



WAGENINGEN
UNIVERSITY & RESEARCH

Plan van aanpak voor het Wageningen Data Competence Center 2018

Openbare versie

Wageningen Data
Competence Center

DATUM
5 maart 2018

AUTEUR
Willem Jan Knibbe

VERSIE
1.0

STATUS
Definitief

Inhoudsopgave

1	Inleiding	5
2	Visie en missie	7
2.1	Visie	7
2.2	Missie	7
3	Doelstellingen	9
4	Strategie en organisatie	11
4.1	Strategie	11
4.2	Organisatie	11
5	De rol van het WDCC binnen WUR	12
5.1	Aanleiding	12
5.2	Ambitie	12
5.3	Scope	13
5.4	Mandaat	13
5.5	Rol van het WDCC	13
6	Research data infrastructuur	15
6.1	Functie	15
6.2	Beleid	15
6.3	Voorzieningen	16
6.4	Communities	16
6.5	Toepassingen	16
6.6	Ontwikkeling en doel	16
7	Programma	17
7.1	Onderzoek	17
7.2	Onderwijs	18
7.3	Waarde creatie	18
7.4	Big data management	19
7.5	Specifieke infrastructuur	19
7.6	Samenhang programmaliijnen	20

1 Inleiding

Dit document beschrijft de aanpak van het Wageningen Data Competence Center (WDCC). Het WDCC is opgericht op basis van het Advies (2016)¹ en het Implementatieplan (2017)², beide goedgekeurd door de Raad van Bestuur. Dit Plan van aanpak is een nadere uitwerking van het Implementatieplan en dient als basis voor de werkzaamheden van het WDCC en voor de verdere werving van het team dat de uitvoering zal begeleiden.

Het WDCC is het antwoord van Wageningen University & Research (WUR) op de zeer snel groeiende hoeveelheid digitale gegevens die beschikbaar is voor onderzoek (big data). Het Advies en het Implementatieplan volgend is dat antwoord niet het oprichten van een apart instituut, een leerstoelgroep, een onderzoeksschool of een andere organisatievorm die op zichzelf staat naast de bestaande organisatie. Het WDCC is daarentegen gericht op de integratie en versterking van de bestaande organisatie van onderwijs en onderzoek.

De hier gepresenteerde aanpak neemt nauwe samenwerking met de kennisenheden en stafafdelingen als uitgangspunt. Voor het onderwijs in de vaardigheden die nodig zijn om onderzoek met big data te kunnen doen (data science) wordt voortgebouwd op de sterke punten die aanwezig zijn in de organisatie. Voor verdere versterking van het onderwijsprogramma dat hieruit volgt zal samenwerking gezocht worden met instituten die aanvullende mogelijkheden kunnen bieden. Door op deze manier studenten op te leiden met de vaardigheden voor de toekomst wordt gewerkt aan het creëren van waarde in de samenleving.

Voor het onderzoek met big data wordt ook aangesloten bij de vele lopende initiatieven op dit gebied. In aanvulling hierop wordt de uitwisseling van kennis en het combineren van verschillende typen data georganiseerd. Beide hebben niet alleen als doel het verhogen van de kwaliteit en efficiëntie van het onderzoek, maar ook het ontwikkelen van nieuwe mogelijkheden, concepten en diensten, gebaseerd op integratie van het brede scala van expertise dat aanwezig is bij WUR. Centrale vraag bij deze nieuwe mogelijkheden is de vraag hoe daarmee duurzaam waarde kan worden gecreëerd. Als aanvulling op het eigen onderzoek wordt gericht samenwerking gezocht met externe partners.

De ambities voor onderwijs en onderzoek vergen een goede infrastructuur. Voor de realisatie hiervan zal een combinatie gezocht worden van eigen middelen, externe middelen, partners en internationale samenwerkingsverbanden. Ten grondslag hieraan ligt weloverwogen beleid op het gebied van datamanagement. Dit beleid wordt waar nodig vormgegeven en wanneer aanwezig uitgedragen. Als uitgangspunt voor zowel infrastructuur als datamanagement worden waar mogelijk de principes van open source code, open data, Fairness-Accuracy-Confidentiality-Transparency (FACT) voor onderzoek en Findable-Accessible-Interoperable-Reusable (FAIR) voor data toegepast.

Het WDCC bestaat uit een klein team dat zich richt op onderwijs, onderzoek, waarde creatie, datamanagement en infrastructuur. De ambities weergegeven in dit plan van aanpak gaan verder dan dit team en betreffen de hele organisatie van WUR. Daarom is de hier gepresenteerde aanpak een succes als we als One Wageningen de samenwerking zoeken voor het realiseren van deze ambities.

¹ Advies Wageningen Data Competence Center, WUR, december 2016.

² Implementatieplan WDCC, WUR, maart 2017.

