

# **Topwetenschap vereist topinfrastructuur**

**Adviesrapport nationale digitale infrastructuur  
voor wetenschappelijk onderzoek**

**Juni 2017**

## Executive Summary

NWO was asked by the Cabinet to provide advice about the resources needed to ensure that the digital infrastructure meets the needs of all research communities, remains at a world-class level, and is capable of keeping up with developments in computer and data science. The 'ICT Subcommittee' was installed by NWO's Permanent Committee for Large-Scale Scientific Infrastructure to provide background and advice over the requirements for a digital infrastructure that meet the ambitions set out in the 2016 National Roadmap Large-Scale Scientific Infrastructure. (See Chapter 1). The ICT Subcommittee was given three tasks, as follows:

1. Provide an investment plan for the national digital infrastructure, within the current budget, including recommendations for how this could be evaluated.
2. Provide advice on what additional resources are needed to ensure that the Dutch national digital infrastructure is at a comparable level to that of other advanced knowledge economies.
3. Advise the Permanent Committee regarding criteria to be used in the evaluation of proposals for the National Roadmap.

In the Netherlands, as in all advanced knowledge societies, digital infrastructure is essential for all scientific and scholarly research, not only for storing and analysing large volumes of data but also for generating data with simulations, for visualisations, and for engaging in real-time collaboration with researchers around the world. The Netherlands enjoys a leading position in the provision of digital infrastructure for research, and for broader societal and economic use. This fundamental infrastructure on which other research infrastructures depend is often taken for granted by researchers and policy makers, even though it is impossible in the 21<sup>st</sup> century to conduct high-quality research without high-quality and robust digital infrastructure. Given the massive explosion of data in recent years, and the dependence on digital technologies by researchers in all disciplines, the demands on the digital infrastructure and those providing it continue to grow. Other developments on the horizon, such as the European Open Science Cloud and the Internet of Things, will lead to even further infrastructural requirements. (See Chapter 2)

Digital infrastructure includes five elements: Networks, Computing facilities; Data, eScience and Authorisation and authentication (cybersecurity). All of these elements include the support of highly skilled professionals. Three organisations are key to providing digital infrastructure at the Dutch national level: SURF, the Netherlands eScience Center, and DANS. But universities, public research organisations, and discipline-specific collaborations also have their own important facilities, and equipment and software may also be provided by commercial organisations. The annual turnover for SURF alone is approximately 75 million euros, which includes not only the costs for infrastructural innovation (the focus of this report) but also the costs for providing the higher education and the basic research infrastructure.

This report focuses on the resources needed for ensuring that the growing demands of data- and compute-intensive research can be met, that the digital infrastructure of the Netherlands is of an internationally competitive level and therefore contributes to the current high level of scientific research in the Netherlands can at least be maintained, and that Dutch-based researchers continue to have the opportunity to participate in European and global collaborative projects. (See Appendix 4.1)

**Task 1. Provide an investment plan for the national digital infrastructure, within the current budget, including recommendations for how this could be evaluated.**

**Conclusion.** The current resources for innovating the national digital infrastructure (21 million euros per annum) are inadequate for meeting both the growing demands from researchers, and for incorporating the continued developments in digital capabilities. If there is inadequate structural support, then all stakeholders involved in providing infrastructure will continue to react on an ad hoc basis in order to meet users' needs. If this situation continues, it will result in sub-optimal infrastructure provision, leading to lower research quality, increased 'brain drain', fragmentation of provision, and risks to data security and privacy. (See Chapter 3.1) The question of evaluation is thus no longer relevant here, and is addressed in Recommendation 2 below.

**Task 2. Provide advice on what additional resources are needed to ensure that the Dutch national digital infrastructure is at a comparable level to that of other advanced knowledge economies.**

**Recommendation 1.** Additional structural funding of 27 million euros per annum (for a total of 50 million euros for innovation) is needed to meet the growing needs of Dutch researchers, and to ensure that the digital infrastructure remains at a world-class level. The additional funding includes (i) 10 million euros identified in earlier advice from *ICTRegie* (2008), plus (ii) 17 million euros to meet the growing demands and ambitions of the Dutch research community, as set out in the National Science Agenda, the National Plan for Open Science (OCW, 2017), and the National Roadmap (NWO, 2016). To date, (i) has been financed on an ad hoc basis, but it must become part of the ongoing, structural funding. (See Chapter 3.2)

**Recommendation 2.** Draw upon the expertise of the national organisations, SURF, the Netherlands eScience Center and DANS, to ensure optimum use is made of both the digital infrastructure and additional investments. Coordinate the existing evaluations of these national organisations to ensure that review of infrastructural provision is adequately addressed, and to ensure that the additional structural funding is invested in such a way as to maintain the integrity and quality of the national digital research infrastructure. (See Chapter 3.2)

**Task 3. Advise the Permanent Committee regarding criteria to be used in the evaluation of proposals for the National Roadmap.**

**Recommendation 3.** Researchers must demonstrate awareness of the material and operational costs of digital infrastructures, and where appropriate include additional costs in their proposals for large-scale infrastructure and research projects. Both domain-specific and, where possible, cross-domain collaboration should be encouraged, not only to reduce costs but also to stimulate the emergence of standards and the joint development of software and tools, and to facilitate learning from shared experience. (See Chapter 3.3)

**Additional recommendations:**

- Strengthen dialogue and coordination between all parties providing and using the national digital infrastructure. (Chapter 2.3)
- Increase the synergy between fundamental computer and data science, and digital infrastructure. (Chapter 3.4)
- Stimulate local support from ICT departments and libraries to assist researchers in understanding their digital needs and options. (Chapter 3.4)
- Contribute to greater digital literacy of all researchers, and stimulate the educational and career opportunities for data scientists. (Chapter 3.4)

# Inhoudsopgave

<b>1</b>	<b>INLEIDING .....</b>	<b>1</b>
<b>2</b>	<b>WETENSCHAP EN DIGITALE INFRASTRUCTUUR .....</b>	<b>3</b>
2.1	TRENDS IN DE WETENSCHAP .....	3
2.2	INFRASTRUCTUUR ALS RANDVOORWAARDE .....	5
2.3	DE NATIONALE DIGITALE INFRASTRUCTUUR .....	7
<b>3</b>	<b>ADVIES.....</b>	<b>9</b>
3.1	INVESTERINGSPLAN BINNEN DE HUIDIGE MIDDELEN .....	9
3.2	DIGITALE INFRASTRUCTUUR IN INTERNATIONALE KOPGROEP .....	18
3.3	AANBEVELINGEN VOOR DE NATIONALE ROADMAP .....	22
3.4	AANBEVELINGEN VOOR FLANKEREND BELEID .....	23
<b>4</b>	<b>BIJLAGEN .....</b>	<b>24</b>
4.1	OVERZICHT NATIONALE DIGITALE INFRASTRUCTUUR.....	24
4.2	NATIONALE ENQUÊTE DIGITALE INFRASTRUCTUUR .....	32
4.3	GROEI WETENSCHAPPELIJKE DATA.....	33
4.4	E-NEEDS QUESTIONS IN ESFRI SUBMISSION FORMAT 2018 .....	34
4.5	VERKLARENDE WOORDENLIJST .....	36
4.6	ACHTERGRONDLITERATUUR .....	38
4.7	SAMENSTELLING VAN DE ADVIESCOMMISSIE ICT-INFRASTRUCTUREN .....	39

# 1 Inleiding

De digitale infrastructuur is een bijzonder element van de wetenschap: vrijwel onzichtbaar zolang ze functioneert, maar duidelijk ontwrichtend wanneer ze faalt. Het vertrouwen erin is zo groot, dat er vaak pas over wordt nagedacht zodra de infrastructuur *niet* werkt. Door de exponentiële toename in gebruik en afhankelijkheid van onderzoekssoftware, supersnelle netwerken en (big) data in alle wetenschappelijke disciplines is de druk op de huidige digitale infrastructuur sterk en snel toegenomen. Bovendien stellen ontwikkelingen rondom cloud computing, open science en open data, security en privacy steeds hogere eisen aan de nationale digitale infrastructuur. Blijvend investeren in een betrouwbare, duurzame en breed toegankelijke hoogwaardige infrastructuur en support zijn dan ook noodzaak. Zonder een krachtige infrastructuur van netwerken, rekenfaciliteiten, data-diensten, onderzoekssoftware en ondersteuningsmechanismen, zullen de wetenschappers van nu en de toekomst hun werk nauwelijks meer kunnen uitvoeren. Het recente Deloitte rapport *Dutch Digital Infrastructure 2016: enabling the digital economy and society* laat bovendien zien hoe de interactie tussen een excellente digitale infrastructuur en een brede waaier aan bedrijvigheid de afgelopen jaren tot bloei is gekomen. Het achterblijven van investeringen in de infrastructuur heeft dan ook schadelijke gevolgen voor de Nederlandse economie en samenleving. De landen om ons heen staan niet stil, waardoor we talentvolle onderzoekers dreigen te verliezen en bovendien niet meer aan zullen trekken.

Tegen de achtergrond van deze snelle en disruptieve digitalisering en toenemende competitiviteit, adviseerde de AWTI het kabinet in 2015 al in het rapport 'Klaar voor de Toekomst' om ten behoeve van toponderzoek in Nederland meer te investeren in een digitale infrastructuur van wereldklasse. Het kabinet vroeg NWO hierop in kaart te brengen welke behoeften wetenschappers hebben aan digitale infrastructuur. Daarnaast zou bezien moeten worden of voldoende middelen beschikbaar zijn om tot de internationale top te blijven behoren. De ICT-commissie (zie bijlage 4.7 voor samenstelling) werd hierop ingesteld door de Permanente Commissie voor Grootschalige Wetenschappelijke Infrastructuur (PC-GWI). De ICT-commissie werd gevraagd 'een advies te geven over de optimale inzet van de beschikbare middelen voor ICT infrastructuur' (in het vervolg 'digitale infrastructuur' genoemd). Hierbij werd speciale aandacht gevraagd om een goede afstemming tussen de investeringen door SURF in de digitale infrastructuur en die van NWO in grootschalige onderzoeksfaciliteiten.

De PC-GWI gaf de ICT-commissie de volgende drie opdrachten:

1. Het opstellen van een duurzaam investeringsplan voor de wetenschappelijke ICT-infrastructuur binnen de huidige beschikbare middelen voor de uitvoering van wetenschappelijk onderzoek voor de komende vijf jaar inclusief een evaluatie-mechanisme.
2. Het geven van een advies over de aanvullingen op de eerste taak die nodig zijn om de Nederlandse ICT-infrastructuur op een vergelijkbaar niveau te brengen/houden in vergelijking met belangrijke andere kenniseconomieën.
3. Het adviseren van de permanente commissie over de formulering van het criterium voor ICT bij de beoordeling van aanvragen voor de Nationale Roadmap.

De zoektocht van individuele wetenschappers, de toenemende digitalisering van de maatschappij en economie – ze vereisen digitale infrastructuren die gepaard gaan met investeringen die de financiële draagkracht van kennisinstellingen of van de individuele gebruikers ver te boven gaat. Iedere Nederlandse onderzoeker heeft tegenwoordig directe toegang nodig tot een breed pakket van betrouwbare diensten om te communiceren en samen te werken, data te verwerken, ze te analyseren, op te slaan en te delen. Een nationale infrastructuur draagt hieraan bij door het ter beschikking stellen van een geïntegreerd aanbod van faciliteiten, dat aangeboden kan worden tegen lagere kosten als gevolg van samenwerking en schaalvergroting. Door vraagbundeling via de nationale digitale infrastructuren kunnen in Nederland bepaalde geavanceerde diensten efficiënt en tijdig worden gefinancierd. Een overheidsbijdrage is noodzakelijk om het solidariteitsmodel in stand te houden: grote investeringen zijn nodig voor een relatief kleine groep high-end users. Pas later volgt de adoptie van nieuwe technologie door de verschillende wetenschappelijke disciplines. De relatief kleine groep koplopers kan de kosten van nieuwe en kostbare infrastructuur niet individueel opbrengen, terwijl wachten tot iedereen bereid is te investeren Nederland binnen korte tijd op een flinke achterstand zal zetten.

Op 13 december 2016 presenteerde de PC-GWI de Nationale Roadmap Grootschalige Wetenschappelijke Infrastructuur. De Roadmap beschrijft een strategisch kader voor grootschalige wetenschappelijke infrastructuur voor de komende vijf jaar. Digitale infrastructuur speelt een steeds belangrijkere rol in het wetenschappelijk onderzoek. In veel gevallen vormt dit het fundament voor vooraanstaand onderzoek. Het is van belang dat Nederlandse onderzoekers toegang hebben tot geavanceerde digitale infrastructuur. Dit vraagt om vervolginvesteringen in een nationale digitale infrastructuur die ook internationaal gezien van hoog niveau is. Hiermee wordt ook tegemoet gekomen aan de ambities van het Nationaal plan voor Open Science (OCW, 2107) en de , de Nationale Wetenschapsagenda, topsectoren en Nationale Roadmap waar te maken, In Nederland zijn SURF, DANS en het Netherlands eScience Center de organisaties met een nationale opdracht op het gebied van digitale infrastructuur naast meer lokale initiatieven zoals het 4TU datacentrum. SURF heeft een coördinerende rol voor de nationale digitale infrastructuur. Lokale- en domein-specifieke faciliteiten zijn doorgaans lokaal of project gebaseerd gefinancierd en hierdoor meestal niet landelijk beschikbaar of niet geschikt voor meerdere wetenschappelijke domeinen. Het is van belang om in de toekomst de afstemming en samenwerking tussen deze initiatieven en de nationale initiatieven te intensiveren.

## 2 Wetenschap en digitale infrastructuur

### 2.1 Trends in de wetenschap

Sinds enkele jaren is er sprake van een exponentiële toename van de data die geproduceerd worden ten behoeve van onderzoek in vrijwel alle wetenschappelijke disciplines. Deze toename wordt veroorzaakt door de mogelijkheden die nieuwe technologieën bieden, zoals die binnen grootschalige onderzoeksfaciliteiten, geavanceerde software, krachtigere rekenfaciliteiten en sensornetwerken. Daarnaast komt er steeds meer data beschikbaar voor onderzoek die niet direct gerelateerd zijn aan grootschalige onderzoeksfaciliteiten. Hierbij kan bijvoorbeeld gedacht worden aan de grote hoeveelheden tekst, foto en video van sociale media en online nieuwssites die worden gebruikt door politicologen en mediawetenschappers. In zorgonderzoek wordt in toenemende mate data gegenereerd door draagbare consumententechnologie, zoals sporthorloges en smartphones.

De meeste grootschalige onderzoeksfaciliteiten op de Nationale Roadmap Grootschalige Wetenschappelijke Infrastructuur (2016) hebben een sterke ICT- en datacomponent. De faciliteiten en samenwerkingsverbanden tussen wetenschappers zoeken de grenzen op van de technische mogelijkheden en kunnen alleen uitgevoerd worden als er goede digitale infrastructuren beschikbaar zijn. In feite illustreren dit type projecten de aankomende ontwikkelingen in bijna alle domeinen, waarin onmiskenbaar eenzelfde ontwikkeling van explosieve datagroei plaatsvindt. Ook de verandering naar meer (internationaal) samenwerken en de wijze van werken van nieuwe generaties onderzoekers vragen om infrastructurele diensten voor alle onderzoekers.

#### *Afhankelijk van digitale infrastructuur*

MinE is een internationaal project om de genetische oorzaken op te sporen van ALS, een dodelijke spierzenuwziekte. Voor dit onderzoek is materiaal nodig van minstens 15.000 patiënten en 7.500 controlepersonen. Na het sequencen (het uitlezen) van 1 DNA-sample heb je een bestand van 75 à 100 gigabyte. Vermenigvuldig dat met 22.500 en je komt op zo'n 2 miljoen gigabyte. Dat komt neer op een stapel harde schijven zo hoog als de Utrechtse Domtoren.

Natuurkundigen over de hele wereld werken samen binnen CERN. Ze bestuderen met behulp van grote detectoren tientallen miljoenen botsingen tussen protonen - per seconde - en de deeltjes die daaruit ontstaan. Een wereldwijd netwerk van computers (grid) filtert, registreert en verwerkt de data uit deze botsingen. Het gaat om miljoenen gigabytes aan data per jaar.

Astronomen werken op het ogenblik de bouwplannen uit voor de Square Kilometer Array (SKA). Nu al is duidelijk dat het deze grootste telescoop ter wereld ook leidt tot enorme data uitdagingen. Stel je de totale hoeveelheid data voor op het internet, over hele wereld. Die hoeveelheid aan data, een miljard gigabyte per dag, komt continu binnen zodra SKA in werking treedt.



