

READER SURF STUDIEREIS STUDIEDATA 2019



SURF

INHOUDSOPGAVE

1. Learning analytics, how can we make it happen?	3
2. Verder lezen	11
3. Samenvattingen	13
I Samenvatting Wolfgang Greller en Hendrik Drachsler <i>'Translating Learning into Numbers: A generic Framework for Learning Analytics'</i>	14
II Samenvatting Hendrik Drachsler and Wolfgang Greller, <i>'Privacy and Learning Analytics – it's a DELICATE issue'. A checklist for Trusted Learning Analytics.</i>	16
III Samenvatting Martin Kurzweil and Mitchell Stevens, <i>'Setting the Table: Responsible Use of Student Data in Higher Education'.</i>	18
IV Samenvatting Simon J. Buckingham Shum and Timothy A. McKay, <i>'Architecting for Learning Analytics: Innovating for Sustainable Impact</i>	19
V Samenvatting Jerrold Grochow, <i>Data-Driven IT Strategic Planning: A Framework for Analysis'</i>	20
VI Samenvatting <i>Hoe data de kwaliteit van het onderwijs kunnen verbeteren</i>	21
VII Samenvatting <i>Learning Analytics in het onderwijs: een onderwijskundig perspectief</i>	22
4. Artikelen	
I Wolfgang Greller en Hendrik Drachsler <i>'Translating Learning into Numbers: A generic Framework for Learning Analytics'</i>	23
II Hendrik Drachsler and Wolfgang Greller, <i>'Privacy and Learning Analytics – it's a DELICATE issue'. A checklist for Trusted Learning Analytics.</i>	40
III Martin Kurzweil and Mitchell Stevens, <i>'Setting the Table: Responsible Use of Student Data in Higher Education'.</i>	51
IV Simon J. Buckingham Shum and Timothy A. McKay, <i>'Architecting for Learning Analytics: Innovating for Sustainable Impact</i>	58
V Jerrold Grochow, <i>Data-Driven IT Strategic Planning: A Framework for Analysis'</i>	72
VI <i>Hoe data de kwaliteit van het onderwijs kunnen verbeteren</i>	87
VII <i>Learning Analytics in het onderwijs: een onderwijskundig perspectief</i>	106

LEARNING ANALYTICS, HOW CAN WE MAKE IT HAPPEN?

Christien Bok

Copyright



4.0 internationaal

Beschikbaar onder de licentie Creative Commons Naamsvermelding 4.0
Internationaal. <https://creativecommons.org/licenses/by/4.0/deed.nl>

LEARNING ANALYTICS, HOW CAN WE MAKE IT HAPPEN?

Christien Bok

Christien Bok is impulsmanager onderwijsvernieuwing SURF. Met de impuls wil SURF een extra stimulans geven aan visievorming en ontwikkeling van onderwijsvernieuwing met ICT.

WOORD VOORAF

Metten voor het beste resultaat

Als impulsmanager onderwijsvernieuwing van SURF, de ICT-samenwerkingsorganisatie van het onderwijs en onderzoek in Nederland, probeer ik de innovatie van het onderwijs met ICT een impuls te geven. Een complex en kansrijk onderwerp dat zo'n impuls verdient, is learning analytics; het gebruik van studiedata. Het gebruik van data is in het dagelijks leven inmiddels al bijna gewoon. Zo ben ik zelf een fietser. Sinds een aantal jaar train ik met een hartslagmeter. Via een groot scherm op mijn sportschool zie ik real-time mijn prestaties en die van de andere sporters. Meten is weten, en daarom heb ik een hartslagmeting gedaan. De standaard rekensom "220 min je leeftijd" om je maximale hartslag te berekenen bleek voor mij niet op te gaan. Ik fiets nu ook met een hartslagmeter als ik in de Alpen fiets, omdat ik precies weet bij welke hartslag ik nog in de zuurstofrijke zone blijf. Op die manier kan ik het klimmen heel makkelijk volhouden. Dat maakt dat ik met nog meer plezier en vertrouwen in de bergen fiets.

Studenten maken in hun onderwijs steeds meer gebruik van digitale leeromgevingen, toetsystemen, discussiefora en allerhande digitale tools. Cijfers en andere gegevens worden opgeslagen in studievoortgangsystemen. Dat betekent dat er steeds meer data over hun studiegedrag en -resultaten beschikbaar zijn. En door beschikbare technologie om data te analyseren is het betaalbaar en bereikbaar geworden om die data te gebruiken om het onderwijs te verbeteren. Het gebruik van studiedata heet learning analytics. Die data bieden docenten bijvoorbeeld inzicht in de kwaliteit van hun leer materiaal. En studenten in hun studiegedrag en -voortgang. Want we willen dat studenten ook optimaal kunnen studeren en hun talenten kunnen benutten, net zoals ik optimaal een berg op wil fietsen.

Studiedata als motivator

Inzicht in gedrag en studievoortgang kan bijdragen aan motivatie. Deelname aan een honoursprogramma ligt misschien wel veel meer binnen bereik dan een student denkt. Als studenten met bijna vergelijkbare studieresultaten succesvol waren, kan dat andere studenten motiveren om zich toch aan te melden. Of misschien geeft een student juist extra gas als hij ziet dat studenten met vergelijkbaar gedrag het risico lopen om uit te vallen. Studiedata bieden ook inzicht in de effectiviteit van de manier waarop het onderwijs is ingericht. Als inzichtelijk wordt waar studenten over struikelen, kan een docent een onderdeel van de cursus of het cursusmateriaal aanpassen. Inzicht in studiegedrag helpt docenten om veel meer real-time feedback te geven. Niet meer na afloop van een cursus op basis van een tentamenresultaat, maar gedurende de cursus. Het gebruik

van al die mogelijkheden is nog lang niet wijdverbreid. Er is nog heel wat werk aan de winkel om dat op een goede, verantwoorde en geaccepteerde manier te doen. Alle redenen om er een brede discussie over te voeren.

SURF Studiereis Studiedata 2019

SURF organiseert in 2019 een Studiereis Studiedata naar Groot-Brittannië voor bestuurders en seniormedewerkers van universiteiten, hogescholen en mbo's om die discussie een impuls te geven. In die discussie verdienen verschillende aspecten van learning analytics de aandacht, zoals welke data je eigenlijk wilt gebruiken, en voor wat voor soort aanwijzingen je die data wilt gebruiken. Wie bepaalt er welke data worden gebruikt? En zie je de AVG als een rem of een hulpmiddel? Gebruik je data voor vrijblijvende of verplichtende feedback? Zijn volledig geautomatiseerde analyses de toekomst of juist niet? En hoe waarborg je het belang van studenten als ook commerciële partijen de mogelijkheden van data in het onderwijs zien? Er zijn geen pasklare antwoorden op deze vragen. In dit artikel schets ik zeven onderwerpen met bijpassende dilemma's waar het onderwijs voor staat. Daarmee ben ik niet compleet, maar ik geef er een eerste aanzet voor de discussie mee.

1 DATA VERZAMELEN

Is het “hoe meer data hoe beter” of niet? In grote hoeveelheden studiedata kan je met behulp van data-analyses correlaties vinden waar je niet naar op zoek was. Dat kan hele nieuwe inzichten opleveren. Big data wordt in allerlei sectoren ingezet. Waarom dan niet ook in het onderwijs? Maar dat betekent ook dat je die grote hoeveelheden data moet bewaren en analyseren, wat opslagruimte en rekenkracht kost. Je kunt door correlaties ook op het verkeerde been worden gezet, als je denkt dat er een causaal verband is dat er niet blijkt te zijn. Het vraagt veel onderzoek om causaliteit in de gevonden patronen te vinden.

Voor heel gerichte data-analyses gebruik je een beperkte set gestructureerde studiedata. Je weet wat je zoekt, en dat vind je ook. Je hoeft alleen de data die je voor je analyse nodig hebt te bewaren, je kan je analyse precies afstemmen op je doelstelling. Maar omdat je niet weet wat je niet weet, krijg je op deze manier niet snel nieuwe inzichten.

In het onderwijs wordt steeds meer gebruik gemaakt van een leeromgeving die is samengesteld uit verschillende losse onderdelen. De data uit één specifiek systeem bieden daarom maar een deel van de informatie. Leveranciers bieden je vooralsnog niet de mogelijkheid om alle data die zij in hun systeem opslaan gemakkelijk zelf te gebruiken. Om data uit verschillende bronsystemen op dezelfde, gestructureerde manier te kunnen opslaan, zijn afspraken nodig, zoals bijvoorbeeld over standaarden. Leveranciers maken nu nog niet gebruik van (dezelfde) standaarden.

DILEMMA: ALLE STUDIEDATA VERZAMELEN OF ALLEEN GERICHT VERZAMELEN?

2 IN-CONTROL

Of je nu veel of weinig data wilt verzamelen, het is belangrijk dat duidelijk is wie de eigenaar is van de data. Die bepaalt namelijk welke data gebruikt mogen worden, en voor welk doel. De AVG regelt dat de controle over data zo laag mogelijk belegd is. Er kunnen uitzonderingen op de regel zijn, als daar bijvoorbeeld een juridische grondslag voor is. Dat geldt bijvoorbeeld voor de opnamen die camera's in openbare ruimte van je maken. In het belang van veiligheid van de openbare ruimte ben je als burger geen eigenaar van deze data. Voor learning analytics betekent dat dat de student in controle is over zijn persoonlijke data en zelf beslist over het wel of niet delen van deze data met een docent of studiebegeleider. De praktijk is soms anders. Bijvoorbeeld doordat een docent gebruik maakt van een leeromgeving van een grote internationale leverancier. Die leeromgeving biedt weliswaar prachtige dashboards waarop de voortgang van studenten kan worden gemonitord, maar het eigenaarschap van de data ligt niet per se bij de student. SURF zorgt voor verwerkingsovereenkomsten met leveranciers, waarin geregeld is dat data eigendom blijft van de eigenaar. Maar grote bedrijven met een monopoliepositie hebben een broertje dood aan zo'n verwerkingsovereenkomst.

De belangen van de instelling en van de student zijn niet in alle gevallen gelijk. Misschien geven alleen gemotiveerde studenten toestemming om hun data te gebruiken. Inzicht in hoe een student presteert ten opzichte van medestudenten geeft dan een vertekend beeld. Overigens moet het volgens de AVG (zie ook dilemma 3) voor de student volstrekt helder zijn waarvoor welke data gebruikt worden.

DILEMMA: CONTROLE HOUDEN OF CONTROLE GEVEN?

3 PRIVACY

Behalve de vraag welke data we willen gebruiken en wie de controle heeft over het gebruik van die data, is er de vraag welke data we mógen gebruiken. Learning analytics kan alleen gebruikt worden binnen de kaders van de AVG. En daarmee helpt de AVG ons om studenten en docenten een veilige leeromgeving te blijven bieden. Nu is de AVG weliswaar helder over wat wel en niet mag, maar er is ook nog een groot grijs gebied waarin er ruimte is voor interpretatie en invulling. En er is nog maar weinig jurisprudentie. Voor het nog relatief nieuwe learning analytics is nog geen gearticuleerd en gedragen ethisch kader. De AVG biedt bescherming, maar er is, juist vanwege het ontbreken van duidelijkheid op veel punten, ook het risico dat de AVG verlamt. We doen maar niets, als we niet zeker weten of we conform de wet handelen. Jurisprudentie volgt alleen op basis van concrete casussen waar rechters uitspraken over doen. Je kunt jurisprudentie afdwingen door te gaan experimenteren. Uiteraard in volledige openheid naar en met toestemming van alle betrokkenen.

Dat heeft het risico dat je af en toe een beslissing neemt waar de rechter het achteraf niet mee eens is. Als je als sector nog verder mee richting wilt geven aan de duidelijkheid over de AVG, kan je een eigen privacykader ontwikkelen, en dat voorleggen aan de Autoriteit Persoonsbescherming.

DILEMMA: DE AVG: AFWACHTEN OF MEDE VORMGEVEN?

4 FEEDBACK

De grootste waarde van learning analytics is dat het de mogelijkheid biedt om zo gericht mogelijk feedback te geven. Voor een docent die vooral hoorcolleges voor groepen van 600 studenten geeft, kan dat heel behulpzaam zijn. En de mogelijkheid om bijna real time feedback te geven is van waarde, niet alleen na afloop van een cursus, maar tijdens het leerproces. Wil je learning analytics gebruiken voor het geven van vrijblijvende feedback, of moet de ontvanger er echt wat mee? In beide gevallen moet allereerst de feedback goed zijn. Feedback is gericht op de ontwikkeling van de ontvanger. En de feedback moet adequaat zijn, passend bij de omstandigheden, prestaties of het gedrag van de ontvanger. Formatieve feedback moet het vooral van zijn waarde hebben om opgevolgd te worden.

Je moet ook nadenken over wie de afzender en de ontvanger van de feedback zijn. Als de feedback volledig geautomatiseerd wordt gegenereerd, en zonder tussenkomst van een mens wordt verstuurd naar alleen degene voor wie de feedback is bedoeld, is dat per definitie vrijblijvender dan wanneer de feedback ook naar bijvoorbeeld de docent of de leidinggevende van de ontvanger wordt verstuurd.

Acceptatie is een belangrijke succesfactor voor learning analytics. Ook al kies je voor het geven van formatieve feedback, realiseer je dat in het onderwijs geen sprake is van gelijke verhoudingen. Zodra er sprake is van een gezagsverhouding, kan de lijn tussen motiveren en afrekenen dun worden. Als dat in de praktijk niet het geval is, kan dat nog steeds zo voelen. Een instelling kan bijvoorbeeld historische studiedata van eerstejaarsstudenten gebruiken om studenten zo goed mogelijk te helpen bij de overgang van de middelbare school. Maar dat kan studenten het schrikbeeld geven dat ze op basis van prestaties uit het verleden worden afgeschreven en geen eerlijke kans krijgen. In het onderwijs is er nu eenmaal ook sprake van gezags- en afhankelijkheidsverhoudingen.

DILEMMA: VRIJBLIJVENDE OF VERPLICHTENDE FEEDBACK?

5 DOELSTELLING

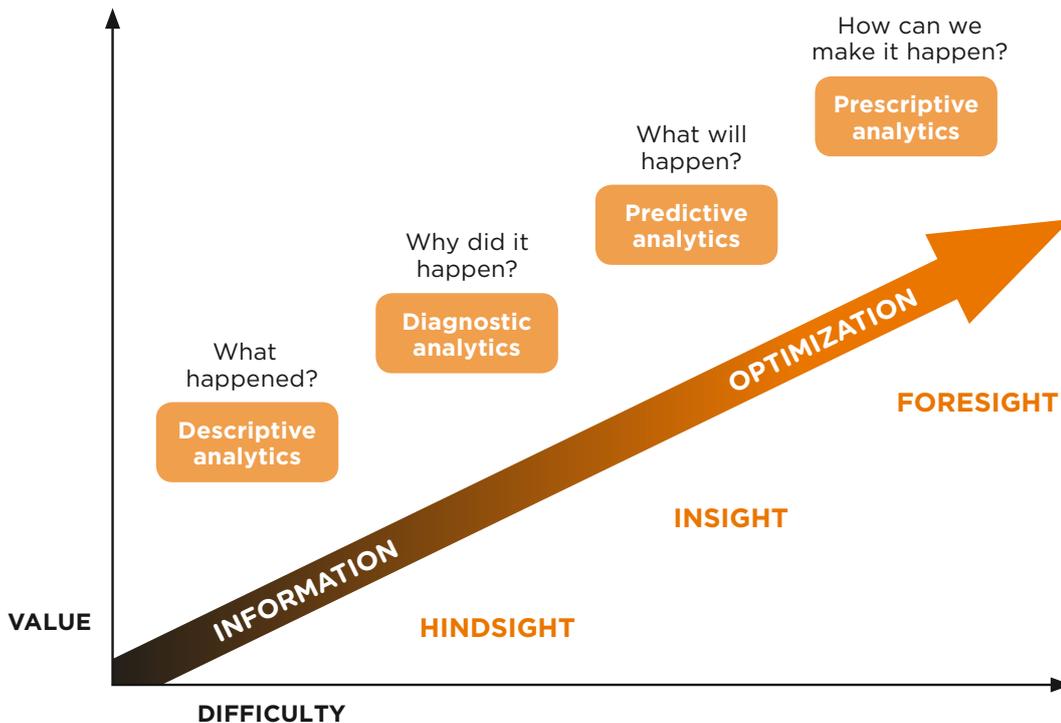
Gartner onderscheidt in het Analytics Maturity Model vier vormen van analytics: descriptive, diagnostic, predictive en prescriptive. Descriptieve en diagnostische analytics kijken terug, predictive en prescriptieve analytics kijken vooruit.

Echt rake voorspellingen zijn goud waard. Het kan een enorme impact hebben op het verminderen van studie-uitval en het verbeteren van studiesucces. Maar het is ook (nog) erg lastig om goede voorspellingen in het onderwijs te doen op basis van studiedata. Er is nog weinig bekend wat betekenisvolle data en patronen voor voorspellingen zijn. Het gevaar van een selffulfilling prophecy ligt op de loer: een advies “met dit studiegedrag zal je waarschijnlijk uitvallen” kan voor de student, die het bijltje er eerder bij neer gooit dan een tandje bijzet, heel demotiverend werken.

Het puur beschrijven van wat er gebeurt of is gebeurd, is minder spectaculair. Het is ook veiliger: er is geen kans op foute voorspellingen. Het kan wel degelijk waarde toevoegen. Experimenten tonen bijvoorbeeld aan dat de terugkoppeling van resultaten in kleuren (groen voor goed, oranje voor voldoende, rood voor onvoldoende), motiverender werken dan het geven van een cijfer. En een bericht dat je niet hebt ingelogd in je leeromgeving kan je motiveren dit alsnog te doen.

DILEMMA: INZICHT BIEDEN OF VOORSPELLEN?

GARTNER ANALYTICS MATURITY MODEL



Bron: ©Gartner

6 AUTOMATISERING

Dankzij Machine Learning en Artificial Intelligence kan het volledige proces van learning analytics worden geautomatiseerd: het verzamelen, opschonen, analyseren en visualiseren van data en het interpreteren van die analyses en visualisaties, en het kiezen en doen van een interventie. Een voordeel van automatisering van data-analyses in het onderwijs is dat het op grote schaal real time feedback mogelijk maakt. LinkedIn gebruikt bijvoorbeeld geautomatiseerde analyses voor de suggestie voor andere banen die je via LinkedIn ontvangt. Die zijn lang niet altijd in de roos. Niet-geautomatiseerde analyses zijn niet per definitie beter zijn dan geautomatiseerde analyses. Mensen hebben een bias en (voor)oordelen, die zij, al dan niet onbewust, laten meewegen. Algoritmen zijn overigens net zo biased, aangezien ze door mensen geschreven worden. Maar een computer kan in korte tijd wel veel meer data verwerken dan mensen.

DILEMMA: MENSENWERK OF VOLLEDIG AUTOMATISEREN?

In het onderwijs wordt gebruik gemaakt van nudging. Een nudge, een duwtje in de goede richting, bestaat in het onderwijs uit een automatisch gegenereerd bericht aan een student. Het lectoraat gedragseconomie van Hogeschool Rotterdam experimenteert daar bijvoorbeeld mee in het project [Nudging the student](#).

7 LEVERANCIERS

Leeromgevingen en toetssystemen bieden tegenwoordig bijna standaard prachtige dashboards die inzicht bieden in bijvoorbeeld studievoortgang. Met name de grote leveranciers beschikken over enorme ontwikkelkracht, zowel in menskracht als in geld, en zijn in staat om snel grote vooruitgang te boeken op de mogelijkheden om data te gebruiken. Maar bedrijven hebben toch vooral een commercieel belang. Bovendien beconcurreren ze elkaar en zijn hun ontwikkelingen alleen gericht op het eigen systeem. Dat brengt standaardisatie en interoperabiliteit niet dichterbij. Binnen onderwijsinstellingen is heel veel expertise op het gebied van data verzamelen, van dataopslag en van data-analyse. Van het interpreteren van analyses, en het maken van visualisaties. De publieke sector heeft het gebruik van open-sourcesystemen, met open access en open data als uitgangspunt. Maar al worden er binnen instellingen mooie experimenten uitgevoerd, de opschaaling van succesvolle innovaties is niet de kernexpertise van universiteiten en hogescholen. Bovendien geldt dat expertise wel in ruime mate aanwezig is, maar ook verspreid is over verschillende instellingen binnen specifieke onderzoeksgroepen.

DILEMMA: BETER ONDERWIJS OF COMMERCIEEL BELANG CENTRAAL?

SURF werkt samen met instellingen aan een ontwerp voor een infrastructuur die het gebruik van data uit verschillende bronsystemen mogelijk maakt. Vanaf 2019 gaat SURF daar met instellingen aan bouwen en mee experimenteren. SURF wil samen met instellingen werken aan open algoritmes en een open architectuur die voor iedereen bruikbaar is. Het versnellingssteam Veilig en betrouwbaar benutten van studiedata van het [Versnellingsplan Onderwijsinnovatie met ICT](#) gaat zich de komende jaren richten op de randvoorwaarden die nodig zijn om learning analytics te kunnen benutten in het onderwijs.

NAWOORD

Deze zeven onderwerpen en dilemma's zijn bedoeld als achtergrond voor de discussie die ruim vijftig senior-managers en bestuurders van universiteiten, hogescholen en roc's zullen voeren in februari en maart 2019, tijdens de SURF studiereis Studiedata 2019 naar Groot-Brittannië. Tijdens die reizen gaan we in gesprek met universiteiten die al gebruik maken van learning analytics. Met de studiereis wil SURF het belang van learning analytics agenderen, de deelnemers informeren over de verschillende aspecten ervan en hen inspireren ook binnen hun eigen instelling met learning analytics te gaan experimenteren. Tijdens de reis staat ook de vraag centraal of samenwerking binnen de sector op dit onderwerp nuttig of zelfs nodig is. Als het antwoord daarop ja is, zal SURF het initiatief nemen om die samenwerking te organiseren en vorm te geven.

Januari 2019

VERDER LEZEN

Verder lezen

Ter voorbereiding op de studiereis vind je in de reader SURF Studiereis Studiedata 2019 een zevental artikelen over verschillende aspecten van learning analytics. In de reader vind je een samenvatting van deze artikelen en de artikelen zelf.

We kozen voor het artikel *'Translating Learning into Numbers: A generic Framework for Learning Analytics'* (I) van Wolfgang Greller en Hendrik Drachsler omdat ze met hun zes factoren die je zou moeten betrekken in het denken over en werken met learning analytics de complexiteit van learning analytics schetsen.

Dezelfde auteurs bieden in het artikel *'Privacy and Learning Analytics - it's a DELICATE issue. A checklist for Trusted Learning Analytics'* (II) een praktisch stappenplan voor privacy by design.

Martin Kurzweil en Mitchell Stevens roepen in *'Setting the Table: Responsible Use of Student Data in Higher Education'* (III) op tot samenwerking met marktpartijen en het organiseren van een gezamenlijke verantwoordelijkheid voor verantwoord gebruik van studiedata.

Simon J. Buckingham Shum en Timothy A. McKay analyseren in *'Architecting for Learning Analytics: Innovating for Sustainable Impact'* (IV) waarom onderwijsinstellingen, met hun enorme schat aan ervaring op het gebied van data-analyse, zelf nog maar nauwelijks bijdragen aan de ontwikkeling van verantwoorde systemen, en onderzoeken manieren waarop het hoger onderwijs dat beter zou kunnen organiseren.

Jerrold Grochow benadrukt in zijn artikel *'Data-Driven IT Strategic Planning: A Framework for Analysis'* (V) het belang van een strategie voor datagebruik.

De SURF-paper *'Hoe data de kwaliteit van het onderwijs kunnen verbeteren'* (VI) biedt praktische handreikingen voor het werken met learning analytics, op basis van ervaringen die instellingen de afgelopen jaren hebben opgedaan.

SURF bracht in 2016 met het rapport *'Learning Analytics in het onderwijs: een onderwijskundig perspectief'* (VII) in kaart hoe learning analytics kan aansluiten bij het onderwijsontwerp, en op welke vragen learning analytics een antwoord kan geven.

SAMENVATTINGEN

SAMENVATTINGEN

I **SAMENVATTING:** Wolfgang Greller en Hendrik Drachsler, 'Translating Learning into Numbers: A generic Framework for Learning Analytics'

Wolfgang Greller en Hendrik Drachsler¹ onderscheiden zes dimensies van learning analytics: stakeholders, instruments, objectives, data, instruments, internal limitations en external constraints.

Alle zes dimensies zijn van belang en ze beïnvloeden elkaar. Voor wie zetten we learning analytics eigenlijk in (*stakeholders*)? Willen we data verzamelen om het curriculum te verbeteren, om docenten inzicht te geven in de effectiviteit van hun lesmateriaal of manier van lesgeven, of om studenten inzicht te geven in hun studievoortgang? Wie moet er vooral profijt hebben van learning analytics? De instelling, docenten of studenten?

En met welk doel zetten we learning analytics in (*objectives*)? Greller en Drachsler onderscheiden *reflection* en *prediction* als doelen. Het gaat bij reflectie vooral om de kritische zelfevaluatie van de instelling, docent of student. Voor prediction wordt gebruik gemaakt van profielen, op basis waarvan al dan niet automatische adviezen kunnen worden gegeven. Hier kleven veel ethische bezwaren aan, en het gevaar van vooroordelen ligt op de loer.

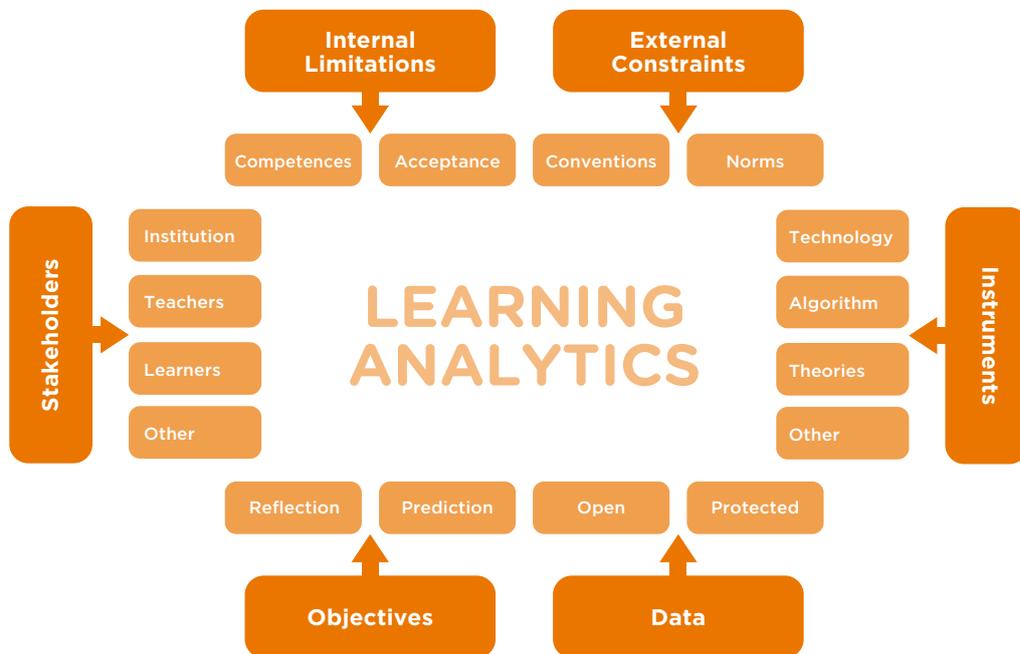
Afhankelijk van doelgroep en doel moet er een keuze gemaakt worden voor de data die gebruikt wordt. (*data*). Welke data zijn beschikbaar? Hoe kunnen we data uit verschillende bronnen gebruiken? En welke data zijn betekenisvol voor de doelen die we willen bereiken? En welke data mogen we gebruiken, onder andere onder de Algemene Verordening Gegevensbescherming? Gebruiken we data geanonimiseerd of niet?

Voor welke data geven studenten toestemming voor gebruik? En welke data gebruiken applicaties als onze digitale leeromgeving eigenlijk? En hoe is daarvoor de toestemming geregeld? Welke instrumenten gebruiken we om die data te analyseren (*instruments*)? Klassieke statistische analyse technieken? Machine learning? Welke algoritmen gebruiken we? En kan een algoritme wel neutraal zijn, of is die per definitie net zo biased als de ontwikkelaar van het algoritme? Verschillende algoritmen geven een totaal ander resultaat op dezelfde dataset. Het is belangrijk dat we ons daarvan bewust blijven.

Tenslotte zijn er de *external constraints* en *internal limitations* waar we rekening mee moeten houden. Externe beperkende factoren zijn bijvoorbeeld juridische en ethische aspecten. Interne beperkingen zijn bijvoorbeeld het ontbreken van de benodigde competenties of geen bereidheid om learning analytics te gebruiken. Het zijn uiteenlopende competenties die nodig zijn: data verzamelen, analyseren, visualiseren en vooral ook het interpreteren van analyses, maar ook juridisch en ethisch begrip.

¹ 'Wolfgang Greller and Hendrik Drachsler,'Translating Learning into Numbers: A generic Framework for Learning Analytics'. In: Educational Technology & Society, 15, (3) 2012, p. 42-57.

CRITICAL DIMENSIONS OF LEARNING ANALYTICS



Bron: ©Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics.

II **SAMENVATTING:** Hendrik Drachsler and Wolfgang Greller, 'Privacy and Learning Analytics - it's a DELICATE issue'. A checklist for Trusted Learning Analytics.

Van de zes dimensies die van belang zijn, zijn ethiek en privacy misschien wel de belangrijkste. Een ethisch kader biedt leidraden voor het nemen van verantwoorde beslissingen. Maar waar wordt de ethische discussie over learning analytics gevoerd? Als het gaat om onderzoek naar learning analytics, gelden de ethische uitgangspunten voor onderzoek: vrijwillige deelname aan onderzoek, expliciete toestemming, experimenteel onderzoek is in maatschappelijk belang, deelnemers mogen niet in gevaar worden gebracht, privacy en vertrouwelijkheid van informatie is gewaarborgd, en deelnemers kunnen altijd kiezen voor opt-out. Voor mensgebonden onderzoek zijn er ethische commissies die onderzoeksprojecten toetsen op deze uitgangspunten. Maar voor experimenten in de onderwijspraktijk zijn ook ethische uitgangspunten nodig.

Bezorgdheid en onzekerheid over wat wel en niet mag zorgen voor stagnatie van learning analytics. Als we gebruik willen maken van de voordelen van learning analytics, is een debat over ethiek en privacy van groot belang, en moeten we richtlijnen ontwikkelen die misbruik van data voorkomen. Universiteiten en hogescholen moeten hun studenten een veilige omgeving blijven bieden. De relatie met hun studenten moet gebaseerd blijven op wederzijds vertrouwen. Om hierbij behulpzaam te zijn, ontwikkelden Drachsler en Greller de DELICATE checklist for trusted Learning Analytics in education.²

² Hendrik Drachsler and Wolfgang Greller, 'Privacy and Learning Analytics - it's a DELICATE issue'. A checklist for Trusted Learning Analytics. April 2016, Conference Paper for 6th Learning Analytics and Knowledge Conference 2016, At Edinburgh.

THE DELICATE CHECKLIST

Implementing trusted Learning Analytics in education

D

DETERMINATION → Why want you to apply Learning Analytics

- What is the added value (Organisational and data subjects)
- What are the rights of the data subjects (e.g., EU Directive 95/46/EC)

E

EXPLAIN → What are the objectives and boundaries

- What data will be collected for which purpose?
- How long will this data be stored?
- Who has access to the data?

L

LEGITIMATE → Why you are allowed to have the data?

- Which data sources you have already (aren't they enough)
- Why are you allowed to collect additional data?

I

INVOLVE → Involve all stakeholders and the data subjects

- Be open about privacy concerns (of data subjects)
- Provide access to the personal data collected (about the data subjects)

C

CONSENT → Make a contract with the data subjects

- Ask for a consent from the data subjects before the data collection
- Define clear and understandable consent questions (Yes/No options)
- Offer the possibility to opt-out of the data collection without consequences

A

ANONYMIZE → Make the individual not retrievable

- Anonymize the data as far as possible
- Aggregate data to generate abstract metadata models (those do not fall under EU Directive 95/46/EC)

T

TECHNICAL → Procedures to guarantee privacy

- Monitor regularly who has access to the data
- If the analytics change, update the privacy regulations (new consent needed)
- Make sure the data storage fulfills international security standards

E

EXTERNAL → If you work with the external providers

- Make sure they also fulfill the national and organisational rules
- Sign a contract that clearly states responsibility for data security
- Data should only be used for the intended services and no other purposes

Bron: ©Hendrik Drachsler and Wolfgang Greller, 'Privacy and Learning Analytics - it's a DELICATE issue'.

III **SAMENVATTING:** Martin Kurzweil and Mitchell Stevens, 'Setting the Table: Responsible Use of Student Data in Higher Education'.

Martin Kurzweil and Mitchell Stevens³ kijken vanuit de Verenigde Staten misschien met enige jaloezie naar onze AVG. Want de wetgeving die betrekking heeft op studie-data is in de VS nog gebaseerd op de tijd dat die alleen op papier werden bewaard. Maar tegelijkertijd zijn ze ook sceptisch over de premisse dat gebruikers kunnen beslissen over wat er met "hun" data gebeurt. Bedrijven als Google, Amazon en Facebook verzamelen zoveel data over ons, en komen daardoor zoveel meer over ons te weten dan een overheid ooit over burgers heeft geweten, dat die individuele autonomie over eigen data een illusie lijkt.

Kurzweil en Stevens ontwikkelden vier principes voor beleid voor learning analytics, waarin zij nadrukkelijk ook leveranciers betrekken. *Shared Understanding* zorgt ervoor dat studenten, docenten, managers en leveranciers allen bijdragen aan learning analytics. Dat betekent dat zij allen geïnformeerd moeten zijn, en dat leveranciers medeverantwoordelijk zijn voor het debat over data-ethiek. Het betekent volgens de auteurs ook een gedeeld eigenaarschap van de data zelf. *Transparency* garandeert dat een student inzicht heeft welke data over/van hem of haar wordt verzameld door een instelling of een leverancier en op welke manier die data worden beoordeeld. Ook biedt het de student de mogelijkheid de beoordeling van die data te laten heroverwegen. *Informed Improvement* draagt instellingen op om door onderzoek het gebruik van studiedata te evalueren en te verbeteren. *Open Futures* tenslotte moet waarborgen dat learning analytics kansen biedt, en niet inperkt.

De discussie over deze principes zou niet alleen door het onderwijs gevoerd moeten worden, maar vooral ook samen met EdTech bedrijven.

³ Martin Kurzweil and Mitchell Stevens, 'Setting the Table: Responsible Use of Student Data in Higher Education'. In: *Educause Review* May/June 2018, p. 18-24.

IV **SAMENVATTING:** Simon J. Buckingham Shum and Timothy A. McKay, **Architecting for Learning Analytics: Innovating for Sustainable Impact**

Simon Buckingham Shum en Timothy McKay betogen dat universiteiten en hogescholen een veel grotere rol moeten spelen in de ontwikkeling van learning analytics.⁴ Hun motivatie daarvoor is niet alleen dat commerciële partijen minder aandacht voor ethiek hebben, maar ook dat zij vooral producten aanbieden gericht op de mainstream, niet op de “future-focused bleeding edge of teaching and learning”. En hoewel universiteiten en hogescholen de beste dataspecialisten, statistici, user interface ontwerpers in huis hebben, richten die zich nauwelijks op het mogelijk maken van verantwoorde inzet van learning analytics. Dat is te verklaren. We rekenen onze onderzoekers immers af op hun peer reviewed publicaties, niet op hun bijdrage aan de onderwijspraktijk. Bovendien zijn ze vrij in het kiezen van hun onderzoeksvraagstukken, en dat zijn zelden de uitdagingen van hun eigen instelling met data-analyse. En die uitdagingen lijken dan ook vooral op het bord van de IT-afdelingen te liggen, niet bij onderzoekers. Onderzoekers krijgen waardering voor innovaties, maar nauwelijks voor het ontwikkelen van schaalbare applicaties. Bovendien zijn de vaardigheden als software-ontwikkeling en interface-ontwerp vaak onvoldoende aanwezig in onderzoeksgroepen. Logisch dus, dat er binnen instellingen zelf nog maar weinig ontwikkeld wordt. Maar hoe neem je als instelling veel meer regie op de ontwikkeling van learning analytics? Je kan de verantwoordelijkheid leggen bij de IT-afdeling. Je maakt gebruik van de functionaliteiten die het LMS biedt. Vanuit Institutional research kunnen LMS-data worden geïntegreerd met data uit andere bronnen. Onderwijskundigen kunnen docenten ondersteunen bij het gebruik van learning analytics. Het nadeel van dit model is dat je als instelling erg leunt op de data en analyses van de producten van leveranciers.

Als onderzoekers, bij voorkeur in samenwerking met de IT-afdeling, aan zet zijn, staat innovatie veel meer centraal en kunnen vaak veel rijkere data worden gebruikt, zowel voor onderzoek als voor feedback voor studenten en docenten. Nadeel van dit model is dat er over het algemeen (te) weinig ontwerp, doorontwikkel- en onderhoudexpertise aanwezig is in onderzoeksgroepen voor inbedding van ontwikkelde tools in de infrastructuur van de organisatie.

