

Ontwerp Centrale Infrastructuur

Parelsnoer



PARELSNOER INITIATIEF

Versie: 1.1
Status: Definitief
Datum: 19 september 2008

Inhoud

1	Inleiding	4
1.1	Positionering van dit ontwerp.....	4
1.2	Uitgangspunten.....	5
1.3	Gerelateerde documenten	5
1.4	Versiehistorie.....	6
2	Context.....	7
2.1	Parelsnoer	7
2.2	Project Parelsnoer Centrale Infrastructuur	7
2.3	Keuzes in de opzet van de centrale voorzieningen	8
2.3.1	Bundelen en Centrale opslag.....	8
2.3.2	Privacy en beveiliging	9
2.3.3	Parelsnoer Informatiemodel.....	9
2.4	Globale werking van de centrale voorzieningen	10
2.4.1	Informeren	10
2.4.2	Bundelen.....	10
2.4.3	Uitleveren	11
3	Gebruiksscenario's	12
3.1	Wat zijn gebruiksscenario's	12
3.2	Actoren	12
3.2.1	Onderzoeker.....	12
3.2.2	Instelling.....	12
3.2.3	Gegevensbeheerder CI.....	12
3.2.4	Beheerder CI	12
3.2.5	Centrale Voorzieningen	13
3.3	Overzicht van de scenario's	13
3.4	Scenario's voor de aanleverende instelling.....	14
3.4.1	Reguliere aanlevering dataset.....	14
3.4.2	Incidentele aanlevering dataset.....	14
3.4.3	Bekijken status van aanleveringen	15
3.4.4	Bekijken logbestand	16
3.4.5	Verwijderen van patiëntgegevens.....	16
3.4.6	De-pseudonimisatie.....	17
3.5	Scenario's voor de (externe) onderzoeker	17
3.5.1	Bekijken informatie over Parelsnoer en de parels	17
3.5.2	Opvragen statistische kentallen van een Parel.....	18
3.5.3	Opvragen data dictionary van een Parel	18
3.5.4	Indienen verzoek voor een dataset	19
3.5.5	Indienen verzoek voor biomateriaal	20
3.5.6	Indienen verzoek voor beeldmateriaal.....	20
3.5.7	De-pseudonimisatie.....	20
3.6	Scenario's voor de gegevensbeheerder CI	21
3.6.1	Verwerken ontvangen datasets in de centrale database	21
3.6.2	Beschikbaar stellen dataset	22
3.6.3	Afhandelen verzoek biomateriaal.....	23
3.6.4	Afhandelen verzoek beeldmateriaal	24
3.6.5	Afhandelen verzoek tot de-pseudonimisatie	24
3.6.6	Inzien log.....	25
3.6.7	Bekijken gearchiveerde verzoeken.....	25
3.7	Scenario's voor de beheerder CI.....	26
3.7.1	Operationeel beheer CI	26
3.7.2	Bijwerken datadictionary	26
3.7.3	Inzien log CI.....	27
3.7.4	Gebruikersbeheer instellingportal	28

3.7.5	Aansluiten instelling.....	28
3.7.6	Afsluiten instelling	29
3.7.7	Management rapportages.....	29
3.7.8	Bewaking aanlevering	29
4	Functionaliteit van de CI	30
4.1	Ontvangst van gegevens: Data Staging	31
4.1.1	Interface	31
4.1.2	Ontvangen van bestand	31
4.1.3	Verwerken van een bestand.....	32
4.2	Ontsluiten catalogus	33
4.3	Samenstellen dataset	33
4.3.1	Interface	33
4.3.2	Zoekfunctionaliteit.....	34
4.3.3	Beschikbaar stellen resultaat dataset	34
4.4	Logging.....	34
4.5	Beheerfuncties	35
4.6	Instellingenportal	35
4.7	Beveiliging	36
4.8	Rapportages	36
5	Gegevens in de CI.....	37
5.1	Parelsnoer Informatie Model	37
5.1.1	IBD	37
5.2	Afgehandeld verzoek en uitgegeven datasets.....	37
5.3	Ontvangen datasets.....	38
5.4	Administratieve informatie.....	38
5.5	Logging.....	39

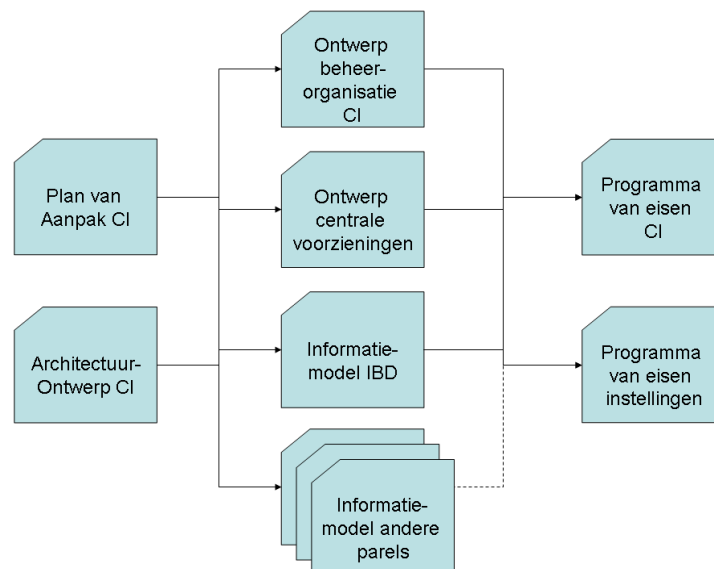
1 Inleiding

Het project Parelsnoer beoogt¹ 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 een ontwerp van de techniek en de beheerorganisatie benodigd. Het voorliggende document betreft het ontwerp van de technische centrale voorzieningen voor Parelsnoer.

Het ontwerp van de beheerorganisatie is opgenomen in een separaat document.

1.1 Positionering van dit ontwerp

Het voorliggende document betreft het ontwerp van de centrale infrastructuur. Dit ontwerpdocument 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.
- *Ontwerp Centrale Voorzieningen*
Dit is het voorliggende document. Dit is het ontwerp van de centrale

¹ Uit "Werkplan Parelsnoer"

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)*
Het PIM is een gegevensontwerp voor het vastleggen en uitwisselen van gegevens over Parels. Het PIM is de basis voor de inrichting van de database van de centrale infrastructuur. Tevens is het PIM de afspraak volgens welke alle instellingen gegevens gaan uitwisselen met de centrale infrastructuur
- *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 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 CI levert een deelverzameling op van de beschikbare gegevensverzameling van een Parel. De CI zal hierop geen bewerkingen of aggregatie uitvoeren, zoals statistische berekeningen. De berekeningen en analyse op de dataset zal de onderzoeker met eigen voorzieningen doen.
- 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 onderzoekers meer geautomatiseerd kan worden zodat de handmatige vertaling overbodig (of uitzonderlijk) wordt.
- Binnen de CI wordt gebruik gemaakt van gepseudonimiseerd BSN. Door de instelling wordt de BSN vastgelegd en bij aanlevering gepseudonimiseerd. Van patiënten die geen BSN hebben worden de gegevens niet uitgewisseld.
- 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.
- De CI kan een onderzoeker datasets opleveren. Op basis van deze dataset kan een onderzoeker statistisch onderzoek uitvoeren. De CI zal geen (statistisch) onderzoek uitvoeren op de gegevens voor een onderzoeker.

1.3 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, Programma Van Eisen Instelling: Parelsnoer PvE Instellingen. Versie 0.3. 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).
- Miranda de Gouw en Sandra Sliepenbeek, Parelsnoer Informatie Model (PIM). Versie 0.0. Amsterdam: Parelsnoer Initiative, Q1 2008 (PM).

1.4 Versiehistorie

Datum	Versie / Status	Opmerkingen
14 dec 2007	0.3	Initiële concept versie
20 dec 2007	0.4	Concept versie ter bespreking met UMC's
8 jan 2008	0.5	Commentaar Gerard van der Hoorn verwerkt
10 maart 2008	1.0	Verwerking ontvangen commentaar CI team en UMC's (in sessie 31 jan)
4 april	1.01	Finaliseren
10 sep 2008	1.1	Overgezet naar parelsnoer 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 Parel² productioneel te zijn. Omdat nog vele onduidelijkheden bestaan binnen Parelnoer 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 Parelnoer 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 Parels en acht UMC's, zal de CI 64 aanleveringen ontvangen en bundelen.

