

Integrale aanpak voor digitalisering in wetenschap

Inleiding

Op alle terreinen is de impact van digitalisering van de wetenschap merkbaar. De manier van inwinning, analyse, deling en presentatie van onderzoeksdata- en bronnen is sterk aan het veranderen: de omvang en complexiteit van datasets is fors toegenomen, en internationale en discipline-doorkruisende samenwerkingen zijn belangrijker en meer gemeengoed dan ooit.

Met de steeds luidere roep om FAIR (Findable, Accessible, Interoperable, Reusable) in het kader van de principes van Open Science ligt er nu een sterke nadruk op het delen en vindbaar maken van data. Onderzoeksdata vormen echter maar één facet van de digitalisering. Hetzelfde geldt voor de enorme groei in de behoefte van rekencapaciteit. Enerzijds is er een steeds verder groeiende behoefte aan snellere supercomputers en anderzijds —aangejaagd door het beschikbaar komen van grote volumes aan metingen— neemt de vraag naar rekenkracht voor nieuwe analysevormen, zoals Machine Learning, vanuit bijna alle wetenschapsgebieden sterk toe. Dataverwerking is vaak afhankelijk van specialistische software. Veel van die software wordt binnen projecten ontwikkeld en gaat na afloop van die projecten verloren. Software en data zijn onlosmakelijk met elkaar verbonden; verduurzaming van software is essentieel om het delen van data een succes te maken.

Een sterk ontwikkelde digitale infrastructuur is essentieel om de digitalisering in goede banen te leiden, en toereikende rekenkracht, opslagruimte en snelle netwerken zijn een belangrijke randvoorwaarde. Om deze infrastructuur toegankelijk te maken voor alle onderzoekers is expertise nodig; laagdrempelig en lokaal waar mogelijk, specialistisch en gecentraliseerd waar nodig. Net als data, software en rekenkracht moet lokale expertise kunnen worden gedeeld.

Bij sommige wetenschapsgebieden is de digitalisering jaren geleden al op gang gekomen en is het delen van data, software en rekencapaciteit gemeengoed. Andere wetenschapsgebieden staan pas aan het begin van deze cruciale digitaliseringsslag. De behoeften lopen dan ook sterk uiteen. Voor sommige wetenschappers is toegang tot high-end computing essentieel, anderen hebben behoefte aan het ontsluiten en/of combineren van publiek beschikbare data en weer anderen hebben behoefte aan het delen van software. In alle gevallen betreft het een combinatie van infrastructuur en expertise om de wetenschap beter te ondersteunen.

Uitgangspunten en draagvlak

Het ICT adviesrapport "Topwetenschap vereist topinfrastructuur", in 2017 opgesteld door de ICT commissie van de Permanente Commissie Grootchalige Wetenschappelijke Infrastructuur (o.l.v. prof. dr. S.M.E. Wyatt) vormt het fundament van de voorgestelde bestedingen in dit advies. Echter, gezien de kleinere hoeveelheid beschikbare middelen dan voorgesteld in dat advies was het noodzakelijk de prioritering te herzien. Er is ook sprake van voortschrijdend inzicht, deels tot stand gekomen via een peiling van de noden bij een brede groep stakeholders, dat ertoe geleid heeft dat subcommissie nu voorstelt ook middelen ter beschikbaar te stellen voor lokale expertise. De categorisering van de bestedingen is ten opzichte van het oorspronkelijke adviesrapport aangepast om te benadrukken waar de middelen gaan landen en waar zij aan ten goede komen. Ambities binnen de categorieën netwerken en cybersecurity worden gerealiseerd middels investeringen in expertise bij SURF, in het gefedereerde systeem, en in lokale en interuniversitaire Digitale Competentie Centra (DCC's, zie het hoofdstuk *Gecombineerde aanpak digitalisering*). Behalve bovengenoemd adviesrapport is ook Thema 2 van het Nationaal Plan Open Science (*hergebruik van data*) een belangrijk uitgangspunt geweest voor de totstandkoming van dit advies. Een aantal in dit advies voorgestelde investeringen komen ten goede aan het realiseren van de ambities die met dit thema samenhangen, m.n. de stimulering van investering in data stewards in lokale DCC's en de stimulering van interuniversitaire DCC's, vele van welke data opslag en uitwisseling volgens de FAIR principes faciliteren.

Voor de totstandkoming van dit advies heeft de subcommissie een brede doorsnede van het wetenschappelijke veld geraadpleegd, en is input verkregen van de belangrijkste relevante bestuurlijke en faciliterende partijen. Voor dit advies heeft de subcommissie inbreng gekregen van SURF, DANS, 4TU, NLeSC, VSNU, VH, NFU, KNAW, de Nationaal Coördinator Open Science, MinOCW en vertegenwoordigers van de faciliteiten op de Roadmap Wetenschappelijke Grootschalige Infrastructuur.

