

Notitie

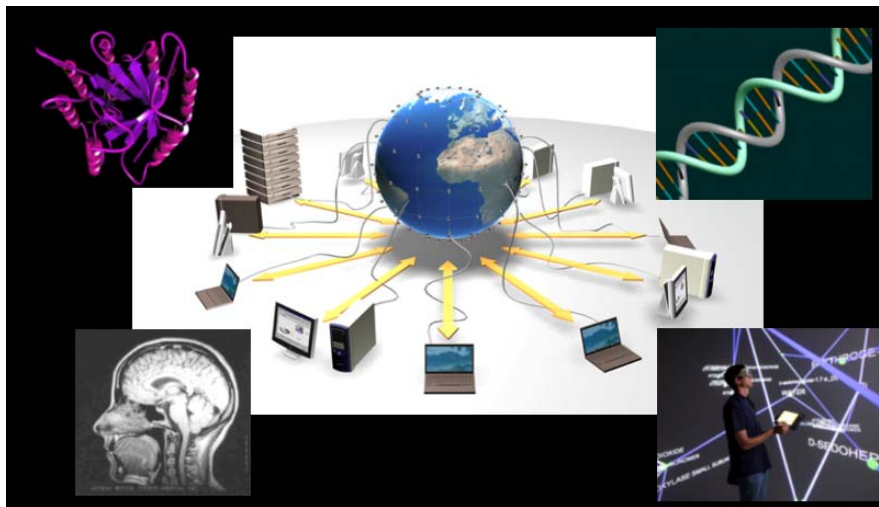
e-Bioscience onderzoeksinfrastructuur in het AMC

Versterking van biomedisch en klinisch wetenschappelijk onderzoek

Antoine van Kampen
Bioinformatica Laboratorium
Klinische Epidemiologie, biostatistiek en bioinformatica
020-5667096 / 06-41768067
a.h.vankampen@amc.uva.nl

13 mei, 2008

Deze notitie is tot stand gekomen in overleg met en bijdragen van Prof. dr. Frank Baas (Neurogenetica), Prof. dr. Damiaan Denys (voorzitter van divisie Psychiatrie), Prof. dr. Kees Grimbergen (Medische Fysica), Prof. dr. Han Laméris (voorzitter van divisie Beeldvormende Specialismen en onderzoek- en behandelcentrum), Dr. Aart Nederveen (Radiologie), Dr. Silvia Olabarriaga (Radiologie/Bioinformatica Laboratorium), Prof. dr. Bert Schadé (voorzitter van divisie Klinische Methoden & Public Health), Prof. dr. Guus Sturk (voorzitter van divisie Laboratoriumspecialismen).



Samenvatting

e-Science (enhanced Science) is een nieuw paradigma in wetenschappelijk onderzoek en creëert effectief een IT-gebaseerde onderzoeksinfrastructuur, die multidisciplinaire en geografisch gedistribueerde wetenschappelijke samenwerkingen faciliteert en versterkt. Op het gebied van de *e-Science* voor medische beeldverwerking heeft het AMC een sterke positie opgebouwd. Deze notitie doet een voorstel om in een *tweede fase* de *e-Science* in het AMC te versterken en te verankeren. Hiertoe wordt een concreet voorstel gedaan voor een organisatiestructuur en voor projecten. Een beperkte eerste geldstroom financiering is gewenst voor de realisatie van deze tweede fase.

1. Inleiding

In moderne wetenschap worden we meer en meer geconfronteerd met complexe vraagstellingen. Voorbeelden hiervan zijn:

- Het begrijpen van de werking van een cel door middel van mathematische modellering (systeembioïologie);
- Ontrafelen van (complexe) ziekten;
- Bestudering van kennis, kennisprocessen en de wijze waarop ons brein kennis representeert en verwerkt (cognitiewetenschap);
- Het analyseren en interpreteren van genomics data;
- Het ontwikkelen van beeldverwerkende methoden voor diagnose, prognose en behandeling.

Door het gebruik van nieuwe experimentele technieken om deze vraagstellingen te beantwoorden, is niet alleen de hoeveelheid wetenschappelijke gegevens explosief toegenomen, maar ook de complexiteit ervan. Die gegevens worden bovendien veelal gegenereerd door verschillende onderzoeksgroepen. Hiernaast zijn methodieken en expertise uit verschillende disciplines noodzakelijk om te komen tot antwoorden op de wetenschappelijke vraagstelling.

e-Science is een nieuw paradigma in wetenschappelijk onderzoek en creëert effectief een IT-gebaseerde onderzoeksinfrastructuur. Deze onderzoeksinfrastructuur faciliteert en versterkt multidisciplinaire en geografisch gedistribueerde wetenschappelijke samenwerkingen verregaand en zorgt er tevens voor dat wetenschappelijke experimenten kunnen worden gecombineerd met computerberekeningen. Om deze onderzoeksinfrastructuur te realiseren is een geïntegreerde benadering van instrumentatie, computerverwerking, opslag en computernetwerken essentieel. Het ontwikkelen van zo'n geïntegreerde infrastructuur is een uitdagend onderzoeksprobleem op zichzelf, dat niet anders dan met multidisciplinair samengestelde onderzoeksteams kan worden aangepakt. Toepassing van *e-Science* in (bio)medisch onderzoek noemen we *e-Bioscience*.

Doel van deze notitie.

