

# Programma van Eisen Instellingen

## Parelsnoer PvE Instellingen



PARELSNOER INITIATIEF

Versie: 1.2  
Status: definitief  
Datum: 19 september 2008

# Inhoud

1	Inleiding .....	3
1.1	Positionering van dit programma van eisen .....	3
1.2	Uitgangspunten.....	4
1.3	Leeswijzer .....	4
1.4	Gerelateerde documenten .....	4
1.5	Versiehistorie.....	5
2	Context.....	6
2.1	Parelsnoer .....	6
2.2	Project Parelsnoer Centrale Infrastructuur .....	6
2.3	Keuzes in de opzet van de centrale voorzieningen .....	7
2.3.1	Bundelen en Centrale opslag.....	7
2.3.2	Privacy en beveiliging .....	8
2.3.3	Parelsnoer Informatiemodel .....	8
2.4	Globale werking van de centrale voorzieningen .....	9
2.4.1	Informeren .....	9
2.4.2	Bundelen.....	10
2.4.3	Uitleveren .....	10
3	Processen.....	11
3.1	Aansluiten op de Centrale Infrastructuur .....	11
3.2	Aanleveren van onderzoeksgegevens.....	11
3.3	Aanleveren van beeldmateriaal .....	12
3.4	Aanleveren van biomateriaal.....	12
3.5	Afhandelen verzoek de-pseudonimisatie.....	13
3.6	Verwijderen van patiëntgegevens uit de CI.....	13
4	Informatie en functionaliteit .....	14
4.1	Samenstellen dataset .....	14
4.2	Pseudonimiseren en de-pseudonimiseren .....	14
4.2.1	Bepalen gepseudonimiseerd BSN.....	14
4.2.2	De-pseudonimisatie .....	15
4.2.3	Algoritme berekenen gepseudonimiseerd BSN.....	15
5	Techniek .....	16
5.1	Netwerkverbinding .....	16
5.2	Tweezijdige SSL.....	16
5.3	Koppelvlak datasets.....	16
5.4	Koppelvlak beeldmateriaal.....	16
5.5	Eisen aan beveiliging certificaten.....	17
6	Afspraken.....	18
6.1	Frequentie van aanlevering .....	18
6.2	Beschikbaarheid.....	18
6.2.1	Kwaliteit van gegevens .....	18

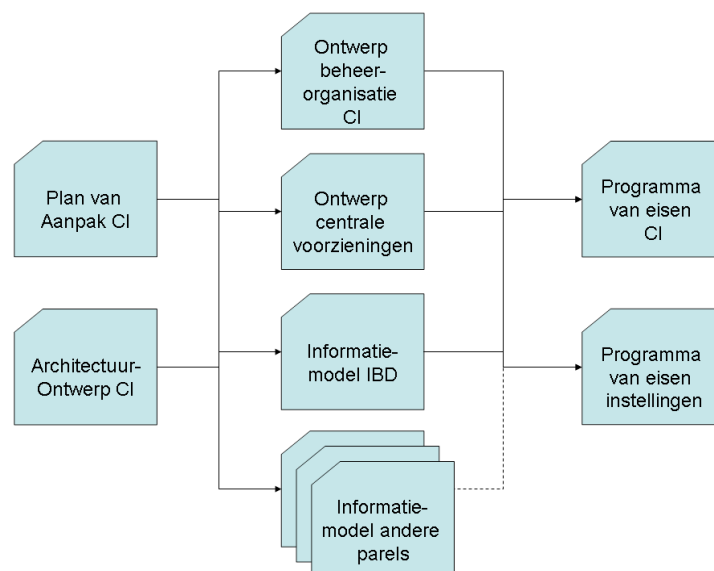
# 1 Inleiding

Het project Parelsnoer beoogt<sup>1</sup> binnen de academische centra een infrastructuur te realiseren voor het verzamelen van klinische data, beeldmateriaal en opzetten van biobanken op interuniversitair niveau. Voor de realisatie van de benodigde centrale voorzieningen is niet alleen een ontwerp van de techniek en de beheerorganisatie benodigd, ook de verwachtingen aan een instelling moeten expliciet gemaakt worden. Het voorliggende document betreft het programma van eisen voor instellingen binnen Parelsnoer.

In dit document wordt de term instelling gebruikt voor de aangesloten academische centra en niet UMC. Er is voor deze generieke term gekozen omdat op termijn ook externe partijen (niet UMC) aangesloten kunnen.

## 1.1 Positionering van dit programma van eisen

Dit programma van eisen is een van de documenten die binnen het Parelsnoer Centrale Infrastructuur project is opgesteld. Onderstaande figuur geeft een overzicht van alle ontwerpdocumenten binnen het project:



In de figuur is de volgorde van de documenten aangegeven. Dit wordt nader toegelicht in het Plan van Aanpak CI.

De volgende documenten zijn opgenomen in bovenstaande figuur:

- *Architectuurontwerp CI*  
Dit architectuurontwerp bevat alle kaders en principes die gehanteerd worden bij de totstandkoming van de centrale infrastructuur. Het architectuurontwerp is uitgangspunt bij het opstellen van het ontwerp.
- *Plan van aanpak CI*  
Dit document beschrijft de aanpak voor de realisatie van de centrale infrastructuur. Het plan beschrijft een incrementele aanpak voor de realisatie van de centrale infrastructuur. Dit ontwerp dient ook gezien te worden als het ontwerp voor de eerste increment. Voor de komende incrementen kan dit plan worden bijgewerkt.