Gecombineerde aanpak digitalisering

Lokale Digitale Competentie Centra

De digitalisering van de wetenschap en de principes van Open Science vereisen een nieuw pakket van kennis en vaardigheden; data moeten worden opgeslagen volgens de FAIR principes, in overeenstemming met geldende (internationale, domein specifieke) standaarden; software moet duurzaam worden vervaardigd, onderhouden en gedeeld kunnen worden, en onderzoekers moeten de weg kunnen vinden in en naar de juiste (lokale, nationale of internationale) infrastructuur, zoals reken- en opslagfaciliteiten en data en software repositories. De hiervoor benodigde kennis is niet altijd in voldoende mate aanwezig en vindbaar bij de universiteiten. Hiertoe beveelt de subcommissie de oprichting en uitbouw van universitaire Digitale Competentie Centra (DCC's) aan. Hier wordt expertise op het gebied van data(stewardship), software en computing gebundeld op één locatie. Sommige universiteiten hebben al belangrijke stappen gezet op weg naar een voldragen DCC.

De subcommissie ziet de volgende functies voor de DCC's:

- Landingsplaats voor projectmatig ontwikkelde software, door eigen personeel, maar ook bijvoorbeeld door het Netherlands eScience Center (NLeSC); vanuit DCC's kan deze software onderhouden en openbaar gemaakt worden
- Kennis- en adviescentrum voor FAIR data en software
- Kennis- en adviescentrum voor lokale ICT-infrastructuur
- Knooppunt in een gefedereerd netwerk voor data, computing en expertise

Voor de oprichting van (en versterking van bestaande) DCC's adviseert de subcommissie NWO tijdelijk middelen beschikbaar te stellen. De subcommissie adviseert NWO hiervoor éénmalig een bedrag van **4,5 M€** als impulsfinanciering beschikbaar te stellen. Hiervoor zal een eenvoudig financieringsinstrument worden ingericht door NWO, waarbij de administratieve lasten voor aanvragers worden geminimaliseerd.

Digitale Competentie Centra in een gefedereerd systeem

Om FAIR te laten slagen is het noodzakelijk dat "lokaal gehuisveste" data en software toegankelijk worden vanaf andere onderzoeksinstellingen. De subcommissie beveelt de oprichting van een gefedereerd netwerk van lokale infrastructuren aan om dit te realiseren. De lokale DCC's worden de knooppunten in dit netwerk. SURF heeft een sleutelrol in het technisch faciliteren en coördineren van dit secure, gefedereerd systeem, dat deze universitaire DCC's onderling zou moeten verbinden. Security in relatie tot FAIR, waarbij toegang tot data wordt verschaft conform de intentie van de eigenaar, is een belangrijke pijler van het systeem. Belangrijk is dat de aangesloten lokale infrastructuren middels het gefedereerde systeem geïntegreerd moeten kunnen worden in de European Open Science Cloud (EOSC), die in ontwikkeling is. Naast het ontwikkelen en onderhouden van de technische aspecten van dit systeem zet SURF hun expertise en ervaring op het gebied van ICT-infrastructuur ook in voor het verzorgen van ondersteuning aan de DCC's. Voor de ontwikkeling, het onderhoud en het leveren van ondersteuning adviseert de subcommissie NWO jaarlijks een bedrag van **2 M€** beschikbaar te stellen aan SURF. De subcommissie beveelt aan dat SURF samen met de CIO's van universiteiten en onderzoeksinstituten een plan opstelt waarin de functionaliteit van het gefedereerde systeem wordt beschreven, en hoe het zal worden gerealiseerd. Om te waarborgen dat het gefedereerde systeem in voldoende mate bijdraagt aan het realiseren van de ambities gesteld in het Nationaal Plan Open Science (NPOS) zouden SURF en de het landelijk CIO overleg, alvorens het plan uit te werken, de hoofdlijnen moeten afstemmen met de stuurgroep NPOS.

Interuniversitaire Digitale Competentie Centra

Er bestaan organisaties die gezien kunnen worden als landelijke, interuniversitaire DCC's: organisaties als DANS, 4TU Data Center, HealthRI, en de Nederlandse knopen van bestaande Europese onderzoeksinfrastructuren, combineren data-, software- en infrastructuurexpertise met hun domein-

specifieke kennis. De subcommissie adviseert NWO voor deze categorie DCC's jaarlijks een bedrag van **2,5 M€** ter beschikking te stellen vanaf 2019. Dit bedrag wordt vrijgegeven op basis van gerichte projectvoorstellen die matching vereisen van academische instituten; op deze manier kunnen de instellingen zelf keuzes maken in het prioriteren van deze DCC's.