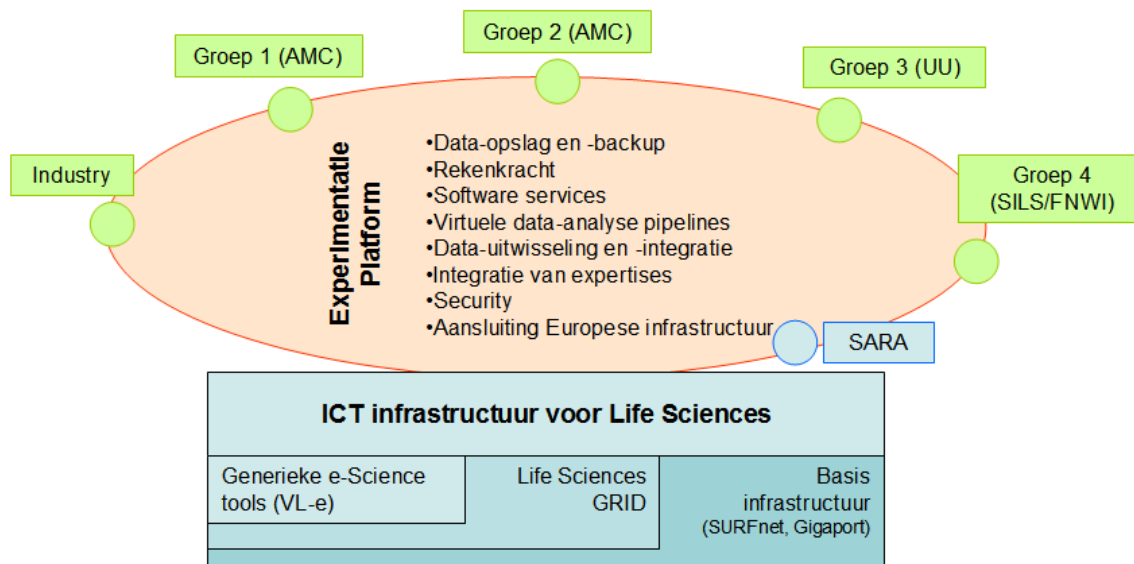
Het AMC heeft internationaal een sterke positie verworven op het gebied van *e-Science* onderzoeksinfrastructuur voor medische beeldvorming. Hierdoor is het AMC in een uitstekende positie om ook voor andere onderzoeksdeelgebieden een *e-Science* onderzoeksinfrastructuur op te zetten om op deze manier dit onderzoek verder te versterken. Dit vereist echter wel dat

- (a) *e-Science* in het AMC op een adequate manier wordt georganiseerd en versterkt;
- (b) huidige activiteiten op het gebied van *e-Science*/beeldverwerking worden verankerd, zodat aanwezige expertise voor het AMC wordt behouden;
- (c) specifieke *e-Science* infrastructuur wordt ontwikkeld voor geselecteerde onderzoeksdomeinen;
- (d) wordt aangesloten bij UvA en (inter)nationale ontwikkelingen op het gebied van *e-Science*.

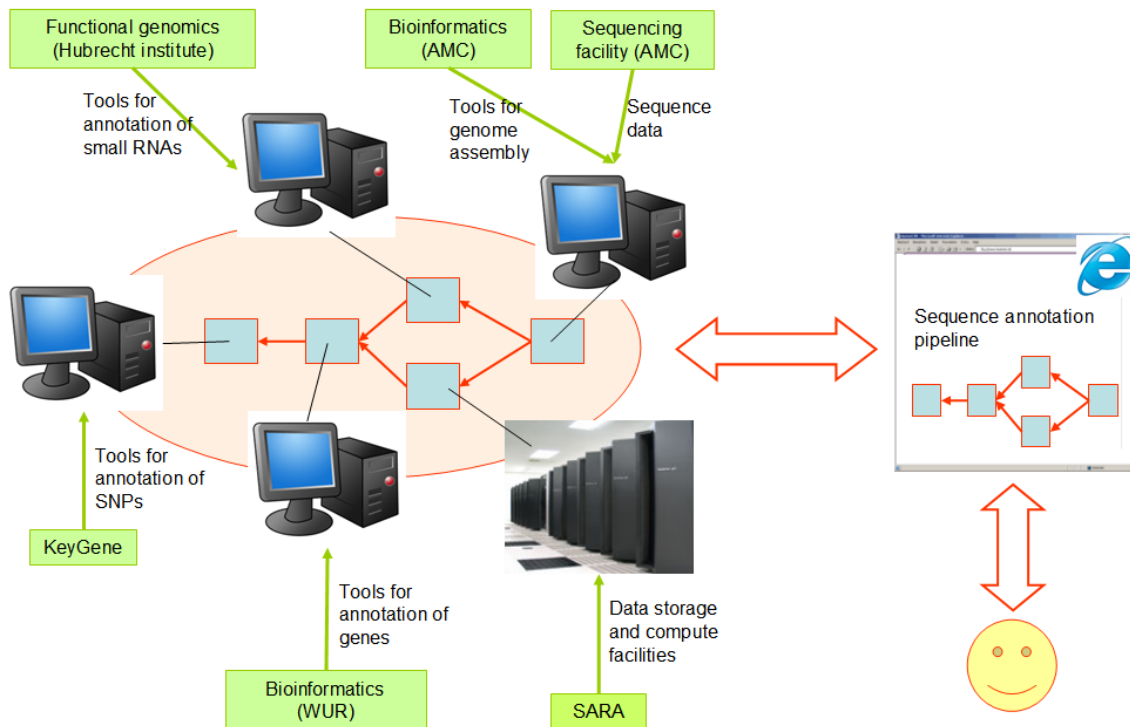
Deze notitie doet hiertoe een voorstel.

2. Wat levert e-Science?

Een e-Science onderzoeksinfrastructuur bestaat uit geografisch gedistribueerde faciliteiten die transparant worden aangeboden aan de eindgebruikers (Figuur 1). Deze faciliteiten kunnen bestaan uit meetapparatuur (MRI, massaspectrometers, sequencers), computerclusters, data-opslag en services (data-analyse software, databases). Deze faciliteiten kunnen middels nieuwe technieken uit de informatica (GRID, web-services, workflow-managementsystemen) worden geïntegreerd, zodat specifieke, maar voorheen lastige taken, in het onderzoek eenvoudig kunnen worden uitgevoerd. Informatiemanagement is een belangrijk onderdeel van de infrastructuur. e-Bioscience realiseert dus *experimentatieplatforms* ter ondersteuning van computationeel en data-intensief biomedisch en klinisch onderzoek. Figuur 2 laat een concreet voorbeeld zien van een experimentatieplatform voor de annotatie van DNA genomesequenties.



