

Focus meeting Research Workspace

Op 16 december 2016 is vanuit het programma Data4LifeScience (D4LS) een focus meeting 'Research Workspace' georganiseerd met medewerking van SURF en DTL. Het betreffende onderwerp is onderdeel van het werkpakket 4, Biomedical Data Sharing & Analysis in Clinical Studies, binnen het programma D4LS. Gastheren van deze meeting waren dan ook de trekkers van dit werkpakket te weten Arnoud van der Maas en Harry Pijl.

Het programma bevatte in grote lijnen twee topics; welke requirements zijn nodig voor het inrichten van een Research Workspace en welke oplossingen (best practices) zijn er reeds beschikbaar op dit gebied. Het volledige programma en de getoonde presentaties zijn te vinden op <https://www.surf.nl/agenda/2015/12/surf-dtl-focusmeeting/index.html>.

Onderstaand een samenvatting van de presentaties met discussies.

Requirements from UMC research ICT (van der Maas en Pijl)

Korte schets waar de focus van het werkpakket op ligt en wat de huidige stand van zaken is. Bestaande best practices identificeren en gebruiken waarbij de wens om de data zoveel als mogelijk lokaal te houden gerespecteerd moet worden. Het gaat hier vooral over een nieuwe manier van samenwerken waarbij in gezamenlijkheid tools die nodig zijn voor Research gedefinieerd moeten worden. Gebruikers vragen om toegang tot data, compute power en multicenter delen terwijl centra eisen vooral veiligheid. Van belang hierbij is dat de infrastructurele laag niet iedere keer opnieuw uitgevonden wordt.

Discussie

Wordt de data centraal of lokaal vastgelegd? Ideaal beeld is dat alle data zoveel als mogelijk lokaal blijft en alleen bij een uitvraag wordt een kopie vrijgegeven. Deze moet in de centrale omgeving (bewerkt of niet) opgeslagen kunnen worden.

Toevoegen van externe data; rechtstreeks of via lokaal? Beide opties zouden mogelijk moeten zijn.

Er zou een centrale omgeving voor data en tools beschikbaar moeten zijn. Het uitgangspunt voor de research Workspace is dat alles wat potentieel te virtualiseren is in de Workspace te vinden moet zijn. Hierbij zeker ook ruimte bieden aan de wens tot het creëren van een samenwerkingsomgeving voor multicenter onderzoeksprojecten.

Workspace for researchers; requirements and wishes of users (Haitjema)

In deze presentatie is een beeld geschetst waar onderzoekers binnen een UMC behoefte aan hebben op het gebied van Workspaces. Belangrijke eisen/wensen zijn het kunnen werken op een eigen device, gebruik kunnen maken van voldoende geheugen, toegang tot eigen softwarepakketten waarbij ook versies beheerd kunnen worden. Verder moet er een netwerk beschikbaar zijn dat aan al deze wensen invulling kan geven. Het support moet ook meer gericht zijn op de onderzoeker en zijn/haar specifieke vragen. Tot slot; onderzoekers dienen serieus genomen te worden. Er wordt, mede ingegeven door de strenge regelgeving rondom de zorgprocessen, te vaak aangegeven dat zaken die voor de onderzoekers noodzakelijk zijn geen doorgang kunnen vinden. Een andere houding is nodig; een transitie van 'dit kan of mag niet' naar 'we waarderen je werk en regelen je IT support'. Onderzoek is een van de drie kerntaken van een academisch centrum en moet op ICT-gebied ook zo behandeld worden.

Discussie

Belangrijk issue is dat de onderzoeksafdelingshoofden onvoldoende kennis hebben van ICT. Terwijl de ICT-afdeling geen aansluiting heeft met de onderzoeker. Hierdoor ontstaat een vacuüm voor de individuele onderzoeker.

Moet een netwerk uit een open en gesloten netwerk bestaan of is dit juist niet zinvol. De reactie uit de zaal is dat vele UMC's bezig zijn de netwerken zorg en onderzoek uit elkaar te trekken, terwijl Utrecht deze juist bij elkaar aan het brengen is. Radboud geeft aan dat zij juist daarom gekozen heeft voor een Cloud-oplossing. Uit elkaar halen van de netwerken heeft voor de onderzoeker veel voordelen op het gebied van vrijheid en multicenter werkzaamheden. Echter brengt dit wel weer andere problemen met zich mee zoals het integreren van zorg en onderzoek. Iets wat de komende jaren meer en meer gaat spelen. Daarnaast brengt die keuze ook een verhoogde beheerlast met zich mee.

Use Galaxy and Transmart in data management (Fijneman)

In deze presentatie is toegelicht op welke wijze het door CTMM opgerichte project TraIT een bijdrage levert aan het op een generieke wijze inrichten en gebruiken van een research Infrastructuur voor complexe samenwerkingsverbanden. CTMM richt zich vooral op het samenbrengen van publieke en private partijen.

Binnen TraIT is de focus gelegd op het verzamelen van data en integreren over verschillende bronnen heen. Ook hier is het uitgangspunt dat het wiel niet door iedereen opnieuw uitgevonden moet worden. Daarbij wordt er binnen TraIT voornamelijk gewerkt met open source oplossingen. Zo biedt TraIT best practices (tools) lokaal aan. Open Clinica (verzamelen data), Transmart (integreren data) en Galaxy (workflow) zijn hier voorbeelden van. Open source is (qua licenties) in principe gratis, dit geldt ook voor Galaxy wat ook veilig en schaalbaar is. TraIT werkt volgens de zogenaamde FAIR en SMURF principes.

