

DE TOEKOMST VAN DE BIOINFORMATICA IN NEDERLAND

Position paper

NWO/KNAW commissie voor de positiebepaling van de bioinformatica in Nederland

9 juli 2002

INHOUD

1. Inleiding	3
2. Samenvatting en aanbevelingen	4
3. Internationale context	5
4. Integratie van bioinformatica in andere onderzoeksgebieden	5
5. Infrastructuur	5
6. Centrale faciliteit	5
7. Onderwijs	7
8. Bioinformatici uit het buitenland	8
Samenstelling commissie	Appendix 1
Verslag gesprek Janet Thornton (Directeur EBI)	Appendix 2
Uitwerking services Bio-ASP	Appendix 3
Verslag workshop "The future of bioinformatics in The Netherlands"	Appendix 4

1. Inleiding

Het moderne onderzoek in de levenswetenschappen kenmerkt zich door een toenemende omvang en complexiteit van de datasets waarmee gewerkt wordt. Onder meer door de vlucht die het genomics onderzoek de laatste jaren heeft genomen zijn (en worden) er enorme hoeveelheden gegevens gegenereerd. De noodzaak om deze data niet alleen efficiënt te kunnen verwerken, maar vooral om er biologisch relevante informatie uit te kunnen destilleren, heeft de ontwikkeling van een nieuw vakgebied gestimuleerd: de bioinformatica. In de bioinformatica worden methoden en technieken uit de informatica en de wiskunde ingezet om biologische vraagstellingen en problemen te adresseren. Hierdoor wordt een compleet nieuwe benadering van onderzoeksvragen mogelijk gemaakt. Gezien de aard en de complexiteit van de nieuwe materie zal de bioinformatica in toenemende mate een sleutelrol vervullen binnen de "nieuwe" biologie.

Het streven van Nederland om internationaal een topositie in te gaan nemen binnen het genomics veld impliceert een navenante aandacht voor de bioinformatica. Om excellent genomics onderzoek te leveren is een bioinformatica infrastructuur die specialistische kennis, hoogwaardige faciliteiten en goed opgeleide mensen combineert essentieel. Op deze gebieden schiet Nederland momenteel te kort om op internationaal niveau competitief te zijn. Een forse inspanning is vereist om de benodigde infrastructuur te versterken. Reden voor de KNAW en NWO om in december 2001 een workshop te organiseren onder de titel "The future of bioinformatics in The Netherlands", waar verschillende buitenlandse bioinformatica experts hun visie op het vakgebied gaven. (zie voor verslag Appendix 4) Naar aanleiding van de workshop hebben NWO en KNAW een initiatief gestart ten behoeve van de positiebepaling van bioinformatica in Nederland, dat als uitgangspunt kan dienen voor het strategisch plan van het Nationaal Regie-orgaan Genomics. De hiervoor samengestelde NWO/KNAW commissie (samenstelling Appendix 1) heeft éénmaal vergaderd, daarnaast is er veel interactie geweest met de programmacommissie van het NWO programma Biomoleculaire Informatica (BMI).

De commissie heeft zich laten leiden door vier concrete onderwerpen, te weten de integratie van de bioinformatica in de (bestaande) onderzoeksprogramma's, de opzet en organisatie van een landelijke bioinformatica infrastructuur, de vormgeving van het bioinformatica onderwijs en de noodzaak om al dan niet bioinformatici uit het buitenland aan te trekken.

Deze position paper geeft de visie van de commissie op de gewenste positie van de bioinformatica binnen Nederland en schetst de initiatieven en acties die noodzakelijk zijn om de Nederlandse bioinformatica te versterken en internationaal een rol van betekenis te laten spelen.

2. Samenvatting

Voor alle aspecten van de bioinformatica geldt dat de internationale context en ontwikkelingen leidend zijn. Nederland moet ernaar streven zich aan te sluiten bij bestaande internationale initiatieven. Daarbij heeft het opzetten van een goede infrastructuur voor de bioinformatica de hoogste prioriteit. Deze infrastructuur moet gevormd worden door een netwerk van een beperkt aantal onderzoekszwaartepunten (samenwerkingsverbanden van onderzoeksgroepen uit universiteiten, instituten en eventueel industrie) rond een centrale faciliteit die beschikt over hoogwaardige hard- en software en bioinformatica experts. Dit centrum verleent services op alle terreinen van de bioinformatica, onderhoudt structurele contacten met internationale instituten, speelt een rol bij het bioinformatica onderwijs en is het ontmoetingspunt voor wetenschappers om ideeën en ervaringen uit te wisselen. De Bio-ASP (Bioinformatics Application Service Provider), een initiatief van de KUN en de UvA onder auspiciën van het NWO programma BMI, is zeer geschikt om uit te bouwen tot een dergelijke faciliteit en dient daarom op korte termijn over de benodigde financiële middelen te beschikken. Daarnaast dient een Raad van Toezicht te worden ingesteld die een goed contact tussen gebruikers en de centrale faciliteit garandeert.

Naast het opzetten van een adequate infrastructuur moet hoge prioriteit gegeven worden aan het opleiden van bioinformatici en het integreren van bioinformatici binnen de bestaande biologische, biomedische en biochemische onderzoeksstructuur. Er moet voor gewaakt worden om te snel de huidige behoefte aan bioinformatici op te vullen als dit ten koste gaat van de kwaliteit. Juist omdat de bioinformatica zich bevindt op het grensvlak van twee bestaande disciplines is het van het grootste belang dat bioinformatici behalve van het vak zelf, ook over gedegen kennis beschikken van tenminste één en bij voorkeur meerdere andere disciplines zoals, biologie, chemie, medische wetenschappen en wiskunde of informatica.