2 Visie en missie

2.1 Visie

Nieuwe en snel opkomende technologieën veranderen de wereld van onderwijs en onderzoek. Het opkomen van big data door de ontwikkeling van internet, mobiele communicatie, toenemende rekenkracht van computers, local computing en internet of things, is zo'n nieuwe en snel opkomende technologie. Big data biedt in combinatie met machine learning nieuwe mogelijkheden voor onderzoek. Het vakgebied data science dat zo is ontstaan vraagt om aanpassingen in het onderwijs. Samen levert dit nieuwe kansen voor het creëren van waarde. Tegelijkertijd leveren nieuwe ontwikkelingen ook nieuwe vraagstukken op aan de organisatie. In dit geval betreft dat vooral datamanagement en de infrastructuur van hardware en software.

Wereldwijd richten universiteiten en onderzoeksinstituten zich op deze ontwikkelingen door het organiseren van aparte instituten en het aanbieden van nieuwe opleidingen. Een mooi voorbeeld hiervan is het Berkeley Institute for Data Science³ dat zich expliciet richt op een flexibele aanpak en samenwerking tussen verschillende expertisegebieden.

Binnen Wageningen University & Research hebben deze ontwikkelingen al veel invloed op het onderwijs en onderzoek binnen de verschillende kennis-eenheden. Onze visie is dat big data en data science voor geheel Wageningen University & Research van groot belang is. Om de verschillende ontwikkelingen te ondersteunen, te coördineren en daarmee te versterken is daarom besloten tot oprichting van het WDCC.

2.2 Missie

We geven het WDCC als missie om voor Wageningen University & Research

'To explore the potential of data to help to improve the quality of life'

door

- Nieuwe mogelijkheden voor data driven onderzoek te ontwikkelen,
- Het onderwijs in data science te versterken,
- Daarmee bij te dragen aan het creëren van waarde,
- Te bouwen aan de voorwaarden hiervoor van datamanagement en infrastructuur,

in nauwe samenwerking met de betrokken instituten en afdelingen.

We willen hiermee bereiken dat Wageningen University & Research op basis van data gecombineerd over alle onderzoeksvelden integraal onderzoek kan uitvoeren, dat studenten worden opgeleid die dit vak beheersen, en dat op deze manier nieuwe methoden voor waarde creatie worden ontsloten. Tegelijk verwachten we dat de coördinatie van gelijksoortige werkzaamheden leidt tot efficiënter werken en dat het delen van kennis de kwaliteit verhoogt en leidt tot nieuwe mogelijkheden voor zowel onderwijs als onderzoek.

³ <https://bids.berkeley.edu>.

3 Doelstellingen

Uitgaande van de visie en de missie richten we ons op de volgende doelstellingen.

Voor onderzoek:

- Het stimuleren van het delen van kennis door het organiseren van communities,
- Het ontwikkelen van platforms en methodieken voor het toegankelijk maken en combineren van diverse soorten data,
- Het ontwikkelen van nieuwe concepten en diensten op basis van gedeelde kennis en gecombineerde data.

Voor onderwijs:

- Het organiseren van een onderwijsprogramma voor data science toegepast op het kennisdomein van WUR voor de opleidingen tot bachelor, master en PhD
- Het organiseren van cursussen in data science voor stafleden

Voor waarde creatie:

- Het organiseren van een proces waarbij vanaf het begin rekening wordt gehouden met de mogelijkheden voor waarde creatie bij de ontwikkeling van nieuwe concepten en diensten gebaseerd op big data

Voor infrastructuur:

- Helderheid over de benodigde infrastructuur voor onderzoek en onderwijs zowel binnen als buiten WUR

Voor datamanagement:

- Helderheid over het beleid en de uitvoering daarvan op het gebied van research datamanagement
- Een data-infrastructuur gericht op efficiënt gebruik, uitwisseling en integratie van interne en externe data

Voor alle onderdelen:

- Het vormgeven van samenwerking op dit gebied met externe partners zodanig dat deze aanvullend zijn op de mogelijkheden en expertise van WUR,
- Het verkennen van mogelijkheden voor aanvullende financiering bij het realiseren van bovenstaande doelstellingen.

De aanpak om deze doelstellingen te realiseren wordt verder uitgewerkt in hoofdstuk 4.

4 Strategie en organisatie

4.1 Strategie

We kiezen als strategie voor de realisatie van onze doelstellingen het vormgeven van een programma. In het programma onderscheiden we verschillende programmalijnen, die worden omschreven in hoofdstuk 7. Elke programmalijn draagt bij aan een of meerdere van de doelstellingen beschreven in hoofdstuk 3. De research data infrastructuur die zo wordt ontwikkeld wordt beschreven in hoofdstuk 6.

Via dit programma kunnen we het werk aan de doelstellingen organiseren zonder dat vooraf alle tussenstappen en randvoorwaarden bekend zijn. Elke stap in de ontwikkeling werken we uit als project. Door elk project vooraf goed te omschrijven en af te stemmen kunnen we beheerst het programma steeds verder invullen. Door de projecten inclusief de aanbevelingen voor het vervolg te evalueren dienen de resultaten als basis voor volgende projecten.