<sup>1</sup> Uit "Werkplan Parelsnoer"

- *Ontwerp Centrale Voorzieningen*  
Dit is het ontwerp van de centrale infrastructuur. Het document vormt samen met het programma van eisen CI de basis voor de realisatie van de CI. Dit document in combinatie met het programma van eisen Instellingen vormt, aangevuld met eigen ontwerpdocumentatie, de basis voor de aansluiting van de instellingen op de CI.
- *Ontwerp Beheerorganisatie CI*  
De CI betreft naast techniek ook een beheerorganisatie voor het uitvoeren van het beheer en het leveren van diensten aan onderzoekers. Deze organisatie is in dit document beschreven.
- *Parelsnoer InformatieModel (PIM)*  
De PIM is een gegevensontwerp voor het vastleggen en uitwisselen van gegevens over Parels. De PIM is de basis voor de inrichting van de database van de centrale infrastructuur. Tevens is de PIM de afspraak volgens welke alle instellingen gegevens gaan uitwisselen met de centrale infrastructuur. Gestart wordt met IBD en daarop zullen snel de andere parels ook volgen.
- *Programma van eisen CI*  
Het programma van eisen CI vormt de basis van de realisatie van de centrale infrastructuur. Het bevat de functionele eisen (die uitvoerig beschreven zijn in het Ontwerp CI) en de niet-functionele eisen (eisen zoals schaalbaarheid, performance, beheerbaarheid, etc.)
- *Programma van eisen Instellingen*  
Het programma van eisen Instellingen, het voorliggende document, is de basis voor de aansluiting van een Instelling op de CI. Het bevat alle functionele en niet-functionele eisen die gesteld worden aan de Instellingen om te kunnen aansluiten op de CI.

## 1.2 Uitgangspunten

- In het architectuurontwerp worden uitgangspunten beschreven voor zowel 2008/2009 als mede voor daarna (geannoteerd met toekomst). Dit ontwerp gaat alleen uit van de uitgangspunten voor 2008/2009.
- De definitie van de PIM is zodanig dat de dataset (met uitzondering van het BSN) voldoende anoniem is. Hierdoor kan worden volstaan met alleen het vervangen van het BSN door een gepseudonimiseerd BSN.

## 1.3 Leeswijzer

Bij het lezen van dit programma van eisen wordt er vanuit gegaan dat de lezer bekend is met het architectuurontwerp en met het ontwerp van de CI.

In hoofdstuk 2 wordt eerst de context van Parelsnoer en de globale werking beschreven. Daarna worden de eisen aan instellingen beschreven vanuit vier verschillende perspectieven: procesmatig (hoofdstuk 3), functioneel/informatie (hoofdstuk 4), techniek (hoofdstuk 5) en ten slotte in hoofdstuk 6 algemene afspraken.

## 1.4 Gerelateerde documenten

- Gerard van der Hoorn, Reinier Balt, Parelsnoer Centrale Infrastructuur: Projectplan 2007-2010. Versie 0.10. Amsterdam: Parelsnoer Initiative, 3 december 2007.
- Reinier Balt, Frans van den Dool, Architectuur Centrale Infrastructuur: Parelsnoer, Versie 1.1. Amsterdam: Parelsnoer Initiative, 30 november 2007.
- Reinier Balt, Joost Schalken, Eelco den Heijer, Ontwerp Centrale Voorzieningen Parelsnoer, versie 0.4. Amsterdam: Parelsnoer Initiative, 20 december 2007.

- Reinier Balt, Joost Schalken, Eelco den Heijer, Programma Van Eisen Instelling: Parelsnoer PvE Centrale Infrastructuur. Versie 0.0. Amsterdam: Parelsnoer Initiative, Q1 2008 (PM).
- Reinier Balt, Joost Schalken, Eelco den Heijer, Ontwerp Beheerorganisatie Centrale Voorzieningen, versie 0.0. Amsterdam: Parelsnoer Initiative, Q1 2008 (PM).
- Miranda de Gouw en Sandra Sliepenbeek, Parelsnoer Informatie Model (PIM). Versie 0.0. Amsterdam: Parelsnoer Initiative, Q1 2008 (PM).

## 1.5 Versiehistorie

Datum	Versie / Status	Opmerkingen
	0.2	Initiële concept versie
20 dec 2007	0.3	Concept versie waarin commentaar Parelsnoer CI-team verwerkt is.
8 jan 2008	0.4	Commentaar Gerard van der Hoorn verwerkt
10 maart 2008	0.9	Nieuw commentaar verwerkt
20 maart 2008	1.0	Laatste scan Gerard verwerkt
4 april 2008	1.01	Finaliseren
10 sept 2008	1.2	Omzetting huisstijl; geen inhoudelijke wijzigingen

## 2 Context

### 2.1 Parelsnoer

In het huidige medisch wetenschappelijke onderzoek speelt de relatie tussen fundamentele kenmerken (genen) en processen in het lichaam (eiwitten, metaboliëten), 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.

