

Toegankelijke en Verantwoorde Waarde Creatie uit Big Data

Evolutie van de technologie, revolutie in de toepassingen

De impact van de explosief groeiende hoeveelheid en diversiteit van gegevensbestanden – Big Data – is zichtbaar in alle sectoren van de samenleving: van wetenschap tot economie, van gezondheid tot overheid, van de financiële sector tot elk facet van het dagelijks leven. Big Data is niet alleen “veel data”; het betekent vooral nieuwe soorten en continue stromen van gegevens, die nieuwe eisen stellen aan infrastructuur en technieken, aan mensen die met deze data werken, en aan de samenleving die ze implementeert. Pas sinds kort hebben we voldoende rekenkracht, computergeheugen, netwerkverbindingen, dataopslag, en vooral slimme algoritmen, analysetechnieken en software om grote gegevensbestanden te kunnen analyseren en de resultaten daarvan te kunnen gebruiken. De urgentie van het onderwerp blijkt ook uit de Nationale Wetenschapsagenda (NWA) zélf; in alle domeinen van de agenda speelt Big Data een rol in de vragen. De route “*Big Data Verantwoord Gebruiken: Zoeken naar Patronen in Grote Gegevensbestanden*” brengt deze vragen samen. Omdat bijna alle andere routes van de NWA afhankelijk zijn van Big Data, hebben investeringen in deze route de hoogste prioriteit. Het is niet voor niets dat een serie van recente adviezen aan de overheid (zoals het AWTI advies “Klaar voor de toekomst”) expliciet stilstaat bij het belang van investeringen in ICT en data science.

Wat betekent Big Data als gamechanger?

De revolutionaire kracht van Big Data ligt in het combineren van een veelheid aan verschillende gegevens. Dit lukt alleen als databestanden open en toegankelijk zijn. Dat betekent echter wel dat we verantwoord met deze gegevens moeten omgaan. Ons doel voor het komende decennium is daarom het ontwikkelen van de juiste instrumenten en infrastructuur voor toegankelijke en verantwoorde waarde creatie vanuit Big Data voor economische, maatschappelijke en wetenschappelijke domeinen.

Hiervoor is een aantal acties noodzakelijk. Bestaande en nieuwe analysetechnieken en automatische beslismethoden die werken met verschillende soorten data (bijvoorbeeld transacties, gesproken of geschreven teksten, beelden, metingen) moeten worden ontsloten via een grote gevarieerde toolbox van software, analysemethoden en de daarbij behorende infrastructuur en kunde (eScience). Experimenteeromgevingen zijn nodig om partijen te leren (samen te) werken met data. Er moet een infrastructuur worden opgebouwd waarin data open, toegankelijk en interoperabel kunnen worden verzameld, opgeslagen, geanalyseerd en gedeeld. Ook de mogelijkheden voor datamarkten, waarin burgers en bedrijven rechten op gebruik van data verhandelen moeten worden onderzocht. Wet- en regelgeving wordt geëvalueerd en geactualiseerd om maximaal meerwaarde uit de combinatie van gegevens mogelijk te maken, met inachtneming van de privacy van burgers. Tot slot zullen de essentiële elementen van data science in alle niveaus van het onderwijs worden opgenomen, om de benodigde tienduizenden data science specialisten te kunnen opleiden.



Drie invalshoeken op de route Big Data die geleid hebben tot de gamechanger “toegankelijke en verantwoorde waarde creatie uit Big Data”.

Deze waarde creatie moet op een verantwoorde manier gebeuren. Het gezamenlijk zoeken naar de balans tussen wat met Big Data kan, mag en wenselijk is, is dé gamechanger in Big Data. Democratische waarden als gelijkheid, openheid, juistheid, volledigheid, betrouwbaarheid en vertrouwelijkheid moeten verankerd zijn in het ontwerp van de Big Data instrumentatie. Toegepaste algoritmen moeten transparant zijn, en geen politieke, regionale, etnische, of gender kleuring hebben. Al deze aspecten moeten aandacht krijgen in huidig en toekomstig onderzoek.