Dat de behoefte aan digitale infrastructuur steeds sterker over de hele breedte van de wetenschap wordt gevoeld, blijkt onder andere uit het ePLAN-onderzoek Infrastructuur Duurzaam op Maat. Dit rapport, gebaseerd op meerdere workshops en een vragenlijst onder meer dan 1000 wetenschappers uit alle wetenschappelijke disciplines, toont aan dat grote zorgen bestaan over het achterblijvende investeringsniveau in de Nederlandse digitale infrastructuur. Wetenschappers verwachten met name dat flink meer behoefte zal ontstaan op het gebied van *data & cloud services*. Ook verwachten wetenschappers meer ondersteuning en support nodig te hebben bij hun dagelijkse data-intensieve onderzoek.

De sterke afhankelijkheid tussen wetenschappelijke ontwikkelingen en digitale infrastructuur heeft de Europese Commissie ertoe bewogen om een extra impuls te geven aan dit ecosysteem. Met het plan voor een European Open Science Cloud wil de commissie uiteindelijk een onderzoeksinfrastructuur van FAIR (Findable, Accessible, Interoperable, Reusable) data en diensten realiseren. Een belangrijke randvoorwaarde hiervoor is een effectieve dialoog tussen wetenschappelijke domeinen en aanbieders van digitale infrastructuur, diensten en eScience.

**Open Science** is een benadering van het wetenschappelijke proces gebaseerd op intensievere samenwerking en kennisdeling door het gebruik van digitale technologieën en nieuwe samenwerkingstools. Onderzoeksdata, methoden en publicaties worden zo transparant, herbruikbaar en toegankelijker voor wetenschappers, bedrijfsleven en maatschappij.

Het ministerie van Onderwijs, Cultuur en Wetenschap presenteerde in februari 2017 de ambities van Nederland in het Nationaal Plan Open Science. Dit plan concludeert dat in alle fasen van het onderzoeksproces ondersteuning nodig is, van infrastructurele diensten tot training, van hands-on support tot de ontwikkeling van onderzoekssoftware. Ook steunt het de ambitie voor een consistent systeem voor het FAIR ontsluiten van onderzoeksdata. Om deze ontwikkelingen te ondersteunen richt het initiatief GO-FAIR zich op de ontwikkeling van sterke nationale of domeinspecifieke *open science cloud nodes*, die een cultuuromslag nastreven door middel van training en de implementatie van FAIR-data en -diensten. Om data-intensief onderzoek mogelijk te maken en initiatieven zoals de European Open Science Cloud, Nationaal Plan Open Science, GO-FAIR te laten slagen, is een forse uitbreiding en vernieuwing van de huidige infrastructuur noodzakelijk. Bij deze en toekomstige grote wetenschappelijke projecten worden de technologische uitdagingen van wetenschappers alleen maar complexer en uitgebreider. Daarnaast zal een groeiend aantal individuele wetenschappers uit een groot aantal wetenschappelijke disciplines meer en meer behoefte hebben aan hoogwaardige en hands-on ondersteuning bij data-analyse en toepassing van software.

## Cloud computing

De ontwikkeling van cloud computing heeft geleid tot veel nieuwe digitale diensten door commerciële aanbieders, maar heeft de behoefte aan nationale digitale infrastructuren en diensten niet weggenomen. Onderzoekers zoeken immers de grenzen op van de technische mogelijkheden, waarvoor dienstverlening, ondersteuning en expertise nodig is die door de markt (nog) niet geboden wordt of niet onder de juiste voorwaarden.

Dit geldt bijvoorbeeld voor de high-end vraag naar High performance computing (HPC), waar wetenschappers een grotere behoefte aan capaciteit hebben dan de markt biedt. HPC cloud computing is een ontwikkeling die in zekere zin nog in de kinderschoenen staat. Kijkend naar de TOP500, dan bestaat de wereldtop uit nationale supercomputers, supercomputers van meteorologische en andere (onderzoeks)instituten en soms van overheden. De enkele bedrijven die hoog in TOP500 staan, zijn geen aanbieders van HPC cloud computing, maar gebruikers die voor hun bedrijfsprocessen een eigen supercomputer nodig hebben (Total, Petroleum Geo-services).

Er zijn verschillende vormen van cloud: publieke cloud, private cloud, community cloud en hybride cloud. Waar de generieke (voor alle afnemers gelijk) functionaliteit van de publieke cloud voldoet, sluit SURF namens de leden contracten af met commerciële aanbieders. Naast de prijs voor de diensten zijn afspraken over privacy, beveiliging, gebruikersauthenticatie en intellectueel eigendom daarbij minstens zo belangrijk. SURF deelt deze ervaringen op Europees niveau, omdat schaalvergroting helpt om tot een beter aanbod en een betere dialoog met cloudleveranciers te komen. Dit heeft recent geleid tot deals met Amazon en Microsoft.

Wanneer privacy en vertrouwen erg belangrijk zijn, kan men kiezen voor een private cloud. Met een private cloud heeft de gebruiker volledige controle over data, beveiliging en kwaliteit van de dienst. De toepassingen worden beschikbaar gesteld binnen de eigen (virtuele) organisatie. Zo hebben overheden, bedrijven, instellingen of (Europese) grootschalige onderzoeksinfrastructuren hun eigen private cloud die lokaal kan staan of bij een cloudleverancier.

De community cloud combineert de vertrouwelijkheid van een private cloud met de schaalvoordelen van een publieke cloud. Gebruikers uit meerdere organisaties werken op dezelfde infrastructuur. Dit kan alleen als ze elkaar voldoende vertrouwen en vergelijkbare eisen stellen. SURF ontwikkelt op verzoek van de leden verschillende community clouddiensten voor onderzoek en onderwijs, die de specifieke behoefte van gebruikers bedient. Daarbij biedt SURF garanties over lokaliteit van de data, over vertrouwen en privacy. Dit is belangrijk bij persoonsgegevens of vertrouwelijke medische onderzoeksdata en als men zeker wil zijn dat de data in eigen land worden opgeslagen.

Relatief nieuw is de hybride cloud waarbij meerdere vormen van cloud samen worden gebruikt. SURF heeft hiervoor samen met enkele early adopters een nieuwe dienst ontwikkeld, waarbij SURF namens de instellingen optreedt als broker en regisseur van de dienstverlening. Instellingen hoeven daardoor alleen nog het gebruik te monitoren en te managen, waardoor de ICT-afdelingen meer tijd hebben om zich te concentreren op hun primaire taken: het ondersteunen van onderwijs en onderzoek door middel van ICT.

## 2.2 Infrastructuur als randvoorwaarde

Een breed toegankelijke hoogwaardige nationale digitale infrastructuur is een randvoorwaarde geworden om competitieve wetenschap op een efficiënte en effectieve manier te faciliteren. De urgentie ervan werd onlangs nog benadrukt door de AWTI. In het rapport *Houd de basis gezond* wordt de ICT-infrastructuur benoemd als één van de 5 prioriteiten voor extra investeringen in onderzoek en innovatie. De AWTI waarschuwt dat Nederland op concurrentiekracht en verdienvermogen achterop dreigt te raken of zelfs onder het internationale gemiddelde dreigt te belanden.

Wat betreft de rol van de digitale infrastructuur, stelt het:

“De ICT-onderzoek infrastructuur vormt de ruggengraat van een kennis- en innovatiesysteem van de 21ste eeuw. Vrijwel alle wetenschappelijke en technologische ontwikkelingen die ertoe doen, zijn afhankelijk van ICT. Communicatie en samenwerking, kennisontwikkeling (rekenkracht, geavanceerde algoritmen en big data) en wetenschappelijke vooruitgang zijn meer dan ooit afhankelijk van een kwalitatief hoogwaardige ICT-infrastructuur.”

Door extra te investeren kan noodzakelijke capaciteitsuitbreiding worden gerealiseerd, de groeiende hoeveelheid onderzoeksdata goed worden geëxploiteerd en de noodzakelijke hard- en software en support worden geleverd om die gereedschappen te gebruiken in onderzoek.

Dat thema's als digitalisering, big data en ICT een steeds centralere en grotere rol hebben gekregen, blijkt ook uit verschillende andere recente publicaties. Zo vragen onder andere VSNU (*Digitale Samenleving*), VNO-NCW (*Investeren in een digitale kwantumsprong*) en Nederland-ICT (*Groei door Digitalisering*) om meer inzet op deze thema's. Digitalisering is de doorsnijdende ontwikkeling die elke sector vooruit kan helpen en hoogwaardige digitale infrastructuren zijn hiervoor een onmisbare basis. In een gezamenlijke verklaring benadrukten de G7 ICT-ministers in het voorjaar 2016 nog eens het belang van de digitale infrastructuren voor onderwijs en onderzoek. Ook de Nationale Wetenschapsagenda (NWA), en in het bijzonder de Big Data Route, laat zien dat het belang van (big) data en daarmee het belang van digitale infrastructuren en specialistische ondersteuning groeit.

De meeste onderzoeksfaciliteiten op de Nationale Roadmap van NWO en die van de KNAW voor de toekomst vereisen hoogwaardige technologische vaardigheden en support. De faciliteiten zijn vaak ingebed in Europese samenwerkingsverbanden zoals ESFRI. Deze internationale samenwerking kan alleen gestalte krijgen door beschikbaarheid van digitale infrastructuur.

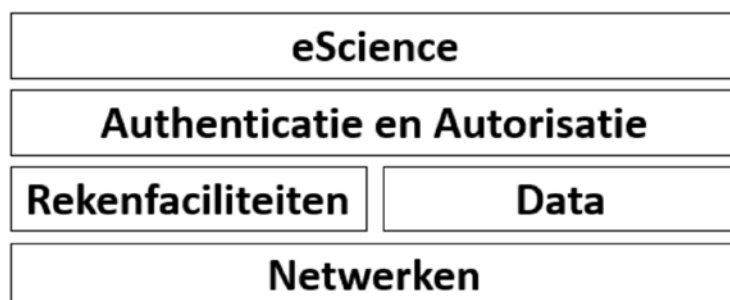
De eerdergenoemde European Open Science Cloud, en de Nederlandse aansluiting hierop, vereisen goede coördinatie en een versterking van de digitale infrastructuur. Hoewel de term Open Science Cloud nog volop in ontwikkeling is, is een aantal zaken duidelijk waar het gaat om de behoefte van wetenschappers: groot gebruiksgemak, integratie, single sign-on, reproduceerbaarheid van analyses en methodes, transparantie van datalocaties, training, passende diensten en experts die onderzoekers kunnen bijstaan. Het succes ervan zal daarom bepaald worden door de robuustheid van het systeem, de drempels die onderzoekers ervaren en natuurlijk het volume en de kwaliteit van de resources die van de cloud deel uit maken. Het realiseren van de visie van de Europese Commissie, maar ook die van Nederland op nationaal niveau, vereist hiermee verschillende uitbreidingen en verbeteringen van de bestaande digitale infrastructuur, inclusief afstemming met gebruikers. De Europese Commissie heeft een indicatie afgegeven hiervoor nu al investeringen van 6,7 miljard Euro, waarvan het meeste geld uit de lidstaten moet komen. Zo moet de European Open Science Cloud ondersteund worden door een European Data Infrastructure van gefedereerde nationale netwerken en reken- en opslagfaciliteiten.

Nederland heeft een goede uitgangspositie om een actieve rol te spelen in deze internationale ontwikkelingen: de publieke digitale infrastructuren zijn goed georganiseerd met nauwe betrokkenheid van de gebruikers. Voor veel landen heeft Nederland daarmee een voorbeeldfunctie. Niettemin blijft het noodzakelijk om ook hier een majeure innovatieslag en coördinatieslag te maken om tegemoet te komen aan de groeiende behoefte aan dienstverlening van wetenschappers.

### 2.3 De nationale digitale infrastructuur

De nationale digitale infrastructuur bestaat uit de faciliteiten en diensten die in principe beschikbaar zijn voor alle wetenschappers, ongeacht het wetenschappelijke domein. Ze vormen daarmee onder andere het fundament voor bovengenoemde onderzoeksactiviteiten en ook voor de uitvoering van de Nationale Wetenschapsagenda (NWA) en de Roadmap. Dit fundament moet ruim voldoende zijn om de wetenschap in de volle breedte te kunnen ondersteunen. Voor nationale digitale infrastructuren is geen financiering vanuit Europa mogelijk. Elk land moet zorgdragen voor zijn eigen nationale netwerken, rekenfaciliteiten, data-opslag en ondersteuning. Voor deelname in internationale infrastructuurprojecten moeten bovendien eigen faciliteiten en financiële middelen worden ingebracht, zoals netwerkverbindingen of een nationale supercomputer.

De commissie heeft gekeken naar digitale infrastructuren en diensten die een nationale en/of internationale rol spelen. In onderstaande figuur worden de vijf verschillende onderdelen van de digitale infrastructuur weergegeven. De organisaties die op nationaal niveau niet-domeinspecifieke digitale infrastructuren bieden zijn SURF, DANS en Netherlands eScience Center. Terwijl SURF actief is binnen alle onderdelen van de infrastructuur, richt het eScience Center zich hoofdzakelijk op onderzoekssoftware en hoogwaardige onderzoekssupport, terwijl DANS diensten biedt op het vlak van data-opslag en -archivering. Een uitgebreidere beschrijving van de verschillende organisaties is te vinden in Bijlage 4.1. Verder bieden RUG-CIT en Nikhef onder regie van SURF diensten op nationaal niveau aan via een gefedereerd model. Zo stemmen de RUG-CIT, Nikhef en SURF gezamenlijk de investeringen, inzet en beschikbaarheid van grid- en clouddiensten af ten behoeve van hun wetenschappelijke gebruikersgemeenschappen. Meer inhoudelijke informatie over de digitale infrastructuur volgt in hoofdstuk drie.



*Figuur 1: Onderdelen digitale infrastructuur.*

Naast de nationale faciliteiten zijn er veel lokale regionale en domeinspecifieke infrastructuur en diensten, zoals campusnetwerken, datacenters (bijv. 4TU datacentrum) en research support-afdelingen, of domeinspecifieke infrastructuur. In dit advies worden ze niet meegenomen, omdat ze met name in een lokale of domeinspecifieke behoefte voorzien of lokaal gefinancierd worden. Dit geldt bijvoorbeeld ook voor KNMI, MARIN en NLR, die weliswaar aanzienlijke rekenfaciliteiten in huis hebben, maar niet nationaal diensten verlenen aan de wetenschap in den brede. Deze faciliteiten worden vaak ook met publieke middelen gefinancierd. In de adviezen is daarom wel aandacht voor deze faciliteiten daar waar het regievoering om de optimale inzet van deze middelen te waarborgen.

Om dezelfde reden blijven de data-infrastructuren van Roadmap-projecten als BBMRI, ELIXIR-NL en ODISSEI in dit advies buiten beschouwing: deze infrastructuren zijn domeinspecifiek georganiseerd en dienen als zodanig niet als fundament voor alle wetenschapsgebieden. Tegelijk wordt in de call-for-proposals voor de Roadmap aangestuurd op een effectief gebruik van bestaande nationale digitale infrastructuren.

Dit advies betoogt een intensieve dialoog tussen de aanstaande onderzoeksfaciliteiten en de nationale digitale infrastructuur, om daarmee het effectief gebruik van de laatste te vergroten. Analoog aan de Europese situatie in ESFRI beoogt het advies dat de aanstaande onderzoeksfaciliteiten op de Nationale Roadmap het benodigde gebruik van digitale faciliteiten beschrijft én begroot, en aangeeft hoe een en ander gerealiseerd wordt in de ontwikkeling en uitvoering van de faciliteit. Hierbij is het belangrijk dat een balans gevonden wordt tussen lokale support en domeinspecifiek maatwerk en de efficiëntie en effectiviteit van (inter)nationale faciliteiten.

#### *Organisatie van de nationale digitale infrastructuren in Nederland*

De nationale digitale infrastructuur is voor een groot deel georganiseerd binnen de coöperatie SURF, waarin universiteiten, hogescholen, universitaire medische centra, onderzoeksinstituten en steeds meer mbo-instellingen op bestuurlijk, beleidsmatig en operationeel niveau samenwerken aan ICT-innovatie en digitale infrastructuur. Het kabinet omarmde in 2009 het advies van ICTRegie van een jaar eerder om alle digitale infrastructuur voor wetenschappelijk onderzoek onder regie van SURF te brengen en zo de directe verantwoordelijkheid voor strategie en innovatie bij de gebruikers neer te leggen. Dit heeft geleid tot een fusie van SARA en SURF en het opgaan van de activiteiten van NCF in SURF.