Het project Parelsnoer beoogt binnen de academische centra een infrastructuur te realiseren voor het verzamelen van klinische data, beeldmateriaal 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. 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 (CVA)
- inflammatoire darmziekten (IBD)

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

### 2.2 Project Parelsnoer Centrale Infrastructuur

Voor het bereiken van de doelstellingen van Parelsnoer is het bundelen van de onderzoeksgegevens van de deelnemende instellingen nodig. Om dit mogelijk te maken is een centrale infrastructuur voorzien die het mogelijk maakt om de verschillende onderzoeksgegevens te bundelen en beschikbaar te stellen aan geautoriseerde onderzoekers. Voor de realisatie van deze centrale infrastructuur is het (deel)project Centrale Infrastructuur gedefinieerd. Dit project heeft als doelstelling:

- Een centrale voorziening in te richten waar onderzoekers terecht kunnen met
  - een geautoriseerd verzoek voor een verzameling met klinische gegevens en/of beeldmateriaal ten behoeve van onderzoek naar bepaalde ziektebeelden. Op basis van dit geautoriseerde verzoek wordt de verzameling samengesteld en aangeleverd aan de onderzoeker

- o een geautoriseerd verzoek voor biomateriaal dat is opgeslagen in de aangesloten instellingen. Op basis van dit geautoriseerde verzoek wordt het biomateriaal opgehaald en afgeleverd bij de onderzoeker. De informatiestroom voor dit proces is onderdeel van dit project. Het fysieke transport van het biomateriaal niet.

Ten behoeve van deze verzoeken wordt een catalogus beschikbaar gesteld waarmee een onderzoeker kan bepalen welke onderzoeksgegevens w.o. beeldmateriaal en biomateriaal beschikbaar is

- Een infrastructuur te realiseren die het mogelijk maakt om de bronnen met klinische onderzoeksgegevens van de verschillende UMC's voor ziektebeelden te ontsluiten ten behoeve van andere onderzoekers
- Een referentiemodel te introduceren voor de structuur van de onderzoeksgegevens die over de infrastructuur worden uitgewisseld. Dit referentiemodel is een integraal informatie model over de ziektebeelden heen en biedt een model voor een specifiek ziektebeeld.

De CI zal niet alleen een technisch platform zijn. Rondom de CI wordt een beheerorganisatie ingericht. Deze organisatie heeft tot doel om onderzoekers te voorzien van onderzoeksgegevens en biomateriaal, het beheren van de technische infrastructuur, het bewaken van de gegevensaanlevering en het beheer van de architectuur en standaarden. In 2008 wordt een eerste invulling gegeven aan deze organisatie. Deze wordt daarna in 2009 verder doorontwikkeld.

In de aanpak van de realisatie van de centrale infrastructuur wordt een incrementele aanpak gehanteerd. De doelstelling is om in 2008 met minimaal een eerste Parel<sup>2</sup> productieel te zijn. Omdat nog vele onduidelijkheden bestaan binnen Parel<sup>2</sup> wordt gestart met een basisplatform voor uitwisseling van onderzoeksgegevens. Dit platform kan vervolgens incrementeel doorontwikkeld worden, mede op basis van gedane ervaringen en voortschrijdende inzichten.

## **2.3 Keuzes in de opzet van de centrale voorzieningen**

Het architectuurontwerp van de centrale voorzieningen schetst de uitgangspunten en principes waaraan deze voorzieningen moeten voldoen. Hierin zijn keuzes gemaakt over

- De wijze van bundelen van onderzoeksgegevens van de instellingen
- Het waarborgen van de privacy van de patiënt en het beveiligingen van patiëntgegevens.
- Het standaardiseren van de vastlegging van de onderzoeksgegevens in de centrale infrastructuur in de vorm van een Parel<sup>2</sup> Informatiemodel (PIM)

Hieronder worden deze onderdelen nader beschreven.

### **2.3.1 Bundelen en Centrale opslag**

In het architectuurontwerp is gekozen voor een centrale opslag van een recente kopie van de onderzoeksgegevens van de instellingen. Het beheer van de onderzoeksgegevens blijft hierbij bij de instellingen.

Regelmatig zullen instellingen onderzoeksgegevens (klinische gegevens, gegevens over bio- en beeldmateriaal) elektronisch aanleveren bij de CI. Dit gebeurt per Parel. De CI zal per Parel de aangeleverde gegevens bundelen en centraal vastleggen. Uitgaande van acht Parel's en acht UMC's, zal de CI 64 aanleveringen ontvangen en bundelen.

---

<sup>2</sup> Er is gekozen om te starten met Inflammatoire darmziekten (Inflammatory Bowel Disease, IBD)

Informatie over de parels als mede statistische kentallen over de beschikbare gegevens per parel zullen in een catalogus gepubliceerd worden voor (externe) onderzoekers.

Een (externe) onderzoeker die beschikking wil krijgen over delen van deze gebundelde gegevens zal hiervoor een verzoek indienen bij een Parelsnoer Commissie. Als deze toestemming verleent, zal het verzoek door de CI worden uitgevoerd op de centraal vastgelegde gebundelde klinische gegevens. Het resultaat hiervan wordt aan de onderzoeker beschikbaar gesteld (gebruiksrecht). Een instelling zal dus ontlast worden van de afhandeling van verzoeken om klinische gegevens.

### **2.3.2 Privacy en beveiliging**

Aangezien het medische gegevens betreft van patiënten, dient zorgvuldig met de onderzoeksgegevens om te worden gegaan. In de architectuur is gekozen voor beveiliging op meerdere lagen.

In de eerste plaats worden alle onderzoeksgegevens gepseudonimiseerd. Dat wil zeggen dat de gegevens in een dataset die afgegeven is aan een externe onderzoeker niet herleidbaar is naar een specifieke patiënt. De pseudonimisatie wordt uitgevoerd door de instellingen voordat gegevens worden aangeleverd. De CI weet daardoor ook niet wie de patiënt is waarvan de gegevens in de centrale database zijn vastgelegd.

