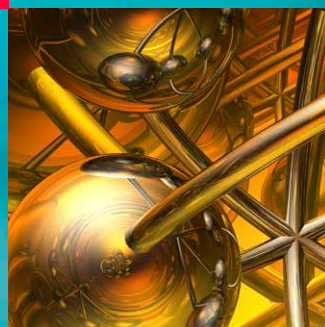
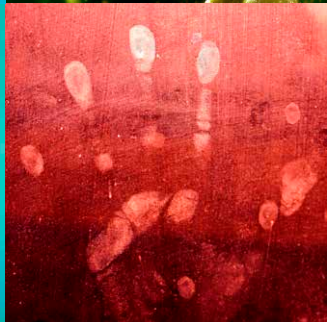
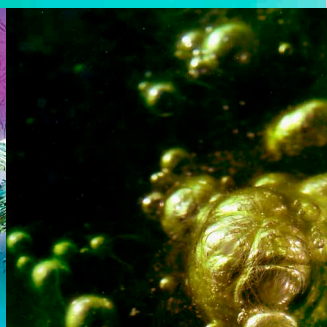
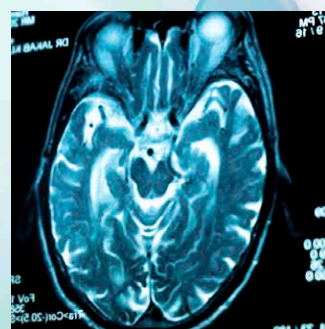
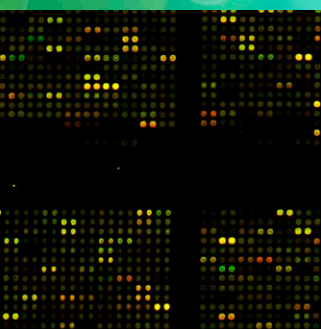


Naar een **datavaardige** **professional** binnen de Applied Sciences

Een inventarisatie en advies op het gebied
van data-onderwijs



Colofon

Onderzoek en tekst

ir. Bo Blanckenburg

Projectcoördinatie

drs. Jacqueline van Erp

Marjolein Wijnker – Schrauwen, MSc (DAS)

drs. Lisette van der Beek (DAS)

Tekstredactie

drs. Marijne Thomas (LOSI, Amersfoort)

Vormgeving

Liesbeth Thomas, BA (t4design, Delft)



Domein Applied Science (DAS)

Postbus 249

2260 AE Leidschendam

070 337 87 72

info@appliedscience.nl

www.appliedscience.nl

© Domein Applied Science, september 2018



Naar een datavaardige professional binnen de Applied Sciences

Een inventarisatie en advies op het gebied van data-onderwijs

Inhoud

Inleiding	5	
Managementsamenvatting	6	
Projectaanpak	7	
Algemene terminologie	8	
Opleidingsgroep	Applied Science	9
Opleidingsgroep	Bio-informatica	10
Opleidingsgroep	Biologie en medisch laboratoriumonderzoek & Biotechnologie	11
Opleidingsgroep	Chemie	16
Opleidingsgroep	Chemische technologie	19
Opleidingsgroep	Forensisch onderzoek	20
Opleiding	Milieukunde	21
Opleidingsgroep	Technische natuurkunde	22
Conclusies en aanbevelingen	24	
Bijlage I	Vragenlijst	27
Bijlage II	Overzicht geïnterviewden	28
Bijlage III	Toelichting veelgebruikte termen	29

Inleiding



‘In de wetenschap is het dus een multidisciplinair geschoolde persoon die de brug kan bouwen tussen enerzijds de wereld van Big Data, en anderzijds de wetenschappelijke methode.’

– 28 juni 2016, Kamerbrief Big Data in Onderwijs, Cultuur en Wetenschap

De groei van automatisering en digitalisering is in de gehele samenleving merkbaar, ook in het werkveld van de Applied Science-professional van nu en de toekomst. Zo worden biomedische analyses geautomatiseerd en worden sequence- en multi-omicsdata steeds sneller en nauwkeuriger gegenereerd – in toenemende mate zelfs zonder tussenkomst van fysieke werknemers. De chemische industrie voorziet een toename van automatische, meer lokale metingen, en bij ingenieursbureaus spelen automatische sensoren een steeds grotere rol.

Al deze data moeten natuurlijk ook ergens naartoe, bij voorkeur doorzoekbaar voor zowel mens als machine. Data stewardship en -management¹ zijn dan ook nieuwe disciplines geworden binnen het werkveld.

Ook de groeiende beschikbaarheid van grote hoeveelheden data in online databases heeft de behoefte aan kennis en vaardigheid op het gebied van het genereren, managen en analyseren van data vergroot.

Hierdoor lijkt bij sommige opleidingen binnen het domein Applied Science een verbreding gaande, waarbij ‘traditionele’ vaardigheden belangrijk blijven, maar de data-analytische vaardigheden een grotere rol spelen dan voorheen. Daarnaast ervaren sommige opleidingen een vermindering

aan beschikbare stageplaatsen² op laboratoriumgebied; wellicht kan meer data-onderwijs hier uitkomst bieden.

Tijdens de DAS-bijeenkomst ‘Professional van de Toekomst’³ van 1 december 2017 werd de omgang met data, automatisering en digitalisering als een van de vijf thema’s voor de toekomst geïdentificeerd⁴. Hoe bereiden wij, als DAS-opleidingen, onze studenten voor op dit toekomstige werkveld?

Voor u ligt een analyse van wat de DAS-opleidingen zelf signaleren aan trends en behoeften op dit gebied en hoe hierop binnen het onderwijs wordt ingespeeld. Daarnaast vindt u een aantal adviezen over hoe deze tendensen verder kunnen worden ondersteund door DAS en de betrokken opleidingen.

MISSIESTATEMENT DATA-ONDERWIJSPROJECT

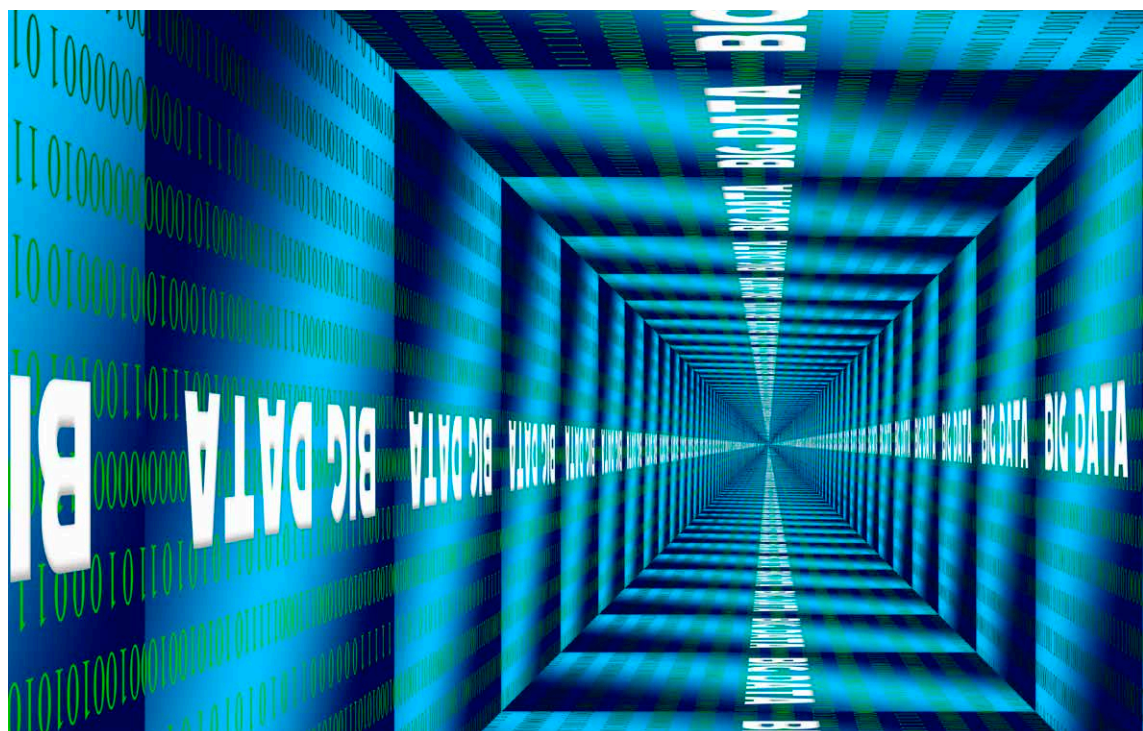
Inventariseren hoe de DAS-opleidingen beginnend beroepsbeoefenaars van voldoende vaardigheden kunnen en willen voorzien op het gebied van datavaardigheid (data-analyse en/of data stewardship).

1. Dutch Techcentre for Life Sciences 18.04.2018

2. Memo DAS algemeen bestuur aan Raad van Advies 25.05.2018

3. Fotoverslag discussiebijeenkomst Professional van de toekomst 01.12.2017

4. De overige thema’s zijn: Duurzaamheid, veiligheid & circulaire economie; Fundament, de basis op orde; Leven lang leren / ontwikkelen, en Lab in verschillende verschijningsvormen



Managementsamenvatting

Recente ontwikkelingen in het werkveld van de Applied Science-opleidingen, zoals de groeiende automatisering van biomedische laboratoria en de toenemende beschikbaarheid van grote hoeveelheden data in online databases, hebben de behoefte aan kennis en vaardigheid op het gebied van het genereren, managen en analyseren van data vergroot.

Voor dit rapport is bij de verschillende opleidingsgroepen geïnventariseerd wat zij aan data-onderwijs aanbieden, waar de groeigebieden en wensen liggen en op welke gebieden DAS dit onderwijs verder zou kunnen ondersteunen. Dit heeft geresulteerd in een vijftal adviezen voor vervolg van dit traject.

1. Aandacht vergroten voor digitale geletterdheid in de curricula, zodat communicatie binnen multidisciplinaire teams op het gebied van datageneratie, analyse en opslag soepel kan verlopen.
2. Faciliteren van uitwisseling tussen met name BML-docenten op het gebied van data-onderwijs, aangezien in dit vakgebied de grootste noodzaak wordt ervaren om afgestudeerden datavaardigheid mee te geven.
3. Stimuleren van het werkveld om specifieke wensen op dit gebied duidelijk naar de opleidingen te communiceren.
4. Het onderwijs, in samenwerking met het werkveld, voorzien van een geschikte dataset om onderwijs mee te ontwikkelen.
5. Data stewardship in beperkte mate opnemen in de onderwijsprogramma's.



Projectaanpak

Om goed te kunnen inventariseren waar de behoeftes van de DAS-opleidingen liggen, is gekozen voor een serie semigeleide gesprekken aan de hand van een vooraf opgestelde vragenlijst (bijlage 1). Het streven was om van elke bij DAS aangesloten opleiding met één persoon te spreken die door de opleiding zelf werd aangewezen als deskundig op dit gebied. Dit is voor vrijwel alle opleidingen gebeurd; een handvol opleidingen gaf aan op het moment niet in de gelegenheid te zijn deel te nemen aan deze inventarisatie. In bijlage 2 is een

lijst opgenomen van geïnterviewden.