Het prioriteren van investeringen in interuniversitaire DCC's vereist een brede, discipline- en domeinoverstijgende visie op digitalisering. De subcommissie adviseert gebruik te maken van de kennis van het NLeSC om de middelen die hiervoor worden gereserveerd te verdelen. De subcommissie beveelt aan dat NLeSC in samenwerking met NWO een financieringsinstrument ontwerpt voor de verdeling van deze middelen. Zij zullen voorstellen welke partijen voor financiering in aanmerking komen en hoe een evenredige verdeling tussen wetenschappelijke domeinen kan worden bewerkstelligd. Gezien de ambities van DANS en SURF ligt een sterkere samenwerking tussen die partijen voor de hand.

Voorwaarden financiering DCC's

Het behoeft nadruk dat de kosten van de digitalisering van de wetenschap uiteindelijk vanuit de wetenschapsbudgetten gedragen zullen worden. Data stewards en research software engineers zullen langzamerhand een vast onderdeel van het ondersteunend personeel gaan uitmaken. Kennisinstellingen zullen zich hieraan moeten committeren wanneer zij gebruik maken van bovenstaande financiering. Voor de bekostiging van versterking van lokale DCC's adviseert de subcommissie het vragen van cofinanciering; naar mening van de subcommissie kan 30% van de kosten voor maximaal drie jaar worden gefinancierd uit de 1,5 M€ die hiervoor jaarlijks ter beschikking wordt gesteld, de overige 70% zou voor rekening van de kennisinstellingen moeten komen. Ook voor financiering van de interuniversitaire DCC's adviseert de subcommissie deze 30/70 verhouding, en er moeten waarborgen voor duurzame financiering door de kennisinstellingen worden afgegeven. De universiteiten zullen moeten zorgen voor dat de impulsfinanciering voor de lokale DCC's ten goede komt aan alle wetenschappelijke domeinen, en NLeSC en NWO zullen de gelijkmatige spreiding van de financiering voor interuniversitaire DCC's over de domeinen moeten waarborgen. Daarnaast doet de subcommissie de aanbeveling dat de VSNU gevraagd wordt te onderzoeken hoe de betrekkelijk nieuwe categorieën ondersteunend personeel die bij lokale DCC's worden gehuisvest een aantrekkelijk carrièreperspectief gegeven kan worden.

Rekenfaciliteiten

Hardware

De nationale supercomputer is een belangrijke faciliteit, en voor een aantal onderzoeksvelden een noodzakelijke voorwaarde om internationaal competitief te kunnen zijn. De subcommissie adviseert NWO een bedrag van **18 M€** beschikbaar te stellen voor het vervangen van de nationale supercomputer Cartesius, om deze onderzoeksvelden ook in te toekomst te kunnen bedienen, en aansluiting te behouden met internationale computing initiatieven als PRACE. De subcommissie adviseert NWO een éénmalige bijdrage ter hoogte van **2 M€** van SURF zelf te vragen om het gewenste ambitieniveau voor de supercomputer te kunnen realiseren. De subcommissie adviseert dat de middelen die hiervoor beschikbaar komen in beheer van NWO blijven totdat SURF overgaat tot aanschaf van de nieuwe supercomputer.

Ook adviseert de subcommissie NWO vanaf 2020 jaarlijks **2,5 M€** ter beschikking te stellen voor de financiering van andere rekensystemen bij SURF, met een eenmalige impuls van **3 M€** in 2019. Door te investeren in een breed spectrum van computing faciliteiten kan SURF ook gebruikers voor wie de supercomputer niet geschikt is voorzien in hun behoeften. Daarnaast adviseert de subcommissie NWO vanaf 2020 jaarlijks **1 M€** te investeren in uitbreiding van de opslag- en archiveringsmogelijkheden in de reken- datacentra van SURF, om tegemoet te komen aan de rappe toename van datavolumes, met een eenmalige impuls van **1,5 M€** in 2019. SURF is de aangewezen partij om te waarborgen dat de Nederlandse kennisinstituten aangehaakt blijven bij de nieuwste technologische ontwikkelingen; daartoe is het belangrijk te investeren in de vernieuwing en ontwikkeling van hun kennis. De subcommissie adviseert NWO vanaf 2019 jaarlijks **3,5 M€** te investeren in het opbouwen en vernieuwen van de voor de rekenfaciliteiten benodigde kennisbasis bij SURF. Deze middelen voorzien voor een deel in het verduurzamen van de financiering van de expertise die reeds aanwezig is bij SURF, en essentieel is voor de instandhouding van een ICT kenniscentrum van wereldniveau.