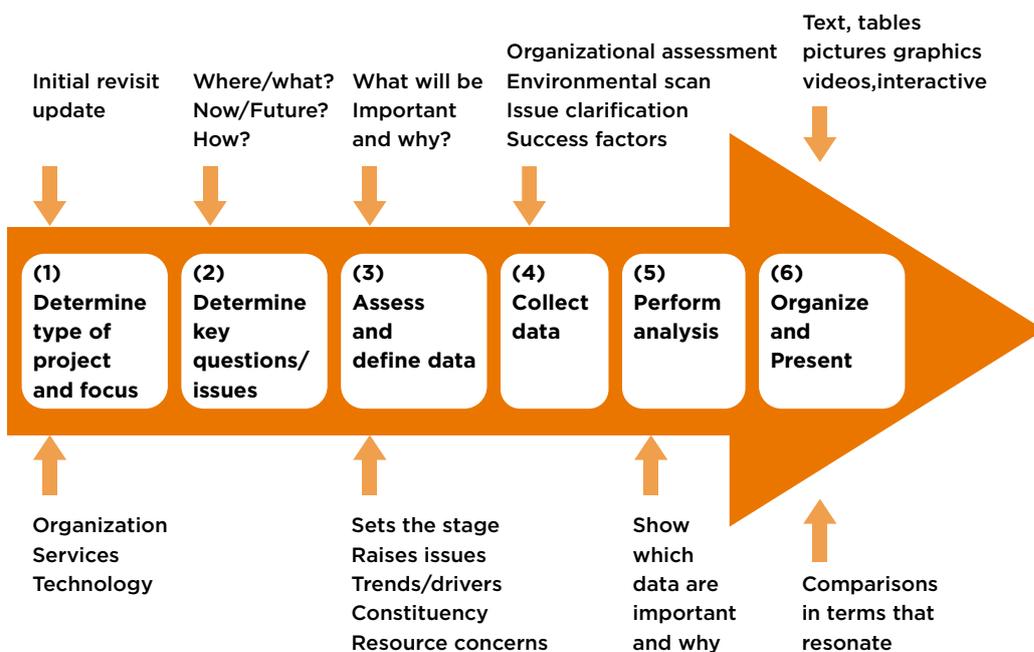
Buckingham Shum en McKay pleiten daarom voor een hybride model voor het werken aan learning analytics. In het zogenaamde Innovation Center Model werken onderzoekers samen met IT-specialisten. Het staat los van de IT-afdeling en de faculteiten. Zowel Shum als McKay heeft er ervaring mee. De University of Technology Sydney van Buckingham Shum heeft het Connected Intelligence Centre, waar onderzoekers werken die niet alleen op hun onderzoekskwaliteiten worden aangenomen, maar vooral ook op vaardigheden als multidisciplinair samenwerken en communicatie. In het centrum wordt gewerkt aan prototypes, kleinschalige pilots, grotere pilots en instellingsbrede implementatie van learning-analyticstoepassingen. De University of Michigan van McKay startte een Digital Innovation Greenhouse voor opschaling van succesvolle innovaties. Het overdragen van die innovaties aan de IT-afdeling werkte niet. Die afdeling heeft ervaring met uitontwikkelde producten, niet met het tot wasdom brengen van innovaties. De expertise die daarvoor nodig is, levert het Digital Innovation Greenhouse. In hun artikel beschrijven de auteurs hun eerste – positieve – ervaringen met het Innovation Center Model om learning analytics te innoveren en om de impact van learning analytics te vergroten.

⁴ Simon J. Buckingham Shum and Timothy A. McKay, ‘Architecting for Learning Analytics: Innovating for Sustainable Impact’. In: *Educause Review* March/April 2018, p. 26-39.

V **SAMENVATTING:** Jerrold Grochow, Data-Driven IT Strategic Planning: A Framework for Analysis

Jerrold Grochow benadrukt in zijn artikel 'Data-Driven IT Strategic Planning: A Framework for Analysis' het belang van een strategie voor datagebruik. Hij benadrukt dat het van belang is goed in kaart te brengen welke impact ambities voor data-analyses hebben op de infrastructuur, maar ook bijvoorbeeld voor de benodigde vaardigheden van personeel. Voordat je data gaat verzamelen, moet je eerst bepalen welk doel je hebt, en welke vragen je wilt beantwoorden, en welke data je dan precies nodig hebt voor het beantwoorden van die vragen. Pas daarna komt het verzamelen, analyseren en interpreteren van de data.

DATA-DRIVEN IT STRATEGIC PLANNING FRAMEWORK



Bron: ©Jerrold M. Grochow / Creative Commons License BY-SA-NC

VI SAMENVATTING: Hoe data de kwaliteit van het onderwijs kunnen verbeteren

In de paper *Hoe data de kwaliteit van het onderwijs kunnen verbeteren*⁵, laten een aantal praktijkvoorbeelden van Nederlandse en Vlaamse universiteiten en hogescholen zien welke keuzen docenten maken in learning analytics experimenten, en wat ze dat oplevert. Docenten ervaren over het algemeen meer inzicht in het leerproces van studenten, wat ze de mogelijkheid geeft hun keuzen daar beter op te kunnen afstemmen. Het biedt ze ook een hulpmiddel om het gesprek met studenten aan te gaan. Het real time aspect van de informatie en feedback zien ze als groot voordeel. Ook biedt het ze goede handvatten om de kwaliteit van hun leermateriaal te evalueren.

Maar er zijn nog voldoende uitdagingen. De docenten benadrukken het belang van het gebruik van data uit verschillende bronsystemen. Omdat te kunnen, is een infrastructuur nodig. En een multidisciplinair team, mensen die databronnen kennen en begrijpen, die data kunnen interpreteren, (voorspellende) modellen kunnen maken, statistische analyses kunnen doen, en interventies kunnen plagen. De AVG lijkt een showstopper, maar eigenlijk is er niet eens zoveel veranderd in de mogelijkheden voor datagebruik. Het is vooral de in- en externe verantwoording over wat je doet die goed geregeld moet zijn.

SURF werkt samen met instellingen aan een infrastructuur voor learning analytics, die het gebruik van data uit verschillende bronnen mogelijk maakt. Maar wil learning analytics echt impact hebben op het onderwijs, moet niet alleen de techniek goed zijn geregeld. Het allerbelangrijkste is dat we inzicht krijgen in welke data betekenisvol zijn, en welke interventies werken en waarom. Daarvoor is nog veel onderzoek nodig.

Ervaringsdeskundigen benoemen het belang om als onderwijssector pro-actief te zijn. We gaan ongetwijfeld toe naar een situatie waarin datagebruik in het onderwijs heel normaal is. Dan is het wel fijn als je zelf ook aan de knoppen zit om te bepalen hoe dat dan gebeurt. Laat het je niet overkomen, en laat de markt niet bepalen wat de randvoorwaarden en mogelijkheden zijn.

Je vindt in de publicaties tips voor het beginnen met learning analytics, en tips voor opschaling. Deze tips zijn gebaseerd op de ervaringen die een heel aantal instellingen de afgelopen jaren hebben opgedaan.

De KU Leuven richt zich vooral op eerstejaarsstudenten, omdat voor hen de overgang van de middelbare school groot is. Ze geven studenten inzicht in hun positionering ten opzichte van medestudenten. Ze gebruiken daarvoor cijfers van (tussen)toetsen en de uitkomsten van een vragenlijst. Studenten krijgen alleen actionable advies, waar ze echt wat mee kunnen doen, bijvoorbeeld via een aanbod van trainingen leer- en studievaardigheden. Dat is ook de reden waarom in de analyses geen gegevens over geslacht en economische status worden meegenomen. Studenten komen vaker op gesprek bij de studieloopbaanadviseur, en ook al vroeg in het semester, en vinden de tips die ze krijgen nuttig.

⁵ SURF, *Hoe data de kwaliteit van het onderwijs kunnen verbeteren*, 2019

VII **SAMENVATTING:** Learning Analytics in het onderwijs: een onderwijskundig perspectief

Waarom zou je learning analytics eigenlijk willen gebruiken? Dat gaat om meer inzicht in het leerproces, gerichte feedback aan studenten en verbetering van het onderwijs, betogen experts van Nederlandse universiteiten en hogescholen in de SURF-publicatie *Learning Analytics in het onderwijs: een onderwijskundig perspectief*.⁶ Maar om learning analytics effectief te kunnen benutten, moet een docent al in de ontwerpfase nadenken welke informatie behulpzaam kan zijn bij goede begeleiding. Het is te laat om tijdens of na afloop van een cursus te bedenken welke data beschikbaar zijn, en welke informatie dat kan opleveren. Daar moet al rekening mee gehouden worden bij het formuleren van leerdoelen en de toetsing, het ontwerpen van de leerfuncties, de inhoud en de leeractiviteiten, en om de keuzen van lesmateriaal, werkvormen en praktische factoren als tijdstip, groepsgrootte, etc.

Niet iedereen kan gemakkelijk “in data” denken. Zorg daarom voor voldoende expertise, en voor concrete onderwijskundige vragen die je beantwoord wilt hebben. Maar er spelen ook nog andere vragen een rol. Welke data mogen onder welke voorwaarden worden gebruikt? Op welke manier kunnen data uit (verschillende) systemen worden gebruikt? Welke visualisaties zijn effectief? Welke effecten heeft dit gebruik van data in het onderwijs?

Learning analytics hebben vooral zin als op basis van analyses effectieve interventies worden ingezet. Effectieve interventies vinden tijdig plaats, gaan uit van een persoonlijke benadering en hebben een doel en een middel. Automatiseren hoeft een persoonlijke benadering niet uit te sluiten, maar het is wel belangrijk om in het ontwerp van automatisch gegenereerde feedback dit uitgangspunt te hanteren.

⁶ SURF, *Learning Analytics in het onderwijs: een onderwijskundig perspectief*. 2016

ARTIKEL 1

‘TRANSLATING LEARNING INTO NUMBERS: A GENERIC FRAMEWORK FOR LEARNING ANALYTICS’

Wolfgang Greller en Hendrik Drachsler

copyrights

Published under open access.

Journal: <https://www.ds.unipi.gr/et&s/issues.php?id=56>

PDF: https://www.ds.unipi.gr/et&s/journals/15_3/4.pdf

Translating Learning into Numbers: A Generic Framework for Learning Analytics

Wolfgang Greller* and Hendrik Drachsler

Open University of the Netherlands, Valkenburgerweg 177, 6419AT Heerlen, The Netherlands //
wolfgang.greller@ou.nl // hendrik.drachsler@ou.nl

*Correspondence author

ABSTRACT

With the increase in available educational data, it is expected that Learning Analytics will become a powerful means to inform and support learners, teachers and their institutions in better understanding and predicting personal learning needs and performance. However, the processes and requirements behind the beneficial application of Learning and Knowledge Analytics as well as the consequences for learning and teaching are still far from being understood. In this paper, we explore the key dimensions of Learning Analytics (LA), the critical problem zones, and some potential dangers to the beneficial exploitation of educational data. We propose and discuss a generic design framework that can act as a useful guide for setting up Learning Analytics services in support of educational practice and learner guidance, in quality assurance, curriculum development, and in improving teacher effectiveness and efficiency.

Furthermore, the presented article intends to inform about soft barriers and limitations of Learning Analytics. We identify the required skills and competences that make meaningful use of Learning Analytics data possible to overcome gaps in interpretation literacy among educational stakeholders. We also discuss privacy and ethical issues and suggest ways in which these issues can be addressed through policy guidelines and best practice examples.

Keywords

Learning analytics, Framework, Educational data mining, Ethics, Domain design, Data for learning

Introduction

In the last few years, the amount of data that is published and made publicly available on the web has exploded. This includes governmental data, Web2.0 data from a plethora of social platforms (Twitter, Flickr, YouTube, etc.), and data produced by various sensors such as GPS location data from mobile devices. In the wake of this, data-driven companies like Google, Yahoo, Facebook, Amazon, etc. are growing exponentially by commercially exploiting such data for marketing or in the creation of new personalised services. The new “data economy” empowers companies to offer an increasing amount of data products at little or no cost to their users (e.g., Google Flu Trends, bit.ly customised URLs, Yahoo Pipes, Gapminder.com). This growth in data also renewed the interest in information retrieval technologies. Such technologies are used to analyse data and offer personalised data products customised to the needs and the context of individual users.

It is already evident that data in combination with information retrieval technologies are not only the basis for the emergent data economy, but also hold substantial promises for use in education (Retalis et al., 2006; Johnson et al., 2011). One example of this is the research on personalisation with information retrieval technologies which has been a focus in the educational field for some time now (Manouselis et al., 2010). The main driver is the vision of improved quality, effectiveness, and efficiency of the learning processes. It is expected that personalised learning has the potential to reduce delivery costs while at the same time creating more effective learning experiences, accelerating competence development, and increasing collaboration between learners.

Not so long ago, for universities and companies alike, gathering data on their users met with substantial limitations in terms of cost, time requirements, scope, and authenticity of the data, as this was typically done using questionnaires or interviews with a selected representative number of stakeholders. The new data economy has made data collection very much an affordable activity. It is based on the highly economic electronic data mining of people’s digital footprints and the automated analysis of behaviours of the entire constituency rather than sampling. Because data mining is not a separate act to normal user behaviour, the information retrieved is also highly authentic in terms of reflecting real and uninterrupted user behaviour. As such, data mining is more comparable to observational data gathering than to intrusive collection via direct methods. This will not make questionnaires and structured interviews

obsolete, but it will greatly enhance our understanding and highlight possible inconsistencies between user behaviour and user perception (Savage and Burrows, 2007).

The proliferation of interactive learning environments, learning management systems (LMS), intelligent tutoring systems, e-portfolio systems, and personal learning environments (PLE) in all sectors of education produces vast amounts of tracking data. But, although these e-learning environments store user data automatically, exploitation of the data for learning and teaching is still very limited. These educational datasets offer unused opportunities for the evaluation of learning theories, learner feedback and support, early warning systems, learning technology, and the development of future learning applications. This leads to the importance of Learning Analytics (LA) being increasingly recognised by governments, educators, funding agencies, research institutes, and software providers.

The renewed interest in data science and information retrieval technologies such as educational data mining, machine learning, collaborative filtering, or latent semantic analysis in Technology-Enhanced Learning (TEL) reveals itself through an increasing number of scientific conferences, workshops and projects combined under the new research term Learning Analytics. Examples are the *1st Learning Analytics conference in Banff, Canada, 2011*; *the 4th International Conference on Educational Data Mining 2011 in Eindhoven, Netherlands*; *the 1st dataTEL workshop on Educational Datasets for Technology-Enhanced Learning at the Alpine-Rendez-Vous conference La Clusaz, France 2011*; *the 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver 2012*; *the 1st Workshop on Learning Analytics and Linked Data (LALD 2012)*; and more. Thus, the increasing amount of dedicated research events and publications make a meta-analysis of the domain timely and needed in order to establish a solid scientific basis which facilitates the development of new learner-oriented services.

Critical dimensions of learning analytics

Despite the great enthusiasm that is currently surrounding LA, it also raises substantial questions for research. In addition to technically-focused research questions such as the compatibility of educational datasets, or the comparability and adequacy of algorithmic and technological approaches, there remain several “softer” issues and problem areas that influence the acceptance and the impact of Learning Analytics. Among these are questions of data ownership and openness, ethical use and dangers of abuse, and the demand for new key competences to interpret and act on LA results. We shall point at these issues in more detail below. This means that the implementation of LA in learning processes requires to be carefully crafted in order to be successful and beneficial.

This necessity motivated us to identify six critical dimensions (soft and hard) of LA, which need to be covered by the design to ensure an appropriate exploitation of LA in an educationally beneficial way. By soft issues we mean challenges that depend on assumptions being made about humans or the society in general, e.g., competences or ethics. They are opposed by the hard challenges of the fact-based world of data and algorithms (cf. also the similar soft-hard distinction in Dron, 2011). In its coverage of soft issues, our framework differs from other, more workflow oriented models for LA, like that by Siemens (2011), although in his presentation he does acknowledge these as of concern. Rather than being a process model such as those collected in Elias (2011), we aim at a description framework that can later be developed into a domain model or ontology.

The critical dimensions highlighted here have been deduced from discussions in the emerging research community using a general morphological analysis (GMA) approach (cf. Ritchey, 2011). In this early formation stage of the LA community, scientific exchanges such as open online courses (MOOC) in Learning and Knowledge Analytics (LAK11, LAK12), or the above-mentioned events and congregations soon began to revolve around a number of key questions, like: Who is the target group for LA? What are we trying to achieve? How do we deal with privacy and data protection? These questions are naturally extended by other on-going debates such as the openness of data, which has been a topic for some time in the EDM and Open Linked Data domain, as well as technical and theoretical questions on achieving meaningful extraction of information from data.

Our chosen approach leading to the proposed framework consisted of a number of gathering and analysis processes. First, as a matter of opinion mining, we scanned the scientific interactions from proceedings and presentations of the conferences and working groups mentioned above. We conducted a brief literature review of abstracts in the field of Learning Analytics and Educational Data Mining. Additionally, we scanned the live discussions on the LA Google Groups (<http://groups.google.com/group/learninganalytics> and <http://groups.google.com/group/LAK11>), as well as

the LAK11 MOOC (presentation chats and social networking exchanges). Furthermore, we looked back at recent RTD projects that contained elements of analytics and the questions and lessons they produced, e.g., the Language Technologies for Lifelong Learning project (<http://www.ltfll-project.org>) contained an analytics approach related to learner positioning and conceptualisation. Following these reviews, we applied cognitive mapping (Ackermann, Eden, and Cropper, 2004) for synthesising and sense making. We analysed these discussions and clustered them into the proposed six fields of attention, which we then presented as the first draft of the framework to a community of commercial and academic experts for evaluation and feedback at the *SURF seminar on Learning Analytics* (Eindhoven, 30-31 August 2011). The number six in the framework is not chosen for any particular reason, and other divisions are of course possible. However, we find the dissection into these six dimensions a useful and easy to follow domain orientation.

With the framework, we take the presumption that responsible designers of analytic processes will not only implement what is technically possible to do and legally allowed (or at least not prohibited), but to consider holistically the outcomes for stakeholders and, even more importantly, the consequences for the data subjects, i.e., the people supplying the data (cf. the section on stakeholders below). The framework intends to be a guide as much as a descriptor of the problem zones. Hence we refer to it as a “design framework” that can and should be used to design LA services from an inclusive perspective. We will argue below that this will help the transferability of LA approaches between different contexts of application and research.

Proposed design framework for learning analytics

Our proposed model for the domain and application of LA in figure 1 below considers six critical dimensions. Each of the dimensions can be subdivided into several instantiations falling into that dimension. For example, the generic “stakeholder” dimension can have instantiations (values) like “learners” and “teachers.” The list of instantiations in the diagram is not exhaustive and can be extended on a case-by-case basis. To stay with the above example, commercial service providers and even automated agents could also function as stakeholders in a LA process. It is useful to note that through connecting various (and also multiple) different instantiations of each dimension, concrete use cases can be constructed. We call the dimensions “critical” in the sense that each of the six fields of attention is required to have at least one instantiation present in a fully formulated LA design. We realise, though, that some dimensions are vaguer than others in this respect.



Figure 1. Critical dimensions of learning analytics

The six dimensions of the proposed LA framework are (cf. Figure 1): stakeholders, objectives, data, instruments, external constraints, and internal limitations. We will discuss each of these dimensions individually in the following and exemplify their instantiations and impact on the LA process and the benefits and opportunities they may determine. We will also elaborate apparent problem zones and limitations that may hinder any envisaged benefits.

Before embarking on the abstract dimensions in detail, we would like to illustrate the purpose and possible usage of the framework on the following sample use case, which is created out of a number of instantiations of the six dimensions. This specific example relates to conducting a social network analysis of students discussing in a forum using the SNAPP tool, based on the work by Dawson et al. (Dawson, 2008; Macfadyen & Dawson, 2010).

Table 1. Sample use case and values for dimensions

Dimension	Values
Stakeholders	<i>Data subjects</i> : a group of learners. <i>Data clients</i> : tutor, discussion moderator.
Objective	<i>Reflection</i> : Analyse student interactions in a forum discussion, identify network connections between students, and identify isolated students to bring them back into the discussion.
Data	<i>Protected dataset</i> : Student interactions and posts in the discussion forum of the LMS. <i>Relevant indicators</i> : Posts published, posts replied to. <i>Time scale</i> : what time frame is applied to the analysis?
Instruments	<i>Pedagogic theory</i> : socio-constructivist, hypothesis is that active participants in a discussion show better learning outcomes. <i>Technology</i> : Social Network Analysis (SNA), statistics. <i>Presentation</i> : network diagram visualisation, stats table.
External limitations	<i>Conventions</i> : (1) <i>Privacy</i> : is the analysis in accordance with privacy arrangements, are the students properly informed? (2) <i>Ethics</i> : What are the dangers of abuse/misguided use of the data? <i>Norms</i> : Are there e.g., legal data protection or IPR issues related to this kind of use of student data? <i>Time scale</i> : will the students still be able to benefit from the analytics outcome? Is the analysis post-hoc or just-in-time?
Internal limitations	<i>Required competences</i> : (1) <i>Interpretation</i> : Do the data clients have the necessary competences to interpret and act upon the results? Do they understand the visualisation or presentation of the information? (2) <i>Critical thinking</i> : Do they understand which data is represented and which data is absent? How will they use this information?

The above use case can be used (1) as a checklist when designing a purposeful LA process; (2) as a sharable description framework to compare context parameters with other similar approaches in other contexts, or for replication of the scientific environment. The framework allows an indefinite number of use cases with the respective value arguments.

Stakeholders

The stakeholder dimension includes *data clients* as well as *data subjects*. Data clients are the beneficiaries of the LA process who are entitled and meant to act upon the outcome (e.g., teachers). Conversely, the data subjects are the suppliers of data, normally through their browsing and interaction behaviour (e.g., learners). It is important to make this distinction in order to understand the impact of the process on individuals. In certain cases, the two types of stakeholder groups can be the same, as is the case if a LA application feeds back information to learners about their own learning rather than to inform the teacher, as would be a common case in informal learning scenarios. In the traditional learner-teacher scenario, the teacher would act as the data client, who receives information gathered from the data subjects, i.e., the learners.

As shown in the framework model (Figure 1), the main stakeholder groups of LA in formal learning situations are learners, teachers, and educational institutions. These may be expanded or substituted by other stakeholder groups, such as researchers, service providers, or governmental agencies. Each of the groups has different information needs and can be provided with tailored views on information using LA.

Information flow between stakeholders can best be exemplified with the common hierarchical model taken from formal education (Figure 2). What the diagram illustrates as an example is by which ways benefits might be obtained from LA. The pyramid encapsulates the academic layers of education and training institutions. In the most direct way, data analysis from the student level, e.g., via a LMS, can inform the above layer, in this case the teachers. Teachers can then use the analytics information to plan targeted interventions or adjust their pedagogic strategies. Institutions can, similarly, retrieve benefits from student and teacher data in order to provide staff development opportunities or to plan policies like quality assurance and efficiency measures. We also want to stress the major benefits LA offers for self-reflection on every level (cf. left side of the diagram). We would like to see institutions enabling and actively encouraging students to reflect on their learning data. But also teachers and institutions can gain new insights by reflecting on their performance. Not immediately involved in the learning processes, researchers (right of the diagram) could harvest data for the purpose of evaluating or innovating teaching processes or learning services. Finally (on top of the diagram), Government agencies may collect cross-institutional data to assess the requirements of Higher Education Institutes (HEI) and their constituencies.

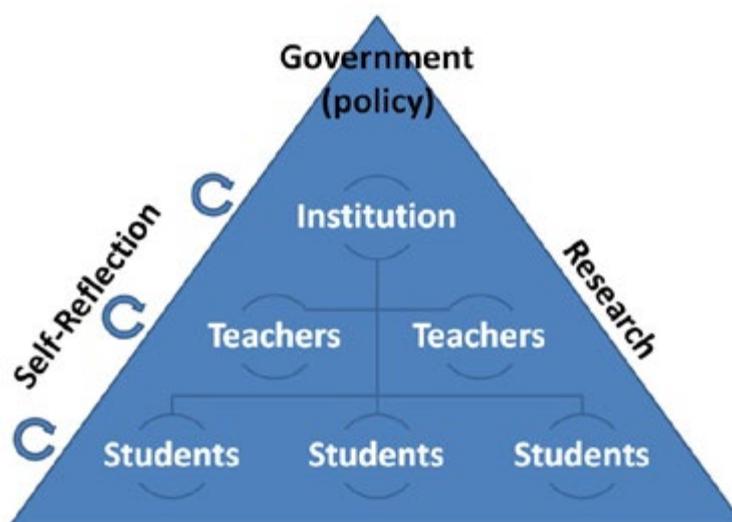


Figure 2. Information flow between LA stakeholders

Although they are the most widespread form in formal education, hierarchies are not the only flow models to describe where benefits can be retrieved. For example, peer evaluation using Personal Learning Environments (PLE) may be another information environment for LA. Peer environments also prevail in academic transactions like conferences or publications that are based on peer review systems. Practical examples for a horizontal peer-related information flow are the various scientific impact measures that exist, e.g., citation indexes. Equally, serious games can provide a non-hierarchical approach and/or team perspective to collaborative learning, e.g., how fast a team completed a level. In each of these, however, lie some issues of dependency and possible legal constraints (cf. further below).

Example opportunities for LA with respect to different stakeholder groups are:

Students can be supported with specific learning process and reflection visualisations that compare their performance to the overall performance of a course. Furthermore, they can be provided with personalised recommendations for suitable learning resources, learning paths, or peer students (Gaviria et al., 2011).

Teachers can be provided with course monitoring systems that inform them about knowledge gaps of particular pupils and thus enable them to focus their attention on those pupils. They can also harvest emergent group models that can lead to shared understanding of domain topics or processes for better curriculum design and on-the-fly adaptations.

Institutions can monitor the performance of students regarding drop-out and graduation rate on a much finer granular level. In this way, they can evaluate their courses and improve outcomes of their courses.

Other stakeholders: We would like to emphasise that stakeholders need not be confined to formal education settings, but include all formal, non-formal, or informal environments, such as professional development (CPD). In these cases, the stakeholders are to be substituted by the relevant entities. For non-formal learning, for example, stakeholders would include a “learner” instantiation with (only) a self-reflection dimension in which feedback is mirrored back to the same person. In work-based learning, employees and line-managers may be the most common stakeholder groups involved. More notably, computer agents can also serve as stakeholders, for example as data clients that take further decisions on the learner’s behalf or trigger an event (e.g., notification e-mail, recommendation of content or peer, etc.).

Objectives

The main opportunities for LA as a domain are to unveil and contextualise so far hidden information out of the educational data and prepare it for the different stakeholders (see above). Monitoring and comparing information flows and social interactions can offer new insights for learners as well as improve organisational effectiveness and efficiency. This new kind of information can support individual learning processes but also organisational knowledge management processes (Butler & Winne, 1995). We can distinguish two fundamentally different objectives: *reflection* and *prediction* (cf. Figure 1 above).

Reflection: Reflection is seen here as the critical self-evaluation of a data client as indicated by their own datasets in order to obtain self-knowledge. Wolf (2009) calls this process the “quantified self”, i.e., self-observation and reacting to one’s own performance log data. There already is a growing number of Personal Informatics Systems, i.e., human-computer interaction systems that support this process (Li & Forlizzi, 2010). However, reflection may also be seen as critical self-evaluation based on *other* stakeholders’ datasets. This would especially be true if, for example, a teacher was led to reflect upon their teaching style as indicated by the datasets of their students. In the above hierarchical flow model (Figure 2), the higher order stakeholder would have the ability to utilise all the datasets from lower constituencies for their own reflection.

On an individual level, LA can support reflection of learning processes and offer personalised information on the progress of the learner (Govaerts et al., 2010). On the institutional level, LA can enhance monitoring processes and suggest interventions or activities for particular students. Greatest care should however be taken not to confuse objectives and stakeholders in the design of a LA process and not to let, e.g., economic and efficiency considerations on the institutional level dictate pedagogic strategies, as this would possibly lead to industrialisation rather than personalisation.

LA is a support technology for decision making processes. Therein also lies one of the greatest potential dangers. Using statistical analytic findings is a quantitative not a qualitative support agent to such decision making. We are aware that aligning and regulating performance and behaviour of individual teachers or learners against a statistical norm without investigating the reasons for their divergence may strongly stifle innovation, individuality, creativity and experimentation that are so important in driving learning and teaching developments and institutional growth.

Prediction: Apart from support for reflective practice, LA can also be used for predicting and modelling learner activities (Siemens, 2011; Verbert et al., 2011). This can lead to earlier intervention (e.g., to prevent drop-out), or to adaptive services and curricula. Using Machine Learning techniques, for example, learner profiles can be built dynamically and automatically, saving the learner filling in and maintaining profile data. In predictive outcomes lies currently much hope for efficiency gains in terms of establishing acts of automatic decision making for learning paths, thus saving teacher time for other more personal interventions. But prediction suffers potentially from big ethical problems (to which more further below), in that judgements about a person, whether originating from another

human or a machine agent, if based on a limited set of parameters could potentially limit a learner's potential. For example, not every learner who has difficulties mastering subject level two, will automatically not master level three. We have to prevent re-confirming old-established prejudices of race, social class, gender, or other with statistical data, leading to restrictions being placed upon individual learners. Furthermore, there are limitations in the use of LA data as a means for supporting the learning process. Learning processes assume the leading role of the learner, rather than that of the teacher. However, the reliability of a LA-supported learner profile and its usefulness to the learners will remain questionable. For example, what LA data can be used in order to define whether a learning activity had a "high" or "low" impact on the learning process of learners, and at which points in the process itself? The diversity of learning makes it also problematic to judge which learning activity was of high value for learner A but of low value for learner B.

With respect to pedagogic theories, we would like to argue that LA does neither support nor ignore specific pedagogic theories, and as an abstract concept is pedagogically neutral. Indeed, we are of the opinion that LA can be used to evaluate different pedagogic strategies and their effects on learning and teaching through the analysis of learner data. This can be defined as a specific pedagogically oriented objective under the current dimension, but, as we will discuss further below, certain technologies are not pedagogically neutral and this will influence the analytics process in one way or another.

Educational data

LA takes advantage of available educational datasets from different Learning Management (LMS) and other systems. Institutions already possess a large amount of student data, and use these for different purposes, among which administering student progress and reporting to receive funding from the public authorities are the most commonly known. Linking such available datasets would facilitate the development of mash-up applications that can lead to more learner-oriented services and therefore improved personalisation.

LA strongly relies on data about learners and one of the major challenges LA researchers are facing is the availability of publicly available datasets to evaluate their LA methods. Most of the data produced in institutions is protected, and the protection of student data and created learning artefacts is a high priority for IT services departments. Nevertheless, similar to Open Access publishing and related movements, calls for more openness of educational datasets have already been brought forward (Drachsler et al., 2010). Anonymisation is one means of creating access to so-called Open Data. Recently, Verbert et al., (in press) presented a state of the art review of existing educational datasets. How open educational data should be, requires a wider debate (cf. section on legal constraints below), but, already in 2010, several data initiatives were started to make more educational data publicly available:

dataTEL challenge—The first dataTEL challenge was launched as part of the first workshop on Recommender Systems for TEL (Manouselis et al., 2010), jointly organized by the 4th ACM Conference on Recommender Systems and the 5th European Conference on Technology Enhanced Learning (EC-TEL 2010) in September 2010. In this call, research groups were invited to submit existing datasets from TEL applications that can be used for LA research purposes and recommender systems for TEL.

dataTEL workshop—The "Datasets for Technology Enhanced Learning" workshop was organised at the third STELLAR Alpine Rendez-Vous in March 2011. During this workshop, related initiatives that are collecting educational datasets, and apply these in data-driven learning applications were presented, and challenges related to privacy and data protection were discussed.

PSLC dataShop (Stamper, 2011) offers an open data repository that provides access to a large number of educational datasets. *dataShop* has data from students derived from interactions with intelligent tutoring systems.

LinkedEducation.org (Dietze et al., 2012) is another initiative that provides an open platform to promote the use of data for educational purposes. At the time of writing, five organizations have contributed datasets.

Despite these pioneering activities, it does, by comparison, still seem somewhat bizarre that in the commercial world with clicking the “register” button, the default access to all user data becomes owned by some company, whereas educational institutions operate on the default that everything is protected from virtually everyone.

Distinguishing educational data by access rights in open and protected datasets (Figure 1) is not as simple as it sounds. Because the technical systems producing and collecting data are typically owned by the institution, the easiest assumption would be that this data belongs to them. However, which employees of the institution exactly are included in the data contract between a learner (or their parents) and the educational establishment, is as yet unresolved. This poses severe constraints on inner-institutional research or wider institutional use. We will bring up some more legal consideration under the point on external constraints below.

Like in related research domains, LA datasets create a new set of challenges for research and practice. These include:

- A lack of common dataset formats like the suggested one from the CEN/ISSS PT social data group (cf. CAM Schema at: <https://sites.google.com/site/camschema/home>; and Wolpers et al., 2007).
- The need for version control and a common reference system to distinguish and point to different datasets.
- Methods to anonymise and pre-process data according to privacy and legal protection rights (Drachler et al., 2010).
- A standardised documentation of datasets so that others can make proper use of it like that promoted by the data-seal-of-approval initiative (cf. <http://www.datasealofapproval.org>).
- Data policies (licences) that regulate how users can use and share certain datasets. For instance, the Creative Commons licensing rights could be considered as a standard way to grant permissions to datasets. DataCite (Brase, 2009) is an organization that enables to register research datasets and to assign licensing rights to them, so that the datasets can be referenced similar to academic articles.

From a technical point of view, idealised datasets probably remain the biggest challenge for analytics. This is to say that the assumption that datasets consist of context-free, meaningful and *only* meaningful data, is highly optimistic. In most natural settings, users “pollute” databases by producing erroneous or incomplete datasets. For example, teachers who want to see their students’ view on LMS courses often set themselves up as “test students” or create “test courses”. These are not always obvious, but need to be removed from the data to be analysed. Therefore empirical findings coming from a specific dataset are almost certainly affected by the context of data collection and processing.

Similarly, data collection often leads to “enmeshed identities” being used for analytics and prediction. A dataset cannot typically distinguish between a single individual and a shared presence in the learning space (group work on a single device). Students who often work together with others on shared devices (laptops, smartphone, lab space, etc.) produce enmeshed fingerprints in their educational data. This may lead to behaviours being attributed to a logged-in identity that may actually have originated from an “invisible” partner. Standardised documentation of datasets can be seen as paramount to raise awareness of this danger.

Additionally, from a pedagogic perspective, it remains an on-going challenge to formulate indicators from the available datasets that bear relevance for the evaluation of the learning process. The selection of specific data and their weighting (under the methods applied in the “instruments” dimension) against the real behaviour of students is of greatest importance, as is the process of relating behaviour pattern data to cognitive developments.

Instruments

Different technologies can be applied in the development of educational services and applications that support the objectives of educational stakeholders. LA takes advantage of so-called information retrieval technologies like educational data mining (EDM; cf. Romero et al., 2008), machine learning, or classical statistical analysis techniques (cf. Figure 1), but other techniques may also be considered relevant, e.g., social network analysis (cf. Buckingham & Ferguson, 2011) or Natural Language Processing (NLP).

Through these technologies, LA can contribute tailored information support systems to the stakeholders and report on demand. For instance, LA could be applied to develop a drop-out alert system. High drop-out rates are a challenging problem in education, especially distance education. Further research on LA can contribute to decrease

the drop-out rate by developing e.g., a *Drop-out Analyser* that notifies the teacher of a course in time which students are in danger of falling behind or dropping out. This could be done by using LMS datasets and train a certain information retrieval technology (e.g., a Bayesian classifier) on the datasets to learn behavioural patterns of students that dropped out. Afterwards, the *Drop-out Analyser* could be applied on a follow-up online course and flag up students that show similar patterns. The teacher of the course could then intervene in an appropriate manner. Preliminary prototypes of such systems are already available, like the Blackboard Early Warning System.

Under the dimension “instruments” in our model (Figure 1), we also subsume conceptual instruments such as theoretical constructs, algorithms, or weightings, by which we mean different ways of approaching data. These ways in the broadest sense “translate” raw data into information. The quality of the output information and its usefulness to the stakeholders depend heavily on the methods chosen. Hildebrandt (2010), quite rightly, warns that “invisible biases, based on ... assumptions ... are inevitably embodied in the algorithms that generate the patterns”.

Competing methods, technologies and algorithms applied to the same set of data, will result in different outcomes, and thus may lead to different consequences in terms of decision making based on these outcomes. LA designers and developers need to be aware that any algorithm or method they apply is reductive by nature in that it simplifies reality to a manageable set of variables (cf. Verbert et al., 2011).