Het Netherlands eScience Center is in 2011 opgericht door SURF en NWO. Het slaat de brug tussen de ambities van wetenschappers en de steeds complexere digitale infrastructuur en ICT, inclusief kennis en kunde uit computer en data science. Het eScience Center ondersteunt onderzoekers door onder andere het ontwikkelen van hoogwaardige onderzoekssoftware, in nauwe samenwerking met de betrokken onderzoeksgroepen.

DANS is een initiatief van KNAW en NWO. Het is het nationaal data-archief en kenniscentrum voor toegang tot onderzoeksdata en wetenschappelijke informatie. DANS bevordert duurzame toegang tot digitale onderzoeksgegevens en stimuleert onderzoekers om gegevens duurzaam te archiveren en te hergebruiken. Ook biedt DANS toegang tot duizenden wetenschappelijke datasets, publicaties en andere onderzoeksinformatie in Nederland en doet onderzoek naar duurzame toegang tot digitale informatie en duurzaamheid van software.

## 3 Advies

### 3.1 Investeringsplan binnen de huidige middelen

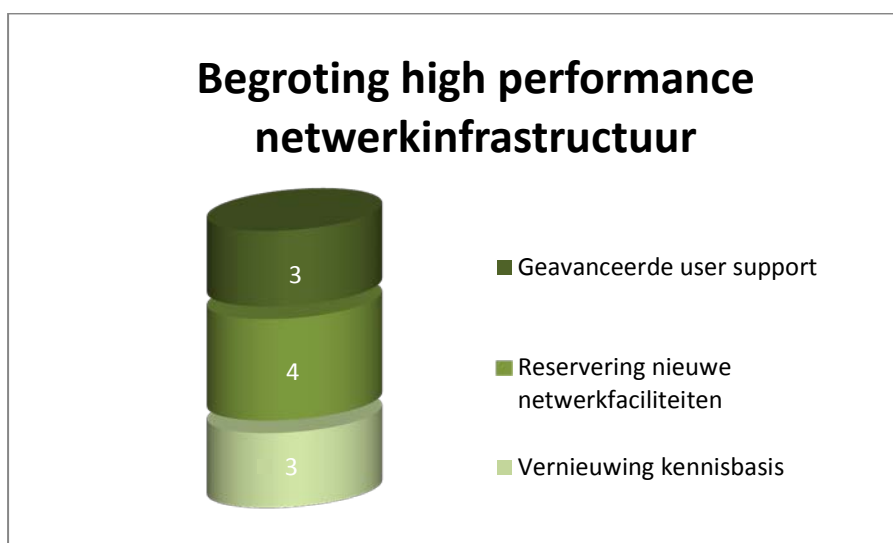
Het opstellen van een duurzaam investeringsplan voor de wetenschappelijke digitale infrastructuur binnen de huidige beschikbare middelen voor de uitvoering van wetenschappelijk onderzoek voor de komende vijf jaar inclusief een evaluatie-mechanisme.

#### **Beschrijving huidige situatie**

De huidige structurele financiering van 21 miljoen Euro voor de nationale digitale infrastructuur gaat naar SURF en het Nederlands eScience Center. Daarnaast is er een separate financiering vanuit NWO en KNAW voor DANS. In Figuur 1 zijn de onderdelen van de digitale infrastructuur geschetst waar deze organisaties verantwoordelijk voor zijn.

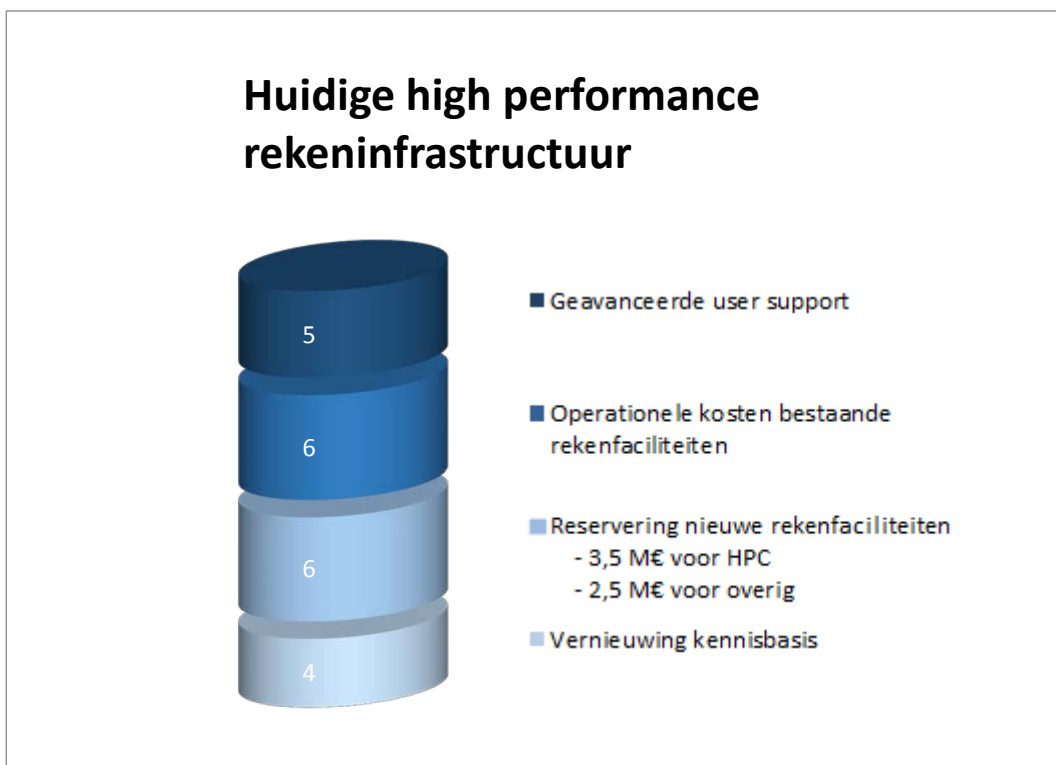
De twee meest kapitaalintensieve onderdelen van de digitale infrastructuur zijn het netwerk en de (*high performance*) rekeninfrastructuur. Uitgangspunt bij investeringskeuzes en de doorontwikkeling van diensten is dat SURF op basis van de ontwikkelingen in de markt een afweging maakt welke kennis, producten en diensten zelf worden ontwikkeld en welke aanbesteed zouden moeten worden in de markt.

- De benodigde investeringen in een het netwerk zijn – gemiddeld - minimaal 10 Meuro per jaar (Figuur 2). Dat leert de langjarige ervaring van SURF met geavanceerde netwerken. Daarin is inbegrepen de vernieuwing van de kennisbasis, het beproeven van nieuwe concepten, het uitvoeren van pilots met geavanceerde gebruikers en de reservering voor de aankoop – de kosten hiervan zijn geschat op ongeveer 6 Meuro (van de 10) per jaar. De migratie, operationele kosten en ad-hoc uitbreidingen worden bekostigd op basis van tariefheffing door de aangesloten onderwijs- en onderzoeksinstituten.



Figuur 2: Begroting high performance netwerkinfrastructuur (bedragen in M€)

- De benodigde investeringen in de huidige *high performance* rekeninfrastructuur zijn – inclusief de gevraagde gebruikersondersteuning - circa 18 Meuro per jaar (zie ook paragraaf 4.1.2 voor een overzicht van de verschillende reken-faciliteiten en hun type gebruikers). Bij dat bedrag inbegrepen is de vernieuwing van de kennisbasis, het beproeven van nieuwe concepten (zoals *deep learning*), het uitvoeren van pilots, de reservering voor de aankoop van een supercomputer vergelijkbaar met de huidige en de overige rekenfaciliteiten (zie paragraaf 4.1.2), en de operationele kosten om het systeem up-and-running te houden (incl. energie en huisvesting). De praktijk is dat er zeer veel behoefte is aan geavanceerde en community support, Deze groeiende behoefte aan support bij met name de reken- en datafaciliteiten zorgt ervoor dat het oorspronkelijk begrootte budget voor support met 4 Meuro per jaar wordt overschreven. Voorbeelden van intensieve gebruikersondersteuning zijn LOFAR, TROPOMI, MinE, BMMRI en Virgo-LIGO ondersteund, waarbij het gros van het benodigde budget ten laste van SURF komt. Maar ook support voor minder grootschalige gebruikers die toch HPC-faciliteiten nodig hebben is sterk groeiend en meer dan destijds begroot. Hierdoor is SURF nu niet in staat om voldoende te reserveren voor een nieuwe supercomputer. Figuur 3 geeft een overzicht van de huidige kosten voor high performance computing.

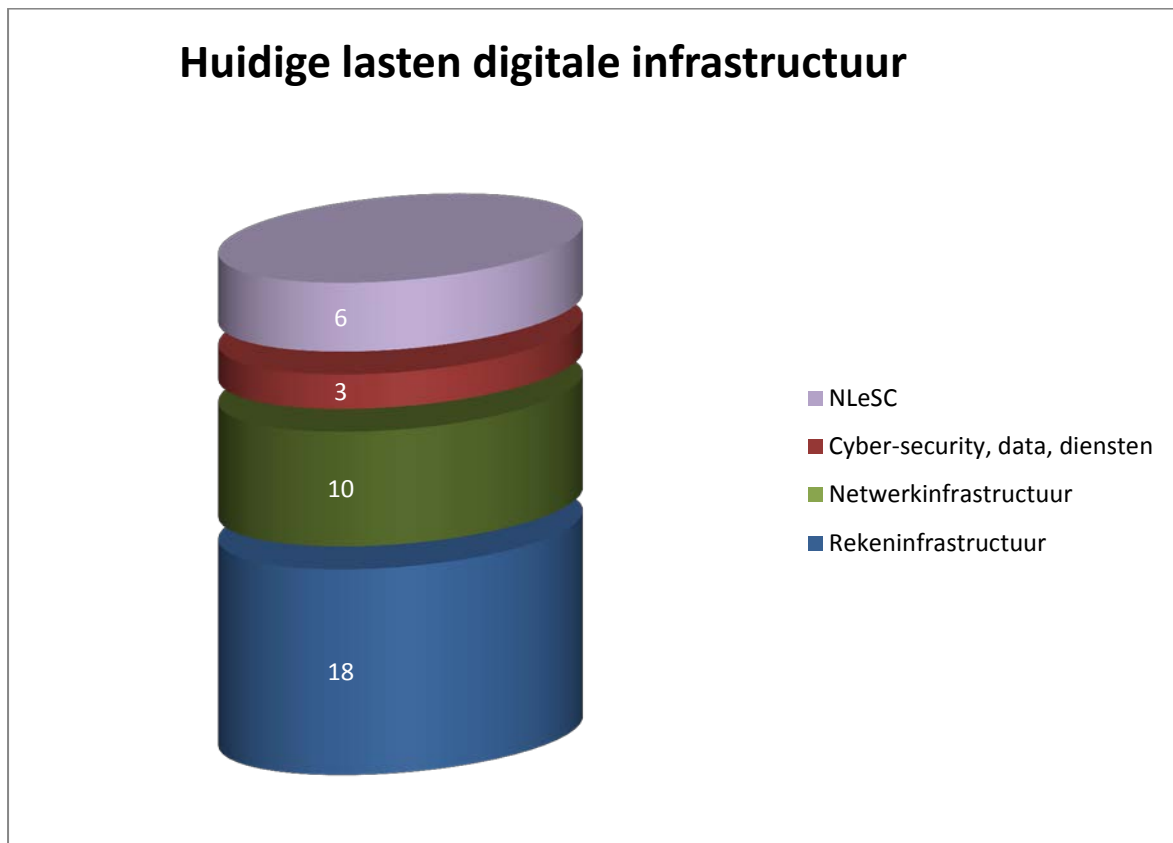


Figuur 3: Huidige high performance rekeninfrastructuur (Bedragen in Meuro)

Naast deze twee faciliteiten wordt geïnvesteerd in kennis en knowhow op het gebied van cybersecurity, inclusief veilige en gebruiksvriendelijke manieren om toegang te krijgen tot dienstverlening. Het belang van een veilige (leer- en) werkomgeving, waar op een veilige (*two-factor authentication*) maar ook laagdrempelige (*single sign-on*) wijze toegang tot wordt verkregen, is vergroot. De complexiteit van de *virtual research environment* neemt alleen maar toe, en daarmee het belang van een hoog-niveau AAI (authenticatie- en autorisatie-infrastructuur – zie ook paragraaf 4.1.4).

De afgelopen jaren zijn geavanceerde data-opslag, *-curation* en *-analytics* daar ook bij gekomen. De investeringen zitten nu rond de 3 Meuro, maar de behoefte is duidelijk groeiend.

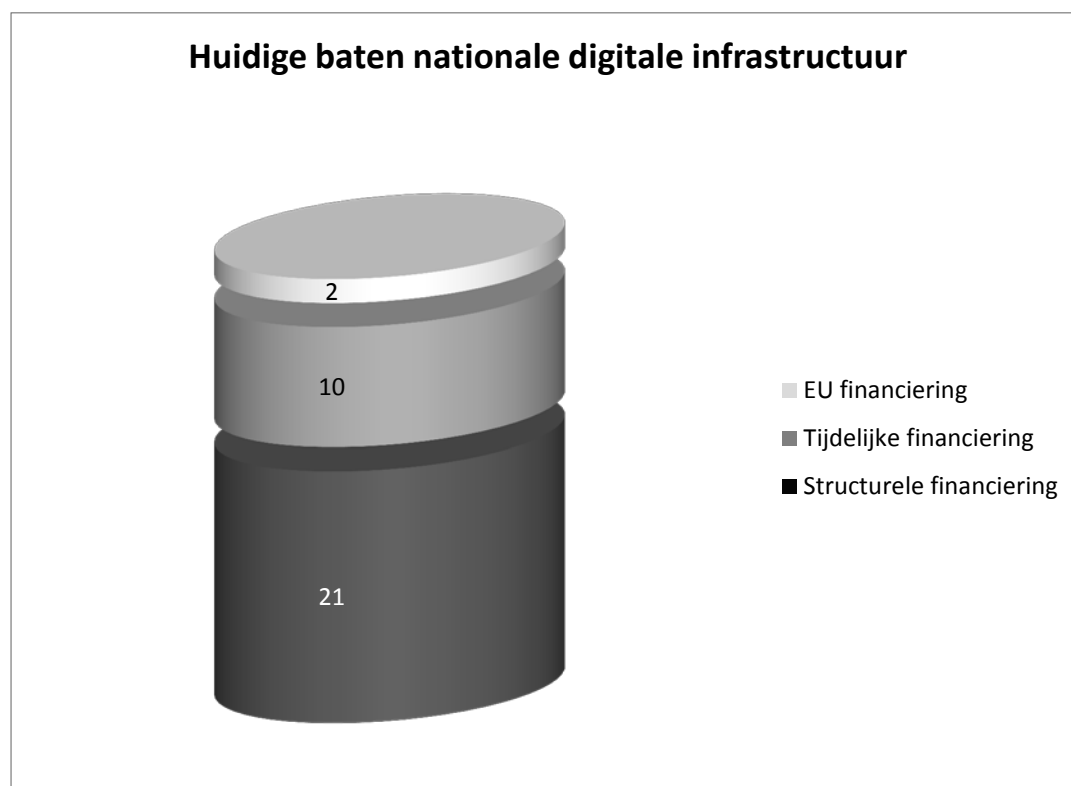
Naast deze kapitaalintensieve faciliteiten is er sinds 2011 het Netherlands eScience Center die zich richt op het onderdeel eScience in Figuur 1. Hiervoor is ruwweg 6 Meuro voor gereserveerd vanuit de structurele middelen. De totale middelen om de huidige digitale infrastructuur te financieren is 37 (18+10+3+6) Meuro (Figuur 4).



*Figuur 4: Huidige lasten nationale digitale infrastructuur (Bedragen in M€)*



In Figuur 5 wordt de opbouw van de huidige beschikbare middelen getoond. Zoals hierboven beschreven is er voor de financiering van de huidige lasten een tekort van 4 Meuro per jaar.



