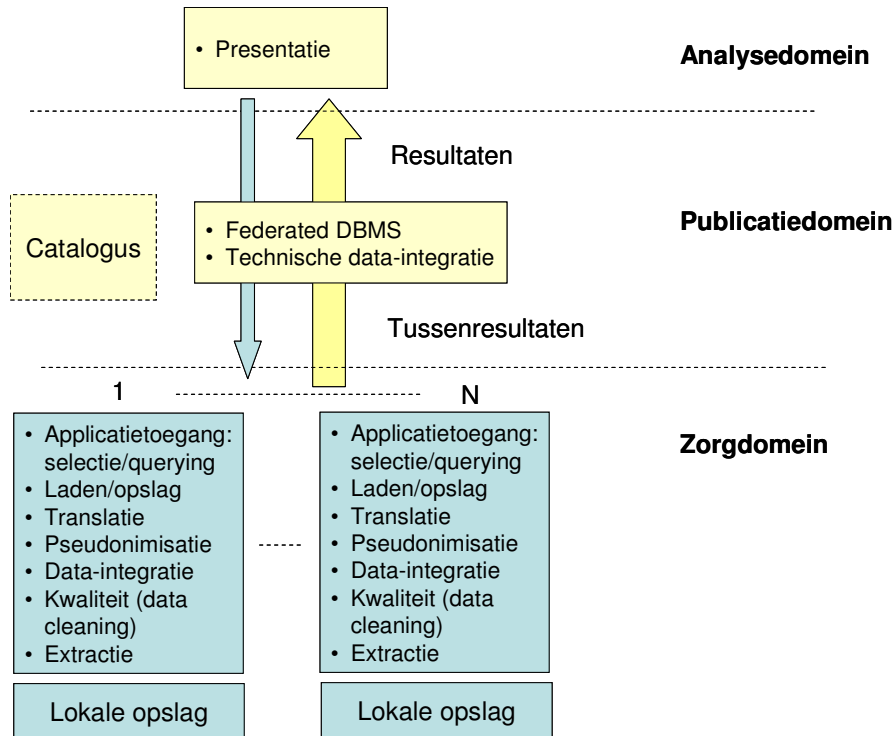


Figuur 3: Opzet met een beperkte centrale voorziening

2.6.3 Federatief model

Tenslotte is er nog de variant mogelijk dat de UMC's functionaliteit beschikbaar stellen teneinde selecties/queries on line uit te voeren op elk van de datasets in hun databases. Hierdoor ontstaat feitelijk een virtuele database/datawarehouse. Deze situatie is weergegeven in **Figuur 4**.

In de centrale infrastructuur is functionaliteit noodzakelijk om de onderliggende UMC databases aan te spreken en zich als een virtuele database te laten gedragen, in de figuur aangegeven met de term "Federated DBMS (Database Management Systeem)". Ook moeten de opgeleverde deelresultaten worden geconsolideerd door middel van data-integratie. Dit alles moet als een centrale functie worden aangeboden om aan de onderzoekers één dataset te kunnen aanleveren.



Figuur 4: Federatief model

Het federatieve model wordt vooral ingezet in bedrijfsomgevingen waar het van belang is actuele bedrijfsinformatie uit diverse bronnen direct beschikbaar te hebben. In een dergelijke situatie voldoet het periodiek uitwisselen van datasets niet meer.

Deze variant vraagt een uitgebreide standaardisatie en afstemming tussen de UMC's en vereist een uitgebreide lokale infrastructuur in alle aangesloten UMC's. De UMC's moeten hun databases op gestandaardiseerde, beveiligde wijze real-time benaderbaar maken voor de CI. De CI zal hierbij centrale regie moeten voeren voor beheer met impact op lokale infrastructuur. Dit is derhalve de meeste complexe van de beschouwde varianten.