Er is gekozen voor pseudonimisatie zodat in uitzonderlijke gevallen, volgens een strikte procedure een patiënt wel achterhaald kan worden. Dit kan alleen met medewerking van de aanleverende instelling, omdat alleen daar de informatie bekend is om te de-pseudonimiseren.

Verder is gekozen voor beveiligde communicatie volgens PKI standaarden (tweezijdige SSL op basis van servercertificaten) tussen de instellingen en de CI voor de aanlevering van gegevens.

Ten slotte is voorzien in het verwijderen van gegevens in lijn met de eisen die in de WBP<sup>3</sup> worden gesteld. Het verwijderen vindt plaats bij de bron, dus bij de instellingen. Zij verwijderen de gegevens uit hun systemen. Bij de volgende aanlevering worden eerst alle oude gegevens uit de CI verwijderd en daarna worden de nieuw aangeleverde gegevens (waarin de te verwijderen patiëntgegevens niet meer voorkomen) in de centrale database geplaatst.

### **2.3.3 Parelsnoer Informatiemodel**

Een belangrijke keuze in het architectuurontwerp is het Parelsnoer Informatiemodel (PIM). Dit is het informatiemodel waarin alle Parels zijn gemodelleerd. Binnen Parelsnoer wordt daarmee met één model gewerkt waarin eenduidig is gedefinieerd hoe onderzoeksgegevens binnen een Parel worden vastgelegd.

De instellingen leveren hun onderzoeksgegevens bij de CI aan in een bestand. Het bestandsformaat voldoet aan deze PIM. Dit wil niet zeggen dat alle systemen van een instelling aangepast moeten worden om te werken volgens de PIM, maar een instelling moet wel haar onderzoeksgegevens kunnen aanleveren conform de PIM. Een instelling kan daarbij bijvoorbeeld kiezen om eerst haar bestaande gegevens te converteren naar het PIM.

Bij het ontwerpen van de Parels wordt het datamodel (syntax en semantiek) gedefinieerd voor de klinische gegevens, de gegevens over beschikbaar biomateriaal en de gegevens

---

<sup>3</sup> Wet Bescherming Persoonsgegevens

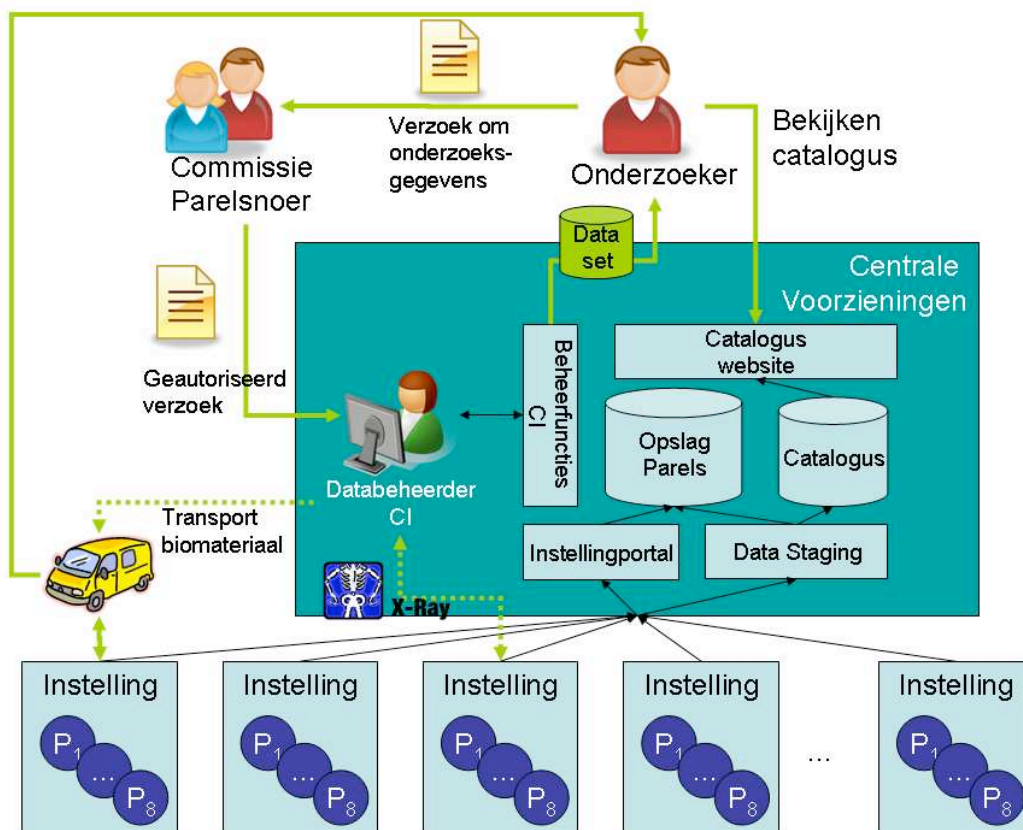
over beschikbaar beeldmateriaal. Verder wordt bij een Parel gekeken naar de kwaliteitsnormen van de vastlegging van de onderzoeksgegevens. De invoering daarvan valt buiten de scope van de CI. De CI gaat er vanuit dat alleen gecontroleerd moet worden of de aanlevering conform het PIM is en zal geen andere kwaliteitscontroles uitvoeren.