Figuur 1. Experimentatieplatform als een gedistribueerde samenwerking tussen groepen/universiteiten en industrie (groene bolletjes). Deze partijen werken gezamenlijk aan een e-Science onderzoeksinfrastructuur ten behoeve van een specifiek wetenschappelijk domein of een specifieke vraagstelling. Het platform levert een selectie van faciliteiten en functionaliteiten. Elke participerende onderzoeksgroep brengt eigen expertise en faciliteiten (software, databases, meetapparatuur, computerclusters, etc.) in en eindgebruikers krijgen toegang tot het platform. De participerende groepen zijn verbonden aan de Nederlands e-Science infrastructuur, die door Big Grid (life sciences GRID) en het Virtual Laboratory for e-Science (VLe; Generic e-Science tools) wordt geïmplementeerd. SARA (blauwe bolletje) onderhoudt deze hardware/e-Science infrastructuur. Figuur 2 geeft een concreet voorbeeld.



Figuur 2. Experimentatieplatform voor de annotatie van DNA genomsequenties afkomstig van de AMC sequencing faciliteit. SARA levert de faciliteiten voor data-opslag en rekenkracht ten behoeve van 'genome assembly'. De participerende onderzoeksgroepen zijn via een lokaal computercluster verbonden aan het life sciences GRID en stellen tools voor de annotatie van de genomsequenties beschikbaar via dit lokale cluster. Deze tools worden geïntegreerd tot een virtuele annotatie-pipeline. De eindgebruiker krijgt toegang tot de annotatie-pipeline via zijn web-browser en kan hiermee zijn eigen genomsequenties assembleren en annoteren. De toegang is transparant; de eindgebruiker hoeft zelf geen tools te integreren, ziet niet dat de gebruikte annotatie-tools door verschillende groepen worden ingebracht en heeft direct beschikking over voldoende rekenkracht en mogelijkheden voor data-opslag. Deze manier van aanpak leidt tot gestandaardiseerde data-verwerkingsprotocollen.

3. e-Science in ontwikkeling

e-Science is zelf ook onderwerp van onderzoek (zoals bijvoorbeeld uitgevoerd in het Virtual Laboratory for e-Science (VL-e) en de NWO STAR E-Science en Jacquard programma's). Dit onderzoek richt zich zowel op de verdere ontwikkeling van de methodologie en infrastructuur als op de toepassing hiervan. e-Science toepassingen worden ontwikkeld voor specifieke onderzoeksdomeinen en -vragen. e-Science levert dus geen kant-en-klare oplossingen.

Nationale ontwikkelingen

Het VLe programma (directeur Prof. dr. Hertzberger), het Nederlandse Bioinformatica Centrum (NBIC; wetenschappelijk directeur Prof. dr. Van Kampen), Big Grid (directeur Prof. dr. Van den Herik) en NWO-NCF dragen zorg voor een belangrijk deel van de financiering en implementatie van het e-Bioscience onderzoek, infrastructuur en toepassing. Op het gebied van de life sciences (inclusief bioinformatica) participeren verschillende UvA-groepen in deze programma's. SARA speelt een belangrijke rol in de implementatie, disseminatie en het beheer van het Nederlandse Life Science grid-netwerk en de generieke e-Science-infrastructuur. Hiernaast heeft SARA een rol in de training van wetenschappers voor het gebruik van deze infrastructuur. Door deze programma's en initiatieven heeft de Faculteit der Natuurwetenschappen, Wiskunde en Informatica (FNWI) van de UvA een internationaal leidende positie verworven. Momenteel worden door de FNWI de eerste stappen gezet om een nationaal e-Science centrum op te zetten, dat moet zorgdragen voor verdere ontwikkeling van e-Science, voor het implementeren van toepassingen en voor het leveren van landelijke ondersteuning.

Internationale ontwikkelingen

e-Bioscience onderzoeksinfrastructuur en toepassingen nemen wereldwijd snel toe. Voorbeelden zijn het Cancer Biomedical Informatics GRID (NCI; cagrid.nci.nih.gov), het Biomedical Informatics Research Network (BIRN; www.nbirn.net), OMII-UK (o.a. bioinformatica; www.omii.ac.uk) en de biomedische activiteiten van het Enabling Grids for E-scienceE programma (EGEE; egee-na4.ct.infn.it/biomed/applications.html). In Groot-Brittannië is een nationaal e-Science instituut opgezet om de nationale projecten te ondersteunen (www.nesc.ac.uk).

4. e-Bioscience in het AMC

e-Bioscience speelt in toenemende mate een rol in het AMC onderzoek (Figuur 3). Inmiddels is het AMC aangesloten op het Nederlandse Life Sciences-GRID. Een internationaal leidende rol is door Radiologie verworven met de e-Bioscience infrastructuur voor biomedische beeldverwerking (projectleider Dr. Olabarriga). In de context van informatiemanagement wordt gewerkt aan nieuwe manieren om biomedische- en klinische-informatie te organiseren in kennisdatabases. In een samenwerking met Genetisch Metabole Ziekten is een start gemaakt met de implementatie van een peroxisoom database (projectleider Prof. dr. Antoine van Kampen). Deze kennisdatabase zal op korte termijn ook worden ingezet voor het beschrijven van biologische systemen ten behoeve van systeembioïologie (projectleider Dr. Van Beek, VU). Samen met de sequencing faciliteit is een start gemaakt met het opzetten van een sequencing platform voor bioinformatica-ondersteuning (projectleider Prof. dr. Van Kampen). Op korte termijn zal samen met Medische Biochemie een start worden gemaakt met het opzetten van een proteomicsplatform voor bioinformatica-ondersteuning (projectleider Prof. dr. Van Kampen). Hierbij zal worden aangesloten bij de proteomics en sequencing e-Bioscience BioAssist projecten van het NBIC. Deze projecten organiseren en leveren bioinformatica-ondersteuning zoals geïllustreerd in Figuur 2. In Appendix 1 worden deze projecten in meer detail beschreven en in hoofdstuk 5.2 wordt de financiering van deze projecten uiteengezet. In Appendix 2 wordt een aantal mogelijke toekomstige projecten kort beschreven.