Het federatieve model is minder geschikt voor toepassing binnen Parelsnoer vanwege de volgende redenen:

- Bij het federatief model zal veel afstemming met de acht UMC's moeten plaatsvinden over standaarden en technieken aangezien het model de koppeling van heterogene databases van mogelijk verschillende leveranciers noodzakelijk maakt. Hoewel de technologie hiervoor beschikbaar is (zogenaamde "wrapper" technologie) is de implementatie hiervan complex, duur en tijdrovend.
- In het analyse domein zijn nog onzekerheden ten aanzien van de frequentie van het gebruik, het soort vragen/queries, de verwachte grootte van de datasets in de toekomst en de vereiste functionaliteit die nu afstemming en standaardisatie bemoeilijken.

- Het model vereist dat alle UMC's zover zijn om op een redelijke termijn hun gegevens "on line" beschikbaar te kunnen en willen stellen.
- De UMC's zullen interne databases voor de diverse parels ofwel moeten integreren en extern toegankelijk moeten maken of alle databases individueel extern toegankelijk moeten maken. Intensieve externe queries zullen de performance van de systemen kunnen beïnvloeden.
- Nieuwe onderzoeksvragen vereisen in de toekomst mogelijk aanpassingen aan datamodel/datamarts om de databases te kunnen optimaliseren voor intensieve bevragingen. Dergelijke aanpassingen moeten door alle UMC's worden doorgevoerd. Bovendien zullen ook alle nieuw aan te sluiten bronnen moeten voldoen aan het federatieve model.
- Ook de beveiliging en autorisatie-functies zijn bij een dergelijk "on line" bevraging naar verwachting complex. Er zullen hiervoor lokaal uitgebreide voorzieningen moeten worden gerealiseerd bij de UMC's.

2.6.4 Evaluatie

Het realiseren van Parelsnoer vraagt veel activiteiten van de UMC's:

- Het afronden van het bestuurlijke proces voor het beschikbaar stellen van financiële middelen en additionele menscapaciteit ("research nurses");
- Het afstemmen van de domeinen ICT, zorg en onderzoek
- Het mogelijk aanpassen van het onderzoek- en registratieproces
- Het mogelijk implementeren van nieuwe kwaliteitsnormen en informatiedefinities
- Het bouwen van een extern koppelvlak

De uitgebreide centrale voorziening en het federatieve model voegen hier nog extra activiteiten voor afstemming en implementatie aan toe. De beheersbaarheid van een dergelijke grootschalig en meerjarig project is voor zowel de uitgebreide centrale voorziening als het federatieve model problematisch en het project kent derhalve een hoog afbreukrisico.

Omdat in 2008 voor Parelsnoer de eerste resultaten moeten kunnen worden getoond (Uitgangspunt 16) vallen vanwege bovengenoemde redenen het federatief model en het uitgebreide centrale voorzieningenmodel af.

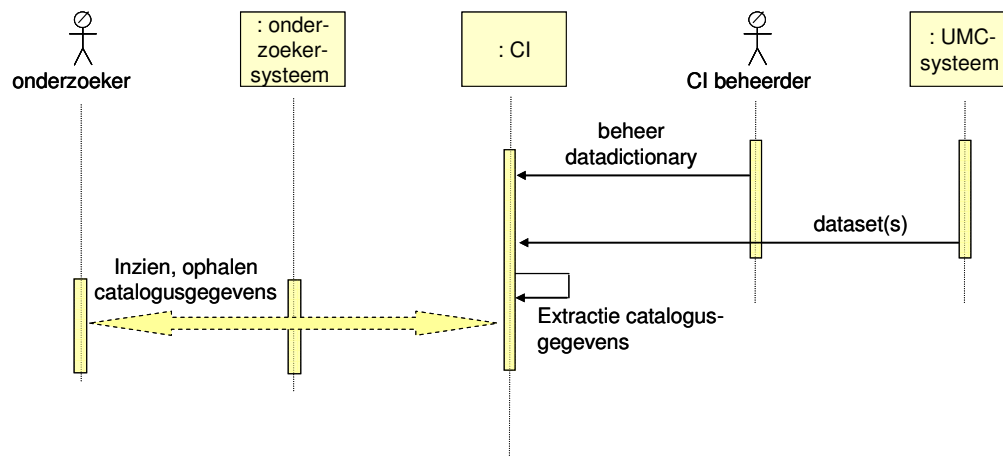
Architectuurprincipe 12: De startarchitectuur voor Parelsnoer wordt gebaseerd op de beperkte centrale voorzieningen variant, gebaseerd op het periodiek "pushen" van datasets.