De programmatische aanpak biedt bovendien flexibiliteit. Dit is een voorwaarde om gemakkelijk nieuwe ontwikkelingen, zowel intern als extern, te kunnen betrekken bij het werk aan de doelstellingen.

4.2 Organisatie

De samenwerking tussen de verschillende onderdelen van WUR en het WDCC organiseren we als volgt. Het programma wordt georganiseerd door het WDCC, dat hiervoor een klein team beschikbaar heeft. De uitvoering van de projecten binnen dat programma gebeurt door dat team in nauwe samenwerking met andere onderdelen van WUR. Hiervoor leveren verschillende onderdelen van WUR inzet voor deze projecten.

Voorstellen, voortgang en resultaten van het programma worden afgestemd met een stuurgroep die hiervoor regelmatig (bv eens per kwartaal) bijeenkomt. De stuurgroep besluit over het programma van het WDCC. De voorzitter van de stuurgroep (algemeen directeur SSG) rapporteert regelmatig over de voortgang in de concernraad wanneer nodig bijgestaan door het hoofd van het WDCC. In deze stuurgroep zijn de kennisseenheden en de relevante stafafdelingen op strategisch niveau vertegenwoordigd samen met de Deans of Research en Education, en de Directeur Waardecreatie. Deze vertegenwoordigers stemmen het programma binnen hun eenheid af. Deze vertegenwoordigers zijn ook aanspreekpunt voor de inzet vanuit hun eenheid voor het WDCC.

Inhoud, koers en relevante ontwikkelingen voor het programma worden per programmalijn afgestemd met een expertgroep die hiervoor regelmatig (bv eens per maand) bijeenkomt. In deze expertgroep zijn de kennisseenheden en relevante stafafdelingen op expertniveau vertegenwoordigd. In overleg met de Deans of Research en Education en met de Directeur Waardecreatie wordt bepaald wie hun expertise hierin kan vertegenwoordigen. De experts ondersteunen het bepalen van de inhoudelijke koers van het WDCC en stemmen deze af met deskundigen binnen hun kenniseenheid. De adviezen van de expertgroep worden via het hoofd WDCC overgebracht aan de stuurgroep WDCC, die zich bij de besluitvorming in de regel hierop zal baseren.

Het is mogelijk dat binnen de expertgroep kennisseenheden door meer dan één persoon zijn vertegenwoordigd. De vertegenwoordiger voor het onderwijsprogramma is bijvoorbeeld niet noodzakelijk dezelfde als voor het onderzoeksprogramma. Deelname aan de bijeenkomsten kan bepaald worden afhankelijk van de daar te bespreken onderwerpen.

De kennisseenheden en stafafdelingen stellen zelf hun vertegenwoordigers voor de stuurgroep en de expertgroep voor, waarna deze groepen vervolgens in overleg met het hoofd WDCC worden samengesteld.

5 De rol van het WDCC binnen WUR

5.1 Aanleiding

De aanleiding voor het oprichten van het WDCC is het toenemend belang van data in de samenleving. Ons beperkend tot onderzoeksgegevens (research data) geldt dit evenzeer voor WUR. Dit levert strategische vragen op, zoals:

- Onder welke voorwaarden stellen we onze data beschikbaar?
- Wat betekenen de prestaties van op data gebaseerde algoritmen in zowel sterk gestructureerde⁴ als minder gestructureerde problemen⁵ voor academisch onderzoek en onderwijs?
- Welke externe samenwerkingsverbanden gaan we aan voor het verkrijgen van data en expertise?

Hierop volgen praktische vragen, zoals:

- Hoe organiseren we het datamanagement en richten we de bijbehorende infrastructuur in?
- Hoe gaan we om met het onderwijs- en onderzoeksprogramma?

Een recente verkenning geeft aan dat organisaties zich aanpassen voor dit type vraagstukken door een nieuwe rol te creëren, die van chief data officer⁶. Het belang overstijgt namelijk de onderdelen van de organisatie en vraagt om een samenhangende aanpak die niet goed past bij de bestaande rollen in organisaties. Dit komt overeen met de doelstellingen van het WDCC, waarbij we het WDCC kunnen zien als samengesteld uit een chief data officer met een bijbehorend team.

Deze rol kan, zoals blijkt uit een recente studie⁷, op verschillende manieren worden ingevuld. Dit betreft ambitie, scope en mandaat van de rol. Voor het slagen van de aanpak is het belangrijk vanaf het begin helderheid te hebben over de rol die we kiezen voor het WDCC.