² 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 WBP³ 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 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

³ Wet Bescherming Persoonsgegevens

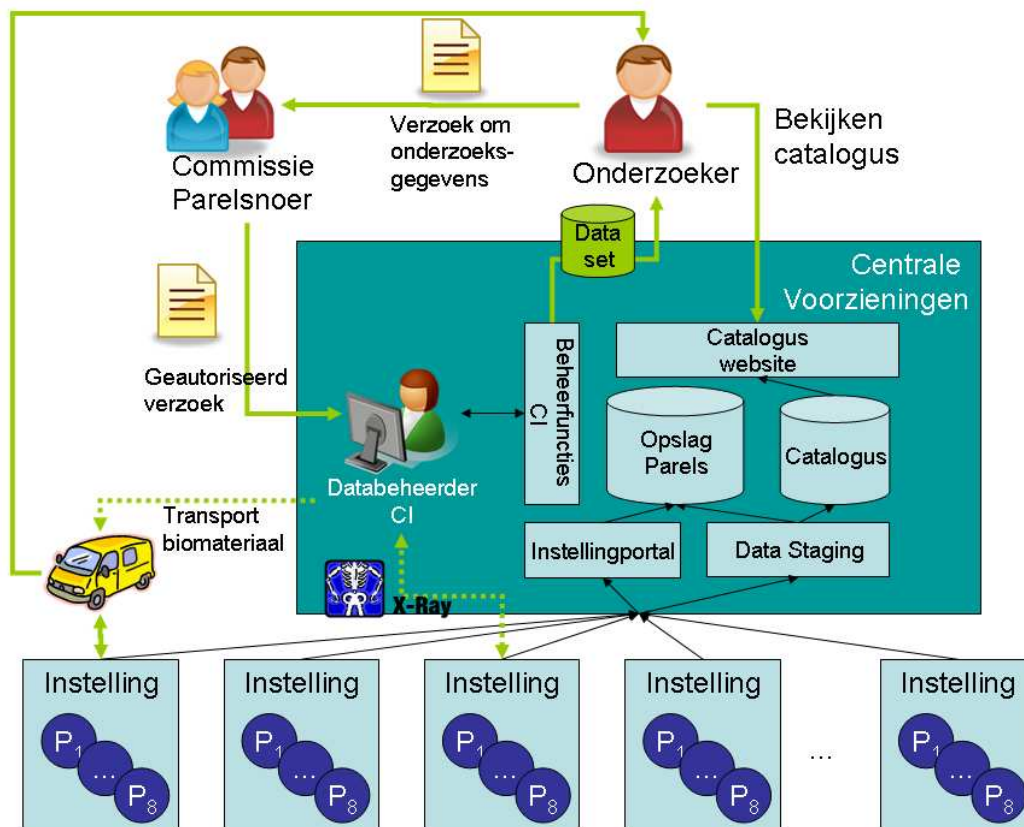
worden of de aanlevering conform het PIM is en zal geen andere kwaliteitscontroles uitvoeren.

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 databeheerder 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 onderzoekers 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 Gebruiksscenario's

3.1 Wat zijn gebruiksscenario's

In de gebruiksscenario's (use cases) worden de stappen die gebruikers met het systeem doorlopen, beschreven. Een use case is een tekstuele weergave van deze communicatie tussen het systeem (in brede zin) en de verschillende actoren. Zowel handmatige stappen als stappen waar het systeem een rol in speelt, komen aan bod, gezien vanuit het perspectief van de gebruiker van het systeem.

Per scenario staat het doel, de actor, de aanleiding en voorwaarden beschreven. Daarna volgt het scenario dat door de actor doorlopen wordt. Een actor kan een eindgebruiker zijn (zoals een onderzoeker in een instelling, of een beheerder) , maar ook een geautomatiseerd proces (bijvoorbeeld in het geval dat twee systemen gegevens met elkaar uitwisselen). De stappen in de use case zijn chronologisch beschreven en genummerd. Naast dit standaard scenario, kunnen zich andere situaties voordoen waardoor andere stappen worden doorlopen. Meestal betreffen deze situaties uitzonderingssituaties of foutsituaties. Deze staan beschreven onder 'alternatieve scenario's'.

De nummering van deze alternatieve scenario's sluit aan op het standaard scenario: bij een stap waar een alternatief bestaat, wordt bij alternatieve scenario's onder hetzelfde nummer het alternatief beschreven. Bijvoorbeeld als bij stap 2 en alternatief scenario bestaat, wordt deze beschreven met het nummer 2. Als het alternatief scenario uit meerdere stappen bestaat, wordt deze genummerd met een toevoegsel. In het voorbeeld: stap 2a, stap 2b, stap 2c. In het geval dat er meerdere alternatieve scenario's zijn bij een specifieke stap wordt dit genoteerd met 2-1a, 2-1b,.. voor het eerste alternatief en 2-2a, 2-2b, ... voor het tweede alternatief.

Daarna staan, voor zover nodig, de opmerkingen of openstaande issues vermeld.

3.2 Actoren

3.2.1 Onderzoeker

De onderzoeker is een persoon die voor het uitvoeren van onderzoek geïnteresseerd is in gegevens die Parelsnoer biedt. Vooral nog wordt er vanuit gegaan dat een onderzoeker werkt voor een van de aangesloten instellingen.

3.2.2 Instelling

Een instelling is een organisatie die voor een of meerdere parels onderzoeksgegevens, biomateriaal en/of beeldmateriaal beschikbaar stelt of wil stellen aan Parelsnoer. Vooral nog wordt er vanuit gegaan dat alleen UMC's als instelling aansluiten op Parelsnoer.

3.2.3 Gegevensbeheerder CI

De gegevensbeheerder van de CI is een rol. Vanuit deze rol wordt een verzoek voor onderzoeksgegevens afgehandeld. Bij deze rol is ook de bewaking van de aanleveringen door instellingen belegd.

3.2.4 Beheerder CI

De beheerder CI is de technisch beheerder van de centrale infrastructuur. In deze rol is de verantwoordelijkheid belegd voor het operationeel houden van de centrale infrastructuur.

3.2.5 Centrale Voorzieningen

De centrale voorzieningen is het geheel aan techniek, organisatie en processen binnen Parelsnoer die onderzoeksgegevens van instellingen kunnen bundelen en beschikbaar kunnen stellen aan onderzoekers.

3.3 Overzicht van de scenario's

Scenario's voor de aanleverende instelling

3.4.1	Reguliere aanlevering dataset;
3.4.2	Incidentele aanlevering dataset
3.4.3	Bekijken status van aanleveringen
3.4.4	Bekijken logbestand
3.4.5	Verwijderen van patiëntgegevens
3.4.6	De-pseudonimisatie

Scenario's voor de (externe) onderzoeker

3.5.1	Bekijken informatie over Parelsnoer en de parels
3.5.2	Opvragen statistische kentallen van een Parel
3.5.3	Opvragen data dictionary van een Parel
3.5.4	Indienen verzoek voor een dataset
3.5.5	Indienen verzoek voor biomateriaal
3.5.6	Indienen verzoek voor beeldmateriaal
3.5.7	De-pseudonimisatie

Scenario's voor de gegevensbeheerder CI

3.6.1	Verwerken ontvangen datasets in de centrale database
3.6.2	Beschikbaar stellen dataset
3.6.3	Afhandelen verzoek biomateriaal
3.6.4	Afhandelen verzoek beeldmateriaal
3.6.5	Afhandelen verzoek tot de-pseudonimisatie
3.6.6	Inzien log
3.6.7	Bekijken gearchiveerde verzoeken

Scenario's voor de beheerder CI

3.7.1	Operationeel beheer CI
3.7.2	Bijwerken datadictionary
3.7.3	Inzien log CI
3.7.4	Gebruikersbeheer instellingportal
3.7.5	Aansluiten instelling
3.7.6	Afsluiten instelling
3.7.7	Management rapportages
3.7.8	Bewaking aanlevering

3.4 Scenario's voor de aanleverende instelling

3.4.1 Reguliere aanlevering dataset

Doel: Het aanleveren van een nieuwe dataset aan de CI

Actor: Instelling. Dit kan een persoon zijn of een geautomatiseerd proces

Aanleiding: In de instelling wordt voortdurend de onderzoeksgegevens aangevuld en bijgewerkt. In de architectuur is gekozen om op regelmatige basis de laatste versie van de volledige dataset van de parels van een instelling aan te leveren bij de CI.

Voorwaarden:

- De instelling is aangesloten op de centrale voorzieningen
- De instelling kan per parel een nieuwe/vernieuwde dataset maken
- De aan te leveren dataset omvat alle gegevens van deze Parel, bestaande uit klinische gegevens, beschikbaar biomateriaal en beschikbaar beeldmateriaal.

Scenario:

1. De instelling genereert per parel een nieuwe dataset conform het Parelsnoer Informatie Model.
2. De instelling maakt verbinding met de CI
3. De CI identificeert en authenticceert de instelling
4. De instelling upload een dataset van een parel
5. De CI plaatst de dataset met informatie over de aanlevering in het component data staging area en verbreekt de verbinding met instelling na bevestiging van de laatste dataset.
6. Stappen 2-6 worden herhaald voor alle datasets van de overige parels.

Alternatieve scenario's:

3. De identificatie of authenticatie lukt niet. De CI geeft een foutmelding en verbreekt de verbinding

Opmerkingen:

- In stap 3 wordt verbinding met de CI gemaakt. Dit kan via secure FTP of via een webservice (HTTPS), zie de beschrijving van de data staging area in 4.1.
- De data staging area wordt beschreven in 4.1. Het verwerken van de gegevensset naar de centrale database wordt beschreven in 4.1.3.
- In stap 5 wordt bij de dataset informatie over de aanlevering (metadata) vastgelegd. Deze is beschreven in 5.3.

Open issues:

- Geen.

3.4.2 Incidentele aanlevering dataset

Doel: Het aanleveren van een nieuwe dataset aan de centrale voorzieningen

Actor: Instelling. Dit kan een persoon zijn of een geautomatiseerd proces

Aanleiding: Als een dataset niet op de regulier wijze aangeleverd kan worden, kan een instelling ook incidenteel de dataset aanleveren.

Voorwaarden:

- De instelling is aangesloten op de centrale voorzieningen
- De instelling kan een enkele dataset maken
- De aan te leveren dataset omvat alle gegevens van deze Parel, bestaande uit klinische gegevens, beschikbaar biomateriaal en beschikbaar beeldmateriaal.

Scenario:

1. De instelling genereert een nieuwe dataset voor een specifieke parel
2. Stappen 2-6 van scenario 3.4.1 worden uitgevoerd om deze enkele dataset bij de CI aan te leveren.

Alternatieve scenario's:

- Geen.

Opmerkingen:

- Geen.

Open issues:

- Geen

3.4.3 Bekijken status van aanleveringen

Doel: Het bieden van een overzicht van uitgevoerde aanleveringen bij de CI door een instelling

Actor: De beheerder van de aanleverende instelling

Aanleiding: De aanleverende instelling wil weten welke aanlevering zijn ontvangen en verwerkt in de CI.