2.7 Het uitwisselingsproces

2.7.1 Vullen, inzien, ophalen van catalogusgegevens

Het UML sequence diagram voor vullen, inzien en ophalen van catalogusgegevens wordt weergegeven in **Figuur 5**. Hieraan wordt aangegeven dat:

1. Gegevens in de CI actueel worden gehouden door:
 - a) V.w.b. de data dictionary voert een centrale CI-beheerder het beheer; de gegevens worden via de catalogus beschikbaar gesteld.
 - b) UMC's leveren op reguliere basis datasets aan; de karakteristieken van de beschikbare datasets, beeldmateriaal en biomateriaal worden binnen CI afgeleid en via de catalogus beschikbaar gemaakt voor onderzoekers/coördinatoren.
2. De catalogusinformatie algemeen toegankelijk is.

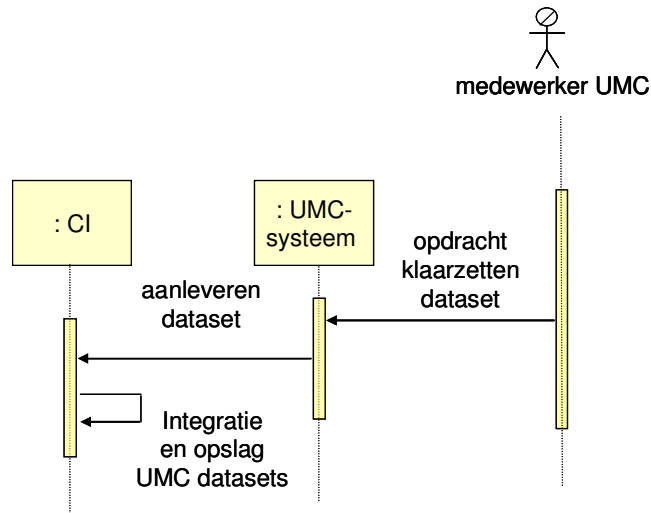


Figuur 5: UML sequence diagram voor vullen, inzien en ophalen van catalogusgegevens

2.7.2 Uitwisseling van datasets

Het vullen van de CI met datasets door de UMC's wordt getoond in **Figuur 6**. Daarin zijn de volgende stappen te onderkennen:

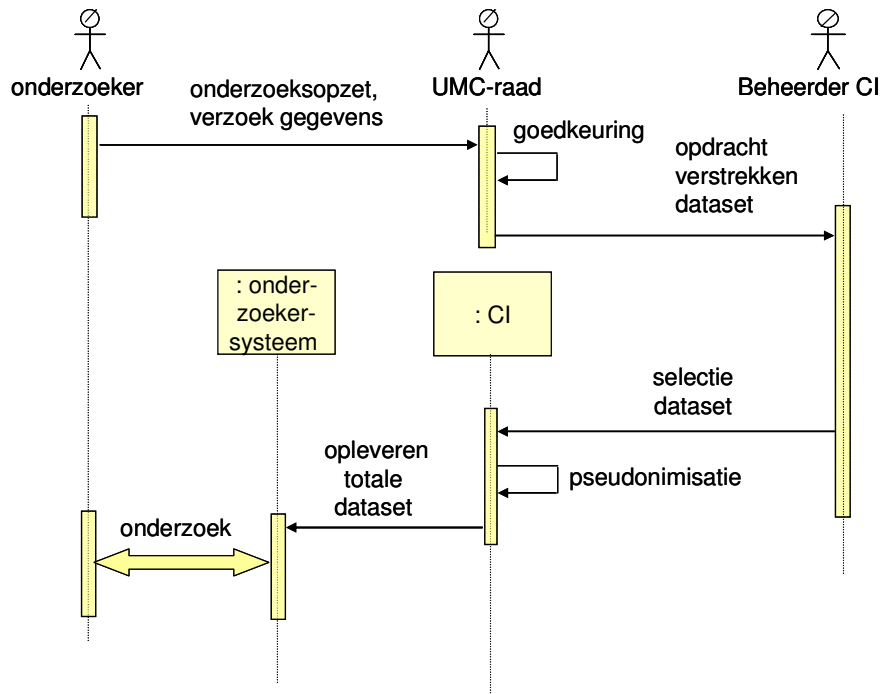
- 1) Zorginformatie wordt door de UMC's lokaal vastgelegd in eigen informatiesystemen.
- 2) De UMC's leveren periodiek nieuwe volledige klinische datasets aan. Aanlevering van datasets vindt plaats conform gemaakte afspraken (formats, tijdstip, etc..). Datasets zijn sec PAREL gerelateerd
- 3) De Centrale Infrastructuur (CI) importeert de datasets van de UMC's en zorgt voor technische integratie van de opgeleverde gegevens.