Om voor de toekomst over voldoende bioinformatici te kunnen beschikken dient er meteen gestart te worden met het integreren van bioinformatica in de WO en HBO curricula. Daarbij dient te worden aangesloten bij de lokale voorstellen tot inrichting van bachelors- en mastersopleidingen in de bioinformatica die nu circuleren. De commissie stelt voor om op korte termijn een landelijke coördinator aan te stellen die de lokale initiatieven inventariseert en onderlinge uitwisseling bewerkstelligt. Tevens dient het aantal AIO en post-doc plaatsen op dit gebied uitgebreid te worden. Om de huidige onderzoekers adequaat in de bioinformatica te trainen moeten er gerichte masterclasses en workshops georganiseerd worden, waarvoor de deelnemers actief geselecteerd worden uit het veld. Het actief aantrekken van bioinformatici uit het buitenland is vooralsnog geen afdoend middel om de Nederlandse leemten op te vullen. Het creëren van een aantrekkelijk klimaat waarbinnen excellent onderzoek verricht wordt is een constructievere manier om wetenschappers van topniveau aan te trekken.

Aanbevelingen

Voor eind 2002 dienen de volgende zaken duidelijk te zijn

- de (vervolg)financiering van de Bio-ASP (centrale faciliteit)
- de directeur van de Bio-ASP
- een landelijk coördinator voor het bioinformatica onderwijs
- extra budget voor onder andere AIO en post-doc projecten binnen het NWO programma BMI
- het creëren van opstartfondsen ten behoeve van bioinformatici

3. Internationale context

De commissie is van mening dat het voor Nederland in de eerste plaats van het grootste belang is om aan te sluiten bij bestaande internationale initiatieven. Dit geldt voor zowel onderzoeksinhoudelijke aspecten als voor onderwijs, soft- en hardware, databases en services. De archivering van primaire bio-data is internationaal geregeld en het is in het algemeen belang dat data die door Nederlandse onderzoekers worden gegenereerd in overeenstemming met de internationale afspraken worden gearhiveerd en beschikbaar gesteld, waarbij de afgesproken en nog te ontwikkelen standaarden strikt worden gevolgd. Voor Nederland is afstemming en samenwerking met het EBI (European Bioinformatics Institute, een outstation van het EMBL, European Molecular Biology Laboratory) daarvoor noodzakelijk. Nederland kan, als tegenprestatie voor het benutten van de diensten die het EBI beschikbaar stelt, bijdragen tot de internationale samenwerking door in bepaalde *niches* van het bioinformatica-onderzoek kennis en data beschikbaar te stellen. Dat kan bijvoorbeeld betrekking hebben op uniek onderzoek, dat in Nederland goed ontwikkeld is, of op algemeen bruikbare methoden en software die in het kader van Nederlands onderzoek is ontwikkeld, of op unieke databases die specifiek in Nederland verzameld materiaal bevatten (bijvoorbeeld genetische data van homogene bevolkingsgroepen). Ook op het gebied van educatie is internationale samenwerking geboden. (zie voor een nadere uitwerking van de relatie met het EBI Appendix 2)

Er is op bioinformatica gebied geen internationale benchmark voor de beste manier om één en ander te structureren, maar geschikte voorbeelden uit het genomics veld zijn voorhanden, zoals bijvoorbeeld de centrale programma's van Denemarken en Frankrijk. De commissie acht het nuttig deze programma's te bestuderen om eventueel praktische handreikingen te vinden voor de Nederlandse situatie.

4. Integratie van bioinformatica in andere onderzoeksgebieden

Essentieel bij een goede integratie van bioinformatica in de bestaande (biologische, bio-medische en biochemische) onderzoeksprogramma's en -structuur is dat de bioinformatici niet geïsoleerd worden. Er moet geen bioinformatica-eiland ontstaan dat losgekoppeld is van de overige disciplines. De bioinformatica kan juist een brug vormen tussen de bestaande gebieden. Om onderzoek van hoge kwaliteit te garanderen moeten bioinformatici duidelijke steunpilaren hebben binnen zowel de biologie als de informatica, gedegen kennis van één van beide disciplines is absoluut noodzakelijk. De bioinformatica moet vanuit de basisdisciplines gevuld worden. Het is geen goede zaak om vanwege de acute concrete behoefte snel het middenveld te vullen, dit zal ongetwijfeld leiden tot onderzoekers van lager niveau die noch van het één noch van het ander voldoende weten om toponderzoek te kunnen leveren.

Een belangrijk middel bij de integratie van de bioinformatica is een goed functionerende centrale faciliteit, waar alle bioinformatici contacten mee onderhouden. Omdat voor genomics-groepen bioinformatici onmisbaar zijn, zal het vakgebied via hen waarschijnlijk soepel integreren in de dagelijkse onderzoekspraktijk. Daarnaast levert de centrale faciliteit het middel om contacten tussen bioinformatici van de verschillende groepen te onderhouden. Intensieve contacten kunnen voorkomen dat decentrale bioinformatici het contact met de overige Nederlandse bioinformatici verliezen. Tegelijkertijd moet men er alert op zijn dat ook de bioinformatici van de centrale faciliteit geen geïsoleerd eiland gaan vormen, maar duidelijk ingebed zijn in het landelijke netwerk van bioinformatici die werkzaam zijn bij universiteiten, instituten en bedrijven.

5. Infrastructuur

Om het bioinformatica onderzoek in Nederland echt van de grond te krijgen moet de eerste prioriteit liggen bij het opzetten van een adequate infrastructuur. Deze dient te zijn opgebouwd uit de volgende componenten:

- Een beperkt aantal onderzoekszwaartepunten, waar bioinformatica onderzoek op een wetenschappelijk hoogstaand en internationaal competitief niveau wordt uitgevoerd.
- Het doordrenken van onderzoeksgroepen die zich bezighouden met biomoleculair onderzoek met de benodigde bioinformatica expertise.