## 2.4 Globale werking van de centrale voorzieningen

De centrale voorzieningen hebben de volgende functionele hoofddoelen:

- Het informeren van onderzoekers en geïnteresseerden via een catalogus
- Het bundelen van onderzoeksgegevens van de aangesloten instellingen
- Het uitleveren van gegevensverzamelingen op basis van een geautoriseerd verzoek van een onderzoeker.

De hierboven beschreven functionele werking van de centrale voorzieningen zijn in de figuur hieronder weergegeven. Midden-boven staat de onderzoeker die de catalogus kan bekijken en een verzoek bij de commissie kan indienen. Na autorisatie komt dit verzoek bij de databeheerder van CI die dit verzoek zal afhandelen. Het resultaat wordt vervolgens bij de onderzoeker afgeleverd.



De hoofdfuncties van de CI wordt in de komende paragrafen nader toegelicht.

### 2.4.1 Informeren

De CI zal de catalogus als website beschikbaar stellen. De catalogus is een publiek toegankelijke website op Internet. Het doel is om geïnteresseerden en onderzoekers te informeren over Parelsnoer, de Parels en over de beschikbare gegevens. De website toont alleen totalen en metagegevens, maar niet de gegevens zelf.

### **2.4.2 Bundelen**

Voor het aanleveren van onderzoeksgegevens is het doel dat instellingen op regelmatige basis een volledige verzameling onderzoeksgegevens aanleveren bij de CI. Elke instelling levert één gegevensverzameling aan per parel. De gegevensverzameling is gestandaardiseerd volgens het Parelsnoer Informatiemodel (PIM).

Stel dat acht instellingen voor alle acht parels gegevensverzamelingen aanleveren, dan krijgt de CI  $8 \times 8 = 64$  gegevensverzamelingen aangeleverd. De CI zorgt voor bundeling van de gegevensverzamelingen in de centrale database.

Een instelling kan zelf bepalen uit welk systeem of systemen de gegevens geëxporteerd worden naar een bestand en met welk systeem of systemen de upload van het bestand wordt uitgevoerd.

### **2.4.3 Uitleveren**

Een onderzoeker die beschikking wil hebben over gegevens van Parelsnoer, zal zijn verzoek moeten indienen bij een wetenschappelijke Parelsnoer commissie. De commissie beslist over het wel of niet beschikbaar stellen van onderzoeksgegevens van Parelsnoer.

Om het verzoek te kunnen opstellen is informatie over de beschikbare onderzoeksgegevens beschikbaar in een catalogus. Met behulp van deze catalogus moet een onderzoeker in staat zijn om een verzoek te kunnen indienen bij de commissie.

Als de CI vanuit de commissie een geautoriseerd verzoek heeft ontvangen, zal de databasebeheerder deze vertalen naar een zoekopdracht op de centrale database. De resulterende gegevens wordt omgezet naar een exportbestand. Dit bestand wordt beschikbaar gesteld aan een onderzoeker.

Momenteel is weinig bekend over de wijze waarop een onderzoeker zijn vraag zal / kan stellen. Voor dit ontwerp wordt dan ook uitgegaan dat een menselijke functionaris de vraag van de onderzoeker moet vertalen naar een zoekopdracht op de database. Op basis van praktijkervaring kan bekeken worden of het aanleveren van de zoekvraag door een onderzoeker meer geautomatiseerd kan worden zodat de handmatige vertaling overbodig (of uitzonderlijk) wordt.

Als het verzoek biomateriaal betreft, zal de CI alle benodigde gegevens hiervoor beschikbaar stellen aan de functionaris die de logistiek van transport van het biomateriaal verzorgt. Deze zal het transport van het betreffende materiaal organiseren vanaf de instelling waar zich het materiaal bevindt tot de locatie van de aanvragende onderzoeker.

Als het verzoek beeldmateriaal betreft, zorgt de gegevensbeheerder van CI voor het opvragen van het beeldmateriaal bij de verschillende instellingen. Dit proces dient nog nader te worden vormgegeven.

## 3 Processen

In dit hoofdstuk wordt beschreven op welke wijze een instelling procesmatig deel kan nemen aan Parelsnoer. Voor een groot deel zijn de processen in de vorm van gebruiksscenario's uitgewerkt in het ontwerp van de CI hoofdstuk 3.

### 3.1 Aansluiten op de Centrale Infrastructuur

Om tot aansluiting te komen op de centrale infrastructuur zal een instelling eerst moeten voldoen aan enkele randvoorwaarden. Vervolgens kan overgegaan worden tot aansluiting op de centrale infrastructuur.

Het aansluitproces is meer gedetailleerd beschreven in het ontwerp van de beheerorganisatie.

#### Randvoorwaarden

Voordat aangesloten kan worden op parelsnoer wordt verwacht dat een instelling voldoet aan de volgende randvoorwaarden:

- Binnen de Parels wordt vastgelegd conform het Parelsnoer InformatieModel (PIM) en conform de afgesproken kwaliteitsnormen.
- Een instelling kan een dataset maken conform de bestandsformaatpecificaties beschreven in het PIM. Zie ook 4.1.
- Een instelling is in staat om deze dataset op regelmatige basis te genereren. Zie 6.1 voor de frequentie hiervan.