External constraints

Many different kinds of constraints can limit the beneficial application of LA processes, some being “softer” than others. It has been suggested to us to identify them as ethical, legal, and social constraints, but also to feature organisational, managerial, and process constraints. This we find a useful subdivision of external limitations, but other divisions look equally logical. In the abstraction of the diagram above (cf. Figure 1), we propose the preliminary distinction of *conventions*, under which we count ethics, personal privacy, and similar socially motivated limitations, and, *norms* that are restricted by laws or specific mandated policies or standards. For reasons of space, we want to elaborate especially on the ethical aspects as this has grown into a field of much recent attention and debate (Bollier, 2010) and even spawned a collaborative effort in the Learning Analytics research community (Siemens, 2012).

New ethical and privacy issues arise when applying LA in education. These are challenging and highly sensitive topics when talking about educational datasets, as described in Drachler et al. (2010). The feeling of endangered privacy may lead to resistance from data subjects toward new developments in LA. In order to use data in the context of LA in an acceptable and compliant way, policies and guidelines need to be developed that protect the data from abuse. Legal data protection may require that data is anonymised before it can be used. At the same time, as much openness of the datasets as possible is desirable (see paragraph on data above).

Personal data enjoys strong legal protection, differing by national laws and sometimes competing with other legal frameworks such as the Freedom of Information Act in the United Kingdom. We will not go into the legal details here, but did already above hint at the predicament that faces any data-related venture when using people’s digital footprints. It concerns the lack of legal clarity with respect to data ownership. In current circumstances, data gathered about a person (before it is anonymised) belongs to the owner of the data collection tool, typically also the data client and beneficiary. Up till now, through direct intervention like questionnaires and sign-up processes, this was not a big problem to the data subjects. However, with the dramatic increase of ambient sensors and new technologies, such as location tracking or biometric face-recognition cameras, etc., more and more parts of individual behaviour are logged without the data subject’s approval or even awareness. That being so, the ethic principle of “informed consent” (cf. AoIR, 2002) is very much under threat. The fundamental question legislators need to ask is: who does a person’s life data belong to? (cf. also Hildebrandt, 2006). We believe that this question may in the near future become more and more elevated in importance and prominence.

On institutional level, educational and student data was traditionally handled separately, and is legally something of a blind spot. Registration data was kept and maintained by registry staff, IT data by IT staff, and learning data by academic staff. To use LA to its full potential, integration of available institutional datasets needs to happen. Universities, for instance, already collect and report socio-economic data such as students’ post codes or ethnic and linguistic background. Institutions are even legally obliged by funding bodies to do so, but integrating this dataset

with educational performance data, would be widely considered unethical or even illegal. As has been already mentioned above, the extent of a student's data contract with an institution and its individual staff representatives in different roles (teacher, administrator, secretary, researcher, IT support staff, Deans and management, etc.) needs to be urgently clarified. At the same time, privacy directives such as the *Data Seal of Approval* supported by the Dutch DANS institute (<http://www.datasealofapproval.org>) and related European data directives like the European Directive on data protection 95/46/EC (Directive, EU, 1995) need to be implemented.

Even where in compliance with the law, educational data can easily be abused for purposes inappropriate for educational institutions or for the data subjects (especially where minors are concerned). By principle, the more access to information about a data subject a data client has, the higher the responsibility is to use this information in a sensitive and ethical way. In an inspired article, Hildebrandt (2010) elaborates the ethically limited applicability of automated pattern recognition to the Law domain, but these limitations can be transferred just as easily to the domain of learning. Among the more obvious ethical risks are the exploitation of such data for commercial and similar purposes, or data surveillance issues (social sorting, cumulative disadvantages, digital stalking) and their ethical implications.

Ethics don't stop at the data gathering and integration. The realisation that we may encounter conflicts in values and interests in and through the analysis of people's behaviour needs to guide the post-analytic decision making process and the conclusions drawn from the approach. It is important to remind stakeholders of LA processes that data can be interpreted in many ways and lead to very different consequent actions. To give a drastic example, imagine being confronted with the insight that children from an immigrant background show reading difficulties, backed by supportive data analysis. This may lead to a wide ranging variety of responses, from developing extracurricular support mechanisms, to segregated classes, up to bluntly racist abuse of various kinds.

Ethics in LA may affect students and teachers alike, especially, where institutions aim to use LA to quality assure the performance of their teaching staff. Data can easily be abused as supporting evidence for exercising inappropriate pressures on data subjects to change otherwise perfectly acceptable or explainable performance behaviour. Institutions are therefore challenged with establishing a set of ethical policies and principles, together with, e.g., complaints procedures and safety nets that secure proper use of educational data in teaching and research. We find the ethical guidelines of the Association of Internet Researchers, AoIR (2002), a useful starting point in this respect, in that it has a purpose-oriented approach that supports ethical pluralism and respects the individual.

Another ethical consideration is the acceptance of divergence in the data constituency (AoIR, 2002). We already touched upon the danger that the result of algorithmic analysis, consequent policies and exercised pressures may aim at uniformity and at mainstreaming learning and teaching processes, thereby greatly harming creative processes and innovation that diverge from the statistical mean. It is one of the principal shortfalls of statistical prediction that it can only predict average behaviour not outliers. As such, LA provides no means of predicting exceptions to a rule, or exceptions to the exception rule.

Internal limitations

In complement to the environmental problem areas contained in the above "external constraints" section, we subsume a number of human factors that enable or may pose obstacles and barriers under the dimension heading "internal limitations." Prominent among these are *competences* and *acceptance*.

It is already becoming clear that the application of learning analytics requires new higher-order *competences* to enable fruitful exploitation in learning and teaching. In order to make LA an effective tool for educational practice, it is important to recognise that LA ends with the presentation of algorithmically attained results that require interpretation (Reffay & Chanier, 2003; Mazza & Milani, 2005). There are innumerable ways to present and to interpret data and base consecutive decisions and actions on it, but only some of them will lead to benefits and to improved learning. Basic numeric and other literacies, as well as ethical understanding are not enough to realise the benefits that LA has to offer. In a recent survey we conducted among LA experts, only 21% of the 111 respondents felt that learners would possess the required competences to interpret LA results themselves and determine appropriate actions/interventions from it (Drachler & Greller, 2012). Therefore, the optimal exploitation of LA data requires some high level competences in this direction, but interpretative and critical evaluation skills (cf. Figure 1)

are to-date not a standard competence for the stakeholders, whence it may remain unclear to them what to do as a consequence of a LA outcome or visualisation.

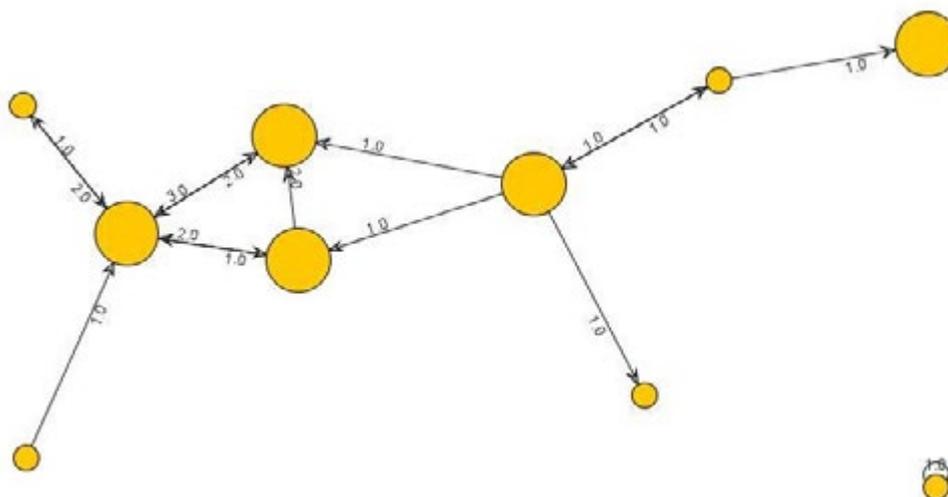


Figure 3. Social network diagrams may look attractive, but may not always be the best way to present information

Interpretation of LA results is often facilitated by enticing visualisations that are aimed to serve as a functional aid (Figure 3). One inherent danger that we perceive is that the simplicity and attractive display of data information may delude the data clients, e.g., teachers, away from the full pedagogic reality. This may negatively affect the pedagogic assessment and grading of a student's performance which should not be based alone on the visualisation of log files from a Learning Management System. To illustrate this danger take the example of marking student essays. An automated spell-check on orthographic mistakes presents itself as a quick and simple to interpret translation of the learner artefact into numbers. This makes it ideal for an efficient, cognitively effortless, and egalitarian grading mechanism. In an education environment that increasingly suffers from time constraints and calls for more efficiency in teacher activities, it is easily imaginable that the traditional qualitative assessment of essays gives way to such quick number crunching being over-proportionally reflected in student marks.

In our model (cf. Figure 1), we include among the key competences for LA, critical evaluation skills, because superficial digestion of data presentations can lead to wrong conclusions. It has to be strongly emphasised that data not included in the respective LA approach, is equally if not more important than the dataset that is included. To judge a learner's performance merely on, e.g., LMS quantitative data is like looking at a single puzzle piece. As learning is more and more happening in a lifelong and diverse ecosystem, an exclusive data view on single elements may provide a stimulus for reflection but not a sound basis for assessment.

The necessary competences notwithstanding, acceptance factors can further influence the application or decision making that follows an analytics process. This can, as is regularly seen in political debates, lead to blunt rejection of the results or applied methods from the constituency or parts thereof. In a learning context, ways to increase acceptance is vitally important also in order to produce usable outcomes. To get a better grasp on this issue, current scientific debate, therefore, should focus on empirical evaluation methods of learning analytics tools (Ali et al., 2012) and on advanced technology acceptance models (cf. Venkatesh & Bala 2008), inspired by the early work in this area (Davis, 1989, 1993). For LA, a revised technology acceptance model (TAM) could be an interesting approach to evaluate the emergent analytic technologies for all stakeholders described in our framework and also the needed implementation requirements to guarantee successful exploitation.

The place of pedagogy in the learning analytics framework

LA holds promises in the context of TEL by offering new methods and tools to diagnose learner needs and provide personalised instructions to better address these needs. It is not yet clear to what extent LA will lead to more personalised learning experiences rather than merely clustering people into behaviouristic "learner models" (e.g., as

“outliers” of mainstream models). Consequently, more empirical evidence is needed to identify which pedagogic theory LA serves best. LA has been effectively used for behaviourist-instructivist style approaches (but see the critical reflection by Pardo & Kloos, 2011), but there is as yet little evidence for the support of constructivist approaches to learning (Duffy & Cunningham, 2001), where learning is seen as an active cognitive process in which learners construct their own concepts of the world around them. In LA, the latter is mostly inferred indirectly, by relating grades of learning outcomes with activities during the learning process (Dawson, 2012). In these correlations, it emerges that active students get better results. However, the role LA plays in this has not yet been conclusively demonstrated. Despite these questions, we would like to maintain that as knowledge and experiences vary considerably among learners, the diversity of learning can more effectively be addressed by LA methods than with current learning environments.

In our model, LA can work in support of a multitude of pedagogic strategies and learning activities as manifested and represented by the available data. This means we can only see pedagogies through the data. Because of this, we do not include them as part of the analytics process (Figure 1) but as *implicitly* contained in the input datasets that encapsulate the pedagogic behaviour of users. As we know, this behaviour depends a great deal on the platform and the pedagogic vision the developers built in (Dron & Anderson, 2011). For example, data from a content sharing platform will have a behaviourist/cognitivist pedagogy attached to the learner behaviour, since this is the pedagogic model underlying the technology. In any case, only the pedagogic patterns exhibited in the dataset can be analysed; and this will vary.

Additionally, pedagogy can be *explicitly* addressed in the goals and objectives that the LA designer sets (“objectives” dimension). The LA method (“instruments” dimension) will determine the outcome of the analysis and together with the interpretation applied may lead to a large variety of options for consequences and interventions. If such pedagogic interventions were applied, they would lead to new behaviours which, once again, could be analysed through the available data (Figure 4). In the graph below, we refer to pedagogic behaviour as learner/teacher behaviour that is motivated by didactic designs (learning designs). Pedagogic consequences, similarly, are adjustments to the didactic strategy or learning design based on the outcomes of the LA process.

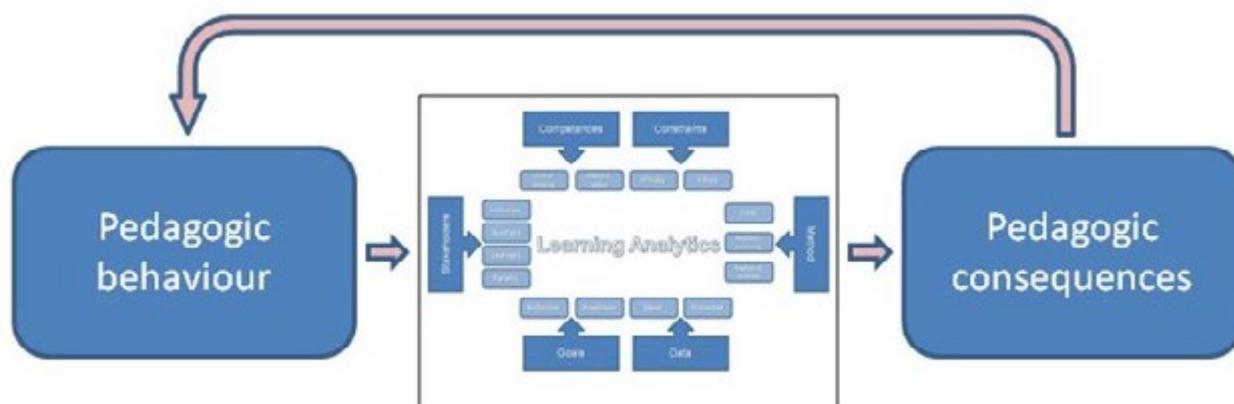


Figure 4. Learning analytics and pedagogy

A simple analogy would be boiling water in a pan. At any time (or continuously) you can immerse a thermometer and measure its temperature. The goal would be to determine whether you need to turn up the heat or not. The result of the analysis can then lead to the actions you want to take. The thermometer is only one method for such an analysis. An alternative would be to observe and wait until the water bubbles. Setting a benchmark (in the objectives design) can inform you when it is time for the teabag to go in. When applied to a learning process, immersing the thermometer into the water equates to the LA data gathering and analysis of learning in progress. It is here where learning is translated into numbers.

It should be noted that the pedagogic input factors are not confined to behaviour alone, but also include beliefs, (societal) values, and implicit theories of knowledge and learning. However, the LA application can only see these in the way they manifest themselves in the data. We also want to point at the possibility to apply computer agents to

determine specific interventions. These could be as simple as sending a notification or recommendation to one of the stakeholders.

An example for pedagogic consequences is the following (Dawson, 2012): Using SNAPP as a tool to do a social network analysis (SNA) on discussion forums in a learning environment, the moderator or tutor might discover that certain changes in the moderation, the organisation, or the task, may lead to more or less engaged discussion among participants. In this way, the information gained through the LA process can support the fine-tuning of pedagogic effectiveness in a particular activity, depending on the desired learning outcome.

The model takes note that pedagogic success and performance is not the only thing that LA can measure. LA collects snapshots taken from educational datasets. These snapshots can be used to reflect or predict, in order to make adjustments and inform interventions, either by a human or by a system. Apart from offering efficiency benchmarking and business information for education providers, new support services for learning and more qualitative personal experiences can be achieved.

Conclusion and outlook

In summary, the proposed framework model in figure 1 above stresses the inherent connections between the six different dimensions and the impact of the analytics process on the end user and the data suppliers. If one of the parameters changes, the outcome and anticipated benefits will change. It is therefore our conviction that only the consideration of all six dimensions in the design process can lead to optimal exploitation of LA. Additionally, substantial work on new ethical guidance, data curation, and ownership needs to happen at universities and in legislation to reduce the risks connected to the application of LA and to protect the data subject, usually the learner.

Because of the inherent dependencies, we argue that all six dimensions are mandatory to be argumentatively present in a fully flexed LA design. We would, therefore, strongly welcome if application developers and researchers would not only make their technical environment known and open, but also describe the contextual environment and expectations from the users (e.g., required competences) along the lines of the framework. This would allow scientific comparison, replication, external analysis, and alternative contextualisation.

To validate the framework as both a descriptive approach as well as a guide to the design process of LA applications, we suggest evaluating the growing number of LA application showcases and testing for consistency in the descriptive values of the model. Additionally, we want to create a selected number of use cases that encompass the six dimensions and their instantiations.

LA is very much at the dawn of its existence and considered by many as one of the technological advances that will bring learning onto the next higher level. While we join in with this chorus of positive expectations, we are also aware that LA shows facets of a double nature: In its most optimistic outlook, learners will be provided with personal information about their current needs, while, at the same time, the educational system will be evolved from a “one-size-fits-all” approach into a highly personal competence-driven educational experience. But this view is not without flaws, because of the real dangers that the extended and organised collection of learner data may not so much bring added benefits to the individual, but instead provides a tool for HEIs, companies, or governments to increase manipulative control over students, employees, and citizens, thereby abusing LA as a means to reinforce segregation, peer pressure, and conformism rather than to help construct a needs-driven learning society. We therefore believe that it will be of critical importance for its acceptance that the development of LA takes a bottom-up approach focused on the interests of the learners as the main driving force.

LA has the potential for new insights into learning processes by making hitherto invisible patterns in the educational data visible to researchers and end users, and to enable development of new instruments for everyday educational practice. However, there are substantial uncertainties about the extent of impact LA will have on education and learning in general. The proposed framework model is motivated by the potential and opportunities that LA offers in its relevance for educational development and opportunities to personalise learning. While we agree with the Horizon report’s forecast and its claim for a prosperous future of LA (Johnson et al., 2011), we also strongly feel that this development should not happen without a guiding framework that combines use of educational data with the protection of individuals and their learning.

Decisions based on LA are of concern, because they determine the usefulness and consequences for the stakeholders as well as the extent of its impact. Data analysis could have dramatic (and unwanted) consequences if not used with the necessary care. It is here where ethics play an enormously important role. Building of trust and confidence throughout the data constituencies has to be a priority from the start, and, here again, this proposed framework hopes to act as a useful guide.

One of the major questions in LA is the relation with theories of learning, teaching, cognition and knowledge. We hinted above at the opportunity that LA may support the evaluation of concrete didactic approaches which in turn may provide supportive evidence for particular pedagogic theories of learning and knowledge. At the same time, technologies are not pedagogically neutral; hence the evaluation will be influenced by the approach chosen. We consider this debate as an on-going one which will require further research and demonstration of applications and the impact they make on the process of learning.

It is still too early to base education fully on LA approaches alone, and we expect it never will be possible to do so. However, at the very least, opportunities this new discipline has to offer are to provide new support for learning activities and stimuli for reflection. In our opinion, it is these opportunities that LA should pursue.

References

- Ackermann, F., Eden, C., & Cropper, S. (2004). Getting started with cognitive mapping. Retrieved from <http://www.banxia.com/pdf/de/GettingStartedWithCogMapping.pdf>
- Ali, L., Hatala, M., Gasevic, D., & Jovanovic, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1), 470-489.
- AoIR (Association of Internet Researchers) Ethics Working Committee, & Ess, C. (2002). Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee. Retrieved from AoIR website: www.aoir.org/reports/ethics.pdf
- Bollier, D. (2010). The promise and peril of big data. Washington, DC: The Aspen Institute.
- Brase, J. (2009). DataCite – A global registration agency for research data. *Proceedings of the 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology* (pp. 257–261). Washington, DC: IEEE Computer Society. doi: 10.1109/COINFO.2009.66
- Buckingham Shum, S. and Ferguson, R. (2011). *Social learning analytics*. (Report No. KMI-11-01). Retrieve from Knowledge Media Institute, The Open University, website: <http://kmi.open.ac.uk/publications/pdf/kmi-11-01.pdf>
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245-281
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340
- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perception and behavioral impacts. *International Journal of Man – Machine Studies*, 38, 475-487.
- Dawson, S. (2012). Interpreting social networks: Informing teaching practice. Learning and Knowledge Analytics massive open online course (LAK12). Retrieved from <https://sas.illuminate.com/p.jnlp?psid=2012-02-28.1237.M.0728C08DFE8BF0EB7323E19A1BC114.vcr&sid=2008104>
- Dawson S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society*, 11(3), 224-238
- Directive, EU (1995). 95/46/EC Protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal, L 281*, 0031 – 0050.
- Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., & Taibi, D. (2012, March). *Linked education: Interlinking educational resources and the web of data*. Paper presented at the 27th ACM Symposium on Applied Computing, Special Track on Semantic Web and Applications, Trento, Italy.
- Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, ... Wolpers, M. (2010). Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning. *Elsevier Procedia Computer Science*, 1(2), 2849–2858.

- Drachslar, H. & Greller, W. (2012, April). *The pulse of learning analytics—Understandings and expectations from the stakeholders*. Paper presented at the Second International Conference in Learning Analytics (LAK12), Vancouver, Canada.
- Dron, J. (2011). *Analytics: Soft and hard. Learning and knowledge analytics 2011 presentation* [Powerpoint slides]. Retrieved August 1, 2011, from <http://www.slideshare.net/jondron/learning-analytics-soft-and-hard>
- Dron, J. & Anderson, T. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning*, 80-97, 80–97
- Duffy, T. M. & Cunningham, D. J. (2001). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology*. New York, NY: Simon and Schuster.
- Elias, T. (2011). Learning analytics: Definitions, processes and potentials. Retrieved from <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
- Gaviria, F., Glahn, C., Drachslar, H., Specht, M., & Gesa, R. F. (2011). Activity-based learner-models for learner monitoring and recommendations in Moodle. In C. D. Kloos et al. (Eds.), *Proceedings of the 6th European Conference on Technology-Enhanced Learning* (pp. 111-124). Heidelberg, Berlin: Springer-Verlag.
- Govaerts, S., Verbert, K., Klerkx, J., Duval, E., (2010, December). *Visualizing activities for self-reflection and awareness*. Paper presented in the 9th International Conference on Web-based Learning, Shanghai University, China.
- Hildebrandt, M. (2006). 'Privacy and Identity,' privacy and the criminal Law. In E. Claes, A. Duff and S. Gutwirth (Eds.), *Antwerpen* (pp. 43–58). Oxford, UK: Intersentia. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1005&context=mireille_hildebrandt
- Hildebrandt, M. (2010, March). *The meaning and the mining of legal texts*. Paper presented at the Computational Turn in the Humanities, Swansea, Wales, UK. Retrieved July 11, 2011, from http://works.bepress.com/mireille_hildebrandt/27/
- Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 horizon report*. Austin, Texas: The New Media Consortium.
- Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. *ACM Conference on Human Factors in Computing Systems*, pp. 557–566. New York, NY: ACM. doi: 10.1145/1753326.1753409
- Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588-599.
- Manouselis, N., Drachslar, H., Vuorikari, R., Hummel, H., & Koper, R. (2010). Recommender systems in technology enhanced learning. In P. B. Kantor, F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 387-415). Berlin, Germany: Springer.
- Mazza, R., & Milani, C. (2005, July). *Exploring usage analysis in learning systems: Gaining insights from visualisations*. Paper presented at the 12th International Conference on Artificial Intelligence in Education, Amsterdam, the Netherlands.
- Pardo, A., & Kloos, C. D. (2011). Stepping out of the box: Towards analytics outside the learning management system. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 163–167). New York, NY: ACM.
- Reffay, C., & Chanier, T. (2003, June). *How social network analysis can help to measure cohesion in collaborative distance learning*. Paper presented at the International Conference on Computer Supported Collaborative Learning, Bergen, Norway.
- Retalis, S., Pappasalouros, A., Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). *Towards networked learning analytics—A concept and a tool*. *Proceedings of the Networked Learning Conference 2006*. Retrieved from <http://www.networkedlearningconference.org.uk/past/nlc2006/abstracts/Retalis.htm>
- Ritchey, T. (2011). General Morphological Analysis—A general method for non-quantified modelling. In T. Ritchey (Ed.), *Wicked Problems—Social Messes*. Retrieved from <http://www.swemorph.com/pdf/gma.pdf>
- Romero, C., Ventura, S. Espejo, P. G., & Hervs, C. (2008). Data mining algorithms to classify students. In R. de Baker, T. Barnes, J. Beck (Eds), *Proceedings of the 1st International Conference on Educational Data Mining* (pp. 8–17). Retrieved from http://www.educationaldatamining.org/EDM2008/uploads/proc/1_Romero_3.pdf
- Savage, M. & Burrows, R. (2007). The coming crisis of empirical sociology. *The Journal of The British Sociological Association*, 41(5), 885–899. doi: 10.1177/0038038507080443
- Siemens, G. (2011). Learning analytics: A foundation for informed change in higher education. Retrieved 1 August, 2011, from <http://www.slideshare.net/gsiemens/learning-analytics-educause>
- Siemens, G. et al. (2012). Learning analytics: Guidelines for ethical use. Shared effort of the learning analytics research community. Retrieved March 23, 2012, from <http://bit.ly/wMDmLW>

Stamper, J. (2011). *EDM and the 4th paradigm of scientific discovery* [Powerpoint Slides]. Retrieved 1 August, 2011, from https://pslcdatashop.web.cmu.edu/about/edm_stamper_2011.html

Venkatesh, V., & Bala, H. (2008), Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273–315.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011, February). *Dataset-driven research for improving recommender systems for learning*. Paper presented in 1st International Conference Learning Analytics & Knowledge, Banff, Alberta, Canada.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (in press). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society*.

Wolf, G. (2009, July 17). Know thyself: Tracking every facet of life, from sleep to mood to pain, 24/7/365. *Wired Magazine*. Retrieved from http://www.wired.com/medtech/health/magazine/17-07/lbnp_knowthyself?currentPage=all

Wolpers, M., Najjar, J., Verbert, K., & Duval, E. (2007). Tracking actual usage: The attention metadata approach. *Educational Technology & Society*, 10(3), 106-121.

ARTIKEL 2

PRIVACY AND LEARNING ANALYTICS - IT'S A DELICATE ISSUE'. A CHECKLIST FOR TRUSTED LEARNING ANALYTICS

Hendrik Drachsler en Wolfgang Greller

Copyrights

Published in the proceedings of the Learning
Analytics & Knowledge Conference 2016 in Edinburgh.
<https://dl.acm.org/citation.cfm?id=2883893>

Privacy and Learning Analytics – it’s a DELICATE issue

Hendrik Drachsler
Open University of the Netherlands
Welten Institute
hendrik.drachsler@ou.nl

Wolfgang Greller
Vienna University of Education
wolfgang.greller@phwien.ac.at

Abstract

The widespread adoption of Learning Analytics (LA) and Educational Data Mining (EDM) has somewhat stagnated recently, and in some prominent cases even been reversed following concerns by governments, stakeholders and civil rights groups. In this ongoing discussion, fears and realities are often indistinguishably mixed up, leading to an atmosphere of uncertainty among potential beneficiaries of Learning Analytics, as well as hesitations among institutional managers who aim to innovate their institution’s learning support by implementing data and analytics with a view on improving student success. In this paper, we try to get to the heart of the matter, by analysing the most common views and the propositions made by the LA community to solve them. We conclude the paper with an eight-point checklist named DELICATE that can be applied by researchers, policy makers and institutional managers to facilitate a trusted implementation of Learning Analytics.

General Terms

Technology enhanced learning, Analytics, Education, Design, Human factors, Legal aspects, Privacy, Ethics

Keywords

Learning analytics, data management, implementation, privacy, ethics

1. Introduction

In 2011, Learning Analytics has been hailed by the Horizon Report [1] as a revolutionary game-changer for teaching and learning. Rapid moves and developments by enthusiasts have, however, slowed down recently due to rising concerns about the impact analytics has on individuals, their identity and integrity, as has already been identified early on by [2]. These second thoughts about Learning Analytics, in our view, originate from and run in parallel to the fears expressed in the wider context of Internet safety, surveillance, and commercial exploitation of data and labour on the Internet.

Since then, vivid academic discussions are taking place on how to provide acceptable approaches for the institutional adoption of Learning Analytics. A number of initiatives have been created to address issues of ethics and privacy in relation to Learning Analytics. At LAK15, the first workshop on Ethics and Privacy in Learning Analytics (EP4LA) has been organised jointly by the EU FP7 project Learning Analytics Community Exchange¹ and the

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the Owner/Author. Copyright is held by the authors.

LAK '16, April 20-24, 2016, Edinburgh, UK,

ACM 978-1-4503-3417-4/15/03.

¹<http://www.lacp1948.org/about-todays-ethics-and-privacy-in-learning-analytics-ep4la/>
<http://dx.doi.org/10.1145/2723576.2723642>

SURF SIG Learning Analytics², who also organised similar events at other conferences in the Netherlands (Utrecht), US, (Washington), and France (Paris).

How pertinent the issue is can be seen in prominent recent examples like the shutting down of inBloom in the US, due to privacy concerns about Learning Analytics and Big Data in education [3] below. In spite of receiving more than \$100m of grant funding by the Gates and Carnegie foundations, and despite the potential benefits, the inBloom system was closed down for good in April 2014, after parents and pressure groups expressed sincere concerns about the misuse of data, the repurposing of data for commercial interests, as well as general safety from cyber-attacks. Another, very similar case concerned the Snappet foundation in the Netherlands which provided tablets to some 400 public primary schools with pre-installed apps for maths and language learning. Snappet repurposed the collected usage data and used it to classify and predict individual student success. It also provided schools with information on educational interventions. However, an investigation by the national organisation for the protection of personal data (CBP, College Bescherming Persoonsgegevens) identified the collected datasets as ‘personal data’ that needed to be treated according to the Dutch laws on privacy about individuals. The fact that the data collected affected young children and provided insights into their performance at school has been used by the CBP to classify this data as ‘highly sensitive’ and demanding the highest privacy standards [4] below.

These two cases demonstrate how sensitive the issue of privacy and ethical use of educational data is, particularly when dealing with underage children, but also with older students in further and higher education, and even with adults. They show that ignoring the fears and public perception of the application of analytics, no matter how benevolent the intent, can lead to a lack of acceptance, protests, and even failure of entire Learning Analytics implementations.

Acceptance of technological solutions using data in education depends to a great extent on the data subjects being sufficiently aware of the consequences of using the system, the validity and relevance of the results obtained, and the level of transparency of the data model – e.g., how data are collected, stored, processed and shared (cf. the technology acceptance model [5]). Even before the Snappet case, it has been clear that the inclusion of minors in technological solutions can have severe ethical implications and may need additional action, such as parental consent, to receive backing. A big challenge for Learning Analytics in this respect is the complexity of the data collection and algorithmic analysis processes. The applied technologies are not trivial and it can be rather difficult to provide non-technical educational stakeholders (students, teachers, managers, and external parties like education authorities or parents) with an understanding of how and what

² <https://www.surfspace.nl/sig/18-learning-analytics/>

data are being collected, how they are processed, and how reliable the results of the analysis are.

The above cases, and others, justify a thorough and open discussion about the issues surrounding the safe and transparent use of Learning Analytics in an educational context. To contribute to this wider academic exchange is the aim and objective of this paper. We approach this complex issue primarily from an institutional and policy perspective with a view to provide some guidance for educational managers, decision makers, and data curators for K12, Higher Education, and workplace learning when implementing privacy-conform and ethically agreed solutions for Learning Analytics. To this end, we developed an eight point checklist named DELICATE that can serve as a reflection aid.

In the remaining parts of this paper, we first summarise the current state of the art on ethics and privacy for Learning Analytics. Thereafter, we give a description of ethics, privacy and the state of legal obligations to clean up the substantial overlap in understanding between these concepts, which often leads to confusion when discussing the implications of Learning Analytics. After clarifying the relevant working concepts, we analyse some of the most prominent fears and doubts about Learning Analytics and try to demystify them. Finally, we conclude the paper with the DELICATE checklist that aims to provide a practical means to support educational organisations in becoming trusted Learning Analytics users.

2. Related work

In an empirical expert study, Scheffel et al. [6] identified ‘data privacy’ as the most important aspect for increasing quality and trust in Learning Analytics. This followed an earlier survey among the emerging Learning Analytics community [7], which showed similar opinions, with about two thirds of the surveyed experts believing that Learning Analytics will affect ‘privacy and personal affairs’. Despite the apparent prominence of the issue in public and academic thinking, only few papers have been published in this area to date (e.g., [11][8][10][11][12]), and, even fewer policies or guidelines regarding privacy, legal protection or ethical implications [14][15] were developed and publicised.

While the law relating to personally identifiable information is widely understood, there has been insufficient attention on privacy from a user-centred perspective. What is more, there are no clearly defined best practices for the anonymisation, reuse, storage and security of educational data. Still, the need for an ethical and privacy-wise acceptable data analysis has been widely acknowledged, including by some national authorities. In November 2015, the SURF foundation of the Dutch Universities released a very comprehensive review of the legal state of educational data in the Netherlands that also considers European law and shows ways on how Learning Analytics can take advantage of educational data without affecting the legal rights of individuals. Kennisnet, the Dutch educational IT infrastructure provider, has also set up an initiative and consultation on “privacy by design” in a user-centric approach³. In previous attempts to organise academic and educational consent in this area, the UK CETIS institute published a consultation paper [16] and the JISC, more recently, developed a code of practice for learning analytics [14].

The user-centric aspect also underlies the most recent paper by [11], which explores the topic from a student vulnerability side

using a substantial set of recent literature. Vulnerability is an interesting perspective to take in the management of privacy and ethics. Unfortunately, in their paper it is not explained what these vulnerabilities actually are in the context of Learning Analytics. What are the concrete risks beyond, for example, that students may get targeted advertisements? How do these vulnerabilities balance against others – non-analytic ones (that is, for example, the vulnerability of dropping out of school education)? Furthermore, it is not clear from their elaboration whether the authors see vulnerabilities as part of the “learning contract” or fiduciary duty of the institution. Nevertheless, they quite rightly state that “privacy management goes beyond the traditional binary of opting in or opting out” and go on to propose a framework for learner agency [ibid.] that explores facets of control such as privacy self-management, timing and focus, and contextual integrity. With this, they clearly demand a more pro-active engagement with students and stakeholders to inform and more directly involve them in the ways individual and aggregated data are being managed and used.

In another state-of-the-art approach, Steiner et al. [9] provide a thorough and clearly articulated piece of work in the context of a project called LEA. After analysing previous propositions in depth, they synthesise them into a privacy and data protection framework with eight foundational requirements: (1) data privacy; (2) purpose and data ownership; (3) consent; (4) transparency and trust; (5) access and control; (6) accountability and assessment; (7) data quality; and, (8) data management and security. They conclude that technology and tools developed and used in learning analytics contexts need to be in line with these foundations and see them as fundamental requirements for a proper code of conduct.

3. Ethics, privacy and the current state of legal frameworks

In order to be able to discuss the challenges of ethics and privacy for Learning Analytics, we first need a better understanding of both concepts as well as their relationship towards each other. Below, we will express our summarised views on what constitutes ethics and what is typically meant when talking about privacy. In short, we can say that ethics is a moral code of norms and conventions that exists in society externally to a person, whereas privacy is an intrinsic part of a person’s identity and integrity. The understanding of what constitutes ethical behaviour varies and fluctuates strongly over time and cultures. Privacy, on the other hand, is first and foremost context bound [16], in that it forms the boundary of one’s person or identity against other entities. Thus, the understanding of privacy can diverge greatly between, e.g., persons living in large one-room family households and people living in large space single-occupancies, even when they belong to the same culture at the same time. Perceived violation of privacy can occur, when the ethical code of the surrounding society conflicts with the personal boundaries. To exemplify this, we can look at the social concept of “divorce”, which had different ethical and social implications at different times and places. It may lead to a personal conflict and infringement of privacy when the expectations of the ethical code contradict the personal priorities derived from the situation in the marriage.

Research ethics have become a pressing and hot topic in recent years, first and foremost arising from discussions around codes of conduct in the biomedical sciences such as the human genome (cf. [20][21]), but also, more recently, in the shape of “responsible

³https://www.kennisnet.nl/fileadmin/kennisnet/publicatie/Personaliseren_in_het_Leren_een_Internationale_Schets.pdf

research and innovation” (RRI) which is being promoted by the European Commission⁴.

In common language and popular thinking, there exists a substantial overlap between ethics and privacy, which sometimes leads to confusion when discussing the effects of Learning Analytics on either of them. Both concepts have in common that they can be approached from various perspectives, especially sociologically, psychologically, and even philosophically and religiously. They manifest themselves differently in a legalistic context and change over time.

3.1 Ethics

Ethics is the philosophy of moral that involves systematising, defending, and recommending concepts of right and wrong conduct. In that sense, ethics is rather different to privacy. In fact, privacy is a living concept made out of personal boundary negotiations with the surrounding ethical environment.

The first basic written principles for ethical research originated from the Nuremberg trials in 1949, and were used to convict leading Nazi medics for their atrocities during the Second World War [16]. This so-called Nuremberg Code is the first manifest for ethical research. It contains ten internationally recognised principles for the experimentation on humans:

1. Data subjects must be voluntary, well-informed, and consent to their research participation.
2. The experiment should aim at positive results for society.
3. It should be based on previous knowledge that justifies the experiment.
4. The experiment should avoid unnecessary physical and mental suffering.
5. It should not be conducted when there is any reason to believe that it implies a risk of death or disabling injury.
6. The risks of the experiment should be in proportion to (that is, not exceed) the expected humanitarian benefits.
7. Preparations and facilities must be provided that adequately protect the subjects against the experiment’s risks.
8. The staff who conduct or take part in the experiment must be fully trained and scientifically qualified.
9. The human subjects must be free to immediately quit the experiment at any point when they feel physically or mentally unable to go on.
10. Likewise, the medical staff must stop the experiment at any point when they observe that continuation would be dangerous.