Voorwaarden:

- De beheerder is geauthenticeerd door de instellingportal van de CI door middel van gebruikersnaam/wachtwoord

Scenario:

1. De beheerder vraagt de instellingportal van de CI om het overzicht van aanleveringen
2. De instellingportal toont een overzicht van de afgelopen aanleveringen. Bij elke aanlevering is aangegeven of de aanlevering goed is ontvangen en of deze is opgenomen in de centrale database
3. De beheerder logt uit

Alternatieve scenario's:

- Geen.

Opmerkingen:

- Geen

Open issues:

- Geen

3.4.4 Bekijken logbestand

Doel: Het bieden van een overzicht van de logregels van/over/voor een instelling

Actor: De beheerder van de aanleverende instelling

Aanleiding: De aanleverende instelling wil inzage in haar gebeurtenissen op de CI,

Voorwaarden:

- De beheerder is geauthenticeerd door de instellingportal van de CI door middel van gebruikersnaam/wachtwoord

Scenario:

1. De beheerder vraagt de instellingportal van de CI om het overzicht van de log
2. De instellingportal biedt de beheerder selectiemogelijkheden om specifieke delen van de log te kunnen zien. Er kan geselecteerd worden op gebeurtenistype
3. Aan-/afmelden beheerder
4. Aanleveringen gegevenssets
5. De instellingportal toont een overzicht van de log op basis van de ingevulde selectiecriteria.
6. De beheerder klikt op een logregel
7. De instellingportal toont uitgebreide details van de logregel
8. De beheerder logt uit.

Alternatieve scenario's:

- Geen.

Opmerkingen:

- Geen

Open issues:

- Geen

3.4.5 Verwijderen van patiëntgegevens

Doel: Een instelling wil zorgdragen dat de gegevens van een patiënt niet meer gebruikt kan worden door onderzoekers die gebruik maken van de CI.

Actor: De instelling

Aanleiding: De patiënt kan een verzoek bij de instelling gedaan hebben om zijn gegevens niet (meer) te gebruiken of de instelling kan andere redenen hebben om de gegevens uit de CI weg te halen

Voorwaarden:

- De instelling is aangesloten op de centrale voorzieningen

Scenario:

1. De instelling verwijdert de gegevens in zijn lokale systemen
2. De instelling maakt een voor elke parel waarin de gegevens van de patiënt voorkomen, een nieuwe dataset
3. De instelling verstuurt deze nieuwe dataset via de reguliere aanlevering of via de incidentele aanlevering volgens scenario 3.4.1 respectievelijk scenario 3.4.2.
4. De CI vervangt de gehele oude dataset door de nieuwe dataset waarin de gegevens van de patiënt niet meer voorkomen. Hierdoor worden de gegevens van de patiënt verwijderd uit de CI.

Alternatieve scenario's:

3. Zie de alternatieve scenario's van scenario 3.4.1 dan wel scenario 3.4.2.

Opmerkingen:

- Deze use case is hier opgenomen om de lezer inzage te geven in de wijze waarop patiëntgegevens verwijderd worden uit de CI. Deze use case zal geen nieuwe functionaliteit introduceren doordat gebruik gemaakt wordt van andere use cases voor aanlevering.
- De CI heeft zelf geen mogelijkheden om gegevens van een patiënt te verwijderen. De CI beschikt ook niet over het BSN van de patiënt om dit te kunnen doen.

Open issues:

- Geen.

3.4.6 De-pseudonimisatie

Het integrale scenario voor de-pseudonimisatie is beschreven in paragraaf 3.5.7.

3.5 Scenario's voor de (externe) onderzoeker

3.5.1 Bekijken informatie over Parelsnoer en de parels

Doel: Informatie verstrekken over parelsnoer en de parels

Actor: De gebruiker: een willekeurige bezoeker of een onderzoeker

Aanleiding: Een gebruiker wil informatie over Parelsnoer en over de parels

Voorwaarden:

- De gebruiker heeft een webbrowser en heeft verbinding met internet

Scenario:

1. De gebruiker surft naar de website van de catalogus van Parelsnoer
2. De website toont een startpagina
3. De gebruiker kiest voor Algemene Informatie
4. De website toont een overzichtspagina met beschikbare informatie. Vanaf deze pagina kan gekozen worden voor Informatie over het project Parelsnoer en er kan gekozen worden voor informatie over de individuele Parels
5. De gebruiker kiest voor een Parel
6. De website toont een pagina met algemene informatie over de Parel en verwijzingen naar relevante informatiebronnen over deze Parel. De website toont ook links naar de catalogus voor deze Parel

Alternatieve scenario's:

- 5a. De gebruiker kiest voor Informatie over het project
- 5b. De website toont een pagina met algemene informatie over het project Parelsnoer. Hierbij zijn ook alle procedures voor de aanvraag van een deelverzameling opgenomen.

Opmerkingen:

- Een overzicht van alle parels is opgenomen in paragraaf 2.1

Open issues:

- Is de catalogus een publieke website?

3.5.2 Opvragen statistische kentallen van een Parel

Doel: Het tonen van verschillende kentallen van één Parel

Actor: De gebruiker: een willekeurige bezoeker of een onderzoeker

Aanleiding: De gebruiker wil informatie over de beschikbare klinische gegevens, biomateriaal en/of beeldmateriaal van een Parel. De catalogus bevat hiertoe statistische kentallen

Voorwaarden:

- De gebruiker heeft een webbrowser en heeft verbinding met internet

Scenario:

1. De gebruiker surft naar de website van de catalogus van Parelsnoer en kies voor informatie over Parels zoals beschreven in 3.5.1.
2. De website toont de pagina met informatie over de gekozen Parel
3. De gebruiker kiest voor statistische kentallen over beschikbare klinische gegevens
4. De website toont de kentallen van beschikbare klinische gegevens van de Parel

Alternatieve scenario's:

3-1a. De gebruiker kiest voor statistische kentallen over beschikbaar biomateriaal

3-1b. Ga naar stap 4

3-2a. De gebruiker kiest voor statistische kentallen over beschikbaar beeldmateriaal

3-2b. Ga naar stap 4

Opmerkingen:

- De kentallen zijn per Parel gedefinieerd. Zie het Parelsnoer Informatiemodel.

Open issues:

- Zijn de kentallen publieke informatie?

3.5.3 Opvragen data dictionary van een Parel

Doel: Het tonen van de gegevensstructuur (data dictionary) van de dataset van een Parel

Actor: De gebruiker: een willekeurige bezoeker of een onderzoeker

Aanleiding: De gebruiker wil inzicht in de beschikbare kenmerken dat van een patiënt vastgelegd is binnen Parelsnoer.

Voorwaarden:

- De gebruiker heeft een webbrowser en heeft verbinding met internet

Scenario:

1. De gebruiker surft naar de website van de catalogus van Parelsnoer en kies voor informatie over Parels zoals beschreven in 3.5.1.
2. De website toont de pagina met informatie over de gekozen Parel
3. De gebruiker kiest voor Gegevenstructuur Dataset
4. De website toont de gegevenselementen en definitie van deze elementen van de Parel.

Alternatieve scenario's:

- Geen.

Opmerkingen:

- De datadictionary is per Parel gedefinieerd. Zie het Parelsnoer Informatie Model en paragraaf 5.1 in het voorliggende document
- Bij nieuwe versies van de datadictionary zal de beheerder van de CI deze bijwerken in de catalogus. Zie scenario 3.7.2.

Open issues:

- Zijn de kentallen publieke informatie?
- Is ook semantische informatie beschikbaar?
- Kan de data dictionary uit de databasestructuur afgeleidt worden, of is dit te technisch voor de doelgroep?

3.5.4 Indienen verzoek voor een dataset

Doel: Het opvragen van een dataset ten behoeve van onderzoek

Actor: De onderzoeker

Aanleiding: De onderzoeker heeft bepaald welke selectie van de beschikbare gegevens hij beschikking over zou willen hebben. Of de onderzoeker heeft bepaald over welk biomateriaal hij zou willen beschikken.

Voorwaarden:

- De onderzoeker is in staat een onderzoeksvraag te formuleren
- De onderzoeker heeft beschikking over een aanvraagformulier en informatie over de aanvraagprocedure.

Scenario:

1. De onderzoeker dient een verzoek in bij de autorisatiecommissie voor de Parel waar hij een dataset van wil
2. De commissie beslist over autorisatie
3. De commissie stemt in met het verzoek. Dit verzoek wordt (al dan niet met aanpassingen van de commissie) aangeleverd bij de CI.
4. De gegevensbeheerder CI zal dit verzoek afhandelen conform scenario 3.6.1.

Alternatieve scenario's:

- 3a. De commissie wijst het verzoek af. Zij informeren de onderzoeker over de argumentatie.
- 3b. De onderzoeker kan het verzoek aanpassen en opnieuw indienen.

Opmerkingen:

- De afhandeling van een verzoek valt buiten scope van dit plan. Het scenario is op hoofdlijnen hier beschreven voor de volledigheid. De CI werkt alleen op basis van door de commissie geautoriseerde verzoeken.
- Het aanvraagproces is hetzelfde voor klinische gegevens, biomateriaal en/of beeldmateriaal. De afhandeling van de aanvraag niet.

Open issues:

- Welke informatie moet een verzoek bevatten?
- Een alternatieve invulling kan zijn dat een onderzoeker via een webformulier een verzoek indient bij de CI. De CI stuurt deze naar commissie voor behandeling. Het formulier kan via een wizard de query samenstellen zodat de afhandeling na besluitvorming door de commissie geautomatiseerd kan verlopen. Hiervoor is het nodig om te bepalen of de vraag van de onderzoeker concreet genoeg gesteld kan worden om te kunnen vertalen naar een zoekvraag op de database.

3.5.5 Indienen verzoek voor biomateriaal

Het indienen en afhandelen van dit verzoek zal op dezelfde wijze plaats vinden als het verzoek voor een dataset. Zie scenario 3.5.4.

