

Data4lifesciences

Een gedeelde data infrastructuur voor biomedisch onderzoek

Verantwoord omgaan met wetenschappelijke data vergt veel van onderzoekers. De hoeveelheid data is enorm toegenomen en er worden steeds hogere eisen aan de omgang met data gesteld. Privacy-eisen, maar ook eisen op het gebied van gebruik en hergebruik. Onderzoeksdata moet verzameld, opgeslagen, bewerkt, geanalyseerd, bewaard en met anderen gedeeld kunnen worden. Om dit te vergemakkelijken bestaan er onderzoeks infrastructuren: generieke oplossingen voor het omgaan met wetenschappelijke data. Zo'n infrastructuur bestaat niet alleen uit technische voorzieningen (hard- en software), maar richt zich ook op processen van kwaliteitsborging en expertise bij de onderzoekers en beheerders.

Achtergrond

De kwalitatief hoogwaardige infrastructuur die nodig is om aan de data-eisen te voldoen, ontstaat bij de individuele onderzoeker en, in steeds belangrijkere mate, ook de afzonderlijke universitair medische centra (umc's). Daarom kiezen de umc's samen met landelijke programma's als TraIT, BBMRI-NL, Parelnoer, DTL, AcZie, Mondriaan en SURF voor een gezamenlijke aanpak onder regie van de NFU: Data4lifesciences. Dit programma zal een innovatieve onderzoeksdata infrastructuur opzetten in, voor, door en tussen de umc's en hun partners.



Biomedisch onderzoek richt zich steeds meer op de individuele patiënt. Via 'personalised medicine' wordt geprobeerd de behandeling precies toe te snijden en deze zo aan te passen aan iedere patiënt, dat deze optimaal geholpen wordt. De technologische revolutie in onder andere de genetica en beeldvormende technieken (MRI, CT, etc.) maakt dit mogelijk, maar zorgt tegelijkertijd voor een explosieve groei aan onderzoeksdata. Data die ook nog eens erg complex zijn. Daarbij speelt ook dat voor personalised medicine onderzoek grote cohorten (groepen patiënten met eenzelfde eigenschap(en), bijvoorbeeld alle mannen geboren tussen 1940-1950) nodig zijn. Die cohorten zijn alleen samen te stellen via (internationale) samenwerking.

Onderzoekers hebben dus te maken met veel meer (complexe) data, maar ook met strengere eisen op het gebied van kwaliteit, beheer en het beschikbaar maken van die data voor anderen. Zoals eerder geschetst is dit de aanleiding om specifieke onderzoeksdata infrastructuren op te zetten, allereerst op kleinere schaal binnen individuele umc's en vakgebieden. Het programma Data4lifesciences beoogt deze lokale initiatieven aan te laten sluiten bij nationale en internationale infrastructuren. Hiervoor zijn een gezamenlijke aanpak, bestuurlijke afstemming en regie op landelijk niveau onontbeerlijk.

Ambitie

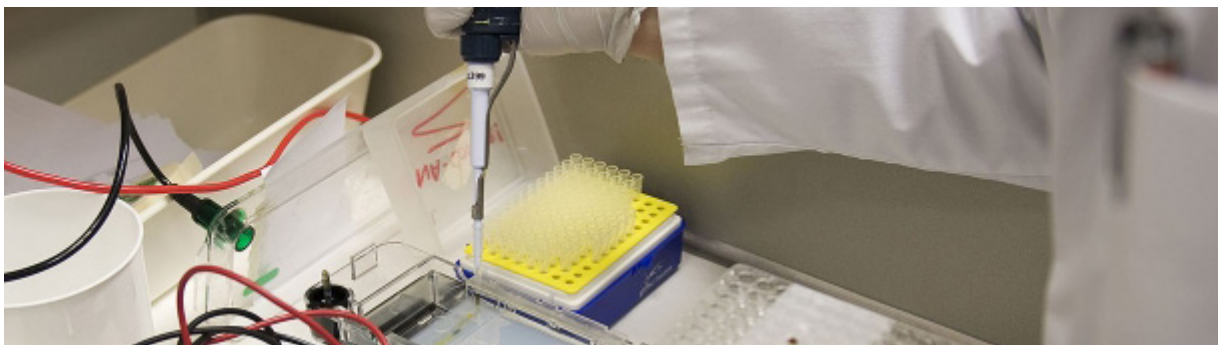
Het Data4lifesciences programma heeft haar doel bereikt als:

Alle aan umc's verbonden artsen en onderzoekers de onderzoeksdata infrastructuur gebruiken om klinische en experimentele data van alle aan umc's verbonden patiënten op te vragen en beschikbaar te stellen. Daarnaast wordt de infrastructuur ingezet voor het vinden en aanvragen van biologisch materiaal. Deze infrastructuur houdt niet op bij de muren van de umc's, maar wordt ook gebruikt en beschikbaar gesteld voor samenwerkingspartners in binnen- en buitenland.

De infrastructuur vormt een nationale virtuele samenwerkingsomgeving waarin data wordt geregistreerd, verwerkt, geanalyseerd, gearchiveerd en gedeeld. De data is FAIR (Findable, Accessible, Interoperable en Reusable) en wordt beschikbaar gesteld in een schaalbare, gedistribueerde omgeving, waarbij de benodigde reken capaciteit voor het verwerken van de data komt van nationale en umc rekenfaciliteiten. De infrastructuur is toegankelijk voor alle onderzoekers en artsen, onafhankelijk van instelling of locatie. De privacy van de patiënt is hierbij gewaarborgd.

Onderzoekers en artsen met vragen over de omgang met data hebben daartoe de beschikking over een uitgebreid data expertise netwerk. Deze experts in de umc's zijn het eerste aanspreekpunt voor deze vragen, waarbij de oplossingen zowel binnen als buiten de umc's wordt gezocht. Educatie is hierbij essentieel.

Waarom nu opportuun?



- Deelprojecten als TraIT, BBMRI-NL, Parelsnoer en DTL hebben veel bereikt op het gebied van data-infrastructuur, maar er is gecoördineerde actie nodig om de behaalde resultaten te behouden en verder uit te bouwen.
- Urgente redenen om de huidige infrastructuur aan te passen zijn de komst van de nieuwe EU-privacyrichtlijn, de implementatie van nieuwe Elektronische Patiënten Dossiers (EPD's) en de hogere eisen die door belangrijke sponsors van onderzoek (bijvoorbeeld NWO, ZonMW en KWF) worden gesteld aan de kwaliteit van datamanagement en herbruikbaarheid van data. Data moeten FAIR zijn: findable, accessible, interoperable en reusable.
- Het gezamenlijk optreden van de umc's versterkt de concurrentiepositie in Nederland en in Europa. De huidige data infrastructuur zorgt voor een goede uitgangspositie, maar die positie kan alleen gehandhaafd blijven als de umc's samen optreden om de infrastructuur regelmatig te vernieuwen. Dit is een noodzakelijke voorwaarde om blijvend substantiële financiering aan te trekken, om nieuwe Europese programma's te kunnen benutten en de concurrentie voor te blijven. Daar komt bij dat Europese infrastructuren een steeds belangrijke rol spelen bij de acquisitie van nieuwe middelen.
- De umc's zijn verantwoordelijk voor de kwaliteit van onderzoeksdata en de zorgvuldigheid waarmee deze worden verzameld, opgeslagen, bewerkt en gearchiveerd, alsmede voor het voldoen aan de relevante regelgeving ten aanzien van privacybescherming en patiëntveiligheid. Een kwalitatief hoogwaardige landelijke infrastructuur is noodzakelijk om de kans op reputatieschade te verminderen die zou kunnen optreden als individuele onderzoekers zich niet houden aan de geldende wetgeving. Dit geldt nog in sterkere mate voor het indammen van risico's van misbruik, wetenschappelijke fraude en beveiligingslekken.