Impact van de gamechanger

Big Data heeft een belangrijke impact op vier maatschappelijke terreinen.

Kwaliteit van leven. Door te zorgen voor meer transparantie, goede bescherming van privacy, en minder foutieve of oneerlijke beslissingen kan Big Data de kwaliteit van leven vergroten in de zorg, in het verkeer, in slimme steden, en in vrijetijdsbesteding. De digitale weerbaarheid van de bevolking zal ook worden vergroot.

Urgente politieke en overheidsvraagstukken. De kwaliteit en efficiëntie van de Nederlandse overheid kan aanzienlijk verbeteren door slim gebruik te maken van de mogelijkheden van Big Data in urgente uitdagingen als efficiëntie in de zorg, digitale dienstverlening aan burgers, opsporen van (cyber)criminaliteit, mobiliteit, versterken van milieudoelstellingen en internationale conflicten.

Economie. Nieuw te ontwikkelen methoden, modellen en technieken op het terrein van Big Data vormen een basis voor economische en intellectuele exportproducten. Het openstellen van publieke data ten behoeve van gemeenten, energievoorziening, openbaar vervoer, en onderwijs creëert nieuwe economische kansen. Het is essentieel voor onze economie om waarde creatie in Nederland en de Europese Unie te borgen in het licht van de langjarige effecten van opkomende (niet-EU) data-monopolies en bijbehorende interventiestrategieën.

Beroepsbevolking en werkgelegenheid. Met het oog op de toenemende rol van Big Data, automatische algoritmen en robotisering in allerlei sectoren, is het versterken van de data science competentie noodzakelijk voor de concurrentiekracht van het Nederlandse bedrijfsleven en het toekomstig groeivermogen van de Nederlandse economie.

Welke rol speelt Nederland in Big Data?

Nederland is bij uitstek geschikt om het voortouw te nemen in onderzoek naar Big Data. Europa is zeer gericht op open wetenschap en open innovatie. Ons land heeft een sterke uitgangspositie met veel internationaal gerenommeerde onderzoekers op data science-gerelateerde terreinen als wiskunde, informatica, statistiek, economie en taalwetenschappen. Daarnaast vraagt het toenemend gebruik van verspreide gegevensbestanden maximale samenwerking tussen wetenschappers, bedrijven en maatschappelijke spelers. Het Nederlandse poldermodel werkt door in het ontsluiten van gegevensbestanden. Meer dan in landen met data monopolies ontstaat samenwerking tussen verschillende data eigenaren en afnemers, elk met verschillend belang om data te delen; deze Nederlandse aanpak biedt kansen voor export. In (topsector)plannen als Commit2Data en eHumanities.NL komen deze partijen in toenemende mate samen.

Nederland heeft een wereldwijd uniek groot en actief ecosysteem op het gebied van Big Data onderzoek en toepassingen. SURF voorziet in de nationale e-infrastructuur (netwerk, dataopslag, rekenen) voor onderwijs en onderzoek. Het Netherlands eScience Center (NLeSc) ontsluit digitale technologie op basis van de vraag uit onderzoektoepassingen. Veel universiteiten en sommige HBOs hebben in de afgelopen jaren data science centra opgericht. Onderwijsprogramma's komen langzamerhand van de grond. En verschillende nieuwe samenwerkingsnetwerken zijn onlangs ontstaan, zoals Data Science Platform Nederland (DSPN), Big Data Alliance (BDA), Big Data Value Association (BDVA) en ePLAN. Ook de ministeries, regionale overheden, rijkskennisinstellingen en andere organisaties ontwikkelen samenwerkingsverbanden om Big Data toepassingen te versterken.

Tot slot heeft Nederland één van de beste digitale infrastructuren van Europa, ook wel de digitale mainport genoemd. Het verantwoord energiezuinig ontwikkelen van deze mainportfunctie is van groot belang voor Nederland. Data centers zijn nu al verantwoordelijk voor 4 procent van het nationale elektrisch energieverbruik.