The Nuremberg Code stimulated a major initiative in 1964 to promote responsible research on human subjects for biomedical purposes in the Helsinki Declaration [23]. It represents a set of ethical principles developed by the World Medical Association, but it is widely regarded as the cornerstone in ethical human research. The Helsinki Declaration has later been developed into the Belmont Report [24] in 1978 by the US National Commission for the Protection of Human Subjects of Biomedical and Behavioral Science. Both documents build on the basic principles of the Nuremberg Code. Together, they provide the foundation for the modern, ethically agreed conduct of researchers mainly in the medical fields. These are nowadays also taken to apply to

technological research and data collection about human subjects. The basic principles can be summarised as:

- Voluntary participation in research;
- Informed consent of the participants, and, with respect to minors, the informed consent of their parents or guardians;
- Experimental results are for the larger good of society;
- Not putting participants in situations where they might be at risk of harm (either physical or psychological) as a result of participation in the research;
- Protected privacy and confidentiality of the information;
- Option to opt-out;

To make it clear what ethical and what unethical research may constitute it is helpful to recall some prominent examples of unethical research from the past. Two infamous cases have been the *Milgram experiment* at Yale University [25], and the *Stanford Prison experiment* [26]. Those experiments are in their nature very different to the evaluation of any data tools in the way they caused harm to their participants. The negative impact of those experiments has been very direct and even physical to the participants of the experiments.

This is rather different to the effects that Learning Analytics research has on its research subjects. For the sake of argument, however, one could envisage dividing a class of students into control and experimental groups for A/B testing, and then, providing a positive stimulus to the participants of group A and a negative stimulus to the others. Such an experimental setup could be considered as unethical, as it would disadvantage group B. Harm in such a case would not be physical, but by deprivation of beneficial learning opportunities.

To prevent such negative impact from research and data usage, ethical committees – or IRB as they are called in the US – are charged with ensuring the protection of fundamental rights of all subjects participating in experiments (i.e., humans, but also animals). In most Western universities, nowadays, any human-subject research has to pass the ethical committee, and research ethics has also become part of the training of young scientists. From this, we dare say, that the risk of having unethical research being conducted at a Western university can be considered rather limited.

This being said, with the rise of Big Data and cloud computing new ethical challenges emerged, spurring the call for ethical committees at Big Data companies like facebook. A recent trigger was the facebook contagion study from 2014 [27], where a team of researchers manipulated the newsfeed of over 650.000 facebook users without notification or informed consent. The reaction to this manipulation has been massive among the user community and beyond. However, ethics is a volatile human made concept and what we see after the facebook study, is, that researchers now discuss the pros and cons of the study. Some people argue that the study has indeed been unethical, but, at the same time, contributed new insights into human behaviour [28]. In this context, it remains questionable whether self-regulation of profit-oriented companies via ethical committees would work in practice.

By contrast, as has been indicated in the previous paragraphs, after decade-long debates and risk assessments, the public education and research systems are much more advanced in applied ethics and reflective practice than private enterprises. Still, as the above cases of inBloom and Snappet have shown, the educational domain is extremely sensitive, and, therefore,

⁴ <http://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation>

universities and also the increasing amount of private companies collaborating within education need to be very careful with using student data for Learning Analytics.

It is important to mention, that ethical approval is typically only required for 'experiments'. If a system is rolled out directly and becomes part of the operational infrastructure of an institution, ethical approval is not always sought. The rollout of the new system then has to comply with the privacy and data protection laws according to national legislation.

3.2 Privacy

The right to privacy is a basic human right and an established element of the legal systems in developed countries. Already in 1890, Warren & Brandeis [29] wrote an article about "*The Right to Privacy*", where they explained that privacy is the "right to be let alone", and focused on protecting individuals. This right is often debated in the context of the yellow press with regards to royals and celebrities.

The concept of privacy as a right to be let alone was further developed by Westin in 1968 [30] who made it clear that new technologies change the balance of power between privacy and societal technologies. From this, Westin went on to specify privacy as the "*right of informational self-determination*" and as a vital part for restricting government surveillance in order to protect democratic processes.

According to Westin [ibid.], each individual is continually engaged in a personal adjustment process in which they balance the desire for privacy with the desire for disclosure and interaction with environmental conditions and social norms. Flaherty [31] took the informational self-determination further and claimed that networked computer systems pose a threat to privacy. He first specified 'data protection' as an aspect of privacy, which involves "*the collection, use, and dissemination of personal information*". This concept forms the foundation for fair information practices used by governments globally. Flaherty promoted the idea of privacy as information control. Roessler [32] later operationalised the right to privacy across three dimensions: 1. Informational privacy, 2. Decisional privacy, 3. Local privacy. It is important to note that privacy is not the same as anonymity or data security. They are related concepts that have an effect on privacy, but do not represent privacy as such.

Another important aspect of privacy especially in the age of Big Data is Contextual Integrity. Contextual Integrity is a concept that has arisen in recent years to provide guidance on how to respond to conflicts between values and interests, and to provide a systematic setting for understanding privacy **Error! Reference source not found.** It is not proposed as a full definition of privacy, but as a framework for evaluating the flow of information between agents (individuals and other entities) with a particular emphasis on explaining why certain patterns of flow provoke public outcry in the name of privacy (and why some do not). Contextual Integrity defines a context specified by roles, activities, norms, and values that interact with one another. The actors in this context are: senders, receivers, subjects and the attributes are data fields.

Contextual Integrity is very much at odds with the Big Data business model that actually aims to collect and integrate as many data sources as possible and gain new insights from those data through overarching mining and analyses. It uses data that has been collected under different pretexts and circumstances. This *repurposing* of data is totally against the concept of Contextual Integrity as described.

In other works, Hildebrand [33] examines the concept of privacy from a legal perspective and concludes that "privacy concerns the freedom from unreasonable constraints that creates the freedom to reconstruct one's identity". Data mining might therefore be seen as impacting the personal development of identity.

3.3 Legal Frameworks

The European position towards Learning Analytics has been expressed in the European Commission's report: "New Modes of Learning and Teaching in Higher Education" [34]. In recommendation 14, the Commission clearly stated: *Member States should ensure that legal frameworks allow higher education institutions to collect and analyse learning data. The full and informed consent of students must be a requirement and the data should only be used for educational purposes*", and, in recommendation 15: "*Online platforms should inform users about their privacy and data protection policy in a clear and understandable way. Individuals should always have the choice to anonymise their data.*"⁵

They base these recommendations on the EU Data Protection Directive 95/46/EC [35], i.e., the European law on personal data protection, comparable to the OECD's Fair Information Practice Principles [36]. Both are widely accepted frameworks and are mirrored in the laws of many U.S. states and other nations and international organisations.

Directive 95/46/EC defines personal data as "*any information relating to an identified or identifiable natural person ('Data Subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.*" The Directive is only applicable if automated processing of 'personal data' is employed or if it is part of a 'filing system' (Article 3). This means, in reverse, that if data is properly anonymised, the EU directive 95/46/EC does not apply. However, experts around the world are adamant that 100% anonymisation is not possible (see Section 4.3 below).

With this Directive, the EC is following the previously mentioned ethical frameworks such as the Nuremberg Code and the Helsinki Declaration by demanding consent, legal compliance, and that research should be beneficial for society. But it also specifies some general obligations in Article 6 such as: *Use of personal data need to be:*

- processed fairly and lawfully;
- for specified, explicit and legitimate purposes;
- safeguarded from secondary use and further processing;
- adequate, relevant and not excessive;
- accurate and up to date;
- stored no longer than necessary;

It is important to highlight that these principles are equally valid whether or not the data subjects provided full consent to access and use their data.

With Article 6, EU Directive 95/46/EC addresses a very important aspect towards Big Data business models [37] in restricting the use of data for limited purposes only. Big Data business models are driven by the collection and storing of infinite amounts of data without an expiry date and for later re-purposing. Thus, most of

⁵http://ec.europa.eu/education/library/reports/modernisation-universities_en.pdf

the time, the data is being collected in different contexts, and found later to be of benefit for other information needs. This is diametrically opposed to existing data legislation like the EU Directive 95/46/EC and the concept of Contextual Integrity as explained in the above section on privacy. The reuse of collected data for other (unspecified) purposes is against the information rights of an individual.

The commercial habit of indiscriminate collection and repurposing of data, therefore, strengthened the idea of adding another legal aspect to the EU Directive 95/46/EC which entails *the right to be forgotten*. This legal concept was put into effect at some occasions in the European Union since 2006. It arose from the desire of individuals to "determine the development of their life in an autonomous way, without being perpetually or periodically stigmatized as a consequence of a specific action performed in the past." [38]. Following the high-profile case of EU vs. Google in 2014, search data can now be removed on request by the user. There are, however, serious challenges connected with the right to removal of personal digital footprints and data shadows – the former being the traces left by users in electronic systems, the latter representing data about an individual left by others (e.g. tagged photos in facebook) [39]. More often than not it is unclear against whom such right can be claimed [ibid.]. Things are being complicated even more as recent national and European legislation works towards data storage for security purposes, as is apparent in the European Data Retention Directive, or, more formally known as Directive 2006/24/EC [40].

Besides the evolution of the legal frameworks that underlie learning analytics, the sector of private education service providers also started initiatives towards self-regulation in a move to appease critics after the inBloom case. Some 187 K12 school service providers in the US have signed a *Privacy Pledge*⁶ to protect data gained from educational data subjects and especially the school sector. In the UK, the JISC recently published a first draft of a Code of Practice [14] for learning analytics. One of the examples mentioned in the Code is "Ethical use of Student Data for Learning Analytics Policy" which has been established at the Open University UK [15]. The JISC Code of Practice and the Policy by the Open University are good starting points and blueprints for other implementations to follow. In the Netherlands too, the SURF foundation recently released a guiding paper how to treat educational data in a privacy conform way [41].

4. Fears towards Learning Analytics

Researchers and institutions dealing with Learning Analytics are facing up to privacy as a big concern. Many people are not really aware of the legal boundaries and ethical limits to what they can do within the sphere of privacy protection. Institutions, on the one hand, have a fiduciary duty and need to demonstrate care for the well-being and positive development of students, leading them to success in studying [17]. On the other hand, there is widespread fear of negative consequences from the application of learning analytics, such as negative press or loss of reputation and brand value, or even legal liabilities. Despite the enormous promise of learning analytics to innovate and change the educational system, there are hesitations regarding, among other things, the unfair and unjustified discrimination of data subjects; violation of personal privacy rights; unintended and indirect pressure to perform according to artificial indicators; intransparency of the learning analytics systems; loss of control due to advanced intelligent

systems that force certain decisions; the impossibility to fully anonymise data; safeguarding access to data; and, the reuse of data for non-intended purposes. This is a non-exhaustive list, but we need to take all concerns seriously if we are to establish proper implementations for learning analytics.

From what was said, one could conclude that privacy and legal rights are a burden which we need to combat in the same "formal arena", perhaps by amending the laws. Some people already argued for an adjustment of existing privacy rights to the new age of Big Data, like facebook founder Mark Zuckerberg when he famously stated: "Privacy is dead!" [45]. True, privacy got heavily affected by the latest technology developments and the ever rising computer processing power. But, in fact, these basic human rights have been established for very good reasons and we should firstly care to build technologies that support human learning within the existing ethical & legal boundaries, created many years before the Big Data boom, in the Belmont report and Directive 95/46/EC [24][35].

Similarly, we would refrain from solving a weakness in a new learning technology by proposing technical fixes or technological solutions, such as standardisation approaches, e.g. the emerging IMS Caliper specification⁷ (cf. also [12]). Instead, we prefer to see this as a "soft" issue, rooted in human factors, such as angst, scepticism, misunderstandings, and critical concerns. Within this paper, we, therefore, aim to provide some answers to the fog of misunderstandings around privacy and legal obligations dealing with the semantics of the concepts and translating them into clear action points through the suggested DELICATE checklist. We hope to show that ethical and privacy supported Learning Analytics are indeed possible, and we would like to encourage the Learning Analytics community to turn the privacy burden into a privacy quality label.

There are a wide variety of anxieties expressed with regards to the analysis, sharing and exploitation of personal data, including learner records and digital traces. These are not confined to education alone. Rather, such fears are transferred from other environments (commercial, governmental, social) with little differentiation of the respective domain. Compared to these, we'd argue, the institutional environments in education can be considered as relatively safe havens. This is due to the considerate nature of the education system being in charge of young people's personal development, the long-standing experience and practice of ethical behaviours (including the promotion of such behaviours to the learners), and the public scrutiny and transparency which goes far beyond any other sector. Learners, therefore, should be able to feel in a safe space, where they are allowed to make mistakes without the fear of consequences or unnecessary publicity.

Below, we summarise the most widespread and most critical topics, concerns or arguments against applied Learning Analytics. These have been collected from participants of the earlier mentioned workshop series EP4LA that took place at different locations between October 2014 and March 2015. The organisers received an overwhelming interest from all over the world, which resulted in six international workshops on ethics and privacy with legal and Learning Analytics experts who discussed current regulations and norms for most pressing questions.

⁶ http://studentprivacypledge.org/?page_id=45

⁷ <http://www.imsglobal.org/activity/caliperram>

4.1 Power-relationship, data/user exploitation

One of the criticisms levelled against analytics and Big Data in general is the asymmetrical power relationship it entails between the data controller and the data subject [17]. This can lead to a feeling of being powerless and exploited. That this concern reaches wider than Learning Analytics is demonstrated by events like the 2015 summit of the International Society of Information Sciences (#IS4IS) in Vienna, dedicated to “the Information Society at the crossroads”⁸. One key argument in this context is the exploitation of digital labour, as is the case with facebook – which is widely regarded as the world’s biggest advertising company – monetising user data and contributions for commercial profits [43]. Another criticism is the fact that data models are not neutral, but reflect and reversely influence an “ideal” identity. This means that data subjects are given a (designed) data identity by others, which can cause friction, especially in the context of the global economic and cultural divisions between Western and developing world: “people are operating within structured power relations that they are powerless to contest” [44]. To a smaller extent there is also such an asymmetry in power between Learning Analytics providers and users.

While these concerns are real and serious, things present themselves differently in an institutional learning context. Naturally, there has always been an asymmetrical relationship between the institution, the teacher, and the learner. However, the benevolent fiduciary mission and walled garden of education should install confidence and trust also in the use of digital assets and data. Nevertheless, challenges here are the increased openness of teaching tools, using cloud services and social networks, as well as the pressures to commercialise and commodify education [45].

4.2 Data ownership

At present, there is no clear regulation for data ownership of any party, i.e. neither the student, the university or a third party provider. Data ownership is a very difficult legal concept. It is assumed, that the digital footprints someone leaves behind belong to the data subject, i.e. the users themselves, as long as it isn’t specified differently in the terms of service of the provider of the digital system. An additional factor is that the data subject cannot manage all those tons of data breadcrumbs on a daily basis. Thus, data subjects are in need of having service providers take care of the data storage and management. But, according to EU Data Protection Directive 95/46/EC article 12, the data subject always has the right to know about all the information that has been collected about them. Just recently, a law student from Austria brought down the so-called Safe Harbour data transfer agreement between the European Union and the United States used by more than 4,000 companies, including Google, Facebook, and IBM⁹. He won a law suit that took more than two years and forced facebook and other Big Data providers to keep all data collected from a data subject in the same country and provide an overview of this data to the users¹⁰.

Data ownership becomes even more complex, however, when we consider the processing of data. If there is a computational model developed from a collection of data traces in a system, do the data

subjects then perhaps have some information rights on this data model? Can a student still opt-out of such a data model that is being generated around their data traces? In search for answers, data ownership remains a very complicated issue that is mainly dominated by the technical power the developers and service providers offer to the data subjects. There are visions to change this power relationship by enabling individuals to carry and curate their own datasets through “personal data stores”. This idea parallels earlier such user-centric solutions, like openID for generic authentication across the web. It could fundamentally turn the whole data movement in education around if it were to become real. One such attempt has already been initiated with the MIT’s openPDS¹¹, which “allows users to collect, store, and give fine-grained access to their data all while protecting their privacy”.

4.3 Anonymity & data security

According to PricewaterhouseCoopers¹², on average, around 117,339 attempts of information heists were made per day in 2014. These statistics reveal the severity of the problem of data security. Such attacks bring even big players like Apple, Sony, or Microsoft into trouble and they suffer from not being able to protect their online systems in an adequate way. Cyber-attacks are perhaps an even greater threat for universities and smaller educational providers who have fewer resources to establish and maintain appropriate data security measures. Naturally, this then becomes an issue for the protection of student and staff learning analytics data.

Anonymisation is often seen as the “easy way out” of data protection obligations. Institutions use various methods of de-identification to distance data from real identities and allow analysis to proceed [46]. Most of the time, though, data controllers consider replacing identifiers as sufficient to make data anonymous. But many studies have shown that this kind of anonymisation can, at best, be considered as pseudonymisation [47]. Anonymised data can rather easily be de-anonymised when they are merged with other information sources. Famous approaches include de-anonymisation of medical datasets: names of patients were not included but demographic information that could be linked to electoral registers allowed the retrieval of names and contact information for the medical records [48]. Among those relatively easy approaches are also more computational ones as presented in [49][50]. They showed how Netflix users could be re-identified from the dataset of a Netflix competition by combining it with data from the movie platform IMDB. Those and other examples show that robust anonymisation is hardly achievable with the increase of computational methods, data integration and calculation power. Some residual risk of identification has to be taken into account for educational data and Learning Analytics. Data security and encryption, therefore, has a vital role to play in the acceptance of such systems and combined security strategies, including anonymisation, can go a long way to protect individual privacy from illegal access.

As a further measure to make data less sensitive over time and to protect privacy, data degradation has been suggested whereby a timestamp is introduced when data should be deleted in order to prevent further use [51]. This approach lets data decay over time

⁸ <http://summit.is4is.org/is4is-summit-vienna-2015>

⁹ <http://www.reuters.com/article/2015/10/07/eu-ireland-privacy-schrems-idUSL8N1272Z820151007>

¹⁰ <http://europe-v-facebook.org>

¹¹ <http://openpds.media.mit.edu/>

¹² <http://www.pwc.com/gx/en/consulting-services/information-security-survey/key-findings.jhtml>

and in that way protects privacy and informational self-determination of data subjects.

4.4 Privacy and Data Identity

It can be argued that our (external) identity is made up by the sum of our actions and appearances. Arora [44] critically contrasts the two notions of system identity versus social identity, the latter encapsulating who we are in our social environment, while the former represents the collected digital data. This has become even more sensitive with the collection of personal biometrical information. Logging, tracking and storing individuals online, therefore, can be considered an intrusion into our identity and a violation of our perceived selves, i.e. the information produced by our actions and appearances – exploited for other purposes by third parties without consent or even knowledge. Profiling actions by companies and governments have been under fire for some time. Their impact on the formation and declaration of a person's own identity, the democratic society, and civic self-determination have been highlighted by some scholars (e.g. [52][33]). Repurposing and fragmentation of personal digital identities isn't the only criticism raised in the context of personal integrity, but the fact that individuals are typecast into data identities not of their own choosing and measured against benchmarks, indicators, or algorithmic stereotypes that are out of their control [44]. What is more, if one doesn't fit into a data model, it is often applied on a probabilistic basis [52].

Here again, the special relationship in education can ease the problem. Students are in a "learning contract" with the institution or training provider they sign up with. For the duration of this relationship, the teacher and institution need to be trusted to act responsibly and in the favour of its clients. This is unlike the ambiguous commercial relationship that for-profit enterprises have towards their investors and shareholders versus customers.

4.5 Transparency and trust

It is often said that lack of transparency can cause unease and concern with data subjects. However, it is rarely defined how this transparency should manifest itself. Commercial providers like Google keep their algorithms secret, and, yet, as long as results are relevant and in line with users' expectations, there is trust in the service, despite it being a black box. On the other hand, Google Takeout, which allows insight and export of the user dataset from all Google services, is of no great use to ordinary end users, as they are unable to understand or re-use that data. Here then, again, we see the asymmetrical power relationship: while transparency of the user leads to commodification of their data, the reverse isn't true for large companies.

The Korean-born German philosopher Byung-Chul Han [53] states that transparency turns social systems of trust into control systems because information is readily available. Learning Analytics is perceived as making learners/teachers transparent and open for criticism, while keeping the system itself opaque and out of scrutiny. There is widespread anxiety in the education community that data and information retrieved from data subjects may be used against "outliers", and, thus, leads to more conformity and uniformity [2]. As such, analytics is perceived by some as an engine for controlling and correcting behaviours.

The issue in education can best be tackled by being clear and open about the purpose of data collection and the compliance with the existing legal frameworks. A code of conduct can clarify to the stakeholders and data subjects what the intentions for analytics are, how long data is being kept, and what internal procedures are available to contest negative consequences. As an additional

measure, the focus of analytics should be put on self-reflection rather than prediction or justification. Playing analytics results back to the data subject and letting them decide for themselves, whether to ask for pedagogic support and intervention or not, puts the learner in control, something that is anyway desirable as an educational outcome.

5. Conclusions

In order to use educational data for Learning Analytics in an acceptable and compliant way, and to overcome the fears connected to data aggregation and processing, policies and guidelines need to be developed that protect the data from abuse and ensure treatment in a trustful way.

The development and implementation of Learning Analytics could benefit from the application of a value-sensitive design approach as specified by Friedman in 1997 [52]. Value-sensitive design is the idea that ethical agreements and existing privacy law need to be embedded where and when it is relevant for the design and usage of a system like Learning Analytics – starting early on in the design and implementation process, and close to where the technology is being rolled out.

We need to see data protection not as a mere legal requirement, but should embed the care about privacy deeply into Learning Analytics tools and increase the trust of data subjects in these systems. Privacy should not be seen as a burden but rather as a valuable service we can offer to build trusting relations with our stakeholders. To build this kind of relationship, a high degree of openness, combined with reassuring explanations referencing relevant legislation like the EU Directive 95/46/EC are needed. Therefore, the "contract" between learners and their educational providers needs to be reviewed and adjusted to reach this level of trust in order to have the backing to release the full potential of Learning Analytics.

To support this new "learner contract" as the basis for a trusted implementation of Learning Analytics, we developed the eight point DELICATE checklist. The checklist is derived from the intensive studying of the legal texts mentioned above and several round-table talks of experts, e.g. at the six EP4LA workshops. It has been put into its final shape by the LACE project and in close cooperation with its associated partners across Europe.

The DELICATE checklist contains eight action points that should be considered by managers and decision makers planning the implementation of Learning Analytics solutions either from their own institution or from any external provider. The eight points are: 1. Determination, 2. Explain, 3. Legitimate, 4. Involve, 5. Consent, 6. Anonymise, 7. Technical aspects, 8. External partners. Figure 1 below shows the full overview of the checklist and all its relevant sub questions.

We hope that the DELICATE checklist will be a helpful instrument for any educational institution to demystify the ethics and privacy discussions around Learning Analytics. As we have tried to show in the article, there are ways to design and provide privacy conform Learning Analytics that can benefit all stakeholders and keep control with the users themselves and within the established trusted relationship between them and the institution.

In conclusion, we believe that Learning Analytics projects should follow a value-sensitive design process, which allows considering ethical and privacy values on the same level as functional requirements. Thereby, the aforementioned ethical considerations are not seen as an unfortunate constraint, but help to develop a system that achieves its aims not only in a technical but also in an



The DELICATE Checklist
Implementing trusted Learning
Analytics in education



- D** **DETERMINATION** – Why you want to apply Learning Analytics
 - ▶ What is the added value (Organisational and data subjects)
 - ▶ What are the rights of the data subjects (e.g., EU Directive 95/46/EC)
- E** **EXPLAIN** – What are the objectives and boundaries
 - ▶ What data will be collected for which purpose?
 - ▶ How long will this data be stored?
 - ▶ Who has access to the data?
- L** **LEGITIMATE** – Why you are allowed to have the data?
 - ▶ Which data sources you have already (aren't they enough)
 - ▶ Why are you allowed to collect additional data?
- I** **INVOLVE** – Involve all stakeholders and the data subjects
 - ▶ Be open about privacy concerns (of data subjects)
 - ▶ Provide access to the personal data collected (about the data subjects)
- C** **CONSENT** – Make a contract with the data subjects
 - ▶ Ask for a consent from the data subjects before the data collection
 - ▶ Define clear and understandable consent questions (Yes / No options)
 - ▶ Offer the possibility to opt-out of the data collection without consequences
- A** **ANONYMISE** – Make the individual not retrievable
 - ▶ Anonymise the data as far as possible
 - ▶ Aggregate data to generate abstract metadata models (Those do not fall under EU Directive 95/46/EC)
- T** **TECHNICAL** – Procedures to guarantee privacy
 - ▶ Monitor regularly who has access to the data
 - ▶ If the analytics change, update the privacy regulations (new consent needed)
 - ▶ Make sure the data storage fulfills international security standards
- E** **EXTERNAL** – If you work with external providers
 - ▶ Make sure they also fulfil the national and organisational rules
 - ▶ Sign a contract that clearly states responsibilities for data security
 - ▶ Data should only be used for the intended services and no other purposes

ethical and humane manner. At the same time, we request that the

education sector as a whole and institutions in particular need to get better in distinguishing themselves from commercial profit-oriented enterprises and advertise their mission of care and support for individuals to revive trust in the system.

The eight-point DELICATE checklist represents the most compact approach for educational managers and policy makers to

ACKNOWLEDGMENT

The efforts of Hendrik Drachsler have been partly funded by the EU FP7 LACE-project.eu (grant number 619424). We thank all the enthusiastic participants around the planet who attended the EP4LA workshop series. Those smart discussions supported us in crafting the DELICATE checklist.

References

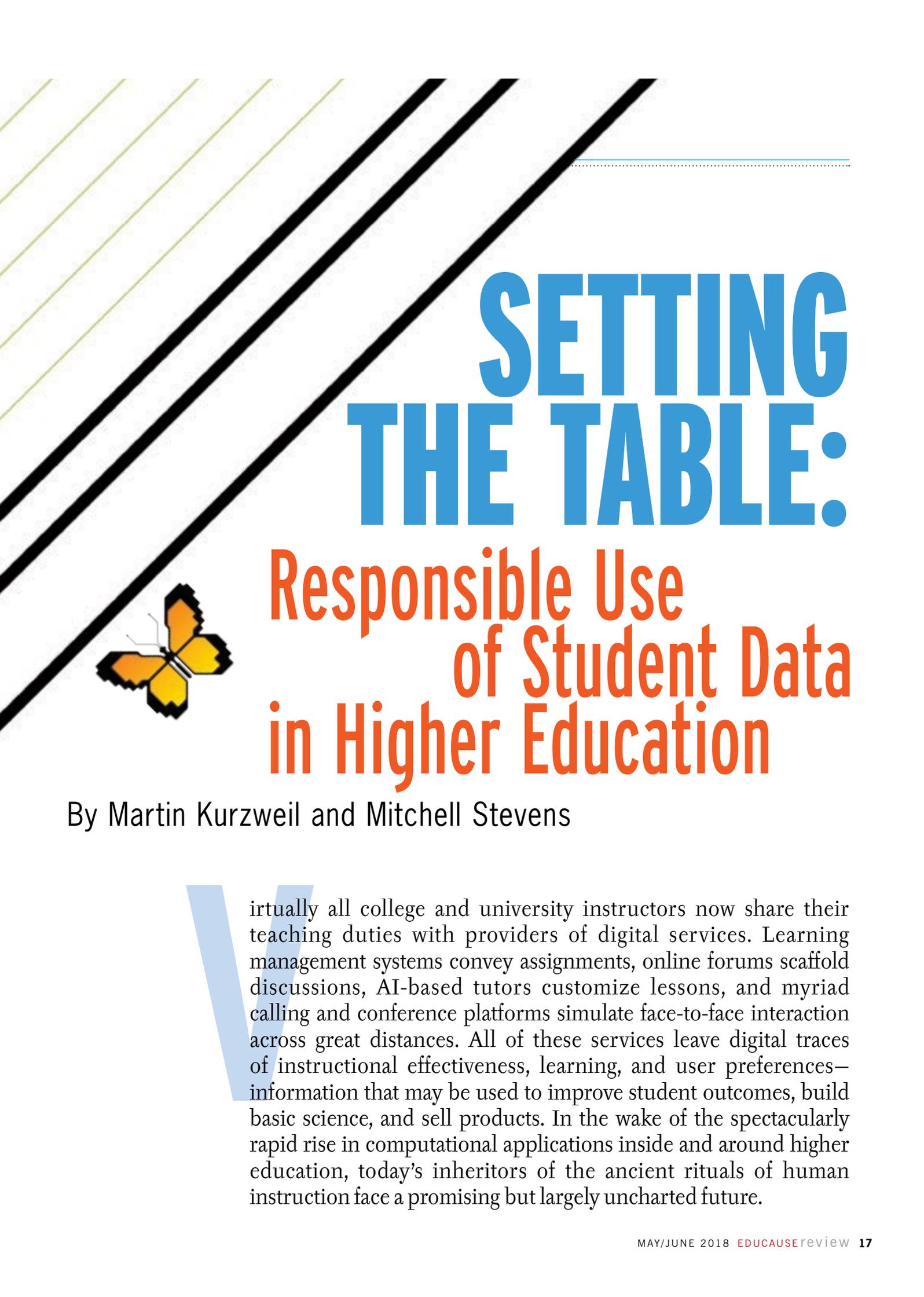
- [1] Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium.
- [2] Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, 15 (3), 42–57.
- [3] New York Times, (2014) InBloom Student Data Repository to Close. April 21, 2014. Available at: http://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/?_r=0
- [4] College Bescherming Persoonsgegevens. Onderzoek CBP naar de verwerking van persoonsgegevens door Snappet. 2014. Available at: https://cbpweb.nl/sites/default/files/downloads/mijn_privacyrap_2013_snappet.pdf.
- [5] Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- [6] Scheffel, M., Drachsler, H., Stoyanov S., & Specht, M. (2014). Quality Indicators for Learning Analytics. *Educational Technology & Society*, 17 (4), 117–132.
- [7] Drachsler, H., & Greller, W. (2012). The pulse of learning analytics: understandings and expectations of the stakeholders. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 120-129). ACM.
- [8] Prinsloo, P. & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in Learning Analytics. Paper presented Learning Analytics conference 2013, Leuven, Belgium.
- [9] Slade, S. & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, 57 (10), pp. 1510–1529.
- [10] Pardo, A. & Siemens, G. (2014). Ethical and privacy principles for Learning Analytics. *British Journal of Educational Technology*, 45(3), 438–450.
- [11] Prinsloo, P. & Slade, S. (2015). Student privacy self-management: implications for Learning Analytics. Paper presented at LAK15, Poohgkeepsie, New York, USA.
- [12] Hoel, T. & Chen, W. (2015). Privacy in Learning Analytics – Implications for System Architecture. In: Watanabe, T. and Seta, K. (Eds.) (2015). *Proceedings of the 11th International Conference on Knowledge Management* (Pre-Print)
- [13] Sclater, N. (2014). Code of practice for learning analytics. A literature review of the ethical and legal issues. JISC report. http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-Literature_Review.pdf.
- [14] Sclater, N., & Bailey, P. (2015). Code of practice for learning analytics. Available at: <https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>
- [15] Open University UK (2014). Policy on Ethical use of Student Data for Learning Analytics. Available: http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.k_students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf
- [16] Kay, D., Korn, N., and Oppenheim, C. (2012). CETIS Analytics Series: Legal, Risk and Ethical Aspects of Analytics in Higher Education, serial number: ISSN 2051-9214 Vol 1, No 6. Available at: <http://publications.cetis.org.uk/2012/500>
- [17] Slade, S., & Prinsloo, P. (2015). Student vulnerability, agency and learning analytics: an exploration (Pre-Print).
- [18] Steiner, C., Kickmeier-Rust, M.D., and Albert, D. (2015). LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox, JLA 2015 (Pre-Print)
- [19] Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review* 79(1). Available: <https://www.nyu.edu/projects/nissenbaum/papers/washingtonlawreview.pdf>
- [20] Lauss, G., Bialobrzeski, A., Korkhaus, M., Snell, K., Starkbaum, J., Vermeer, A.E., Weigel, J., Gottweis, H., Helén, I., Taupitz, J. & Dabrock, P. (2013). Beyond Genetic Privacy. Past, Present and Future of Bioinformation Control Regimes. Available at: http://private-gen.eu/uploads/media/PRIVATE_Gen_FINAL-REPORT_2013_02.pdf
- [21] Lauss, G., Snell, K., Bialobrzeski, A., Weigel, J., Helén, I. (2011). Embracing complexity and uncertainty: An analysis of three orders of ELSA research on biobanks. In: *Genomics, Society and Policy*, vol.7: pp.47-64. Springer. Available at: <http://www.lsspjournal.com/content/pdf/1746-5354-7-1-47.pdf>
- [22] Weindling, P. (2001). The origins of informed consent: the international scientific commission on medical war crimes, and the Nuremberg Code. *Bulletin of the History of Medicine*, 75(1), 37-71.
- [23] World Medical Association. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79(4), 373.
- [24] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Bethesda, MD.

- (1978). The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. ERIC Clearinghouse.
- [25] Milgram, Stanley. *Das Milgram-Experiment*. Reinbek: Rowohlt, 1974.
- [26] Zimbardo, P. G., Maslach, C., & Haney, C. (2000). Reflections on the Stanford prison experiment: Genesis, transformations, consequences. Obedience to authority: Current perspectives on the Milgram paradigm, 193-237.
- [27] Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788-8790.
- [28] Kleinsman, J., & Buckley, S. (2015). Facebook Study: A Little Bit Unethical But Worth It? *Journal of bioethical inquiry*, 1-4.
- [29] Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard law review*, 193-220.
- [30] Westin, A. F. (1968). Privacy and freedom. *Washington and Lee Law Review*, 25(1), 166.
- [31] Flaherty, D. H. (1989). Protecting privacy in surveillance societies: The federal republic of Germany, Sweden, France, Canada, and the United States. UNC Press Books.
- [32] Roessler, B. 2005. *The Value of Privacy*, Oxford: Polity Press.
- [33] Hildebrandt, M. (2006). 3. Privacy and Identity. *Privacy and the criminal law*, 43.
- [34] European Commission (2014). *New modes of learning and teaching in higher education*. Luxembourg: Publications Office of the European Union 2014, 68 pp. ISBN 978-92-79-39789-9 doi:10.2766/81897 http://ec.europa.eu/education/library/reports/modernisation-universities_en.pdf
- [35] European Union: EU Data Protection Directive 95/46/EC. <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31995L0046>.
- [36] Organisation for Economic Co-operation and Development. (2002). *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD Publishing. <http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm>
- [37] Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- [38] Mantelero, A. (2013). The EU Proposal for a General Data Protection Regulation and the roots of the 'right to be forgotten'. *Computer Law & Security Review* 29 (3): 229–235. Available at: https://www.academia.edu/3635569/The_EU_Proposal_for_a_General_Data_Protection_Regulation_and_the_roots_of_the_right_to_be_forgotten
- [39] Koops, B. J. (2011). Forgetting footprints, shunning shadows: A critical analysis of the 'right to be forgotten' in big data practice. In: *SCRIPTed*, Vol. 8, No. 3, pp. 229-256, 2011; Tilburg Law School Research Paper No. 08/2012. Available at: <http://ssrn.com/abstract=1986719> or <http://dx.doi.org/10.2139/ssrn.1986719>
- [40] European Union: EU Data Retention Directive 2006/24/EC <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32006L0024>
- [41] Engelfriet, A., Manderveld, J., Jeunink, E., (2015). Learning Analytics onder de Wet bescherming persoonsgegevens. SURF report. Presented 9th of November at <http://www.deonderwijsdagen.nl/programma/785-programma-pre-conferentie-9-november/>
- [42] The Guardian (2010), Privacy no longer a social norm, says facebook founder. (January, 11,2010) <http://www.theguardian.com/technology/2010/jan/11/facebook-privacy>
- [43] Fuchs, C. (2015). *Digital Labour and Karl Marx*. Routledge.
- [44] Arora, P. (2015). Bottom of the Data Pyramid: Big Data and the Global South. In: *Discover Society*, vol. 23. Available at: <http://discoversociety.org/2015/08/03/bottom-of-the-data-pyramid-big-data-and-the-global-south/>
- [45] Casey, J., & Greller, W. (2015). Jane Austen and the Belly of the Beast Part 2 - Language and Power: Commodification, Technology and the Open Agenda in Higher Education. In: *ISIS Summit Vienna 2015—The Information Society at the Crossroads*. Multidisciplinary Digital Publishing Institute. Available at: <http://sciforum.net/conference/isis-summit-vienna-2015/paper/2910/download/pdf>
- [46] Tene, O., & Polonetsky, J. (2012). Privacy in the age of big data: a time for big decisions. *Stanford Law Review Online*, 64, 63. Available at: <https://www.stanfordlawreview.org/online/privacy-paradox/big-data>
- [47] Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, 57 *UCLA L. Rev.* 1701
- [48] Sweeney L. (2000). Simple Demographics Often Identify People Uniquely. Carnegie Mellon, Data Privacy Working Paper 3. <http://dataprivacylab.org/projects/identifiability/>.
- [49] Narayanan, A., Shmatikov, V. (2008). Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset). The University of Texas at Austin February 5, 2008. Te vinden op: <http://arxiv.org/pdf/cs/0610105.pdf>. (laatst bekeken op 25-03-2015)
- [50] Narayanan, A., and Felten, E.W. (2014). No silver bullet: De-identification still doesn't work. <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>Manuscript.Last accessed: Saturday, 24 October 2015, 10:25 PM
- [51] Heerde, H.J.W. van (2010). Privacy-aware data management by means of data degradation - Making private data less sensitive over time. Ph.D. thesis, University of Twente, Enschede. <http://www.vanheerde.eu/phdthesis.pdf>. (laatst bekeken op 25-03-2015).
- [52] Hildebrandt, M. (2008). Profiling and the identity of the European citizen. In: *Profiling the European citizen* (pp. 303-343). Springer Netherlands.
- [53] Han, B.-C. (2015). *The Transparency Society*. Stanford University Press. ISBN: 9780804794602
- [54] Friedman, B., ed. (1997). *Human values and the design of computer technology*. Cambridge University Press.