5.2 Ambitie

Wat betreft ambitie, is er onderscheid mogelijk tussen die van een

- Data Integrator: gericht op het implementeren van een moderne data-infrastructuur, met innovaties waar mogelijk, om flexibel gebruik van te kunnen maken zonder veiligheidsrisico's;
- Business Optimizer: gericht op steeds efficiënter gebruik maken van de data-infrastructuur met geavanceerde applicaties voor het verwerken van data;
- Market Innovator: gericht op het ontwikkelen van nieuwe producten en diensten en tegelijk daarvoor ruimte makend door het overbrengen van niet-specifieke taken naar externe partners.

Binnen WUR is momenteel op onderdelen een moderne data-infrastructuur beschikbaar. Er is geen standaard voorziening, inclusief operationele procedures, waar elke onderzoeker en docent snel en betrouwbaar alle mogelijke data eenduidig kan opslaan, ophalen en verwerken.

Binnen WUR worden veel geavanceerde applicaties gebruikt voor het verwerken van data en wordt er in een aantal gevallen ook gewerkt aan het uitwisselen van ervaringen. Er is beperkt overzicht van de bestaande toepassingen en het is mogelijk dat er optimalisaties, in beheer en toepassing, haalbaar zijn.

Binnen WUR worden verschillende initiatieven genomen voor het ontwikkelen van nieuwe producten en diensten. Ervaringen worden uitgewisseld maar er is geen samenhangend programma gericht op het

⁴ David Silver e.a., "Mastering the game of Go without human knowledge", *Nature* 550, nr. 7676 (19 oktober 2017): 354–59.

⁵ Ying Chen, Jd Elenee Argentinis, en Griff Weber, "IBM Watson: How Cognitive Computing Can Be Applied to Big Data Challenges in Life Sciences Research", *Clinical Therapeutics* 38, nr. 4 (april 2016): 688–701, <https://doi.org/10.1016/j.clinthera.2015.12.001>.

⁶ <https://www.gartner.com/newsroom/id/3190117>.

⁷ Bruce Tyler, Cortnie Abercrombie, Rebecca Shockley, "The Chief Data Officer playbook", *IBM Institute for Business Value*, (mei 2016).

combineren van de verschillende expertisegebieden binnen WUR. Er wordt beperkt onderzocht welke niet-specifieke taken eventueel overgebracht kunnen worden naar externe partners.

Op basis van deze overwegingen is het zinnig om zowel de ambities van Data Integrator, Business Optimizer, en Market Innovator te kiezen. Het combineren van verschillende ambities levert het risico op van het verdelen van de inspanningen en daardoor minder snel bereiken van resultaat. Het alternatief, namelijk ons beperken tot een van de ambities, kent echter ook nadelen. Dit kan namelijk leiden tot een onevenwichtige ontwikkeling waarbij de een voorloopt op de ander. Daarom kiezen we ervoor om in het programma voor het WDCC de drie ambities te verwerken en via regelmatige afstemming te zorgen voor een evenwichtige ontwikkeling.

5.3 Scope

Wat betreft scope, is er onderscheid mogelijk tussen

- Data: de zorg, bescherming, het eigenaarschap, de kwaliteit, verwerving, identificatie en koppeling, en het beheer van de data zelf;
- Technology: de infrastructuur van hard- en software waarmee de data wordt opgeslagen en verwerkt;
- Analytics: de methodologie, toepassingen, en algoritmen in gebruik voor de interpretatie van data.

Het is duidelijk dat Data (nog steeds beperkt tot onderzoeksgegevens) onderdeel is van de scope van het WDCC. Gegeven de vraagstelling aan het begin van dit hoofdstuk, en de ambities hierboven geschetst, is ook duidelijk dat de beperking van de scope tot Data de vragen niet zal beantwoorden en de ambities niet zal realiseren. Daarom rekenen we Technology en Analytics ook tot de scope.

5.4 Mandaat

Wat betreft mandaat, zijn er gradaties mogelijk tussen sturend en faciliterend. Het is belangrijk dat het mandaat past zowel bij de organisatie als bij de ambities en de scope. Binnen WUR is de RvB sturend.

Technology en Analytics, als onderdelen van de scope van het WDCC, maken delen uit van de deelgebieden van onderwijs, onderzoek, en technische infrastructuur die elk binnen WUR hun eigen organisatie kennen. Dit geldt ook voor de ambities van Data Integrator en Business Optimizer. In aanvulling daarop maakt de ambitie van Market Innovation deel uit van waarde creatie, waarvoor ook een eigen organisatieonderdeel bestaat. Het mandaat voor deze gebieden ligt niet bij het WDCC. Op deze gebieden richt het programma van het WDCC zich dan ook faciliterend op het organiseren van afstemming om mogelijkheden en knelpunten zichtbaar te maken en kennis te delen. In het belang van de ambities zal het WDCC hier wel een adviserende rol op zich nemen.

Dit geldt ook voor Data, als onderdeel van de scope van het WDCC en beperkt tot onderzoeksgegevens, wat binnen WUR is belegd als onderdeel van het domein onderzoek. Om de organisatie te faciliteren kan het WDCC hiervoor in de praktijk fungeren als aanspreekpunt binnen WUR.