- Een centrale faciliteit, die onderzoekers toegang biedt tot software en databases, als vraagbaak dient en als plaats waar onderzoekers en ideeën vanuit verschillende achtergronden samenkomen.

6. Centrale faciliteit

Een centrale faciliteit dient te beschikken over hoogwaardige hard- en software en, niet in de laatste plaats, over experts op deze terreinen. Deze faciliteit moet fungeren als centraal punt voor alle services op het gebied van bioinformatica. Onderzoekers uit verschillende disciplines moeten via de diensten van deze centrale faciliteit toegang kunnen krijgen tot databases, grootschalige berekeningen uit kunnen voeren en advies kunnen inwinnen over praktische problemen (zie Appendix 3 voor een nadere uitwerking van de services van de centrale faciliteit). Het is de verantwoordelijkheid van de centrale faciliteit om continu up-to-date te blijven met de laatste ontwikkelingen en te zorgen voor state-of-the-art hard- en software. Deze faciliteit moet technologische antwoorden vinden voor de vraagstelling van de onderzoekers en nieuwe technologieën implementeren. Ze moet gangbare technologieën en softwarepakketten stroomlijnen en voor iedereen bruikbaar maken (onder meer door het aanbieden van gebruikerscursussen) en irrelevante zaken of zaken die decentraal opgelost kunnen worden afstoten (dus geen helpdesk zijn voor alledaagse computerproblemen). Naast de technologische verantwoordelijkheden, zal de centrale faciliteit tevens verantwoordelijk zijn voor het aanbieden van onderwijs voor onderzoekers uit het veld, zoals gerichte workshops, cursussen en trainingen (zie hiervoor tevens punt 7). De centrale faciliteit zal structurele relaties opbouwen met internationale instituten en het aanspreekpunt zijn voor bioinformatica in Nederland. Tenslotte moet de centrale faciliteit fungeren als de ontmoetingsplaats bij uitstek voor onderzoekers, waar men ervaringen en ideeën uit kan wisselen.

Naast de primaire dienstverlenende taak van een centrale faciliteit moet er voldoende structurele ruimte zijn voor eigen onderzoek. Dit is noodzakelijk om het kennisniveau continu te verbeteren en om voldoende goede onderzoekers aan te trekken en te behouden. Om te voorkomen dat het eigen onderzoek gaat interfereren met de dienstverlenende verantwoordelijkheid dient het eigen onderzoek gelimiteerd te zijn en te volden aan strenge kwaliteitseisen.

De commissie is van mening dat de opzet en structuur van de Bio-ASP (Bioinformatics Application Service Provider), een samenwerking tussen de KUN en de UvA, voldoet om te gaan fungeren als de centrale faciliteit. De Bio-ASP begint zich echter nog in de beginfase, er zal nog veel geïnvesteerd moeten worden (zowel in financiële middelen als in tijd van de betrokkenen) om aan de hierboven beschreven eisen te voldoen. De Bio-ASP is echter niet de enige plaats in Nederland waar bioinformatica onderzoek plaatsvindt. De verschillende onderzoeksgroepen die zich nu reeds met bioinformatica bezighouden dienen daarom nauw betrokken te worden bij de plannen rond de Bio-ASP, die een centrale plaats binnen dit netwerk van groepen zal innemen.

Organisatiestructuur Bio-ASP

Er moet goed operationeel management zijn om ervoor te zorgen dat de eerste prioriteit van de Bio-ASP goede dienstverlening is en blijft. Momenteel is drs. Jan Willem Tellegen van het HPCN als interim-directeur aangesteld, gezien de termijn van deze interim-aanstelling dient de functie eind 2002 definitief vervuld te zijn.

Om de directeur daadwerkelijke zeggenschap over de gang van zaken te geven, is het noodzakelijk dat de Bio-ASP op termijn overgaat in een onafhankelijk instituut (waarschijnlijk gevestigd op meerdere locaties), waarbij duidelijke afspraken bestaan over de relatie tussen de Bio-ASP en de KUN, de UvA en de eventuele overige locaties. Voor de korte termijn moet de nadruk echter liggen op het snel opbouwen van de faciliteiten op beide locaties, waarbij in Nijmegen de nadruk zal liggen op de service en in Amsterdam op de techniek. Vervolgens dient de Bio-ASP uitgebreid te worden met personen die continu contact houden met de gebruikers om de behoeftes in het land te peilen en deze door te spelen aan de experts binnen de Bio-ASP. Er moet in de organisatie voor gewaakt worden dat de onderzoekers binnen de Bio-ASP niet (teveel) belast zijn met organisatorische aangelegenheden.

Raad van Toezicht Bio-ASP

Een belangrijke taak van de directeur zal zijn om continu in contact te staan met betrokken onderzoekers in het land om af te stemmen en te toetsen of de diensten van de Bio-ASP nog voldoen aan de behoeften van de afnemers. Goede terugkoppeling met de gebruikers in het land is cruciaal om de Bio-ASP succesvol te laten zijn. De commissie pleit voor het instellen van een Raad van Toezicht, waarin afnemers van de diensten van de Bio-ASP zitting hebben en die verantwoordelijk is voor het doorlopend evalueren van de diensten van de Bio-ASP. De directeur is verantwoording verschuldigd aan de Raad en beslissingen van de Raad moeten gevolgen kunnen hebben voor de activiteiten en financiering van de Bio-ASP.