Figuur 6: UML sequence diagram voor het vullen van de CI met datasets door de UMC's

De stappen die leiden tot het beschikbaar stellen van datasets aan onderzoekers worden getoond in **Figuur 7**. Daarin staat aangegeven:

- 1) Een onderzoeker kan een onderzoeksvoorstel indienen bij wetenschappelijke commissie van een Parel met een verzoek om hiervoor een bepaalde onderzoeksdataset beschikbaar te stellen
- 2) Na expliciete toestemming wordt door de beheerder CI de juiste selectie aan de desbetreffende onderzoeker beschikbaar gesteld.
- 3) De onderzoeker ontvangt de gevraagde dataset en wordt gebruiker van deze gegevens.



Figuur 7: UML sequence diagram voor het beschikbaar stellen van datasets aan onderzoekers

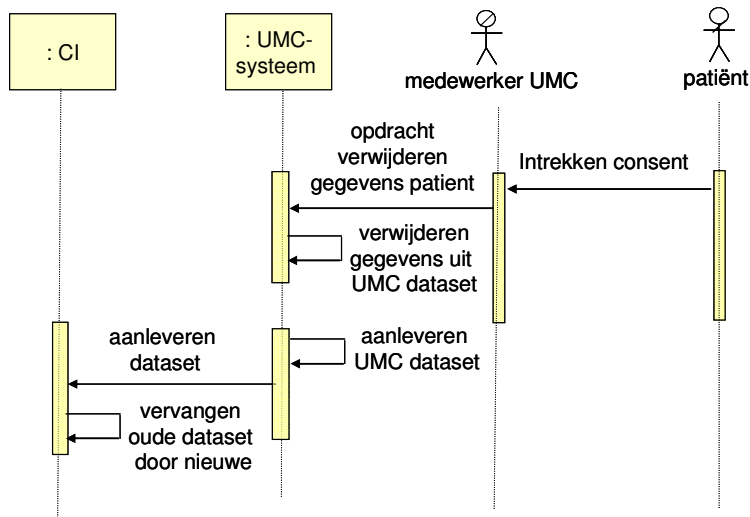
2.7.3 Verwijderen patiëntgegevens uit UMC-dataset

In het UML sequencediagram van **Figuur 8** zijn de volgende stappen te onderkennen:

- 1) De patiënt trekt zijn toestemming in voor het gebruik van zijn gegevens;
- 2) De verantwoordelijke zorgverlener/medewerker UMC geeft opdracht de desbetreffende gegevens te verwijderen uit de klinische dataset(s) van de UMC;
- 3) Bij de eerstvolgende aanlevering van dataset(s) van de desbetreffende UMC aan de CI zijn de gegevens van de desbetreffende patiënt niet meer in de dataset aanwezig.

Architectuurprincipe 13: UMC's leveren per Parel en periodiek een complete klinische dataset aan. Bij elke nieuwe aanlevering wordt de oude dataset uit de CI verwijderd en vervangen door de nieuwe dataset.

Na lokale verwijdering van patiëntgegevens wordt deze automatisch verwijderd in de CI met de eerstvolgende aanlevering. Dit komt doordat de oude dataset wordt verwijderd en vervangen door de nieuwe dataset.