Domain	Domain researcher	e-Bioscience researcher (core group)	IT Collaborator
Biomedical imaging (started)	AMC Radiology •Dr. Aart Nederveen •Sanna Gevers, MD (AiO) •Matthan Caan (AiO) •Remi Soleman	AMC Bioinformatics Laboratory •Dr. Silvia Olabarriga (postdoc) •Dr. Tristan Glatard (postdoc) •Kamel Boulebiar (sci. programmer)	Informatics UvA NIKHEF SARA Univ. Cardiff, UK CNRS, France
Peroxisome Knowledge Base (started)	AMC Genetic Metabolic Disease •Prof. dr. Ronald Wanders Bochum University, Germany •Prof. dr. Ralf Erdmann AMC Bioinformatics Lab. •Dr. Gerbert Jansen (post-doc)	AMC Bioinformatics Laboratory •Prof. dr. Antoine van Kampen •Dr. Aviral Vatsa (post-doc) •Marcel Willemsen (AiO) SILS UvA •Serge Barth (sci. programmer) •Position (post-doc)	Informatics UvA SARA
High-throughput Sequencing (started)	AMC Genomics Facility •Prof. dr. Frank Baas	AMC Bioinformatics Laboratory •Prof. dr. Antoine van Kampen •Barbera Danz-van Schaik (junior researcher) •Angela Luyf (sci. programmer)	Informatics UvA SARA
Proteomics (start in near future)	AMC Medical Biochemistry •Prof. dr. Chris de Koster AMC Biostatistics •Prof. dr. Koos Zwinderman AMC Bioinformatics Lab. •Dr. Perry Moerland SILS •Prof. dr. Chris de Koster •Prof. dr. Age Smilde AMOLF •Prof. dr. Ron Heeren	AMC Bioinformatics Laboratory •Prof. dr. Antoine van Kampen •Position (sci. programmer)	Informatics UvA SARA

Figuur 3. Vier concrete e-Bioscience projecten in het AMC. In de kolommen staan de direkt betrokkenen vermeld. De 'e-Bioscience researchers' hebben een belangrijke rol als intermediair tussen de domeinexperts en de informaticaonderzoekers en vormen samen de e-Bioscience kerngroep (zie hoofdstuk 5). De projectleiders zijn onderstreept. Luyf is tevens verantwoordelijk voor applicatiebeheer op het AMC computercluster.

De (toekomstige) toegevoegde waarde van e-Bioscience voor het AMC wordt samengevat door de volgende punten:

- Het faciliteren van gedistribueerde wetenschappelijke (UvA) samenwerkingen die sneller leiden tot wetenschappelijk resultaat door het gemeenschappelijk ontwikkelen van experimentatieplatforms;
- Door integratie van expertise komen complexere wetenschappelijke vraagstellingen binnen het bereik van onderzoek;
- Vermijden van redundantie en lagere kosten door delen en hergebruik van services en faciliteiten;
- Standaardisatie in onderzoek door o.a. standaard analyse-protocollen en gebruik van *de facto* standaarden uit (bio)informatica;
- Het creëren van een krachtige e-Science onderzoeksinfrastructuur met transparante toegang tot lokale en externe onderzoeksfaciliteiten;
- Aansluiting op Nederlandse (Life Sciences Grid) en Europese (EGEE) GRID netwerken om daarmee toegang te krijgen tot elders aangeboden faciliteiten en services;
- Aansluiting bij grootschalige Europese onderzoeksinfrastructuren (ESFRI) voor o.a. biobanking en bioinformatica.

5. Organisatie en inbedding van e-Bioscience in AMC onderzoek

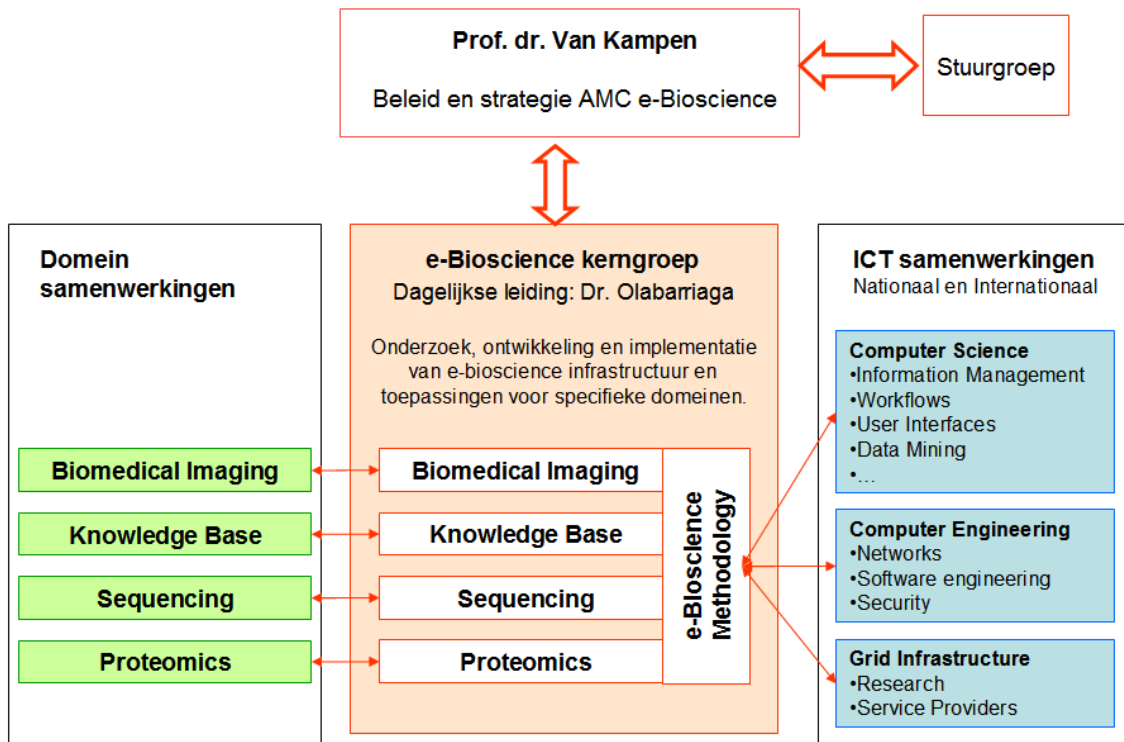
De ontwikkeling en toepassing van e-Bioscience onderzoeksinfrastructuur voor AMC onderzoek is geen sinecure en vereist dat e-Bioscience in het AMC op een adequate manier wordt georganiseerd en verankerd om:

- De benodigde e-Bioscience- en technologie-expertise verder te kunnen ontwikkelen;
- e-Bioscience beschikbaar te maken voor verschillende toepassingen binnen biomedisch en klinisch onderzoek in het AMC;
- Voldoende kritische massa en interactie tussen AMC domeinexperts, e-Bioscience onderzoekers en informatica onderzoekers te bewerkstelligen;
- Aansluiting te krijgen bij FNWI e-(bio)science activiteiten en het beoogde nationale e-Science instituut;
- Aansluiting te vinden bij (inter)nationale e-Bioscience programma's;
- Resultaten te realiseren in de vorm van infrastructuur, software, wetenschappelijke publicaties;
- Verwerven van additionele subsidies voor verdere ontwikkeling van e-Bioscience infrastructuur;
- Aantrekken van toponderzoekers;
- Onderwijs en training.

5.1 Organisatiestructuur e-Bioscience

Figuur 4 geeft de beoogde organisatiestructuur van e-Bioscience in het AMC; deze is inmiddels voor een groot deel gerealiseerd. Recent zijn de e-Bioscience onderzoekers voor medische beeldverwerking ondergebracht bij het Bioinformatica Laboratorium. Hiermee is *de facto* een e-Bioscience-kerngroep (Figuur 4) binnen het Bioinformatica Laboratorium gecreëerd. Deze kerngroep krijgt de volgende taken:

- Het opzetten en onderhouden van nauwe samenwerking met AMC domeinonderzoekers (van o.a. radiologie, sequencingfaciliteit, proteomicsfaciliteit) om wensen te identificeren en ten behoeve van onderzoek nieuwe applicaties te ontwikkelen en te evalueren;
- Toepassen, ontwikkelen, evalueren, en uitrollen van innovatieve informaticatechnieken die geavanceerd biomedisch- en klinischonderzoek mogelijk maakt of verder versterkt;
- Het opzetten en onderhouden van nauwe samenwerkingen met informatica onderzoekers (o.a. Instituut voor Informatica/UvA) en service providers (o.a. SARA, NIKHEF).



Figuur 4. Organisatie van e-Bioscience in het AMC. De e-Bioscience-kerngroep zorgt voor de ontwikkeling en toepassing van e-Bioscience voor verschillende AMC onderzoeksdomeinen. De kerngroep heeft nauwe technologische en inhoudelijke samenwerkingen met (inter)nationale groepen en initiatieven. De e-Bioscience kerngroep staat onder dagelijkse leiding van Olabbarriaga. Van Kampen is verantwoordelijk voor het algemene AMC e-Bioscience-beleid en -strategie en verantwoordelijk voor aansluiting bij (inter)nationale initiatieven zoals NBIC, Big Grid, VLe, NGI consortia, EGEE, ESFRI en het beoogde nationale e-Science instituut. De stuurgroep draagt mede zorg voor aansluiting bij AMC-strategie en -beleid.

Dr. Olabbarriaga krijgt de dagelijkse leiding over de e-Bioscience kerngroep om coherentie tussen de domeinen en projecten te waarborgen, zorg te dragen voor verdere ontwikkeling van de benodigde informatica- en e-Bioscience-expertise, bij te dragen aan de AMC e-Science onderzoeksinfrastructuur en (inter)nationale samenwerkingen te initiëren. Hiernaast blijft Dr. Olabbarriaga vanuit haar achtergrond en expertise de wetenschappelijke trekker van biomedische beeldverwerking.

Prof. dr. Van Kampen (Bioinformatica Laboratorium, KEBB) is verantwoordelijk voor ontwikkeling en toepassing van e-Bioscience binnen AMC onderzoek. Via zijn aanstelling bij de FNWI (SILS) kan eenvoudig aansluiting worden verkregen bij FNWI e-(bio)science activiteiten die voor het AMC van belang zijn. Vanuit zijn positie als wetenschappelijk directeur van NBIC is hij betrokken bij de e-Bioscience bioinformaticaprojecten van deze organisatie, waardoor het AMC eenvoudig aansluiting kan vinden bij de BioAssist experimentatieplatformen. Hiernaast is Van Kampen betrokken bij Big Grid, VLe, en het beoogde nationale e-Science centrum.

De stuurgroep voert overleg met Prof. Van Kampen en Dr. Olabbarriaga over o.a. selectie en voortgang van projecten en aansluiting bij AMC-strategie en -beleid, zoals deze wordt vormgegeven door bijvoorbeeld de Research Council en de ICT raad (voor onderzoek).

5.2 Huidige capaciteit en financiering

In Tabel 1 is een overzicht gegeven van de huidige capaciteit en financiering van de AMC e-Bioscience projecten zoals weergegeven in Figuur 3.

	Naam	Functie	Financiering	einddatum
Coördinator e-Bioscience AMC	Dr Van Kampen	Groepsleider Bioinformatica Laboratorium	1 ^{ste} geldstroom	
Medical Imaging	Dr Olabarriaga	Post-doc ^{1,2}	BSIK VLe	31-12-2009
	Dr Glatard	Post-doc ¹	BSIK VLe	31-12-2009
	Boulebiar (MSc)	Wetensch. Programmeur ¹	BSIK VLe	31-12-2009
Kennisdatabases	Willemsen (MSc)	OiO	1 ^{ste} geldstroom	
	Dr Vatsa	Post-doc	BioRange/EU-Peroxisome	31-2-2009
	Dr Jansen	Post-doc	EU-Peroxisome	31-12-2008
Sequencing platform	Danz - van Schaik	Junior onderzoeker	1 ^{ste} geldstroom	
	Luyf		1 ^{ste} geldstroom	
Proteomics platform	Position ³	Wetensch. programmeur	NGI NPC-NBIC	3 jaar

Tabel 1. Huidige capaciteit voor AMC e-Bioscience projecten. Deze personen zijn ondergebracht bij het Bioinformatica Laboratorium (KEBB). ¹Recent ondergebracht bij het Bioinformatica Laboratorium. ²Het Bioinformatica Laboratorium heeft de intentie om Dr. Olabarriaga een vaste positie als UD te geven. ³Voor deze positie wordt hoogstwaarschijnlijk op korte termijn financiering verkregen van NPC-NBIC.