Per opleidingsgroep is een samenvatting gemaakt van de ontwikkelingen en behoeften die de contactpersonen aangaven. Hierbij is bij elke opleiding vermeld wat er al aan datagerelateerd onderwijs wordt gegeven, waar de opleidingen groeigebieden zien en waar nog wensen liggen. Hoe DAS hier een rol in zou kunnen spelen, stond centraal. Daarbij is ook aandacht besteed aan de mogelijkheid om de gezamenlijk gedragen DAS-competenties op dit gebied uit te breiden.

Deelnemers per opleidingsprofiel

Opleidingsprofiel	Deelnemende contactpersonen	Gebruikte afkorting
Biologie en medisch laboratoriumonderzoek en Biotechnologie	10	BML, BT
Chemie	8	C
Chemische technologie	3	CT
Applied science	2	AP
Technische natuurkunde	3	TN
Forensisch onderzoek	2	FO
Bio-informatica	3	BI
Milieukunde	1	MK

De majortrack Forensisch laboratoriumonderzoek van de Avans Hogeschool is meegeteld bij de opleidingsgroep Forensisch onderzoek.

Algemene terminologie

Competentiegericht onderwijs

In de DAS-competenties worden de vaardigheden beschreven die een Bachelor of Science als beginnend beroepsbeoefenaar zou moeten beheersen. Een competentie is een combinatie van kennis, vaardigheden en attitude, die nodig is voor het uitvoeren van een taak of functie in een bepaalde context.

De competenties van de Bachelor of Science in het domein Applied Science zijn:

1. Onderzoeken
2. Experimenteren
3. Ontwikkelen
4. Beheren | coördineren
5. Adviseren | in- en verkopen
6. Instrueren | coachen
7. Leidinggeven | managen
8. Zelfsturing

In dit rapport komt met name de competentie 'Beheren' ter sprake, waarin op het moment de nadruk ligt op laboratoriummanagement. De bereidheid om databeheer in deze competentie op te nemen, is besproken.

Deze competenties en hun onderliggende handelingsindicatoren worden vastgesteld voor alle DAS-opleidingen, maar per opleidingsgroep wordt in de Landelijke Overlegorganen in samenwerking met het desbetreffende werkveld vastgesteld welke niveaus behaald worden voor welke competentie.

Meer details over de competenties is te vinden in de DAS profielbeschrijving *Bachelor of Applied Science, een competentiegericht profielbeschrijving* van 2016 en het bijbehorende addendum van januari 2018.

Body of Knowledge en Skills (BoKS)

Elke opleiding beschikt over een zogeheten Body of Knowledge en Skills, waarin per opleiding inhoudelijke onderwerpen op landelijk niveau zijn vastgesteld die de minimale basis voor elke opleiding vormen. Per opleidingsgroep wordt de inhoud hiervan in Landelijke Opleidingsoverleggen en in samenwerking met het desbetreffende werkveld vastgesteld.

Data stewardship

Onder data stewardship wordt verstaan, dat er binnen een onderzoek(sgroep) wordt nagedacht over opslag, beheer, uitwisselbaarheid en vindbaarheid van de data tijdens en na een project. Een data steward kan al in de voorbereidingsfase van een project adviseren en blijft ook lang na het project verantwoordelijk voor bepaalde acties rond de data. Bij grote EU-subsidies wordt een dergelijk Data Management Plan (DMP) als wenselijk beschouwd. Aangezien analisten binnen onderzoeksgroepen vaak een stabiele factor zijn, is het denkbaar dat zij hierbij in de toekomst in toenemende mate betrokken zullen worden.

Veelgebruikte termen

In bijlage 3 wordt een aantal veelgebruikte termen, zoals PCA en machine learning, nader toegelicht.



Opleidingsgroep Applied Science

Werkveldbehoefte

Deze opleidingen geven aan dat het werkveld erg divers is, waardoor het moeilijk is om specifiek aan te wijzen hoeveel en waar precies meer behoefte is aan datavaardigheid. De meeste afgestudeerden komen ermee in aanraking, en het idee heerst wel dat er meer behoefte is aan datavaardigheid in het algemeen (Z). Fontys is aan het inventariseren bij het werkveld en de eigen alumni wat de precieze wensen zijn.

Bijzonderheden

Onder docenten wordt vaak de discussie gevoerd: waar begint de bio-informatica? Bovendien: verder is het curriculum al vrij vol, dus wat moet er dan weg? (F). Vakken lijken nu erg over wiskunde te gaan en lijken los te staan van de belevingswereld van de student (Z).

Data stewardship

Fontys geeft aan hier nog niet aan toe te zijn, en eerst een duidelijk beeld te willen hebben van de werkveldwensen aan datacreatie en -analyse,

voordat data stewardship aan bod kan komen.

Zuyd Hogeschool geeft aan dat zij het opnemen hiervan in de competentie 'beheren' zouden verwelkomen, en onderwijs hierover in de keuzeruimte zouden willen opnemen.

Wensen

Onderwijsmateriaal voor een programmeervak (R-statistics of Matlab) en scholing voor docenten (F). Gezamenlijke keuzeminor big data ontwikkelen (Z).

Advies

Voor deze opleidingen is het advies om mee te liften met de adviezen aan andere opleidingen. Stuur indien gewenst een afgevaardigde naar de samenwerkingen bij BML en/of Chemie, en wissel onderling onderwijsmateriaal uit. De unieke uitdaging binnen deze samengestelde opleiding – dat er ruimte moet zijn voor allerlei basisprincipes – behoeft interne discussie. Wellicht dat alle toekomstig afgestudeerden kunnen profiteren van een basisvak data-analyse.

Geïnterviewden:

Charlotte Creusen

Zuyd Hogeschool (Z)

Ilse Vondervoort

Fontys Hogeschool TNW (F)

Aandeel data-onderwijs in de opleiding

	EC	Indicatie van behandelde onderwerpen	Software in verplicht curriculum	Bijzonderheden
Fontys	10-14 EC	Univariate statistiek, wiskunde	Excel, SPSS	Bio-informatica keuzevak (2EC) met NCBI-tools
Zuyd Hogeschool	± 13-15 EC	Univariate statistiek, wiskunde, kalibratie en validatie, experimental design	Excel	Bio-informatica keuzevak over simulaties ism Maastricht University (curriculumherziening gaande)



Opleidingsgroep Bio-informatica

Geïnterviewden:

Michiel Noback

Hanzehogeschool
Groningen (G)

Martijn van der Brugge

Hogeschool van Arnhem
en Nijmegen (N)

Bo Blanckenburg

Hogeschool Leiden (L)
(tevens auteur van dit
document)

Vergeleken met de andere opleidingen is Bio-informatica de enige opleiding waarbij data-onderwijs de kern vormt van de opleiding, in plaats van één van de onderdelen. Daarom is voor deze opleidingsgroep gekozen voor een andere weergave, waarbij alleen de opvallende ontwikkelingen op dit gebied worden benoemd.

Werkveldbehoefte

De HAN signaleert een toename aan afstudeerprojecten waarbij R-statistics wordt gebruikt. In Leiden en Groningen signaleert men een toename aan afstudeerprojecten waarbij machine learning een rol speelt. De opleidingen zijn zich bewust van de big data buzz, en zijn nog bezig te inventariseren wat hiervan in het curriculum thuishoort. Er zijn geen indicaties dat de bio-informatica-studenten specifieke nieuwe vaardigheden zouden moeten bijleren.

Aandeel van de opleidingen aan data-onderwijs

Bio-informaticastudenten zijn onder andere vaardig in het gebruik van NCBI-tools, programmeren in Python, R, Java en soms Perl, pipelines opzetten, werken met servers en met grote hoeveelheden data. Ze worden onderwezen in uni- en (een deel) multivariate statistiek, machine learning, databases, geavanceerde analyses van NGS-data en RNA-seq data uitvoeren en software valideren.

Bijzonderheden

De opleidingen in Leiden en Groningen bieden specifieke data-onderwijsminoren aan voor studenten Biologie en medisch laboratoriumonder-

zoek. De Hanzehogeschool Groningen ontwikkelt op dit moment een hbo-master op dit gebied, die in collegejaar 2019-2020 zou moeten starten. Hierin komt ook aandacht voor data stewardship.

Data stewardship

In Nijmegen is een gastcollege data stewardship onderdeel van het curriculum. In Leiden zijn er (afstudeer)projecten geweest die hiermee te maken hebben. Leiden onderzoekt de mogelijkheid dit op te nemen in het curriculum of de minor informatica. Toevoeging van data stewardship aan de competenties zou prima zijn. Groningen heeft nog geen data stewardship aandeel, maar wil dit wel graag.

Wensen

Voor de bio-informaticaopleidingen zou het prettig zijn als inzichtelijk werd gemaakt wie wat doet op het gebied van big data science in Nederland.

Advies

Samenwerking met Dutch Techcentre for Life Sciences (DTL) voor data stewardship.

In samenwerking met de Data Stewards Interest Group (onderdeel van DTL) zou een gastcollege datamanagement ontwikkeld kunnen worden. Zodra er een framework voor een dergelijk gastcollege ligt, kan er op elke hogeschool die daar behoefte aan heeft, een lokale data steward dit gastcollege komen geven. DTL en de Data Stewards Interest Group hebben al aangegeven dat zij hier graag bij zouden willen helpen.

Opleidingsgroep

Biologie en medisch laboratoriumonderzoek & Biotechnologie

Werkveldbehoefte

Alle BML-opleidingen geven aan dat er vanuit het werkveld een behoefte gesignaleerd wordt aan meer datavaardigheid bij BML-afgestudeerden. Deze behoefte heeft vooral te maken met een groter aandeel van data-analyse tijdens afstudeerstages door snellere en meer geautomatiseerde apparatuur (zoals Next Generation Sequencing-data) en de behoefte aan vaardigheid met bio-informaticatools, ook wel dry lab genoemd. Datavisualisatie speelt daar ook een belangrijke rol bij. Onder big data wordt bij de BML-opleidingen meestal verstaan: -omics data in medisch/biologisch kader die te groot zijn om met Excel te verwerken, of die niet geschikt zijn om in dit platform te verwerken.

Deze toename geldt niet voor alle studenten, maar wel voor het grootste deel van de studenten binnen deze opleiding, met name aan de researchkant. De verwachting is dat de behoefte van het werkveld aan deze vaardigheden alleen maar zal toenemen, en daarom is het van belang hier in het onderwijs op te anticiperen. (Rudimentair) programmeren wordt slechts door enkele opleidingen genoemd als nuttige toe-

voeging aan deze populatie (HU, L, G, IH). Wel merkt dit deel van de opleidingen op dat voor de samenwerking tussen bio-informatici/informatici en BML-afgestudeerden, meer wederzijds begrip van elkaars vakgebied wenselijk is.