Figuur 8: UML sequence diagram voor het verwijderen patiëntgegevens uit UMC-dataset

2.7.4 Aanvragen biomateriaal

P.m.

2.7.5 Aanvragen beeldmateriaal

P.m.

2.7.6 Architectuurinvulling m.b.t. het uitwisselingsproces

Architectuurprincipe 14: Per UMC wordt altijd de gehele klinische dataset per Parel aangeleverd.

Motivatie: dit reduceert de werkzaamheden per UMC aangezien niet per onderzoeksvraag een nieuwe selectie moet worden gemaakt

Architectuurprincipe 15: De CI leidt de karakteristieken van de klinische datasets af van de aangeleverde datasets en werkt de catalogus zelf bij.

Architectuurprincipe 16: Verschillende gegevens van dezelfde persoon die vanuit verschillende bronnen zijn aangeleverd kunnen centraal gekoppeld worden op basis van de door de bronnen berekende pseudocode.

Architectuurprincipe 17: In de CI worden alle transacties geregistreerd (gelogd). Daarbij worden ook de generatiedatum en versienummer van ontvangen en uitgegeven datasets vastgelegd. Ook de pseudonimisatiegegevens worden bewaard zodat een terugvertaling naar het bijbehorende BSN mogelijk blijft

Merk hierbij op dat de uiteindelijke vertaling naar BSN door de UMC's gedaan moet worden. De CI beschikt niet over BSN's.

Uitgangspunt 32: CI produceert managementinformatie, zoals bijvoorbeeld:

1. Per PAREL het gebruik, de kwaliteit etc.;
2. Gebruiksstatistieken van de catalogus;
3. Statistiek van aanleveringen door UMC's.

Architectuurprincipe 18: De onderzoeker krijgt één dataset aangeleverd. De onderzoeker heeft een eigen onderzoekssysteem en zal alle analyses en bewerkingen van de data in zijn onderzoeksomgeving (geen onderdeel van CI) uitvoeren.

3 Informatie (systeem) architectuur

3.1 Inleiding

Betekenis en structuur van gegevens (semantiek en syntax) en de functionaliteit van applicaties voor informatieverwerking worden in de informatie (systeem) architectuur gedefinieerd.

3.2 Functionaliteit van de gemeenschappelijke voorzieningen

Op hoofdlijnen zal de volgende functionaliteit van de gemeenschappelijke voorzieningen in de Centrale Infrastructuur noodzakelijk zijn:

- **Ontvangen, verwerken en beheren van klinische datasets van UMC's**
Door UMC's aangeleverde datasets worden geïntegreerd tot 1 dataset per PAREL. Deze wordt vastgelegd in de CI.
- **Beschikbaar stellen van onderzoeksdatasets.**
Na goedkeuring door de wetenschappelijk commissies van de PAREL wordt in de CI de geaccordeerde selectie uitgevoerd om de gewenste gegevens beschikbaar te kunnen stellen aan de onderzoeker(s).
- **Catalogus**
In deze catalogus worden de karakteristieken van de beschikbare datasets bijgehouden. Ook worden de details van het Parelsnoer Informatie Model (datadictionary) via de catalogus beschikbaar gesteld.
- **Voorzieningen voor beveiligde communicatie**
Er dient gebruik te worden gemaakt van beveiligde communicatie tussen de verschillende domeinen voor de ondersteuning van privacybescherming en informatiebeveiliging.

Architectuurprincipe 19: de Centrale Infrastructuur biedt de volgende diensten aan de UMC's:

- Ontvangen, verwerken en beheren van datasets;
- Beschikbaar stellen van datasets aan onderzoekers;
- Een catalogus met informatie over de beschikbare datasets, beeldmateriaal en biomateriaal;
- Voorzieningen voor beveiligde communicatie.
- Management rapportages

Architectuurprincipe 20: CI dient Identificatie, Authenticatie en Autorisatie te verzorgen van systemen die datasets beschikbaar stellen en datasets ontvangen.

Architectuurprincipe 21: CI dient alle interacties vast te leggen in een logbestand.