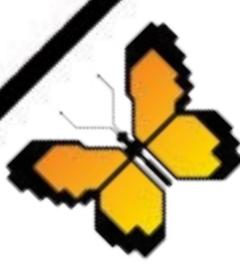
ARTIKEL 3

‘SETTING THE TABLE: RESPONSIBLE USE OF STUDENT DATA IN HIGHER EDUCATION’

Martin Kurzweil and Mitchell Stevens



SETTING THE TABLE:



Responsible Use of Student Data in Higher Education

By Martin Kurzweil and Mitchell Stevens

Virtually all college and university instructors now share their teaching duties with providers of digital services. Learning management systems convey assignments, online forums scaffold discussions, AI-based tutors customize lessons, and myriad calling and conference platforms simulate face-to-face interaction across great distances. All of these services leave digital traces of instructional effectiveness, learning, and user preferences—information that may be used to improve student outcomes, build basic science, and sell products. In the wake of the spectacularly rapid rise in computational applications inside and around higher education, today’s inheritors of the ancient rituals of human instruction face a promising but largely uncharted future.

Which streams of data about learners are properly and positively integrated with one another, and which are best kept distinct?

Which streams of data about learners are properly and positively integrated with one another, and which are best kept distinct? Should the information be kept forever, or if not, under what conditions should it be erased? Does the information produced through digital platforms impose any obligations on those who have access to it? Who is entitled to make money off these data, and what responsibilities does such business entail? These are among the many questions facing educators and vendors about the ethics and politics of information.

Inherited guidelines give everyone little to go on when answering these questions. US government regulations pertaining to student records were drafted under the assumption that the most enduring traces of instructional exchange were kept on paper. Grades were recorded in letters and translated into metrics by hand. Most evaluation required human eyes and human thinking. Integrating information held by different offices of the same organization was cumbersome and costly. Perhaps most important, instructors were presumed to be singularly sovereign over what took place in “their” classrooms.

None of the above obtains today—except for the Family Rights and Privacy Act of 1974, which serves the digital present about as well as a bicycle serves a kangaroo. Many Americans look to the somewhat more sophisticated rules for data use developed by the European Union as a potential framework for US practice; however, the EU program is built on the premise that users can be the final

arbiters of the disposition of “their” data. In a world in which the owners of digital platforms (e.g., Alphabet, Amazon, and Facebook) already have reaped incalculable profits from the production and aggregation of data describing users, in the process amassing more information about people than any government in world history, the presumption of individual data propriety is wishful thinking.

It is time instead for a frank and forward-focused discussion of how to define ethical information practice in academia. This is the context in which we created the Stanford CAROL and Ithaca S+R project on Responsible Use of Student Data in Higher Education. Our goal was simple, but challenging: to articulate first principles that might frame institutional policies on the use of student data in the digital era.¹ In our view, four core premises ought to be at the heart of this inquiry.

First, education is fundamentally a human endeavor. It can be richly supported and enhanced by technologies (algorithms, blackboards, machines, paper), but it cannot be fully accomplished independently of human action. Second, education is only partially a business activity. It is also a civic act: the practice of shaping people, communities, and societies and of transmitting cultural inheritance across generations. Third, retention of the humane and civic character of education cannot be taken for granted. They are fragile, and their preservation requires active, diligent, sustained effort. Fourth, with information and knowledge comes responsibility. Awareness of educational practices that

are suboptimal and of available ways to improve those practices *requires* educators—whether or not they are part of businesses—to proactively change what they do. It is in this spirit of responsibility that we survey the current landscape and offer a framework for ambitiously leveraging digital innovations for critical improvement in higher education.

Emerging Uses of Student Data

Higher education institutions are using student data in many innovative ways.² Let’s start with admissions and enrollment management, an area that has long utilized data-driven practices. Today the steeply diminished costs of computation have coupled with fierce competitive pressures in the postsecondary ecology to make student recruitment and selection a rapidly evolving technology domain. As colleges and universities gain access to more data about students and augment their analytic capacity, they can ever more precisely predict which students will attend and which will succeed. Sophisticated algorithms now inform recruitment campaigns, admissions decisions, and financial aid offers worldwide.

But recruitment is hardly the crest of the campus technology wave. Many institutions now base myriad business decisions on data describing student outcomes. Between 2003 and 2014, Georgia State University (GSU) increased its graduation rate from 32 percent to 54 percent by using data to discover and address problems of retention and completion. For example, after mining historical data to identify courses in which students consistently performed poorly,



administrators created a supplemental instruction program with peer advisors for those courses. Further observation showed that although there was improvement in passing rates in many of the courses targeted for supplemental instruction, introductory mathematics courses in algebra, pre-calculus, and statistics remained stumbling blocks. GSU administrators and math faculty responded by redesigning those courses in a flipped format and saw the DFW (drop-fail-withdrawal) rate fall from 43 percent in 2006 to 19 percent in 2014.³

Or consider GSU's Panther Retention Grant program, created in 2011. After analysis revealed that hundreds of students in good academic standing and within three semesters of graduating were dropping out, administrators investigated and determined that many of these students were unable to register for courses because of small, unpaid balances on their term bills—a restriction codified in state law. To address this, GSU created a targeted grant program offering an average of \$900 to students in those circumstances. Of Panther Retention Grant recipients (who otherwise would not have been able to register), 88 percent graduated or were still enrolled twelve months later, and the tuition revenue from those retained students more than covered the cost of the program.⁴

Predictive analytics also are being put into the hands of instructors, advisors, and students themselves. Early-alert systems aggregate and analyze data from multiple sources (gradebooks, learning management system [LMS] log-files, student information systems) to automatically flag student behavior associated with lower rates of academic success. Advisor-facing systems such as Arizona State University's eAdvisor integrate LMS information about student activity with registration data and student background characteristics. Advisors are notified when a student gets off track, and they are encouraged to intervene. eAdvisor also uses data describing individual academic performance to make

registration suggestions to students and advisors.

Systems often called *dashboards* are designed to provide instructors or students with aggregated information that might help them improve performance. Rio Salado College's RioPACE is a well-known example. The tools merge student demographic information and academic history with LMS log-file data to predict students' likelihood of success in a given course. Those predictions are conveyed to instructors, who can run custom analyses on demand and use what they learn to support particular learners. ASU's eAdvisor includes a student-facing dashboard as well. At the University of Michigan, E2Coach, a tool used in introductory STEM courses, automatically sends students personalized course-performance messages based on a continually updated algorithm.

Evidence of the effectiveness of such programs is limited but promising. A randomized study of student coaching supported by predictive analytics found that the service, offered by the company InsideTrack, improved retention rates by 3 to 5 percentage points compared with control groups whose members did not receive the coaching. Two randomized trials currently in the field are seeking to validate these findings at scale.⁵

Other innovations fall under the umbrella of *adaptive courseware*. These systems are digital platforms that collect information on student activity—time spent on task, task performance, and level of engagement, for example—to create “personalized learning paths” for students. Adaptive courseware systems offer dashboards and analytics tools enabling instructors to see where individual students and entire classes are struggling. Some systems include dashboards for students, enabling them to better understand their own progress and roadblocks. Although adaptive courseware is still a

relatively new technology, there is some promising anecdotal evidence of its efficacy. Findings from a 2016 study of the Bill & Melinda Gates Foundation's Adaptive Learning Market Acceleration Program suggests that implementation strategies make a difference with adaptive courseware and that the most (perhaps the only) effective outcomes accrue with full-scale course redesign.⁶

Uneven Adoption

While analytics programs are becoming much more common, only a minority of colleges and universities have systematically deployed them. According to a KPMG survey of senior administrators in July 2015, only 41 percent of respondents were using student data for predictive analytics, and just 29 percent reported having the internal capacity to analyze their own student data. Even those who are making efforts feel they are coming up short. The 2016 Campus Computing Survey revealed that less than one-fifth of respondents rated their institutions' data analytics investments as “very effective.” In a 2015 Ithaca S+R survey of a representative sample of four-year college faculty, a minority of respondents reported using any form of technology in instruction, although 63 percent said they would like to do so. In the EDUCAUSE Center for Analysis and Research (ECAR) 2017 study of faculty and information technology, between 16 and 28 percent of faculty responded that they did not have access to data-based

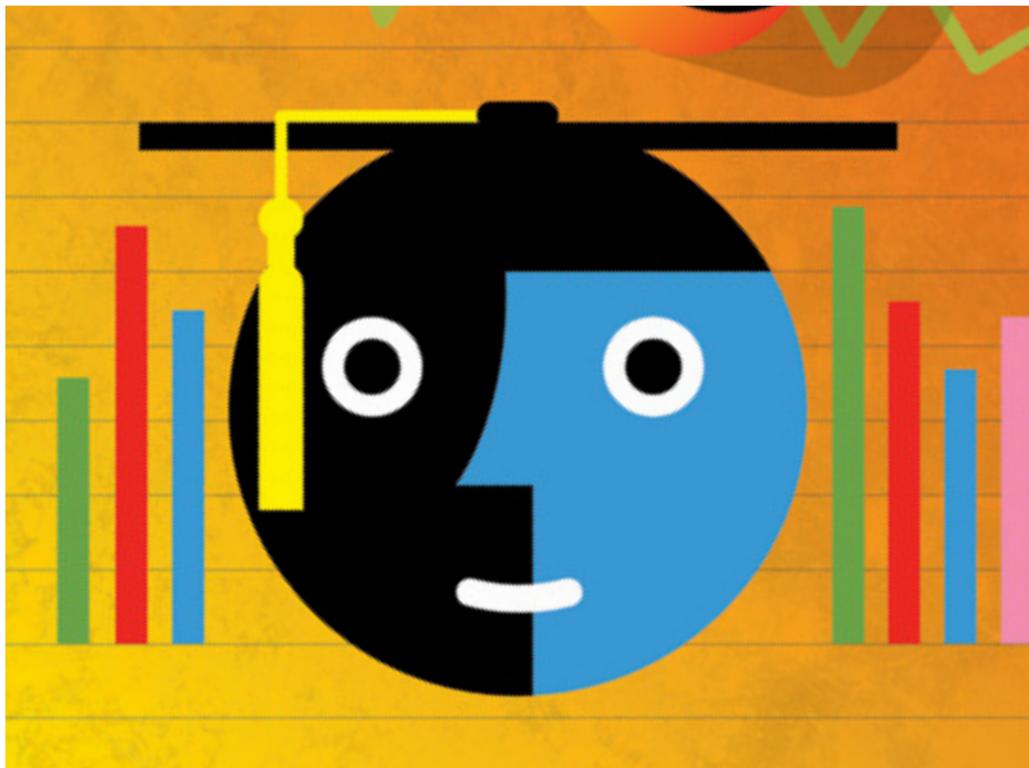
To achieve adoptions at scale, campuses must sustain a culture that embraces data-driven practices among administrators, instructors, and student-support staff.

planning and advising services, while between 23 and 34 percent of faculty have access but apparently choose not to use these services.⁷

Incompatible data systems are a significant drag on intramural change. The information needed for sophisticated analytics is typically dispersed and differentially formatted in student information systems, registrar records, and LMS log-files. Some colleges and universities have the technical, financial, and human resources to merge this data. Many do not.

Even at institutions that have overcome the logistical challenges, innovations frequently remain at the margins. To achieve adoptions at scale, campuses must sustain a culture that embraces data-driven practices among administrators, instructors, and student-support staff. This is no easy task. In the 2015 Ithaca S+R faculty survey, only 35 percent of respondents reported that they would be rewarded or recognized for modifying their pedagogy with technology.⁸

Despite the great promise of digital technologies to scaffold and improve instruction, a very deep political current pushes in the other direction: faculty sovereignty. The long-standing legacy of faculty autonomy over classrooms and curriculum gives those instructors with faculty appointments, particularly tenured ones, a great deal of power and prestige. After decades of decline in the number of tenure-track appointments and simultaneous growth in the ranks of student-services and IT personnel, people with faculty appointments often believe they have good reason to defend the turf remaining to them. In such a



context, the latest innovation heralded by the campus technology initiative is easily interpreted by the professoriate as further erosion of the borders marking what was long their own privileged domain.

Risky Business

Aside from campus turf skirmishes, educators have substantive reasons to be cautious in their embrace of computational learning technologies. Most important is the fuzzy line between prediction and prescription of academic futures. Advocates of the new learning analytics invariably emphasize the promise of using prior data about learners to target instruction in ways that best serve students' individual futures. Yet only rarely do these same advocates invoke the long and unsavory tradition of academic tracking, which justified the categorical tiering of academic opportunities on the basis of supposedly objective, "scientific" measures of students' abilities. The fact that academic tracking has paralleled and indeed reinforced

inequalities of race and social class is an important counterweight to the nearly uniform optimism of those in the edtech (educational technology) sector.⁹

Of course this optimism is essential to the business models of venture-backed startups, which rely on the potential of new platforms and algorithms to substantially improve individual and organizational behavior. Promises of dramatic performance spikes are part of the pitches that new firms make to investors and clients. The fact that major education philanthropies are increasingly funding private-sector players adds to the hype. But the hard truth is that meaningful gains in individual learning and organizational improvement are almost always incremental. The difference in the timetables of doing good business and building good educational practices is real, and the peculiar commingling of Silicon Valley swagger and academic caution is one of the defining features of the global edtech community. Whether this commingling will be for the good or ill of higher education in the

long run is an open question, but in the short term it makes for lots of crossed signals and reciprocal misunderstandings between those on different sides of the business/academia divide.

Another tension is between proprietary and fiduciary control of knowledge and the information that underlies it. Technology firms rely on ownership of their intellectual property and its rising value as user communities grow. Data describing instructors and students is often key to their business proposition, enabling firms to improve algorithms and customize operations competitively. Data may also have commercial value in its own right as a marketing resource or as the basis for commoditized consulting expertise.

Yet colleges and universities inherit a long-standing obligation to hold student credentials information securely and into perpetuity. When the information is covered under government statute, this obligation has the force of law. Additionally, academic research increasingly requires shared access to data to enable verification or disconfirmation of findings for scientific progress. At present, the domains of edtech and learning analytics are without commonly shared routines for adjudicating conflicts of interest in data use for academic, commercial, and scientific purposes.

Finally, transparency of evaluation and the possibility of revisiting academic evaluations are signal ideals of higher education. Colleges and universities have strong traditions of enabling students (and instructors!) to seek reconsideration of evaluations and request independent review. These traditions may be challenged when evaluation is shared with proprietary firms whose systems are computationally opaque, private property, or both. Such barriers to independent review may also make it difficult to determine whether computational systems reproduce bias or historically inequitable academic pathways and outcomes. Careful monitoring and mechanisms for overriding computational decisions can mitigate such risks

but may also undermine the reliability and general efficacy of these systems.

Colleges and universities, and their myriad subunits, have managed these challenges differently, leading to an uneven and highly uncertain ethical and procedural terrain. Coupled with the tech world's famous "bias toward action" is the perennial risk of a data use that will cross poorly articulated and still-in-draft ethical lines. But procedural caution has its own ethical risk: the failure to act in light of accumulating knowledge. This is why every field of professional endeavor maintains an ethical tradition of dual obligation. Do no harm, but do not hesitate to act on awareness of suboptimal practices and outcomes.

Principles of Responsible Use

Rapid movement at the cutting edge of edtech has far outpaced changes in the laws, institutional policies, and ethical frameworks that were crafted to inform responsible use of educational information in the twentieth century. This makes for a jarring recognition, but also an opportunity to revisit and rearticulate guiding ideals of responsible academic practice.

With this opportunity in mind, Stanford CAROL and Ithaca S+R convened colleagues from across higher education at the Asilomar Conference Grounds in Pacific Grove, California, in June 2016. The site was meaningful. In 1975, a group of 140 biologists, lawyers, and physicians met at Asilomar to write voluntary guidelines for ensuring the safety of recombinant DNA technology. An additional precedent for our work was the 1978 meeting at the Belmont Conference Center in Elkridge, Maryland, which produced a document informing ethical research with human subjects.

Through our preparatory work and the robust discussion at the convening, four basic tenets for the use of student data emerged: Shared Understanding;

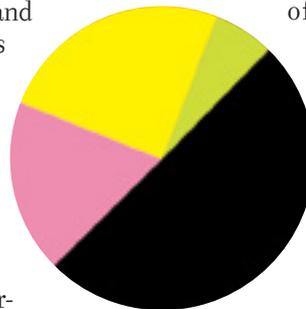
Transparency; Informed Improvement; and Open Futures.

Shared Understanding. *Instructors, administrators, students, and third-party vendors all contribute to the process of data production. All of these parties deserve to have a shared understanding of the basic purposes and limits of data collection.* Here we recognize the fundamentally plural character of digital data. Although most conversations about data ethics grant primary data ownership to the persons the data describes, we propose instead that all digital data be regarded as *joint ventures*. They require not only the contributions of students and instructors, but also the investment of those who create and maintain digital platforms and who hold that data in trust, whether as nonprofit universities or private firms. In this view, the information describing a particular student's learning interactions belong not just to the student. Rather, the student participates in ownership with the other parties contributing to the production of the information. All

those involved in a joint venture of teaching and learning deserve a shared definition of informational use and its limitations.

Transparency. *Clarity of process and evaluation is a hallmark of humane education systems and must be maintained even while those systems grow more complex.*

Students are entitled to (1) clear representations of the nature and extent of the information that describes them and that is held in trust by their institution and relevant third-party organizations; (2) an explication of how they are being assessed; and (3) the ability to request that assessments be reviewed through a clearly articulated governance process. Here we recognize the hallmark academic and scientific value of independent review. Sustaining this value brings new challenges in the era of machine learning, when computational systems routinely produce decisions through processes that are opaque even to system creators. We believe that the ideal of academic



The higher education community has a long tradition of adapting governance to safeguard the autonomy and integrity of the academic enterprise.

and scientific transparency is absolute and is essential to the legitimacy of any judgment on the basis of empirical evidence. In applications of digital technology to academic activity, transparency should be a design and engineering imperative.

Informed Improvement. *Learning organizations have an obligation to study student data in order to make their own educational environments more effective and to contribute to the growth of general knowledge.* Here we recognize that just as academic tradition obliges transparency, so too does it oblige action in the face of evidence. Instructors and academic administrators have vast stores of information describing instructional processes and outcomes. There is no question that some of that information will reveal bad news: particular instructors who disproportionately reward or discourage certain kinds of students; courses or entire programs that produce few measurable learning gains. Whereas diffusely distributed or nonexistent information may have hidden such news in the past, contemporary data management systems will surface it routinely. The ethic of informed improvement presumes that instructors and administrators will seek to remedy any problematic circumstances revealed by accumulating evidence.

Open Futures. *Education should enable opportunity, not foreclose it. Instructional, advisement, and assessment systems must always be built and used in ways that enable students to demonstrate aptitude, capacity, and achievement beyond their own or others' prior accomplishments.* Here we recognize the promise of digital technology to improve lives through

- learning, even while we remember
- that those same technologies can
- be used to block opportunity. We
- believe it is essential to create a guid-
- ing ethic wherein educators default
- to an ideal of opportunity creation
- rather than preemptive prescrip-
- tion. Predictive analytics should
- enable, not track—and it is precisely
- because the distinction between
- those two things is hard to specify
- that decision making must constan-
- tly be guided by the priority of
- open futures.

Conclusion

We view the four principles from the Asilomar convening as an initial contribution to an ongoing conversation that will include a wide range of stakeholders. People from business must be at the table, because technology firms and the holders of private capital supporting them will play only larger roles in the provision of postsecondary opportunity going forward. But all of us in higher education must set that table. Notwithstanding its reputation for resistance to change, the higher education community has a long tradition of adapting governance to safeguard the autonomy and integrity of the academic enterprise. It is time to incorporate new colleagues into that tradition and enlist their help in defining responsible use of student data in a rapidly changing world. If educators do not do this for themselves, others will. ■

Notes

1. The project was organized as a peer review. After working with colleagues to generate several white papers mapping the landscape of digital innovations in postsecondary provision, we convened academic and industrial scientists, senior university administrators, government officials, and representatives from major educational philanthropies at the Asilomar Conference Grounds in Pacific Grove, California, to consider an ethical framework for the responsible use of student data in higher education. The corpus of written work from the project to date is assembled at our website; in this article we attempt a more synoptic view.
2. Detailed descriptions of these efforts and others are included in Rayane Alamuddin, Jessie Brown, and Martin Kurzweil, *Student Data in the*

Digital Era: An Overview of Current Practices (New York: Ithaka S+R, September 6, 2016).

3. Martin Kurzweil and D. Derek Wu, *Building a Pathway to Student Success at Georgia State University* (New York: Ithaka S+R, April 23, 2015).
4. Jamaal Abdul-Alim, "Retention Rank Keeping Dreams Alive at Georgia State," *Diverse: Issues in Higher Education*, April 14, 2016; "Georgia State Launches Pilot Program to Help Retain Students," *OPBlog: Higher Ed Junction*, University of Washington, May 24, 2013.
5. Co-author Martin Kurzweil is the independent evaluator of one of these trials. For the earlier study, see Eric P. Bettinger and Rachel B. Baker, "The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising," *Educational Evaluation and Policy Analysis* 36, no. 1 (2014).
6. Louise Yarnall, Barbara Means, and Tallie Wetzel, "Lessons Learned from Early Implementations of Adaptive Courseware," *SRI Education* (2016).
7. Milford McQuirt, David Gagnon, and Rosemary Meyer, *Embracing Innovation: 2015–2016 Higher Education Industry Outlook Survey* (KPMG, 2015); Kenneth C. Green, *The 2016 Campus Computing Survey* (Encino, CA: Campus Computing Project, November 21, 2016); Christine Wolff-Eisenberg, Alisa B. Rod, and Roger C. Schonfeld, *Ithaka S+R US Faculty Survey 2015* (New York: Ithaka S+R, April 4, 2016); Jeffrey Pomerantz and D. Christopher Brooks, *ECAR Study of Faculty and Information Technology, 2017* (Louisville, CO: EDUCAUSE, October 13, 2017).
8. Wolff-Eisenberg, Rod, and Schonfeld, *Ithaka S+R US Faculty Survey 2015*.
9. Adam Gamoran, "Tracking and Inequality: New Directions for Research and Practice," WCER Working Paper No. 2009-6 (Madison: Wisconsin Center for Education Research, 2009), published in Michael W. Apple, Stephen J. Ball, and Luis Armand Gandin, eds., *The Routledge International Handbook of the Sociology of Education* (New York: Routledge, 2010).

© 2018 Martin Kurzweil and Mitchell Stevens.

The text of this article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.



Martin Kurzweil (martin.kurzweil@ithaka.org) is Director, Educational Transformation, for Ithaka S+R.

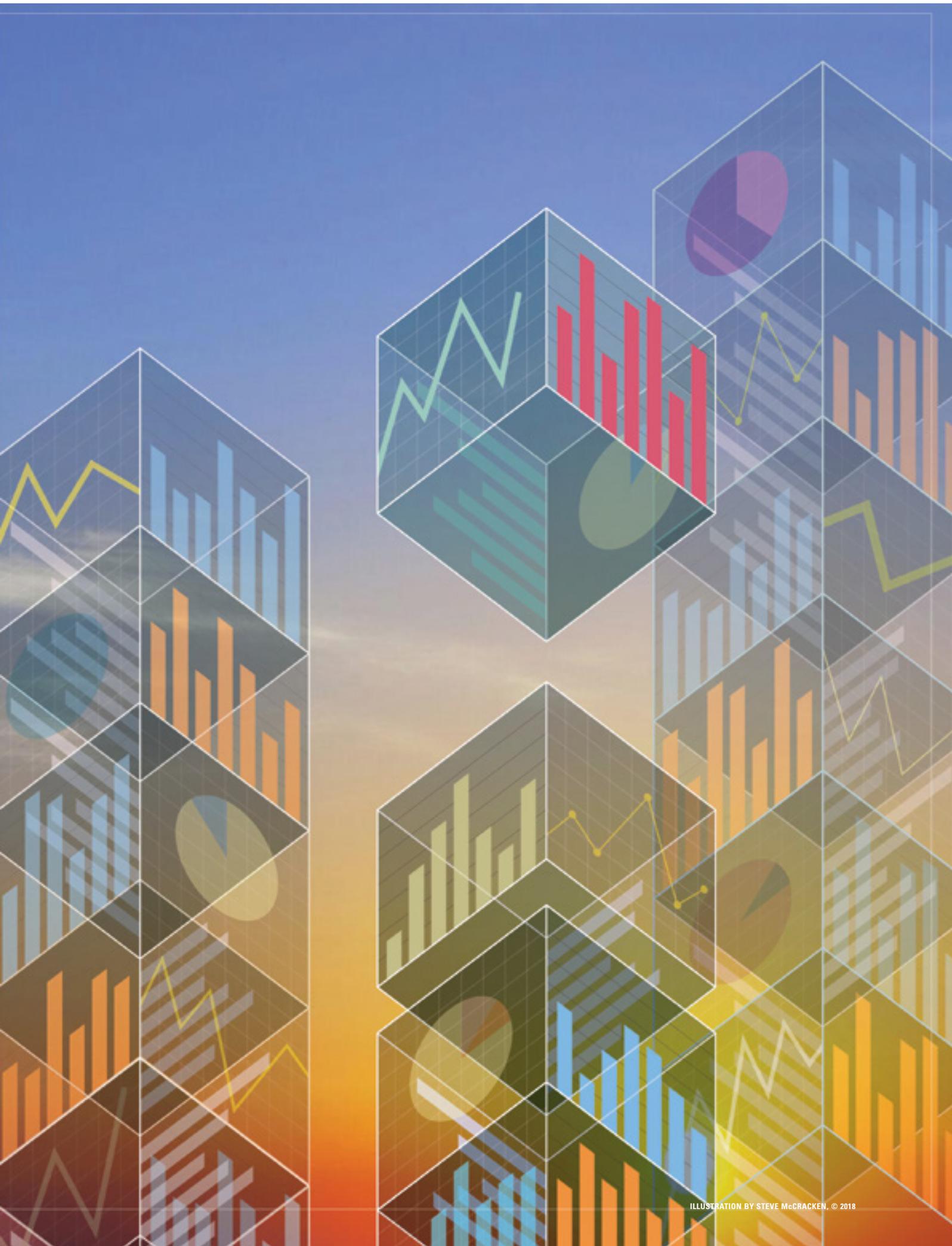


Mitchell Stevens (mitchell.stevens@stanford.edu) is Associate Professor and Director, Center for Advanced Research through Online Learning (CAROL), at Stanford University.

ARTIKEL 4

ARCHITECTING FOR LEARNING ANALYTICS: INNOVATING FOR SUSTAINABLE IMPACT

Simon J. Buckingham Shum and Timothy A. McKay



The focus of learning analytics is the application of analytics approaches to gain insight into educational data to improve teaching and learning. Learning analytics will undoubtedly be an advance if this provides new “power tools” for academic researchers—who have of course studied teaching and learning data in nondigital ways for decades. Such research is a necessary first step to validate the approaches. However, the true potential of learning analytics will be realized when we move from research to development and create human-computer systems that automate this analysis process—from data capture to visualization and recommendation—offering better (i.e., more timely, precise, and actionable) feedback to educators, students, instructional designers, and the other stakeholders who constitute the whole learning system.¹

The word *automation* conjures up many meanings. It is important to emphasize that automation does not necessarily mean that *diagnoses, decisions, and actions* are fully automated, taking human educators “out of the loop.” Automation may “simply” (it is still complex) make the process of data capture, cleaning, analysis, and visualization into a commodity service—a cycle that previously required skilled but scarce researchers or analysts. The responsibility for making sense of that feedback and acting on it can remain fully with the human student, educator, or analyst, or it can be shared (e.g., the analytics system may suggest areas of concern to help users prioritize their scarce attention or may recommend courses of action). It may also be that feedback and advice are fully scripted by expert instructors but are personalized at scale through tailored communication.²

To summarize, the potential of the “data revolution” in teaching and learning, just as in other sectors, is to create much more timely feedback loops for tracking the effectiveness of a complex system. In a field where feedback is already well established as a vital process for both students and educators,

the question is how this potential can be realized through effective human-computer systems.

The Innovation/Impact Tension

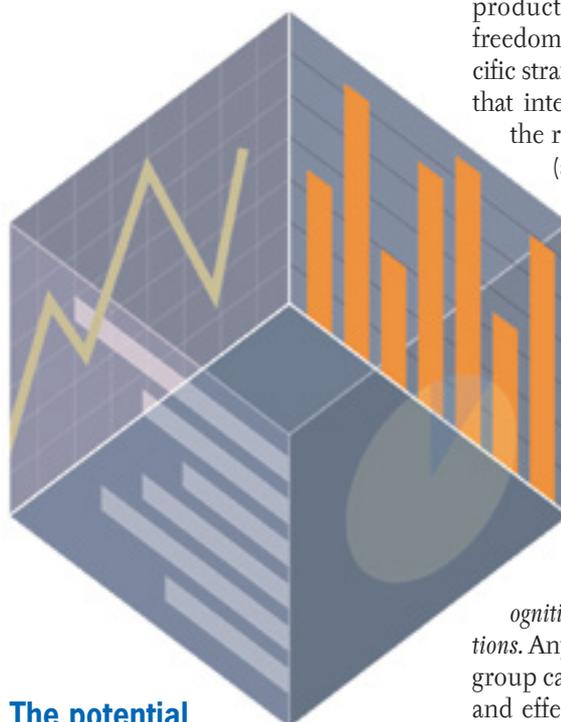
Despite hosting some of the brightest academics in data science, statistics, user interface design, and organizational innovation, a college or university may be far from innovating on these fronts when it comes to tracking, analyzing, and feeding back information to improve teaching and learning. Although paradoxical to an outsider, this apparent dysfunction is all too familiar to insiders: the incentives are

not there for academics to work on their own institution’s strategic teaching and learning problems. As a result, research-active analytics groups are generally not responsive to their institution’s analytics needs. Academics do not want to be branded with the dreaded badge of *service center*, which has connotations of not being research-worthy. Various tensions are in play here.

First, academics are under pressure to conduct novel research. They must engage in analytics innovation that is worthy of peer-reviewed publications and grants because they are making evidence-based claims, about data grounded in rigorous methodology, often using cutting-edge technology well beyond that in current products. They value their academic freedom, so they *choose* to engage in a specific strain of learning analytics research that interests them. Academics reserve the right not to try to solve “boring”

(albeit serious) data challenges in the institution, generally because they feel the challenges are mundane and/or don’t fit their research narrative and because they don’t want to be dictated to. Funds are spent on researchers, dissemination, and equipment. Time is spent on grant writing, training PhDs, and writing and reviewing papers.

Second, academics receive little recognition for developing scalable applications. Any competent academic research group can invent novel, well-grounded, and effective analytics at a small scale, but it reserves the right to move on to the next interesting challenge. Researchers receive little reward for validating more widely or for pushing an innovation through to mainstream deployment. They certainly don’t consider it their job to fix the institution’s broken data systems; that’s an IT or business intelligence (BI) job. In addition, crossing the chasm between innovation and infrastructure requires a suite of skills not often present in research groups—



The potential of the “data revolution” in teaching and learning, just as in other sectors, is to create much more timely feedback loops for tracking the effectiveness of a complex system.

skills that include professional software development, user experience and interface design, innovation advocacy, and behavioral science. It also requires a steady focus on the needs of the user community—what the Design-Based Implementation Research (DBIR) community calls the “persistent problems of practice.” Mainstream deployment of analytics tools also entails extensive cross-campus liaison to build commitment from other academics and units. This is a long-standing challenge for learning technology research and development (R&D).³

Finally, colleges and universities want to see an impact on students. This impact comes in many forms: increasing student learning, providing more efficient instructional delivery processes, addressing inequities in student outcomes, maintaining finances, and improving retention and graduation rates. These practical outcomes are sometimes perceived by academics as counterincentives.

Academics may be cautious about claiming causal connections or about being pushed to justify their work in financial terms. When the mainstream application of analytics requires scalable infrastructure that academics can’t deliver, what gets rolled out is commercial products. These may provide rudimentary dashboards that give analytics a bad name and that academics in turn don’t want to associate themselves with, either as end-users or researchers. Products typically target the larger mainstream, conventional markets rather than the future-focused bleeding edge of teaching and learning.

Organizational Architectures

Our focus here is on organizational architectures that a college or university’s leadership can consider in order to advance innovative analytics for its own mission and context. We are seeking to open a dialogue on organizational architectures and processes as a way to address educational challenges that often require systemic thinking and change. Such challenges may be

faced by many colleges and universities, opening up collaboration opportunities. Moreover, if the innovation-diffusion challenges facing one institution can be taken as a microcosm for the challenges facing the learning analytics field as a whole, organization-level insights may scale to consortia or more open networks.

Surveying the current landscape, we see three broad organizational models that are being used to deliver learning analytics. These three models are largely role-aligned: (1) the *IT Service Center* model (primarily professional services staff); (2) the *Faculty Academics* model (primarily faculty researchers); and (3) the hybrid *Innovation Center* model (a mix of professional services staff and faculty researchers).

To what extent can these three different organizational models deliver both production-grade services and innovation with sustainable impact? We will start by discussing the two “standard” models before moving on to the much less common third model.

The IT Service Center Model

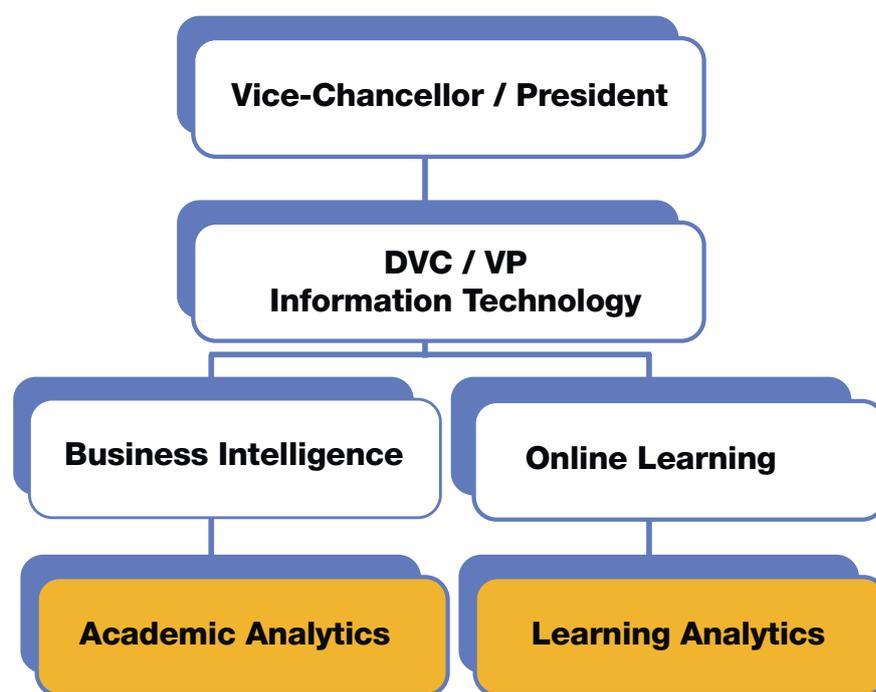
In this model, an IT service center delivers analytics from an enterprise platform. Examples include the following:

- The team delivering the learning management system (LMS) drives analytics provision, using/configuring the product’s dashboards for academics (and possibly students).
- The enterprise data warehouse, BI, or institutional research (IR) team provides analytics by integrating data from the LMS and other data sources.
- A team in the teaching and learning support center works with the above units to help academics make use of analytics.

Pros and Cons

➕ The center is equipped to deliver production-grade analytics services, with close to 24/7 uptime and system support, available to all or a majority of students and staff as a core platform integrated with other institutional systems. The analytics typically cover “academic

Archetypal Organization Chart, IT Service Center Model



analytics” (conventional student demographics, enrollment, and grades) and various forms of learning analytics (finer-grained, midcourse student progress and activity data).

➕ Staff will innovate within the scope of what products can do and how they integrate with existing infrastructure.

➕ End-users will typically be academics, because to date, most LMS and other products deliver reporting dashboards to help educators track student progress.