Aandeel van de opleidingen aan data-onderwijs

Bij alle BML-opleidingen wordt aandacht besteed aan basisvaardigheden op het gebied van data-analyse, zoals werken met Excel, kalibratielijnen maken en betrouwbaarheidsintervallen bepalen. Sommige opleidingen breiden dit uit met het leren werken met statistische pakketten zoals SPSS of specifieke Excel plug-ins. Statistiek vormt in toenemende mate een belangrijk onderdeel van de opleiding.

Alle opleidingen besteden aandacht aan de dry lab- of bio-informaticakant van het werkveld. Studenten moeten in staat zijn om gebruik te maken van met name NCBI-tools en tools die het labwerk ondersteunen, zoals Clone Manager, NEBcutter of Primer3plus. Hoeveel tijd daaraan wordt besteed, verschilt per opleiding en is niet gemakkelijk te vergelijken, omdat veel van dit werk wordt verweven met het labwerk. →

Geïnterviewden:

Eddy van der Linden

Hogeschool Rotterdam (HR)

Taco Graafsma

Saxion (SA)

Romeo Lascaris (BML/BT)

Hogeschool Inholland (IH)

Rens van Leeuwen

NHL Stenden Hogeschool,
locatie Emmen (ST)

Kees Rodenburg

Avans Hogeschool (A)

Ivo Horn

Hogeschool Leiden (L)

Jurre Hageman

Hanzehogeschool
Groningen (G)

Jeroen van Kilsdonk

Hogeschool van Arnhem
en Nijmegen (N)

Marc Teunis

Hogeschool Utrecht (U)

Gerrit van der Steege (BML/BT)

NHL Stenden Hogeschool/VHL,
locatie Leeuwarden (H)



Figuur 1: de toenemende behoefte aan datavaardigheid binnen de BML-opleidingen

Aandeel data-onderwijs in de opleiding.

Het ~-teken geeft aan dat dit onderwijs deels ingebed is in ander onderwijs, waardoor het aantal EC een indicatie is.

	Statistiek en dataverwerking	Bio-informatica tools/ dry lab-aandeel	Software in verplicht curriculum	Bijzonderheden
Avans Hogeschool	~4EC	~4 EC	Excel, NCBI-tools, R (commander), labondersteuningstools	Minor Bio-informatica, 30 EC
Hanze-hogeschool Groningen	~6 EC	~7 EC	Excel, Excel Statistical Analysis pack, SPSS, NCBI-tools, GIMP, Ensembl, UniProt, ExPASy, Clustal X, labondersteuningstools, beetje programmeren	Minor bio-informatica, hbo-master Data science (vanaf collegejaar 2019-2020)
Hogeschool Inholland	BML: ~6 EC BT: ~7 EC	BML: ~6 EC BT: ~10 EC	Excel, NCBI-tools, labondersteuningstools	Minor Informatics, 5 EC alleen voor Biotechnologie groen: Galaxy, BioViz genome-browser, beetje programmeren
Hogeschool Leiden	~8 EC	~10 EC	Excel, NCBI-tools, Galaxy, Geneious, Mascot, Ensembl, BioMart, lab-ondersteuningstools, beetje programmeren	Minor Genomics, 15-30 EC Keuzeproject big data 15 EC
Hogeschool Rotterdam	~9 EC	~ 9 EC	Excel, Excel Statistical Analysis, NCBI-tools, PyMOL	Keuzeproject, 14 EC met R, specialisatie richting 'Complexe apparatuur en data-analyse'
Hogeschool van Arnhem en Nijmegen	~8 EC	~14 EC	Excel, NCBI-tools, lab-ondersteuningstools, RNA-seq analyse	
Hogeschool Utrecht	~5EC	~10 EC	Excel, NCBI-tools, lab-ondersteuningstools, SPSS, Perl en Bash Scripting voor Life Sciences	Keuzeminor 'Bio-informatica' in ontwikkeling; twee keuzevakken waarin R voor Life Sciences centraal staan
NHL Stenden Hogeschool, loc. Emmen	~9 EC	~3 EC	Excel, NCBI-tools, labondersteuningstools	
NHL Stenden Hogeschool/VHL, loc. Leeuwarden	~5 EC	~7 EC	Excel Statistical Analysis pack, SPSS, NCBI-tools, labondersteuningstools	Ontwikkeling van major (15 EC) Biological data sciences, waarbij programmeren in R wordt ontwikkeld; start vanaf sept. 2018
Saxion	~9 EC	~9 EC	Excel, SPSS, NCBI-tools, labondersteuningstools,	Biomedische research en diagnostiek met keuzemogelijkheid, 6 EC programmeren in Perl

Sommige opleidingen (G, A, L, HU) bieden specifieke bio-informatica- of big data-vakken aan binnen het verplichte curriculum. Bij de overige opleidingen is dit onderwijs geïntegreerd in ander onderwijs. Het snelle veranderen van technieken en tools op dit gebied wordt als complicerend ervaren voor het ontwikkelen en onderhouden van onderwijs op dit gebied. Hoofdzaak is om de student flexibele vaardigheid op dit gebied bij te brengen en niet per se dieptekennis op het gebied van een specifieke tool.

‘Sommige studenten gaan actief op zoek naar afstudeeronderwerpen waar een grote data-analyse-component in zit’

– Marc Teunis, Hogeschool Utrecht

Daarnaast wordt de didactische waarde van bio-informaticaonderwijs genoemd: omdat het virtueel plaatsvindt, is het een nuttige situatie om onderzoeksvaardigheden te oefenen, zonder dat er een heel practicum aan vooraf hoeft te gaan (IH). Er zijn enkele opleidingen die in het kader van wederzijdse begripvorming tussen BML en bio-informatica een aantal programmeerlessen aanbieden (HU), verplicht stellen voor studenten biotechnologie en als keuze aanbieden voor studenten BML (IH) of verplicht stellen voor alle BML studenten (L, G).

De meeste opleidingen geven aan dat onderwijs in datavaardigheden ingebed kan worden in bestaande modules. Een enkeling merkt op dat het curriculum al vrij vol is, en dat er een afweging gemaakt zou moeten worden waarvoor dit dan in de plaats kan komen.

Naast het verplichte curriculum kunnen studenten BML zich op verschillende hogescholen specialiseren met een minor, of met een specialisatie op het grensvlak van BML en bio-informatica (zie tabel op blz. 12).

Hanzehogeschool Groningen ontwikkelt op dit moment een hbo-master op dit gebied, die in collegejaar 2019-2020 zou moeten starten.

Data stewardship

Er zijn in deze opleidingengroep nog geen opleidingen die specifiek aandacht voor data stewardship

hebben binnen de opleiding, afgezien van het gebruik van E-labjournals (G, L, IH, A, N, HR) en de signalering dat dataoverdracht door studenten nog niet altijd even soepel gaat. Wel staan vrijwel alle opleidingen open voor enig bewustwording op dit gebied (SA, G, L, IH, HR, HU, N, A, H), bij voorkeur in de vorm van een gastcollege of in sommige gevallen in de vorm van praktijkervaring in projecten. Daarnaast geeft de meerderheid van de BML-opleidingen (L, G, IH, SA, HR, HU, H) aan het zinnig te vinden om iets over hedendaags databeheer op te nemen in de competentie ‘Beheren’.

Een aantal opleidingen staat open voor opname van big data-analyse in de BoKS (A, SA, HU, L, H, IH, HR). Hiervoor moet dan wel goed gedefinieerd worden wat onder big data wordt verstaan binnen dit vakgebied.

Discussiepunten

Vaak wordt genoemd dat het aandeel bio-informatica van het vakgebied lijdt onder het verschijnsel

Data science voor alle BML-studenten

Jurre Hageman, docent Biologie en medisch laboratoriumonderzoek aan de Hanzehogeschool Groningen:

‘Al onze BML-studenten krijgen op het moment, naast de benodigde statistiek, drie bio-informaticamodules in de hoofdfase. Zo ontwikkelen ze een goede basis op het gebied van bijvoorbeeld het werken met de NCBI- en Ensembl-databanken, eiwitpredicties, alignments, primer-ontwerp en simulatie van kloneringen. Ook mogen ze allemaal proeven aan programmeren en besteden we aandacht aan datavisualisatie. Deze modules geven de BML-docenten zelf, met zelf ontwikkelde dictaten. Hiervoor zijn we mede door de bio-informaticaopleiding hier bijgeschoold, voor zover we dat nodig en interessant vonden. Het is een andere tak van sport, maar met wat oefening prima te doen. Ik ben zelf van oorsprong bioloog, maar geef nu bijvoorbeeld ook eerstejaars-

vakken Pythonprogrammeren bij bio-informatica.

Bij sommige studenten zie je gewoon dat de dry lab/datascience-kant van het veld heel goed past. Zij kunnen in jaar 3 bij ons kiezen voor de minor Bioinformatics for Life Science students, waarmee ze de diepte in gaan: zelfstandig NGS-analyses uitvoeren, een RNA-seq dataset analyseren en data analyseren en visualiseren met behulp van de statistische taal R en de scripttaal Python.

Zijn ze dan helemaal om, dan kunnen ze vanaf februari in collegejaar 2018-2019 meedoen aan de hbo-master Data Science for Life Science, waaraan we samenwerken met onder andere de Rijksuniversiteit Groningen en het UMC van Groningen. In die master leren studenten met meer geavanceerde analysemethodieken en complexe datastructuren, big data, te werken. We zetten vol in op multi-omicsintegratie – daar valt prachtig onderzoek te doen!’

‘onbekend maakt onbemind’, zowel bij studenten als bij docenten van de opleidingen. In sommige gevallen wordt verwacht dat deze kant van het werk ‘moeilijk’ is of dat je heel handig met computers moet zijn om dit te (kunnen) leren. Om deze onderwerpen nader in de opleidingen te kunnen integreren, is mogelijk een vorm van pr op dit gebied nodig. Avans Hogeschool heeft geconstateerd dat de benaming ‘dry lab’ beter aanslaat bij deze groep dan de term ‘bio-informatica’, hoewel er vaak hetzelfde mee wordt bedoeld. Mogelijk kan ook het op laagdrempelige wijze kennis laten maken met bio-informatica-analyses of programmeren dit soort ideeën wegnemen.

‘We moeten studenten
zo opleiden dat ze verder denken
dan het bandje in de gel’

– Romeo Lascaris, Inholland

Wensen

Gevraagd naar de wensen van de opleidingen op dit gebied, worden de volgende ideeën genoemd. Ze zijn geordend van vaak naar minder vaak genoemd.