De commissie acht het van groot belang om eerst te zorgen dat deze centrale faciliteit en de voorgestelde structuur goed van de grond komen. Op de korte termijn zou de nadruk moeten liggen op verbetering van de kennis en mogelijkheden van de Bio-ASP en het aanhaken bij internationale initiatieven, gevolgd door een regelmatige evaluatie van het functioneren in landelijk verband.

Financiering Bio-ASP

Vanuit het NWO programma BMI is EUR 0,91 miljoen beschikbaar gesteld tot januari 2003 voor de Bio-ASP. Uiterlijk eind 2002 dient er duidelijkheid te zijn over de vervolfinanciering. Deze kan mogelijk voortkomen uit ICES/KIS 3, want de plannen voor de Bio-ASP passen namelijk zeer goed binnen de gestelde criteria. Verwacht wordt dat het Nationaal Regie-orgaan Genomics een voorstel voor ICES/KIS 3 zal voorbereiden. Gezien de belangstelling die een aantal bedrijven hebben getoond dient tevens het vormen van een publiek-privaat consortium ten behoeve van de Bio-ASP serieus te worden overwogen.

7. Onderwijs

Om Nederland mee te laten draaien met de wereldtop op bioinformatica gebied moet het aantal goede bioinformatici drastisch toenemen. Eén van de belangrijkste middelen om meer bioinformatici te genereren is het onderwijs. Daarbinnen zijn twee doelgroepen te onderscheiden; studenten (HBO en WO) en AIO's, post-doc's en overige onderzoekers.

Studenten:

Om voor de toekomst goede bioinformatici op te leiden is het van groot belang om bioinformatica een volwaardige plaats te geven binnen het WO en HBO onderwijsprogramma, bijvoorbeeld in de vorm van een aparte bioinformatica-masters opleiding. Daarnaast is het essentieel om ook studenten biologie, geneeskunde en overige levenswetenschappen basiskennis rond bioinformatica bij te brengen, omdat zij in hun latere beroepspraktijk hier zeer waarschijnlijk mee te maken krijgen. Een deel van hen zal zich gaan specialiseren in de bioinformatica. Voor de overigen geldt dat het in toenemende mate van belang zal zijn om goed te kunnen communiceren met bioinformatici.

Landelijke coördinatie

De inhoud van het curriculum is de verantwoordelijkheid van universiteiten en hogescholen. Het is daarom moeilijk om direct invloed uit te oefenen op de diverse onderwijsprogramma's of om bepaalde onderdelen (mede) te financieren. Momenteel worden al veel initiatieven ontplooid op onderwijsgebied voor beide doelgroepen. Om te voorkomen dat er dubbel werk wordt verricht stelt de commissie voor een persoon aan te wijzen die zal optreden als een landelijk coördinator van het bioinformatica onderwijs en die de betrokkenen van universiteiten en hogescholen bij elkaar brengt om taken te verdelen en materiaal uit te wisselen. Met dit doel wordt er vanuit NWO BMI in juni een eerste bijeenkomst georganiseerd.

AIO's, post-doc's en overige onderzoekers

Het onderwijs aan deze groep kan het beste plaatsvinden in de vorm van gerichte PhD-cursussen, workshops en trainingen. Het vereiste cursusaanbod is grofweg in twee groepen te verdelen. Enerzijds zijn er wiskundigen, informatici en fysici die zich willen specialiseren in de bioinformatica, en die opleiding behoeven in geavanceerde methoden op een niveau dat zij daar zelf aan kunnen (gaan) bijdragen, en die tevens een vrij elementaire maar brede opleiding in

biologie en genetica behoeven. Anderzijds zijn er moleculair-biologen, biochemici, farmaceuten en medici, die vooral een hands-on opleiding in de methoden nodig hebben, maar zeer geavanceerd opgeleid moeten worden in de toepassingen.

De cursussen dienen de verantwoordelijkheid van de Bio-ASP te worden, die daarvoor ook gebruik zal maken van reeds ontwikkeld materiaal van buitenlandse initiatieven. Om het huidige kennisniveau snel te verhogen, moeten de deelnemers actief geselecteerd worden en kosteloos aan de cursus of workshop kunnen deelnemen. Op deze, internationaal zeer gebruikelijke, manier is de drempel om deel te nemen laag en wordt de juiste doelgroep aangesproken. De financiering van deze onderwijsactiviteiten dient te worden opgenomen in het programma van het Nationaal Regie-orgaan Genomics.

Vanuit het NWO-BMI programma worden al verschillende activiteiten ondernomen. Tevens komt het aantal bioinformatica-AIO's vanuit dit programma goed op gang, waarbij aangetekend moet worden dat het aantal AIO en post-doc plaatsen nog aanzienlijk uitgebreid moet worden. De commissie acht het wenselijk dat voor dit doel extra budget beschikbaar komt. Het is de verwachting dat vanuit deze hoek geleidelijk een bioinformatica onderzoeksschool zal groeien, waarbinnen een goed opleidingsprogramma een duidelijke plaats heeft.

8. Bioinformatici uit het buitenland

Het zal nog enige jaren duren voordat er binnen Nederland voldoende bioinformatici opgeleid zijn om in de behoefte te kunnen voorzien. De vraag rijst dan ook of Nederland actief bioinformatici uit het buitenland moet aantrekken. De commissie is van mening dat door eerst te zorgen voor een voldoende kennis en expertise niveau binnen Nederland, er een katalyserende werking kan ontstaan. Als de Bio-ASP op volle sterkte is en goed draait wordt het vanzelf aantrekkelijk voor buitenlandse bioinformatici om zich hier te vestigen, hetzij bij de Bio-ASP zelf, hetzij bij universitaire onderzoeksgroepen die van de diensten van de Bio-ASP gebruik maken. Er moet een klimaat ontstaan waarbinnen onderzoekers mee kunnen draaien met de wereldtop en het voor buitenlanders aantrekkelijk wordt om naar Nederland te komen. De commissie acht het niet wenselijk om met louter financieel aantrekkelijke maatregelen mensen naar Nederland te halen. Dit sluit uiteraard niet uit dat in voorkomende gevallen actief gezocht moet worden naar buitenlandse kandidaten die in aanmerking komen voor een vacature in Nederland.