Open issues:

- Dient het systeem ook informatie te bevatten over de bronlocatie waar het materiaal staat?
- Dient het systeem ook informatie te bevatten over de voorraad (hoeveelheid) van het materiaal?

3.5.6 Indienen verzoek voor beeldmateriaal

Het indienen en afhandelen van dit verzoek zal op dezelfde wijze plaats vinden als het verzoek voor een dataset. Zie scenario 3.5.4.

Open issues:

- Dient het systeem ook informatie te bevatten over de bronlocatie waar het materiaal staat?

3.5.7 De-pseudonimisatie

Doel: Het herleiden van een specifieke patiënt vanuit gepseudonimiseerde patiëntgegevens.

Actor: De onderzoeker

Aanleiding: Een onderzoeker wil vanuit zijn onderzoek de patiënt herleiden. Hij heeft geen beschikking over identificerende gegevens van de patiënt en zal dit moeten verzoeken bij de instelling.

Voorwaarden:

- De onderzoeker moet een goede reden hebben om de patiënt achter de gepseudonimiseerde gegevens te herleiden. Dit wordt door de instelling die eigenaar is van de betreffende patiëntgegevens beoordeeld.

Scenario:

1. De onderzoeker verstuurt een aanvraagformulier voor de-pseudonimisatie naar de CI. Dit aanvraagformulier bevat
2. Contactgegevens onderzoeker
3. Referentie naar het geautoriseerde verzoek om een dataset waarin de gegevens zijn opgenomen van de patiënt die herleid moet worden
4. De id van de patiënt in de dataset. Dit is de versleutelde, gepseudonimiseerde BSN.
5. Motivatie
6. De CI controleert het formulier.
7. De versleutelde, gepseudonimiseerde BSN wordt eerst door de CI ontsleuteld (dan resteert de gepseudonimiseerde BSN) op basis van de sleutel die destijds gebruikt is door de CI bij het genereren van de dataset.
8. de CI bepaalt welke instelling de eigenaar is van de betreffende patiëntgegevens
9. de CI verzoekt de betreffende instelling om het verzoek tot de-pseudonimisatie af te handelen. Hierbij wordt de instelling voorzien van de ontsleutelde, gepseudonimiseerde BSN.

Alternatieve scenario's:

2. De CI constateert dat het formulier niet goed is ingevuld. De onderzoeker wordt hiervan op de hoogte gesteld. Het scenario wordt vervolgens beëindigd.
3. De CI constateert dat de versleutelde, gepseudonimiseerde BSN niet (meer) bekend is bij de CI. Dit kan zijn omdat de onderzoeker deze foutief heeft aangeleverd of omdat de patiëntgegevens niet meer aanwezig is in de CI omdat de instelling deze

heeft verwijderd.

De onderzoeker wordt hiervan op de hoogte gesteld. Het scenario wordt vervolgens beëindigd.

4. Als er meerdere instellingen gevonden worden op basis van de gegevens in het formulier, dan kiest de beheerder welke instelling verzocht wordt om het verzoek af te handelen zoals beschreven in stap 5.

Opmerkingen:

- De afhandeling van het verzoek is de verantwoordelijkheid van de instelling als eigenaar van de patiëntgegevens. De CI fungeert alleen als centraal ontvangstpunt voor de verzoeken en voor het ontsleutelen van de gepseudonimiseerde BSN.

Open issues:

- Moet de CI ook de afhandeling bewaken? De CI is verder niet betrokken in de afhandeling door de instelling en is daarmee niet goed in staat om de afhandeling te monitoren.

3.6 Scenario's voor de gegevensbeheerder CI

3.6.1 Verwerken ontvangen datasets in de centrale database

Doel: De aangeboden dataset wordt gecontroleerd en opgeslagen in de centrale database

Actor: CI

Aanleiding: Periodiek leveren aangesloten instellingen een vernieuwde dataset aan bij de CI.

Voorwaarden:

- De datasets zijn reeds aangeleverd bij de CI volgens scenario 3.4.1 of 3.4.2.

Scenario:

1. De CI opent de verzameling Parels die ontvangen zijn van de gegevensleveranciers en die klaar staan in de *data staging area*. sinds de laatste consolidatieslag.
2. De CI doorloopt per ontvangen Parel de volgende stappen:
3. De CI controleert de structurele integriteit van de Parel, in andere woorden of de Parel voldoet aan de PIM en de afspraken over het data formaat.
4. De CI vindt geen afwijkingen van de standaard
5. De CI controleert of de teller in de dataset groter is dan de teller in de dataset van de vorige dataleverantie.
6. De teller is groter dan die van de vorige dataleverantie
7. De CI verwijdert de data van de vorige dataleverantie uit de centrale gegevensset.
8. De CI integreert de data van de huidige dataleverantie in de centrale gegevensset.
9. De CI logt het resultaat van de verwerking.
10. De CI regeneert indexen op de data.
11. De CI regeneert statistische kengetallen over de data voor de catalogus.
12. De CI publiceert kengetallen over de data en de catalogus op de website van de catalogus van de centrale infrastructuur.

Alternatieve scenario's:

1. Indien er geen gegevens zijn aangeleverd, dan kan de use case gestopt worden.
- 2a-a. De gegevensbeheerder constateert afwijkingen in de dataset ten opzichte van de standaard.
- 2a-b. De gegevensbeheerder weigert de dataset. De beheerder van de betreffende instelling wordt hiervan op de hoogte gesteld per email. In de email is de reden van afwijzing opgenomen.

De instelling kan vervolgens de foutmelding onderzoeken en de dataset aanpassen. Vervolgens voert de instelling het scenario 3.4.2 uit om de aangepaste dataset als nog aan te leveren.

2d. Idem 2a.

Opmerkingen:

- In stap 2a worden integriteitcontroles uitgevoerd. Deze zijn beschreven in het parelsnoer informatie model.
- De statistische kengetallen in stap 4 zijn ook beschreven in het parelsnoer informatie model.

Open issues:

- Moet de controle van de dataset plaatsvinden tijdens de leverantie of tijdens de consolidatieslag?
- Alleen feedback per email?
- Na aanlevering van de dataset wordt de verbinding met de instelling verbroken. Daarmee is de identiteit van de instelling verloren (asynchroon). In de verdere afhandeling van de dataset dient de identiteit van de aanleverende instelling geborgd te worden.

3.6.2 Beschikbaar stellen dataset

Doel: De aangevraagde dataset wordt aan een onderzoeker beschikbaar gesteld.

Actor: Gegevensbeheerder CI

Aanleiding: Een ontvangen, geautoriseerd verzoek voor de dataset is door de gegevensbeheerder ontvangen van de Parelsnoer Commissie.

Voorwaarden:

- Het verzoek is door de beoordelingscommissie geautoriseerd.
- Het verzoek is volledig ingevuld

Scenario:

1. De gegevensbeheerder ontvangt een geautoriseerd verzoek om gegevens van een onderzoeker.
2. De gegevensbeheerder controleert de autorisatie van het verzoek om gegevens.
3. De gegevensbeheerder gebruikt de beheerfunctionaliteit van de CI om het verzoek te archiveren
4. De gegevensbeheerder stelt een database query op, op basis van het verzoek om gegevens.
5. De gegevensbeheerder archiveert de query met de beheerfunctionaliteit van de CI.
6. De gegevensbeheerder voert de query uit met de beheerfunctionaliteit van de CI.
7. De beheerfunctionaliteit genereert een unieke sleutel voor deze resulterende dataset en versleutelt de gepseudonimiseerde BSN's. Deze sleutel wordt bewaard.
8. De gegevensbeheerder archiveert de geconverteerde dataset.
9. De gegevensbeheerder converteert de resultaten naar het door de onderzoeker gevraagde XML formaat conform PIM
10. De gegevensbeheerder stelt de geconverteerde dataset beschikbaar aan de onderzoeker

Alternatieve scenario's:

6. Er is geen data beschikbaar dat voldoet aan de zoekcriteria. De onderzoeker wordt hiervan op de hoogte gesteld.

Opmerkingen:

- Overwogen wordt om op termijn te bekijken of formaten voor tools die statistische analyse kunnen uitvoeren ook ondersteund kunnen worden (bij stap 10).
Mogelijkheden hiervoor zijn:
- StatDataML (voor R en Matlab)
- XML Mapping (voor SAS)

Open issues:

- Zie ook 3.5.4, kan dit meer geautomatiseerd worden?
- Hoe wordt de correctheid van het omzetten van vraag naar query gegarandeerd?

3.6.3 Afhandelen verzoek biomateriaal

Doel: Het aangevraagde biomateriaal wordt aan een onderzoeker beschikbaar gesteld.

Actor: Gegevensbeheerder CI

Aanleiding: Een ontvangen, geautoriseerd verzoek voor biomateriaal is door de gegevensbeheerder ontvangen van de Parelsnoer Commissie.

Voorwaarden:

- Het verzoek is door de beoordelingscommissie geautoriseerd.
- Het verzoek is volledig ingevuld

Scenario:

1. De gegevensbeheerder ontvangt een geautoriseerd verzoek om biomateriaal
2. De gegevensbeheerder controleert de autorisatie van het verzoek.
3. De gegevensbeheerder gebruikt de beheerfunctionaliteit van de CI om het verzoek te archiveren
4. De gegevensbeheerder stelt een database query op, op basis van de het verzoek. Deze query resulteert in een overzicht van benodigd materiaal per aangesloten instelling
5. De gegevensbeheerder archiveert de query met de beheerfunctionaliteit van de CI.
6. De gegevensbeheerder voert de query uit met de beheerfunctionaliteit van de CI.
7. De gegevensbeheerder overlegt het verzoek en de uitwerking naar locatie van het biomateriaal met de transportfunctionaris. De beheerfunctionaliteit maakt hiertoe een email aan met alle details
8. De transportfunctionaris plant een route uit voor het ophalen van het biomateriaal bij de instellingen en het transporteren hiervan naar de onderzoeker.
9. De transportfunctionaris transporteert het biomateriaal en bevestigt aflevering bij de gegevensbeheerder
10. De gegevensbeheerder informeert de onderzoeker dat zijn verzoek is afgehandeld.