- Een grote biomedische dataset, waartoe het onderwijs vrij toegang heeft, die goed gekarakteriseerd is (door het vakgebied) en waarvan bekend is wat er aan kennis uit te halen valt (IH, G, L, SA, HR, N, H). Avans Hogeschool en de Hogeschool Utrecht hebben aangegeven over dit soort datasets te (kunnen) beschikken en open te staan voor samenwerkingen. In het ideale geval zou over deze dataset ook onderwijs ontwikkeld zijn (of worden) in de vorm van opdrachten voor studenten, docent-handleidingen en nakijkmodellen. De opleidingen zouden delen hiervan direct kunnen inzetten voor het onderwijs.
- Docentscholing op het gebied van analyse van grote datasets zoals NGS-analyses of geavanceerde statistiek (HR, ST, A, IH, L).
- Specifiek onderwijs dat datavisualisatie aan studenten leert (G, L, HU), wellicht in combinatie met exploratory data-analyse.
- Een gemeenschappelijk te ontwikkelen minor op dit gebied (A, HR, N).
- Onderwijs over grote MS-analyses (IH, HR).
- Een modulair te gebruiken onderwijsbron voor statistieklessen, waarbij geschikte datasets

beschikbaar zijn, met name over niet-parametrische toetsen (IH, dit is ook door chemie-opleidingen genoemd).

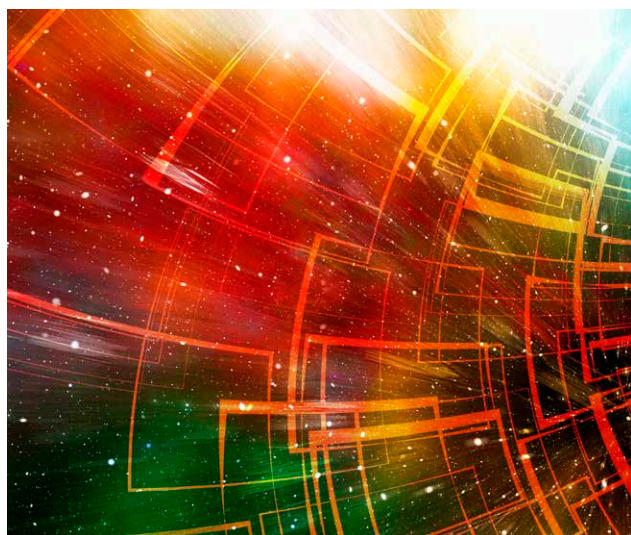
- De verweving van statistiek onderwijs en big data-onderwijs (L).

Adviezen

Advies 1: uitwisseling van onderwijsmateriaal

Verschillende hogescholen hebben aangegeven behoefte te hebben aan uitwisseling van onderwijs op dit gebied. DAS zou dit op verschillende manieren kunnen faciliteren. Bijvoorbeeld:

- samenbrengen van geïnteresseerde docenten op soortgelijke wijze als de groep digitalisering;
- organiseren van een materiaaluitwisseling-‘cursus’. Hierbij zou elke deelnemende docent een scholingsdag (deel) kunnen verzorgen waarin materiaal van zijn opleiding wordt gedoceerd aan de andere deelnemers. Dit zou in een maandelijkse bijeenkomst kunnen, of bijvoorbeeld in een volle week achter elkaar. Voordeel is dat ook docenten kunnen aanhaken die slechts bepaalde onderdelen willen volgen, of die (nog) geen materiaal te delen hebben;
- Marc Teunis, hoofddocent aan de Hogeschool Utrecht, organiseert samen met SURFsara een Applied Research Data Boot Camp voor docenten op 2 november 2018. Hiervoor ontvangen alle geïnteresseerden een uitnodiging. Wellicht zou hierin ook in de toekomst gefaciliteerd kunnen worden door DAS, in combinatie met de SURF sharekit waarover DAS nu al in gesprek is.



Advies 2: een dataset om mee te spelen

Het beschikbaar maken van een grote biomedische dataset, waartoe het onderwijs vrij toegang heeft, die goed gekarakteriseerd is (door het vakgebied) en waarvan bekend is wat er aan kennis uit te halen is. Hogescholen Avans en HU hebben aangegeven over dit soort datasets te (kunnen) beschikken en open te staan voor samenwerkingen. Hiervoor zou gezamenlijk onderwijs ontwikkeld kunnen worden in de vorm van opdrachten voor studenten, docentenhandleidingen en nakijkmodellen. Daarin zouden data-analyse, inzet van bio-informaticatools en datavisualisatie, en statistiek bij een relevante casus aan de orde moeten komen. De opleidingen zouden delen hiervan direct kunnen inzetten als onderwijs. Een en ander zou op soortgelijke wijze aangepakt kunnen worden als de veiligheidsmodule die onlangs door DAS is gefaciliteerd. Dan is wel van belang dat er voldoende tijd voor wordt vrijgemaakt door het betrokken management.

Advies 3: samenwerking met DTL voor data stewardship

In samenwerking met de Data Steward Interest Group (onderdeel van DTL) zou een gastcollege datamanagement ontwikkeld kunnen worden. Zodra er een framework voor een dergelijk gastcollege ligt, kan er op elke hogeschool die daar behoefte aan heeft, een lokale data steward dit gastcollege komen geven. DTL en de Data Stewards Interest Group hebben al aangegeven dat zij hier graag bij zouden willen helpen.

Big Data module

Roel Schelland, docent Moleculaire biologie, Hogeschool Leiden:

‘In onze module Big Data krijgt de student een indruk wat er komt kijken bij de analyse van big data. Hiervoor wordt hij of zij geïntroduceerd in programmeren met Python en wordt er dieper ingegaan op het gebruik van verschillende genome-

browsers. Daarnaast analyseert de student zelf een grote NGS-dataset met twee verschillende tools, Galaxy en Geneious. Ook leert de student proteomicsdata te analyseren met Mascot. Tot slot wordt de student uitgedaagd om een onderbouwde ethische beslissing te nemen voor terugkoppeling van big data van een patiënt.’

Advies 4: uitgebreidere werkveldinventarisatie

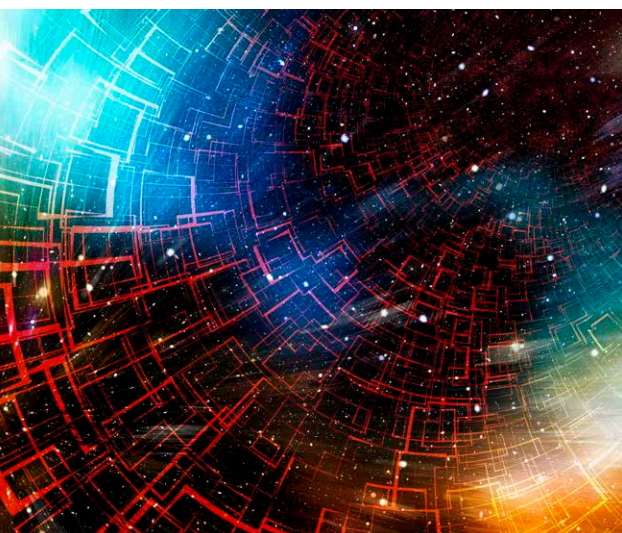
Om het onderwijs goed te laten aansluiten aan de behoeften van het werkveld, moet duidelijk worden wat er binnen de kaders van een BML-afgestudeerde precies wordt bedoeld met (big) data-analysevaardig. Dit werd expliciet genoemd door ST, IH en SA.

Uit de gesprekken met de andere hogescholen bleek ook dat er nog wel verschil bestaat in wat men denkt dat precies de extra datavaardigheden van een BML-afgestudeerde zouden moeten zijn. Vaak wordt NGS-analyse genoemd, maar in verschillende gradaties (met behulp van tools en pipelines, of juist helemaal zonder dergelijke hulpmiddelen); ook proteomics wordt door sommige hogescholen onder big data geschaard, terwijl andere dat juist niet ‘big’ vinden. Het zal geen gemakkelijke klus zijn om hier helderheid in te scheppen, aangezien er ook in het werkveld nogal wat discussie bestaat over wanneer data ‘big’ zijn en wanneer niet.

Ook zal het per bedrijf/instelling dat werknemers op dit gebied aanneemt, verschillen welke vaardigheden essentieel zijn en welke optioneel. Hier zou een apart onderzoek naar kunnen worden gedaan. Aangezien elke opleiding over een eigen werkveldadviescommissie beschikt, zou dit ook op opleidingsniveau onderzocht kunnen worden. Dit laatste heeft als voordeel dat er dan ook ruimte is voor specialisatie van de verschillende opleidingen op dit gebied, toegespitst op dit specifieke werkveld.

Advies 5: de competentie beheren updaten

De competentie beheren uitbreiden, zodat deze de hedendaagse status op het gebied van e-labjournals en vindbare data (opslag) reflecteert.





Opleidingsgroep Chemie

Geïnterviewden:

André van Roon

Hogeschool Leiden (L)

Ton Rulman

Hogeschool Rotterdam (HR)

Ad Dekkers

Avans Hogeschool (A)

Taco Graafsma

Saxion (SA)

Marcel van der Berge

HZ University of Applied Sciences (HZ)

Patrick Krooshof

Hogeschool Utrecht (HU)

Roel Baars

Hogeschool Inholland (IH)

Han Suelman

NHL Stenden Hogeschool/
VHL, loc. Leeuwarden (H)

Werkveldbehoefte

Bij organische chemie speelt dit onderwerp niet of nauwelijks, zo geven de opleidingen aan. Bij analytische chemie speelt de toename van data en de behoefte aan interpretatievaardigheden hiervan nu nog een kleine rol. Maar, geven eigenlijk alle opleidingen aan, pas afgestudeerde studenten zijn hierop voldoende voorbereid tijdens de opleiding (L, A, HZ, HU, SA, HR). Studenten brengen soms meer kennis mee dan in het bedrijf aanwezig is; bijvoorbeeld design of experiments is niet overal gebruikelijk, maar bedrijven zien er de voordelen wel van (A). Aan de VHL en NHL/Stenden wordt nu geïnventariseerd of hun werkveld er meer behoefte aan heeft (H). De opleidingen geven aan dat ze voorzichtige toekomsttrends signaleren op het gebied van multivariaatanalyse en geautomatiseerde metingen (SA, L, HR, HZ, HU).

Aandeel data-onderwijs in de opleiding

Gehanteerde werkvormen: computerlessen, werkcolleges, beroepsgerelateerde casussen uitwerken. De meeste opleidingen stappen af van hoorcolleges over dit onderwerp ten faveure van instructiecolleges, digitale colleges en praktijklessen op de computer.

Data stewardship

Enige kennis op het gebied van datamanagement en data stewardship klinkt als een nuttige toevoeging, vooral op het gebied van dataoverdracht tussen groepjes (L, A, SA, HZ). Hogeschool Rotterdam merkt op dat het een idee zou kunnen zijn om de studenten kennis laten maken met een LIMS, en hen de data zo te leren structureren dat een IT-er ermee verder zou kunnen. Sommige opleidingen

lijkt het zinnig om databeheer in de competenties of BoKS op te nemen (L, SA, HR, H).