APPENDIX 1

Samenstelling NWO/KNAW commissie ten behoeve van de positiebepaling van bioinformatica in Nederland

Prof.dr. P.C. van der Vliet (voorzitter)
Prof.dr. H.J.C. Berendsen (namens KNAW)
Prof.dr. T. Berns (namens KNAW)
Prof.dr. P. Folstar (namens Nationaal Regie-orgaan Genomics)
Prof.dr. W.J. Stiekema (namens NWO)
Prof.dr. J. de Vlieg (namens NWO)

Dr. M. Brandsma (ZonMw, coördinatie)
E.R. Thole (position paper)

APPENDIX 2

*Verslag van het gesprek van Herman Berendsen met Janet Thornton, Directeur EBI, over de plaats van een op richten Nederlands bioinformatica instituut in de internationale context en een mogelijke samenwerking met het EBI.
23 april 2002*

In het gesprek kwam het aan de orde welke rol een nationaal instituut kan vervullen tussen de lokale gebruikers en de internationale instituten in. Thornton ziet, evenals de commissie, zo'n rol wel als zinvol, enerzijds om gebruikers specifieke services te verlenen waarvoor instituten als het EBI niet geëquipeerd zijn, en anderzijds om een centrale rol te vervullen op het gebied van educatie. Een dergelijk instituut zou goede banden moeten hebben met het EBI, zich conformeren aan standaarden van data definitie en -exchange, en de bereidheid hebben data die zij genereert of beheert, ook internationaal ter beschikking te stellen. Op het gebied van educatie, waar het EBI nu een centrale rol vervult door het organiseren van cursussen, maar daar eigenlijk nauwelijks capaciteit voor heeft, ziet Thornton de mogelijkheid tot effectieve samenwerking, onder meer in gezamenlijke EU projecten in het kader van het zesde framework programma. Thornton gaf duidelijk aan dat het EBI een taak ziet in het organiseren van cursussen voor *opleiders* en niet voor eindgebruikers. Daarvoor hebben zij eenvoudig de middelen niet.

APPENDIX 3

Nadere uitwerking services Bio-ASP

Specifieke services

Een centrale faciliteit als de Bio-ASP kan de volgende services verlenen:

1. *Toegang van gebruikers regelen tot op internet beschikbare databases* . Daarbij gaat het niet in de eerste plaats om de databases die via EBI en NCBI beschikbaar zijn, maar vooral om het opsporen, evalueren en toegankelijk maken van andere data die elders beschikbaar worden gesteld (de zg. *boutique databases*). Die kunnen voor specifiek Nederlands onderzoek van belang zijn. Deze wereld is aan snelle verandering onderhevig en vereist actief zoeken en evalueren, wat het beste door een gespecialiseerd instituut kan gebeuren, dat goede contacten met gebruikers onderhoudt. Ook gaat het om het beschikbaar stellen aan gebruikers van data uit commerciële databases , waarbij een centraal instituut de nodige regelingen op landelijk niveau kan treffen.
2. *Faciliteiten bieden voor het uitvoeren van grootschalige analyses en berekeningen*. Hiervoor moet programmatuur en apparatuur, of de toegang daartoe, centraal beschikbaar worden gesteld. De centrale faciliteit moet goede contacten onderhouden met lokale ontwikkelaars van programmatuur en die programmatuur geschikt maken voor landelijk (en in sommige gevallen internationaal) gebruik. Ook kan specifieke programmatuur in de nationale faciliteit worden ontwikkeld, wanneer dat aansluit op behoeften van gebruikers. De centrale faciliteit kan in het bijzonder een belangrijke rol vervullen wanneer de toepassing informatica-expertise vereist (die in de centrale faciliteit geconcentreerd is), bijvoorbeeld om methodieken en standaardisaties te vrijwaren van amateuristische ontwikkelingen en om moderne web-based interfaces te definiëren.
3. *Databases onderhouden voor gebruikers*. Gebruikersgroepen zullen in een aantal gevallen specifieke databases voor hun onderzoek nodig hebben, die hetzij samengesteld worden van internet data, hetzij uit eigen onderzoek gegenereerd worden. Het is van belang dat zulke databases professioneel worden ingericht en beheerd, en aan de internationale afspraken voldoen. Ook het inbrengen in de internationale community kan beter vanuit een centrale faciliteit gebeuren.
4. *Aanbieden van cursussen en vervaardigen van cursusmateriaal*. Een centrale faciliteit met een eigen onderwijscoördinator kan enerzijds cursussen organiseren voor opleiders in de lokale instellingen en instructiemateriaal voor decentrale cursussen ter beschikking stellen, en anderzijds ook cursussen organiseren voor eindgebruikers die zich specialeren in de bioinformatica. Daarbij is intensieve samenwerking met het EBI, dat een cursusprogramma voor opleiders heeft, van groot belang, ter wille van efficiëntie en optimale benutting van aanwezige expertise. Vanuit Nederland kunnen ook specifieke cursussen in het internationale opleidingskader worden ingebracht.