3.3 Informatiemodellen: semantiek en syntax

Architectuurprincipe 22: Gegevens worden uitgewisseld conform het Parelsnoer Informatie Model (PIM). Dit is een gemeenschappelijk referentie informatie model binnen

Parelsnoer; voor opstellen van dit model wordt zoveel mogelijk gebruik gemaakt van bestaande standaarden zoals HL7v3 RIM (inclusief tooling en templates), SNOMED-CT⁵ en LOINC⁶.

Architectuurprincipe 23: Voor afgifte van de dataset aan de onderzoeker wordt in de CI het gepseudonimiseerde BSN nogmaals gepseudonimiseerd met een voor dat onderzoek uniek te gebruiken sleutel.

Bij het modelleren van de dataset wordt rekening gehouden dat in de CI geen andere algemene patiëntgegevens worden vastgelegd waaruit direct afgeleid zou kunnen worden welke patiënt het betreft, zoals adres en woonplaats.

Architectuurprincipe 24: Ten behoeve van het versiemangement wordt in elke datasets een generatiedatum en een versienummer opgenomen.

Architectuurprincipe 25: Voor de uitwisseling van de datasets tussen zorgdomein en CI en tussen CI en analysedomein wordt gebruik gemaakt van een gestandaardiseerd bestandsformaat

Uitgangspunt 33: Er wordt gestreefd naar één technische protocol op basis van webtechnologie voor uitwisseling van onderzoeksgegevens. Voor 2008/2009 wordt aangenomen dat tijdelijk meer dan één protocol mogelijk kan zijn, afhankelijk van de mogelijkheden van de UMC's.

Architectuurprincipe 26: Ondersteuning van de volgende formaten wordt voor 2008/2009 minimaal ondersteund:

- Comma Separated Value (CSV) en/of SPSS file structuur;
- Gestandaardiseerde XML-berichten zomogelijk conform CDISC⁷.

In de CI zal gecontroleerd worden of de aangeleverde bestanden voldoen aan de structuurafspraken (syntaxt). Er worden geen semantische controles voorzien die de CI uitvoert op een ontvangen bestand.

3.4 Eisen aan koppelvlakken met instellingssystemen

Architectuurprincipe 27: De UMC's beschikken over één of meer informatiesystemen waarmee interactie plaatsvindt conform de vastgestelde specificaties van technische uitwisselingsstandaarden en semantische en syntactische definities.

3.5 Functionaliteit voor beveiliging en beheer.

Architectuurprincipe 28: Vertrouwelijkheid van de gegevens wordt gewaarborgd door toegangsbeveiliging in combinatie met transportbeveiliging (versleuteling) bij overdracht.

⁵ SNOMED CT is de 'Systematized Nomenclature Of MEDicine, Clinical Terms'.

⁶ LOINC staat voor Logical Observation Identifiers Names and Codes.

⁷ Clinical Data Interchange Standards Consortium.

Architectuurprincipe 29: Indien wordt gecommuniceerd met systemen dient de authenticiteit van deze systemen te worden vastgesteld (systeemauthenticatie).

3.6 Functionaliteit voor communicatie

Architectuurprincipe 30: In de communicatie tussen het zorgdomein en CI wordt de volgende functionaliteit in de CI voorzien:

- Monitoring van het proces van aanlevering;
- Controle op aanlevering;
- Logging.

Architectuurprincipe 31: In de communicatie tussen CI en onderzoekers wordt de volgende functionaliteit in de CI voorzien:

- Samenstellen van datasets voor onderzoekers;
- Pseudonimisatie van datasets voor onderzoekers;
- Transport van datasets naar onderzoekers;
- Logging.