‘Ik heb toevallig onlangs voor het eerst een minorstage goedgekeurd die puur uit data-analyse bestond’

– André van Roon, Hogeschool Leiden

Discussiepunten

Statistiek en dataverwerking worden door studenten niet als leuk ervaren, een punt waar meerdere opleidingen mee worstelen (L, H, A). Aan VHL en NHL/Stenden gebruikt de docent ‘verkeerde voorbeelden’ uit de literatuur om de stof meer te laten leven voor studenten. Welke analyse bij welk soort data gebruikt moet worden, is de essentie van statistiek en studenten blijven dit als ingewikkeld ervaren. Sommige opleidingen gebruiken beslisschema’s hiervoor, maar docenten geven aan dat het moeilijk blijft een en ander beeldend naar de beroepspraktijk te vertalen. Iedereen geeft aan open te staan voor uitwisseling van met name casuïstiek, in de hoop dat dit de lessen bruikbaar en interessanter maakt voor de studenten. HR geeft aan meer in projectvorm te willen werken.

Wensen

Op dit gebied zouden meerdere opleidingen multivariaatanalyse als ‘nice to have’ aan hun curriculum willen toevoegen. Het gaat dan om multivariate cursussen/boeken binnen de chemie, die goed te behappen zijn voor de studenten (L, HZ, A, HR, H). Daarnaast is er uitwisseling mogelijk

Aandeel data-onderwijs in de opleiding

	Statistiek en data-verwerking	Indicatie van behandelde onderwerpen	Gebruikte software	Bijzonderheden
Avans Hogeschool (A)	6+ EC	Univariate statistiek, betrouwbaarheidsintervallen, ANOVA, lineaire regressie, DoE, fit, factorial design, chemometrie	Excel, Labarchives, Minitab	Keuzeonderwijs waarin met datawarrior wordt gewerkt voor visualisatie, praktijkvak waarin katapulten worden gebouwd voor DoE-vaardigheden
Hogeschool Inholland (H)	8 EC	Univariate statistiek, ANOVA, betrouwbaarheidsintervallen, lineaire regressie, variatiecoëfficiënt, validatie	Excel	
Hogeschool Leiden (L)	± 20 EC	Univariate statistiek, ANOVA, lineaire regressie, DoE, factorial design, validatie van apparatuur	Excel, Minitab	Minor Metabolomics met multivariaatanalyse op gebruikersniveau, onderwijs met katapulten voor DoE-vaardigheden
Hogeschool Rotterdam (HR)	± 16 EC	Univariate statistiek, lineaire regressie, ANOVA, factor-analyse, detectielimiet	Excel, E-labjournal	Minoronderdeel waarin met Matlab wordt gewerkt en Arduino's worden geprogrammeerd
Hogeschool Utrecht (HU)	± 15 EC	Univariate statistiek, betrouwbaarheidsintervallen, chemometrie, DoE	Excel, Matlab	Keuzeminor 'Bio-informatica' in ontwikkeling
HZ University of Applied Sciences	5 EC	Univariate statistiek, ANOVA	Excel	Biochemiemodule met NCBI-skills
NHL Stenden Hogeschool/ VHL, loc. Leeuwarden	± 5 EC	Univariate statistiek, betrouwbaarheidsintervallen, experimenteel design		Ontwikkeling van Biological Data Science majortrack
Saxion (SA)	18 EC	Univariate statistiek, betrouwbaarheidsintervallen, chemical fingerprinting inclusief gegevensbeheer	SPSS, Excel, intro in Python	Multivariaatanalyse met R- en Perl-programmeren in minor

Experimental design

Ad Dekkers, hogeschooldocent Avans:

‘Ik geef met veel plezier statistiek aan alle Applied Sciences-opleidingen hier. Ze krijgen de theorie in een behoorlijk tempo in gespecialiseerde modules aangeboden, maar het moet natuurlijk echt tot uiting komen in praktijksituaties, zowel hier als tijdens hun stages. Die vertaalslag maken – dat vinden studenten soms best lastig, en dat heeft er misschien ook wel mee te maken dat het niet ieders favoriete vak is. Door zo snel mogelijk met relevante casussen aan de slag te gaan, proberen we ze vroeg te laten zien wat de waarde van dit vakgebied is. Maar soms moet je een beetje creatief zijn bij dit soort dingen!

Mijn favoriete onderdeel is design of experiments (DoE). De studenten maken in groepjes van vier zelf een katapult met minimaal vijf factoren die vijf standen hebben. Uiteindelijk laten ze zien dat de katapult een door mij opgegeven afstand binnen een bepaald interval kan raken, binnen drie kansen. Zo brengen ze de concepten van experimental design in de praktijk en gaat het echt voor ze leven. Het werkt zelfs zo goed dat onze studenten het ook gaan toepassen tijdens stage en afstuderen en DoE zo ook bij de bedrijven (weer) onder de aandacht wordt gebracht.’

op het gebied van chemometrie (A, HZ, HU). HR wil graag meer doen aan geavanceerde chemometrie, zoals PCA en machine learning.

**‘Onze studenten brengen soms
zelfs statistische kennis
bij hun werkgever binnen’**

– **Ad Dekkers**, Avans Hogeschool

Twee opleidingen zouden meer richting het grensvlak van data, software en hardware willen gaan, en onderwijs willen uitwisselen op het gebied van automatisering (autosamplers/sensoren programmeren/Arduino's) (L, HR). Twee opleidingen geven aan het nut van R-statistics te willen verkennen (L) of uitbreiden met chemische casuïstiek (SA).

Adviezen

Nice to have: de opleidingen die behoefte hebben aan de toevoeging van meer hardware automatisering (L, HR) en machine learning (HR) zouden met de opleidingen Technische natuurkunde onderwijsmateriaal kunnen uitwisselen.

De invulling van overige wensen kan door onderlinge materiaaluitwisseling, eventueel gefaciliteerd door de DAS, worden bereikt.



Opleidingsgroep Chemische technologie

Werkveldbehoefte

Op het gebied van de chemische technologie signaleren alle opleidingen dat er toenemende vraag is naar datavaardigheid bij afgestudeerden. Die behoefte varieert van het inzetten van de geleerde statistische methoden in een praktijksetting, tot het leren omgaan met hele grote datasets gegenereerd door automatische sensortechnologie. Verder wordt opgemerkt dat studenten zelf aangeven meer behoefte te hebben aan (big) data-vaardigheden. De opleidingen geven aan dat dit voor alle CT-studenten een toenemend belangrijke vaardigheid wordt.

Data stewardship

Op het gebied van data stewardship zien de opleidingen geen noodzaak tot opname in het curriculum. Wel is het belangrijk dat studenten hun data goed leren opslaan; dit zou bijvoorbeeld in

projecten het beste ondervonden kunnen worden.

Discussiepunten

Avans geeft aan dat het curriculum al vrij vol is: waar zou dit onderwijs dan voor in de plaats moeten komen?

Wensen

Alle opleidingen melden dat zij het werken met grote datasets meer in het curriculum zouden willen verwerken. De Haagse Hogeschool geeft aan deze datavaardigheden graag te willen verwerken in onderzoeksvaardigheden in het algemeen. Hierbij zouden zijn graag 2 EC per periode aan deze bredere vaardigheden willen gaan besteden. Avans zou dit onderwerp graag meer aandacht geven (2->4 EC); dit geldt ook voor Rotterdam (richting 7 EC). Andere wens: representatieve datasets uit het werkveld om onderwijs op te ontwikkelen (A, HR). Avans wil graag experimental design toevoegen aan het onderwijspakket, binnen een projectsetting.

Advies

Deze opleidingen zouden het meest geholpen zijn met een representatieve dataset uit het werkveld, waarmee studenten de problematiek al tijdens de opleiding kunnen ervaren. Bij voorkeur is deze dataset goed gekarakteriseerd en is duidelijk welke informatie het werkveld hier graag uit geëxtraheerd zou zien. Hierop kan dan onderwijs ontwikkeld worden door de opleidingen, eventueel gezamenlijk door DAS gefaciliteerd.

De CT-opleidingen die behoefte hebben aan uitbreiding op DoE-gebied, kunnen dit onderwijs bij de Chemie-opleidingen vinden.

Geïnterviewden:

Urjan Jacobs

Hogeschool Rotterdam (HR)

Richard van Ommeren

De Haagse Hogeschool (HA)

Peter Bressers

Avan Hogeschool (A)

Aandeel data-onderwijs in de opleiding

	Statistiek en dataverwerking	Indicatie van behandelde onderwerpen	Gebruikte software	Bijzonderheden
Avans Hogeschool	~2 EC	Univariate statistiek	Excel, Minitab	Keuzemogelijkheid datawarrior voor datavisualisatie
De Haagse Hogeschool	~2EC	Veel wiskunde, statistiek vrij beperkt	Excel	
Hogeschool Rotterdam	~7EC	Univariate statistiek, regressie, betrouwbaarheidsintervallen	Excel	Er is een verkenning gaande op het gebied van machine learning; curriculumaanpassingen voor vergroting van datavaardigheden is gaande

Opleidingsgroep Forensisch onderzoek

Geïnterviewden:

Anne Koster

Hogeschool van Amsterdam
(HvA)

Ben de Rooij

Avans Hogeschool (A)

Werkveldbehoefte

De forensische opleidingen signaleren dat het werkveld het waardeert als de afgestudeerden statistisch goed onderlegd zijn. Daarnaast neemt het aandeel van datascience toe, en dus is het wenselijk dat de afgestudeerden Forensisch onderzoek zich bewust zijn van de veiligheids- en privacyaspecten van databeheer. Daarnaast speelt het combineren van informatie uit verschillende databases, het opzetten van geschikte databases en machine learning voor profilering een rol.

Data stewardship

Bij beide opleidingen is nog geen aandacht voor data stewardship. HvA geeft aan dat een vorm van bewustwording hier wenselijk is: forensisch wetenschappers werken binnen beveiligde systemen waarin veel schillen zijn afgeschermd. Het is moeilijk hiermee te oefenen binnen een onderwijssetting. Avans geeft aan dat men graag zou willen uitbreiden op het gebied van statistiek: meer Bayesiaanse statistiek en chemometrie.

Dit betreft de majortrack Forensisch Laboratorium Onderzoek aan de Avans Hogeschool, te kiezen bij het basisprogramma van Biologie en medisch laboratoriumonderzoek, of Chemie.

Avans Hogeschool geeft aan het niveau van big data niet met de studenten Forensisch Laboratorium Onderzoek te bereiken. Daarom is op het moment niet nodig om uitbreiding op dit gebied te zoeken.

Bijzonderheden

Daar Forensisch onderzoek een heel brede opleiding is, speelt hier in hoge mate mee dat het curriculum al vrij vol is. Als er meer statistiek/meer data-analyse in zou worden opgenomen, moet er eigenlijk ook iets uit – maar wat dan?

FO is de enige opleiding waar het leren analyseren van digitale sporendragers een rol speelt. Hierin heeft de opleiding raakvlakken met de informatica-opleiding Forensisch ICT. Maar het bewustzijn van welke apparaten als digitale sporendrager kunnen worden aangemerkt en wat daar dan mee gedaan kan worden, is ook voor de studenten Forensisch onderzoek van belang.