*Figuur 5: Huidige baten nationale digitale infrastructuur (Bedragen in M€)fl*

SURF en het eScience Center lopen bij het ontwikkelen van een duurzaam investeringsplan aan tegen het feit dat (i) het op dit moment ongebruikelijk is dat onderzoekers begroten en betalen voor geavanceerde digitale dienstverlening (en eScience), en (ii) de groeiende behoefte aan support. Dit leidt tot druk op de c.q. een tekort aan reserveringen voor de reken- en data-infrastructuur, en een laag slagingspercentage (10 procent) bij het eScience Center. Met dit advies – dat in lijn loopt met de internationale aanbevelingen die bijv. in het kader van de ESFRI-roadmap worden gedaan – wil de commissie deze eerste uitdaging (zie i) adresseren waar het gaat over grootschalige wetenschappelijke infrastructuren. SURF ondersteunt op dit moment diverse bestaande infrastructuren, uit eigen middelen.

De financiering van de digitale infrastructuur in Nederland is ontoereikend om de huidige lasten te dekken. Daarnaast wordt een derde van de beschikbare middelen via ad hoc tijdelijke financiering gerealiseerd. Ook de financiering van DANS staat onder spanning door de toenemende behoefte aan dienstverlening rondom data- en archivering. Er is zeer beperkt ruimte voor het matchen van Europese subsidies.

**De commissie concludeert dat voorzien in een robuuste en duurzame digitale infrastructuur voor competitief wetenschappelijk onderzoek met internationale ambities binnen de huidige middelen niet mogelijk is.**

## Gevolgen bij gelijkblijvende financieringsniveau

De explosieve toename, complexiteit en afhankelijkheid van (big) data, analytics en onderzoekssoftware in alle wetenschappelijke disciplines vereist substantiële ingrepen in de nationale digitale infrastructuur. Het huidige investeringsniveau is onvoldoende om Nederland in de pas te laten lopen met snelle huidige en toekomstige technologische ontwikkelingen. Wil Nederland mee blijven doen in de internationale top van de wetenschap, dan zijn forse investeringen in innovatie, capaciteit, ondersteuning, veiligheid en betrouwbaarheid onvermijdelijk (zie Hoofdstuk 3.2).

De afgelopen jaren kende een sterke groei wat betreft de behoeften aan datadiensten, onderzoekssoftware en hoogwaardige wetenschappelijke ondersteuning. Dit geldt in sterke mate voor wetenschapsgebieden die recentelijk sterk gedigitaliseerd zijn (bv. door de aanwezigheid van digitale bronnen en geavanceerde analysesoftware), maar geen sterke traditie van ICT-toepassing hebben binnen het vakgebied. De sterke behoefte aan nationale infrastructuur met sterke ondersteuningsmechanismen wordt gevoeld over de volle breedte van de wetenschap, zoals onder andere blijkt uit het rapport uit 2016 op basis van de nationale ePLAN-enquête onder meer dan 1000 Nederlandse wetenschappers. Daarnaast zijn er in de afgelopen jaren belangrijke ontwikkelingen geweest op het snijvlak van ICT en wetenschap die vooralsnog onvoldoende of niet in de volle breedte ondervangen kunnen worden door de wetenschappelijke digitale infrastructuur, zoals virtualisatie, Internet of Things, cloud, authenticatie en autorisatie, cybersecurity, datamanagement, software sustainability en open science.

Wellicht de belangrijkste ontwikkeling is de sterke stijging van de hoeveelheid data in de wetenschap - dit werd een aantal jaren geleden nog niet voorzien. Doordat de groei van data sneller gaat dan de ontwikkeling van mogelijkheden om deze kosteneffectief op te slaan lopen we tegen grote uitdagingen aan. Het gaat hier bovendien in de meeste gevallen om *big data*: data die complex, snel en groot is en bovendien niet toepasbaar voor wetenschappelijk onderzoek zonder grondige filtering en analyse. Voor globale klimaatdata alleen al is een zeer conservatieve schatting dat de hoeveelheid data in de periode 2015-2030 minimaal verzevenvoudigt. De hoeveelheid data van de LOFAR-radiotelescoop is in minder dan vijf jaar tijd, een factor 25 gegroeid: van minder dan een petabyte naar 25 petabyte. Compute Canada heeft op basis van schattingen van onderzoekers berekend dat in de komende vijf jaar de rekenbehoefte met een factor zeven toeneemt en dat er vijftien keer zoveel behoefte aan opslagcapaciteit zal zijn (zie Bijlage 4.4).

De commissie adviseert ook extra middelen vrij te maken omdat er geen andere financieringsbronnen zijn. Sinds het wegvallen van FES zijn er geen calls voor impulsfinanciering en financiering vanuit Europa is niet mogelijk voor een nationale digitale infrastructuur. Omdat het niet wenselijk is als een domeinonafhankelijke digitale infrastructuur voor *alle* wetenschappers in competitie zou moeten gaan met projecten van gebruikers, maakt de nationale digitale infrastructuur geen onderdeel uit van de Roadmap 2016.

Hoewel de capaciteit (per euro) van ICT-hardware toeneemt, betekent dit niet dat de digitale infrastructuur met minder middelen toe zou kunnen. Zoals elke volgende generatie laptops of smartphones niet goedkoper wordt, maar wel meer capaciteit en mogelijkheden per euro biedt, zo dalen ook bij de high-end infrastructuur de kosten niet. Terwijl de capaciteit stijgt, stijgen de eigen en hoeveelheid gebruikers met een hoger tempo. Bovendien dalen de kosten niet voor kortdurende en langdurige intensieve ondersteuning en softwareontwikkeling.

Het handhaven van het huidige investeringsniveau zal niet alleen verregaande consequenties hebben voor individuele wetenschappers, maar resulteert ook in diverse financiële en inhoudelijke risico's voor de wetenschap als geheel:

#### *Kwaliteitsverlies*

De kwaliteit van veel wetenschappelijk onderzoek hangt in grote mate af van de kwaliteit en capaciteit van de digitale infrastructuur. Onvoldoende investeren op nationaal niveau betekent onvoldoende snelle netwerken of rekencapaciteit ten opzichte van buitenlandse partijen. Daarmee wordt impliciet de keuze wordt gemaakt om geen steun meer te verlenen aan wetenschap van internationaal competitief niveau. Het betekent bijvoorbeeld dat Nederland haar positie zal moeten afstaan in de internationale top van data-intensieve wetenschap als natuur- en sterrenkunde.

#### *Investeringsniveau in Finland, Zwitserland en Zweden*

Dat het investeringsniveau in Nederland achter dreigt te gaan lopen, blijkt ook wanneer wordt gekeken naar andere Europese lidstaten. In Finland wordt bijvoorbeeld bijna twee keer zoveel geïnvesteerd (geschaald naar bevolkingsomvang) in CSC, dat de Finse digitale infrastructuur voor zijn rekening neemt. Opvallend aan de investeringen in Finland is vooral dat er verhoudingsgewijs flink meer wordt geïnvesteerd op het vlak van dienstverlening en support op het vlak van rekenfaciliteiten en computational science. Ook in vergelijking met Zwitserland is het gat groot. De investeringen in alleen al high performance computing (no 8. van de wereld, en de hoogste Europese notering) is ongeveer een factor 8-10 hoger dan in Nederland en Finland. Uitgestreken over 4 jaar zijn de bijdragen vanuit de overheid alleen op het gebied van computing rond de 43 miljoen Euro per jaar. Tot slot helpt het om te kijken naar de investeringsplannen in Zweden. Daar wordt de komende jaren rond de 20 miljoen Euro per jaar aan high performance computing besteed. Dit bedrag is bijna gelijk aan de geadviseerde 21 miljoen Euro per jaar voor high performance computing in Nederland. Momenteel is dat in Nederland 16 miljoen Euro, waarvan bovendien 5 miljoen bestaat uit tijdelijke middelen.

#### *Isolatie*

Deelname aan grote internationale projecten en infrastructuren zal niet meer mogelijk zijn. Door achterblijvende investeringen in netwerk en rekenfaciliteiten, kan niet worden voldaan aan de matchings- en participatie-eisen van Europese projecten en *Research Infrastructures*. Zo zal Nederland bijvoorbeeld niet langer binnen PRACE, de internationale Research Infrastructure voor Supercomputing, toegang kunnen krijgen tot de snelste supercomputers van Europa. Twee voorbeelden van huidig gebruik van PRACE-faciliteiten zijn simulaties van de interactie tussen eiwitten en nieuwe medicijnen en simulaties van weersextremen in het toekomstige klimaat. Ook zal er onvoldoende support zijn om de gemeenschappelijke en ondersteunende basisinfrastructuur te bieden voor wetenschappers en universiteiten die willen participeren in de Europese en nationale ontwikkelingen op het gebied van Open Science. Meer algemeen zal opvallen dat internationale wetenschappelijke samenwerkingsverbanden in gevaar gekomen door het ontbreken van voldoende toegevoegde waarde vanuit Nederland.

## Kwaliteitsverlies & isolatie in de praktijk

*gevolgen bij achterblijvende investeringen voor...*

### Klimaatonderzoek

Nederland is kwetsbaar voor klimaatverandering. Vooral de dreiging van te veel en te weinig water is relevant. Om uitspraken over de ontwikkeling van het klimaat te kunnen doen worden complexe computermodellen gebruikt. Het is essentieel om details van de atmosfeer, oceaan en ijs te representeren in die modellen. De ontwikkeling van het klimaat hangt af onder andere af van wolkenvorming en van stromingen in de oceaan. Rond 2023 zullen de eerste Exascale-computers operationeel zijn. Die zijn tot 1000 keer krachtiger dan de volgende generatie van de nationale supercomputer. Daarmee kan het klimaat in meer detail en beter gesimuleerd worden. Het Amerikaanse DOE sorteert daarop al voor met simulaties die wolken in detail simuleren en ook in Europa wordt geïnvesteerd. Al die modellen genereren veel data, naast de data uit satellieten, grondobservaties en vele andere bronnen.

De analyse van die data is een grote uitdaging en leidt tot nieuwe infrastructuren waar de dataopslag en de computers voor berekeningen geïntegreerd zijn. In het Verenigd Koninkrijk is al geïnvesteerd in zo'n vernieuwende infrastructuur bij STFC. Bij het Engelse UK Met Office staat een supercomputer die meer dan tien keer sneller is dan de Cartesius van SURFsara, voor alleen weerprognoses en klimaatonderzoek in het VK. In Duitsland staat alleen voor klimaatonderzoek (bij het DKRZ in Hamburg) een supercomputer (Mistral) die een factor vier groter/snelser is dan de Cartesius. De huidige investeringen in Nederland lopen achter ten opzichte van de grote buurlanden en de VS, met als concreet gevolg dat Nederlandse academische partijen geen toegang hebben tot de beste infrastructuur en deelname in competitieve internationale onderzoeksconsortia onder druk komt.<sup>1</sup>

*gevolgen bij achterblijvende investeringen voor...*

### Sterrenkunde

Astronomie beantwoordt de meest fundamentele menselijke vragen: "waar komen we vandaan?" en "zijn we alleen?". Nederlandse sterrenkunde presteert in *absolute* getallen in de top-3 van de wereld. Door nationale samenwerking slagen onze astronomen er dus in globaal ver boven hun maat te presteren. Daarbij is de volledige keten van telescoopontwikkeling, data-analyse en grootschalige simulaties onontbeerlijk.

De grootste telescopen, van over 1 miljard Euro, worden in Europese en globale samenwerkingen ontworpen en gebouwd. Bij Nederlands wetenschappelijk leiderschap verwerft Nederlandse industrie daar grote orders. Zulk leiderschap steunt op technologische innovatie en prestaties, aangezien de beeldverwerking van moderne telescopen zeer veeleisend is. Zo kan de Square Kilometer Array (SKA) in 2022 alleen *zien* nadat een Exaflops supercomputer de antennes combineert. Missies en telescopen als Euclid, JWST, SKA en de ELT produceren vervolgens gemiddeld 100 PB aan nieuwe data per jaar voor Nederlandse sterrenkundigen. Die analyseren zulke data in eenheden van honderden Petaflops aan *capacity* computing.

<sup>1</sup> <https://www.hpcwire.com/2016/09/07/exascale-computing-project-awards-39-8m-22-projects/>  
<http://www.metoffice.gov.uk/news/releases/archive/2014/new-hpc>  
<https://www.primavera-h2020.eu/>  
[http://cordis.europa.eu/project/rcn/197542\\_en.html](http://cordis.europa.eu/project/rcn/197542_en.html)  
<http://www.jasmin.ac.uk/>

Ons begrip van het heelal waarin we ontstaan zijn wordt pas volledig als onze beste simulaties de waarnemingen uiteindelijk reproduceren en verklaren. Alleen als Nederlandse wetenschappers toegang hebben tot Exascale *capability* computing kunnen zij de internationale concurrentie voor blijven. Uitblijven van structurele investeringen in de digitale infrastructuur zou het globale leiderschap van de Nederlandse astronomie op *elk van deze drie gebieden* direct bedreigen.

*gevolgen bij achterblijvende investeringen voor...*

## **Levenswetenschappen**

De levenswetenschappen richten zich internationaal op het ontwikkeling van kennis die direct invloed heeft op ons dagelijks leven, van gezondheid tot voeding, en van landbouw tot biodiversiteit. Nederlandse levenswetenschappers behoren tot de wereldtop, maar hun vorderingen worden steeds meer geremd door beperkingen op ICT-gebied. Het vakgebied heeft daarom dringend behoefte aan een meer hoogwaardige ICT-infrastructuur. Versnippering van investeringen op e-infrastructuurgebied zou voor de levenswetenschappen een regelrechte ramp betekenen.

Datasets in de levenswetenschappen nemen enorm toe in hoeveelheid en variëteit. Momenteel produceren onderzoekers al exabytes aan imaging- en omics-data (met name genomics en metabolomics) per jaar en de hoeveelheid data verdubbelt elke zes maanden. Deze data wordt daarbij op zeer veel plaatsen gelijktijdig gegenereerd. Er is daarom wereldwijd een geweldige behoefte aan hoogwaardige capaciteit voor data-opslag, uitwisseling en analyse.

Een extra uitdaging in het mensgebonden levenswetenschappelijk onderzoek is het privacygevoelige karakter van veel data. Dat vraagt om high performance, distributed analytics (ook wel aangeduid met 'distributed learning'). Hierbij worden eerst lokaal, in vele gedistribueerde databronnen, de analyses uitgevoerd, die vervolgens op meta-niveau gecombineerd worden om conclusies te trekken. Er is daarom een infrastructuur nodig die bestaat uit grote, centrale reken- en opslagcapaciteit (op dit moment verzorgd door SURF) en een sterk verbonden netwerk van lokale HPC-capaciteit, die in toenemende mate bij grotere onderzoeksinstellingen zal worden beheerd en kennis en kunde om dit in te zetten voor levenswetenschappen (zoals eScience dat doet). Deze opzet vereist veel aanpassingen op het gebied van software en security, ook terreinen waar Nederland grenzen verlegt.

Daarnaast maken levenswetenschappen steeds intensiever gebruik van data uit andere domeinen zoals aardwetenschappen, sociale wetenschappen en voedingsleer. Het is van groot belang dat datasystemen uit dergelijke verschillende domeinen kunnen communiceren door zich te richten op internationale data principes die zorgen dat data, standaarden en systemen Findable, Accessible, Interoperable en Reusable (FAIR) zijn voor mens en machine. Dit vereist samenwerking tussen alle partijen in het Nederlandse onderzoeksveld, een sterke coördinatie en afstemming van lokale en nationale ICT-investeringen. Alleen op deze manier zal Nederland een zeer vooraanstaande rol kunnen blijven spelen in het toekomstige onderzoek in de brede levenswetenschappen.

*gevolgen bij achterblijvende investeringen voor...*

## **Sociale & geesteswetenschappen**

Dankzij de *data-driven society* zijn zowel de bronnen als de methoden voor onderzoek gedurende de laatste tien jaar ingrijpend gewijzigd. Voor de geesteswetenschappen geldt dat veel erfgoed is gedigitaliseerd en dat er dagelijks een grote hoeveelheid 'born digital' materiaal ter beschikking komt. De sociale wetenschappen beschikken over nieuwe methoden voor het afnemen van surveys (bijvoorbeeld gebruik van sensors) en eveneens over een vloed aan data zoals die bijvoorbeeld via sociale media worden gegenereerd.

Roadmap-projecten als CLARIAH (PLUS) en ODISSEI bevorderen de verspreiding van innovatie in deze domeinen, maken daarbij gebruik van de huidige infrastructuur en vullen deze innovatief aan. Flankerende investeringen zijn op twee vlakken nodig voor het bereidstellen van een meegroeierende en structureel gefinancierde infrastructuur (de basis) en voor capaciteit om de duurzaamheid van tijdelijke en project-gefinancierde diensten (tools, data) te waarborgen.