Het is belangrijk om te waarborgen dat de infrastructuurvernieuwing van SURF nauw aansluit bij de wensen en noden van de instituten die zij bedienen. De subcommissie beveelt daarom aan dat SURF de CIO's van de universiteiten en onderzoeksinstituten een rol geeft in het sturen van de vernieuwing bij SURF, om de doelmatigheid te waarborgen. De subcommissie vraagt SURF en het landelijk CIO-overleg te beschrijven hoe zij hun samenwerking op het front van infrastructuurvernieuwing gaan intensiveren in het plan voor het gefedereerde systeem.

Gezien de groeiende behoefte aan rekenkracht zal de opvolger van Cartesius naar verwachting in 2025 vervangen moeten worden. De subcommissie adviseert NWO vanaf 2020 jaarlijks **4 M€** uit deze infrastructuurmiddelen te reserveren voor toekomstige high performance computing (HPC) behoeften. Deze reserveringen zouden moeten worden beheerd door NWO. Om ook in de toekomst, wanneer de nieuwe supercomputer end-of-life is, in de behoeften van gebruikers van de supercomputer te kunnen voorzien, adviseert de subcommissie NWO vanaf 2020 een jaarlijkse bijdrage van **2 M€** aan deze reserveringen van SURF te vragen, te realiseren door tariefinkomsten van academische gebruikers en commerciële exploitatie van de machine. Uit deze reserveringen kan, naast een uiteindelijke vervanger, ook een mid-life upgrade voor de opvolger van Cartesius worden gefinancierd.

De verwachting is dat de opgebouwde reserveringen in 2025 niet toereikend zullen zijn om een nieuwe nationale supercomputer van een redelijk ambitieniveau te financieren. Mogelijk kunnen internationale HPC samenwerkingen, zoals het ambitieuze EuroHPC, maar mogelijk ook minder grootschalige "lokale" internationale samenwerkingen, kosteneffectieve en strategisch opportune aanvullingen op de Nederlandse HPC ambitie bewerkstelligen. De subcommissie steunt de suggestie van MinOCW om een tripartiet overleg, geleid door MinOCW en verder bestaande uit NWO en SURF, met ondersteuning door een groep bestaande en potentiële HPC gebruikers, te laten vaststellen op welke wijze en in welke mate de HPC reserveringen ten goede moeten komen aan internationale HPC samenwerkingen. Hierin moet oog zijn voor het behoud van de waardevolle expertise die HPC centrum SURFsara huisvest. De uitkomsten van dit overleg zullen leidend zijn in het bepalen van hoe de reserveringen voor toekomstige HPC behoeften zullen worden besteed. De subcommissie adviseert dit overleg zodanig in te richten dat een besluit hierover tijdig kan worden genomen, zodat indien dat wenselijk blijkt nog aan in EuroHPC call kan worden geparticipeerd. Het behoeft nadruk dat een eventuele deelname aan EuroHPC ten koste zal gaan van mogelijkheden om in nationale HPC capaciteit te investeren. De kosten en baten van deze keuze dienen zorgvuldig te worden gewogen.

Expertise

Om het potentieel van alle bovengenoemde hardware optimaal te benutten is het nodig dat SURF de expertise die benodigd is voor het ondersteunen van de gebruikers kan aantrekken en behouden. Hiervoor adviseert de subcommissie vanaf 2019 jaarlijks **2 M€** ter beschikking te stellen. Gebruikersondersteuning is een noodzakelijke voorwaarde voor het ontsluiten en laagdrempelig toegankelijk maken van de digitale infrastructuur. Daarnaast adviseert de subcommissie, zoals eerder genoemd in de paragraaf *Digitale Competentie Centra in een gefedereerd systeem*, dat ondersteuning vanuit SURF aan lokale DCC's wordt bekostigd uit de **2 M€** die volgens de eerder genoemde aanbeveling jaarlijks ter beschikking gesteld zouden moeten worden voor het gefedereerde systeem.

eScience

Om in internationaal verband een toonaangevende rol te kunnen blijven spelen, spelen de kwaliteit, de uitwisselbaarheid en de analyse van de data een centrale rol. We zien dan ook een enorme groei in de vraag naar kwalitatief hoogwaardig beheer en naar ontwikkeling van software tools voor analyse van de data en naar meer en meer diverse reken capaciteit voor de analyse. Ook constateren we dat domein-specifieke kennis van essentieel belang is voor het beheer en analyseren van de data.