APPENDIX 4

Verslag KNAW/NWO Workshop "The future of bioinformatics in the Netherlands"
7 december 2001

- Integratie van bioinformatica in de onderzoeksinfrastructuur
- Uniforme standaarden
- Onderwijs
- Bioinformatica in Nederland
- Uitdagingen voor de bioinformatica

Verslag voor het opstellen van een *position paper*

Integratie van bioinformatica in de onderzoeksinfrastructuur

Om daadwerkelijk tot een multidisciplinaire infrastructuur te kunnen komen is het van belang dat de rol en toegevoegde waarde van de bioinformatica duidelijk is voor de "anderen" in het team. In veel gevallen wordt de bioinformatica gezien als een ondersteunend vakgebied, met de (bio)informaticus in een serviceverlenende rol. Op veel plaatsen is er bovendien gebrek aan voldoende technisch ondersteunend personeel, waardoor de bioinformaticus vaak wordt gezien als een soort helpdesk medewerker. Scheidingen tussen verschillende groepen wetenschappers ("bloedgroepen") moeten vermeden worden en de bioinformaticus moet een volwaardige plaats hebben binnen het onderzoeksteam met ruimte voor zijn/haar specifieke expertise en onderzoeksvragen.

Volgens Lee Hood heeft het Human Genome Project de visie op de biologie en geneeskunde veranderd, er is een geheel nieuwe benadering tot stand gekomen. Biologie is een informatiewetenschap geworden, die nieuwe strategieën te bieden heeft aan de wiskunde en informatica, terwijl vanuit deze gebieden de *tools* komen die het mogelijk maken om relevante biologische informatie te kunnen destilleren uit de enorme hoeveelheid beschikbare data. In zijn visie is het essentieel dat wiskundigen en (bio)informatici een diepgaand inzicht verwerven in de biologie. In het Institute for Systems Biology krijgt daarom iedereen een (basis)training biologie.

Victor Jongeneel schetst de noodzaak om medici de toegevoegde waarde van de bioinformatica te laten zien, aanhaken bij de dagelijkse medische praktijk is daarbij de beste manier. Concreet laten zien wat bioinformatica te bieden heeft, bijvoorbeeld op het gebied van nieuwe drug targets, pharmacogenomics ('patient profiling', 'personalized medicine'), genetic counselling, moleculaire diagnostiek en pathofysiologie. Het is niet de bedoeling om van artsen experts in dit veld te maken, maar zij moeten wel de waarde van de bioinformatica kunnen inschatten, weten wanneer de bioinformatica een bijdrage kan leveren en de geleverde gegevens juist kunnen interpreteren. De

moderne geneeskunde vereist een bioinformatica infrastructuur, in zowel het onderwijs als de medische praktijk.

In de groep van Bob Franza wordt er naar gestreefd om, binnen de centrale onderzoeksvraag, iedere onderzoeker een nieuwe uitdaging in zijn/haar eigen gebied te geven. Voor een wiskundige wordt er bijvoorbeeld gezocht naar een manier om niet alleen toepassingsgericht bezig te zijn maar ook op wiskundig gebied tot nieuwe inzichten te komen.

Samengevat kan worden geconcludeerd dat er nog veel energie gestoken moet worden in het verspreiden van kennis van de biologie en bioinformatica onder de betrokken onderzoekers om een daadwerkelijk geïntegreerde aanpak van de (nieuwe) biologische vragen op te leveren.

Uniforme standaarden

Een probleem waar veel onderzoekers tegenaan lopen en dat door alle sprekers genoemd werd, is het gebrek aan wereldwijde, uniforme standaarden met betrekking tot weergave en annotatie van de onderzoeksdata. Door de enorme vlucht die het genomics onderzoek de laatste jaren heeft genomen zijn er grote hoeveelheden data gegenereerd, en deze datastroom is nog steeds dermate groot dat het interpreteren en controleren van de (nieuwe) gegevens ver achter ligt. Het gevolg is een aanzienlijke "vervuiling" van diverse databanken. Er zal nog zeer veel werk verzet moeten worden om alle huidige bestanden op te schonen en een uniforme, consistente weergave van de data in te stellen.

Een belangrijke initiatief in dit verband is het Human Proteomics Initiative van het EBI (European Bioinformatics Institute), dat als missie heeft het samenbrengen van alle bekende proteoom informatie. De voor dit doel opgezette database SWISS PROT bevat momenteel zo'n 103000 eiwitsequenties. Om vervuiling van deze databank zoveel mogelijk te voorkomen is een stapsgewijs proces opgezet, waarbij nieuw aangeleverde gegevens eerst in een ander bestand worden opgeslagen, gecontroleerd en de annotatie zo compleet mogelijk wordt gemaakt alvorens de gegevens in SWISS PROT opgenomen worden. De grote bottleneck is ook hier de annotatie, er is een enorme stroom van ruwe data beschikbaar waarvan de kwaliteit niet gecontroleerd is en annotatie onjuist, onbetrouwbaar of zelfs afwezig is. Om structuur aan te brengen in deze data chaos wordt bij het EBI de functionele annotatie in vier stappen uitgevoerd:

1. automatische classificatie (InterPro: groeperen van eiwitten, gericht op het herkennen van kenmerken van eiwitten en deze relateren aan bestaande eiwitfamilies)
2. automatische karakterisatie (Gene Ontology consortium: eiwit wordt gekarakteriseerd op basis van 1. de moleculaire functie, 2. het biologische proces waarin het eiwit een rol speelt (bijv. mitose) en 3. de cellulaire component waarin het eiwit zich bevindt.)
3. automatische annotatie (het succes van deze stap is zeer afhankelijk van de betrouwbaarheid van de data die door stap 1 en 2 gegenereerd zijn, onbetrouwbare data leiden vaak tot onjuiste annotatie in deze stap)
4. handmatige volledige annotatie (vanwege de moeizame controle van alle data blijft volgens Rolf Apweiler handmatige annotatie nog steeds de meest betrouwbare, maar tijdrovende, methode)

Een uniforme standaard is niet alleen van toepassing op de annotatie en de wijze waarop data opgeslagen worden in databanken, maar ook op andere fronten is grote behoefte aan een meer

uniforme aanpak. Een voorbeeld is het verschil in de manier waarop gegevens over de structuur van een eiwit gepresenteerd worden in artikelen en de wijze waarop diezelfde gegevens in een databank opgeslagen zijn. Het is essentieel dat gegevens expliciet, accuraat en computer vriendelijk zijn. Alle leden van de onderzoeksgemeenschap hebben een rol te vervullen in het formuleren van adequate standaarden, niet alleen de onderzoekers maar ook redacteurs en uitgevers.