Gezamenlijke infrastructuur



Een data infrastructuur is een generieke oplossing voor een specifiek aspect van de omgang met wetenschappelijke data, bedoeld om onderzoekers te ontlasten en optimaal gebruik te laten maken van de vele mogelijkheden die IT-technologie het onderzoek kan bieden. Deze infrastructuur kan van alles zijn; een online catalogus van monsters in een biobank, een standaard methode waarop data uit een EPD beschikbaar wordt gemaakt, privacyrichtlijnen, de wijze waarop de IT wordt ingericht binnen umc's, een handleiding waarin wordt uitgewerkt hoe onderzoekers het best met data om kunnen gaan ('data stewardship'), een generieke manier waarop data kunnen worden uitgewisseld, of een expert die een onderzoeker ondersteunt bij datavraagstukken. De beoogde hoogwaardige data infrastructuur die Data4Lifesciences zal opleveren omvat dus niet alleen technische voorzieningen (hard- en software), maar ook systemen en processen voor het borgen van de kwaliteit, voor de vereiste expertise bij onderzoekers en beheerders. Daarnaast zullen onderzoekers ondersteund worden door experts in zogenoemde 'shared service centers'.

Data4lifesciences waarborgt bestuurlijke afstemming, zodat lokale voorzieningen en expertisenetwerken aansluiten bij landelijke en internationale infrastructuren en vice versa. Ook algemene informatie over data-aspecten van onderzoek en de antwoorden op bijvoorbeeld onderstaande vragen zijn straks gemakkelijk te vinden voor de onderzoeker.

1. Ik doe een EU-studie. Mag ik de DNA-sequenties van deelnemers delen 'in the cloud'?
2. Ik combineer zorg- en onderzoeksdata in het EPD. Hoe krijg ik die er weer uit?
3. Ik moet voor een subsidieaanvraag een datamanagementplan schrijven. Wat moet daar in staan?
4. Ik heb tijdelijk veel rekencapaciteit nodig. Waar haal ik die vandaan?
5. Zijn de bloedmonsters die ik nodig heb al ergens beschikbaar in een biobank, zodat ik niet opnieuw patiënten hoeft te benaderen?
6. In mijn onderzoek is een groot aantal ziekenhuizen betrokken. Hoe breng ik de benodigde data bij elkaar en wat moet ik daarvoor regelen?

Programmaliijnen binnen Data4lifesciences



Om tot een gezamenlijke onderzoeksdata infrastructuur te komen, worden in de periode tot eind 2015 de volgende deelprojecten gerealiseerd:

1. *Harmonisatie van richtlijnen voor databeheer ('data stewardship')*.
Deze programmalijn ontwikkelt een richtlijn voor data stewardship beleid, in de vorm van een website met pointers naar lokale en landelijke expertise. Een beleid voor, en gedragen door alle umc's, is een belangrijke voorwaarde voor delen en hergebruik van onderzoeksdata, en de uiteindelijke gezamenlijke infrastructuur. Het vormt in feite de basis voor implementatie van de infrastructuur in de umc's. De richtlijnen betreffen onder andere patiëntveiligheid, bescherming van de privacy, betrokkenheid van patiënten, ethiek, betrouwbaarheid en herkomst van data, monitoring kwaliteit, en wet- en regelgeving.
Het Erasmus MC coördineert deze programmalijn, uiteraard in samenwerking met verschillende specialisten van de umc's, universiteiten, kennisinstituten, bedrijven en financiers.
2. *Harmonisatie van processen en IT-inrichting (architectuur)*
Deze programmalijn inventariseert bestaande IT-architecturen en richtlijnen voor het verzamelen, bewerken en uitvoeren van data binnen de umc's en de samenwerkingspartners.
De informatie wordt gedeeld met alle betrokkenen, waarna een referentie architectuur zal worden gekozen voor die deelgebieden die cruciaal zijn voor kwalitatief hoogwaardige onderzoeksdata. De coördinatie ligt in handen van de bestaande architectuurwerkgroep van CTMM/TraIT en de NFU Special Interest Group PRIMA (waarin de IT-architecten van de huizen deelnemen).
3. *Toegang tot data en samples (catalogus)*
Deze programmalijn wil de diverse collecties van biomedische samples (weefsel, bloed, urine, etc.) en data sets toegankelijk maken voor onderzoekers via een gezamenlijke catalogus. De registers van bijvoorbeeld BBMRI-NL (inclusief umc-biobanken), LifeLines, Parelsnoer, TraIT en PALGA worden verbonden, waardoor er op nationaal niveau naar samples en data gezocht kan worden. Ook de uitgifte procedures voor data en samples worden op elkaar afgestemd. Aspecten als beveiliging van privacy en informed consent worden hierbij goed geborgd. BBMRI-NL coördineert de programmalijn vanuit een nationale cataloguswerkgroep met vertegenwoordigers uit de bovengenoemde initiatieven aangevuld met bijdragen uit Europese projecten als BioSHaRE en BioMedBridges.
4. *Delen en analyse van biomedische data*
Deze programmalijn richt zich op de uitdaging die grote nationale onderzoeken hebben bij het verzamelen en integreren van data uit de verschillende ziekenhuizen en het vervolgens gezamenlijk analyseren van die data. Hiertoe wordt op basis van praktijkvoorbeelden een gezamenlijk research data platform beschikbaar gesteld. Dit gebeurt op basis van bestaande 'best practices', niet alleen van de umc's maar ook van landelijke programma's als Parelsnoer, CTMM-TraIT, BBMRI-NL en Mondriaan.
5. *Gebruik van digitaal opgeslagen patiëntgegevens voor onderzoek*
Rechtstreeks hergebruik van zorgdata voor biomedische onderzoek gebeurt nog maar mondjesmaat; overtypen van zorgdata in onderzoekssystemen is nog steeds de norm. Deze programmalijn beoogt hierin verandering te brengen en kent daartoe een nauwe samenhang met het NFU-programma Registratie aan de bron waarin eenmalige en eenduidige vastlegging van patiëntgegevens wordt geregeld. De eerste resultaten zullen bestaan uit een serie pilotprojecten samen met Registratie aan de bron, CTMM-TraIT en Parelsnoer.
6. *Goed onderzoek in de praktijk*
Privacy is een heet hangijzer binnen het wetenschappelijk onderzoek. De patiënt moet er op kunnen vertrouwen dat medische gegevens niet openbaar worden, maar de technologische ontwikkelingen in 'big data' maken dat steeds lastiger. Deze programmalijn implementeert procedures en technologieën voor privacy en beveiligingsaspecten volgens de richtlijnen gespecificeerd in programmalijn 1 (data stewardship). Hierbij valt te denken aan pseudonimisatie van persoonsgegevens via Trusted Third Parties (TTP), encryptie van burgerservicenummers, en beveiligingsstandaarden.

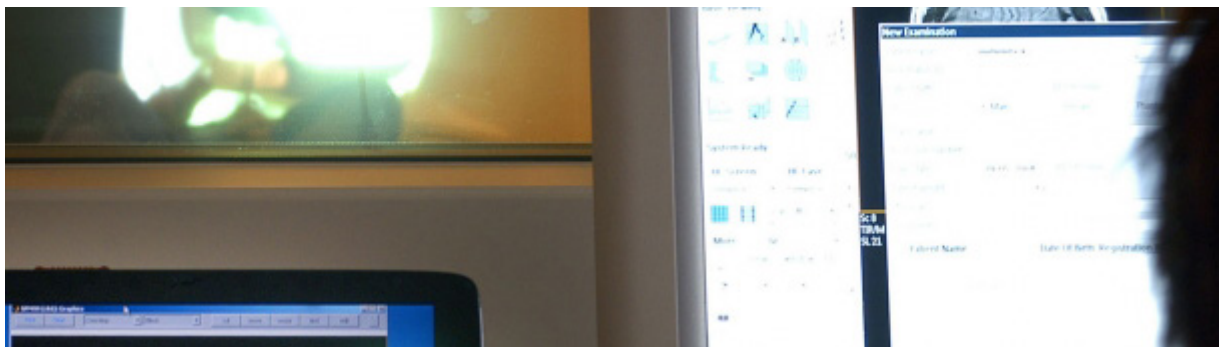
7. *Faciliteiten voor kwalitatief hoogwaardige dataverwerking*