De formatie voor de kennisdatabases is op dit moment toereikend en de verwachting is dat de continuïteit van dit project kan worden gewaarborgd vanuit de eerste geldstroom en door nieuwe externe subsidies. De formatie voor het sequencingplatform is voldoende gezien het feit dat hier wordt aangesloten bij het NBIC BioAssist platform. De continuïteit van dit platform is door eerste geldstroom financiering gewaarborgd. De formatie voor het proteomicsplatform is minimaal, maar een start kan worden gemaakt door aansluiting op het BioAssist proteomicsplatform.

5.3 Knelpunten AMC e-Bioscience

De situatie voor de biomedische beeldverwerking is precair, omdat dit project volledig uit de derde geldstroom wordt gefinancierd (VLe, t/m 2009) en continuïteit is niet gewaarborgd. Om deze activiteiten te verankeren, is het wenselijk om de financiering van het huidige team (Olabarriaga, Glatard, Boulebiar) voor biomedische beeldverwerking te continueren vanuit de eerste geldstroom. Het Bioinformatica Laboratorium heeft de intentie om vanuit haar huidige exploitatiebudget Dr. Olabarriaga een vaste aanstelling als UD te geven. Additionele financiële middelen uit de eerste geldstroom moeten gevonden worden om na 2009 de posities voor de postdoc en wetenschappelijk programmeur te waarborgen. Het is evenwel van belang om hierover op korte termijn uitsluitel te krijgen om het bestaande personeel (en expertise) voor het AMC te behouden, door uitzicht te bieden op een vaste aanstelling. Deze constructie waarborgt tevens, dat opgebouwde e-Bioscience expertise via de kerngroep kan worden ingezet voor verschillende AMC projecten.

6. Fase 2 voor e-Bioscience in het AMC

Om de (inter)nationale positie van het AMC op het gebied van e-Bioscience te versterken en om zorg te dragen dat expertise en e-Bioscience onderzoeksinfrastructuur breder beschikbaar komt, zullen de volgende stappen worden genomen:

- a) **e-Bioscience in het AMC wordt op een adequate manier georganiseerd (Figuur 4).**
Hiertoe accorderen het AMC en betrokken divisies de organisatieopzet. Hiernaast zal definitief vorm worden gegeven aan de kerngroep en zal invulling worden gegeven aan de stuurgroep.
- b) **Huidige e-Science/beeldverwerking activiteiten worden verankerd, zodat aanwezige expertise voor het AMC wordt behouden.**
Het Bioinformatica Laboratorium heeft de intentie om Dr. Olabbarriaga een vaste aanstelling als UD te geven. Hiernaast is het gewenst om eerste geldstroom financiering te vinden voor de huidige postdoc en wetenschappelijk programmeur (beide voor medische beeldverwerking).
- c) **Specifieke e-Bioscience infrastructuur wordt ontwikkeld voor geselecteerde onderzoeksdomeinen.**
e-Bioscience infrastructuur wordt verder ontwikkeld voor de projecten die zijn aangeven in Figuur 3. Hiertoe worden concrete en duidelijke wetenschappelijke vraagstellingen gedefinieerd. In overleg met de stuurgroep wordt bepaald of verdere uitbreiding hiervan gewenst is (Appendix 2).
- d) **Het AMC sluit verder aan bij (inter)nationale ontwikkelingen op het gebied van e-Bioscience.**
De huidige (inter)nationale contacten en samenwerkingen zullen in het kader van concrete projecten verder worden uitgebreid. Er zal worden aangesloten bij nationale e-Science initiatieven.

Appendix 1. AMC e-Bioscience projecten

Biomedische beeldverwerking

Het "Virtual Lab for functional MRI" (VL-fMRI; <http://www.science.uva.nl/~silvia/vlfmri/>) is een project dat in 2004 is gestart door Dr. Olabbarriaga en omvat onderzoek en ontwikkeling van geavanceerde IT infrastructuur voor "enhanced" medische beeldanalyse. Prof. dr. Ard den Heeten en Prof. dr. ir. Kees Grimbergen hebben het VL-e Medical subprogramma geïnitieerd in het AMC. Een belangrijk doel van dit programma is het beschikbaar maken van e-Science infrastructuur voor medische beeldanalyse in onderzoek door AMC radiologen, neurowetenschappers en fysici. Hiervoor is de nauwe samenwerking tussen de eindgebruikers (AMC), de ontwikkelaars van de technologie (Instituut voor Informatica UvA) en grid service providers (SARA, NIKHEF) essentieel. Recente nationale en internationale samenwerkingen voor de uitwisseling van software en expertise, met o.a. VUmc, TUDelft, CNRS Sophia Antipolis, FR; Computer Science, Cardiff University, UK; Computer Science, University of Manchester, UK, hebben de ontwikkelingen in het VL-fMRI veel versneld. Ook verschillende projecten op het gebied van neuroimaging met CT en MRI (f-MRI en Diffusion Tensor Imaging) hebben bijgedragen aan de resultaten. Zie Appendix 3 voor een selectie van publicaties.

De ontwikkelde infrastructuur is gebaseerd op GRID-technologie en wordt ingezet voor de opslag van het grote volume (beeld) data, voor toegang tot voldoende rekenkracht ten behoeve van beeldanalyse, voor het beheer van complexe en grote experimenten en het faciliteren van data-uitwisseling en samenwerking tussen onderzoekers. Deze infrastructuur wordt gebruikt door de gebruikers van de 3.0 Tesla MRI faciliteit in het AMC, zowel voor neurowetenschap als voor klinisch onderzoek. De infrastructuur zelf is generiek en kan in de toekomst als basis dienen voor andere toepassingen zoals microscopie- en cognitie-onderzoek.