Waarom en waarin investeren?

De combinatie van verantwoord gebruik en de vergroting van toegankelijkheid van Big Data vereist dat universiteiten, kennisinstellingen, infrastructuur providers, overheden, bedrijven en burgerorganisaties intensief samenwerken. Veel van de huidige Nederlandse samenwerkingen op het terrein van Big Data zijn relatief kleinschalig, en ontstaan binnen bredere onderzoeksprogramma's (NWO, NLeSc, Horizon2020). Ambitieuze programma's met focus zoals eHumanities-NL (big data in geesteswetenschappen), Commit2Data (big data in topsectoren) en diverse big data life sciences programma's krijgen te weinig geld om de beoogde impact te bereiken, en richten zich nauwelijks op de synergie van de twee kenniselementen in de gamechanger: verantwoorde en drempelverlagende waarde creatie uit Big Data. Verbrokkeling van financieringsbronnen en budgetten maken slagvaardig opereren moeizaam. Om bruggen te kunnen slaan tussen juridische, ethische, culturele, sociale, (bedrijfs)economische, informatiekundige, en technologische aspecten, en om de slagkracht van de Nederlandse Big Data inspanningen aanzienlijk te vergroten, is een extra kennisinvestering van 60 M€ per jaar noodzakelijk. Minstens een kwart hiervan zal naar verwachting uit private middelen komen. Deze kennisinvestering valt uiteen in drie delen.

- 20 M€ per jaar voor het stimuleren van waarde creatie in toepassingsdomeinen. Hiermee worden proeftuinen en grootschalige privaat-publieke valorisatieprojecten opgezet. Het Big Data ecosysteem wordt duurzaam versterkt door toolbox projecten, technologie transfer initiatieven, personele mobiliteit, en multidisciplinaire disseminatie.
- 20 M€ per jaar voor multidisciplinair onderzoek. Met dit budget wordt het onderwijs en wetenschappelijke fundament van Big Data versterkt en verbreed in de richting van data science, technologisch, juridisch, sociaal en economisch onderzoek. Universiteiten en HBO's trekken nieuw (internationaal) talent aan – van studenten en promovendi tot postdocs en wetenschappelijk personeel – die zich gezamenlijk op de uitdaging richten van het verantwoord ontsluiten van Big Data.
- 20 M€ per jaar om bestaande infrastructuur en ontsluiting daarvan voor toepassingen in Nederland blijvend te vernieuwen. Deze investering wordt aangewend om met mensen en systemen de digitale mainport functie te versterken richting Big Data onderzoek en toepassing. Dit omvat zowel investeringen in high-end SURF faciliteiten (denk aan up-to-date gecentraliseerde rekenkracht, snelle netwerken, ene gespecialiseerde data stewards) als investeringen in open en interoperabele infrastructuren die ondersteunend zijn voor Big Data in de volle breedte van wetenschap en economie (bijvoorbeeld via het NLeSc dat domein-specifieke gegevensbestanden en analyses ondersteunt).

Relatie tot de NWA clustervragen

In onderstaande tabel zijn de relaties tot de NWA clustervragen compact samengevat.

Waarde creatie	Vergroten van toegankelijkheid	Verantwoorde manier
008 (klimaat)	112 (grote dataset en benutten daarvan)	036 (gevolgen maatschappelijke veranderingen als digitalisering)
064 (sociale wetenschappen)	115 (bouwen en onderhouden van software van morgen)	057 (balans tussen informatievrijheid en privacy)
081 (genetica)	123 (onvoorspelbaarheid van complexe netwerken)	114 (lager energieverbruik)
098 (biomedische)		
105 (zorg)		
109 (effectiviteit bestuur en rechtstaat)		
111 (erfgoed)		
124 (stromingsleer)		
131 en 132 (astronomie)		