➕ However, generic LMS products are beginning to provide student-facing dashboards (although poorly grounded in the learning sciences⁴), and if niche products are deployed (e.g., an adaptive tutor for a specific topic), this may deliver feedback to students as well, since there is such a rich model of the curriculum and students’ mastery levels.

➕ If products permit report/dashboard customization, or facilitate data

export, and if there is coding/analytical capacity in the center for subsequent analysis and visualization, there is scope to provide bespoke user experiences.

➖ Staff typically work only with data that products can provide, delivered via predefined user interfaces. It is very unlikely that a participatory design model has helped end-users to shape a product, with the risk that analytics services are procured with limited consultation and are then poorly received.

➖ Staff rarely have expertise in educational research, user experience, learning design or advanced analytics techniques, so the scope for analytics innovation within the center is limited accordingly. Such expertise must come from other groups, and most IT service centers have little heritage of collaborating broadly in their work.

The Faculty Academics Model

In the second model, faculty academics (possibly partnering with an IT service center) conduct applied research. Examples include the following:

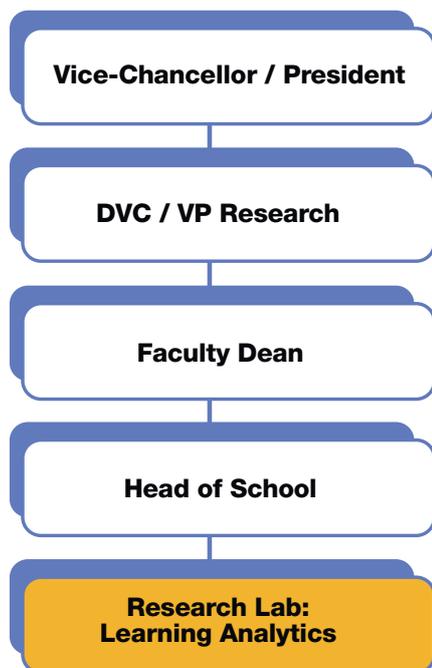
- Faculty academics develop innovative learning technologies (often externally funded) to support specific forms of learning not well supported by the LMS. This generates much richer data than typically comes from generic (i.e., discipline-agnostic) LMS products. This may serve as learning sciences research data, as well as data for feedback to educators and students.
- Faculty academics study the acceptance of analytics delivered by their institutional LMS and/or BI teams. They recruit early-adopter academics interested in piloting such analytics, study student responses, and perhaps engage other groups, such as student support teams.

Pros and Cons

➕ This model provides ample scope for radical innovation by academics, who can experiment with exotic sensors and advanced analytics techniques beyond current products.

➕ Empirical evidence is gathered about the adoption of vendor products,

Archetypal Organization Chart, Faculty Academics Model



typically identifying obstacles often relating to staff readiness, pedagogical practices that are incongruent with the analytics, or other organizational factors.

➕ Evidence-based claims are likely to be made with a high degree of rigor, meeting the standards of the human-research ethics board and, when published, of peer-reviewed research.

➖ The analytics are advanced but require corresponding researcher expertise to design, deploy, and maintain. This is often a scarce expertise, available on only a temporary basis.

➖ Since the analytics are advanced, they attract early-adopter academics in pilot studies, but if those academics move on from teaching a subject, the trials end.

➖ Typically, studies are relatively small scale, and/or use specific student cohorts, and/or are brief in duration (e.g., possible only until external funding expires).

➖ Researchers are less likely to consider users' needs in design ("customer discovery"), so although prototyping analytics are conceptually interesting, they either are unusable or fail to address widely recognized needs.

➖ Once the lead researcher or key project staff leave, there is nobody driving the vision for the analytics service. Vision is needed so that resources code maintenance, further grants, strategic visibility, and the all-important partnerships can sustain the innovation.

➖ Developing a combined software/pedagogical innovation into an enterprise-wide infrastructure is a development task, which is not often rewarded in research metrics and which requires skills that research groups lack.

The Innovation Center Model

In our third model, a hybrid, autonomous innovation center is created to service the entire institution. These innovation centers operate outside—but in close partnerships with—faculty academics, college/university IT/BI/LMS teams, and other stakeholders. Fewer institutions appear to be experimenting with this model, which we now introduce in the context of our own two examples:

■ An innovation center located outside the faculties, and autonomous from institutional IT/analytics, is staffed by research-active academics and data scientists, supported by professional

staff. (Connected Intelligence Centre, University of Technology Sydney)

■ An innovation center located outside the faculties, and autonomous from institutional IT/analytics, is dedicated to maturing and mainstreaming successful analytics innovations invented by academics, as well as innovating its own analytics services. (Digital Innovation Greenhouse, University of Michigan)

Connected Intelligence Centre, University of Technology Sydney

The Connected Intelligence Centre (CIC) at the University of Technology Sydney (UTS) is an innovation center charged with building the university's capacity to gain insights from analytics tools and techniques—spanning teaching and learning, research, and operational units.⁵

Staffing

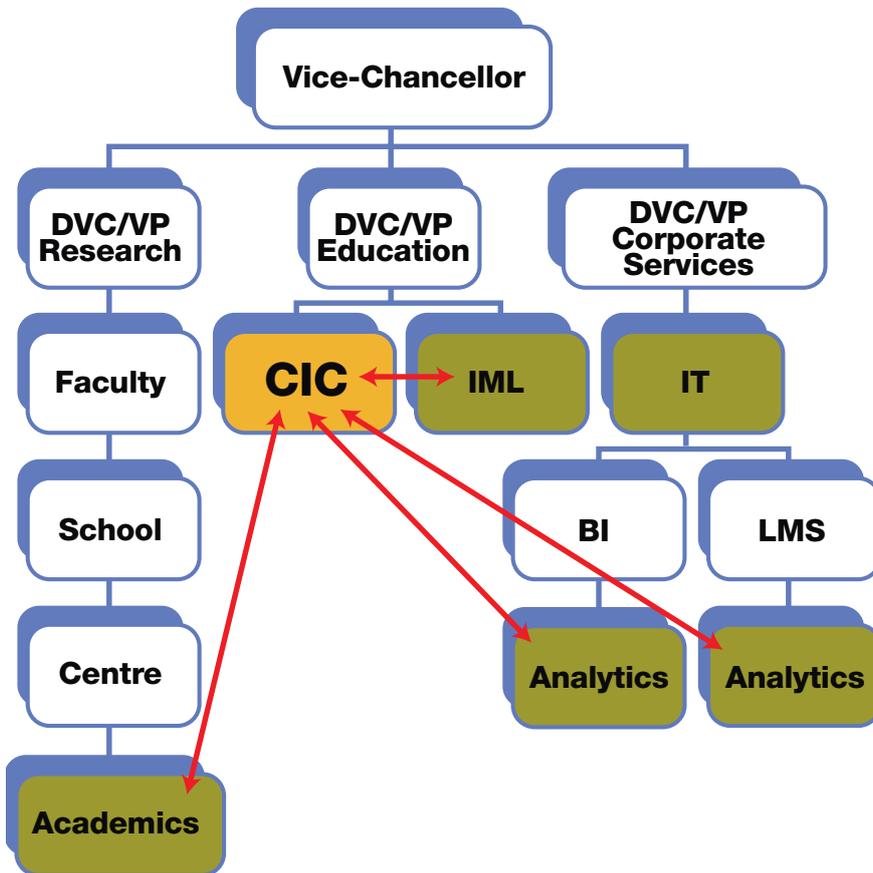
CIC is a small center of about twenty staff (not all full-time). In many ways it looks like an academic group, with researchers at all levels, from PhDs to professors, plus professional support staff.

However, while CIC may look like an academic research group—since its mission is to create research-inspired, sustainable innovation within UTS—the academics are recruited not only for their research capability but also for their team-working ability, transdisciplinary outlook, and communication skills. The academics are also talented educators: CIC designed and launched the Master of Data Science & Innovation (MDSI) degree program in 2015, which is coordinated and largely taught by CIC's academics. This was the only degree program at UTS running outside a faculty until 2017, when UTS launched its Faculty of Transdisciplinary Innovation. This faculty will take over the administration of the MDSI program in 2018, but the current staff will continue to teach it.

CIC has thus operated like a mini-faculty, running the MDSI and an elective subject in quantitative literacy



Connected Intelligence Centre, University of Technology Sydney



and critical thinking (“Arguments, Evidence & Intuition”) and launching its own Learning Analytics Doctoral Program in 2016. CIC’s Academic Board of Studies is drawn from across UTS to reflect its transdisciplinary nature. These teaching programs generated revenue that gave CIC the capability to grow beyond the baseline funding from the university.

CIC reports to the Deputy Vice-Chancellor for Education. The CIC director thus has the privilege of being able to talk to the Directors of IT, Teaching & Learning Innovation, Student Support, and Library to overcome obstacles to accessing data, get servers running, convene meetings, and more. This opens opportunities at a strategic level that a faculty-based team would typically not access.

Navigating from prototype to (1) small-scale pilots, to (2) pilots with several hundred students, and on to (3) a mainstream rollout to thousands of students will be critical transitions for CIC. The center has managed the first two transitions and this year will move to the third. Moreover, students are starting to

Academic Team	Professional Services Team
Director & Professor	Senior Manager
Professor	Course Manager
Associate Professor	Marketing/Communications Officer
Senior Lecturer x2	Administrative Assistant
Lecturer	Cloud Integration Specialist
Research Fellows x2	Web Developer
Data Scientist	Software Developer
PhD students x4	Undergraduate Interns x2
2–3 Visiting Researchers/Students	

expect certain services to be always on, and they complain if there is downtime. CIC is typically pushing the envelope of new technologies (e.g., Amazon Web Services), with encouragement from and in partnership with the IT Division (ITD). But this type of exploration must maintain security and also requires a constructive, friendly relationship with ITD staff, who are finding that servicing CIC's requirements is preparing them for what faculty members will likely be requesting in the future. For example, CIC and ITD co-funded a cloud specialist to assist in supporting MDSI students, in the expectation that before long, other degree programs will be calling on this specialist for cloud services.

Cultivating Research-Grade Innovation in a Non-Faculty Center

Attracting and retaining high-caliber researchers for this work requires the creation of an academically stimulating culture that provides the opportunities and trajectory needed by researchers at different career stages. This includes generating national and international visibility through research conferences, the chance to bid for competitive grants, and time to think and write.

However, running a hybrid academic/service operation requires controlling the innovation/impact tensions mentioned at the beginning of this article. For instance, when is there “enough” evidence, by academic research standards, to scale a prototype

that is exciting interest? CIC researchers understand that their work must be designed to add value into UTS, as the primary client, while working in close partnership with faculties and other client groups. Thus, all PhDs are in partnership with one or more faculty academics, providing authentic testbeds but introducing risk factors to a doctoral plan. Software is breaking new ground but is designed as early as possible in consultation with ITD staff, who are not used to any other unit running 24/7 student-facing systems. Analogous to the R&D centers in companies, the CIC mission is to enhance the efficiency of current services, as well as prototype future services, but it cannot chase “blue sky” ideas simply out of curiosity. CIC was created to build sustainable capacity in UTS staff and students to use data science and data analytics tools.

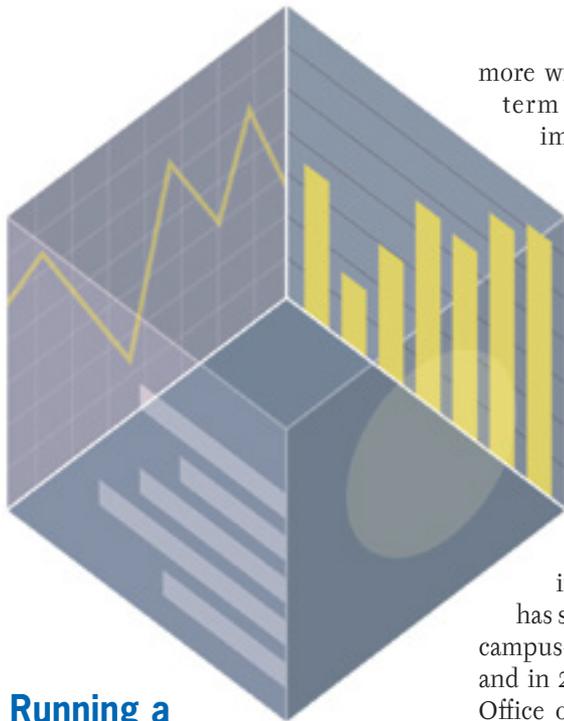
Digital Innovation Greenhouse, University of Michigan

Learning analytics activity at the University of Michigan (UM) began to emerge as a coherent theme of work in 2011, with SLAM: Student Learning and Analytics at Michigan (<http://www.crlt.umich.edu/SLAM>). This seminar series combined a forum for exchange of ideas and information among on-campus faculty and staff with an opportunity to connect with external speakers. In 2012, interest in SLAM prompted the UM provost to launch a three-year, faculty-led Learning Analytics Task Force (LATF),

charged with promoting expanded use of data on campus and funding a series of learning analytics projects. These projects took two forms: analysis of data to inform policy and practice; and invention of tools designed to put data to work in support of teaching and learning. Several of these tools entered the classroom as pilot projects and attracted external research support.⁶

The Digital Innovation Greenhouse (DIG) was proposed in 2014 to solve a recurring problem that LATF projects had encountered. Faculty innovators and their research teams had designed, developed, and tested analytics-driven tools designed to improve teaching and learning on campus. These innovations were typically tested in the researchers' home environments, often in courses they had been assigned to teach. When news of their existence and impact became known on campus, interest in expansion to other areas emerged. The founding research groups, while often interested in this expansion, lacked the resources, skills, and/or reward systems needed to mature a seedling innovation into an element of the campus infrastructure.

Several of these projects attempted to grow to scale by handing off their tools to the campus-wide Information Technologies Service (ITS) unit. ITS as an organization is very good at staging and supporting *mature* software systems. Unfortunately, the unit's skills and organizational approaches are ill-matched to adopting the loose, rapid, “duct tape and chewing gum” development methods of researchers. Finally, to take advantage of the vibrant innovation community emerging on campus, DIG was created as a pilot in 2015. Its mission was to adopt a series of existing digital-engagement innovations from the research labs they had outgrown, carry them across the innovation “valley of death,” and deliver them to ITS as infrastructural tools that could be used campus-wide. In doing this, DIG achieved both the immediate goal of making existing research tools much



Running a hybrid academic/service operation requires controlling innovation/impact tensions.

more widely available and the longer-term goal of demonstrating the importance of this greenhouse approach to the development of 21st-century digital-engagement tools.

To create a home for DIG, the university turned to the recently formed Office of Digital Education and Innovation. This unit, which reports to the Vice Provost for Academic Innovation, was created in 2013, initially to provide a home for UM's newly emerging involvement in MOOCs. It has since become the focal point for campus-wide educational R&D efforts, and in 2016 the unit was renamed the Office of Academic Innovation (OAI). Today, OAI is home to teams working on three major themes: designing, developing, delivering, and experimenting with online and hybrid education; growing analytics-driven educational innovations to scale; and promoting gameful design of educational experiences. These teams often work in collaboration

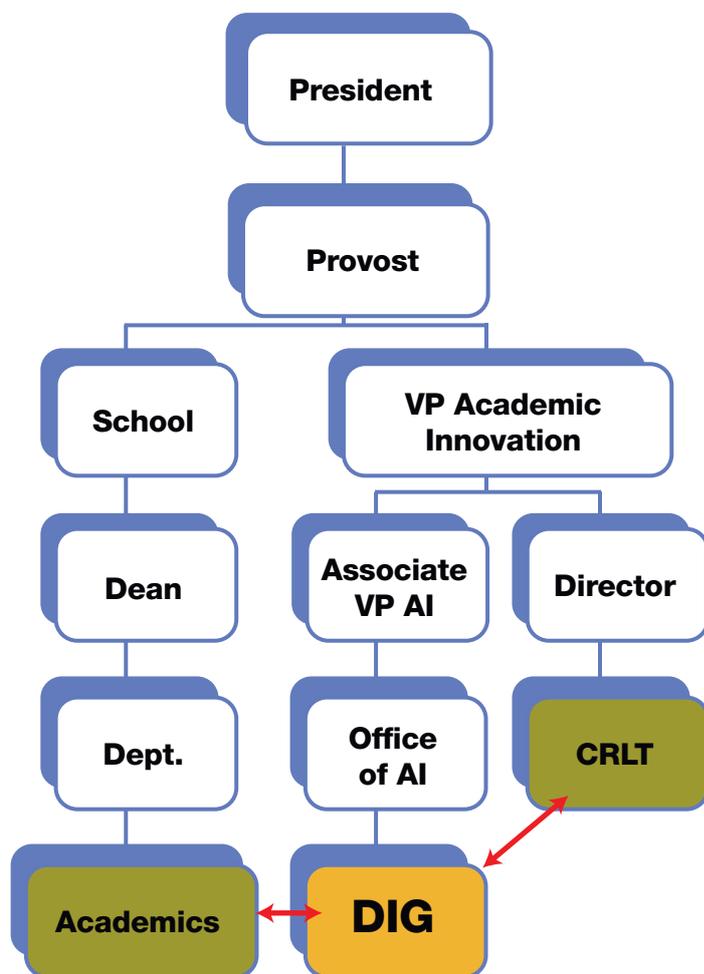
with the campus's long-standing Center for Research on Learning and Teaching (CRLT), especially on projects that focus on residential learners.

Staffing

DIG projects advance in a highly collaborative environment including at least four elements:

1. *Faculty champions and their research teams.* Each project enters DIG with a principal advocate drawn from the faculty, usually accompanied by members of the research team. DIG is now also exploring how to support innovations backed by students and staff. Innovators help to drive the vision for each project. In addition, they regularly lead efforts to research innovation design and impact, often with external support.
2. *The DIG team of software developers, user experience and interface designers, behavioral scientists, and innovation advocates.* This group provides the professional skills and reward system needed to mature innovations technically, in close interaction with the expanding

Digital Innovation Greenhouse, University of Michigan



user community. Full-time DIG staff are supported by an array of graduate and undergraduate student fellows who provide additional effort, a fresh design view, and close contact with DIG tools.

3. *An expanding campus community of users, from early adopters to those hoping to use DIG tools as infrastructure.* Continuous, intensive interaction with this community of faculty, staff, and students is essential to the success of DIG

Academic Team	DIG Staff
Faculty Director	Lead Software Developers x4
AI Operations Director	Software Developers x4
Faculty Champions x7	Lead Behavioral Scientist
	Behavioral Scientist
	Lead Innovation Advocate
	Data Scientist
	UX/UI Designers x3
	Student Fellows x10–15

tools. The DIG team forms the liaison between faculty innovators and this community, putting in the time and effort required to establish and maintain deep and mutually beneficial relationships.

4. *The UM ITS organization.* The DIG team interacts with ITS both to obtain the required infrastructure support (e.g., servers, single-sign-in authentication, access to data) and to ensure that the development cycle for DIG tools stays within campus ITS structures.

From these four groups, only the core DIG team is officially employed within the OAI. Originally established as a team of three senior software developers, the DIG team now includes a continuously growing, full-time staff of fifteen.

Funds for this staff come from three sources: a UM Third Century Initiative grant that provided for the launch of DIG; additional investments made from the university through the OAI budget; and project funds obtained from various sources (NSF and other grants). Excellent administrative support comes from OAI and has proven essential to creating and maintaining a nimble, responsive organization within an occasionally leisurely and conservative campus community.

Engaging Faculty in Academic Innovation

DIG was established within the OAI as a staff unit, without official appointments for any members of the faculty. The model is predicated on close collaboration between faculty champions, their research groups, DIG staff, and the broader community of educational practitioners across campus. Although this model has worked well, it has created significant tensions for some faculty members, particularly those for whom education is not a research focus. They find themselves doing this work *in addition to* their existing responsibilities for research, teaching, and service. Championing an analytics-driven innovation as it expands across campus is not a small

task, even when receiving the extensive professional support the DIG team provides. Indeed, the rapid pace of development and expansion that the DIG team can deliver makes the job of the faculty champion *more* challenging. For this reason, DIG has come to think of the activities taking place in the OAI as educational R&D—applied research aimed at reinventing higher education for an information age. With this lens, DIG and OAI can be seen as a research institute, akin to UM’s long-standing Institute for Social Research or its Life Sciences Institute. These units provide associated faculty members with appointments of varying kinds, ranging from 100 percent salaried positions to 0 percent affiliated faculty status. In 2018, OAI will explore offering appointments of this kind to ensure that faculty champions have the support they need to accelerate innovations to scale.

Reflections on the Innovation Center Model

CIC and DIG started at different points and in response to different drivers. CIC has been home to research-active academics from its launch, running its own master’s and PhD programs, but it now needs to build developer capacity as demand grows for its analytics tools. In contrast, DIG launched with technical staff to scale innovations from existing faculty academics, but it is now considering new models for engaging academics. Today, DIG and CIC seem to be moving toward common ground: *they are both autonomous centers reporting to a VP/DVC, tasked with innovating data and analytics infrastructure to tackle strategic teaching and learning challenges, while working in partnership with faculties, teacher professional development, and IT services.*

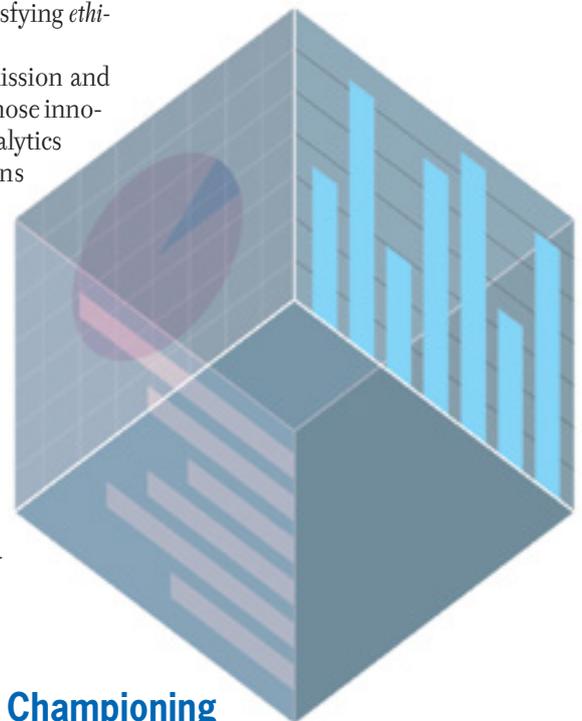
At this point, it is too early to declare these centers to be the blueprint for success; however, we can summarize the CIC and DIG hallmarks:

- Having the center report directly to a senior leader at the VP/DVC level provides the strategic positioning

needed to effect change around, for instance, access to data and the provisioning of campus-wide analytics services or innovations.

- The center either incorporates academics and research students (CIC) or works very closely with them in their faculties (DIG). This enables the center to conduct *research-inspired* innovation, in analytics that are *grounded in educators’ and students’ needs*, leading to *evidence-based claims* and satisfying *ethical standards*.
- The center has the core mission and the capability to translate those innovative ideas into robust analytics services, to specifications approved by the IT division (e.g., security, architecture), integrating with the institution’s enterprise infrastructure. As CIC’s prototypes begin the transition to more widely scaled internal “products,” it is recognizing the importance of having a DIG-scale software design, development, evaluation, and communications team.
- The center can supplement its institutional funding with other income including externally funded joint projects with faculty, internal strategic grants with faculty, student fees from formal teaching programs, and faculty buyouts of teaching time from the center’s academics.
- By creating a campus-wide focal point for the development of analytics-driven tools, the center can create significant efficiencies of design and synergies of operation. For example, for access to campus data, the DIG team has developed a shared, extensible framework that can be used by all tools, preventing the need to re-create this framework

for each tool. It has also worked to coordinate a toolkit of elements generally useful for education at scale in large foundational courses, increasing the adoption of the full suite of tools. Similarly CIC, having better understood the diverse needs of academics, has re-architected the writing-feedback tool to expand the range of services it can offer.



Championing an analytics-driven innovation as it expands across campus is not a small task, even when receiving extensive professional support.

- If (like CIC) the center is hosting its own academics and doctoral students, it looks very much like faculty research groups. Therefore, the center must be very clear that it is not simply doing research “business as usual” but is working on data challenges faced by the institution and is adding value through generic/customized analytics services for academics, students, and professional

business units from across the campus. All of the center's innovations must be conducted in collaboration with institutional "clients."

- If (like DIG) the center houses only staff and collaborates with faculty who are offered no official role in the organization, it may be placing a significant burden on both faculty innovators and early adopters from its community of practice. The reward systems for faculty members at research universities are finely tuned, leaving little allowance for the recognition of new forms of activity. Unless these innovative types of activities are accepted as either research or service by faculty members' home departments, the faculty must contribute effort without recompense.
- The center complements but does not duplicate the work of the campus division that is leading academic professional development in teaching and learning. These well-established

It is perhaps not a coincidence that both CIC and DIG have converged independently on common strategically important teaching and learning challenges.

teaching and learning centers bring expertise in pedagogy and academic development but do not have the capacity to develop the college or university's learning analytics capability. For example, CIC has worked very fruitfully with the UTS Institute for Interactive Media in Learning, whose *Academic Learning and Literacy* specialists have advised on the CIC automated writing-feedback tool and have coauthored research papers. DIG and its leaders are collaborating closely with UM's Center for Research on Learning and Teaching on the launch of a Foundational Course Initiative and with the Sweetland Center for Writing on the M-Write project.

- The center complements but does not duplicate the work of the IT division. Colleges and universities already have an established LMS and BI unit, but the center's focus is on modes of learning, forms of data, and analytics user groups not covered by such units. The center must liaise closely with other units and divisions that are concerned with data governance and may be developing novel services to test frameworks. Similarly, the center may well be the first non-IT group provisioning 24/7 student- or staff-facing software applications, requiring solid IT partnership to ensure security, network services, responsive maintenance, and so forth. The IT collaboration should be mutually beneficial. For instance, CIC and DIG have provided IT staff with secondment opportunities to work in an environ-

ment more akin to a startup, on different projects developing new skills.

- The center starts to develop platforms that advance the work of faculty researchers, in domains other than teaching and learning. For example, DIG's ECoach tool has been used as a platform for conducting experiments in social psychology, online engagement, and the visual display of quantitative information. Similarly, CIC's text analytics platform, social media activity aggregator, and multimodal collaboration analytics can be generalized to non-educational contexts, to benefit other UTS academics.

Finally, it is perhaps not a coincidence that both CIC and DIG have converged independently on common strategically important teaching and learning challenges. We are excited about the potential for collaboration around the role that analytics can play:

- *Personalized messaging enabling feedback at scale.* Although all the educational evidence points to the importance of timely, actionable, and personalized feedback for effective learning, providing this feedback is particularly challenging in large classes. Both universities have developed platforms that permit academics to craft coaching messages to students contingent on their progress (e.g., over a week): UM developed the earlier noted ECoach platform, whereas UTS has been running its own personalized messaging platform for a decade and is a partner in the Australian national OnTask project developing an open-source tool. The role of the analytics is to analyze students' activity profiles from multiple data sources and compile the feedback into a personalized email, with a growing evidence base that this feedback is well received by students and improves their outcomes.⁷
- *Text analytics for student writing feedback.* Critical, persuasive, reflective, academic writing is hard to learn,



hard to teach, and hard to give rapid feedback on. Both universities are developing applications of Natural Language Processing to give instant formative feedback (not summative grades) on students' drafts as a way to encourage revision and reflection. These applications require scaleable text analytics platforms, tuned to the specific writing features that will enable actionable feedback. Common to both efforts (the UTS Academic Writing Analytics tool and research program and the UM M-Write initiative) is the recognition that such tools are most effective when aligned with good "learning design"—that is, congruent with the curriculum, assignment activities, and grading rubrics.

- **Human-centered analytics.** Software design has gradually shifted from being technology-driven to human-centered, and it is no coincidence that both universities have academic professionals from human-computer interaction on their teams, as well as designers who care about the user experience. The human dimensions of learning analytics are diverse, from deciding what the overall user needs are, to designing the user interface and evaluating how users engage, to considering the ethics of data, algorithms, and visualizations. Finding ways to involve stakeholders early on, through participatory design methods, is critical to achieving these aspirations.⁸

Conclusion and Invitation

We have introduced the dilemma that higher education institutions face: how can they architect themselves organizationally to both innovate learning analytics (the traditional province of faculty researchers) and see the learning analytics deliver sustainable impact through mainstreamed services (the traditional province of LMS/IT units)—which, furthermore, are rigorously evaluated (a conventional faculty role)? We have documented two examples of a new model: the hybrid learning analytics innovation center, reporting to senior leadership and working in close partnership with faculties and service centers.

The hallmarks of our centers are from just two exemplars, and we recognize that these do not reflect the diversity of institutional contexts. Please treat this as an invitation to respond. We offer these reflections as a conversation opener, and we welcome feedback, including examples of other organizational models tackling this challenge. ■

Notes

1. Simon Knight and Simon Buckingham Shum, "Theory and Learning Analytics," chapter 1, and Ruth Crick, "Learning Analytics: Layers, Loops and Processes in a Virtual Learning Infrastructure," chapter 25, both in Charles Lang, George Siemens, Alyssa Wise, and Dragan Gašević, eds., *Handbook of Learning Analytics* (Beaumont, Alberta: Society for Learning Analytics Research, 2017).
2. Madeline Huberth, Patricia Chen, Jared Tritz, and Timothy A. McKay, "Computer-Tailored Student Support in Introductory Physics," *PLoS ONE* 10, no. 9 (September 9, 2015).
3. Barry J. Fishman, William R. Penuel, Anna-Ruth Allen, and Britte Haugan Cheng, eds., *Design-Based Implementation Research: Theories, Methods, and Exemplars*, National Society for the Study of Education Yearbook (New York: Teachers College, Columbia University, 2013); Eileen Scanlon et al., *Beyond Prototypes: Enabling Innovation in Technology-Enhanced Learning* (London: Technology Enhanced Learning Research Programme, [2013]).
4. Ioana Jivet, Maren Scheffel, Marcus Specht, and Hendrik Drachler, "License to Evaluate: Preparing Learning Analytics Dashboards for Educational Practice" (preprint), in *Proceedings LAK18: International Conference on Learning Analytics and Knowledge* (Sydney, March 5–9, 2018).
5. For more details on the organizational history leading to the creation of CIC and the focus of its analytics R&D, see: Rebecca Ferguson et al., "Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption," *Journal of Learning Analytics* 1, no. 3 (2014); Riina Vuorikari and Jonatan Castaño Muñoz, eds., *Research Evidence on the Use of Learning Analytics: Implications for Education Policy*, European Commission Joint Research Centre (JRC) Science for Policy Report (2016).
6. For more details on the organizational history leading to the creation of DIG and its focus on growing analytics-driven innovations to scale, see Steven Lonn, Timothy A. McKay, Stephanie D. Teasley, "Cultivating Institutional Capacities for Learning Analytics," *New Directions for Higher Education*, no. 179 (Fall 2017).
7. Mary C. Wright et al., "Better Than Expected: Using Learning Analytics to Promote Student Success in Gateway Science," *Change: The Magazine of Higher Learning* 46, no. 1 (2014); Huberth et al., "Computer-Tailored Student Support in Introductory Physics."
8. UTS and UM work relating to human-centered analytics includes the following: Michael Geoffrey Brown, R. Matthew DeMonbrun, and Stephanie D. Teasley, "Don't Call It a Comeback: Academic Recovery and the Timing of Educational Technology Adoption," *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (New York: ACM, 2017); Simon Knight et al., "Designing Academic Writing Analytics for Civil Law Student Self-Assessment," *International Journal of Artificial Intelligence in Education* 28, no. 1 (March 2018); Steven Lonn, Stephen J. Aguilar, and Stephanie D. Teasley, "Investigating Student Motivation in the Context of a Learning Analytics Intervention during a Summer Bridge Program," *Computers in Human Behavior* 47 (June 2015); Carlos G. Prieto-Alvarez et al., "From Co-Design to Learning Analytics Deployments: Bringing together Educators, Students, Researchers and Developers," in Jason M. Lodge, Jared Cooney Horvath, and Linda Corrin, eds., *From Data and Analytics to the Classroom: Translating Learning Analytics for Teachers* (London: Routledge, in press, 2018).

© 2018 Simon J. Buckingham Shum and Timothy A. McKay. The text of this work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.



Simon J. Buckingham Shum is Director of the Connected Intelligence Centre at the University of Technology Sydney.



Timothy A. McKay is Founder of the Digital Innovation Greenhouse at the University of Michigan.

ARTIKEL 5

DATA-DRIVEN IT STRATEGIC PLANNING: A FRAMEWORK FOR ANALYSIS'

Jerrold Grochow

Why IT Matters to Higher Education

EDUCAUSEreview

Data-Driven IT Strategic Planning: A Framework for Analysis

by **Jerrold Grochow** ⌚ Monday, June 4, 2018 ⭐ Editors' Pick

A new framework offers a system for defining, collecting, analyzing, and organizing data for IT strategic planning.



Institution-wide IT strategy is #3 on EDUCAUSE's list of **Top 10 IT issues for 2018**. IT strategic planning is a regular topic at EDUCAUSE conferences and in EDUCAUSE

publications. The 2016 EDUCAUSE Center for Analysis and Research working group report ***New Approaches to Higher Education IT Strategic Planning*** included recommendations to "identify, review, and discuss both internal and external change drivers" and "collect data on technology needs and desires and incorporate those into the plan." Further, data-driven IT strategic planning was the **EDUCAUSE Strategic Planning constituent group's theme**  at the **2016 EDUCAUSE Annual Conference**.

While these activities all aim to improve the IT strategic planning process, the idea of true data-driven strategic planning is still discussed only at the general level without a detailed implementation framework. This article provides such a framework.

Strategic planning projects require significant upfront investment — including for gathering materials, developing agendas, and surveying the community — to support project participants in their efforts. Key to that support is developing appropriate, carefully analyzed, and thoughtfully presented data to help produce a plan that best serves the community. Those data provided must

- portray the current state of systems and technology,
- highlight important internal and external trends and issues in how IT services are used and provisioned,
- demonstrate how IT impacts the institution's administration and its research and teaching missions, and
- reveal questions that must be resolved to establish the university's best IT strategic direction.

Strategic planning that focuses on and effectively uses all of these types of data is "data-driven strategic planning."

Strategic Planning Overview

A *strategic plan* is a framework for describing the organization's key processes, people, and technologies at a specific time in the future (typically 3–5 years) and how the organization can reach that future, given its current processes, people, and technologies. To ensure that the plan serves its purpose, the institution — as embodied in its strategic planning team — must fully understand its current state, as

well as the factors that will impact and drive it into the desired future state. Figure 1 shows the major components of the strategic planning process.



Figure 1. Components of the strategic planning process

"If you don't know where you are going, any road will get you there," paraphrases the Cheshire Cat from Lewis Carroll's *Alice's Adventures in Wonderland*. And if you don't know where you are now, you won't be able to find the right road to where you want to go.

So, the first step in any strategic planning process is to determine where you are now, or the *current state*. Once you know that, you can begin to look at factors that are driving you to the future and begin discussing how those drivers will (or should) impact the institution.

For example:

- Will changes in teaching pedagogy (such as the flipped classroom) significantly affect the IT support infrastructure for faculty and students?
- Will new research areas such as genomics or cosmology require research computing capability beyond what the university currently possesses?
- Will the move to cloud-based computing change the basis for IT budgeting across university departments?
- Will changes in the federal funding picture alter the focus of funding sources and thereby impact IT strategy?

Issues and drivers such as these will impact an institution's IT strategic plan. And they all require data if they are to be discussed in detail sufficient to produce that plan.

Collecting and Developing Data

Strategic planning project teams consist of committee members and resource staff providing data and other materials to the committee. The strategic planning resource staff must collect a wide variety of data and make those data available to the strategic planning committee to support each component of the strategic planning process (see figure 1). Data must be collected and developed to highlight trends, raise strategic issues (including resource concerns), and distinguish the needs of different community stakeholders (such as students, faculty, administration, and alumni).

Data must provide details as well as general statistics; merely identifying a general lack of funding (a perennial resource concern!) doesn't provide much basis for discussing the issues. Likewise, simply identifying an absence of qualified people to perform future tasks or an absence of currently acceptable technology is also too general.