Een deel van het VL-fMRI onderzoek richt zich op het ontwerpen, ontwikkelen en evalueren van e-Science infrastructuur voor (bio)medische imaging. Dit onderzoek moet meer inzicht geven in welke componenten van deze IT-infrastructuur belangrijk zijn voor het beeldanalyseplatform en hoe dit platform optimaal kan worden ingezet en hoe de werkomgeving van klinisch onderzoekers kan aansluiten bij (inter)nationale infrastructuren. Hiernaast wordt specifiek gekeken naar toepasbaarheid, gebruikersvriendelijkheid en beveiliging van de e-Science infrastructuur en hoe deze kan worden ingezet ten behoeve van complexe toepassingen. Het ander deel van het onderzoek richt zich op de toepassing zelf, met speciale aandacht voor fMRI. Het doel is vooral de evaluatie van de toegevoegde waarde van de e-Science infrastructuur voor de eindgebruiker van MRI beeldanalyse voor klinisch onderzoek, patiëntenzorg en neurowetenschappen. Vooral wordt gekeken naar de infrastructuur als faciliterende factor voor complexere studies (meer data en/of complexere analyse pipeline), voor collaborative en multicenter onderzoek en voor disseminatie van geavanceerde (nieuwe) toepassingen in de kliniek.

Informatiemanagement: peroxisoom kennisdatabase.

Informatiemanagement is een essentieel onderdeel van e-Bioscience. Met name het ontwikkelen van methodologiën die het mogelijk maken om domeinkennis en data uit verschillende bronnen (semantisch) te integreren is van groot belang binnen gedistribueerde samenwerkingen zoals die in e-Bioscience projecten plaatsvinden. Het kennisdatabase onderzoek van het Bioinformatica Laboratorium (projectleider Prof. dr. Antoine van Kampen) richt zich op de ontwikkeling van een generiek raamwerk voor het opzetten van kennisdatabases voor biomedische systemen, zoals biologische netwerken en gerelateerde ziektes. Specifiek wordt gekeken naar de representatie van kennis door middel van ontologieën en Web Ontology Language (OWL). Hiernaast worden

bijbehorende methodieken ontwikkelt voor (grafische) presentatie van domeinkennis, integratie van domeinkennis met genomics en klinische data, navigatie, en de analyse/interpretatie van genomics data in de context van biologische netwerken.

In samenwerking met Prof. dr. Ronald Wanders (Genetisch Metabole Ziekten) wordt gewerkt aan de ontwikkeling van een geïntegreerde peroxisoomkennisdatabase die in detail het peroxisomale metabolisme en gerelateerde ziektes beschrijft. In samenwerking met Prof. dr. Ralf Erdman (Bochum Universiteit) wordt de kennisdatabase uitgebreid met informatie over peroxisoombiogenese. Samenwerkingen met Dr. Scott Marshall (sementatische netwerken; Instituut voor Informatica, UvA), Prof. Age Smilde (grijze statistische modellen; SILS), en Dr. Robert Stevens (ontologieën; University Manchester) worden momenteel besproken.

- A.M. Willemsen, G.A. Jansen, J.C. Komen, S. van Hooff, H.R. Waterham, P.M.T. Brites, R.J.A. Wanders and A.H.C. van Kampen (2008) Organization and integration of biomedical knowledge with concept maps for key peroxisomal pathways. *Bioinformatics*. Accepted for publication.

NBIC BioAssist Sequencing platform

Recently so-called “next generation sequencing” (NGS) machines have been developed that can produce raw sequence data into the giga-base per day range. These include the 454 GS FLX (Roche Diagnostics), the 1G genome analyzer (Illumina) and the SOLiD sequencer (Applied Biosystems). Several Dutch research institutes and companies have acquired one or more NGS machines and many more will follow. The scale of data generation, the nature of the sequences produced (short reads), and the variety of applications that the platforms facilitate, all pose challenges for data management, bioinformatics, and statistical analyses. High-throughput sequencing is no longer the exclusive domain of large genome centers with a dedicated IT- and bioinformatics infrastructures. However, smaller NGS laboratories face the same bioinformatics challenges without having similarly advanced, in-house computational infrastructures. A NBIC BioAssist experimentation platform will be implemented for the data management, analysis, and visualization of NGS results (project leader Dr. Roeland van Ham (WUR)).

NGS requires implementation of a computational infrastructure that scales to the data volumes produced. Data processing and analysis pipelines must be developed that are robust, flexible, extendable in design, and re-useable. The platform will focus on:

- Data management, data storage, and access to high performance computing infrastructure;
- Data standardization;
- Pre-processing (quality processing and quality control);
- Data assembly and annotation.

Multiple sequencing applications will be supported, including reference genome mapping, de novo assembly, mixed technology assembly, repeat resolution, paired-end sequencing, tag-based approaches, automated small RNA discovery and profiling, ChIP/Seq, digital gene expression analysis, analysis of sequence and structural polymorphisms, meta-genomics, and comparative genome analysis.

The genomics facility of the AMC recently acquired the 454 GS FLX. The AMC genomics facility and the Bioinformatics Laboratory will both participate in the definition, development and use of this BioAssist sequencing platform.

NBIC BioAssist proteomics platform

Proteomics data are inherently heterogeneous due to large variation in instrumentation and experimental design, and much effort will be needed to solve this problem. This BioAssist proteomics platform aims at the development of a pipeline for MS-based proteomics data for (pre)processing, analysis and storage (project leader Bas van Breukelen, UU).

This experimentation platform will be based on generic data standards which are currently developed at the HUPO-PSI and the European Bioinformatics Institute (EBI). A start was made to implement algorithms for data (pre)processing and 2D-LC-MSMS time-alignment algorithms for label free quantitative proteomics (RUG). Further functionalities will include:

- Database. Proteomics data will be stored in PRIDE (developed at the EBI). It is planned to extend the PRIDE database to handle the more specific needs of our project partners, in collaboration with the MOLGENIS team at the RUG;
- Data analysis. Labelled and label-free quantitative (2D-LC-MS(MS)) and peptide/protein quantification); MS data filtering;
- Identification. Peptide search algorithms and de novo peptide identification.
- Statistics. Biomarker discovery and the evaluation of (MS) data and time-alignment analysis.

The AMC departments Medical Biochemistry (proteomics facility) and the Bioinformatics Laboratory will both participate in the definition, development and use of this BioAssist proteomics platform.