5.5 Rol van het WDCC

Het WDCC zal zich richten op de ambities van Data Integrator, Business Optimizer en Market Innovator. Het WDCC kan binnen de scope van Data fungeren als aanspreekpunt, terwijl het WDCC faciliterend en adviserend is binnen de scope van Technology en Analytics. Het programma zoals in het vervolg van dit document wordt beschreven is erop gericht deze rol in nauwe samenwerking met andere organisatieonderdelen te realiseren.

6 Research data infrastructuur

De doelstellingen van het WDCC kunnen weergegeven worden in de vorm van een research data infrastructuur. Dit dient om onderwijs en onderzoek te faciliteren bij de realisatie van toepassingen van big data. In dit hoofdstuk worden de verschillende onderdelen van de research data infrastructuur geschetst.

Het overzicht van een research data infrastructuur helpt om inefficiëntie (bv als gevolg van overlap en dubbelingen) en inconsistenties (bv door tegenspraak in beleid en toepassingen) te voorkomen. Het kan verder gebruikt worden voor het onderwijsprogramma (namelijk door gericht aan te geven welke onderdelen daarvoor van belang zijn) en het onderzoeksprogramma (door gericht ontwikkelingen op elkaar aan te laten sluiten).

Een research data infrastructuur is relevant voor externe ontwikkelingen. Externe dienstverleners bieden inmiddels vergaande faciliteiten die een groot deel van de huidige onderzoeks-infrastructuur kunnen vervangen^{8,9}. Commerciële partijen richten hun eigen data-infrastructuur in¹⁰. Tegelijk wordt er op Europees niveau gevraagd om de ontwikkeling van een internationale research infrastructuur¹¹. Het is noodzakelijk om als organisatie een eigen research data infrastructuur ontwikkeld te hebben om hierop in te kunnen spelen.

Tenslotte zal een gemeenschappelijk beeld in de vorm van een research data infrastructuur de samenwerking tussen de verschillende onderdelen van WUR bevorderen. Hiermee kan namelijk gewerkt worden aan een gedeelde basis van data en tools voor de ontwikkeling van gezamenlijke toepassingen.

6.1 Functie

De research data infrastructuur dient ter ondersteuning van het proces van onderzoek & onderwijs voor zover daarbij data wordt gebruikt.

Zowel onderzoek als onderwijs stellen verschillende voorwaarden aan het gebruik van data: het moet bijvoorbeeld snel beschikbaar zijn, betrouwbaar en gedocumenteerd. Dit zijn de randvoorwaarden voor het inrichten van de voorzieningen. Er zijn echter ook voorwaarden verbonden aan het gebruik van data: bijvoorbeeld, vanwege privacy of eigendom kan niet alle data worden gedeeld, vanwege richtlijnen moet data een aantal jaar worden bewaard. Deze voorwaarden volgen uit het beleid. Kennis van de aanwezigheid en mogelijke toepassing van data levert ideeën voor onderzoek & onderwijs. Deze kennis kan worden gedeeld in communities.

Het proces van onderzoek & onderwijs hangt dus nauw samen met het beleid, de voorzieningen en de communities. Deze drie onderdelen van de research data infrastructuur worden hieronder kort beschreven.

6.2 Beleid

Het beleid voor research data management vormt het kader voor de research data infrastructuur en dient om een heldere afweging te maken van verschillende belangen, met als doel het optimaliseren van het proces van onderzoek & onderwijs, voor zover daarbij data wordt gebruikt.

⁸ C. J. Radford, "Challenges and solutions protecting data within Amazon Web Services", *Network Security* 2014, nr. 6 (2014): 5–8.

⁹ Noel Gorelick e.a., "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone", *Remote Sensing of Environment*, juli 2017, <https://doi.org/10.1016/j.rse.2017.06.031>.

¹⁰ <https://www.forbes.com/sites/bruceupbin/2013/10/02/monsanto-buys-climate-corp-for-930-million/#72b02196177a>.

¹¹ http://ec.europa.eu/research/infrastructures/index_en.cfm.

Het beleid kan beperkt blijven tot noodzakelijke voorwaarden zoals bijvoorbeeld wettelijk vereist. Het beleid kan (als alternatief) verder gaan en principes formuleren ter bevordering van bepaalde doelstellingen. Het beleid kan ook best practices binnen de organisatie verheffen tot standaard met bijvoorbeeld als doel het verhogen van de kwaliteit of de efficiëntie van onderzoek en onderwijs.

Voor de werking is het van belang dat het beleid helder is, ondersteund wordt, en actueel wordt gehouden.

6.3 Voorzieningen

De voorzieningen vormen het fysieke deel van de research data infrastructuur en dienen voor de uitvoering van het proces van onderzoek & onderwijs en bestaan uit data, hardware en software.