Onderwijs

De situatie op het gebied van opleidingen en trainingen in de bioinformatica is vergelijkbaar met de ontwikkelingen in het gehele genomics en bioinformatica veld. Er is een veelheid aan losse initiatieven, iedereen ontwikkelt een eigen cursus of programma die voor zijn/haar toepassingsgebied of technologie belangrijk is, maar er is geen structurele aanpak waarneembaar. Te oordelen naar de verschillende sprekers is er grote behoefte aan bioinformatica opleidingen en trainingen, en aan gekwalificeerde personen om dergelijke programma's te onderwijzen. Vanwege het multidisciplinaire karakter van het bioinformatica onderzoek is het zaak verschillende niveaus en doelgroepen te hanteren bij het opstellen van bioinformatica onderwijs.

Er zijn in dit verband vier doelgroepen te identificeren:

1. Studenten biologie en geneeskunde die ontvankelijk gemaakt moeten worden voor bioinformatica en de rol van dit vakgebied binnen de *life sciences* en de geneeskunde
2. Studenten die als bioinformaticus willen gaan werken, bijvoorbeeld in de (centrale) technologische onderzoeksfaciliteiten
3. Onderzoekers die al actief zijn op het gebied van genetica, biologie, etc., en die behoefte hebben aan meer kennis van bioinformatica
4. Wiskundigen, informatici en technici die basiskennis biologie moeten verwerven

Volgens Jongeneel zal het onderwijs voor groep 1 en 2 decentraal georganiseerd moeten worden, onder meer omdat op die manier de drempel voor studenten lager is om aan bioinformatica onderwijs deel te nemen. Echter, hij onderkent dat dit, zeker op korte termijn, niet haalbaar is voor ieder instituut afzonderlijk. Daarom zou er begonnen moeten worden met een centraal programma wat stapsgewijs uitgebreid wordt naar meerdere universiteiten en instituten. Hij is van mening dat de meer vakgerichte trainingen voor groep 3 en 4 het beste centraal georganiseerd kunnen worden. Tevens schetst hij, evenals Philip Bourne, de grote behoefte die er bestaat aan mensen die bioinformatica onderwijs kunnen verzorgen. Er zijn internationaal al veel initiatieven op onderwijs gebied, het is belangrijk om daarbij aan te sluiten en zoveel mogelijk te profiteren van de beschikbare kennis en ervaring op dit gebied. Een voorbeeld is de International Society for Computational Biology (www.scb.org), maar ook vanuit het EBI wordt veel ondernomen op het gebied van training.

Met betrekking tot onderwijs aan studenten stelt Bourne dat er een duidelijke behoefte is aan formele academische programma's en dat het hierin essentieel is om meteen voor de geïntegreerde aanpak te kiezen, onderwijs over de verschillende biologische niveaus (van atoom naar ecosysteem) heen. De biologische context moet de drijfveer achter de vraagstellingen zijn, zowel in onderzoek als onderwijs. Tegelijkertijd realiseert hij zich dat dit op de korte termijn nog niet haalbaar is en dat voorlopig nog steeds de ontwikkelingen op technologisch gebied de handvatten zullen leveren voor een geïntegreerde training.

Bioinformatica in Nederland

Suggesties Bob Franza (CSI)

Franza suggereert het opzetten van een 'virtueel adaptief' instituut met een focus op onderzoek naar cellulaire systemen. Dit instituut bestaat uit twee componenten: een groep experts (Groep 1) die zorgen voor de technologische en rekenkundige faciliteiten (zowel hardware als software) en de groep van individuele of groepen universitaire, institutionele en industriële onderzoekers (Groep 2) die een wetenschappelijke interesse hebben in het functioneren van levende systemen.

Groep 1 is verantwoordelijk voor opzetten van de analytische platforms en de kennisinfrastructuur, uiteraard via een netwerk toegankelijk, die Groep 2 nodig heeft om het onderzoek te kunnen uitvoeren. Voor de duidelijkheid, Groep 2 is geen eenheid, het is de overkoepelende paraplu van groepen onderzoekers binnen het veld, waarvan de samenstelling voortdurend in ontwikkeling is. Voorstellen van samenwerkende groepen binnen Groep 2 zullen in een *peer review* proces beoordeeld worden, alvorens men aanspraak kan maken op de faciliteiten die Groep 1 levert. De wetenschappers en technici binnen Groep 1 zijn niet louter dienstverleners, deze groep is verantwoordelijk voor het genereren van steeds geavanceerdere datamanagement systemen, modelling methodes, analytische software en instrumentatie, allemaal activiteiten die meestal binnen een academische setting niet goed uit de verf komen.