Wensen

- Een minor dataspecialisatie.
- Contact met andere opleidingen die aan data-onderwijs doen voor de uitwisseling van materialen.
- Een pasklare dataset van een relevante casus met bijbehorend onderwijsmateriaal (zie ook BML).
- Docentenbijscholing op het gebied van validatie, datamanagement en beheer.

Adviezen

De major track FO bij Avans Hogeschool zou in de geadviseerde samenwerking voor BML kunnen meedraaien op het gebied van een casus, en bij chemie op het gebied van chemometrie. De opleiding FO aan de HvA heeft wellicht meer baat bij contacten met de bio-informaticaopleidingen, of met Forensisch ICT.

Aandeel data-onderwijs in de opleiding

	Statistiek en dataverwerking	Bio-informaticatools/ dry labaandeel	Software in verplicht curriculum	Bijzonderheden
Avans Hogeschool	8 EC	5 EC	Excel	Keuzevak bio-informatica, 3 EC
Hogeschool van Amsterdam	5 EC	6 EC	Excel	Databasewonderwijs



Opleiding Milieukunde

Werkveldbehoefte

Milieukunde heeft te maken met een zeer breed werkveld, waarin een toename van datavaardigheid bij afgestudeerden als wenselijk wordt beschouwd door de adviesraad van de opleiding. Daarnaast is datavisualisatie erg belangrijk, aangezien afgestudeerden vaak in een adviserende rol werkzaam zullen zijn.

‘Studenten zouden
niet moeten schrikken van
een grote dataset’

– Femke Brouwer, lecturer Environmental
Technology, Avans Hogeschool

Aandeel van de opleidingen aan data-onderwijs

Er wordt statistiek gegeven op basisniveau (univariate analyse), en de studenten krijgen te maken met datasets die nog gemakkelijk in Excel passen, zoals luchtkwaliteitsmetingen. Momenteel ondergaat de opleiding een curriculumwijziging waar meer aandacht zal zijn voor grotere datasets. Hierbij zijn met name ook experimentele opzetten

van belang: hoe verzamelt de student de juiste (hoeveelheid) data om de onderzoeksvraag goed te beantwoorden? Gevolgd door de juiste verwerking achteraf, met als eindresultaat een duidelijk figuur waarmee de boodschap in een beleidssetting overgebracht kan worden.

Data stewardship

Op dit gebied signaleert de opleiding nog geen behoefte.

Wensen

- Toevoeging van big data en geïntegreerde statistiek (R-statistics) aan het curriculum.
- Een dataset vanuit het werkveld, passend bij een voor studenten aansprekende casus. Wellicht read outs van automatische sensortechnologie.

Advies

Samenwerking met Technische natuurkunde, Chemie en eventueel Chemische technologie op het gebied van sensortechnologie. Hieruit zou eventueel ook een dataset kunnen komen die door het onderwijs uitgewerkt en gebruikt kan worden. Voor vaardigheid op het gebied van R kan contact gezocht worden met bio-informaticaopleidingen of minoren van de BML-opleidingen op dit gebied.

Geïnterviewde:

Femke Brouwer
Avans Hogeschool



Opleidingsgroep Technische natuurkunde

Geïnterviewden:

Rick Walraven

Fontys Hogeschool TNW (F)

Derek Land

De Haagse Hogeschool (H)

Ruud Sniekers

Saxion (S)

Werkveldbehoefte

Binnen het werkveld van Technische natuurkunde is niet direct behoefte aan uitbreiding van de datavaardigheid, of het werken met big data. De studenten zijn voldoende geschoold om tijdens hun afstuderen en als beginnend beroepsbeoefenaar met data aan de slag te gaan. Een enkele keer komt een afstudeerder in aanraking met big data. Met kleinere datasets (past op een scherm) wordt veel gewerkt; hier sluit het onderwijs op aan. De opleidingen hechten

veel waarde aan statistische onderbouwing van de resultaten, soms zelfs meer dan de bedrijven of instellingen waar de (afstudeer)stages plaatsvinden.

Aandeel van de opleidingen aan data-onderwijs
De basis van data-analyse zit bij alle opleidingen verwerkt in praktijk- en projectonderwijs (4-12 EC). Daarnaast geven deze opleidingen een verplicht vak Statistiek (2 EC) en Design of Experiments (4-6 EC), wat keuze of verplicht kan zijn.

Masterclass machine learning

Rick Walraven, docent Technische Natuurkunde op Fontys Hogeschool:

‘Machine learning is een interessant groeigebied voor onze studenten. Daarom zijn we dit jaar met een masterclass hierover gestart, met groot succes! Dertien excellente studenten Technische natuurkunde uit alle leerjaren, verdeeld over vijf groepen, gebruikten de sensoren uit hun mobiele telefoons om de actuele beweging van een testpersoon te classificeren. Ze maakten onderscheid tussen staan, liggen, zitten, lopen, traplopen en rennen, plus nog twee bewegingen naar keuze.

Ze hebben zelf een database van ruwe data opgebouwd, de nuttige informatie eruit gehaald (de zogenaamde features), en met Matlab een model gebouwd. Ondersteunende colleges hielpen de studenten inhoudelijk op weg, maar het bleek een pittige opgave. De beweging

“huilen in een hoekje” werd dan ook serieus als keuzebeweging overwogen. Voor de lagerejaars-studenten was het extra moeilijk, omdat ze een deel van de benodigde wiskunde nog niet hadden gehad.

Maar uiteindelijk hebben alle groepen een mooi resultaat neergezet.

Bij de eindwedstrijd werden docenten ingezet als testpersonen. Het uitvoeren van de foetushouding was geen probleem, maar bij “opdrukken met handjeklap” vloede er toch wat bloed.

De winnende groep wist hun stage-ervaring en opgedane kennis van programmeren, Fourier-transformaties en signaalverwerking te combineren en ieders verwachting te overtreffen. Ze wisten 96% van hun dataset juist in te delen en dit ook nog eens weer te geven in een mooi uitgewerkte GUI.’

Alle opleidingen geven aan dat hun studenten het moeilijk vinden om, buiten het specifieke vak statistiek om, te beoordelen welke data-analyse geschikt is voor de data die zij verzamelen, of in andere situaties moeten analyseren. Hiervoor zou wellicht nog meer aandacht kunnen komen.

Data stewardship

Er is op dit moment in de opleidingen nog geen aandacht voor data stewardship, en gezien de geringe omvang van de gebruikte datasets lijkt dit niet op korte termijn te veranderen. Er is binnen de opleidingen geen behoefte aan uitbreiding op dit gebied, en opnemen in de competentie 'beheren' wordt niet aangeraden.

Gebruikte/onderwezen software:

Excel (allen), Minitab (S). Python (H), Matlab (H+F), Labview, SPSS (F)

Bijzonderheden

Bij deze opleidingen speelt het programmeren van hardware als een van de weinige een duidelijke rol. Twee van de drie opleidingen maken gebruik

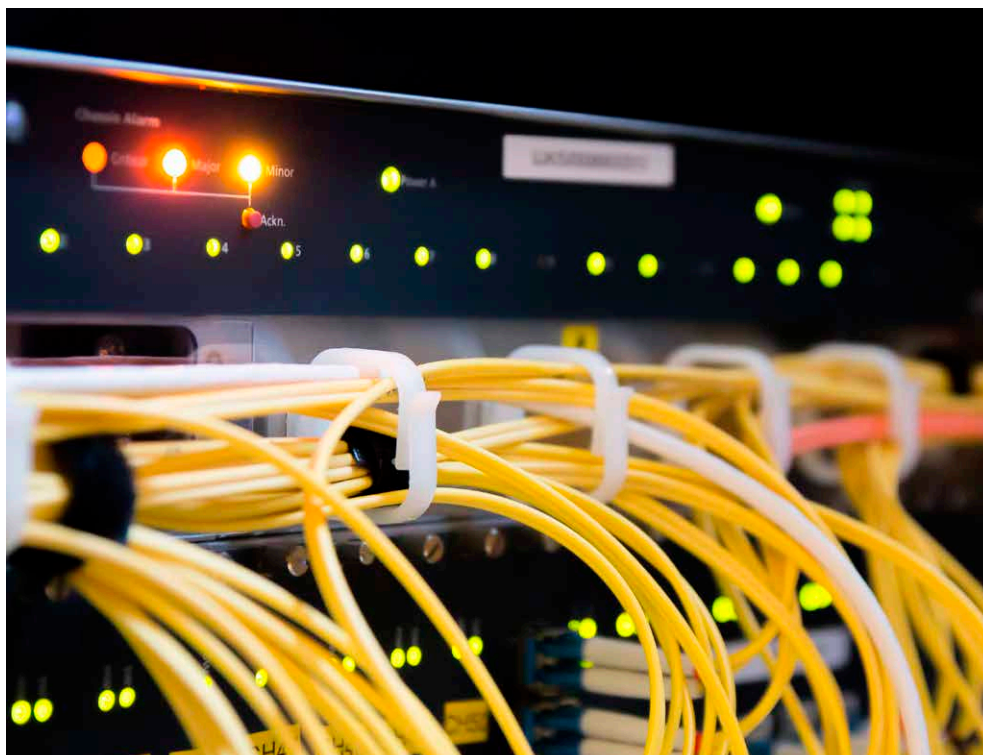
van Arduino en Raspberry Pi voor de acquisitie van datasets door middel van sensoren.

Wensen

De gebieden Lineaire regressie, Machine Learning en (meer) Programmeren worden door de verschillende opleidingen genoemd als mogelijke toevoegingen aan het curriculum. Big data-analyse zou wellicht in een minor kunnen worden opgenomen. Daarnaast zou Design of Experiments wellicht verplicht kunnen worden in alle drie de opleidingen, waarna opname in de BoKS mogelijk wordt. Tot slot is er de behoefte aan data van een goede case study uit het werkveld waarvoor onderwijs ontwikkeld zou kunnen worden.

Advies

Contact faciliteren tussen deze opleidingen zodat zij, indien gewenst, op dit gebied onderwijsmateriaal kunnen uitwisselen. Daarnaast contact faciliteren tussen deze opleidingen en de analytische chemieopleidingen, voor kennisuitwisseling op het gebied van Design of Experiments en sensortechnologie.



Conclusies en aanbevelingen

Data-onderwijs is, en blijft, een belangrijk onderdeel van alle Applied Science-opleidingen. Uit deze inventarisatie is gebleken dat de behoefte aan uitbreiding op dit gebied sterk verschilt per opleidingsgroep. Als grootste groeigebied kan de opleiding Biologie en medisch laboratoriumonderzoek worden aangemerkt. Hier zijn de opleidingen al volop aan de slag gegaan om in deze werkveldbehoefte te voorzien. Voor de andere opleidingen ligt de behoefte iets verder in de toekomst, maar wordt goed gemonitord.