Alternatieve scenario's:

6. Er is geen data beschikbaar dat voldoet aan de zoekcriteria. De onderzoeker wordt hiervan op de hoogte gesteld.

Opmerkingen:

- Er wordt hier vanuit gegaan dat de Commissie rekening heeft gehouden met beschikbaarheid van materiaal. De CI voert slechts de opdracht tot transport uit.

Open issues:

- Geen

3.6.4 Afhandelen verzoek beeldmateriaal

Doel: Het aangevraagde beeldmateriaal wordt aan een onderzoeker beschikbaar gesteld.

Actor: Gegevensbeheerder CI

Aanleiding: Een ontvangen, geautoriseerd verzoek voor beeldmateriaal is door de gegevensbeheerder ontvangen van de Parelsnoer Commissie.

Voorwaarden:

- Het verzoek is door de beoordelingscommissie geautoriseerd.
- Het verzoek is volledig ingevuld

Scenario:

1. De gegevensbeheerder ontvangt een geautoriseerd verzoek om beeldmateriaal
2. De gegevensbeheerder controleert de autorisatie van het verzoek.
3. De gegevensbeheerder gebruikt de beheerfunctionaliteit van de CI om het verzoek te archiveren
4. De gegevensbeheerder stelt een database query op, op basis van de het verzoek. Deze query resulteert in een overzicht van benodigd beeldmateriaal per aangesloten instelling
5. De gegevensbeheerder archiveert de query met de beheerfunctionaliteit van de CI.
6. De gegevensbeheerder voert de query uit met de beheerfunctionaliteit van de CI.
7. De gegevensbeheerder verzoekt de betreffende instellingen om het beeldmateriaal aan te leveren bij de CI. De beheerfunctionaliteit kan hiertoe naar elke instelling een email sturen met het verzoek om beeldmateriaal aan te leveren.
8. De instellingen leveren het beeldmateriaal aan bij CI
9. De gegevensbeheerder bundelt het aangeleverde beeldmateriaal van de verschillende instellingen en stelt dit beschikbaar aan de onderzoeker
10. De onderzoeker haalt het beeldmateriaal op van de CI

Alternatieve scenario's:

6. Er is geen data beschikbaar dat voldoet aan de zoekcriteria. De onderzoeker wordt hiervan op de hoogte gesteld.
8. Een instelling levert het beeldmateriaal niet aan. De gegevensbeheerder neemt contact op met de contactpersoon. Bij uitblijven van aanlevering wordt geëscaleerd naar de onderzoekscommissie.

Opmerkingen:

- Momenteel wordt op verzoek het beeldmateriaal opgevraagd bij de instellingen en de CI zal geen beeldmateriaal opslaan. Op termijn kan overwogen worden om dit beeldmateriaal ook centraal op te slaan om instellingen te ontlasten van verzoeken van onderzoekers.

Open issues:

- De hele uitwerking van het beeldmateriaal wordt nog nader uitgewerkt binnen Parelsnoer.
- Onderzocht moet worden of het nodig is om de informatie in/over het beeldmateriaal gepseudonimiseerd moet worden.

3.6.5 Afhandelen verzoek tot de-pseudonimisatie

Het integrale scenario voor de-pseudonimisatie is beschreven in paragraaf 3.5.7.

3.6.6 Inzien log

Doel: Het bieden van een overzicht van de logregels van de Centrale Instelling

Actor: Gegevensbeheerder CI

Aanleiding: De gegevensbeheerder van de CI wil inzage in de gebeurtenissen op de CI.

Voorwaarden:

- De beheerder is geauthenticeerd door de portal van de CI door middel van gebruikersnaam/wachtwoord

Scenario:

1. De beheerder vraagt de portal van de CI om het overzicht van de log
2. De portal biedt de beheerder selectiemogelijkheden om specifieke delen van de log te kunnen zien. Er kan geselecteerd worden op gebeurtenistype, tijdstip en ernst van de logmelding.
 - Inloggen/uitloggen gegevensbeheerder CI
 - Aanleveringen gegevenssets totaal
 - Aanleveringen gegevenssets per instelling
 - Laatste aanlevering gegevensset (per parel) van iedere instelling
 - Falen verwerking gegevenssets totaal
 - Falen verwerking gegevenssets per instelling
 - Falen integratie van correcte gegevenssets in centrale database
 - Falen extractie van catalogus gegevens uit centrale database
 - Falen update catalogus informatie
3. De instellingportal toont een overzicht van de log op basis van de ingevulde selectiecriteria.
4. De beheerder klikt op een logregel
5. De instellingportal toont uitgebreide details van de logregel
6. De beheerder logt uit.

Alternatieve scenario's:

- Geen.

Opmerkingen:

- Geen

Open issues:

- Geen

3.6.7 Bekijken gearchiveerde verzoeken

Doel: Het bieden van een overzicht van gearchiveerde verzoeken voor datasets, biomaterialen en beeldmateriaal vanuit de Centrale Instelling

Actor: Gegevensbeheerder CI

Aanleiding: De gegevensbeheerder van de CI wil inzage in de verstrekte datasets, biomaterialen en beeldmateriaal vanuit de Centrale Instelling.

Voorwaarden:

- De gegevensbeheerder is ingelogd op de CI

Scenario:

1. De beheerder vraagt het systeem van de CI om het overzicht van de gearchiveerde verzoeken voor datasets.

2. Het systeem biedt de beheerder zoekmogelijkheden om specifieke verzoeken uit het archief te kunnen zien.
Er kan gezocht worden op:
 - Exacte indiendatum van het verzoek.
 - Bereik indiendatum van het verzoek (alle verzoeken na datum X en voor datum Y)
 - Exacte afhandeldatum van het verzoek.
 - Bereik afhandeldatum van het verzoek (alle verzoeken na datum X en voor datum Y)
 - Woorden in de titel van het verzoek.
 - Parels die geraadpleegd zijn.
 - Naam van een van de onderzoekers die het verzoek heeft ingediend.
 - Naam van een van de instellingen van de onderzoekers die het verzoek heeft ingediend.
3. Het systeem toont een overzicht van de verzoeken op basis van de ingevulde zoekcriteria.
4. De beheerder klikt op een specifiek verzoek.
5. Het systeem toont uitgebreide details van het verzoek, in ieder geval worden de volgende gegevens getoond:
 - Contactgegevens van de betrokken onderzoekers.
 - Datum van indienen van het verzoek door de onderzoeker.
 - Datum van beoordeling over het verzoek door de commissie.
 - Of het verzoek goedgekeurd, afgekeurd of nog in behandeling is.
 - De originele vraag van de onderzoeker.
 - De originele onderbouwing van de vraag van de onderzoeker.
 - Het resultaat van de vertaling van de vraag van de onderzoeker in een query die op de database kan worden uitgevoerd.
 - Eventueel de datum waarop de dataset is aangeleverd.
 - Eventueel de inhoud van de aangeleverde dataset en informatie op tweede pseudonimisatieslag ongedaan kan worden.
 - Overzicht van de specifiek aangevraagde biomaterialen
 - Overzicht van de specifiek aangevraagde beeldmaterialen
 - De betrokken parels die data hebben aangeleverd aan de onderzoeker.
6. De beheerder sluit het systeem af.

Alternatieve scenario's:

- Geen.

Opmerkingen:

- Geen

Open issues:

- Geen.

3.7 Scenario's voor de beheerder CI

3.7.1 Operationeel beheer CI

Het operationeel beheer van de centrale infrastructuur wordt uitgevoerd op basis van de beheerstandaarden ITIL (technisch beheer) en ASL (applicatiebeheer) zoals beschreven in het ontwerp van de beheerorganisatie. Deze procedures worden hier niet verder uitgewerkt.

3.7.2 Bijwerken datadictionary

Onderdeel van het applicatiebeheer is het beheer van de datadictionary in de catalogus. Een nieuwe versie van de PIM impliceert het bijwerken van de datadictionary.

N.B. Naast het bijwerken van de catalogus tot een nieuwe versie van de PIM zal een wijziging in de PIM veel meer impact hebben (bijv. het migreren van gegevens in het oude model naar het nieuwe model). Deze migratieproblematiek valt buiten scope van dit ontwerpdocument.

Openstaand issue:

- Kan de datadictionary geautomatiseerd afgeleid worden uit de databasestructuur? Dit kan technisch wel, maar is dit leesbaar voor de doelgroep?

3.7.3 Inzien log CI

Doel: Het bieden van een overzicht van de logregels van van de Centrale Instelling

Actor: Operationeel beheerder CI

Aanleiding: De operationeel beheerder van de CI wil inzage in de gebeurtenissen op de CI.

Voorwaarden:

- De beheerder is geauthenticeerd door de portal van de CI door middel van gebruikersnaam/wachtwoord

Scenario:

1. De beheerder vraagt de portal van de CI om het overzicht van de log
2. De portal biedt de beheerder selectiemogelijkheden om specifieke delen van de log te kunnen zien. Er kan geselecteerd worden op gebeurtenistype, tijdstip en ernst van de logmelding.
 - Inloggen/uitloggen gegevensbeheerder CI
 - Inloggen/uitloggen operationeel beheerder CI
 - Inloggen/uitloggen beheerder van aanleverende instellingen
 - Mislukte inlogpogingen van beheerders
 - Aanleveringen gegevenssets totaal
 - Aanleveringen gegevenssets per instelling
 - Laatste aanlevering van iedere instelling
 - Falen verwerking gegevenssets totaal
 - Falen verwerking gegevenssets per instelling
 - Falen integratie van correcte gegevenssets in centrale database
 - Falen extractie van catalogus gegevens uit centrale database
 - Falen update catalogus informatie
3. De instellingportal toont een overzicht van de log op basis van de ingevulde selectiecriteria.
4. De beheerder klikt op een logregel
5. De instellingportal toont uitgebreide details van de logregel
6. De beheerder logt uit.