Specifiek voor de software is het van essentieel belang om vernieuwing binnen de Informatica en nieuwe ontwikkelingen binnen specifieke domeinen dicht bij elkaar te brengen. De aanpak van de calls van NLeSC, waarin zij hun generieke computing kennis implementeren in specifieke wetenschappelijke vraagstukken in nauwe samenwerking met domein-specifieke partijen, is hiervoor een uitstekend middel. De subcommissie adviseert NWO daarom vanaf 2020 jaarlijks **2,5 M€** uit deze middelen te investeren in het NLeSC, met een impulsfinanciering van **3 M€** in 2019. Deze financiering is bestemd voor het uitbreiden van de open calls van het NLeSC, en zou moeten worden toegekend op voorwaarde van een

positieve uitkomst van de periodieke evaluatie die op het moment van schrijven in opdracht van de Raad van Bestuur van NWO en de directie van SURF wordt uitgevoerd.

De focus van NLeSC ligt op het ontwikkelen van geavanceerde analysesoftware, en gezien die focus en hun beperkte capaciteit ligt het niet voor de hand het lange-termijn beheer van deze tools bij hen te leggen. De eerder genoemde Digitale Competentie Centra dienen de verduurzaming van data-opslag en bijbehorende software voor hun rekening te nemen. Het gefedereerde systeem waar SURF de lead voor neemt, en de stimulering van lokale DCC's zullen dit ondersteunen.

Evaluatie en herijking

Het wetenschappelijke landschap is sterk in beweging. De ambities op het gebied van FAIR zijn groot, maar om deze te realiseren zijn er nog omvangrijke organisatorische hindernissen te nemen. De behoefte aan rekenkracht neemt toe, maar ontwikkelingen in analysemethoden kunnen die behoefte in verschillende richtingen sturen. Internationale computing initiatieven als EuroHPC kunnen een belangrijke rol gaan spelen in het spectrum van rekenfaciliteiten dat tot de beschikking staat van Nederlandse onderzoekers. Vanwege deze dynamiek acht de subcommissie het noodzakelijk om in 2022 het effect van de nieuwe investeringen te evalueren en indien nodig te herijken.

Een aantal van deze investeringen komen ten goede aan SURF, en zijn aanvullend op middelen die al jaarlijks door NWO aan SURF worden verstrekt voor haar innovatieprogramma. NWO is voornemens in het vervolg jaarlijks een begroting en een eindafrekening aan SURF te vragen voor deze 'lump sum'. De subcommissie adviseert NWO om jaarlijks een gecombineerde begroting te vragen voor de innovatieprogramma's van SURF, met daarin de voorgenomen bestedingen uit de huidige structurele financiering van 16 M€, de in dit advies voorgestelde financiering (totaal 12 M€ in 2019, 11 M€ vanaf 2020), en alle incidentele financiering die beschikbaar is voor de innovatieprogramma's. De subcommissie biedt aan NWO te adviseren over deze jaarlijkse begroting.

Samenvattend adviseert de subcommissie NWO de volgende investeringen te doen:

	Impulsfinanciering 2018/2019	Subtotalen 2018/2019	Structurele financiering vanaf 2020	Subtotalen vanaf 2020
Rekenfaciliteiten				
Supercomputer	18 + 2 ¹			
Jaarlijkse reserveringen toekomstige HPC behoeften			4 + 2 ²	
Overige rekenfaciliteiten	3		2,5	
Opslaghardware	1,5		1	
Kennisvernieuwing rekenfaciliteiten	3,5		3,5	
Subtotaal		26		11
Digitalisering				
Impulsfinanciering Digitale Competentie Centra lokaal	4,5			
Stimulering interuniversitaire Digitale Competentie Centra	2,5		2,5	
NLeSC	3		2,5	
Subtotaal		10		5
Expertise en gebruikersondersteuning SURF				
Expertise en ondersteuning rekenfaciliteiten en datacentra SURF	2		2	
Inrichten en onderhouden gefedereerd netwerk, ondersteuning lokale DCC's	2		2	
Subtotaal		4		4
Totaal		40		20

Tabel : Overzicht voorgestelde bestedingen. Genoemde bedragen zijn in M€

¹ Voor de vervanging van supercomputer Cartesius zal SURF een bedrag van 2 M€ bijdragen.

² SURF wordt gevraagd jaarlijks 2 M€ bij te dragen aan de reserveringen voor toekomstige HPC behoeften, te vergaren uit tariefinkomsten van academische gebruikers en commerciële exploitatie van de opvolger van supercomputer Cartesius. De reserveringen wordt door NWO beheerd.