Advies 1: Digitale geletterdheid, een 21st century skill
Alle betrokken partijen⁵ zijn het erover eens dat (toenemende) datavaardigheid voor de DAS-professional in de toekomst een rol zal spelen, al wordt ook erkend dat dit vakgebied nog in ontwikkeling is. Daardoor is het in dit stadium vooral van belang dat de afgestudeerden van de opleidingen niet bang zijn om met flinke datasets aan de slag te gaan. Hierbij is een leergierige en flexibele werkhouding van groter belang dan vaardigheid met gedefinieerde tools of situaties. De Applied Science-professional hoeft niet alles te kunnen op ICT-gebied, maar op eigen vakgebied inhoudelijk kunnen communiceren met professionals uit ICT-disciplines (bio-informatica, informatica), is wel essentieel. Daarvoor is onderlinge kennismaking met deze andere disciplines, bij voorkeur al tijdens de opleiding, gewenst. Zo kan de datavaardige professional op zijn toekomstige werkplek als bruggenbouwer optreden tussen een lab- of onderzoeksafdeling en de ICT-afdeling.

Om dit te faciliteren, wordt de verschillende opleidingen aangeraden om te onderzoeken waar een en ander ingebed zou kunnen worden in het

bestaande onderwijs, in aansluiting op de overige adviezen.

Advies 2: Onderlinge uitwisseling data-onderwijs

Een van de meest genoemde wensen is het samenbrengen van docenten in het kader van 'een kijkje in de keuken'. Successen delen, van elkaar leren, zeker op een zich zo snel ontwikkelend vakgebied als data science. Het geeft opleidingen die nog aan diepgang binnen dit onderwerp moeten beginnen een laagdrempelig startpunt, en voorkomt dat het wiel door iedereen opnieuw uitgevonden moet worden. Met name bij de BML-opleidingen bestaan veel initiatieven op dit gebied, zoals data science tracks, minoren en zelfs uitgebreide plannen voor een hbo-master.

De waarschijnlijk succesvolste combinaties zijn weergegeven in figuur 2.

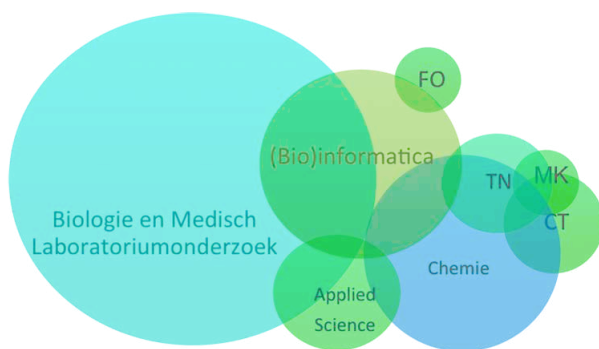
Hierbij zou een combinatie tussen de BML-opleidingen, Applied Science en wellicht Bio-informatica een groep kunnen vormen. Gezamenlijk beschikken deze opleidingen over onderwijs op het gebied van NGS-analyses, RNA-seq, proteomics, introducties in programmeren, data-handling en datavisualisatie.

Het tweede duidelijke samenwerkingsverband is Chemische technologie, Chemie en Technische natuurkunde, wellicht met Milieukunde en Applied Science waar gewenst. Deze groep kan eventueel een beroep doen op de Bio-informaticaopleidingen voor kennis in het kader van R of machine learning. Onderling kunnen zij inspringen op de aankomende sensortechnologische ontwikkelingen, waarbij veelvuldig en automatische (chemische) metingen moeten worden verwerkt.

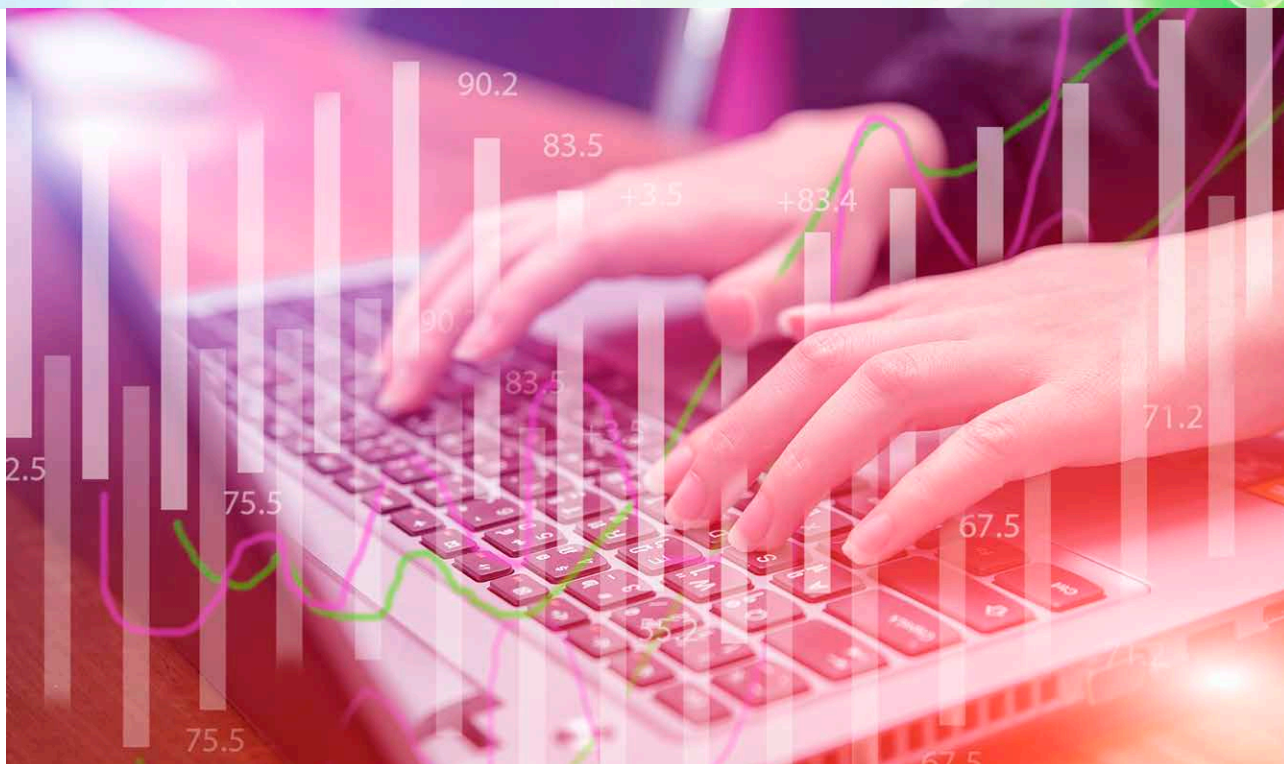
Periodieke datadag voor scholing en uitwisseling
Als startpunt voor de samenwerking hebben de contactpersonen die voor dit rapport zijn gesproken, aangegeven ervaringen te willen delen. Hierbij is van groot belang dat het ontwikkelde onderwijsmateriaal gedemonstreerd kan worden door de ontwikkelaars, ook omdat er soms infrastructuur voor op de hogeschool aanwezig moet zijn (servertoegang).

Marc Teunis, hoofddocent aan de Hogeschool Utrecht, organiseert samen met SURFSara een Applied Research Data Boot Camp voor docenten op 2 november 2018. Hiervoor ontvangen alle contactpersonen een uitnodiging. Wellicht zou hierin ook in de toekomst gefaciliteerd kunnen worden

5. De opleidingscontactpersonen, deelnemers DAS-conferentie workshop, DAS-werkgroep Onderwijs, werkveldvertegenwoordiging bij de professional van de toekomst, Raad van Advies



Figuur 2: model voor onderlinge samenwerking



door DAS, in combinatie met de SURFsharekit waarover DAS nu al in gesprek is binnen het Landelijk Overleg Digitalisering.

Daarnaast zou DAS een periodieke datadag kunnen faciliteren, waarbij telkens een andere hogeschool of groep van hogescholen als gastheer zou kunnen optreden. Deze opleiding of groep zou dan zijn onderwijs in vogelvlucht kunnen demonstreren aan geïnteresseerde docenten van de andere hogescholen, zodat uitwisseling kan plaatsvinden.

Advies 3: *Werkveldinventarisatie*

Om het onderwijs goed te laten aansluiten op de behoeften van het werkveld, moet duidelijk worden wat er binnen de kaders van elke afstudeerrichting precies wordt bedoeld met (big) data-analysevaardig.

Vanuit de bijeenkomst *Professional van de toekomst* kwam datavaardigheid in de breedste zin bij twee van de vijf kernonderwerpen naar voren. Hier werd de nadruk gelegd op de noodzaak in te spelen op de toenemende automatisering in laboratoria, waardoor het takenpakket van de DAS-professional van de toekomst ook meer data-analyse zal herbergen.

Deze onderwerpen worden ook bij de opleidingen genoemd; met name bij de BML-opleidingen wordt op verschillende plaatsen onderwijs ontwikkeld om hieraan tegemoet te komen. Bij de Chemie-opleidingen bleek tijdens de gesprekken dat er hier en daar wel een toename wordt gesignaleerd op het gebied van chemometrie, en dat automatisering in het lab als toekomstig gebied wordt

gesignaleerd. Maar tegelijkertijd geven de opleidingen aan dat het werkveld nog niet dusdanig om specifieke, nieuwe vaardigheden vraagt, dat de opleidingen daar sterk op moeten inspelen.

Opvallend hier is dat hierop tijdens de presentatie van de voorlopige conclusies van dit onderzoek aan de DAS Werkgroep Onderwijs en de Raad van Advies, verwonderd werd gereageerd. Deze organen geven aan dat zij deze noodzaak wel degelijk signaleren, ook bij de (organische) Chemie afgestudeerden. Hier is dus een verschil in beleving tussen de opleidingen en deze beleidsorganen te constateren.

Een specificatie vanuit het werkveld wat precies de kennis en vaardigheden zijn die zij wenselijk achten voor de (toekomstig) afgestudeerden van elke opleiding, zou meer duidelijkheid scheppen in dit onderwerp.

De overige opleidingsgroepen zouden hier ook baat bij hebben, want de definities van 'data-vaardig' en 'big data' lopen uiteen. Zo noemen verschillende BML-opleidingen NGS-analyse als een vorm van big data, maar definiëren andere dit pas als zodanig als er honderden patiënten samen worden genomen in een onderzoek. Wellicht zijn er bedrijven te identificeren die op dit gebied vooroplopen. Als de opleidingen dat als benchmark voor hun onderwijs kunnen gebruiken, zouden studenten de benodigde kennis wellicht bij de overige bedrijven kunnen meebrengen.

In acht nemend dat het werkveld van DAS, ook binnen de gedefinieerde opleidingsgroepen, altijd zeer divers zal blijven, zou een dergelijke inventarisatie

helpen om het concept datavaardig te concretiseren. In de eerste plaats zijn de landelijke overleggen hiervoor het startpunt, wellicht in combinatie met de Raad van Advies.

Maar aangezien er ook op lokaal niveau met werkveldadviescommissies overleg plaatsvindt, zou ook op dat niveau geïnventariseerd kunnen worden. Dit heeft als voordeel dat er dan ook ruimte is voor specialisatie van de verschillende hogescholen op dit gebied, toegespitst op hun lokale werkveld.