Voorzieningen kunnen lokaal georganiseerd worden of centraal, intern WUR of extern WUR. De keuze hieruit volgt uit een afweging van de voorwaarden vanuit onderzoek & onderwijs in combinatie met het geldend beleid.

Er is onderscheid te maken in verschillende soorten voorzieningen, zoals bijvoorbeeld bronnen (voor het verwerven van data), nodes (hardware voor berekeningen of voor opslag van data), connectors (verbindingen tussen nodes al dan niet voorzien van faciliteiten die verder gaan dan transport van gegevens), en tools (faciliteiten voor het bewerken van gegevens in de vorm van software al dan niet geïnstalleerd op nodes of geïmplementeerd in connectoren).

Voorzieningen kunnen verkregen worden vanuit de markt of door eigen ontwikkeling, en beschikbaar gesteld worden voor eigen gebruik of (ook) voor de markt.

6.4 Communities

De communities vormen het actieve deel van de infrastructuur en dienen voor het delen van kennis, het oplossen van problemen en het genereren van nieuwe ideeën.

Communities kunnen zich richten op verschillende onderwerpen. Bij de samenstelling en uitvoering van het onderzoeksprogramma is het bijvoorbeeld van belang dit af te stemmen met onderzoekers uit de verschillende kenniseenheden. Om expertise te delen kunnen specialisten in data science of op het gebied van de technische infrastructuur regelmatig nieuwe ontwikkelingen bespreken. Studenten die in een proeftuin samen onderzoek kunnen doen naar verschillende typen data vormen ook een community die kan zorgen voor inspirerend onderwijs.

6.5 Toepassingen

De toepassingen vormen als het ware de kroon op de research data infrastructuur, in de zin dat hiermee via onderzoek & onderwijs waarde wordt gecreëerd.

6.6 Ontwikkeling en doel

Een research data infrastructuur dient stabiel te zijn maar kan niet altijd ongewijzigd blijven. Het programma van het WDCC, zoals beschreven in hoofdstuk 7, dient om de research data infrastructuur stapsgewijs verder te ontwikkelen. Daarmee is het doel van het programma van het WDCC te vertalen in het ontwikkelen van helder beleid, een passende infrastructuur, en actieve communities, leidend tot veelbelovende toepassingen.

7 Programma

Voor het werkprogramma van het WDCC definiëren we vijf programmalijnen:

1. Onderzoek
2. Onderwijs
3. Waarde creatie
4. Big datamanagement
5. Specifieke infrastructuur

Deze programmalijnen worden hieronder beknopt beschreven. Een nadere beschrijving voor het opstarten van deze programmalijnen is gegeven in de bijlagen.

7.1 Onderzoek

Het doel van deze lijn is het tot stand brengen van een aansprekend onderzoeksprogramma voor toepassingen van big data dat dwarsverbanden legt tussen de verschillende onderzoeksgroepen. Het programma is gericht op het realiseren van een structurele samenwerking op dit gebied tussen de onderzoeksgroepen (ontschotten). Bij het vormgeven van deze programmalijn wordt nauw samengewerkt met het lopende investeringsthema Big Data ter voorbereiding op een nieuw investeringsthema.

Met het beschikbaar komen van enorme gegevenssets (big data) ontstaat de mogelijkheid om onderzoeksvragen te beantwoorden die voorheen niet goed konden worden onderzocht. Bij data driven onderzoek zijn deze gegevenssets als het ware het startpunt op basis waarvan onderzoeksvragen worden geformuleerd.

Dit vergt een wat andere aanpak voor onderzoek dan de 'klassieke' aanpak, door de data te nemen als uitgangspunt bij het opstellen van de vraagstelling. Juist de combinatie van gegevens uit verschillende disciplines biedt hierbij perspectief op nieuwe mogelijkheden. Deze lijn richt zich dan ook vooral op het ontwikkelen van mogelijkheden voor onderzoek door gegevens vanuit verschillende onderzoeksgroepen te combineren. Door interne en externe data uit verschillende bronnen te combineren kunnen we de unieke combinatie van expertisevelden binnen Wageningen optimaal gebruiken. Tegelijk laat dit de waarde zien van One Wageningen.

Niet alle data komt overigens zonder meer in aanmerking om gecombineerd te worden. Het stadium van het onderzoek waarbinnen de gegevens zijn verzameld, de daarbij betrokken partijen, of voorwaarden onder welke de data alleen kan worden gebruikt kunnen hiervoor beperkingen opleveren. Het is van belang hiermee zorgvuldig om te gaan en zo mogelijk wel de beschikbaarheid van deze data voor eventueel later gebruik te borgen.

Deze lijn wordt opgezet rond concrete toepassingen van data driven onderzoek. Op basis van de toepassingen worden de benodigde datasets bepaald en beschikbaar gemaakt voor het onderzoek. Het doel is vervolgens niet alleen het realiseren van de toepassing maar ook het onderzoeken hoe de data gemakkelijk beschikbaar gesteld kan worden voor het onderzoek, op welke manier verschillende typen data aan elkaar kan worden gekoppeld, welke infrastructuur aan hard- en software hiervoor nodig is, en welke modellen en algoritmen gebruikt worden.