Deze programmalijn richt zich op de benodigde rekencapaciteit voor het verwerken van onderzoeksdata, voornamelijk via High Performance Computing Cloud systemen. Zaken die hier spelen zijn onder andere het makkelijk kunnen inschakelen van extra piekcapaciteit, het delen van best practices en het gezamenlijk optrekken richting leveranciers. De coördinatie is in handen van SURF. Het programma is (inter)nationaal ingebed via BBMRI-NL/EU, EYR, CTMM/TraIT, en ELIXIR.

8. *Toegang tot experts, training en ondersteuning*

Onderzoekers met data stewardship vragen (zie ook programmalijn data stewardship) kunnen terecht bij lokale experts in de umc's. Hiertoe wordt een landelijk expertise netwerk opgericht, waarin de huidige umc data desks een belangrijke schakel vormen. DTL heeft een uitstekend netwerk om kennisuitwisseling en – integratie tussen de umc's en andere stakeholders te bevorderen en zal deze programmalijn coördineren.

Organisatie en informatie



Data4lifesciences is een programma waarin de umc's samen met vele specialistische samenwerkingspartners participeren. Onder aansturing van voorzitters Frank Miedema (UMC Utrecht) en Folkert Kuipers (UMCG) en programmamanager Jan Willem Boiten (CTMM) werken de programmacommissie, vele werkgroepen en andere samenwerkingsverbanden aan de realisatie van de geïntegreerde onderzoeksdata infrastructuur.

Het operationeel bestuur bestaat uit:

Naam	Organisatie	Mede namens programma
Jan-Willem Boiten (programma manager)	Lygature	TraIT
Jeroen Beliën	VUmc	TraIT
Hans van den Berg	AMC	
Rob Cornelisse	LUMC	
André Dekker	Maastro Clinica	TraIT
Petra Drankier	NFU	
Erik Flikkenschild	LUMC	PSI
Rob Hooft	DTL	
Arnoud van der Maas	Radboudumc	
Irene Nooren	SURFsara	
Bert van Ooijen	Erasmus MC	
Petra van Overveld	DTL	
Harry Pijl	UMCU	
Ronald van Schijndel	VUmc	
Igor Schoonbrood	MUMC+	
Morris Swertz	UMCG	BBMRI-NL
Jan Jurjen Uitterdijk	UMCG	

De programmacommissie bestaat uit:

Naam	Organisatie	Mede namens programma
Frank Miedema (voorzitter)	UMC Utrecht	
Jaap Verweij (voorzitter)	Erasmus MC	
Ameen Abu Hanna	AMC	
Edvard Beem	ZonMw	
Jan-Willem Boiten (programma manager)	Lygature	TraIT
Arjen Brussaard	VUmc	
Petra Drankier	NFU	
Alain van Gool	Radboudumc	
Jan Hazelzet	Erasmus MC	Registratie aan de Bron
Olaf Klungel	Universiteit Utrecht	Mondriaan
Ruben Kok	DTL	
Gabriël Krestin	Erasmus MC	Population Imaging
Karel van Lambalgen	LUMC	Vz AcZIE
Gerrit Meijer	NKI	TraIT, EATRIS
Frits van Merode	MUMC	
Barend Mons	LUMC	ELIXIR
Anwar Osseyran	SURFsara	
Ronald Stolk	UMCG	Mondriaan, PSI
Cisca Wijmenga	UMCG	BBMRI-NL

Informatie

Voor meer informatie over Data4lifesciences zie www.nfu.nl

of neem contact op met Carolien Bouma, Coördinator bij de NFU: bouma@nfu.nl of 030-2739 632.

NFU Nederlandse Federatie van Universitair Medische Centra
Oudlaan 4, 3515 GA / Postbus 9696, 3506 GR Utrecht
T 030 273 98 80 nfu@nfu.nl

Disclaimer © 2016