Alternatieve scenario's:

- Geen.

Opmerkingen:

- Geen

Open issues:

- Geen

3.7.4 Gebruikersbeheer instellingportal

Doel: Het beheren van usernaam en wachtwoorden voor gebruikers van de instelling portal

Actor: Beheerder CI

Aanleiding: Een nieuwe gebruiker van een instelling meldt zich aan, een gebruiker meldt zich (als zijnde gebruiker van de instellingportal), een gebruiker heeft zijn wachtwoord vergeten.

Voorwaarden:

- De instelling van de gebruiker is geautoriseerd

De usernamen en de wachtwoorden van de instellingen worden door de Centrale Infrastructuur versleuteld opgeslagen. De beheerder kan usernamen en wachtwoorden toevoegen middels een webpagina van de beheerapplicatie. Om de beheerinspanningen te minimaliseren zal ook een pagina worden gemaakt waarop gebruikers (onderzoekers van instellingen) aan kunnen geven dat ze hun wachtwoord zijn vergeten. In dat geval wordt een geheime vraag gesteld die ze hebben ingevoerd bij registratie. Voorbeelden van dit soort vragen zijn "Wat is de meisjesnaam van uw moeder" en "Wat is de geboorteplaats van uw vader". Indien het juiste antwoord is gegeven op de geheime vraag wordt het wachtwoord opnieuw gegenereerd, en per e-mail naar de gebruiker gestuurd.

Scenario:

1. Een gebruiker meldt zich via een schriftelijke procedure aan; hij of zij wil gebruik maken van de instellingportal. De procedure voor het aanmelden als gebruiker is beschreven in het architectuurontwerp van de centrale infrastructuur. De beheerder maakt voor de gebruiker een gebruikersnaam/wachtwoord aan en verstuurt deze per email naar de aanvrager.
2. De beheerder maakt een gebruiker aan; een gebruiker heeft de volgende velden: usernaam, wachtwoord, secret question, e-mail adres, volledige naam. Als de beheerder de gebruiker heeft aangemaakt, zal het CI automatisch een e-mail sturen naar de nieuwe gebruiker met een gegenereerd wachtwoord plus een link om in te loggen. Het gegenereerde wachtwoord is eenmalig geldig. Bij de eerste keer inloggen moet de gebruiker zijn of haar wachtwoord wijzigen. De wachtwoorden zijn versleuteld opgeslagen door middel van een zogenaamde MD5 hash; hierdoor zijn wachtwoorden nergens direct op te vragen (dus ook niet voor de beheerder).
3. De gebruiker is zijn of haar wachtwoord vergeten, en klikt op de link 'Wachtwoord vergeten'. In de inleiding op deze use case is beschreven hoe dit verder verloopt.
4. Een coördinator of manager deelt de beheerder van het CI mede dat een bestaande gebruiker niet langer bevoegd is om in de instellingportal in te loggen. De beheerder kan via de beheer applicatie de gebruiker verwijderen, zodat deze niet meer in staat is om in te loggen.

3.7.5 Aansluiten instelling

Dit proces is bedoeld om een nieuwe instelling te laten aansluiten op de centrale infrastructuur. Dit proces wordt nader beschreven in het ontwerp van de beheerorganisatie en wordt hier niet nader uitgewerkt.

In het proces worden onder meer de volgende stappen uitgevoerd:

- Vastleggen administratieve gegevens
- Uitreiken Certificaat
- Toegang configureren in CI (ip-nummer, certificaat info)

3.7.6 Afsluiten instelling

Dit proces is bedoeld om een aangesloten instelling af te sluiten van de centrale infrastructuur. Het afsluiten van een instelling betekent dat alle gegevens van deze instelling uit de centrale database van de CI wordt verwijderd. Vervolgens wordt de autorisaties van de instelling verwijderd zodat toegang door de instelling niet mogelijk is. Toegang tot de catalogus zal wel mogelijk blijven.

Dit proces wordt nader beschreven in het ontwerp van de beheerorganisatie en wordt hier niet nader uitgewerkt.

3.7.7 Management rapportages

De centrale infrastructuur zal verschillende rapportages opleveren:

- Rapportages over het operationeel beheer. Dit is de SLA-rapportage
- Rapportages over het applicatie beheer. Hieronder valt rapportages over de aanlevering door instellingen van datasets van de parels. Ook de kwaliteit van deze datasets wordt gerapporteerd.
- Rapportages over het gebruik van de dienstverlening van de CI.

Deze rapportages zijn nader beschreven in het ontwerp van de beheerorganisatie en ze worden hier niet nader uitgewerkt.

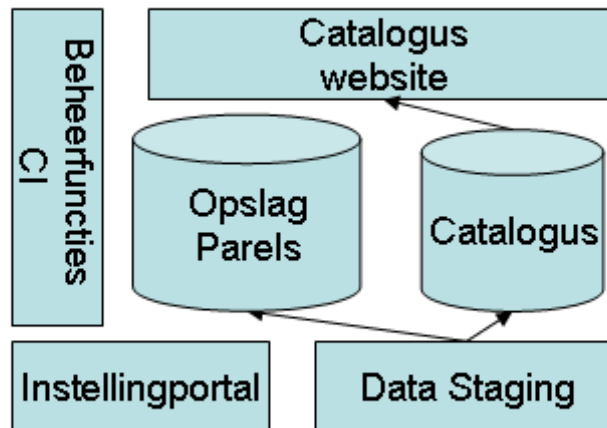
3.7.8 Bewaking aanlevering

De CI biedt overzichten voor de beheerder van de CI wanneer welke instelling welke dataset heeft aangeleverd. Tevens wordt inzichtelijk gemaakt welke instellingen een bepaalde aanlevering heeft gemist.

Op termijn kan deze controle meer geautomatiseerd uitgevoerd worden.

4 Functionaliteit van de CI

In paragraaf 2.4 is de globale werking van de CI beschreven. Hierin zijn de volgende componenten beschreven:



In meer detail betreft het de volgende functies:

- Ontvangst van gegevens: Data Staging.
Hier worden de dataset van de instellingen ontvangen en vervolgens verwerkt
- Ontsluiten catalogus
Dit is de website waarmee geïnteresseerden kunnen kijken wat Parelsnoer is en welke gegevens binnen Parelsnoer beschikbaar zijn. De website toont alleen totalen en metagegevens, niet de gegevens zelf.
- Beheerfuncties CI
 - Samenstellen en beschikbaar stellen dataset
Deze functie wordt gebruikt om een geautoriseerd verzoek om gegevens af te handelen
 - Aan-/afsluiten van een instelling
Deze functie wordt gebruikt als een instelling voor een Parel wordt aangesloten of afgesloten
 - De-pseudonimisatie
Deze functie wordt gebruikt om te achterhalen bij welke instelling een specifieke patiënt zich bevindt als hiervoor een verzoek is ontvangen.
- Instellingportal waar instellingen hun logging en status van aanleveringen kunnen inzien

Naast deze primaire functies van de CI zal de CI ook functies bevatten voor:

- Logging
Alle handelingen in en met de CI worden gelogd
- Rapportages
Het betreft hier management rapportages voor Parelsnoer
- Beveiligde communicatie
Alle transport tussen instellingen en de CI verloopt over beveiligde netwerkverbindingen

In dit hoofdstuk worden deze functies nader beschreven.

4.1 Ontvangst van gegevens: Data Staging

Zoals beschreven in de scenario's in 3.4 zullen instellingen op regelmatige basis een bestand per parel aanleveren bij de CI. Hierbij geldt dat alle gegevens worden aangeleverd (geen verschillenbestand) Het bestand wordt eerst aangeleverd bij de Data Staging component. Vervolgens wordt door de CI dit bestand gecontroleerd en verwerkt in de centrale database.

4.1.1 Interface

De aanlevering vindt plaats door het versturen van een bestand per Parel van de instelling naar de CI over een beveiligde verbinding. Hierbij worden de gegevens aangeleverd in een XML bestand. Het bestandsformaat is gedefinieerd in de PIM. In de PIM is een gedetailleerde beschrijving van het formaat beschikbaar.

Voor het transport wordt gebruik gemaakt van een beveiligde SSL verbinding, zie het architectuurontwerp van de centrale infrastructuur. Over deze SSL verbinding zijn twee protocollen mogelijk voor bestandsaanlevering:

- FTP
- SOAP over HTTP

Voor beide protocollen wordt asynchrone verwerking toegepast. Dit betekent dat een instelling een verbinding opbouwt, het bestand aanlevert en de verbinding afsluit. Daarna zal de CI het bestand controleren en verwerken. Hierna zal een hiervan terugkoppeling naar de instelling worden gestuurd per email.

4.1.2 Ontvangen van bestand

Bij ontvangst van een bestand is het van belang om de volgende informatie te ontvangen:

- Identiteit van de aanleverende instantie
- Het bestand
- Datum/tijd van aanlevering
- Volgnummer bestand. Dit volgnummer is opgenomen in het bestand. Dit volgnummer wordt door een instelling bijgehouden. De CI kan bijhouden wat het volgnummer is van de laatst verwerkte dataset. Als het volgnummer van het aangeleverde bestand lager is dan de laatst verwerkte, is er mogelijk sprake van een verkeerde volgorde van aanlevering.
- Identificatie van de Parel waarvoor dit bestand wordt aangeleverd

FTP

Doordat gebruik gemaakt wordt van tweezijdige SSL is er eigenlijk sprake van secure FTP (SFTP). De FTP server wordt zodanig ingericht dat na de upload van het bestand de bovengenoemde informatie is vastgelegd. Complexiteit hierbij is dat de FTP-server meestal geen functies heeft om iets specifiek te doen met de informatie in het certificaat.