Investeringen in infrastructuur zijn noodzakelijk omdat big data-onderzoek steeds grotere rekencapaciteit vergt en die trend zal zeker doorzetten. Veel onderzoeksprojecten zijn nu al afhankelijk van HPC-computing. Op basis van beschikbare voorzieningen wordt nieuwe technologie ontwikkeld. Recente voorbeelden zijn het project SHICO dat het volgen van culturele concepten door de tijd mogelijk maakt, of Golden Agents dat door het verknopen van verschillende bronnen laat begrijpen waar de innovatiekracht van de Gouden Eeuw vandaan kwam.

Een gebrek aan structurele investeringen in de digitale infrastructuur zal deze projecten belemmeren of ertoe leiden dat hun resultaten niet gewaarborgd kunnen worden. En dat verhindert op zijn beurt dat erop wordt voortgebouwd. In dat geval missen de Nederlandse geestes- en sociale wetenschappers de aansluiting met het buitenland.

### *Braindrain*

De aantrekkelijkheid van Nederland als vestigingsland voor onderzoekers, technologiestartups en toonaangevende ICT-bedrijven, zal snel verdwijnen bij achterblijvende investeringen. Data-intensieve wetenschappelijke onderzoekers, data scientists en innovatieve kennis zal verdwijnen naar landen om ons heen waar substantieel meer wordt geïnvesteerd in (wetenschappelijke) digitale infrastructuur.

### *Fragmentatie*

Door ontbreken van nationale investeringen, zullen individuele onderzoeksgroepen eigen oplossingen bouwen, die versnipperd, niet-interoperabel, minder kostenefficiënt en dus veel minder krachtig zijn. Ook kan door het ontbreken van nationale regie op verschillende plekken ongebruikte capaciteit ontstaan.

### *Privacy- en veiligheid*

Onvoldoende investeringen in betrouwbare en beveiligde netwerken en data-faciliteiten, zorgt voor afname in vertrouwen in de veiligheid van data en de privacybescherming van persoonsgegevens, bijvoorbeeld bij medisch onderzoek. Omdat er niet genoeg vertrouwen onder gebruikers is dat commerciële cloud computing en storage voldoende garantie kunnen geven over de veiligheid van onderzoeksdata, kan het ontbreken van publieke nationale infrastructuur betekenen dat grote beperkingen optreden in wetenschappelijke innovatie.

Dit zal onder andere het geval zijn bij medisch - en sociaalwetenschappelijk onderzoek, waar grote hoeveelheid persoonsgegevens worden verwerkt.

### *ICT-innovatie*

Door het niet laten meegroeien van investeringen in de digitale infrastructuur, of het zelfs nalaten om nieuwe digitale infrastructuur aan te schaffen zoals HPC, zal het niveau van ICT-onderzoek en de ICT-innovatie, en de hoeveel ICT-onderzoekers (waaronder veel data scientists) afnemen. Excellente infrastructuren werken als een magneet op (buitenlands) toptalent. Door het verdwijnen van human capital op dit snelgroeibende gebied zal Nederland wat betreft aanwezigheid van kennis en talent snel een achterstand oplopen die lastig meer in te lopen is.

De vraag over een evaluatiemechanisme wordt in 3.2 geadresseerd.

## 3.2 Digitale infrastructuur in internationale kopgroep

Het geven van een advies over de aanvullingen op de eerste taak die nodig zijn om de Nederlandse digitale infrastructuur op een vergelijkbaar niveau te brengen/houden in vergelijking met belangrijke andere kenniseconomieën.

Voor Nederland als top kenniseconomie is een hoogwaardige digitale infrastructuur essentieel. De Nederlandse instellingen hebben de ambitie om:

- een brede en diverse groep geavanceerde gebruikers goed te bedienen, aansluitend op hun behoeften. Hiervoor is het nodig om voortdurend en na consultatie van de gebruikers te investeren in topfaciliteiten. Samen met lokale instellingen wordt een brede set aan faciliteiten gecreëerd voor de (geavanceerde) gebruikers van digitale infrastructuur. Dit is geen statisch gegeven – vandaar dat continu investeren nodig is.
- strategisch partner te zijn voor topspelers in de markt, zodat zij in een vroege fase tegen gunstige prijzen producten beschikbaar stellen voor experimenteren en pilots door geavanceerde gebruikers. Dit kan alleen gerealiseerd worden met een infrastructuur die zich kan meten met de top – de combinatie van topkennis en topfaciliteiten zorgt dat SURF een strategische gesprekspartner op niveau kan zijn.

Om deze ambities te realiseren is een excellente kennisbasis nodig die zich continu vernieuwt. Omdat de ontwikkelingen razendsnel gaan, is een regelmatige vernieuwing van de infrastructuur noodzakelijk (eens per 4 à 5 jaar) op basis van de eisen die de wetenschappelijke gemeenschap naar voren brengt. Tenslotte is geavanceerde support nodig om de wetenschappers te ondersteunen bij het gebruik van deze faciliteiten.

**Om deze ambities waar te maken adviseert de commissie een totaal investeringsniveau van 50 Meuro per jaar voor de digitale infrastructuur in Nederland. De commissie verwacht dat via de middelen van de Nationale Roadmap en het verwerven van Europese middelen een bedrag van 4 Meuro verworven kan worden. De commissie adviseert daarom een extra investering van 25 Meuro per jaar te realiseren. Hiervan is 10 Meuro bedoeld om de huidige incidentele middelen structureel te maken en 4 Meuro om het tekort op de exploitatie van de huidige supercomputer aan te vullen. De overige middelen 11 Meuro worden ingezet om een extra ambitie op het gebied van de ondersteuning van de onderzoekers een investering in e-science en onderzoekssoftware en het realiseren van een extra impuls op het gebied van high performance computing.**

	Huidige besteding		EU projecten	Totaal huidig	Additionele financiering o.b.v. huidig advies	Gewenste omvang ICT infra
	Basis	Tijdelijk NL overheid				
Netwerken	5	4	1	10	1	11
Cybersecurity (AAI), Data	2	1	-	3	3	6
Rekenfaciliteiten	8	5	1	14	8	22
NL eScience	6	-	-	6	5	11
<b>Totaal</b>	<b>21</b>	<b>10</b>	<b>2</b>	<b>33</b>	<b>17</b>	<b>50</b>

*Tabel 1. Huidige besteding en advies additionele middelen (in miljoen €/jaar)*

*Noot. De bedragen onder 'Huidige besteding' zijn exclusief het huidige budget van DANS. DANS is wel meegenomen in de additionele geadviseerde financiering.*

#### *4 Meuro/jaar extra voor Netwerken, Cybersecurity en Data*

Dit bedrag is nodig om de structurele financiering op een niveau te brengen waarin de Nederlandse ambities, zoals verwoord in o.a. de Roadmap en de Nationale Wetenschapsagenda, te faciliteren. De beantwoording van vraag 1 laat zien dat er op dit moment een schreeuwend structureel tekort is. Daarnaast zien we dat de datagroei en digitalisering flink hoger liggen dan ten tijde van eerdere adviezen was voorzien. Excellente ICT-faciliteiten zijn een voorwaarde voor succesvol participeren in internationale onderzoeksinfrastructuren, zowel om dit soort infrastructuur in het buitenland te kunnen bereiken halen als om die in naar Nederland te kunnen halen. Deze investering beslaat een mix van netwerk- en rekenfaciliteit, dataopslag en de benodigde support voor grootschalige projecten die niet op de Roadmap staan. Tevens zal er matching voor veel nationale en Europese projecten nodig zijn om met infrastructuur bij te kunnen dragen. Dit deel van het advies sluit ook aan bij een andere ambitie van Nederland: een digitale mainport als basis voor een goed vestigingsklimaat (zie 2.2).

#### *8 Meuro/jaar extra om de Nederlandse HPC op niveau te brengen met andere vooraanstaande kenniseconomieën in Europa*

De beantwoording van vraag 1 (zie 3.1) liet zien dat de behoefte aan support omvangrijk is en ten koste gaat van de reserveringen voor een nieuw HPC-cluster. Daarnaast zal een extra investering in high performance computing (HPC) nodig zijn om in de pas te blijven van de landen om ons heen. Landen om ons heen (Duitsland, Frankrijk, Zwitserland) werken toe naar exascale-computers (zie o.a. EuroHPC), waarmee wetenschappers ongekennde mogelijkheden tot hun beschikking krijgen om grote maatschappelijke uitdagingen aan te gaan (bijvoorbeeld betere klimaatsimulaties voor vragen rondom de opwarming van de aarde) en huidige berekeningen kosten-efficiënter te kunnen uitvoeren. Nederlandse onderzoekers kunnen alleen toegang krijgen tot deze grootste Europese rekenfaciliteiten als Nederland ook nationale rekenfaciliteiten van een minimale omvang van typisch 10 procent van de grootste rekenfaciliteiten in Europa. Tevens moeten wetenschappers hun modellen kunnen testen op nationale systemen voordat ze gebruik maken van de grootste Europese faciliteiten. Dit was ook een van de conclusies van de ePLAN enquête. Daarnaast is er een toenemende behoefte aan een diversiteit van rekenfaciliteiten. Grootschalige simulaties in de astronomie, klimaat en materiaalkunde vereisen een andere rekenfaciliteit dan grootschalige



analyse van data in genomics bijvoorbeeld. Een grotere diversiteit van infrastructuur aanbod is nodig. Hierbij kan ook gebruik worden gemaakt van commerciële cloud-oplossingen. De geavanceerde rekenfaciliteiten zullen op termijn ook extra gebruikersondersteuning vragen. De middelen zullen daarom eerst worden ingezet om de aanschaf van een nieuwe supercomputer mogelijk te maken die borgt dat Nederland in de pas blijft lopen met omliggende landen. Vervolgens wordt een deel van de middelen (circa M 2 per jaar) ingezet om de verwachte toename van de gebruikersondersteuning op te vangen. Uiteindelijk zal Nederland zo een jaarlijkse reservering van circa 8 Meuro per jaar voor vervanging van de high performance computing faciliteiten, voor gebruikersondersteuning 7 Meuro per jaar, voor operationele kosten 6 Meuro per jaar en voor vernieuwing van de kennisbasis 1 Meuro realiseren.

#### *5 Meuro/jaar extra voor eScience en onderzoekssoftware*

De commissie adviseert om 5 Meuro/jaar extra te investeren in eScience en onderzoekssoftware. Dit is nodig om te kunnen voorzien in de groeiende behoefte van gebruikers aan dit soort specialistische diensten en ondersteuning. Hoewel slechts een klein deel van de Nederlandse onderzoekers bekend is met het eScience Center (~20% volgens ePLAN), kan door de beperkte middelen slechts 10% van de aanvragen bij het eScience Center worden gehonoreerd. Dit percentage zal naar verwachting verder dalen, gezien de sterke groei van data- en software-intensief onderzoek in alle wetenschappelijke disciplines. Het gevolg is dat wetenschappers zelf een groot deel van hun kostbare en beperkte onderzoekstijd moeten besteden aan het ontwikkelen en toepassen van software en gebruik van digitale infrastructuur. Dit levert niet alleen vertraging op voor onderzoekers, maar ook gaat het ten koste van de kwaliteit en duurzaamheid van de ontwikkelde software. Daarom wordt dringend geadviseerd meer middelen beschikbaar te stellen om wetenschappers te ondersteunen met nog te ontwikkelen specialistische onderzoekssoftware en deze nationaal en actief aan te bieden aan de wetenschappelijke community.

De commissie adviseert ook te investeren in de ontwikkelingen en het onderhoud van een duurzame infrastructuur voor onderzoekssoftware. Vooral wordt in Nederland enkel voorzien in een duurzame en algemene infrastructuur voor netwerken (SURF), rekencapaciteit (SURF) en data & archivering (SURF en DANS). Met een infrastructuur voor software is met het eScience Technology Platform reeds een begin gemaakt. Hiermee wordt ontwikkelde software uit projecten toegankelijk gemaakt volgens de principes van Open Science, wordt samenwerking tussen wetenschappers gestimuleerd en wordt fragmentatie (zelfde software die op meerdere plekken wordt ontwikkeld) verminderd. Additionele investering op het gebied van eScience levert bovendien de noodzakelijke versteviging op tussen het Nederlandse ICT-onderzoek (computer en data science) en de nationale infrastructuur enerzijds en het data-intensieve wetenschappelijke onderzoek anderzijds. Hiermee wordt het gebruik en bezettingsgraad van de digitale infrastructuur gestimuleerd, terwijl ook intensievere multidisciplinaire samenwerking tussen ICT-onderzoek en andere wetenschapsdomeinen mogelijk wordt gemaakt.

#### *Zorg voor eenheid en focus in beleid*

Na het advies van ICTregie heeft SURF de regie over de digitale infrastructuren voor wetenschappelijk onderzoek gekregen, waardoor de directe verantwoordelijkheid voor strategie en innovatie nog meer bij de gebruikers kwam te liggen.

De commissie adviseert om ook regie te voeren over de publieke investeringen in ICT en digitale infrastructuren voor onderwijs en onderzoek. Stel daartoe een ambtelijk regieoverleg in tussen de ministeries van Onderwijs Cultuur en Wetenschap, Defensie, Economische Zaken, Infrastructuur en Milieu en Volksgezondheid, Welzijn en Sport en betrek SURF hierbij. Deze ministeries zijn betrokken bij investeringen in digitale infrastructuren voor onderwijs en onderzoek, zoals voor KNMI, MARIN, NLR, maar ook voor de KB en universiteitsbibliotheken. In dit regieoverleg moeten ministeries elkaar informeren over voorgenomen investeringen en bezien waar synergie te behalen valt.

Daarnaast adviseert de commissie aan NWO om bij investeringen in grootschalige onderzoeksfaciliteiten te bezien of deze faciliteiten gebruik kunnen maken van nationale digitale infrastructuren en hoe eventuele investeringen in digitale infrastructuren ook gefedereerd door een bredere groep gebruikers benut kunnen worden. Meer hierover in 3.3.

### Ontwikkel een visie op nieuwe thema's en regisseer de uitvoering

Er zijn diverse nieuwe thema's waarvoor eerst een visie op het gebied van onderzoek en innovatie nodig is. Binnen de huidige en gevraagde middelen kan, onder regie van SURF, de inzet en innovatie van de nationale digitale infrastructuur worden vastgesteld. In navolging van de ePLAN-enquête wordt voorgesteld om jaarlijks de behoeftes onder domeinonderzoekers peilen door ePLAN. Nieuwe technologische ontwikkelingen, zoals ontwikkelingen van sensornetwerken en Internet of Things, zijn voorbeelden van behoeftes aan nieuwe faciliteiten en diensten. De totstandkoming van het Nationaal Plan Open Science is (ook) een goed recent voorbeeld hoe zo'n uitvraag kan werken. De commissie vraagt speciale aandacht voor de verdere uitwerking van dit Nationaal Plan Open Science.

Het is niet effectief om parallel vergelijkbare diensten te ontwikkelen. Dat vraagt om samenwerking tussen de betrokken organisaties, onderzoeksfaciliteiten en de dienstverleners van ICT infrastructuur. De commissie adviseert hiervoor scenario's te onderzoeken. Binnen het huidige advies kan een deel van de Open Science ambities worden gerealiseerd, in het bijzonder door ontwikkeling van een grotere diversiteit aan faciliteiten en ontwikkelen van diensten en support.

### Evaluatiemechanisme: sluit aan bij bestaande evaluaties

De commissie adviseert om de bestaande evaluatiemechanismen van SURF, het eScience Center en DANS op elkaar af stemmen zodat het aanbod van digitale infrastructuur aan de Nederlandse wetenschap afdoende en in zijn geheel kan worden onderzocht. Bovendien zorgt het ervoor dat de doelmatigheid van additionele investering in de nationale infrastructuur goed kan worden gecontroleerd.

Toekenning van meerjarige financiering voor SURF gebeurt op basis van een door de aangesloten instellingen goedgekeurd meerjarenplan, waarin investeringen, innovatie, dienstontwikkeling, dienstverlening en ondersteuning beschreven wordt. Halverwege vindt een mid-term review plaats door een evaluatiecommissie met ook internationale leden. De uitkomsten worden gedeeld met de leden van SURF, inclusief NWO.

De toekenning van meerjarige financiering van eScience Center gebeurt op basis van een samenwerkingsovereenkomst tussen NWO en SURF. eScience Center wordt elke zes jaar extern geëvalueerd door een onafhankelijke evaluatiecommissie, waarna ook SURF en NWO de resultaten ontvangen. DANS wordt geëvalueerd volgens het Standard Evaluation Protocol van NWO, KNAW en VNSU (SEP) met elke zes jaar een externe onafhankelijke evaluatie.