Advies 4: *Een dataset om mee te spelen*

Het onderwijs heeft behoefte aan één of meerdere datasets, beschikbaar gesteld door praktijkpartners of lectoraten, waartoe het vrij toegang toe heeft, die goed gekarakteriseerd zijn (door het vakgebied) en waarvan bekend is wat er aan kennis uit te halen is. Hoe helderder de casus, hoe bruikbaar voor het onderwijs, juist als de data ongestructureerd zijn.

Per opleidingengroep zou een dergelijke dataset gezocht of gecreëerd moeten worden met praktijkpartners: een biomedische set (bij voorkeur menselijke data) voor de BML-opleidingen (Topsector LSH, Nederlandse Vereniging Innovatieve Geneesmiddelen), een metabolomics/chemometrische set voor Chemie (Nederland Metabolomics Centre, COAST), een (proces)technologische set voor CT en TN, enzovoorts.

In samenwerking met de praktijkpartners die een dergelijke set goed kennen, over de benodigde

expertise beschikken en kunnen aangeven wat afgestudeerden hiermee zouden moeten doen, kan dan door de betrokken opleidingen gezamenlijk onderwijs ontwikkeld worden. Idealiter gebeurt dit in de vorm van opdrachten voor studenten, docenthandleidingen en nakijkmodellen. Hierin zouden data-analyse, inzet van relevante tools, datavisualisatie en statistiek aan de orde moeten komen bij een relevante casus. De opleidingen kunnen dit, of delen hiervan, direct inzetten in hun onderwijs. Gezien de snelle ontwikkelingen op dit vakgebied, raadt de Raad van Advies aan om dit onderwijs niet meer vast te leggen dan strikt noodzakelijk. Het vermogen om flexibel met data om te leren gaan, wordt door alle partijen als essentieel gezien; de specifieke tools die hiervoor worden gebuikt, in mindere mate.

Om dergelijke onderwijsontwikkeling mogelijk te maken, is wel van belang dat hier voldoende tijd voor wordt vrijgemaakt door het betrokken management, en dat de benodigde faciliteiten zoals serverruimte of rekencapaciteit op clusters beschikbaar worden gesteld aan alle deelnemende opleidingen.

Om dit onderwijs nog meer richting de toekomst te tillen, zou ingezet kunnen worden op multi-omicsintegratie. Hiervoor is (nog) geen geschikte dataset voor het onderwijs beschikbaar. Deze zou in samenwerking met praktijkpartners verzameld of gecreëerd kunnen worden. Hierbij kan de Data Science-master die in Groningen door de Hanze-hogeschool wordt ontwikkeld voor collegejaar 2019-2020 een voortrekkersrol spelen.

Advies 5: *Data stewardship*

In samenwerking met de Data Stewards Interest Group (onderdeel van Dutch Techcentre for Life Sciences) zou er een gastcollege datamanagement ontwikkeld kunnen worden. Zodra er een framework voor een dergelijk gastcollege ligt, kan op elke hogeschool die daar behoefte aan heeft, een lokale data steward dit gastcollege komen geven. DTL en de Data Stewards Interest Group hebben al aangegeven dat zij hier graag bij zouden willen helpen.

Daarnaast is het merendeel van de opleidingen ervoor om de competentie 'beheren' uit te breiden, zodat deze de hedendaagse status op het gebied van e-labjournals en vindbare data(opslag) reflecteert.



BIJLAGE I

Indicatievragenlijst voor opleidingsgesprekken

Inschatting signalen werkveld bij opleidingen

1. Heeft u een beeld van de behoefte van het werkveld van uw opleiding aan kennis en vaardigheden op het gebied van dataverwerking en -analyse? Zo ja, kunt u deze beschrijven?
2. Zouden alle afgestudeerden meer datavaardigheden moeten hebben dan nu, of een gedeelte van hen, of is wat ze nu krijgen voldoende?

Inventarisatie wat van er al is

3. Hoe maken uw studenten kennis met dataverwerking en -analyse?
4. Waar bestaan die analyses/verwerkingen uit?
5. Welke onderwijsvormen gebruiken jullie voor data-onderwijs?
6. Welke software is verplicht/keuze in het onderwijs?
7. Hoe belangrijk is statistiek bij uw opleiding? Wellicht andere modellen?
8. Hoe leren de studenten welke methode past bij welke soort data?
9. Zou dit nog uitbreiding behoeven?
10. In welke vorm?

Data stewardship

Onder 'data stewardship' wordt verstaan dat er binnen het onderzoek wordt nagedacht over de opslag, beheer, uitwisselbaarheid en vindbaarheid van de data tijdens en na het project. Een data steward kan al in de voorbereiding van een project adviseren en blijft ook lang na het project verantwoordelijk voor bepaalde acties rond de data. Bij grote EU-subsidies wordt een dergelijk Data Management Plan (DMP) als wenselijk beschouwd. Omdat analisten vaak een stabiele factor zijn

binnen onderzoeksgroepen, is het mogelijk dat zij hierbij in toenemende mate betrokken zullen raken in de toekomst.

11. Is er binnen de opleiding aandacht voor data stewardship?
12. Zo ja, in welke vorm?
13. Zo nee, zou u dat als een nuttige toevoeging beschouwen?
14. Welke vorm zou dat idealiter aannemen?
15. Zou opnemen in de competentie 'beheren' wellicht zinnig zijn?

Big data

16. In hoeverre komen uw studenten/afstudeerders in aanraking met zogenaamde Big data?
17. Zo ja, in welke vorm?
18. Zo nee, zou dat in uw ogen een nuttige toevoeging zijn? In welke vorm?
19. Zouden analyses van big data in de curricula moeten worden opgenomen, of niet? Zou hiervoor plaats zijn in de competenties/BOKS?

Afsluitend

20. Hoeveel EC zou wenselijk zijn dat er globaal aan data onderwijs besteed kunnen gaan worden binnen uw opleiding?
21. Als er uit dit onderzoek komt dat een gemeenschappelijke datacomponent wenselijk wordt gevonden door de verschillende opleidingen binnen uw landelijk overleg, welke vorm zou u dan ideaal vinden voor dit data-onderwijs?
22. Moeten docenten nog iets bijleren en zo ja, wat?

BIJLAGE II

Overzicht geïnterviewden per opleidingsgroep

1. Applied Science:

Fontys Hogeschool TNW
Zuyd Hogeschool

Ilse Vondervoort
Charlotte Creusen

2. Bio-informatica:

Hanzehogeschool Groningen
Hogeschool Leiden
Hogeschool van Arnhem en Nijmegen

Michiel Noback
Bo Blanckenburg
Martijn van der Brugge

3. Biologie en medisch laboratoriumonderzoek & Biotechnologie:

Avans Hogeschool
Hanzehogeschool Groningen
Hogeschool Inholland
Hogeschool Leiden
Hogeschool Rotterdam
Hogeschool van Arnhem en Nijmegen
Hogeschool Utrecht
NHL Stenden Hogeschool, loc. Emmen
NHL Stenden Hogeschool/VHL, loc. Leeuwarden
Saxion

Kees Rodenburg
Jurre Hageman
Romeo Lascaris
Ivo Horn
Eddy van der Linden
Jeroen van Kilsdonk
Marc Teunis
Rens van Leeuwen
Gerrit van der Steege
Taco Graafsma

4. Chemie:

Avans Hogeschool
Hogeschool Inholland
Hogeschool Leiden
Hogeschool Rotterdam
Hogeschool Utrecht
HZ University of Applied Sciences
NHL Stenden Hogeschool/VHL, loc. Leeuwarden
Saxion

Ad Dekkers
Roel Baars
André van Roon
Ton Rulman
Patrick Krooshof
Marcel van der Berge
Han Suelman
Taco Graafsma

5. Chemische technologie:

Avans Hogeschool
De Haagse Hogeschool
Hogeschool Rotterdam

Peter Bressers
Richard van Ommeren
Urjan Jacobs

6. Forensisch onderzoek:

Avans Hogeschool
Hogeschool van Amsterdam

Ben de Rooij
Anne Koster

7. Milieukunde

Avans Hogeschool

Femke Brouwer

8. Technische natuurkunde

De Haagse Hogeschool
Fontys Hogeschool TNW
Saxion

Derek Land
Rick Walraven
Ruud Sniekers

BIJLAGE III

Toelichting veelgebruikte termen

ANOVA

Statistische test om groepen te vergelijken.

Bash scripting

In Linux programmeren; wordt vaak gebruikt om automatisch na elkaar programma's aan te roepen.

DTL

Landelijk platform en expertisenetwerk van kennisinstellingen en bedrijven in de levenswetenschappen, gericht op data-analyse en data stewardship in onderzoek en innovatie. DTL faciliteert kennisuitwisseling, stimuleert samenwerking en vormt een landelijk schakelpunt voor training en onderwijs (www.dtls.nl).

GUI

Graphical User Interface: grafische gebruikersvriendelijke omgeving voor computerprogramma's, te bedienen met bijvoorbeeld een muis en knopjes, in plaats van alleen met tekstuele instructies.

LIMS

Lab Information Management System. Digitaal systeem om herkomst, gebruik en opslag van samples te bewaken.

Machine learning

Verzamelnaam voor computerprogrammatuur waarbij de computer patronen of verbanden in grote datasets vindt.

Matlab

Programmeertaal, veel gebruikt voor wiskundig modelleren.

MS-analyses

Massa-spectrometrie-analyses: een techniek om eiwitten (proteomics) en kleine chemische moleculen (metabolomics) te bestuderen.

Multi-omicsdata

Data waarbij meerdere -omicstechnieken worden

gecombineerd. Een veelgebruikt voorbeeld is als van een sample zowel het DNA (genomics), het RNA (transcriptomics) alsook de eiwitten (proteomics) of de kleine chemische moleculen (metabolomics) bekend zijn.

NGS-analyse

Next generation sequencing analyse. Verzamelnaam voor methoden die de lettervolgorde van grote hoeveelheden DNA-samples kunnen bepalen.

PCA

Principle Component Analysis. Mathematische bewerking om data met vele dimensies om te zetten naar een kleiner aantal samengestelde dimensies.

RNA-seq

Het analyseren van RNA door de nucleïnezuurvolgorde te bepalen door te sequencen.

Sequence-data

Gegevens van (meestal grote hoeveelheden) afgelezen DNA, of RNA (zie ook RNA-seq). Deze bevatten in elk geval de meest waarschijnlijke volgorde van de letters ATCG van het gesequenste sample, en kwaliteitsgegevens, afhankelijk van de techniek die gebruikt is om de data te verkrijgen. De grootte van de sequencedata kan variëren van enkele MB's voor een enkele DNA-streng, tot terabytes of meer als er meerdere grote organismen worden bestudeerd.

SPSS

Statistisch softwarepakket.

NCBI-tools

Verzamelnaam voor functionaliteiten aangeboden op de website van het National Center for Biotechnology Information.

R-statistics

Open source programmeertaal, veel gebruikt voor data-analyse en het creëren van figuren.