Met de resultaten van de verschillende toepassingen kan zo worden gebouwd aan een heldere omschrijving van de mogelijkheden van en voorwaarden voor onderzoek met big data.

De resultaten worden regelmatig besproken in bijeenkomsten die worden georganiseerd voor onderzoekers uit de verschillende kenniseenheden. Op deze manier wordt gewerkt aan het opbouwen van een research community op het gebied van data science.

7.2 Onderwijs

Het doel van deze lijn is het tot stand brengen van een aansprekend onderwijsprogramma bij WUR op het gebied van toegepaste data science in de brede zin van het woord. Dit dient te leiden tot afgestudeerden met 'data skills' voor hun verdere loopbaan, zowel in de opleiding tot bachelor, master als PhD. Het onderwijsprogramma dient ook te voorzien in opleidingen voor stafleden die zich op dit gebied willen ontwikkelen.

Om te kunnen werken met grote gegevenssets zijn allerlei vaardigheden nodig. Dit varieert tussen technische kennis van software, databases en programmeervaardigheden aan de ene kant en kennis van modelmatige berekeningen, visualisaties en geavanceerde analysetechnieken aan de andere kant.

Deze vaardigheden worden bij verschillende onderdelen van de organisatie al ruimschoots toegepast. Er zijn dan ook intern volop mogelijkheden voor stages van studenten. In het algemeen zijn de studenten bij WUR echter beperkt voorbereid op dit type onderzoek. Met een onderwijsprogramma dat hierin voorziet kan een betere aansluiting tussen onderwijs en onderzoek bereikt worden.

Momenteel worden er binnen Wageningen University & Research al verschillende vakken gegeven en zijn er ook initiatieven genomen voor het realiseren van toegepast onderwijs. Deze programmaliijn richt zich op het uitbouwen hiervan tot een compleet programma van toegepaste vaardigheden. Waar mogelijk wordt daarbij aangesloten bij lopende ontwikkelingen.

7.3 Waarde creatie

Het doel van deze programmaliijn is het ontwikkelen van een proces zodat als deel van de onderzoeken op dit gebied vanaf het begin rekening wordt gehouden met de mogelijkheden voor duurzame waarde creatie. Waarde creatie varieert daarbij tussen het leveren van een bijdrage aan maatschappelijke ontwikkelingen enerzijds tot het vergroten van de omzet door het op de markt brengen van bij WUR ontwikkelde technologie anderzijds. Met duurzaam wordt hier bedoeld dat in principe niet gestreefd wordt naar eenmalige verdiensten maar naar continue opbrengsten.

Onderwijs in data science biedt de mogelijkheid om studenten bij WUR een nieuw vak te leren, namelijk de toepassing van computer science op de specifieke expertisegebieden van WUR. Juist deze combinatie kan WUR als een van de weinige partijen ontwikkelen. De waarde die hiermee gecreëerd kan worden bestaat uit goed opgeleide mensen die vervolgens hun bijdrage kunnen leveren aan de maatschappij, zoals in het onderzoek en binnen het bedrijfsleven. De inspanningen hiervoor zijn georganiseerd binnen de lijn Onderwijs. De lijn Waarde creatie beperkt zich dan ook tot Onderzoek.

De toepassing van big data biedt ruimschoots mogelijkheden voor waarde creatie. Dit varieert tussen het tegen betaling beschikbaar stellen van eigen data tot het bieden van online services op een platform voorzien van data en modellen voor abonnees. Er zijn binnen WUR al diverse ontwikkelingen in deze en in andere richtingen gaande.

Het proces dat wordt voorgesteld maakt deel uit van het uit te voeren onderzoek. Het bestaat eruit dat bij elk onderzoeksvoorstel vanaf het begin een (al is het eenvoudig) verdienmodel wordt opgesteld. Dit verdienmodel geeft aan hoe het resultaat van het onderzoek kan worden gebruikt voor het creëren van waarde, in maatschappelijke of in financiële zin. Hierbij wordt rekening gehouden met de toepassing van het onderzoek, verwachte afnemers, de kosten benodigd voor het realiseren en eventueel onderhouden van de toepassing, en de verwachte opbrengsten.

Om dit als onderdeel van de onderzoeken uit te kunnen voeren wordt binnen deze programmaliijn gewerkt aan de ontwikkeling en het beschikbaar stellen van verschillende verdienmodellen voor verschillende toepassingen van big data. Op deze manier wordt voorkomen dat de ontwikkeling van het

verdienmodel als extra onderdeel binnen het onderzoek moet gebeuren. Tegelijk wordt zo verkend welke nieuwe mogelijkheden er zijn voor waarde creatie.