### 3.3 Aanbevelingen voor de Nationale Roadmap

Het adviseren van de permanente commissie over de formulering van het criterium voor ICT bij de beoordeling van aanvragen voor de Nationale Roadmap.

De Roadmap is al gepubliceerd en heeft enkele goede aanbevelingen op het gebied van ICT gedaan (in het bijzonder 4.3 & 4.6 van de Roadmap). Desondanks wil de commissie nog enkele aanvullende aanbevelingen doen voor de beoordeling van ICT bij de aanvragen voor de Nationale Roadmap.

#### Gebruik de e-Needs vragen van ESFRI bij de call voor de Nationale Roadmap

De commissie adviseert om in de aanstaande call voor de Roadmap de ESFRI-2018-vragen (e-Needs) te hanteren. Het gebruik van de e-Needs-vragenlijst zorgt voor eenduidigheid in de aanvragen die onderzoekers doen in Nederland en in Europa. Bovendien geven de antwoorden inzicht in de relatie tussen project en nationale of internationale digitale infrastructuren en eventueel te behalen synergie.

#### Zorg voor voldoende ICT-expertise in de Roadmap-beoordelingscommissie

De commissie onderschrijft de aanbeveling van de Permanente Commissie om bij de samenstelling van de beoordelingscommissie te borgen dat er ook voldoende expertise aanwezig is om de ICT-component te kunnen beoordelen. ESFRI heeft ervoor gekozen om leden van de e-IRG te vragen voor beoordeling van de Europese calls. In lijn hiermee adviseert de commissie om in Nederland SURF, DANS en eScience Center om de benodigde expertise te laten leveren.

#### Vraag ook in andere calls naar de behoefte aan digitale infrastructuur

De commissie vindt het een goede eerste stap om in de Roadmap call onderzoekers te vragen naar hun gebruik van digitale infrastructuur en ondersteuning. Dit initiatief moet verder worden uitgebreid naar andere nationale en Europese calls voor onderzoeksfinanciering. Op die manier maken de calls onderzoekers in alle domeinen meer bewust van de mogelijkheden van digitale infrastructuren en ondersteuning én kunnen deze infrastructuren beter benut worden.

#### Stimuleer het federatief inzetten van digitale infrastructuren

Niet altijd zijn de beschikbare nationale digitale infrastructuren voldoende voor de eisen van de gebruiker. Wanneer de beoordelingscommissie in dat geval verantwoord acht dat de aanvrager voor zijn domein of cluster een eigen of aanvullende digitale infrastructuur ontwikkelt, moet direct bekeken worden of deze op een andere manier een breder bereik kan hebben. De commissie adviseert dan ook om het federeren van digitale infrastructuren onder regie van SURF te stimuleren. Een huidig voorbeeld is Nikhef, dat samen met SURF investeringen, inzet en beschikbaarheid van grid- en clouddiensten afstemt ten behoeve van de wetenschappelijke gebruikersgemeenschappen.

#### Stimuleer samenwerking tussen domeinspecifieke infrastructuren

Verschillende nationale grootschalige onderzoekersfaciliteiten hebben een sterke infrastructurele datacomponent, waaronder BBMRI en ELIXIR-NL (Life Sciences), CLARIAH (geesteswetenschappen), ICOS (klimaat) en ODISSEI (sociale wetenschappen en economie). De commissie adviseert om bij dit soort onderzoeksfaciliteiten de samenwerking met andere onderzoeksfaciliteiten te stimuleren om onderling en met de nationale dienstverleners samen te werken aan het ontwikkelen van tools en

software, standaarden en beleid. Zo wordt kennis ook buiten het eigen cluster gedeeld en kan men gezamenlijk werken aan standaarden. Tenslotte is voor enkele onderzoeksfaciliteiten aansluiting bij de initiatieven GO-FAIR en Research Data Netherlands (RDNL) nuttig.

### 3.4 Aanbevelingen voor flankerend beleid

Omdat de digitale infrastructuur niet op zichzelf staan en er meer nodig is om de ambities uit de Nationale Wetenschapsagenda en de Roadmap te behalen, vraagt de commissie aandacht voor randvoorwaardelijk beleid dat buiten de directe invloedssfeer ligt van de aanbieders van de nationale digitale infrastructuur en ondersteuning.

#### Versterk de samenhang tussen ICT-onderzoek en de digitale infrastructuur

De commissie adviseert om de samenhang tussen ICT-onderzoek en de digitale infrastructuur te versterken. Nederland heeft als digitale mainport veel belang bij kennis en innovatie op het gebied van digitale infrastructuur. Besteed daarom bij de wetenschappelijke programmering meer aandacht aan onderwerpen die nieuwe inzichten of technieken opleveren voor de nationale digitale infrastructuur. Zorg dat eScience Center een rol kan vervullen om samen met ICT-onderzoekers nieuwe methoden en technieken op te pakken en te toetsen op hun bruikbaarheid voor de nationale digitale infrastructuur. En zorg tenslotte voor implementatie van bruikbare concepten en technieken door de aanbieders van de nationale digitale infrastructuur.

#### Organiseer lokale onderzoeksondersteuning voor onderzoekers

Onderzoekers hebben een groeiende behoefte aan digitale ondersteuning en expertise in het onderzoeksproces, zeker in disciplines waar het omgaan met software en grote hoeveelheden data nieuw is. Het is niet schaalbaar en praktisch om deze ondersteuning enkel op nationaal niveau te regelen, omdat in veel gevallen om kortdurende hands-on support gaat en bovendien om niet al te complexe uitdagingen. De commissie adviseert de VSNU daarom om in het kader van hun visie 'De Digitale Samenleving' bij de instellingen te stimuleren dat ze hun ICT-afdelingen en/of bibliotheken omvormen of uitbreiden tot research support-afdelingen. Deze afdelingen zijn dan het eerste aanspreekpunt voor de onderzoeker en kunnen ook als schakel dienen naar de nationale ondersteuning en de nationale digitale infrastructuur.

#### Leid data scientists, data stewards en digitale wereldburgers op

Digitalisering van wetenschap vereist digitaal geschoolde werknemers. Tekorten aan ICT-professionals zetten ons op achterstand. De commissie adviseert daarom om digitale skills een prominente plaats in alle niveaus van het onderwijs te geven, ook in het wetenschappelijk onderwijs. Maak data- en programmeervaardigheden onderdeel van alle wetenschapsdisciplines en -opleidingen, zoals statistiek dat als basis ook is.

Leid voldoende data scientists en data stewards op: in alle sectoren ontstaat een grote behoefte aan data scientists en stewards, bij bedrijven, de overheid en in de wetenschap zelf. De Europese High Level Expert Group voor de European Open Science Cloud schat dat Europa een half miljoen data stewards nodig heeft. De commissie adviseert de VSNU in het kader van hun visie 'De Digitale Samenleving' na te denken over hoe universiteiten een aantrekkelijk carrièreperspectief aan data stewards kunnen bieden. Dit sluit goed aan bij het advies van Platform Onderwijs 2032 en van de KNAW in 2012.

## 4 Bijlagen

### 4.1 Overzicht nationale digitale infrastructuur

SURF, DANS (initiatief KNAW en NWO) en eScience Center (initiatief SURF en NWO) zijn de organisaties met een nationale opdracht op het gebied van digitale infrastructuur. Zij bieden verschillende typen infrastructurele diensten, die hieronder worden beschreven: een beschrijving van de infrastructuur, het nut ervan voor onderzoekers en een korte analyse van welke ontwikkelingen er de komende jaren nodig zijn om onderzoekers goed te kunnen faciliteren.

Hoewel er ook lokale – en domeinspecifieke netwerken en rekenfaciliteiten zijn in Nederland, zijn deze lokaal of projectgebaseerd gefinancierd en in de meeste gevallen niet landelijk beschikbaar of niet geschikt zijn voor meerdere wetenschappelijke domeinen. Ze worden daarom beperkt genoemd. Internationale infrastructuren worden wel kort beschreven, omdat voor deelname hieraan inbreng van nationale middelen vereist is en ze noodzakelijk zijn voor internationale samenwerking van onderzoekers.

#### 4.1.1 Netwerken

Het netwerk verbindt de biobanken van BBMRI-NL met elkaar, transporteert data van CERN naar computing- en datacenters in de hele wereld, biedt onderzoekers toegang tot de gedistribueerde data in CLARIAH, laat onderzoekers online samenwerken en koppelt (commerciële) clouddiensten met de instellingen. Reguliere commerciële netwerkaanbieders voldoen wat betreft kwaliteit, snelheid, veiligheid en capaciteit voor de meeste data-intensieve wetenschappers niet.

#### *Organisatie*

Toegang tot netwerken gebeurt lokaal via de campusnetwerken van de instellingen. Deze zijn verbonden met het nationale netwerk van SURF. SURF werkt op zijn beurt samen met Europese en andere internationale partners om onderzoekers in hun wereldwijde samenwerking te faciliteren. SURF werkt op zijn beurt samen met Europese (GÉANT) en andere internationale partners (NRENS) om onderzoekers in hun wereldwijde samenwerking te faciliteren.

#### *Ontwikkeling vast netwerk*

Capaciteit is jarenlang de belangrijkste factor geweest bij nieuwe netwerkgeneraties. Dit is nog steeds belangrijk, maar het blijkt ook dat de basis van het huidige internet niet ontworpen is voor het veranderende gebruik. Onderzoekers slaan grote hoeveelheden data niet meer op in een eigen datacentrum, maar sturen ze over publieke verbindingen naar gezamenlijke databanken en de cloud. Privé-netwerken bieden een oplossing voor grote hoeveelheden data of voor data waar beveiliging belangrijk is, zoals medische onderzoeksdata. Dit kan door snelle en veilige lichtpaden op te zetten tussen een grootschalige onderzoeksfaciliteit en een data- of computingcenter.

Daarom is het minstens zo belangrijk om de netwerkdiensten geavanceerder en flexibeler te maken door nieuwe intelligente technieken (virtuele en programmeerbare netwerken) in bouwblokken beschikbaar te stellen.

Zo kunnen dienstenaanbieders en gebruikers zelf nieuwe toepassingen maken op de geboden bouwblokken. Vergelijk het met een smartphone, waarbij derden applicaties bouwen op de functionaliteiten die de smartphoneproducent biedt. Door creatieve ideeën realiseerbaar te maken versnelt innovatie en ontstaat een ecosysteem van aanbieders en afnemers waarin sneller en beter ingespeeld kan worden op de behoeften van onderzoekers.

#### *Ontwikkeling draadloze netwerken*

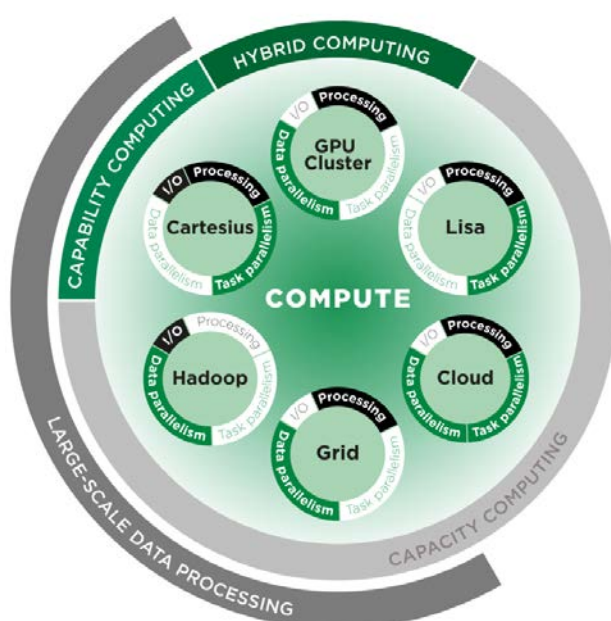
De eindpunten van het netwerk worden in toenemende mate draadloos. Dit geldt voor de gebruiker die via draadloze systemen toegang zoekt tot het netwerk. Door de hoge eisen van de gebruiker aan bandbreedte en kwaliteit van het draadloze netwerk hebben instellingen meer behoefte aan geavanceerde kennis en ondersteuning bij de uitrol van deze netwerken.

Maar ook de instrumenten worden massaal mobiel. Denk aan sensornetwerken, Internet of Things. Er zijn sensoren die eenmaal per dag data versturen en sensoren die continu alert moeten zijn op de omgeving, zoals bij zelfrijdende auto's. Voorbeelden van sensornetwerken in de wetenschap zijn netwerken voor fijnstofmetingen, aardbevingssensoren en smart cities. Een enorme groei van sensornetwerken is te verwachten, met daarbij hoge eisen van onderzoekers aan betrouwbaarheid, veiligheid, bandbreedte, dekking etc.

#### 4.1.2 Rekenfaciliteiten

Voor simulaties en voor het verwerken en analyseren van data hebben onderzoekers computerfaciliteiten nodig. Klimaat- of genetische modellen worden berekend met behulp van grote computerclusters, maar ook taalwetenschappers en cartografen maken in toenemende mate gebruik van geavanceerde computerfaciliteiten.

Veelal maken wetenschappers voor berekeningen gebruik van nationale computerfaciliteiten, zoals SURF die biedt. Omdat niet elke toepassing om eenzelfde type reksysteem vraagt, bestaat de nationale computerinfrastructuur van SURF uit verschillende typen computing. Het huidige aanbod, type gebruik en type gebruikers is hieronder in twee schema's weergegeven.



Figuur 1 Portfolio SURF

Infrastructuur	Typical usage	Typical users	Use cases
Cartesius	<ul style="list-style-type: none"> <li>• Capability computing</li> <li>• MPI applications</li> </ul>	<ul style="list-style-type: none"> <li>• Research groups</li> <li>• Consortia</li> </ul>	<ul style="list-style-type: none"> <li>• Climate modeling</li> <li>• Fluid dynamics</li> <li>• Galaxy simulation</li> </ul>
LISA	<ul style="list-style-type: none"> <li>• Job farming</li> <li>• MPI applications</li> </ul>	<ul style="list-style-type: none"> <li>• Research groups</li> <li>• Individual researchers</li> </ul>	<ul style="list-style-type: none"> <li>• GWAS studies</li> <li>• Magnetic field research</li> </ul>
Grid	<ul style="list-style-type: none"> <li>• Job farming</li> <li>• I/O heavy applications</li> </ul>	<ul style="list-style-type: none"> <li>• Consortia</li> </ul>	<ul style="list-style-type: none"> <li>• LHC experiments</li> <li>• LOFAR</li> <li>• NGS data analysis</li> </ul>
HPC Cloud	<ul style="list-style-type: none"> <li>• 3<sup>rd</sup> party PaaS</li> <li>• Courses</li> <li>• High Memory apps</li> <li>• Microsoft Windows (&amp; other licensed software)</li> </ul>	<ul style="list-style-type: none"> <li>• Individual researchers</li> <li>• Consortia</li> </ul>	<ul style="list-style-type: none"> <li>• Galaxy platform</li> <li>• AMC compute course</li> <li>• De novo sequence alignment</li> <li>• Stock exchange ticker data</li> </ul>
Hadoop	<ul style="list-style-type: none"> <li>• Pattern recognition / data mining</li> <li>• Web crawls</li> <li>• Information retrieval</li> </ul>	<ul style="list-style-type: none"> <li>• Individual researchers</li> </ul>	<ul style="list-style-type: none"> <li>• Twitter mining</li> <li>• TwiNL</li> <li>• SETI analysis</li> </ul>

Figuur 2 Gebruik SURF portfolio

**Organisatie**

Onder regie van SURF werkt de digitale infrastructuur met partners binnen onderzoeksinstituten aan een gefedereerde infrastructuur. Zo stemmen de Nikhef, RUG-CIT en SURF gezamenlijk de investeringen, inzet en beschikbaarheid van grid- en clouddiensten af ten behoeve van hun wetenschappelijke gebruikersgemeenschappen. Bij enkele grootschalige onderzoeksfaciliteiten met een exceptionele dataproductie maken HPC-clusters onderdeel uit van de faciliteit om de eerste dataselectie te doen (zoals LOFAR, LHC). Na die eerste selectie worden data verspreid over de wereld en kunnen onderzoekers nationale rekenfaciliteiten gebruiken voor onderzoek op de data.

Enkele publieke instellingen beschikken over een kleinere HPC-faciliteit: KNMI (weerprognoses), MARIN (hydrodynamische simulaties) en NLR. Deze faciliteiten bieden capaciteit op nationaal niveau. Om nieuwe theorieën en concepten te beproeven is er voor het ICT-onderzoek in Nederland de DAS-5 (Distributed ASCI Supercomputer 5) beschikbaar, die als flexibele onderzoeksinfrastructuur dient.