#### Stappen tot aansluiting

Om tot aansluiting te komen op de centrale infrastructuur moeten de volgende stappen worden doorlopen:

- Aanvragen aansluiting  
In deze stap worden administratieve gegevens uitgewisseld. Hiervoor dient een aanvraagformulier ingevuld te worden en bij de CI aangeleverd te worden. Deze aanvraag is voorzien van gegevens van contactpersonen voor de CI en technische gegevens voor aansluiting (zoals IP-nummer e.d.)
- De instelling ontvangt vervolgens aansluitgegevens van de CI. Dit zijn contactgegevens van de beheerder van de centrale infrastructuur, technische gegevens over netwerkadressen, maar ook een PKI-certificaat dat gebruikt moet worden voor beveiligde netwerkverbindingen met de CI
- De instelling en de CI realiseren technische aansluiting. De instelling installeert hierbij het door de CI uitgereikte PKI-certificaat. Gezamenlijk wordt getest of een verbinding opgezet kan worden
- De instelling en de CI testen bestandsuitwisseling. Dit is zowel een technische test (het transport van de dataset) als een functionele test (kan de centrale infrastructuur overweg met de aangeleverde dataset)
- Vervolgens wordt overgegaan tot "productie"

### 3.2 Aanleveren van onderzoeksgegevens

Voordat een dataset met onderzoeksgegevens aangeleverd kan worden bij de CI moet een instelling deze eerst aanmaken. Er geldt dat per Parel een aparte dataset wordt aangeleverd. Momenteel zijn er acht Parels. Dit betekent dat een instelling acht datasets aanlevert.

Voor een dataset geldt dat:

- Het bestand moet voldoen aan de PIM en de bestandsdefinitie. Zie hiervoor het Parelsnoer Informatie Model.
- De BSN's van de patiënten zijn vervangen door een gepseudonimiseerd BSN.
- In paragraaf 4.2.3 is beschreven volgens welk algoritme het gepseudonimiseerd BSN berekend moet worden.  
Merk op dat de definitie van de PIM zodanig is dat de dataset (met uitzondering van het BSN) voldoende anoniem is. Hierdoor kan worden volstaan met alleen het vervangen van het BSN door een gepseudonimiseerd BSN.
- Voor het kunnen de-pseudonimiseren (zie 3.5) is het nodig dat een instelling een vertaaltabel bijhoudt voor BSN en het bijgehorende gepseudonimiseerd BSN. Dit is nodig omdat het algoritme voor het berekenen van een gepseudonimiseerd BSN één richting op werkt (zie architectuurontwerp).

Als het bestand is aangemaakt volgende de hierboven beschreven voorwaarden, dan kan de instelling deze aanleveren. Het doel is om regelmatig een dataset aan te leveren, zie 6.1 voor de frequentie hiervan. De wijze van aanleveren is beschreven in gebruiksscenario 3.4.1 van het ontwerp van de CI.

Er is altijd de mogelijkheid om tussentijds een incidentele aanlevering te doen van een dataset, bijvoorbeeld omdat de reguliere aanlevering niet gelukt is. Dit is beschreven in 3.4.2 van het ontwerp van de CI.

Als een instelling een dataset heeft aangeleverd, zal deze niet direct verwerkt worden. De CI zal batchgewijs alle aanleveringen van alle instellingen verwerken. Een instelling krijgt altijd terugkoppeling per email over de verwerking van de aangeleverde datasets.

Ten slotte is bij de CI een instellingportal beschikbaar. Hierin is het mogelijk om de status van alle aanleveringen te bekijken. Tevens is inzage in de log van de CI mogelijk voor gebeurtenissen die betrekking hebben op de instelling.

### **3.3 Aanleveren van beeldmateriaal**

In de dataset van een Parel levert een instelling ook een overzicht aan van beschikbaar beeldmateriaal voor de Parel. Dit overzicht betreft alleen metadata van het beeldmateriaal, niet het beeldmateriaal zelf.

Een externe onderzoeker kan bij de CI een verzoek indienen om beschikking te krijgen over beeldmateriaal. Als dit verzoek is geautoriseerd door de Parel commissie, zal de CI herleiden in welke instelling het verzochte beeldmateriaal zich bevind. De CI zorgt voor bewaking van de afhandeling hiervan. Daarna zorgt de CI voor bundeling van alle beelden en het aanbieden hiervan aan de onderzoeker.

N.B. De wijze van opvragen en beschikbaar stellen van beeldmateriaal wordt nog nader onderzocht. Ook eventuele pseudonimisatie voor beeldmateriaal wordt daarbij nader bekeken.

Het koppelvlak hiervoor is nader beschreven in hoofdstuk 5

### **3.4 Aanleveren van biomateriaal**

Een externe onderzoeker kan Parelsnoer verzoeken om biomateriaal beschikbaar te stellen voor onderzoek. Als dit verzoek door de Parelsnoer Commissie is goedgekeurd zal de CI de afhandeling hiervan uitvoeren.

In de dataset die per Parel aangeleverd wordt door een instelling is een overzicht opgenomen van beschikbaar biomateriaal en de locatie daarvan. Deze informatie wordt gebruikt om een logistiek proces uit te voeren om het materiaal op te halen op locatie en af te leveren bij de onderzoeker. Dit logistieke proces staat buiten scope van dit document.