Appendix 2. Mogelijke toekomstige e-Bioscience projecten

- *Cognitie*. Op dit moment zijn er plannen om binnen het Spinoza Centre for Imaging een nauwere samenwerking te bewerkstelligen tussen het Nederlands Instituut voor Neurowetenschappen (NIN; KNAW), de afdeling psychologie van de UvA en het AMC geneeskunde. Functionele MRI zal in combinatie met andere modaliteiten in dit centrum worden gebruikt voor wetenschappelijk onderzoek. Het VLe-fMRI platform dat binnen het AMC is ontwikkeld, kan verder worden uitgebreid en afgestemd met cognitieonderzoek om hieraan een bijdrage te leveren.
- *Biobanking (Parelsnoer)*. e-Bioscience infrastructuur biedt mogelijkheden voor informatiemanagement (o.a. data-integratie, -organisatie en -annotatie) ten behoeve van biobanking. Het Nederlandse life sciences GRID kan de koppeling van patiëntcohort databases van verschillende academische ziekenhuizen faciliteren. NBIC en Big Grid investeren in een biobanking platform (projectleider Prof. dr. GJ van Ommen, LUMC). Hiernaast is overleg vanuit NBIC (van Kampen) en Big Grid met Parelsnoer (M. Ros en Prof. dr. D. Legemate) om mogelijkheden tot samenwerking te inventariseren.
- *Systeembioologie*. Als onderdeel van het Netherlands Genomics Initiative (NGI) systeembioologie consortium (NCSB; directeur Prof. dr. R. Van Driel, UvA) zal een platform worden opgezet voor de integratie van data- en computationele modellen. Dit platform kan ook voor het AMC benodigde faciliteiten voor systeembioologie leveren. Prof. dr. A. van Kampen participeert in dit platform vanuit zijn positie bij het Swammerdam Institute for Life Sciences (SILS).
- *Microscopie*. Als onderdeel van Biomedical imaging zal ook microscopie een rol kunnen gaan spelen. NBIC zet samen met het NKI en het VLe-medical subprogramma een platform op voor de management en verwerking van microscopie-data. Dit platform is potentieel interessant voor het Centrum voor Microscopisch Onderzoek (CMO; Dr. J. Van Marle, Dr. J. Aten; AMC) en Centrum voor Advanced Microscopy (Prof. dr. D. Gadella, Moleculaire Cytologie, UvA) en kan verdere samenwerking tussen deze centra faciliteren.

Appendix 3. Selected Publications of VL-e Medical

Journals

S.D.Olabarriaga, J.G. Snel, C.P. Botha, R.G. Belleman, "Integrated Support for Medical Image Analysis Methods: from Development to Clinical Application", *IEEE Transactions on Information Technology in Biomedicine*, Volume 11, Issue 1, Jan. 2007 Page(s):47 - 57

M.L. de Win, L. Reneman, G. Jager, S.D. Olabarriaga, C. Lavini, I. Bisschops, C.B.L.M. Majoie, J. Booi, G.J. den Heeten and W. van den Brink, "Converging evidence for sustained neurotoxic effects of ecstasy on the thalamus", *British Journal of Psychiatry*, accepted for publication.

M.L. de Win, L. Reneman, G. Jager, E.-J. P. Vlieger, S.D. Olabarriaga, C. Lavini, I. Bisschops, C.B.L.M. Majoie, J. Booi, G.J. den Heeten and W. van den Brink, "A Prospective Cohort Study on Sustained Effects of Low-Dose Ecstasy Use on the Brain in New Ecstasy Users", *Neuropsychopharmacology*, 2007, v. 32, p.458-470

Conferences

T. Glatard, K. Boulebiar and S. D. Olabarriaga. "Workflow integration in VL-e medical", *21st IEEE International Symposium on Computer-Based Medical Systems*, Finland, June 2008

S.D. Olabarriaga, T. Glatard, K. Boulebiar, and P. de Boer. "From 'low-hanging' to 'user-ready': initial steps into a healthgrid", *Healthgrid*, Chicago, June 2008

G. van Noordende, M. Koot, S.D. Olabarriaga. "Privacy and Trust for Grid-based Medical Applications", *First Workshop on Security, Trust and Privacy in Grid Environments*, 8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'08), France, June, 2008.

S. Caton, M. Caan, S.D. Olabarriaga, O. Rana and B. Batchelor. "Using Dynamic Condor-based Services for Classifying Schizophrenia in Diffusion Tensor Images" *8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'08)*, France, June, 2008.

K. C.Maheshwari, S. D. Olabarriaga, C. P. Botha, J. G Snel, J. Alkemade and A. Belloum. "Problem Solving Environment for Medical Image Analysis" *IEEE Intl. Symp. on Computer-Based Medical Systems (CBMS 2007)*, Maribor, Slovenia, June 2007.

S.D. Olabarriaga, A.J. Nederveen, Breannan O' Nuallain, "Parameter Sweeps for Functional MRI Research in the Virtual Laboratory for e-Science Project", *Fifth International Workshop on Biomedical Computations on the Grid (BioGrid'07)*, as part of the 7th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'07), Rio de Janeiro, May 2007, pp. 685-690

S.D. Olabarriaga, P.T. de Boer, K. Maheshwari, A. Belloum, J.G. Snel, A.J. Nederveen, M. Bouwhuis, "Virtual Lab for fMRI: Bridging the Usability Gap", *2nd IEEE Conference on e-Science and Grid Computing*, Amsterdam, December 2006.

S.D.Olabarriaga, A. Nederveen, J.G. Snel, R.G. Belleman, "Towards a Virtual Laboratory for fMRI Data Management and Analysis", *HealthGrid 2006*, Valencia, June 2006.

J. G. Snel, S. D. Olabarriaga, J. Alkemade, H. G. van Andel, A. J. Nederveen, C. B. Majoie, G. J. den Heeten, M. van Straten, R. G. Belleman, "A Distributed Workflow Management System for Automated Medical Image Analysis and Logistics", *IEEE CMBS, special track on Grids for Biomedical Informatics*, Salt Lake City, June 2006.