4 Technische architectuur

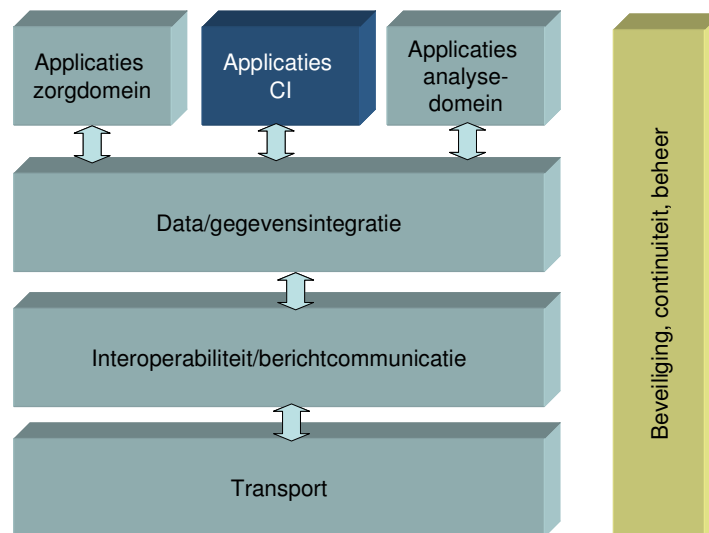
4.1 Inleiding

Dit hoofdstuk bevat een eerste kaderzetting voor de invulling van de technische architectuur.

Uitgangspunt 34: Voor 2008/2009 is de relatie met BIG GRID out of scope. In het voorjaar 2008 start een verkennend onderzoek voor het uitwisselen van beeldmateriaal.

4.2 Het servicemodel

Als basismodel wordt gebruik gemaakt van een 4-laags service model, een gelaagde hiërarchie waarbij componenten uit een hogere laag een 'service' vragen aan een lagere laag.



Figuur 9: Het servicemodel

De getoonde lagen hebben de volgende functionaliteit:

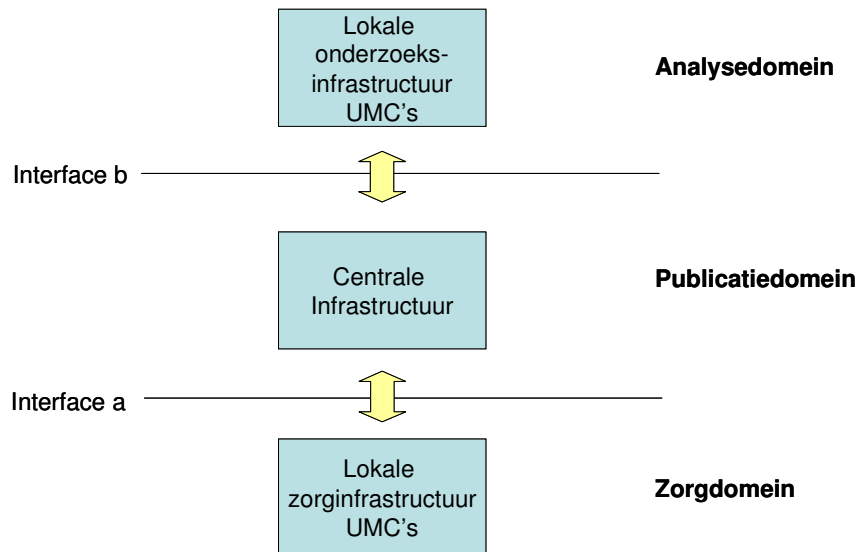
1. Toepassingen (ook wel de applicaties genoemd), onder te verdelen in:
 - a) applicaties zoals die bij de UMC's worden gebruikt, zowel in het zorgdomein als in het analysedomein;
 - b) infrastructurele toepassingen, d.w.z. functies in de CI.
2. Data/gegevensintegratie voor consolidatie van datasets van de Parels
3. Interoperabiliteit/berichtcommunicatie, voor het ondersteunen van interacties tussen systemen.
4. Transport, de onderliggende netwerktechnologie.

Daarnaast zijn er bepaalde gebieden, beveiliging, continuïteit en beheer, die niet specifiek aan een bepaalde laag kunnen worden toegewezen, maar juist in hun

samenhang voor het gehele ICT-bouwwerk moeten worden beschouwd. Dit is in het model aangegeven door een aparte kolom naast de vier lagen te plaatsen.

4.3 Interfacemodel

In **Figuur 10** wordt het Interfacemodel van Parelsnoer getoond. In de figuur wordt aangegeven dat er sprake is van twee verschillende Interfaces a en b.