### **3.5 Afhandelen verzoek de-pseudonimisatie**

Onder strikte voorwaarden (zoals beschreven in het ontwerp CI) is het mogelijk dat gepseudonimiseerde gegevens van een patiënt gede-pseudonimiseerd wordt. De CI houdt geen gegevens bij om dit te doen. De instelling wel: de instelling dient gegevens bij te houden zodat de-pseudonimisatie mogelijk wordt:

Het afhandelen van dit verzoek zal daarom een tweetraps raket zijn. Eerst zal het verzoek bij de CI komen. Het verzoek is voorzien van een versleuteld en gepseudonimiseerd BSN. De CI zal deze ontsleutelen. Vervolgens wordt de instelling die de onderzoeksgegevens van deze patiënt heeft aangeleverd (dit wordt achterhaald op basis van het ontsleutelde maar nog gepseudonimiseerd BSN) verzocht om het verzoek verder af te handelen. Hiertoe houdt de instelling een vertaaltabel bij om bij een gepseudonimiseerd BSN de BSN van de patiënt te kunnen afleiden.

N.B. De verdere eisen aan dit proces dienen nog nader te worden uitgewerkt

### **3.6 Verwijderen van patiëntgegevens uit de CI**

Als een patiënt verzoekt om zijn gegevens te verwijderen (of indien de instelling om een andere reden de gegevens van een patiënt wil verwijderen), is het aan de instelling om dit uit te voeren.

De instelling verwijdert lokaal de gegevens van de patiënt. Bij het genereren van een nieuwe dataset zal deze gegevens dan niet meer worden opgenomen. De CI krijgt bij de eerst volgende aanleveringen een nieuwe dataset zonder deze patiëntgegevens.

De CI verwijdert bij een nieuwe dataset altijd eerst de oude dataset en vervangt deze door de nieuwe dataset. Omdat de nieuwe dataset de gegevens van de betreffende patiënt niet meer bevat, zullen deze gegevens ook niet meer voorkomen in de CI. Zie verder 3.4.5 uit het ontwerp CI.

## 4 Informatie en functionaliteit

Van een instelling wordt verwacht dat zij op regelmatige basis een bestand aanlevert bij de CI. Dit bestand moet voldoen aan het afgesproken bestandsformaat, zoals gedefinieerd in het ontwerp van de CI en de PIM. Aandachtspunt hierbij is dat de bestand geen BSN's bevat van de patiënten, maar een gepseudonimiseerd BSN. Dit pseudonimiseren is een taak van de instelling.

### 4.1 Samenstellen dataset

Van een instelling wordt gevraagd om haar gegevens aan te leveren conform het Parelsnoer InformatieModel (PIM). Hierbij wordt niet verwacht dat een instelling in haar software het PIM gaat hanteren, maar dat er wel een exportbestand gemaakt kan worden conform dit model. Hierbij kan het noodzakelijk zijn dat een instelling haar software wel aanpast bijvoorbeeld omdat zij nu bepaalde gegevens die conform de PIM worden verwacht, nog niet vastlegt.

De PIM is in een apart document gespecificeerd. Hierin is ook het bestandsformaat gedefinieerd. Hierbij is gekozen voor de XML standaard.

### 4.2 Pseudonimiseren en de-pseudonimiseren

In het architectuurontwerp is ervoor gekozen dat de CI geen weet heeft van de identiteit van de patiënt waarvan gegevens in de CI beschikbaar zijn. Dit wordt deels in de PIM opgelost doordat de vastgelegde gegevens voldoende geanonimiseerd zijn.

Wel is het doel dat gewerkt wordt met de BSN van de patiënt. Om vast te houden aan het architectuurprincipe dat de CI geen weet heeft van de identiteit van de patiënt, wordt gewerkt met een gepseudonimiseerd BSN. De instelling wordt gevraagd dit gepseudonimiseerd BSN in de dataset op te nemen in plaats van het BSN zelf.

#### 4.2.1 Bepalen gepseudonimiseerd BSN

In het architectuurontwerp is verder gekozen om er voor te zorgen dat als twee instellingen gegevens van dezelfde patiënt aanleveren, dat deze gekoppeld kunnen worden in de CI. Daarvoor is het nodig dat de gepseudonimiseerde BSN voor hetzelfde BSN hetzelfde is voor beide instellingen. Om dit te bewerkstelligen is het nodig dat elke instelling hetzelfde algoritme gebruikt voor het berekenen van het gepseudonimiseerde BSN.

In het architectuurontwerp is beschreven dat dit algoritme een cryptografische one-way algoritme moet zijn (architectuurprincipe 6).

Het BSN bestaat uit 9 cijfers. Dit betekent dat er 1.000.000.000 potentiële BSN's zijn. Het BSN voldoet verder aan de 11-proef waardoor de getalruimte kleiner wordt<sup>4</sup>, tot zo'n 90 miljoen.

Omdat het algoritme bekend is, kan een dictionary (woordenboek) opgezet worden waarin alle mogelijk BSN's worden uitgezet tegen de versleutelde BSN.

---

<sup>4</sup> Zie verder <http://nl.wikipedia.org/wiki/Burgerservicenummer> en <http://nl.wikipedia.org/wiki/Sofinummer>

Om deze aanpak fors te bemoeilijken wordt gekozen om de geboortedatum<sup>5</sup> van de patiënt voor het BSN te plakken. Hierdoor wordt de getalruimte vergroot tot ongeveer  $30 \times 12 \times 100^6 \times 90$  miljoen = 3,24 triljoen.

Op basis van de concatenatie van geboortedatum en BSN wordt het SHA-2 algoritme gebruik voor de berekening van het gepseudonimiseerd BSN. Het SHA-2 algoritme dient hierbij gebruik te maken van 256 bits digest. Er wordt hierbij<sup>7</sup> ook wel gesproken over SHA-256.