**Ontwikkeling computing**

Met de enorme groei van data en de diversiteit en complexiteit van data neemt de behoefte aan geavanceerde computerfaciliteiten en –diensten toe. Dit geldt niet alleen voor de supercomputer, maar voor alle typen rekendiensten.

De rekencapaciteit moet blijven groeien en minstens in de pas blijven lopen met de internationale ontwikkelingen, ook om de Nederlandse onderzoekers toegang te bieden aan nog grotere Europese supercomputers. Duurzaamheid is daarbij een steeds belangrijker aspect. De rekenfaciliteit wordt veel sterker verbonden met de dataopslag om data-intensief onderzoek mogelijk te maken. Ook de behoefte aan specialistische software- en statistiekpakketten groeit, die het analyseren van grote complexe datasets mogelijk maken.

#### 4.1.3 Data-opslag en -archivering

Onderzoekers willen hun ruwe en verwerkte onderzoeksdata bewaren, ook als het om grote, diverse en complexe data gaat. De data moeten immers herbruikbaar zijn voor verder onderzoek of voor controle van de onderzoeksresultaten. Zeker bij data uit kostbare of langdurige onderzoekstrajecten is het wenselijk dat andere onderzoekers ze kunnen hergebruiken.

Een uitdaging voor datadiensten is dat data veilig moeten worden opgeslagen, maar tegelijkertijd beschikbaar moeten zijn voor de onderzoeker. Er zijn datadiensten op verschillende niveaus: dicht bij de opslagvoorzieningen, gericht op een optimale, veilige en snelle toegang; diensten op een hoger aggregatieniveau en domeinspecifieke diensten.

##### *Organisatie*

Naast de lokale datadiensten bij instellingen zijn er enkele spelers die op regionaal en nationaal niveau datadiensten aanbieden: SURF en DANS doen dit nationaal, het 4TU.Centre for Research Data voor de 4TU's.

DANS bevordert duurzame toegang tot digitale onderzoeksgegevens en stimuleert onderzoekers om gegevens duurzaam te archiveren en te hergebruiken. Ook biedt DANS toegang tot duizenden wetenschappelijke datasets, publicaties en andere onderzoeksinformatie in Nederland. SURF biedt verschillende diensten op het gebied van big data, waaronder veilige en langdurige opslag, eenvoudig opslaan en delen, bigdata-analyse en codering om data vindbaar te maken.

DANS, SURF en 4TU.Centre for Research Data werken samen binnen Research Data Netherlands (RDNL) en het Landelijke Coördinatiepunt Research Data Management. Ze opereren als zogenaamde *backoffices*, waarbij de *frontoffices* (bijvoorbeeld de universiteitsbibliotheken of universitaire ICT-afdelingen) in direct contact staan met de onderzoekers. Verder zijn er domeinspecifieke datadiensten, onder andere geleverd door grootschalige onderzoeksfaciliteiten.

##### *Ontwikkeling datadiensten*

Het delen van data is nog geen algemeen gebruik onder onderzoekers, maar met de politieke wens van *open data* (in navolging van *open access*) is het een kwestie van tijd dat onderzoeksfinanciers hieraan eisen gaan stellen. Ook de explosieve groei van data is algemeen erkend, zeker met de komst van talloze sensornetwerken. Uit de enquête van ePLAN blijkt dat wetenschappers een enorme groei verwachten in hun behoefte aan *data-resources*. DANS werkt in toenemende mate voor alle disciplines waarin behoefte bestaat aan datadiensten.



Opslag van, toegang tot en het delen van data moet nog beter ondersteund worden. Een belangrijk initiatief is GO-FAIR, om onderzoekers te helpen om data *Findable, Accessible, Interoperable and Reusable* (FAIR) te maken. Data, en in de eerste plaats big en open data, is een domeindoorsnijdende uitdaging die alle onderzoeksinfrastructuren raakt. Data verliezen hun betekenis zonder de software waarmee data gelezen, bewerkt of geïnterpreteerd kunnen worden. Daarom richt DANS zich ook steeds meer op aspecten van duurzaamheid van software.

#### 4.1.4 Authenticatie- en autorisatie-infrastructuur (AAI) (of 'cyber-security')

Onderzoekers werken steeds meer nationaal en internationaal samen. Onderzoeksfaciliteiten en data worden gedeeld. Om te bepalen of deze onderzoeker hiertoe toegang heeft, is het belangrijk om de identiteit vast te stellen. De identiteit moet betrouwbaar zijn en tegelijkertijd de privacy waarborgen, zeker als het gaat om toegang tot interne of externe online-diensten.

##### *Organisatie*

Hiertoe heeft SURF een authenticatie- en autorisatie-infrastructuur (AAI) ontwikkeld die ook voor docenten en studenten beschikbaar is. Onderzoekers loggen in met één account en hebben via *single-sign-on* veilig toegang tot diensten die door nationale en internationale partners en (commerciële) aanbieders beschikbaar worden gesteld, terwijl de uitwisseling van persoonsgegevens tot het absolute minimum is beperkt. Inmiddels zijn meer dan 130 clouddiensten aan SURFconext gekoppeld. Dit zijn diensten op het gebied van online samenwerking, storage, video en onderzoek, bijvoorbeeld Google Apps, N@tschool, Kluwer Navigator.

##### *Ontwikkeling AAI*

Ook voor gevoelige data, zoals eHRM, student-informatiesystemen en applicaties met patentgevoelige onderzoeksdata of privacygevoelige patiëntinformatie, maken instellingen steeds meer gebruik van clouddiensten. Voor deze diensten zijn sterkere vormen van authenticatie dan gebruikersnaam/wachtwoord gewenst, om de kans op beveiligingsincidenten te beperken.

Op dit moment is het account van de onderzoeker nog verbonden aan de instelling. Er wordt aan gewerkt om de onderzoeker in de toekomst een eigen betrouwbare identiteit te geven. De onderzoeker krijgt hierdoor meer vrijheid en kan onafhankelijk van zijn instelling gebruik maken van diensten, waardoor hij accounts eenvoudig kan meenemen als hij voor een andere instelling gaat werken. Ook blijft er nog veel aandacht nodig voor de privacy van de gebruiker en voor de vraag hoe je de betrouwbaarheid van de identiteit goed kunt waarborgen.

Internationale standaardisatie van AAI is noodzakelijk. Oude methoden voor het vaststellen van een identiteit middels certificaten zijn niet schaalbaar bij grote communities. AAI-oplossingen worden daarom ook steeds vaker toegepast bij grootschalige onderzoeksinfrastructuren. De digitale infrastructuren werken daarom nauw met deze communities samen om bestaande oplossingen te testen, op elkaar af te stemmen en te implementeren.

#### 4.1.5 eScience

De transitie naar data- en rekenintensief onderzoek vereist van onderzoekers om optimaal gebruik te kunnen maken van geavanceerde onderzoekssoftware. Waar voorheen veel onderzoekers hun dataverwerking en -analyses nog uitvoerden met relatief eenvoudige, lokale, hardware en software, zijn de vereisten de afgelopen jaren enorm toegenomen. Steeds makkelijker kunnen met behulp van onderzoekssoftware rekenfaciliteiten op afstand worden gebruikt. Ook de mogelijkheden van cloud-gebaseerde computing en data-diensten zijn enorm toegenomen.

##### *Organisatie*

Om de digitale infrastructuur en de inzichten uit de computer- en datascience beter te benutten hebben NWO en SURF in 2011 het Netherlands eScience Center opgericht. Het eScience Center fungeert als schakel tussen onderzoekers en bestaande digitale infrastructuur en ondersteunt onderzoekers bij het ontwikkelen van onderzoekssoftware om grote en complexe datasets te verwerken en te analyseren. Onderzoekers dienen bij eScience Center een projectaanvraag in. Bij toekenning ontvangen ze (een combinatie van) in-kind ondersteuning van eScience Research Engineers en in-cash bijdragen.

Het eScience Center coördineert op nationaal niveau ePLAN, het platform voor eScience/data research-organisaties. Op Europees vlak gebeurt dat via PLAN-E, waarin de eScience Centers van Europa samenkomen.

##### *Ontwikkeling eScience en onderzoekssoftware*

Vrijwel alle wetenschappelijke disciplines hebben in toenemende mate te maken met complexe uitdagingen wat betreft de analyse van grote hoeveelheden teksten, foto's, sensordata, klimaatdata, satellietdata, social media-data, databases of surveydata. Het effectief inzetten van de digitale infrastructuur en nieuwe inzichten uit de computer- en data science dragen bij aan de wetenschappelijke impact. In de groeiende behoefte aan ondersteuning hierbij moet worden voorzien.

Onderzoekssoftware wordt nu hoofdzakelijk op projectbasis ontwikkeld binnen het eScience Center in nauwe samenwerking met gebruikers. Er is behoefte aan een structurele voorziening om onderzoekssoftware breed beschikbaar te stellen, te onderhouden en door te ontwikkelen. Daartoe is in 2016 het eScience Technology Platform geïnitieerd. Dit online platform moet uitgroeien tot een nationale digitale software-infrastructuur, waar ook handleidingen en andere eScience- en softwaredocumentatie beschikbaar worden gesteld aan onderzoekers.

#### 4.1.6 Dienstverlening en Skills

Ondersteuning en expertise, human capital en dienstverlening zijn een integraal onderdeel van de hiervoor beschreven digitale infrastructuren. Zonder mensen met de juiste kennis en wetenschappelijke ervaring op het gebied van beheer, advies en toepassing van ICT kunnen wetenschappers geen gebruik maken van bestaande infrastructuur.

### *Organisatie*

Ondersteuning en expertise is bij alle organisaties belegd die zich bezighouden met digitale infrastructuur, dienstverlening voor en ondersteuning van onderzoekers. Dit gebeurt op lokaal niveau door universiteiten en onderzoeksinstellingen en op nationaal niveau bij de aanbieders van nationale digitale infrastructuur en (eScience) diensten, zoals DANS, SURF en eScience Center.

SURF ondersteunt onderzoekers bij het kiezen van de juiste digitale hulpmiddelen en organiseert trainingen over uiteenlopende onderwerpen, zoals actuele ICT-thema's en hands-on systeemtrainingen. SURF werkt daarin samen met onder andere het eScience Center en NIKHEF in het Support4Research programma.

Waar het bij SURF doorgaans om korte adviestrajecten gaat, levert het eScience Center intensieve en hoogwaardige eScience ondersteuning, zoals wetenschappelijke ondersteuning en advies voor de inzet van digitale infrastructuur en het ontwikkelen van onderzoekssoftware. Ook biedt het eScience Center samen met SURF trainingen aan voor wetenschappers op het gebied van programmeren, data-analyse en gebruik van computingfaciliteiten.

DANS verzorgt training en consultancy en doet onderzoek naar duurzame toegang tot digitale informatie. Bij 4TU.Centre for Research Data kan de onderzoeker in alle stadia van zijn onderzoek terecht voor advies en ondersteuning met betrekking tot databeheer. Dit reikt van het opzetten van een datamanagementplan, het faciliteren van een data-lab tot het uiteindelijk deponeren van de data in het archief.

### *Ontwikkeling*

De Open Science Cloud zal een prominente plaats gaan innemen in de onderzoekswereld. Onderzoekers krijgen behoefte aan ondersteuning bij open research, open notebooks, open source, open data, open access en data stewardship. Ondersteuning van onderzoekers is nog volop in ontwikkeling, ook bij het gebruik van digitale infrastructuur en het beschikbaar maken van onderzoekssoftware. Er ligt een uitdaging om de ondersteuning schaalbaar te maken en ervaringen uit de ene discipline toe te passen in een andere. Belangrijk is in ieder geval dat de onderzoeker centraal staat en dat er een multidisciplinaire aanpak komt.

#### 4.1.7 Internationale digitale infrastructuur

Hoewel internationale digitale infrastructuur geen onderdeel zijn van de nationale infrastructuur, volgt hieronder toch een korte beschrijving van de belangrijkste Europese infrastructuur. Andersom is het namelijk wel het geval: de nationale infrastructuur maken deel uit van een grotere internationale infrastructuur, die noodzakelijk zijn voor het faciliteren van internationale samenwerking tussen onderzoekers. Er is een bijdrage van de nationale organisaties vereist, wat dus integraal onderdeel uitmaakt van de totale begroting van de nationale digitale infrastructuur.

## **GÉANT**

GÉANT beheert een Europese backbone tussen de 41 nationale onderzoeksnetwerken (NREN's). De NREN's werken binnen GÉANT ook samen aan innovatie, standaardisatie, harmonisatie, AAI en nieuwe dienstverlening. SURF is partner in het project. SURF draag voor de exploitatie van het netwerk bij volgens een model dat gebaseerd is op de afname van bandbreedte en op het Bruto Binnenlands Product.

## **PRACE**

PRACE biedt Europese onderzoekers in competitie toegang tot de grootste Europese HPC-systemen (Tier-0). Op dit moment zijn er enkele Tier-0-systemen: in Duitsland, Frankrijk, Italië, Spanje en Zwitserland. Onderzoekers kunnen alleen toegang krijgen als er door hun land ook nationale rekenfaciliteiten van een minimale omvang (tier-1) in het project worden ingebracht én het nationale rekencentrum een financiële bijdrage levert volgens model dat gebaseerd is op basis van historisch gebruik en het Bruto Binnenlands Product. SURF brengt de nationale supercomputer in.

## **EGI**

EGI biedt geavanceerde computerdiensten voor wetenschappers, multinationale projecten en research infrastructures. EGI is voortgekomen uit de grid-infrastructuur, die is gebouwd voor de verwerking van de data van de LHC-experimenten. EGI is een federatie van meer dan 300 data en computing datacenters verspreid over 22 landen van Europa en de rest van de wereld. In Nederland zijn SURF, Nikhef en het RUG-CIT onderdeel van de EGI-faciliteit. EGI biedt diensten op het gebied van (gefedereerde) high throughput computing, storage en cloud. Het hoofdkantoor van EGI is gevestigd in Amsterdam.

## **EUDAT**

EUDAT biedt generieke datadiensten, opslag en support voor data aan meerdere onderzoekscommunities, maar ook aan individuele onderzoekers, via de 35 Europese partners. De data wordt opgeslagen in vijftien Europese landen, nabij de krachtigste supercomputers van Europa. EUDAT zorgt voor toegang en opslag, van informeel data delen tot langetermijn-archivering, identificatie en vindbaarheid van zowel *long-tail* als big data gedurende de hele life cycle van onderzoeksdata. SURF, DANS en KNMI zijn partner in EUDAT.

## 4.2 Nationale enquête digitale infrastructuur

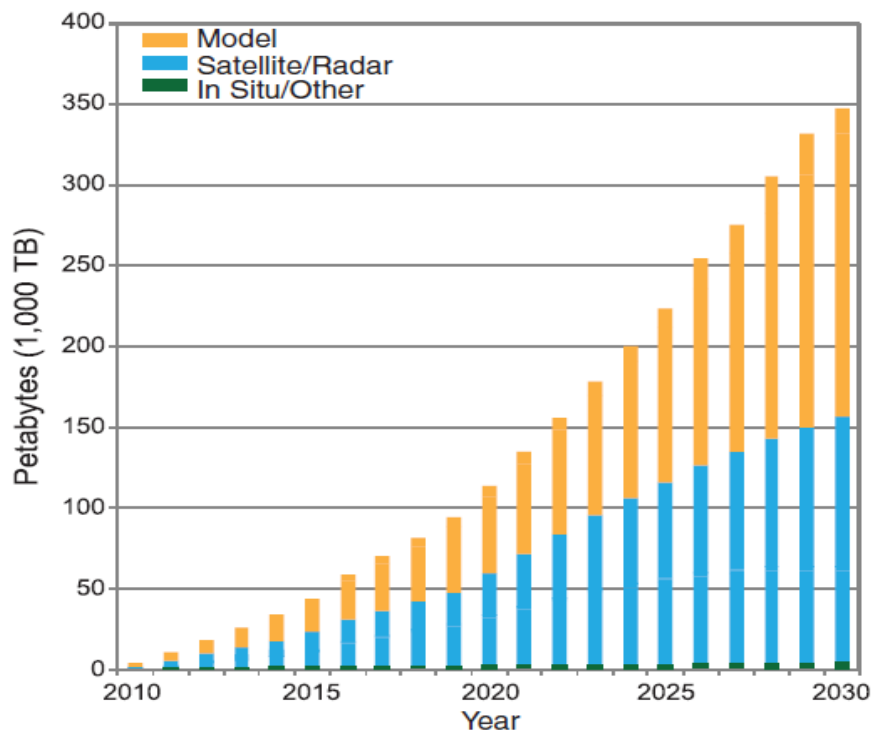
Om een goed beeld te krijgen van de huidige en toekomstige behoeften aan digitale infrastructuur van wetenschappers, is eind 2015 is door het nationaal platform van eScience- en data science organisaties een enquête gehouden onder meer dan 1000 wetenschappers (UD, UHD en hoogleraren) verspreid over alle disciplines, waarbij 242 respondenten namens hun onderzoeksgroep reageerde. In aanvulling hierop zijn in 2016 vier gerichte workshops gehouden onder wetenschappers uit de Life Sciences, Humanities, Physics en Environment.