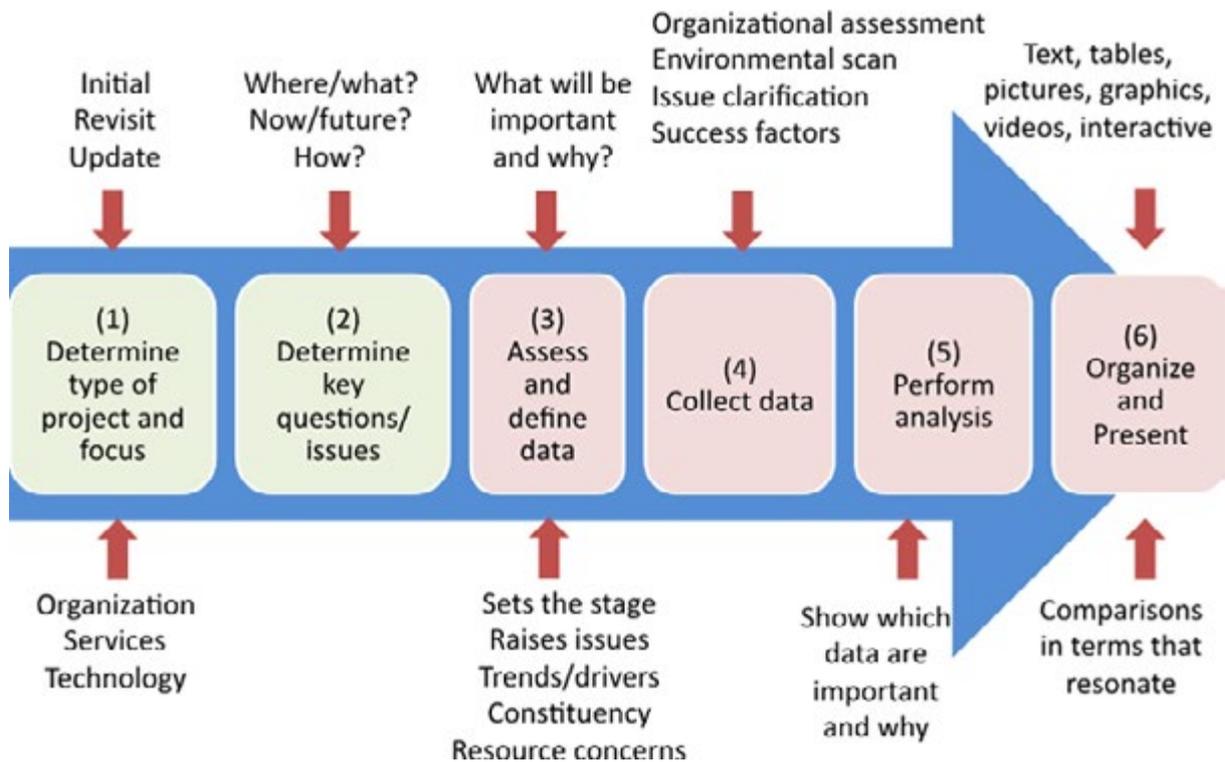
With the move to cloud computing, for example, IT departments are universally concerned about the different types of skills their staff members will need in order to work in an external service-driven environment rather than a locally supplied environment. Coupling this with the specific number of personnel available and the number expected to be needed, as well as possible training approaches, allows for a richer discussion of both people and funding.

On the technology front, data must show the technologies currently available, as well as what can reasonably be expected in the future. For example, when the **Large Hadron Collider** [↗] (LHC) was being planned, physicists designed their distributed computing environment in anticipation of computer networking capability levels of multiple gigabits per second, which no university had yet implemented. Data showing the current and expected increase in network bandwidth (and the reduction in costs) allowed physicists to work with university IT departments to ensure that LHC expectations could be met, offering an example of effective data-driven IT strategic planning.

Determining the range and scope of data that will allow institutions to develop IT strategic plans for all constituencies will itself require planning.

The Data-Driven Planning Framework

The data-driven strategic planning framework consists of two parts; the first part is a sequence of tasks to be executed during data collection and presentation (see figure 2).



© Jerrold M. Grochow / Creative Commons License BY-SA-NC

Figure 2. Data-driven IT strategic planning framework

Prior to assessing which data should be collected, the strategic planning committee should

- determine the strategic planning project's type and focus (task 1 in figure 2), and
- determine key issues and questions arising from that focus (task 2).

Strategic planning projects range from totally new activities — such as an organization's initial strategic plan — to relatively straightforward annual or biennial updates. In between are planning projects that revisit an existing strategic plan because something significant has changed; examples include a major shift in

educational policy, such as focusing on online education for remote students or a shift from in-house to cloud computing. Such projects require more than a simple update and often need significant planning and data collection. The "Data Definition" sidebar discusses how to define the data you want to collect.

Data Definition

A fundamental consideration of data-driven strategic planning is ensuring that you carefully define each type of data you want to collect. If you cannot specifically define these data, you cannot appropriately analyze or draw inferences from them.

For example, data about applicants' place of origin may seem clear, but do your data show where applicants were born, where they spent most of their lives, where they currently reside, or which high schools they attended? Data elements that are seemingly straightforward nevertheless require clear definition, as do data that may not come from a limited answer set. Imagine the difficulty in (and importance of) defining data elements about faculty needs for high-performance computing: How do you define *need*? (If it costs a million dollars an hour, is it still a need?) What is *high performance*? (Yesterday's high-performance computers are hardly comparable to today's.) And, for that matter, how do you define *faculty*?

Some projects have a narrow scope, such as those focusing on organizational resources and culture (severe budget cutbacks, for instance); specific IT service delivery (such as moving to the cloud); or a specific technology (incorporating artificial intelligence). Other projects have a broader scope that covers all areas. Understanding the project's focus and scope is important in determining the types of data you need to collect to support it. If the project's purpose is to examine specific issues, such as the impact of funding decreases, listing those issues can guide the initial data collection.

Once these two preliminary activities have been accomplished, the data-driven IT strategic planning framework focuses on the next four tasks in figure 2:

- Determining which data to collect to assess the current state, raise issues (including resource concerns and issues affecting a particular constituency), and identify trends and drivers that might affect the future (task 3)
- Developing a systematic approach to data collection (task 4)

- Analyzing the data (task 5)
- Organizing and presenting the data (task 6)

Supporting the entire strategic planning process means collecting data to help in each of the key strategic planning tasks: understanding the current state, understanding drivers of the future state, analyzing issues, and specifying the future state. It also means collecting data to address the issues of multiple constituencies, from students, faculty, researchers, and staff to alumni and applicants. Presenting data that ignore any of these tasks or stakeholders leaves the planning process open to question and perhaps criticism. Using the second part of the framework — the strategic planning data-organizing matrix described later — can help you track different data categories and purposes and expose areas where you might look for additional data.

Sometimes, strategic planning committee members do not have a complete picture of IT's current state on campus. A principle source of these data is current service usage data, such as the number of computers on campus; Wi-Fi usage (by time, location, and type of user); the number of different software packages supported and used; the amount of space devoted to computing resources, including networks; the number of data elements in the data warehouse; and help desk interactions. The list of data items that demonstrate service usage can get quite long. These numbers can be augmented with financial data showing the total cost and the cost per unit for as many of these services as possible.

Current service usage data provide the *volume* picture; to understand the *performance* picture, you will need other data, such as outage statistics, and data forms, such as user satisfaction surveys. Surveys can be augmented with user testimonials (good and bad) and other qualitative and descriptive performance data.

If specific issues have already been identified (either by the IT organization or in the charge to the strategic planning team), you should collect data to support discussion and analysis of those issues. For example, if a significant cut in funding is expected in the next several years, you should present data that show the impact of cuts in various services, as well as whether cutting central IT budgets will simply result in additional departmental spending.

Presenting raw data, even well-defined raw data, is clearly less valuable than providing data with analysis. Analysis can take many forms, ranging from simple

statistics to sophisticated factor analysis to qualitative discussions that introduce related data. Say, for example, that moving a certain type of research computing to centralized equipment will save \$600,000 per year. Faculty may initially focus on the loss of control that would be a consequence of this move, but they might take a second look if you also give them data showing that \$600,000 could support 10 additional graduate research assistants. Statistics and charts that bring in related data (such as the cost of a graduate research assistant) provide significant value.

Further, in the student demographics example above, even if you collect data on all the various types of origin, will you have information on student needs for IT training and capabilities when they arrive on campus? You may need to combine data on the high schools that students are coming from with socioeconomic and academic-standards data to get a more complete picture. Although this is a task that the strategic planning committee will not necessarily have the time or resources to take on, the strategic planning resource team should.

In deciding the type and level of analysis to perform, consider how data will be used (either in describing the current state, analyzing issues, or determining the future state or action). How you present the data analysis is equally important. Presenting data in contexts and forms that are meaningful to planning-process participants will help them understand the data's strategic significance.

Graphs, charts, and other visual forms add value to tables of numbers (which should always be made available as well). It is the strategic planning resource team's job to provide data in a variety of formats, without bias and without obscuring the original data (see the "Data Quality" sidebar). Members of the strategic planning committee also need to know why the resource team thinks the data are important, but they must make up their own minds as to whether they agree.

Data Quality

Data presented to decision makers should always be scrubbed to be as *accurate* and *free of bias* as possible (two measures of data quality). Although high-quality data don't always lead to high-quality decisions, poor-quality data often lead to poor decisions. Data quality should be thoroughly addressed, particularly when the person or group collecting those data has a vested interest in the decisions that may be made using them.

Data quality is a question not just of whether the raw data are accurate or precise but also of whether they are appropriately presented in context and without bias. For example, you may have determined the accurate dollar amount for total IT expenditures for each of the past five years, but you need context to discuss whether IT expenditures were too low or too high. One way to provide context is to consider how IT expenditures supporting research, teaching, and administration have changed in each area over the five-year period. Showing how your institution's spending compares with that of peers on a per capita basis can provide additional information. Also, adding metrics such as the relative research intensity of the institutions should make the data even more useful and thus of higher quality for decision makers.

Data Classification

How do you know if you have covered all the bases in collecting data for the strategic planning committee? Have you missed any important data elements or sets that will be critical to committee deliberations? The data-driven IT strategic planning approach classifies data types in various ways to ensure there are no holes in your data collection effort. To ensure you have the data you need, you can use the second part of the framework: the strategic planning data organizing matrix (see figure 3).

Type of project (Initial, Revisit, Update): _____ Project focus (Org, Service, Tech): _____

Data:	Strategic Questions:	Strategic Importance:	Data Category:	Constituency Interest:	Source:	What Kind of Data / How to Present:
Name of data element	Where we are What drives us Where we want to be How to get there	Sets the stage Raises issues Highlights trends Constituency Resource concern	Financial Demographic Operational Performance Environmental (Issue focused)	Faculty/ Researchers/ Students/ Administrators /Other staff	Primary: - Survey/ Quest - Interviews Secondary: - Instrument - Public - Other	Text/ Numbers/ Pictures Table/Graph/ Chart/Picture/ Video/ Interactive

© Jerrold M. Grochow / Creative Commons License BY-SA-NC

Figure 3. Strategic planning data organizing matrix

Data characteristics to consider include:

- The strategic question the data support (current state, future driver, future state, and implementation)
- The data's strategic importance (current state, raises issues, highlights trends, distinguishes constituencies, presents resource concern)
- The data category (financial, demographic, operational, performance-related, external environment, descriptive)
- The constituency most interested in the data (faculty, students, administration, etc.)
- The data's source, including primary (collected for the current study), secondary (previously collected), and external (public databases)

You can create a spreadsheet with data items as rows and these characteristics as columns to classify each; you can then look for "holes" in your data and decide whether you need to find data with other combinations of characteristics. Figure 4 shows an example.

Type of project (Initial, Revisit, Update): *Revisit* **Project focus (Org, Service, Tech):** *All three*

Data:	Strategic Questions:	Strategic Importance:	Data Category:	Constituency Interest:	Source:	What Kind of Data / How to Present:
Name of data element	Where we are What drives us Where we want to be How to get there	Sets the stage Raises issues Highlights trends Constituency Resource concern	Financial Demographic Operational Performance Environmental (Issue focused)	Faculty/ Researchers/ Students/ Administrators / Other staff	Primary: - Survey/ Quest - Interviews Secondary: - Instrument - Public - Other	Text/ Numbers/ Pictures Table/Graph/ Chart/Picture/ Video/ Interactive
Staffing/skills	Current state (Where we are)	Sets the stage Raises issues	Operational, financial, performance	Staff	Secondary	Tables, graphs
Focus group results	Future state (Where we want to be)	Raises issues Resource concern	Operational	All	Primary	Charts
Spending (by purpose) for several years	Current state Drivers	Raises issues Highlight trends Resource concern	Financial	All	Secondary	Charts, tables

© Jerrold M. Grochow / Creative Commons License BY-SA-NC

Figure 4. Strategic planning data-organizing matrix example

Strategic Planning Example

Although it may soon be relegated to the "oldies but goodies" list as we move to cloud services, a favorite topic of strategic planning committees has been how best to provide email services to an institution's various constituencies. Which data are needed to make that discussion most productive? Current usage and cost data are obvious choices, but industry trends and other projections are also useful. The following example shows how these data types can raise issues for discussion. To simplify, we'll focus on usage data for students and faculty only (see table 1).

Table 1. Usage data — emails per month per person over a five-year period

Group	2013	2014	2015	2016	2017
Students	1,500	1,700	2,000	2,200	2,350
Faculty	700	800	925	1,100	1,250

The first question to ask is whether these data are well defined. For example, does "year" refer to a calendar year or an academic year? (Here, we'll assume it's an academic year.) Does "students" include full-time, part-time, residential, nonresidential, and online students? The list of such questions goes on, and a full data definition should answer all such questions — before they are asked.

Now let's look at the data table itself. Plotting the table data, as in figure 5, offers another view and more information. We can clearly see from the data table that students send and receive approximately double the number of emails as faculty and that the number of emails has grown significantly each year for both groups. Is this what we expected? If so (or not), why (or why not)? What does it say about student and faculty communication behavior? Further, if we look more closely at the table (or simply look at the plot in figure 5), we can see that student emails seem to be leveling off, while faculty emails are still growing, although even that growth is

slowing. This raises questions as to what is going on and whether it shows a trend that will continue or even accelerate in the future. These are clearly issues to explore further, perhaps by collecting additional data.

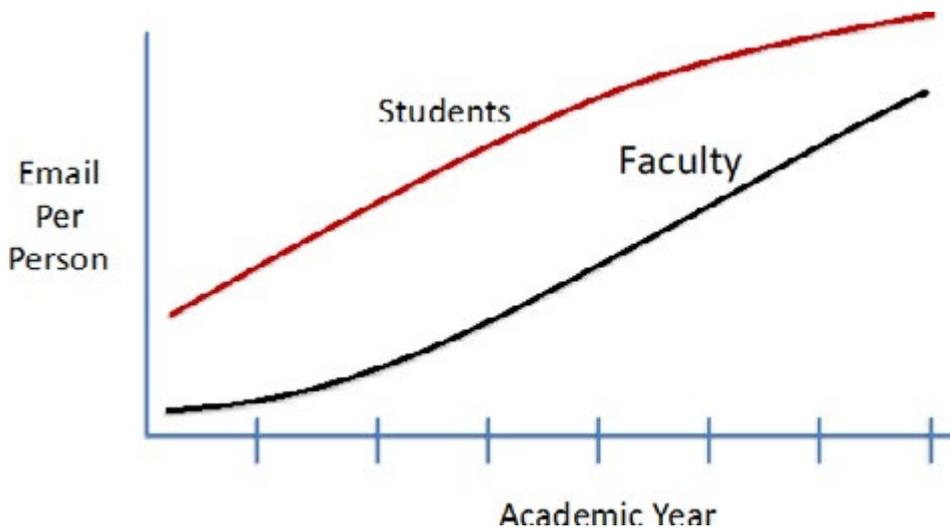


Figure 5. Student and faculty email usage over time

If we collect additional data — say, of social media posts — we might find an explanation for the slowing growth in email (see figure 6). In fact, the social media plot would be off the charts for most students! Adding data on social media postings could make the common issue of what is happening to student–faculty communication on campus even more of a concern (given faculty reliance on email!), perhaps one requiring a specific survey. Assuming the posited change — that students are increasingly using social media in place of email — is true, are faculty members noticing this shift in their students' communication habits ("students don't even read most emails from me anymore")? Should faculty be encouraged to use social media in addition to (or in place of) email? What does the increasing use of social media for on-campus communications mean for the IT department, given that it doesn't have anything to do with the social media platforms that students use? Also, what does this mean for resources and the types of people IT must deal with if issues occur? Other factors — such as the integrity, security, and availability of communications — all need to be discussed.

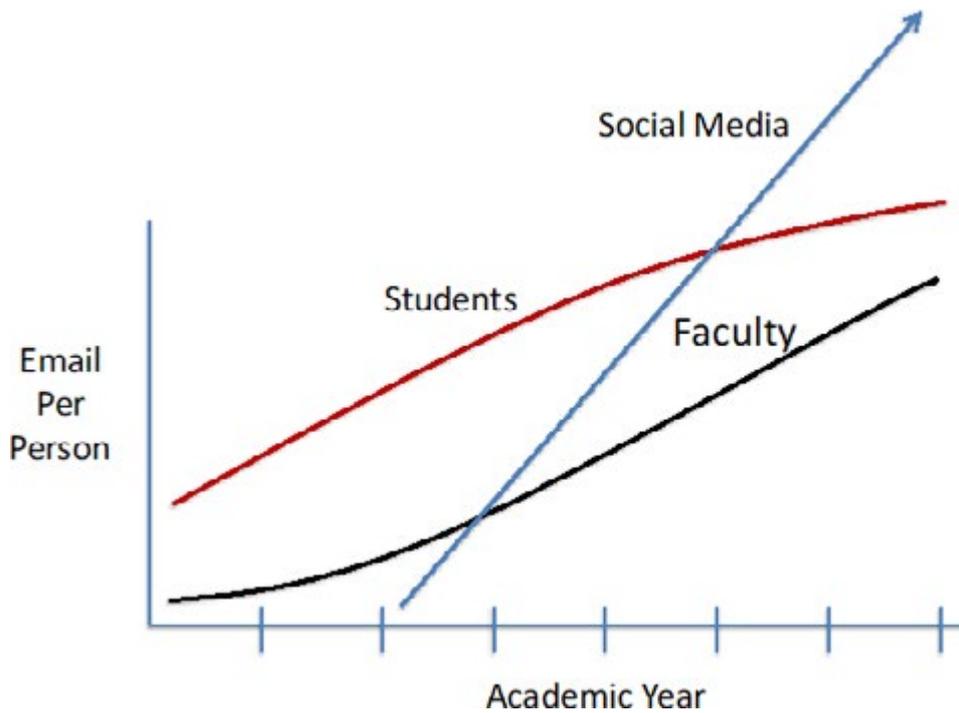


Figure 6. Email and social media usage example

As this example shows, presenting data about the *current state* (number of emails) by *constituency* (faculty and students), and factors *influencing the future* (trends in email versus social media use) can lead to fruitful strategic discussions about where campus communication is headed and what IT will need to do to support it. Filling in the data-organizing matrix (figure 7) also highlights some missing data (financial information, for example) that can be added to further the analysis. This simple example also demonstrates the value of presenting data in multiple forms — here, a table and plot, but there are of course many other options.

Type of project (Initial, Revisit, Update): *Revisit* **Project focus (Org, Service, Tech):** *All three*

Data:	Strategic Questions:	Strategic Importance:	Data Category:	Constituency Interest:	Source:	What Kind of Data / How to Present:
Name of data element	Where we are What drives us Where we want to be How to get there	Sets the stage Raises issues Highlights trends Constituency Resource concern	Financial Demographic Operational Performance Environmental (Issue focused)	Faculty/ Researchers/ Students/ Administrators / Other staff	Primary: - Survey/ - Quest - Interviews Secondary: - Instrument - Public - Other	Text/ Numbers/ Pictures Table/Graph/ Chart/Picture/ Video/ Interactive
Email usage	Current state (Where we are)	Sets the stage Raises issues Highlights trends	Operational (Financial)	Faculty, student, (administratio n), (alumni)	Secondary (existing data)	Table, graph
Social media usage	Current state (Driver)	Sets the stage Raises issues Highlights trends (Resource concern)	Operational (Financial)	Faculty, student, (administratio n), (alumni)	Primary (collected for this project)	Table, graph

© Jerrold M. Grochow / Creative Commons License BY-SA-NC

Figure 7. Strategic planning data organizing matrix for the email/social media example

Strategy and Systematic Planning

Successful IT strategic plans result from discussions of a broad variety of issues related to the future provision and use of IT services. These discussions are based on a thorough understanding of the current state, industry and technology trends, and related information.

Ensuring that appropriate data are available to the strategic planning committee for these purposes requires a systematic approach to identifying, defining, collecting, analyzing, organizing, and presenting strategic planning data.

Acknowledgment

This article is based on the author's 2017 EDUCAUSE Annual Conference seminar, "**Data-Driven IT Strategic Planning for Data-Driven IT Leaders.**"

Jerrold M. Grochow is a Strategic Consultant on IT in higher education and a research affiliate in cybersecurity at MIT's Sloan School of Management. He was previously Vice President for Information Services & Technology at MIT and Interim Vice President at Internet2.

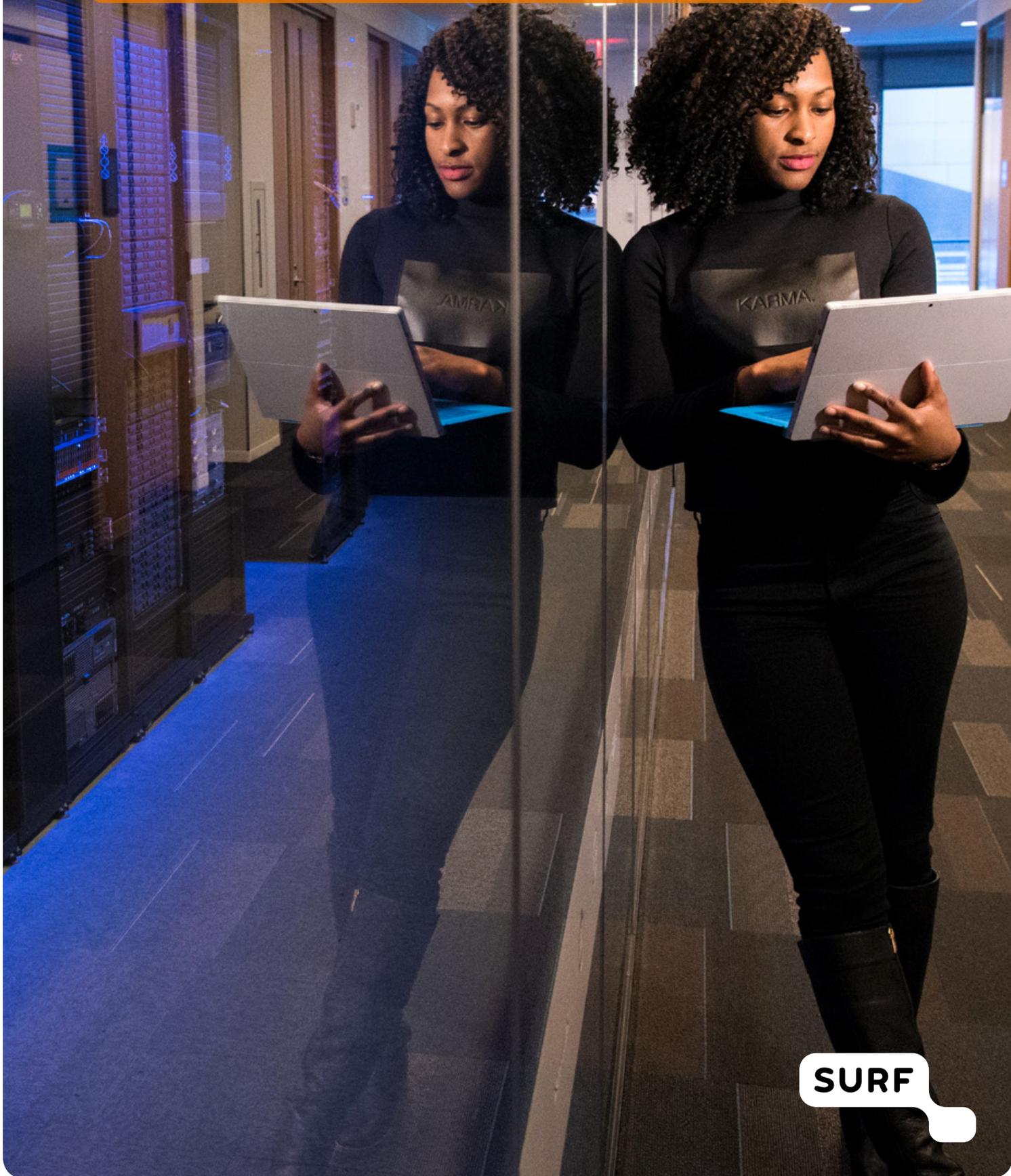
© 2018 Jerrold M. Grochow. The text of this work is licensed under a **Creative Commons BY-NC-ND 4.0 International License** .

- **Data Administration and Management, Data Governance, Data Management Planning, IT Organization, IT Strategic Planning, Planning, Portfolio and Project Management, Strategic Planning**

ARTIKEL 6

HOE DATA DE KWALITEIT
VAN HET ONDERWIJS KUNNEN
VERBETEREN

HOE DATA DE KWALITEIT VAN HET ONDERWIJS KUNNEN VERBETEREN



SURF

INHOUD

Inleiding	3
1. De potentie van learning analytics voor het Nederlandse hoger onderwijs	5
<i>Praktijkvoorbeeld: KU Leuven</i>	<i>7</i>
<i>Praktijkvoorbeeld: Vrije Universiteit</i>	<i>8</i>
<i>Praktijkvoorbeeld: SURF learning analytics experimenten</i>	<i>9</i>
2. Uitdagingen van learning analytics	10
<i>Praktijkvoorbeeld: Open Universiteit</i>	<i>14</i>
3. Beginnen met learning analytics: 10 tips	16
4. Opschalen van learning analytics: 15 tips	17

INLEIDING

De afgelopen jaren zijn er in het Nederlandse hoger onderwijs veel inzichten opgedaan over de mogelijkheden van datagebruik voor het onderwijs. Sommige van die inzichten zijn een direct resultaat van experimenten met learning analytics, andere komen voort uit onderzoek óver learning analytics.

Online onderwijs brengt een aantal grote veranderingen met zich mee. Eén ervan is dat docenten minder zicht hebben op wat studenten doen. In een lokaal zien docenten hoe het met de studenten gaat. Het is duidelijk of ze er zijn of niet, of ze meedoen en of ze hun opdrachten hebben gemaakt. In een online onderwijsomgeving ontbreekt dat overzicht grotendeels. Dat is lastig, want iedere docent vindt het belangrijk om de studenten goed te kunnen begeleiden. Het gebruik van data in het onderwijs kan hiervoor een oplossing bieden. De overkoepelende term hiervoor is learning analytics. Learning analytics is het verzamelen, analyseren en rapporteren van data van studenten en hun omgeving om zo het onderwijs en de omgeving waarin dat onderwijs plaatsvindt te begrijpen en te verbeteren.¹

Met learning analytics heeft de docent meer grip op online studiegedrag, zodat hij studenten kan ondersteunen en uitval te voorkomen. Daarnaast helpt learning analytics bij het verbeteren van online materiaal. Data over hoe vaak, hoe lang, wanneer en hoe materiaal wordt gebruikt, geven inzicht in de vraag of het materiaal wel doet waarvoor het is bedoeld. Eigenlijk is learning analytics een logisch gevolg van online onderwijs. Het is een middel om zicht te krijgen op wat er online gebeurt.

Bestaande data koppelen

Datagebruik kan ook breder worden ingezet. Hogeronderwijsinstellingen beschikken over heel veel verschillende databronnen: over de vooropleiding van de student, het oriëntatietraject dat hem naar de instelling bracht, zijn leeftijd, waar hij is opgegroeid, studieresultaten, et cetera. Mogelijk zijn er daarnaast WiFi-trackers binnen de instelling geïnstalleerd, of hangen er warmtesensoren in de collegezalen. Koppel een aantal van zulke data aan elkaar en je verkrijgt inzicht in het leergedrag van de studenten, de kwaliteit van het onderwijs en de effectiviteit van de instelling. Op instellingsniveau biedt learning analytics daarmee nieuwe mogelijkheden om trends te bestuderen, studieuitval te beperken en efficiënter te werken. Informatie die je toch al hebt, kun je gaan koppelen, omdat die informatie tegenwoordig digitaal beschikbaar is. Daardoor heb je bijvoorbeeld de mogelijkheid om risicostudenten te *nudgen*, een zetje te geven. Afhankelijk van het antwoord kun je al dan niet actie ondernemen.

Gericht zoeken naar verbanden

Nog maar enkele jaren geleden leek de belofte van big data dat het verzamelen van zo veel mogelijk data zou leiden tot het vinden van steeds meer verbanden. Nog afgezien van de vraag of dat klopt, is het een werkwijze die slecht samengaat met de AVG en de manier waarop hogeronderwijsinstellingen met hun studenten en medewerkers willen omgaan. Learning analytics kan beter onderwijs opleveren voor de student doordat het meer inzicht oplevert. Dat inzicht komt echter altijd voort uit een concrete vraag. De eerste vraag is dan ook wat je wil bereiken met het gebruik van data. Waar wil je precies achterkomen en waarom? De vervolgvraag is: welke data heb je beslist nodig om een antwoord te vinden

¹ Deze definitie is afkomstig van de Learning Analytics Community Exchange (LACE). Zie ook <http://www.laceproject.eu/faqs/learning-analytics/>

op je vraag? Niet onbegrensd grazen in data, maar gericht zoeken in een geminimaliseerde dataset, is anno 2019 het credo van learning analytics. Met een strak ingericht proces kun je uitstekend tot nieuwe inzichten komen, zo laat het praktijkvoorbeeld van de VU in deze uitgave zien.

Een andere ontwikkeling waarover een paar jaar geleden druk werd gespeculeerd, was dat instellingen met behulp van learning analytics konden gaan voorspellen welke eerstejaars weinig kans zouden maken om de studie te voltooien. In Europa bestaat vooralsnog geen animo om studenten te selecteren op basis van computergegenereerde voorspellingen. Het Nederlandse en Belgische hoger onderwijs is diep doordrongen van het belang van een ethische, transparante omgang met data van studenten en medewerkers. Statistieken kunnen wel worden ingezet om studenten een realistisch beeld te geven van hun studievoortgang, zoals de KU Leuven dat doet.

Leeswijzer

Wat is de potentie van learning analytics voor het Nederlandse hoger onderwijs? Waar hebben studenten en docenten daadwerkelijk wat aan? Daarover gaat hoofdstuk één. Aan de hand van een aantal praktijkvoorbeelden beschrijven we wat Nederlandse en Belgische hogeronderwijsinstellingen na een paar jaar van experimenteren zien als de belangrijkste meerwaarde. Hoofdstuk twee gaat over de didactische, juridische, ethische en technische uitdagingen en over de mogelijke impact op het onderwijs. In dit hoofdstuk worden de inzichten benoemd die voortkomen uit een eerder learning analyticsexperiment dat SURF met de instellingen uitvoerde. Ook laten we zien hoe deze inzichten leiden tot het inslaan van een nieuwe richting, met meer mogelijkheden voor ethisch datagebruik. De whitepaper bevat naast tips om te starten met het gebruik van data in het onderwijs ook tips voor instellingen die de experimentele fase zijn ontstegen en zich willen wagen aan het opschalen van learning analytics.

Verschillende vormen van analytics

De Learning Analytics Community Exchange (LACE) hanteert een brede definitie van *datagebruik in het onderwijs*: het verzamelen, analyseren en rapporteren van data van studenten en hun omgeving om zo het onderwijs en de omgeving waarin dat onderwijs plaatsvindt te begrijpen en te verbeteren. Dat zegt echter nog niets over het type data dat wordt gebruikt. Kijkt men vooruit, terug of kijkt men *real-time* mee?

In deze whitepaper hanteren we consequent de term 'learning analytics' voor alle vormen van analytics. Volgens sommigen zegt *learning analytics* echter alleen iets over wat er tijdens het leerproces plaatsvindt. Learning analytics biedt inzage in het studieproces, met als doel het beter kunnen helpen of bijsturen van de student. Volgens iets bredere definities kunnen voor learning analytics ook *multimodal data* worden gebruikt, zoals data afkomstig van fitbits, WiFi-trackers of detectiepoortjes binnen de instelling.

Student analytics maakt gebruik van bestaande data uit verschillende bronnen. De data worden toegankelijk gemaakt voor docenten en studiebegeleiders om bijvoorbeeld uitspraken te doen over mogelijke

risicostudenten of studenten die meer uitdaging kunnen gebruiken.

Academic analytics gaat om het analyseren van trends en ontwikkelingen via data op instellings- of instituutniveau. Denk aan rendementcijfers, maar ook bijvoorbeeld aan prognoses over de kans dat verschillende groepen studenten een opleiding afronden.

Machine learning is een ontwikkeling die zich nog veelal buiten het onderwijs afspeelt. Het komt nog weinig voor dat hogeronderwijsinstellingen een grote hoeveelheid data aan een algoritme voeren, in de hoop dat het algoritme zelfstandig nieuwe verbanden legt.

Over de precieze invulling van de termen bestaat discussie. Belangrijk is vooral dat er diverse manieren zijn om met data in het onderwijs om te gaan. Het is belangrijk om na te denken op welke manier je learning analytics wilt inzetten: wil je weten wat er gebeurt tijdens het leerproces van de studenten, of wil je trends analyseren aan de hand van bestaande bronnen? Hoe je die van vorm van datagebruik vervolgens noemt, is aan jou.

DE POTENTIE VAN LEARNING ANALYTICS VOOR HET NEDERLANDSE HOGER ONDERWIJS

LACE verzamelt wereldwijd bewijzen voor de effectiviteit van de inzet van learning analytics in het onderwijs. In de [LACE Evidence Hub](#) is een grote hoeveelheid mooie voorbeelden te vinden van instellingen die met learning analytics aan de slag zijn gegaan. Waar ligt de meerwaarde voor het Nederlandse hoger onderwijs? In interventies en onderwijs op maat? Of moet er op de korte termijn toch meer naar verbetering van het onderwijsmateriaal worden gekeken?

Justian Knobbout onderzoekt de effecten van learning analytics voor de Hogeschool Utrecht. Hij ziet een grote behoefte aan *low level analytics*. Studenten en docenten zijn vaak geholpen met eenvoudige dashboard, die inzicht geven in wat hun studenten doen in de digitale leeromgeving. “Zet learning analytics in voor het evalueren van het gebruik van onderwijsmateriaal,” zegt hij. “In het ideale geval krijgen de gebruikers daardoor veel beter inzicht in het leerproces dan nu het geval is. Dat kan zich vertalen in een situatie waarin de docent precies weet hoe hij verschillende studenten moet aansturen, of de student wat de volgende stap in het leerproces zou moeten zijn.”

Maar het ene vak is de andere niet; waar de ene docent bijvoorbeeld behoefte heeft aan een samenvatting van de meest besproken onderwerpen op het online forum, wil de andere docent vooral weten hoe zijn videomateriaal wordt bekeken. “Een systeem voor learning analytics moet flexibel zijn en in die verschillende wensen kunnen voorzien,” zegt Knobbout. Dat is nog niet altijd het geval.

Waar datagebruik in het onderwijs in de breedte aan een gestage opmars bezig zijn, blijft de praktijk in de diepte nog achter op de mogelijkheden, signaleert hij. Zo ziet hij dat instellingen bij het implementeren van een nieuwe digitale leeromgeving (DLO) learning analytics nog beschouwen als een ‘nice to have’. “Het is veel beter om het meteen uit te rollen, in plaats van achteraf te bedenken hoe je data wil onttrekken uit verschillende applicaties die in één omgeving samenwerken. Dat is namelijk bijna onmogelijk. Je moet vanaf de start een infrastructuur voor data-opslag hebben, met de mogelijkheid om die voorlopig nog uit te zetten zolang je er nog niet aan toe komt.”

Binnen het in 2016 opgerichte Amsterdam Center for learning analytics (ACLA), een onderzoeksgroep van de Vrije Universiteit, wordt onderzoek gedaan naar de effectiviteit van de inzet van learning analytics in het onderwijs. Directeur is Chris van Klaveren, tevens hoofddocent Onderwijswetenschappen & Economie aan de VU. Samen met Ilja Cornelisz en Martijn Meeter richtte hij ACLA op, omdat hij inzag dat learning analytics nauwe samenwerking vereist tussen verschillende disciplines. De onderzoeksgroep combineert inzichten uit de wetenschappelijke vakgebieden economie, computerwetenschappen, kunstmatige intelligentie, psychologie en onderwijskunde. Hij is enthousiast over de inzichten die uit de samenwerking voortkomt. “Je hebt niet zelf de kennis om alles op te lossen. Door de dialoog aan te gaan, erken je elkaars blinde vlekken en kun je die aanpakken.”

Learning analytics is in de eerste plaats mensenwerk. “Ik heb nog nooit data een vraag zien beantwoorden en algoritmes ontstaan niet vanzelf, maar worden bedacht en geprogrammeerd door mensen,” zegt Van Klaveren. “De meerwaarde van learning analytics verschilt niet van andere tools die het onderwijs individualiseren of personaliseren: het

heeft meerwaarde als het iets verbetert, of dat nu de leerbeleving, de leeromgeving of de leeruitkomst is." Veelbelovend is volgens hem het real time karakter ervan. "Je monitort de juiste informatie en de lerende kan die informatie op elk moment tot zich nemen. Dat kan in potentie positieve gevolgen hebben."

Vooraf studenten die al betrokken zijn bij het eigen leerproces, hebben veel baat bij learning analytics, blijkt uit onderzoek van Knobbout. "Zij reflecteren op hun leergedrag en daarbij komt een dashboard goed van pas. Maar eerstejaars die nog niet helemaal op hun plek zitten, zijn met andere zaken bezig dan hun leerproces."