De wetenschappers binnen Groep 2 moeten in staat zijn innovatieve onderzoeksvorstellen te formuleren waarin maximaal gebruik wordt gemaakt van de middelen die Groep 1 kan leveren. Het 'virtueel adaptieve' aspect van deze structuur naar voren omdat:

1. er continue interactie is tussen Groep 1 en 2
2. de voorstellen vanuit Groep 2 de drijfveer zijn voor het verleggen van technologische grenzen door Groep 1 (die daarmee weer Groep 2 stimuleert tot het verleggen van de grenzen van de onderzoeksvragen)
3. de experts in Groep 1 geacht worden ook zelf continu de grenzen van de technologie te verleggen
4. er een strikt beoordelingsproces is voor het werk binnen beide groepen, zodat iedereen er van doordrongen blijft dat innovatie, pionieren en het nemen van risico's essentieel is
5. de onroerende faciliteiten van Groep 1 op maximaal twee locaties gehuisvest zijn
6. de middelen van Groep 2 niet geografisch gebonden zijn, deze zijn verbonden aan de individuele onderzoekers binnen de samenwerkingsverbanden

Door duidelijk onderscheid te maken tussen onroerende, kapitaal intensieve technologische faciliteiten en (niet-plaatsgebonden) onderzoeksvorstellen is het mogelijk om een multi-inzetbare unit te vormen voor observatie van analyse van cellulaire systemen die de vereiste middelen kan bieden voor de onderzoeksvragen door de diverse samenwerkingen binnen Groep 2 geformuleerd worden. Er is een duidelijke overeenkomst met de Hubble-telescoop, het continue onderhoud en verbeteren van de telescoop door vele experts maakt het mogelijk dat nog meer wetenschappers hun "eigen" onderzoek kunnen uitvoeren. Deze onderzoekers hoeven niet zelf bij de Hubble betrokken te zijn om wel te kunnen werken met de data die centraal gegenereerd worden.

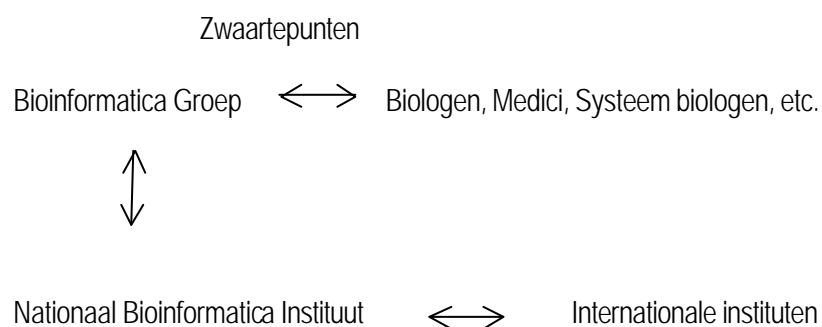
Franza stelt voor om de beschikbare gelden te gebruiken om biologie om te vormen van een wetenschap gericht op het verzamelen, beschrijven en determineren naar een wetenschap van levende systemen. Gebruik de beschikbare middelen om een vaste operationele entiteit op te zetten die los staat van de (wisselende) multidisciplinaire samenwerkingsverbanden die een beroep doen van de faciliteiten van deze entiteit. Zorg ervoor dat er geen onderlinge technologische competitie ontstaat, geen middelen toewijzen aan ad hoc groepen die (kleine stukken) software of algoritmen zelf willen gaan ontwikkelen. De bovengenoemde operationele entiteit heeft als één van haar belangrijkste taken het bijhouden van de ontwikkelingen in het veld om zo een goede inschatting te kunnen maken van de beschikbare oplossingen voor een probleem. Pas na een grondige verkenning zal deze entiteit besluiten zelf nieuwe methoden te ontwikkelen.

De eerste belangrijke stap is het fysiek samenbrengen van de topwetenschappers en technologen die de operationele structuur van de grond krijgen. Deze groep verdient eerste prioriteit, daarna volgen de personen die op een meer virtuele basis betrokkenen moeten worden bij een dergelijke structuur.

Voorgestelde structuur (voortgekomen uit de afsluitende discussie)

Binnen de (Genomics) Zwaartepunten werken biologen, medici en overige *life science* onderzoekers in een iteratief proces samen met de bioinformatica groep om de biologische vraagstellingen te definiëren en te onderzoeken. De bioinformatica groep staat in nauw contact met een nationaal bioinformatica instituut. Dit instituut beheert de benodigde technologische faciliteiten, zoals supercomputers, databestanden etc. die ten dienste staan van de gehele Nederlandse onderzoeksgemeenschap. Vanuit dit instituut worden tevens trainingen aangeboden voor gebruikers. Bioinformatica onderwijs aan studenten zal binnen iedere universiteit afzonderlijk georganiseerd worden, al dan niet vanuit een gemeenschappelijk geformuleerd plan. Het nationale instituut onderhoudt de (technologische) contacten met internationale groepen zoals het EBI, Cell System Initiative, Institute for Systems Biology etc.

In schema:



De toekomst: uitdagingen voor de bioinformatica

Korte termijn (initiatieven zijn al in gang gezet)

- Bioinformatica beschikbaar maken voor biologen en medici
- Volledige en up-to-date databestanden met een minimum aan redundantie
- Maximale informatie (annotatie) over iedere sequentie
- Coherente, uniforme opslag/weergave van de gegevens
- Schoonmaak van de huidige databestanden
- Annotatie wordt ook een high throughput technologie: nieuwe rekenmodellen vereist

Langere termijn

- Meer focus op vergelijken van genen (*comparative genomics*) en structuren dan op het genereren van nieuwe data
- Integratie van gegevens over de biologische niveaus heen: multidisciplinaire aanpak essentieel
- Integratie van wetenschappers uit verschillende disciplines
- Effectieve industriële partnerships en spin-offs

Continue aandachtspunten

- Up-to-date houden van technologische faciliteiten (vergt geld en onderhoud)
- Up-to-date houden van rekenmodellen, algoritmen, software etc.
- Koppelen van de verschillende databestanden