De meest relevante conclusies voor dit advies worden hieronder weergegeven.

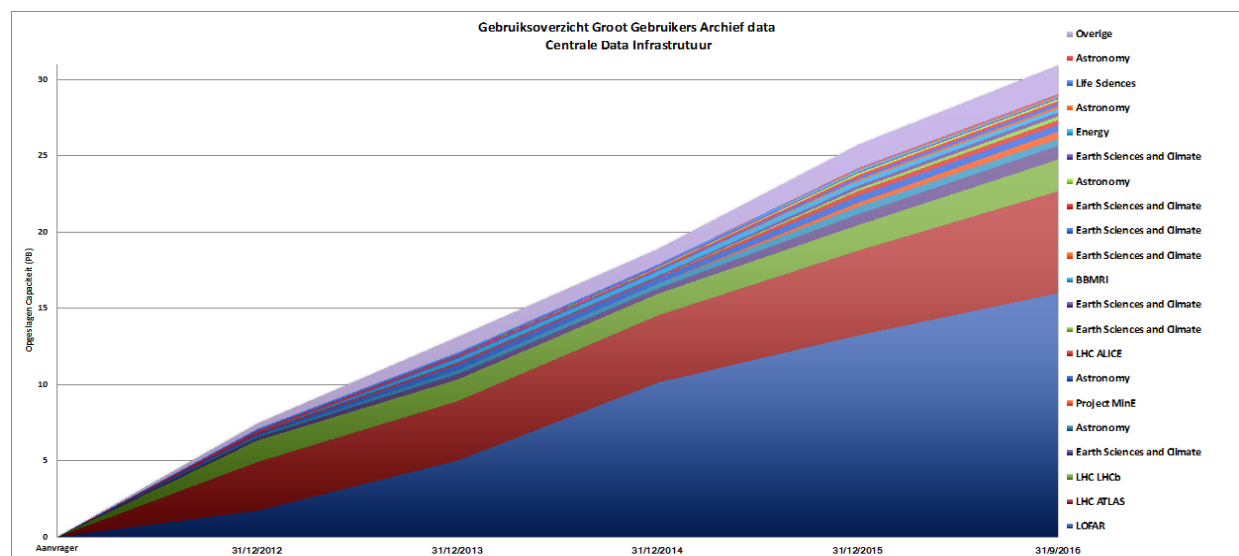
- De ontwikkelingen in het onderzoeksdomein worden door een overzienbaar aantal kernwoorden gekarakteriseerd: vooral meer data, maar ook rekenen, speciaal ook meer echt *grootschalig* rekenen, meer beeldverwerking en visualisatie, meer wensen voor (ondersteuning bij) analyse van data en zorgen over security en privacy;
- Er vindt veel multidisciplinaire samenwerking plaats, met een sterk accent op internationale samenwerking;
- De toename en het belang van en het volume aan data bij het onderzoek betreft alle disciplines en alle types data;
- Er is een concrete verwachte enorme groei in behoefte aan data-resources (naar het Exabytes domein) en –diensten;
- Grote zorgen worden geuit over het algemeen achterblijven van het investeringsniveau van de nationale ICT-infrastructuur in relatie tot de sterk toegenomen behoefte aan alle soorten resources en in vergelijking met het aanbod elders in de wereld waarmee men in competitie is. Ook het gebrek aan reserveringen voor toekomstige investeringen, onder meer in een nieuwe nationale supercomputer is reden voor die zorg;
- De afstand van een nationale HPC-voorziening tot de internationale top zou niet groter moeten zijn dan een factor 10 voor onderzoekers om op competitief niveau te kunnen meekomen. Dit ook nodig om (recht op) toegang te houden tot Europese voorzieningen, zoals PRACE, de Europese infrastructuur voor supercomputers;
- Van de meerderheid die aangeeft meer te zullen gaan rekenen geeft een derde ook aan dat dit geheugenintensief werk betreft.
- Er is, naast een supercomputer, behoefte aan een robuuste tweede laag van clustervoorzieningen, centraal of gecoördineerd bij de instellingen;
- Over het algemeen zou de beschikbaarheid van een nationale clouddienst, of een op Europese grondslag, worden gewaardeerd, in verband met de verwachting dat privacy en security beter kunnen worden geregeld dan via commerciële diensten;
- Er is een grote wens naar verbetering van het gebruiksgemak van de ICT-infrastructuur en de toegang daartoe.
- Ondersteuning door partijen als SURF en het eScience Center wordt genoemd als waardevolle en noodzakelijke voorziening voor een efficiënte inzet van ICT-infrastructuur in het onderzoek;
- Ruim meer dan de helft van de respondenten geeft aan nog nooit gehoord te hebben van de termen data stewardship, research datamanagement en/of software sustainability. Maar wel vindt 90% (!) deze onderwerpen van groot belang, groot genoeg dat hiervoor landelijk beleid op zijn plaats is (80%).

### 4.3 Groei wetenschappelijke data

(voor opmaak: bijgaande illustraties ter illustratie door hele stuk heen of als 'mozaiek' op 1 of 2 pagina's)



Figuur 2 Wereldwijde omvang klimaatdata (Science, 2011)



Figuur 3 SURFSara Centrale Data Infrastructuur. Gebruiksoverzicht groot gebruikers archief data

## 4.4 E-Needs questions in ESFRI submission format 2018

### A) E-NEEDS questions for new proposals

4.1. Outline the Data Management Plan (DMP) and data access policy of the RI. If applicable, describe how data would become accessible to the public: (maximum 3000 characters with spacing)

4.2. Describe and quantify what e-infrastructure services - e.g. resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc. - your RI will need: (maximum 2000 characters with spacing)

4.3. Describe how the e-infrastructure services needed by your RI will be implemented, specifying the potential need of external e-infrastructure resources and the relations to external e-infrastructures: (maximum 2000 characters with spacing)

4.4. Describe how the RI will contribute to the development of the European and global e-infrastructure landscape at all levels (institutional, regional, national, international) - including e.g. the e-infrastructure commons and the European Open Science Cloud (EOSC): (maximum 2000 characters with spacing)

### B) E-NEEDS questions for 2008 & 2010 PROJECT MONITORING

4.1. Summarise the data management, data access and data security policies of your RI - including procedures for ensuring the sustainability of data – and clarify to what degree they are approved: (maximum 3000 characters with spacing)

4.2. Describe the technical design and operations of the e-infrastructure services -e.g. resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc. - for your RI and clarify to what degree it is approved *and implemented*: (maximum 2000 characters with spacing)

4.3. Describe the use of external e-infrastructure resources and the relations to external e-infrastructures - including agreements with external partners delivering core e-infrastructure services for your RI – and clarify to what degree this use is approved within your consortium: (maximum 2000 characters with spacing)

4.4 Describe how the RI contributes to the development of the European and global e-infrastructure landscape at all levels (institutional, regional, national, international) - including e.g. the e-infrastructure commons and the European Open Science Cloud (EOSC): (maximum 2000 characters with spacing)

4.5 If available, upload the approved Data Management Plan (DMP): (upload with limit 1 MB)

**C) E-NEEDS questions for LANDMARKS**

4.1. Summarise the data management, data access and data security policies of your RI - including procedures for ensuring the sustainability of data – and clarify to what degree they are approved *and implemented*: (maximum 3000 characters with spacing)

4.2. Describe the technical design and operations of the e-infrastructure services -e.g. resources for storage, computing, networking, tools for data management, security, access, remote analysis, etc. - for your RI and clarify to what degree it is approved *and implemented*: (maximum 2000 characters with spacing)

4.3. Describe the use of external e-infrastructure resources and the relations to external e-infrastructures - including agreements with external partners delivering core e-infrastructure services for your RI – and clarify to what degree you have *signed agreements*: (maximum 2000 characters with spacing)

4.4. Describe how the RI contributes to the development of the European and global e-infrastructure landscape at all levels (institutional, regional, national, international) - including e.g. the e-infrastructure commons and the European Open Science Cloud (EOSC): (maximum 2000 characters with spacing)

4.5 Upload the approved Data Management Plan (DMP): (upload with limit 1 MB)



## 4.5 Verklarende woordenlijst

### **BiG Grid zelfevaluatie**

Ten behoeve van de eindevaluatie van het project Big Grid is een zelfevaluatie gedaan volgens het beoordelingskader van de Impuls Grootschalige Onderzoeksfaciliteiten.

### **Compute Canada**

Compute Canada faciliteert de versnelling van onderzoek en innovatie door het aanbieden van state-of-the-art computing systemen, opslag en software aan onderzoekers. Compute Canada voorziet Canadese onderzoekers en hun partners in andere sectoren van hoogwaardige rekenfaciliteiten en ondersteuning. Voor meer informatie: <https://www.computecanada.ca/home/>

### **European Open Science Cloud**

De European Open Science Cloud beoogt een geïntegreerde Europese Infrastructuur te realiseren inclusief toegangs-, besturings- en financieringsmechanismen, die optimaal hergebruik van onderzoeksdata vergroot en open science in den brede faciliteert. Voor meer informatie: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

### **GO-FAIR initiatief**

GO-FAIR is een door Nederland gecoördineerd internationaal initiatief dat een revolutionaire transformatie ondersteunt van Open, data intensieve wetenschap en innovatie, met een wereldwijde maatschappelijke impact. Het beoogt een schaalbare implementatie omgeving om een Internet of FAIR data en diensten te realiseren, waar analyse-tools gecombineerd kunnen worden met relevante data omdat zowel de tools als de data *Findable*, *Accessible*, *Interoperable* en *Reusable* (FAIR) zijn.

### **ICTRegie advies (2008)**

ICTRegie bracht onder verantwoordelijkheid van NWO in 2008 op verzoek van de ministeries Onderwijs, Cultuur en Wetenschap en Economische Zaken advies uit over de digitale infrastructuur in Nederland. Het advies was tweeledig: ICTRegie deed aanbevelingen voor een nieuwe governance, waarbij SURF de regie kreeg over de digitale infrastructuren. Daarnaast deed ICTRegie aanbevelingen voor investeringen in de digitale infrastructuur.

### **Landelijk Coördinatiepunt Research Data Management (LCRDM)**

Het Landelijk Coördinatiepunt Research Data Management zorgt voor de nationale coördinatie rondom research datamanagement-beleid en faciliteert daarmee een landelijke aanpak van research datamanagement in Nederland. Nauwe samenspraak met onderwijs- en onderzoeksinstituten is daarbij cruciaal. Zo kan Nederland tot een efficiënte en effectieve ontwikkeling en uitvoering van research datamanagement beleid komen. SURF vervult deze coördinerende rol, de VSNU-stuurgroep onderzoek en valorisatie vervult een klankbordfunctie.

Samen met de instellingen zijn 5 vraagstukken geïdentificeerd die het Landelijk Coördinatiepunt in werkgroepen aanpakt: faciliteiten en data-infrastructuur, juridische aspecten en zeggenschap, financiën, onderzoeksondersteuning en advies en bewustwording voordelen.

De werkgroepen van het LCRDM, elk met experts uit het veld, brengen deze vraagstukken verder, en ondersteunen zo de ontwikkeling van landelijk research datamanagement-beleid. Dit gebeurt in nauwe en sensitieve afstemming met alle relevante stakeholders en research communities en in aansluiting op actuele ontwikkelingen op Europees Niveau. Zie voor meer informatie: [www.lcrdm.nl](http://www.lcrdm.nl)

### **Nationaal Plan Open Science (NPOS)**

Het Nationaal Plan Open Science is een nationaal plan waarin de ambitie van Nederland, de huidige activiteiten en de mogelijk vervolgstappen voor open science staan beschreven. Door middel van het ondertekenen van het Nationaal Plan Open Science zullen alle stakeholders hun intentie uitspreken gezamenlijk verder te werken aan de open science ambities.

### **Research Data Netherlands (RDNL)**

Research Data Netherlands is een samenwerkingsverband tussen de dienstverleners 4TU.Centre for Research Data, Data Archiving and Networked Services (DANS) en SURF. Met deze coalitie, die ook voor andere partijen open staat, bundelen de drie data-archieven hun krachten op het gebied van duurzame data-archivering. Voor meer informatie: [www.researchdata.nl](http://www.researchdata.nl)

### **Taskforce NWO (2010)**

De Taskforce NWO uit 2010 was een vervolg op het advies van ICTRegie. Enerzijds beschreef de Taskforce een versoberd plan ten opzichte van het ICTRegie advies, anderzijds bevatte het een voorstel hoe de door het kabinet onderschreven plannen van ICTRegie voor de infrastructuur gefinancierd zouden kunnen worden. NWO, het ministerie van Onderwijs, Cultuur en Wetenschap, het ministerie van Economische Zaken en SURF maakten deel uit van de Taskforce.

## 4.6 Achtergrondliteratuur

- AWTI (2015). *Klaar voor de toekomst. Naar een brede strategie voor ICT.*
- AWTI (2016). *Houd de basis gezond. Prioriteiten voor extra investeringen in onderzoek en innovatie.*
- Commission High Level Expert Group on the European Open Science Cloud (2016). *Realising the European Open Science Cloud.*
- Deloitte (2016). *Dutch Digital Infrastructure 2016: Enabling the Digital Economy and Society.*
- ePLAN (2016). *Infrastructuur duurzaam op maat. Een analyse van de vraag naar ICT-infrastructuur en services in Nederland.*
- ICTRegie (2008). *Verkenning naar de gewenste ICTinfrastructuur voor het onderzoek in Nederland.*
- Kenniscoalitie (2015). *De Nationale Wetenschapsagenda: Vragen, verbindingen, vergezichten.*
- KNAW (2016). *KNAW-Agenda grootschalige onderzoeksfaciliteiten.*
- KNAW (2012). *Digitale geletterdheid in het voortgezet onderwijs.*
- National Academies of Sciences, Engineering, and Medicine (2016). *Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020.*
- Nederland ICT (2016). *Groei door digitalisering.*
- NL next level (2016). *Investeren in een digitale kwantumsprong.*
- NWO (2016). *Nationale Roadmap Grootschalige Wetenschappelijke Infrastructuur.*
- OCW (2017). *National Plan Open Science.*
- Platform Onderwijs 2032 (2016). *Ons onderwijs 2032. Eindadvies.*
- Staatssekretariat für Bildung, Forschung und Innovation SBFI (2015). *Schweizer Roadmap für Forschungsinfrastrukturen im Hinblick auf die BFI-Botschaft 2017-2020.*
- Taskforce Financiering Wetenschappelijke ICT-Infrastructuur (2010). *Hoe de nationale ICT-infrastructuur voor research en hoger onderwijs te financieren.*
- VSNU (2016). *De Digitale Samenleving. Nederland en zijn universiteiten: internationale pioniers in mensgerichte informatietechnologie.*

## 4.7 Samenstelling van de adviescommissie ICT-infrastructuren

### **Voorzitter**

Sally Wyatt (Maastricht University en Huygens ING, KNAW)

### **Commissieleden**

Patrick van Beers (op persoonlijke titel)

Henk Dijkstra (Universiteit Utrecht)

Erik Fledderus (SURF)

Bernard de Geus (Virtual Laboratory Plant Breeding)

Wilco Hazeleger (Netherlands eScience Center)

José de Kruif (Universiteit Utrecht)

Joeri van Leeuwen (ASTRON)

Andrea Scharnhorst (DANS)

Ana-Lucia Varbanescu (Universiteit van Amsterdam)

Jacob de Vlieg (Technische Universiteit Eindhoven)

### **Dankwoord**

De adviescommissie is veel dank verschuldigd aan Tom Bakker (eScience Center) en Ilse Koning (SURF) voor het vele werk dat zij verricht hebben bij het voorbereiden van dit rapport. Ook willen wij Yvette Tuin (NWO) bedanken voor haar ondersteuning van de activiteiten van de adviescommissie.