Ten behoeve van de-pseudonimisatie is het vervolgens nodig dat een instelling een vertaaltabel bijhoudt tussen BSN en het met SHA-2 berekend nummer.

#### **4.2.2 De-pseudonimisatie**

Om de mogelijkheid te hebben de identiteit van een patiënt te achterhalen, wordt een procedure opgezet voor het de-pseudonimiseren van een patiëntrecord. Omdat de CI geen weet heeft van het BSN, maar alleen van het gepseudonimiseerd BSN, zal de uitvoering van de procedure bij de instelling plaats vinden, zie ook 3.5.

Om te kunnen de-pseudonimiseren is het nodig dat een instelling een vertaalfunctie heeft om een gepseudonimiseerd BSN terug te vertalen naar een BSN.

#### **4.2.3 Algoritme berekenen gepseudonimiseerd BSN**

De procedure voor het berekenen is als volgt:

voor elke patiënt P in de dataset

    Bepaal concatenatie  $c = P \rightarrow \text{BSN}$  en  $P \rightarrow \text{geboortedatum}$

    Bereken hiervan de SHA-2:  $s = \text{SHA-2}(c)$

    Bewaar in de vertaaltabel het BSN en s

    Vervang in de dataset het BSN van P met s

---

<sup>5</sup> Bij patiënten waarvan alleen het geboortjaar bekend is, wordt 1-jan gebruikt als dag en maand

<sup>6</sup> 30 dagen x 12 maanden x 100 jaar

<sup>7</sup> Zie <http://en.wikipedia.org/wiki/SHA-2> voor meer informatie over SHA

## 5 Techniek

### 5.1 Netwerkverbinding

De verbinding tussen een instelling en de CI verloopt over het publieke Internet. Deze verbinding wordt met SSL beveiligd.

Op termijn wordt overwogen om communicatie ook via SURFnet te laten verlopen.

### 5.2 Tweezijdige SSL

Voor de beveiligde verbinding is gekozen voor tweezijdige SSL. Hierbij wordt SSL 3.0 of TLS 1.0 toegepast. Zie ook 4.7 in het ontwerp CI. Hierbij dient zowel de CI als de instelling een certificaat te hebben. Een verbinding komt alleen tot stand als beide partijen elkaars certificaat accepteren. Deze acceptatie bestaat uit de volgende controles die beide partijen op het certificaat van de andere partij uitvoeren:

- het certificaat is onveranderd
- het certificaat is uitgegeven door een vertrouwde CA<sup>8</sup>
- het certificaat is niet verlopen
- het certificaat is niet ingetrokken (staat niet op de lijst met ingetrokken certificaten: de Certificate Revocation List ofwel CRL<sup>9</sup>)

Dit zijn standaard controles binnen SSL en PKI en worden ondersteund door de gangbare SSL-software

Daarnaast zal de CI controleren of de instelling geautoriseerd is op basis van de identificatie van de instelling in het certificaat bij de verbinding wordt gebruikt. Een instelling dient hetzelfde te doen met de naam in het certificaat van de CI.

### 5.3 Koppelvlak datasets

Het aanleveren van een dataset kan via FTP en via http, zie ook 4.1.1 in het ontwerp CI. Beide protocollen worden gebruikt over een SSL-verbinding zoals beschreven in de vorige paragraaf.

Voor een instelling die FTP wil gebruiken betekent dit dat zij een FTP-client moet gebruiken die in staat is om een tweezijdige SSL verbinding op te zetten.

Voor een instelling die HTTP wil gebruiken, zal zij gebruik moeten maken van een HTTP-client die tweezijdige SSL ondersteunt. De meeste gangbare webbrowsers ondersteunen dit. Ook de standaard Java en .NET omgevingen bieden deze ondersteuning.

### 5.4 Koppelvlak beeldmateriaal

Op verzoek van de CI stelt een instelling beeldmateriaal beschikbaar. De precieze wijze waarop dit zal plaatsvinden zal nog nader worden bepaald.

---

<sup>8</sup> CA staat voor certification authority. Dit is de dienstverlener die verantwoordelijk is voor de uitgifte van de certificaten

<sup>9</sup> De CRL is een bestand dat gedownload kan worden bij de CA

## **5.5 Eisen aan beveiliging certificaten**

Een instelling wordt voorzien van een certificaat waarmee de instelling zich identificeert en authenticceert voor het aanleveren van datasets (voor tweezijdige authenticatie). Een instelling moet voorkomen dat deze (en dan met name de private key) door onbevoegden kan worden benaderd. Hiervoor kunnen reguliere beveiligingsmethoden worden toegepast (zie bijvoorbeeld de adviezen van NEN7510 voor informatiebeveiliging)

## **6 Afspraken**

### **6.1 Frequentie van aanlevering**

Eén keer per maand wordt voor elke Parel een bijgewerkte dataset aangeleverd.

Ook als er geen wijzigingen hebben plaatsgevonden wordt de dataset aangeleverd. (tbv uitvaldetectie)

### **6.2 Beschikbaarheid**

Bevestiging ontvangst verzoek binnen 2 werkdagen  
Aanleveren beeldmateriaal 5 werkdagen  
Afhandelen verzoek de-pseudonimisatie 10 werkdagen

#### **6.2.1 Kwaliteit van gegevens**

De dataset voldoet altijd aan de syntactische afspraken.

Verder voldoet de dataset aan de kwaliteitsnormen die binnen Parels zijn afgesproken