De uitwerking van de verdienmodellen brengt het complete proces in kaart van realisatie van de toepassing tot en met de opgeleverde waarde. Hierbij zullen omwille van de eenvoud waar nodig een aantal aannames worden gedaan. Op deze manier wordt ervaring opgedaan met het uitwerken van verdienmodellen. Bovendien levert het zicht op toekomstige mogelijkheden voor waarde creatie.

7.4 Big data management

Het doel van deze lijn is het optimaliseren van het databeheer binnen WUR, specifiek voor toepassingen van big data in onderwijs en onderzoek. Hiervoor wordt beleid geformuleerd en ondersteuning geboden. De lopende activiteiten op het gebied van Research Data Management vormen hiervoor het uitgangspunt. Beleid en ondersteuning vormen onderdeel van de data-architectuur die in grote lijnen de structuur en organisatie beschrijft voor het verzamelen, bewaren, verwerken en verstrekken van data.

Het beheer van data vormt een basisvoorziening voor het uitvoeren van onderzoek en onderwijs. Beheer betreft o.m. de methodes voor inzameling, verstrekking en opslag van gegevens, het waarborgen van privacy en security, helderheid scheppen over eigendom en daarmee mogelijkheden voor gebruik en verstrekking, het documenteren en daarmee inhoudelijk toegankelijk maken van gegevens, het mogelijk maken van zinvolle koppelingen van verschillende typen gegevens, en het bepalen van de eisen die het gewenst gebruik van de gegevens stelt aan hard- en software en aan de organisatie voor het gegevensbeheer.

Beleid wordt geformuleerd op hoofdlijnen en wordt beperkt worden tot de zaken die noodzakelijk zijn of het werken binnen de organisatie efficiënter maken. De data-architectuur dient om de diversiteit aan data inzichtelijk te maken, te werken aan eenduidig en efficiënt gebruik van data, en om sleutels te kunnen introduceren voor het combineren van verschillende soorten data.

7.5 Specifieke infrastructuur

Het doel van deze lijn is het aangeven van de functionele voorwaarden die er zijn aan de infrastructuur van hard- en software voor onderwijs en onderzoek op het gebied van big data binnen WUR. Hiermee kan vervolgens worden bepaald welke mogelijkheden er zijn voor het realiseren hiervan. Een goede technische infrastructuur vormt de basis voor zowel eigen onderwijs en onderzoek als voor externe samenwerking.

Om deze infrastructuur praktisch te kunnen realiseren is het (net zoals dit geldt voor onderzoek, onderwijs en data management) nodig inzicht te hebben in de huidige praktijk (als startpunt voor elke verandering), de knelpunten die dat oplevert, en op de verwachte toekomstige behoeften. Deze behoeften kunnen worden gebaseerd op concrete ontwikkelingen in onderzoek en onderwijs. Op basis hiervan kan vervolgens een visie worden ontwikkeld en uitgewerkt in een architectuur. Deze architectuur beschrijft in grote lijnen de structuur van de benodigde hardware en software.

Het doel van een dergelijke schets van de infrastructuur is te komen tot afgestemde en daardoor efficiëntere oplossingen. Mogelijke technische oplossingen kunnen vervolgens hieraan worden getoetst voordat ze worden geïmplementeerd. Op deze manier kan gebouwd worden aan een infrastructuur die efficiënt is en geschikt voor de toekomst. Een belangrijke vraag bij het ontwikkelen van de infrastructuur is in welke mate dit binnen WUR wordt gerealiseerd, en in welke mate een beroep wordt gedaan op de markt of op samenwerkingsverbanden.

7.6 Samenhang programmalijnen

Het WDCC heeft als doel om de mogelijkheden van big data voor WUR te ontsluiten en te bouwen aan een duurzame toepassing hiervan. Vooraf is niet volledig in te schatten wat de verschillende mogelijkheden op gaan leveren. Daarom zal via het uitwerken van concrete toepassingen en randvoorwaarden gaandeweg kennis hierover worden opgebouwd, via de hierboven beschreven programmalijnen.

Elke stap die is afgerond geeft nieuwe informatie over opbrengsten, nieuwe mogelijkheden en benodigde inspanningen. Het is van groot belang om in dit stadium deze kennis te delen. Daarom zullen de resultaten worden gepresenteerd op bijeenkomsten met betrokken medewerkers zodat gebouwd wordt aan een community die ervaringen, inzichten en ideeën hierover deelt. De resultaten dienen ook als achtergrond voor de expertgroep bij het beoordelen van nieuwe voorstellen.

Door het uitwisselen van kennis ontstaat breed verspreid inzicht in de verschillende mogelijkheden van big data, de voor- en nadelen van verschillende analysemethoden, ervaringen met infrastructuur en data management, en de verschillende manieren waarmee waarden kan worden gecreëerd. Op deze manier organiseren we een cyclus gericht op continue creatie van waarde.