Dit betekent voor de inrichting van de FTP server:

- Bij het inloggen op de FTP server moet gebruik gemaakt worden van de certificaat van de instelling: identificatie en authenticatie op basis van het certificaat. Dit stelt eisen aan de FTP client (bij de instellingen) en de FTP server (CI) voor het ondersteunen van deze vorm van authenticatie.
- Voor elke aangesloten instelling wordt een eigen directory gebruikt. Hierdoor kan de identiteit van de aanleverende instelling worden vastgehouden
- De datum/tijd van het vastgelegde bestand wordt gebruik als datum/tijd van de aanlevering (Datum/tijd van het bestandssysteem)

SOAP

Bij SOAP is het eenvoudiger om gebruik te kunnen maken van het certificaat dat door de instelling gebruikt wordt in de SSL-verbinding. Deze SSL-verbinding is nog actief als de server het bestand ontvangt.

Voor de inrichting geldt:

- De CI onthoudt welke instelling een bestand heeft aangeleverd op basis van het certificaat van de instelling in de SSL-verbinding
- De CI onthoudt de datum/tijd van aanlevering

De wijze van vastlegging wordt hier niet nader gespecificeerd. Dit kan het bestandssysteem zijn (vergelijkbaar met FTP inrichting) of in een database.

4.1.3 Verwerken van een bestand

Op regelmatige basis controleert de CI of er bestanden zijn aangeleverd en klaar staan in de Data Staging. Is dit het geval dan wordt *in volgorde van binnenkomst* de volgende stappen gedaan:

- Controle van integriteit van het bestand
- Verwerken van het bestand in de centrale database

Nadat alle bestanden op deze wijze zijn verwerkt wordt de catalogus bijgewerkt.

Controle van integriteit van het bestand

De controle houdt in dat de syntax van het bestand wordt gecontroleerd. De basis hiervoor is de formaatdefinitie in de PIM.

De structuur van het XML bestand wordt aan de hand van de XSD gecontroleerd.

Verder wordt gecontroleerd of de teller die in het bestand is opgenomen groter is dan de teller die in het bestand van de vorige aanlevering was opgenomen. Deze volgorde teller is bedoeld om te voorkomen dat om wat voor reden dan ook een oudere dataset een nieuwere overschrijft.

Verwerken gereedstaande aangeleverde dataset

Als de bestanden gecontroleerd zijn kan deze ingelezen worden. De strategie is hierbij om eerst de oude gegevens te verwijderen en daarna de nieuwe gegevens in te lezen. Mocht er ondanks de controle een probleem zijn met de nieuwe dataset kan de situatie ontstaan dat de oude gegevens weg zijn en de nieuwe gegevens niet ingelezen kunnen worden. Daarom de volgende aanpak

1. De nieuw aangeleverde gegevens worden eerst ingelezen in een tijdelijke tabel die dezelfde definitie kent als de tabel waarin de gegevens uiteindelijk terecht moeten komen.
2. Van elk bestand is bekend door welke instelling deze is aangeleverd en voor welke Parel het gegevens bevat. Op basis van deze gegevens kunnen de gegevens van de vorige aanlevering uit de centrale database verwijderd worden. Oude aanleveringen worden dus niet historisch bijgehouden.
3. Vervolgens worden de nieuw aangeleverde gegevens ingelezen en verwerkt in de centrale database door deze te kopiëren van de tijdelijke tabel naar de definitieve.

Ten slotte wordt informatie over deze verwerking vastgelegd:

- wie heeft aangeleverd,
- wanneer aangeleverd (datum aanlevering, datum verwerking),
- wat is aangeleverd (aantal records)
- het volgnummer van de aanlevering.

Terugkoppeling bestandsverwerking

Er wordt een email gestuurd naar de aanleverende instelling met daarin een terugkoppeling of de aanlevering en verwerking goed is verlopen. Aan de titel van de email is duidelijk te herkennen of deze aanlevering en verwerking goed of fout is gegaan.

Bijwerken catalogus

Als alle bestanden zijn verwerkt kan voor elke Parel die voorzien is van bijgewerkte gegevens de statistische kentallen herberekend worden. De kentallen zijn per Parel bepaald in de PIM.

4.2 Ontsluiten catalogus

De catalogus is een publiek toegankelijke website. 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, niet de gegevens zelf. De website bevat de volgende onderdelen

- Welkomst pagina, contact pagina
- Gedeelte met algemene informatie over het project Parelsnoer
- Gedeelte met informatie over de verschillende Parels. Hier zijn ook de kentallen per parel te vinden, voor gegevensset, biomateriaal en/of beeldmateriaal. Ook is de datadictionary van een parel beschikbaar.
- Gedeelte met informatie (w.o. aanvraagformulier) voor onderzoekers over indienen verzoek voor een gegevensset, biomateriaal en/of beeldmateriaal. Het aanvraagformulier is een PDF bestand of iets dergelijks dat geprint of gemaïld kan worden. Ook is een webformulier mogelijk.

De website houdt zich aan de webrichtlijnen van de overheid. Daarnaast wordt de Parelsnoer huisstijl toegepast.

4.3 Samenstellen dataset

4.3.1 Interface

Er zijn twee 'interfaces' voor dit functionele component. Een binnenkomend interface: het verzoek van een onderzoeker en een uitgaande interface: het bestand met het resultaat van de zoekopdracht.

Verzoek

Voor het verwerken van een verzoek van een onderzoeker wordt vanuit de CI gestart met een door de Parelsnoer commissie geautoriseerd verzoek.

Er is nog weinig bekend over de wijze waarop een onderzoeker een verzoek op stelt. Voor dit ontwerp is uitgegaan dat een onderzoeker de catalogus gebruikt om te bepalen welke gegevens hij wil ontvangen. Op basis van de datadictionary in de catalogus kan de onderzoeker aangeven welke tabellen en welke kolommen hij wil ontvangen. Deze informatie is opgenomen in zijn verzoek.

Resultaat dataset

Voor het exportbestand zijn verschillende formaten mogelijk:

- Een XML bestand, opgemaakt conform PIM. Hierin worden alleen de tabellen en kolommen uit de SQL query opgenomen
- Een CSV bestand. Per tabel wordt een apart bestand gegenereerd. De eerste regel bevat de kolomnamen van de tabel.

Onderzocht wordt of ook andere formaten hier toegepast kunnen worden. Mogelijke formaten kunnen zijn: StatDataML (voor R en Matlab) en XML Mapping (voor SAS)

4.3.2 Zoekfunctionaliteit

Deze functionaliteit bestaat uit het opzoeken van onderzoeksgegevens die voldoen aan het verzoek en daarnaast de functionaliteit om het resultaat van de zoekopdracht te exporteren.

Zoeken

Voor het basisplatform wordt er vanuit gegaan dat de gegevensbeheerder van de CI gebruik maakt van SQL om de dataset samen te stellen.

De databeheerder van de CI vertaalt het ontvangen formulier naar SQL. Deze SQL en het geautoriseerde verzoek worden in de CI vastgelegd. De SQL wordt uitgevoerd en daarna wordt het resultaat geëxporteerd.

N.B. Op basis van de PIM kan bekeken worden of er een eenvoudigere interface te maken is voor een databeheerder. De PIM was nog in ontwikkeling ten tijde van het schrijven van deze versie van het ontwerp.

Versleutelen van gepseudonimiseerd BSN en Exporteren

Na het uitvoeren van de SQL op de centrale database wordt het resultaat geëxporteerd naar een bestand. In het architectuurontwerp is gekozen om de gepseudonimiseerde BSN nog eens te versleutelen (om het bundelen van verschillende datasets voor niet geautoriseerd onderzoek onmogelijk te maken).

Voor het versleutelen van de gepseudonimiseerde BSN worden de volgende stappen uitgevoerd

- Genereren unieke RSA sleutel voor dataset. Elke afgegeven dataset krijgt eigen sleutel. Deze sleutel wordt vastgelegd bij de aanvraag in de CI, zie 4.5
- Vervangen gepseudonimiseerd BSN door versleuteld, gepseudonimiseerd BSN

Voor het versleutelen wordt het RSA algoritme toegepast met een sleutelgrootte van 2048

4.3.3 Beschikbaar stellen resultaat dataset

De resulterende dataset wordt aan de onderzoeker beschikbaar gesteld via publicatie op een besloten gedeelte van de Parelsnoer Website. Op basis van de contactgegevens in het verzoek zal de CI de onderzoeker informeren over de locatie van de dataset en de toegangscode om hierbij te kunnen.

De dataset zal gedurende twee weken hier beschikbaar zijn, daarna wordt de dataset weer verwijderd. Deze periode wordt in de email naar de aanvrager opgenomen.

In het kader van de implementatie kan onderzocht worden of – technisch gezien – de dataset als bestand wordt klaargezet of dat deze *on demand* wordt gegenereerd als de onderzoeker deze opvraagt. In de laatste situatie krijgt de onderzoeker altijd de laatste gegevens.

N.B. Als datasets heel groot worden, kan op termijn overwogen worden om dit ook via FTP beschikbaar te stellen. Het proces verloopt verder vergelijkbaar.

4.4 Logging

Elke handeling van elke actor op de CI wordt gelogd.

Via het instellingenportaal (zie 4.6) kan een instelling de log voor zijn instelling inzien. Deze toont alle handelingen die zijn gedaan met zijn dataset: aanleveringen en verwerkingen van aanleveringen. Het gebruik van dit portaal is beschreven in 3.4.3.

De beheerder van de CI kan de totale log inzien.