Figuur 10: Interfacemodel

Netwerk/transportniveau

Architectuurprincipe 32: Voor transport wordt voor beide interfaces gebruik gemaakt van Internetstandaarden (TCP/IP).

Motivatie: op het netwerkniveau is de huidige situatie dat de partijen al op basis van Internet-protocollen (TCP-UDP/IP) communiceren.

Interoperabiliteit/berichtcommunicatie

Voor 2008/2009 wordt aangenomen dat meer dan één protocol voor het opsturen en ophalen van datasets noodzakelijk is, afhankelijk van de mogelijkheden van de UMC's. Op termijn wordt gestreefd naar één protocol op basis van webtechnologie.

Architectuurprincipe 33: Voor 2008/2009 wordt gedacht aan de protocollen HTTP en FTP. Voor kleine aantallen kan ook SMTP overwogen worden.

Architectuurprincipe 34 (toekomst): Het uitwisselen van datasets zal worden gebaseerd op web services.

De keuze voor web services betekent uitwisseling van XML berichten met gebruik van SOAP en transport via http en/of smtp. De concrete invulling van de web services uitwisselingsstandaard dient nog nader te worden uitgewerkt.

4.4 Beveiliging, continuïteit en beheer

Naast het creëren van een veilige netwerkomgeving en veilig en betrouwbaar berichtenverkeer dienen de infrastructurele voorzieningen en lokale systemen te voorzien in adequate beveiligings-maatregelen.

Architectuurprincipe 35: Voor invulling van de voorzieningen voor beveiligde communicatie wordt het gebruik van certificaten conform X.509 voorzien. Hierbij wordt gebruik gemaakt van bestaande Certificate Authorities.

Architectuurprincipe 36: Transportbeveiliging bij uitwisseling van datasets vindt plaats op basis van SSL/TLS met tweezijdige authenticatie en gebruik van systeemcertificaten.

Uitgangspunt 35: SSL biedt een voldoende beveiligingsniveau voor het uitwisselen van datasets, er zijn geen aanvullende beveiligingsmaatregelen noodzakelijk om te kunnen voldoen aan de gestelde beveiligingseisen.

Architectuurprincipe 37: Voor browsertoegang tot de catalogus wordt gebruik gemaakt van enkelzijdig SSL/TLS op basis van het certificaat van de gemeenschappelijke voorzieningen.

Architectuurprincipe 38: Vanwege de zeer verschillende beveiligingseisen en toegangs-mogelijkheden dient de functionaliteit voor catalogus en voor opslag en verwerking van datasets op fysiek gescheiden systemen te worden geïmplementeerd

Architectuurprincipe 39 (toekomst): Bij web services interacties vindt transportbeveiliging en authenticatie plaats op basis van tweezijdig SSL/TLS met gebruikmaking van een systeemcertificaten.

5 Ontwikkelingsscenario's

5.1 Inleiding

In dit architectuurontwerp wordt gekozen voor een pragmatische invulling waarmee op afzienbare termijn de eerste concrete resultaten kunnen worden getoond. Voor de langere termijn wordt een uitbreiding van de mogelijkheden voorzien. Deze worden in dit hoofdstuk kort geschetst.

5.2 Rijkere zorg-omgeving

1. Ondersteunende gedistribueerde tools/faciliteiten voor onderzoekers
2. Koppelen met externe onderzoeksdatabases/biobanken
3. Integratie met het EPD

5.3 Rijkere publicatie-omgeving

1. Integratie op applicatieniveau:
 - a) Afleveren aan onderzoeker
 - b) Ondersteunende tools/faciliteiten voor onderzoeker
2. Ondersteuning van de mogelijkheden voor datamining
3. Centrale invoer en onderhoud van de catalogus voor biomaterialen
4. Verrijking van de gegevens in de catalogus

6 Literatuur

- [1]. Werkplan Parelsnoer Initiatief, Aanpak, activiteiten en begroting behorende bij het uitvoeringsprogramma van het Parelsnoer Initiatief, versie 1.1
- [2]. Reglement waarin ethische en juridische aspecten zijn geadresseerd. Versie 5 nu beschikbaar; definitieve document op te leveren eind oktober.