Discussie

Hoe generiek zijn de oplossingen; de aangeboden en gebruikte tools zijn generiek opgebouwd. Wel is het zo dat TraIT erg gericht is op het gebruik van specifieke multicenter projecten. Kan TraIT ook als virtuele omgeving aangeboden worden? Ja, dit is mogelijk.

RUG/UMCG Lifelines research workspace (Stolk en Uitterdijk)

Lifelines is een grootschalig populatie cohort in het Noorden van het land. Binnen dit project is een infrastructuur neergezet waarbij vanuit een veilige gesloten data verzameling omgeving naar een meer open Universiteit Groningen (UG) omgeving gewerkt is waarin geen ID-gegevens meer te vinden zijn. De data blijft vervolgens in de UG-omgeving.

Zowel in het UMC als UG werkt met dezelfde infrastructuur, echter wel met verschillende look and feels zodat het voor de gebruiker herkenbaar is en blijft. Qua inrichting zijn er default Workspaces ontwikkeld met verschillende rollen individueel en voor groepen. De rollen bepalen toegang tot applicaties, data, studies, etc. Infrastructuur is op meerdere OS mogelijk dus zowel op Linux als Windows. Het betreft hier een Infrastructuur As A Service vooral gericht op een hoge mate self-service. Onderzoeker kan zelf via een portal voorzieningen aanvragen. Zelf installeren is echter niet mogelijk. Verschillende workflows zijn uitgewerkt (bv toevoegen nieuwe user).

Discussie

Hoe schaalbaar is het Lifelines model? Er zijn nu de eerste stappen gezet om deze manier van werken voor het hele UMC Groningen beschikbaar te maken. Daaruit moet blijken hoe schaalbaar de oplossing is. De verwachtingen zijn dat dit zeker mogelijk is.

Digital Research Environment for health research (Vreeken en Barnes)

Radboud Medisch Centrum heeft een tender uitgeschreven met als doel het inrichten van een Research IT-oplossing. Uitgangspunt is het ontwikkelen van een generiek platform met een zo open als mogelijke opzet. Samenwerking is van levensbelang. Vancis heeft de aanbesteding gegund gekregen. Hiertoe is een consortium samengesteld van Mgrid, Aridhia, Care IQ, Porta Vita en Xomnia. Er zijn op dit moment drie niveaus gedefinieerd te weten Catalogus, Data Ponds en Workspaces.

Vervolgens is door Barnes een demonstratie gegeven van het Software As A Service platform wat door Aridhia is ontwikkeld en al in Schotland (NHS) operationeel is. In het concept zijn zowel Workspace databases als Workspace File systems ingericht die ook via een virtuele desktop benaderbaar zijn. Alle activiteiten op het platform worden gelogd en zijn daarmee herleidbaar. De Workspace is oneindig schaalbaar en werkt met de meest belangrijke analyse tools van dit moment. Er zijn op dit moment meerdere actieve projecten waarbij vooralsnog de insteek gericht is op kleine onderzoeken/long-tail groepen die snel aan de slag kunnen.

OpenEarth, an example from the environmental sciences (van Koningsveld)

Van Oord is wereldwijd actief als toonaangevende aannemer van bagger-, waterbouwkundige en offshore energieprojecten (olie, gas en wind) met innovatieve oplossingen voor maritieme uitdagingen. Van Oord heeft een infrastructuur ontwikkeld die wat betreft het verwerken van data en het uitvoeren van analyses vergelijkbaar is aan het zorg/onderzoek landschap. De structuur is schaalbaar en zelfs zonder internet werkend, iets wat in de baggerwereld van groot belang is.

De baggerwereld heeft expliciet gekozen om modellen en tools rondom de meest basale parameters en kerngegevens met elkaar (NL/Be) in een omgeving te delen. Standaardisatie is hierbij de key. Het probleem dat indien je de historie van data niet kent je ook geen gebruik gaat maken van de data kan daarmee overwonnen worden. De filosofie is re-use van data; trust versus transparency. Wees zo transparant mogelijk en richt de infrastructuur hierop in om dat te kunnen delen. Ontwikkeld gezamenlijk met de gebruikers community.

Einddiscussie

Vraagstuk over haalbaarheid en wenselijkheid van een Nederlandse totaaloplossing. Streven naar een Nederlandse oplossing is lovenswaardig, maar er zal breder gekeken moeten worden (internationale samenwerkingsverbanden). Daarnaast moeten we ervoor waken dat we wel slagvaardig blijven. Verkenning van het inzetten van private initiatieven zoals het Radboud MC zijn interessant. Standaardisatie bijvoorbeeld bij het gebruik van compute power of netwerken is, vergelijkbaar met Nutsvoorzieningen, dan wel noodzakelijk.

Belangrijkste conclusie is dat iedereen het belang van een gezamenlijke oplossing onderstreept maar dat we vooral de wensen/eisen van de onderzoekers niet uit het oog moeten verliezen en dus kleine stappen moeten zetten. Het blijven delen van waar eenieder in de verschillende huizen al mee bezig is, zal daarbij van groot belang zijn.