4.5 Beheerfuncties

Voor de gegevensbeheerder zijn functies beschikbaar voor:

- Het afhandelen van een de-pseudonimisatie verzoek
- Het voeren van sleutelbeheer (voor versleuteling bij uitgifte datasets)
- Administratieve handelen voor aan-/afsluiten instelling

De-pseudonimisatie

- Gegeven een referentie of id van een afgegeven dataset en een versleutelde, gepseudonimiseerde BSN kan de CI sleutel opzoeken die gebruikt was om te versleutelen. Met die sleutel wordt ontsleuteld en wordt daarmee de gepseudonimiseerde BSN bepaald.

Sleutelbeheer

- Sleutels worden versleuteld opgeslagen in de database. Hiervoor is een speciale beheerderssleutel bedoeld
- De beheerderssleutel wordt goed beveiligd opgeslagen

Aan-/Aansluiten instelling

Voor het aansluiten van een instelling op de CI zal een aansluitproces worden doorlopen, zoals globaal beschreven in het programma van eisen Instellingen. In de CI worden de gegevens van een instelling vastgelegd. Hiervoor biedt de CI een beheerdersinterface. Via deze interface kan ook een unieke identificatie worden toegekend aan een instelling.

Daarnaast zal vanuit de CI een SSL-certificaat verstrekken aan de instelling. In dit certificaat is de uitgegeven unieke identificatie van een instelling opgenomen. Er wordt er vanuit gegaan dat voor certificaatbeheer een standaard dienst wordt afgenomen van een dienstverlener (bijv. KPN/GPR of Diginotar, etc.)

Bij het afsluiten van de instelling wordt de autorisatie van de instelling ingetrokken en worden alle gegevens van die instelling uit de centrale database verwijderd. Eventuele gegevens van de instelling in gearchiveerde uitgegeven datasets worden niet verwijderd.

Het uitgegeven certificaat wordt, als nodig, op de CRL van de certificaat leverancier gezet.

Ten slotte is er de mogelijkheid om gebruikers-/wachtwoordbeheer uit te voeren, zoals beschreven in 3.7.4.

4.6 Instellingenportaal

De instellingenportaal is een website voor instellingen waar zij inzage kunnen krijgen in

- De log. Hierin staan alle handelingen van de instelling op de CI.
- Resultaten van aanlevering en verwerking datasets.

4.7 Beveiliging

In de CI dienen voldoende maatregelen te worden genomen om de vertrouwelijkheid van gegevens in de CI te kunnen waarborgen. Dit zijn de standaard maatregelen voor bescherming van externe koppelingen (firewall). Daarnaast dient de beheerorganisatie procedures in werking te stellen om een beveiligde beheeromgeving te realiseren.

Voor de verschillende ingangen in de CI zijn verschillende beveiligingseisen, zoals beschreven in onderstaande tabel.

Web applicatie	Transport protocol	Beveiligingseisen	Authenticatie
Catalogus	HTTP	Eenzijdige SSL, alleen server	N.v.t. (publiek toegankelijk)
Data staging	HTTP of FTP	Tweezijdige SSL	PKI certificaat (X.509)
Data management	HTTP	Eenzijdige SSL	Gebruikersnaam, wachtwoord
Algemeen beheer	HTTP	Eenzijdige SSL	Gebruikersnaam, wachtwoord
Instellingportal	HTTP	Eenzijdige SSL	Gebruikersnaam, wachtwoord

Voor de transportbeveiliging tussen instellingen en de CI wordt gebruik gemaakt van SSL 3.0/TLS 1.0

De toegang tot een uitgegeven dataset door een onderzoeker gebeurt doordat de onderzoeker een uniek gegenereerd token opgestuurd krijgt naar het emailadres in het geautoriseerde verzoek. Met deze token kan de dataset gedownload worden. Hierbij wordt op transportniveau gebruik gemaakt van eenzijdige SSL (servercertificaat in de CI)

Beheerders van de CI kunnen met gebruikersnaam/wachtwoord toegang krijgen tot de CI.

Bij SSL dienen partijen zich te vergewissen dat het certificaat van de andere partij valide is:

- het certificaat is onveranderd
- het certificaat is uitgegeven door de juiste CA
- het certificaat is niet verlopen
- het certificaat is niet ingetrokken (staat niet op de lijst met ingetrokken certificaten: de Certificate Revocation List of CRL)

Dit zijn standaard controles binnen SSL en PKI.

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.

4.8 Rapportages

De databeheerder kan vanuit de CI verschillende management rapportages genereren, zoals beschreven in 3.7.7. Deze rapportages kunnen geëxporteerd worden naar RTF (inleesbaar in Word of andere tekstverwerkers) of PDF. Waar mogelijk kan ook geëxporteerd worden naar CSV (inleesbaar in spreadsheets)

5 Gegevens in de CI

5.1 Parelsnoer Informatie Model

N.B. Ten tijde van het schrijven van deze versie van het ontwerp was de PIM nog niet gereed. Deze paragraaf dient daarom nog aangevuld te worden.

Het Parelsnoer Informatie Model is uitvoerig beschreven. De centrale database zal worden ingericht conform dit model. Hierbij worden de volgende uitgangspunten gehanteerd:

- Op het model wordt versiebeheer gevoerd
In een aanlevering van onderzoeksgegevens is in het bestand de versie van de PIM opgenomen volgens welke de gegevens zijn gespecificeerd.
- De PIM draagt zorg voor het voldoende anonimiseren van de gegevens, met uitzondering van de BSN.
- Bij de aanlevering van een bestand zijn in het bestand door de instelling alle BSN's vervangen door gepseudonimiseerde BSN's. De CI ontvangt nooit een BSN.
- In het PIM worden entiteiten die binnen verschillende Parels voorkomen eenmalig gedefinieerd. De definitie wordt gedeeld tussen de Parels en deze zal, waar mogelijk, gebaseerd zijn op open internationale standaarden (zoals beschreven in de architectuurbeschrijving: HL7v3 RIM)

Op hoofdlijnen bevat de PIM definities voor de volgende Parels:

- IBD
- ...

5.1.1 IBD

Klinische gegevens

Binnen IBD worden de volgende entiteiten vastgelegd:

- Patiënt
- ...

Biomateriaal

Binnen IBD zijn de volgende entiteiten vastgelegd omtrent biomateriaal:

- ...

Beeldmateriaal

Binnen IBD zijn de volgende entiteiten vastgelegd omtrent beeldmateriaal:

- ...

5.2 Afgehandeld verzoek en uitgegeven datasets

Van elk verzoek dat door de CI in behandeling is genomen wordt het volgende vastgelegd:

- Datum van ontvangst formulier
- Het ontvangen formulier met het verzoek, w.o. de contactgegevens van de onderzoeker
- De SQL die is gebruikt om de dataset samen te stellen
- De RSA sleutel die is gebruikt bij het versleutelen van het gepseudonimiseerd BSN
- Datum/tijd van generatie van de dataset
- Datum/tijd versturen email naar onderzoeker

- Email die verstuurd is naar onderzoeker
- Datum/tijd van ophalen van dataset door onderzoeker
- De verstuurde dataset

5.3 Ontvangen datasets

Bij het vastleggen van aangeleverde datasets is het nodig om

- Gegevens over de aanlevering vast te leggen, zie 4.1.
- Per record vast te leggen bij welke aanlevering dit record is aangeleverd.
- Per record vast te leggen welke instelling deze heeft aangeleverd. Hiervoor kan de unieke identificatie van de instelling gebruikt worden, zie 4.5.

Er is overlap mogelijk waarbij van een patiënt voor hetzelfde gegeven meerdere aanleverende instelling kent. In dit geval wordt er vanuit gegaan dat voor dat gegeven meerdere records worden aangemaakt, voor elke aanleverende instelling één.

Van een ontvangen dataset worden over de aanlevering gegevens vastgelegd:

- Datum/tijd van aanlevering
- Datum/tijd van de verwerking door CI
- Instelling die dataset heeft aangeleverd (unieke identifier)
- Parel waarbij de dataset hoort
- Volgnummer aanlevering

De gegevens in de dataset zijn gedefinieerd in het parelsnoer informatie model. Hierin wordt tevens opgenomen

- Gepseudonimiseerd BSN.
Het BSN wordt door de instelling voor de versturing gepseudonimiseerd
- Versie PIM
De versie van het PIM volgens welke dit bestand is opgemaakt
- Identifier instelling
Unieke identificatie van de aanleverende instelling
- Volgnummer van aanlevering
Een instelling houdt per bestand een teller bij die bij elke aanlevering wordt opgehoogd. De CI kan hierdoor in de gaten houden dat een bestand een hoger volgnummer heeft dan een eerder ontvangen bestand.

5.4 Administratieve informatie

Instellingen

- Algemeen contactpersoon voor Parelsnoer
- Uniek nummer voor de instelling. Wordt tevens opgenomen in certificaat
- Domeinnaam / ip-nummer van de instelling
- Gebruikersnaam en wachtwoord voor toegang tot instellingportal. Eventueel krijgen meerdere personen in een instelling toegang.

Per Parel

- E-mail en Naam contactpersoon aanleveringen
- E-mail en Naam contactpersoon opvragen beeldmateriaal
- E-mail en Naam contactpersoon opvragen biomateriaal
- E-mail en Naam contactpersoon opvragen klinische data

5.5 Logging

Bij het loggen wordt een soort audit trail vastgelegd; in principe worden alle relevante gebeurtenissen binnen de CI vastgelegd.

De volgende gegevens worden vastgelegd:

- datum en tijdstip van de actie
- instelling die de actie heeft uitgevoerd
- ID van de dataset (indien van toepassing)
- ID van de parel (indien van toepassing)
- Soort actie (bijvoorbeeld 'upload dataset')
- Omschrijving van de actie
- IP adres van de oorsprong van het verzoek
- Identiteit van de instelling