PRAKTIJKVOORBEELD

ACTIONABLE FEEDBACK: DATA WAAR DE STUDENT IETS MEE KÁN

KU LEUVEN

De overgang van de middelbare school naar het hoger onderwijs is vaak groot. Daarom besloot de KU Leuven vanaf 2015 te onderzoeken hoe bestaande data kunnen worden ingezet om studenten tijdens deze overgang feedback te geven. Eerstejaars studenten ontvangen nu vier keer per jaar een dashboard dat inzicht geeft in hun leer- en studievaardigheden en academische resultaten. De dashboards geven de studenten ook concrete tips om zo nodig verbeteringen te treffen.

De data zijn afkomstig van twee dataspooren. Allereerst de punten van studenten, afkomstig van tussentijdse toetsen, de examens en de ijkingsstoets, de toets die voor de start van het academiejaar wordt afgenomen. Het tweede dataspoor zijn zelfgerapporteerde gegevens. Alle eerstejaars vullen een gevalideerde vragenlijst in over hun leer- en studievaardigheden, de zogeheten LASSI-vragenlijst (Learning and Study Strategies Inventory). Naar aanleiding van de vragenlijst ontvangen ze aan het begin van het studiejaar een eerste dashboard. De dashboards die ze later in het jaar ontvangen tonen onder meer hun positionering ten opzichte van studiegenoten. Ook krijgen ze te zien hoe studenten die in het verleden een vergelijkbaar aantal punten haalden het ervan af brachten. Tinne De Laet, hoofd van de dienst studentenbegeleiding van de faculteit ingenieurswetenschappen, vertelt: “We geven de studenten informatie die ze voorheen niet hadden, maar die eigenlijk heel gemakkelijk ontsloten wordt uit de databanken van een hogeronderwijsinstelling. Het is geen ingewikkeld algoritme, geen geavanceerde technologie. Iedereen kan dit gemakkelijk implementeren. De echte uitdaging ligt in het ontwerp en de integratie in de begeleiding van de student.” Want wat wil je de studenten precies tonen en hoe spoor je ze aan om actie te ondernemen als dat nodig is?

De Laet: “We proberen studenten advies te geven over iets dat *actionable* is, iets waar ze iets aan kunnen doen. Daarom nemen we het geslacht of de sociaal-economische status niet mee in het model, want daar kan de student niets aan veranderen. We koppelen de student analytics aan tips en verwij-

zingen waarmee ze zich kunnen verbeteren. Over leer- en studeervaardigheden biedt KU Leuven een heleboel trainingen aan.”

De aanpak blijkt succesvol. Naar aanleiding van het dashboard over de leer- en studeervaardigheden komen de studenten vaker op gesprek bij de studieloopbaanadviseur, nog vroeg in het semester. Ook geven ze aan dat ze de learning analytics en de tips nuttig vinden.

De grote ethische uitdaging is om studenten niet in de put te sturen door de statistieken. Daarvoor is intern een aantal richtlijnen opgesteld. Zo worden de studenten in de statistieken opgedeeld in drie groepen, waarbij voor elke groep toch nog studenten succesvol zijn in hun studie. De Laet: “Het kan altijd zijn dat de student er in slaagt om iets te realiseren dat nog nooit iemand hem of haar heeft voorgedaan, dus zullen we nooit communiceren dat dit niet mogelijk zou kunnen zijn. Het dashboard kan een hard signaal zijn, maar wel een genuanceerd signaal.” Het is eerlijk over de onzekerheid die op de voorspelling zit. En er wordt nooit gesproken over kansen. “Het is flauwekul om te zeggen dat iemand tachtig procent kans heeft om de bachelor in drie jaar te halen,” benadrukt De Laet. “We zeggen wel: van de studenten die dezelfde punten hadden als jij, heeft tachtig procent zijn bachelor in drie jaar gehaald. Maar dit is op basis van wat wij van jou weten en wij hebben geen compleet beeld van wie jij bent.”

Studenten zijn over het algemeen enthousiast over het project. Dat blijkt wel uit het feit dat de dashboards ondertussen op hun verzoek zijn uitgerold over de gehele bachelor. Sowieso zet de opmars van learning analytics binnen de KU Leuven gestaag door. Vanuit de faculteit ingenieurswetenschap is het project uitgerold naar de onderwijsgroep Wetenschap en Technologie. Afgelopen academiejaar was het de beurt aan 26 opleidingen binnen KU Leuven. Er zijn 12.000 studenten bereikt. De volgende stap is een instellingsbreed project.

PRAKTIJKVOORBEELD

ONVERWACHT PROFIJT OP INSTELLINGSNIVEAU

VRIJE UNIVERSITEIT

Als onderdeel van een groot project om de gehele onderwijsadministratie te digitaliseren, startte Theo Bakker, destijds werkzaam bij Deloitte, een deelproject met als doel om met gecombineerde data studenten meer inzicht te bieden in hun studievoortgang. Met Deloitte ontwikkelde de VU een model dat een prognose gaf van de kans dat eerstejaars hun eerste studiejaar zouden halen. Het model werd ingezet in een pilot om studieadviseurs en tutoeren een extra handvat te geven bij de studiebegeleiding. De VU ontwikkelde hiervoor een uitgebreide Code of Practice voor studenten. De pilot was geen onverdeeld succes, maar de dataset levert nog wekelijks interessante inzichten op.

In totaal heeft de VU een dataset van ruim 1500 kenmerken per student, afkomstig uit 80 databronnen. Bakker is inmiddels strategisch beleidsmedewerker en projectleider Student Analytics bij de VU en aanvoerder van de zone 'Veilig en betrouwbaar benutten van studiedata' van het Versnellingsplan Onderwijsinnovatie met ICT. Hij vertelt: "Wij kijken naar transactionele data, dus data die al aanwezig zijn in administraties. Van welke scholen komen de studenten, wanneer hebben ze zich aangemeld voor hun studie, zijn ze bij open dagen of meeloopdagen geweest, hebben ze matching gedaan? Ook gebruiken we data van de taaltoets die we al onze studenten afnemen. En we gebruiken achtergrondkenmerken, zoals vooropleiding, leeftijd, geslacht. Vervolgens bekijken we welke resultaten ze halen aan de VU."

Afhankelijk van de bron kan Bakker elke week nieuwe gegevens leveren. Die snelheid noemt hij een belangrijke meerwaarde van learning analytics. "Studieadviseurs komen in actie als een student langskomt. Dit soort instrumenten heeft de belofte dat je proactief

kunt werken en bijsturen." Voor de pilot bleek dat niet op te gaan; voor de studieadviseurs voegden de data weinig toe aan wat de student zelf al meldde. Bovendien waren ze bang om de studenten te ontmoedigen met de prognoses. Mede daarom is het begeleiden van individuele studenten op basis van prognoses voorlopig van de baan. Er zijn wel plannen om in 2020 in nauw overleg met studenten, studieadviseurs en tutoeren een nieuw project op studenten-niveau uit te voeren.

Maar de pilot had een belangrijke nevenvangst. In grote hoeveelheden bleken de data veel inzichten te bieden over de instelling. Bakker: "Opeens ontdekten wij dat van al onze eerstejaars studenten een groot deel al een studie had gedaan en helemaal geen VWO-student was. Dat gaf ons een argument om tegen Den Haag te zeggen dat onze bekostiging wezenlijk lager is dan die van andere universiteiten." Dergelijke inzichten wekten de interesse van het College van Bestuur. Inmiddels vinden er volop beleidsaanpassingen plaats op basis van data. "Bijna elke week doen we nieuwe ontdekkingen," zegt Bakker. "Ik was verrast hoe slecht de VU zichzelf eigenlijk kende. Drie jaar geleden werd veel beleid gemaakt op ideeën. Nu is dat zo veel mogelijk evidence-based."

In 2019 wordt VU Analytics een formeel team. VU Analytics richt zich dan ook op nieuwe beleidsterreinen, zoals werving en de arbeidsmarkt. Ook wil het team meer ad hoc vragen van docenten kunnen beantwoorden. Bakker: "Ik hoop dat er steeds meer onderzoeksgroepen komen die deze data zelf gaan onderzoeken."

PRAKTIJKVOORBEELD

SURF LEARNING ANALYTICS EXPERIMENTEN

EXPERIMENTEN GEVEN DOCENTEN HANDS-ON ERVARING

SURF ontwikkelde samen met enkele onderwijsinstellingen een experimenteeromgeving waarbinnen een aantal instellingen hun eerste schreden op het gebied van learning analytics hebben gezet. In 2016 vonden de eerste experimenten plaats, met docenten van de Vrije Universiteit en Hogeschool Utrecht. In 2017 deden de Vrije Universiteit, Hogeschool Windesheim, Technische Universiteit Eindhoven, Zuyd Hogeschool en ROC Noorderpoort mee. In totaal ging het om twaalf docenten van zes instellingen. De ervaringen van de instellingen staan beschreven in de [geleerde lessen](#).

VU

Ilja Cornelisz, mede-oprichter van Amsterdam Center for learning analytics (ACLA) en docent Statistiek aan de VU, was betrokken bij de ontwikkeling van de experimenteeromgeving voor learning analytics van SURF. Als docent nam hij twee keer deel aan het experiment, om meer inzicht te krijgen in het leerproces en het onderwijsaanbod beter te laten aansluiten op de behoeften van studenten. Hij benadrukt dat learning analytics signaleert, maar niet direct een oplossing biedt. Het is een hulpmiddel om het gesprek aan te gaan met studenten en docenten.

HU

De Hogeschool Utrecht onderzoekt hoe learning analytics kan bijdragen aan het verbeteren van het onderwijsproces. Het experiment was voor drie docenten een manier om learning analytics op een makkelijke manier in de praktijk te brengen. Zij zien het als een mooie manier om het onderwijs te monitoren.

Windesheim

Yvonne van Vooren, docent bij Hogeschool Windesheim, gebruikte learning analytics om meer inzicht te krijgen in het leergedrag en de leervaardigheden van haar studenten. Op basis daarvan ging ze met hen in gesprek over verbeterpunten aan de kant van de student en mogelijke aanpassingen aan de lessen. Ze concludeert dat learning analytics hiervoor een mooie tool biedt.

TU/e

De Technische Universiteit Eindhoven werkte aan een beleidskader over learning analytics en dat maakte deelname aan het experiment interessant. De deelnemende docenten wilden meer inzicht krijgen in het gebruik van het online leer materiaal binnen het vak Calculus, door het verzamelen van zo veel mogelijk data. Dat bleek een arbeidsintensieve en niet erg vruchtbare methode.

Zuyd

Het doel van het experiment van Zuyd Hogeschool was om docenten te laten kennismaken met learning analytics. Deze insteek was wat te vrijblijvend, concludeert de hogeschool, die naar aanleiding hiervan aanbeveelt dat learning analytics vraaggericht moet worden ingezet.

Noorderpoort

ROC Noorderpoort is de eerste mbo-instelling die ervaring opdeed met learning analytics in de experimenteeromgeving van SURF. Het inzetten van learning analytics bij het vak Digitale Vaardigheden leverde docent Jacob Poortstra een schat aan informatie op en hij hoopt op een mogelijk vervolg.



UITDAGINGEN VAN LEARNING ANALYTICS

Didactische uitdagingen

De toepassing van learning analytics begint met een goede onderzoeksvraag, maar dat is niet nieuw. De vraag ‘wat is het beste onderwijs dat ik wil aanbieden met behulp van de mij beschikbare hulpmiddelen?’ is tijdloos. Omdat de hulpmiddelen steeds geavanceerder worden, moet je echter beter nadenken hoe je ze wil inzetten. Begin daarmee lang voordat je onderwijs start, anders blijken data bijvoorbeeld niet beschikbaar, of niet te interpreteren.

Vervolgens is er een aantal randvoorwaarden nodig om didactische meerwaarde te bereiken. De Schot Niall Sclater, als consultant verbonden aan Jisc, de Britse zusterorganisatie van SURF, somt op: “Heb je mensen in huis die de data begrijpen, die de databronnen kennen en ze kunnen interpreteren, die voorspellende modellen kunnen maken of statistische analyses kunnen doen, en nog belangrijker, als al het andere achter de rug is, interventies kunnen plegen met de studenten? Je kunt statistieken produceren tot je een ons weegt, maar als je niet in staat bent om interventies te plegen, dan kun je er net zo goed niet eens aan beginnen.”

De docenten hoeven zelf niet over ál deze vaardigheden te beschikken, maar een bepaalde handigheid met ICT is, naast kritisch denkvermogen en het vermogen om analyses te interpreteren, wel een vereiste. Didactische uitdagingen zijn niet los te zien van de technologische uitdagingen. Beginnen met learning analytics betekent allereerst dat de digitale leeromgeving naar behoeven functioneert en dat docenten beschikken over voldoende ICT-vaardigheden. Veel instellingen zijn nog niet zo ver.

Juridische uitdagingen

De komst van de Algemene Verordening Gegevensbescherming (AVG) heeft veel instellingen huiverig gemaakt voor het gebruik van learning analytics. Toch is er strikt genomen niet zo veel veranderd in de mogelijkheden op het gebied van datagebruik. SURF probeert op verschillende manieren om de onzekerheden van instellingen weg te nemen, bijvoorbeeld door middel van een onderzoek naar passende maatregelen bij de toepassing van learning analytics, dat in het voorjaar van 2019 wordt afgerond.

In vergelijking met de Wet bescherming persoonsgegevens (Wbp) moeten instellingen die aan learning analytics doen onder de AVG iets meer aan interne en externe verantwoording doen, zegt Niels Westerlaken, IT- en privacy-advocaat bij Project Moore Advocaten, dat het onderzoek uitvoert. “De andere stappen zijn vrijwel hetzelfde gebleven, maar juist van deze stap schrikken projectverantwoordelijken nog wel eens, omdat je daadwerkelijk moet laten zien wat je doet.” De vraag welke maatregelen vervolgens passend zijn, is niet eenduidig te beantwoorden, omdat de impact van een onderzoek waarin learning analytics worden ingezet flink kan verschillen. Westerlaken: “Er is een aantal algemene beginselen waaraan je altijd moet voldoen: genoeg informatie verstrekken, veel controle geven over de data, duidelijk maken wat je wil gaan doen. Bij projecten die impact hebben op individuen, waarin je misschien ook data gaat samenvoegen om profielen te maken, zou het vragen van toestemming een passende maatregel kunnen zijn.”

Een instelling zou kunnen redeneren dat het gebruik van data het onderwijs altijd ten goede komt en daarmee past binnen de wettelijke plicht. Met die redenering maakt hij korte metten. “Instellingen zullen een goed doel hebben met de onderzoeken die ze doen,

maar dat is geen vrijbrief om learning analytics direct toe te passen. Er zal namelijk een belangenafweging moeten plaatsvinden tussen de doelen van de instellingen en de privacy van individuen.”

Voor juristen zal een eerste uitspraak van een rechtbank op dit onderwijsterrein heel interessant zijn, omdat het meer duidelijkheid biedt over hoe de rechtelijke macht de wet interpreteert. Of een maatregel ‘passend’ is, is namelijk geen wet van meden en perzen. “De verwachtingen van degene over wie het gaat moeten aansluiten op wat er daadwerkelijk gebeurt,” legt Westerlaken uit. “Over vijf of tien jaar verwachten studenten wellicht al dat er learning analytics worden toegepast. Dan zul je een andere schaal van maatregelen nemen dan in de fase waarin het redelijk nieuw is. Nu zul je echt goed je best moeten doen om het voordeel ervan duidelijk te krijgen. Niet om jezelf in te dekken, maar vooral om het vertrouwen te krijgen van degene van wie je de gegevens verwerkt.”

De instelling heeft de verplichting om verantwoording af te leggen over welke gegevens worden verzameld en met welk doel dat gebeurt. Op dit moment komen de verwachtingen van studenten niet altijd overeen met de realiteit, met name als het gaat om het verzamelen van data. Dat men bijvoorbeeld het onderwijs verbetert aan de hand van de proeftoetsen in de digitale leeromgeving, dat verwachten veel studenten wel. Maar dat daarvoor bijvoorbeeld op individueel niveau wordt gemonitord welke fouten er worden gemaakt, is hen vaak veel minder duidelijk.

Westerlaken raadt instellingen aan om experimenten met learning analytics goed te documenteren. Voor verantwoording naar buiten en binnen, maar ook om de resultaten te delen buiten de eigen instelling. Hoe pas je de grondslagen toe? Welke doelen zijn vooraf vastgelegd? “Als je dat met andere instellingen kan delen, werk je toe naar een standaard. Dat is voor iedereen leerzaam.” Denk bijvoorbeeld aan uitkomsten van [Data Protection Impact Assessments](#), de [Code of conduct](#) van Jisc en de [Code of Practice](#) van de VU.

Privacy en andere ethische uitdagingen

Learning analytics zal een vanzelfsprekend onderdeel worden van de leeromgeving van een hogeronderwijsinstelling. Dat is de verwachting van Hendrik Drachslers, professor aan het Welten-instituut van de Open Universiteit en lid van het bestuur van de [Society of Learning Analytics Research \(SoLAR\)](#). In 2015 zei hij: “Ik denk dat we over vijf à tien jaar niet meer over learning analytics praten, maar het gewoon gebruiken. De docent haalt dan een rapportage uit het systeem om te gebruiken in het vak, zonder er verder bij stil te staan dat hij aan learning analytics doet.” In 2018 gelooft hij dat nog steeds, “het duurt alleen allemaal iets langer dan ik had verwacht vanwege de nieuwe privacywetgeving, maar wij zien meer en meer learning analytics in het dagelijks gebruik.”

Een van Drachslers grootste nachtmerries is dat een van zijn studenten naar de rechter stapt omdat hij bezwaar maakt tegen het gebruik van zijn gegevens voor learning analytics. Tegelijkertijd zou die student waarschijnlijk een hoog cijfer krijgen, want Drachslers ziet in de toepassing van learning analytics onder meer een geschikte manier om studenten kritische vragen te leren stellen over datagebruik in het algemeen. *Data literacy*, noemt hij dat. “In het Duits zou het *Daten-Mündigkeit* heten. Dat gaat nog wat verder dan literacy. Een data-mündig mens is iemand die zich heeft laten informeren over het gebruik van zijn data, die daartegen in opspraak komt en daar ook recht toe heeft. Ik zie *trusted learning analytics*, zoals ik het noem, als een mogelijkheid om jonge mensen daarin te trainen.”

Hogeronderwijsinstellingen die learning analytics willen toepassen, zullen allereerst een visiedocument moeten opstellen en beleid maken, zegt Drachslers. Transparantie staat daarbij voorop. Welke gegevens worden verzameld? Wie heeft er toegang tot de gegevens? Welke mogelijkheden heeft de eigenaar ervan? Drachslers: “Een visie vraagt om een innovatie-

programma om mensen bewust te maken van waar ze naartoe willen en de instrumenten waarmee ze dat willen bereiken. Vaak is daar verandermanagement voor nodig. Er moet nog veel training voor docenten plaatsvinden. Ook studenten hebben meer kennis nodig. Learning analytics vraagt om een langdurig organisatieproces.”

Er is een aantal hulpmiddelen beschikbaar die instellingen ondersteunen bij een transparante omgang met data. Zo is er de DELICATE checklist, door Drachsler en anderen ontworpen binnen LACE. Jisc onderscheidt maar liefst 86 ethische en juridische kwesties rondom learning analytics. Niall Sclater, die de lijst opstelde, zegt: “De allergrootste uitdaging is de cultuurinstelling; zijn de mensen er klaar voor om beslissingen te nemen op basis van data?”

Technologische uitdagingen

De grootste technische uitdaging van learning analytics is het omgaan met grote hoeveelheden data uit verschillende bronnen. Data kunnen voortkomen uit een studentinformatiesysteem (cijfers en leerpaden), de digitale leeromgeving (leeractiviteiten), maar ook bijvoorbeeld aanwezigheidsdata kunnen een databron vormen. Data worden verzameld in een zogenaamde Learning Record Store. In deze store zijn de data gestructureerd opgeslagen zodat ze kunnen worden gebruikt voor analyse en visualisatie. Om de verschillende bronsystemen waaruit de data afkomstig zijn met de Learning Record Store te laten communiceren, is het gebruik van een protocol nodig. Twee protocollen vechten om de standaard in de markt te worden: xAPI en de IMS Caliper Analytics.

De technologie voor learning analytics heeft zichzelf redelijk uitgekristalliseerd, maar twee ontwikkelingen zorgen voor nieuwe vraagstukken. Enerzijds verbreedt de doelgroep van learning analytics zich: naast studenten, docenten en studiebegeleiders willen onderzoekers de onderwijsdata inzetten en ook op instellingsniveau neemt de interesse in data-analyse toe. Alle doelgroepen willen hun eigen vragen met behulp van learning analytics beantwoord zien. Ook de hoeveelheid databronnen wordt alleen maar groter. Studenten leven niet in een digitale leeromgeving, maar vooral ook daarbuiten. Door data uit de digitale leeromgeving te koppelen aan additionele databronnen, zoals gezondheidstrackers en aan bestaande gegevens, ontstaat een completer beeld en is het mogelijk om een beter antwoord te geven op een onderzoeksvraag.

Tegelijkertijd is er de verplichting om aan de AVG te voldoen. Dat vraagt onder meer om doelmatigheid: je verzamelt alleen data voor een vooraf vastgesteld doel. Probleem is echter dat er met de huidige technologie al snel te veel data worden opgeslagen.

Aanpassingen aan de learning analytics-architectuur van SURF

Deze ontwikkelingen maken het noodzakelijk om de learning analytics-architectuur van SURF te veranderen. Drie overwegingen leiden momenteel tot aanpassingen:

- 1. Eigenaarschap.** De gegevens worden verzameld van de student, voor de student. De student is in controle over zijn persoonlijke data en kan zijn persoonlijke data met een docent of studiebegeleider delen. Geanonimiseerde data, die niet te herleiden zijn tot de individuele student, kunnen worden gebruikt door anderen in de instelling om hun learning analytics-vragen te beantwoorden.
- 2. Dataminimalisatie.** De gegevens worden specifiek voor een bepaalde vraag verzameld. Stel dat een docent wil weten hoe vaak zijn video wordt bekeken. Is het dan relevant dat hij ook ziet dat een student er om 3 uur 's nachts en om 8 uur 's ochtends naar keek? Eigenlijk niet. Dus moet er alleen worden geteld en mogen er geen tijdstippen worden geregistreerd. De uitdaging is om precies genoeg data te verzamelen voor het goed kunnen beantwoorden van de vraag. Dit is de doelbinding², die het mogelijk maakt om learning analytics binnen de AVG uit te voeren. Dataminimalisatie is één van de belangrijkste

privacy-by-design-principes die SURF wil hanteren. Andere privacy-by-design-principes die SURF inzet, zijn het scheiden van de data, het zoveel mogelijk anonimiseren van de data en het abstraheren van de data, zoals in het voorbeeld hierboven.

3. Vertrouwen en transparantie. Door studenten controle te geven over hun data en alleen data te verzamelen die nodig is voor het beantwoorden van een specifieke vraag, is de basis gelegd voor een nieuw platform, dat is gebaseerd op vertrouwen. Daarnaast is er transparantie nodig over de te gebruiken vragen en de algoritmes die de visualisaties bij de vraag verzorgen (open algoritmes). Ook wil SURF volledig transparant zijn over de werking van het nieuwe platform (open architectuur). Is hieraan voldaan, dan verwacht SURF dat learning analytics een nieuwe impuls kan krijgen. Bij het ontwikkelen van de vragen en de open algoritmes worden experts in het veld betrokken, zodat voor iedere stakeholder duidelijk is waar de vraag en het antwoord in de visualisatie vandaan komen. Er komt een pilot met een platform waarop onderzoekers en ontwikkelaars zelf onderzoeksvragen formuleren en databronnen koppelen, al dan niet met ondersteuning van ICT-ers van de instelling. Het platform biedt verschillende services, specifiek gericht op de vragen die de verschillende stakeholders hebben. De services kunnen tussen instellingen worden gedeeld, waardoor iedere instelling baat heeft bij nieuwe ontwikkelingen.

Deze overwegingen hebben een grote impact op de architectuur. De belangrijkste aanpassing is een nieuwe beveiligingslaag voor gecontroleerd databeheer. Deze laag zorgt ervoor dat altijd aan de AVG wordt voldaan. Als een stakeholder een vraag formuleert en gebruikmaakt van een service op het platform, checkt deze laag of het gebruik valt binnen het doel. Voor vragen waarvoor alleen geanonimiseerde gegevens nodig zijn en voor sommige gepersonaliseerde vragen waar de instelling een gerechtvaardigd belang heeft, zal het systeem automatisch toestemming geven. Voor andere vragen met een specifiek doel of gebruik van een specifiek data type kan een consent-scherm voor de student nodig zijn. Er wordt dus niet globaal toestemming gevraagd voor het gebruik van gegevens, maar voor specifieke vragen op het moment dat het zich voordoet. In 2019 zullen de eerste pilots met de nieuwe architectuur van start gaan.

De Open Universiteit houdt zich met dezelfde overwegingen bezig als SURF, zo blijkt uit het praktijkvoorbeeld. De ontwikkeling van een systeem voor *trusted learning analytics* vertoont grote gelijkenissen met de architectuur waaraan SURF bouwt.



² Volgens de AVG moet de verwerking van persoonsgegevens gebonden zijn aan specifieke verzameldoelen. Dit heet doelbinding. Zie voor meer informatie de Handleiding Algemene verordening gegevensbescherming en Uitvoeringswet Algemene verordening gegevensbescherming: <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/handleidingalgemeneverordeninggegevensbescherming.pdf>

PRAKTIJKVOORBEELD

OP WEG NAAR TRUSTED LEARNING ANALYTICS

OPEN UNIVERSITEIT

De brede uitrol en adoptie van learning analytics in Europese hogeronderwijsinstellingen stagneert door ongerustheid over privacy en ethiek met betrekking tot persoonlijke data en de AVG. Die onzekerheid leidt ertoe dat innovatiemanagers twijfelen over implementatie van learning analytics in hun instelling. Om de verworvenheden van tien jaar onderzoek naar learning analytics voor docenten en studenten op een veilige manier beschikbaar te maken, werkt de Open Universiteit (OU) samen met het DIPF - Leibniz Institute for Research and Information in Education in Duitsland - aan een zogenoemde Trusted Learning Analytics[1] (TLA) infrastructuur. Deze infrastructuur pakt het huidige gebrek aan vertrouwen en transparantie in learning analytics aan. Hendrik Drachsler, een van de initiatiefnemers voor TLA, zegt: “Het TLA consortium probeert het ‘contract’ tussen studenten en onderwijsaanbieders te vernieuwen, niet alleen door een hoog vertrouwensniveau te genereren, maar ook door de datawijsheid en betrokkenheid van studenten te vergroten. Het ontwerp van de TLA-infrastructuur en -tools is gebaseerd op ontwerpprocessen die hetzelfde belang toekennen aan ethische en privacy-onderwerpen als aan functionele eisen. Het resultaat is een systeem dat niet alleen technische, maar ook ethische en humane doelen bereikt.”

Trusted learning analytics moet worden ontwikkeld vanuit een gebruikersperspectief. De gebruikers zouden volledige controle moeten hebben over hun data en geen enkele wantrouwen moeten voelen ten opzicht van het systeem. Om dit niveau van vertrouwen te bereiken, is het cruciaal om zo veel mogelijk ‘black box’ data en systemen te veranderen in een ‘white box’. Het consortium probeert zo veel mogelijk open te zijn over de toegepaste algoritmes en transparant over de gebruikte methodieken. Een belangrijk doel van een TLA-systeem is het geven van feedback, in plaats van dat het wordt gebruikt voor

het maken van beslissingen. Het TLA-systeem zou volledige toegang moeten bieden aan de ‘datasubjecten’ - degenen wiens data worden geanalyseerd - en hen moeten toestaan om data te corrigeren, becommentarieer en verwijderen. Ook moeten zij zelf kunnen beslissen wie er toegang heeft tot persoonlijke data. In huidige learning analytics infrastructuur is dit vaak niet het geval. Drachsler: “Voor de OU legt de TLA-aanpak de basis voor een nieuwe manier van omgaan met onderwijsdata en het aanbieden van diensten en activiteiten voor stakeholders.”

Een TLA-systeem vraagt van het datasubject, dus de student of de docent, om een hoog niveau van vaardigheden, waaronder kritisch denken, het analyseren van informatie en zelfregulering. Het open source TLA-systeem probeert een omgeving te creëren waar ze deze cruciale vaardigheden kunnen trainen.

³ Drachsler, H. (2018). Trusted Learning Analytics. Synergie, 06, Fachmagazine für Digitalisierung der Lehre. <https://synergie.blogs.uni-hamburg.de/ausgabe-06-beitrag-drachsler/>

Impact op het onderwijs

Data beantwoorden geen vragen als er geen goed onderzoeksdesign aanwezig is. Evenmin bieden ze kant-en-klare oplossingen. Het teruggeven van informatie over gedrag, wil nog niet zeggen dat mensen hun gedrag vervolgens gaan verbeteren. Chris van Klaveren van Amsterdam Center for learning analytics (ACLA) zegt: “In de huidige vorm ziet learning analytics er mooi uit en je kunt heel veel ‘voetstapjes’ terugzien, maar ik vermoed dat daarmee alleen nog geen effecten worden bereikt. Als we onderwijs op een effectieve manier willen personaliseren, is het van belang dat we niet alleen onderzoeken wat gemiddeld genomen de beste oplossing is. Wat dat betreft zijn Bayesiaanse Bandit modellen⁴ interessant, omdat deze vertrekken vanuit het idee dat we willen weten hoe jÓuw dynamiek werkt. Op deze manier kunnen we dus ook leren welke interventies niet werken en waarom niet.” Hij benadrukt de noodzaak hiervan. “Het is twee voor twaalf. De technologie wordt steeds mooier, met VR, dashboards, eye tracking, maar ik zeg oprecht dat het niet beter gaat werken in het verkrijgen van betere leeruitkomsten. Niet zolang het diagnostisch element ontbreekt.”

Hendrik Drachsler van de Open Universiteit sluit zich daar tot op zeker hoogte bij aan. “De meeste dashboards zijn onvoldoende geëvalueerd en voldoen niet aan educatieve processen, blijkt uit onderzoek. Het bewijs dat het dashboard daadwerkelijk impact heeft op het leerproces ontbreekt, want juist dat wordt niet onderzocht.” Hij ziet nog te veel instellingen die denken dat ze er met het aanbieden van een dashboard wel zijn. “Het is niet genoeg om studenten of docenten een dashboard voor te schotelen, het gaat om een permanent begeleidingsproces. Alleen zo verbeter je de kwaliteit van het onderwijs.” Het SHEILA-project (Supporting Higher Education to Integrate Learning Analytics) ondervroeg vierduizend Europese studenten, docenten en experts over hun verwachtingen van learning analytics. Dat leverde een framework op dat het implementatietraject ondersteunt.

Sommige zaken zijn ook gewoon nog onbekend. Onderzoek naar learning analytics was de eerste jaren vooral gericht op de techniek. Inmiddels is er meer aandacht voor interventies. Justian Knobbout van de Hogeschool Utrecht is benieuwd wat de volgende fase zal opleveren. Hij zegt: “Het learning analytics-proces is cyclisch. Er zijn nog weinig voorbeelden van onderzoek naar het opnieuw uitvoeren van de cyclus. Wat gebeurt er wanneer we studenten blijven volgen en oude analytics gebruiken om opnieuw te meten, zodat we gerichtere interventies kunnen doen?” Onderzoek naar datagebruik in het onderwijs is nog jong. Veel zal in de komende jaren moeten blijken.

Om substantiële stappen te zetten met datagebruik in het Nederlandse hoger onderwijs is de zone ‘Veilig en betrouwbaar benutten van studiedata’ van het Versnellingsplan onderwijsinnovatie met ICT in het leven geroepen. Ook de SIG Learning analytics speelt een belangrijke rol bij de kennisdeling onder de instellingen. In de toekomst ziet SURF mogelijk een rol voor zichzelf weggelegd als de aanbieder van een basissysteem voor learning analytics dat voldoet aan de AVG.

⁴ Deze blogpost geeft een korte uitleg over de Bayesian Bandit: https://medium.com/@haydar_ai/learning-data-science-day-20-bayesian-bandit-problem-f7e87acfc2b2

BEGINNEN MET LEARNING ANALYTICS

Beginnen met learning analytics is vooral een kwestie van doen. 10 tips voor instellingen die willen experimenteren met learning analytics.

1. Top down of bottom up? Voor beide is iets te zeggen. Begin klein, raden een aantal de kundigen aan, met gratis software en kleine onderzoeksvragen. Maar, zeggen anderen, een groot project heeft meer impact en biedt meer mogelijkheden om learning analytics binnen de organisatie geaccepteerd te krijgen.
2. Begin met een vraag. Welk probleem zou je willen oplossen, of wat zou je willen weten over je onderwijs? Maak het proces interactief. Vraag aan docenten en studenten welke vragen ze met behulp van learning analytics opgelost willen zien. Tinne De Laet van KU Leuven zegt: "Wij vroegen de studieloopbaanbegeleiders wat we zouden kunnen maken. Daarop zijn we met iets gekomen dat ze niet verwacht hadden, maar waar ze zich wel heel erg thuis in voelen."
3. Betrek de ICT-afdeling al vanaf het begin van je project.
4. Regel de zaken op het gebied van privacy en ethiek vooraf. Betrek juridische specialisten voor advies.
5. Vermijd jargon. Studenten van KU Leuven weten bijvoorbeeld niet dat ze aan learning analytics doen, of dat ze gebruikmaken van learning dashboards.
6. Wees duidelijk tegenover de studenten over welke data je analyseert en waarom. Laat ze merken dat de data van hen zijn. Zorg voor vertrouwen in het systeem.
7. Benoem een toezichthouder op centraal niveau. Leiderschap is van groot belang. Instellingen die met succes experimenteren, beschikken vaak over een sterke leidersfiguur, die de positie heeft om learning analytics binnen de gehele instelling te introduceren.
8. Maak bekend dat je met learning analytics bezig bent. Binnen instellingen zijn vaak al verschillende initiatieven waar lang niet altijd van iedereen van op de hoogte is en daardoor worden kansen op samenwerking gemist.
9. Neem de tijd. Het koppelen van alle verschillende databases, het goed regelen van privacyzaken en het vinden van technische oplossingen kost tijd.
10. Een belangrijke uitdaging is de cultuuromslag: is men klaar om beslissingen te nemen op basis van data? Geef de medewerkers tijd om aan het idee te wennen. Ondersteun ondertussen de early adopters.

OPSCHALEN VAN LEARNING ANALYTICS

We hebben experts gevraagd naar hun beste adviezen om learning analytics een goed verankerd onderdeel van de organisatie te maken. 15 tips voor het opschalen van learning analytics.

1. Zorg dat het management van de universiteit op het hoogste niveau aan boord is.
2. Neem daarin ook medezeggenschap mee: de studentenraden en OR.
3. Laat het project besturen door een stuurgroep met een brede vertegenwoordiging van stakeholders in een instelling: wetenschappers, studenten, studieadviseurs, IT-leveranciers, beleidsmedewerkers.
4. Maak een jaarkalender van strategisch projecten binnen de organisatie, zodat je op tijd aan kunt haken op actuele beleidsthema's.
5. Lever niet alleen inzichten op beleidsvragen, maar denk ook actief mee in werkgroepen over adviezen die voortvloeien uit beleid.
6. Laat leden van het team met elkaar bedenken waar ze aan willen werken en in welke volgorde, waarbij je ook de prioriteiten van de instelling meeneemt.
7. Leg in een Code of Practice vast voor welke doeleinden je analyses wel of niet zult gaan gebruiken, op welke grondslag dit gebeurt, wie toegang heeft tot de data, wie eindverantwoordelijk is, welke data je precies hebt verzameld, voor welke beleidsadviezen het al gebruikt is, hoe lang de data bewaard gaat worden, welke rechten studenten hebben, en hoe omgegaan met bijzondere persoonsgegevens. Koppel dit aan ethische uitgangspunten die de instelling heeft.
8. Houd er bij het eerste ontwerp al rekening mee dat het schaalbaar moet zijn.
9. Start met de data die er zijn. Die zijn intrinsiek heel schaalbaar.
10. Leg goed vast welke datasets gebruikt worden, wat daarin staat, wanneer ze geleverd worden, wie daarvoor verantwoordelijk is.
11. Maak gebruik van een stijlgids waarin precies wordt afgesproken hoe variabelen in datasets heten en hoe code wordt gedocumenteerd.
12. Automatiseer zo veel mogelijk uitvoerend werk waar voor medewerkers maar weinig plezier aan beleefd wordt, zoals het inlezen van bestanden.
13. Test elke stap die databewerking doorloopt met behulp van automatische testscripts. Daarmee blijft de kwaliteit van je data op orde en ook de betrouwbaarheid van analyses.
14. Blijf de stakeholders bij het project betrekken. De studieloopbaanbegeleiders van KU Leuven kunnen zelf de inhoud van de dashboards bepalen. In de praktijk passen ze de teksten niet veel aan, maar ze voelen hierdoor wel eigenaarschap over het project en dat zorgt voor acceptatie.
15. Investeer in communicatie. Alle stakeholders moeten worden voorzien van de resultaten.

COLOFON

Auteur

Marjolein van Trigt
Jocelyn Manderveld

Projectleiding**Vormgeving**

Vrije Stijl, Utrecht

Fotografie cover

Christina Morillo / pexels.com

Januari 2018

Copyright

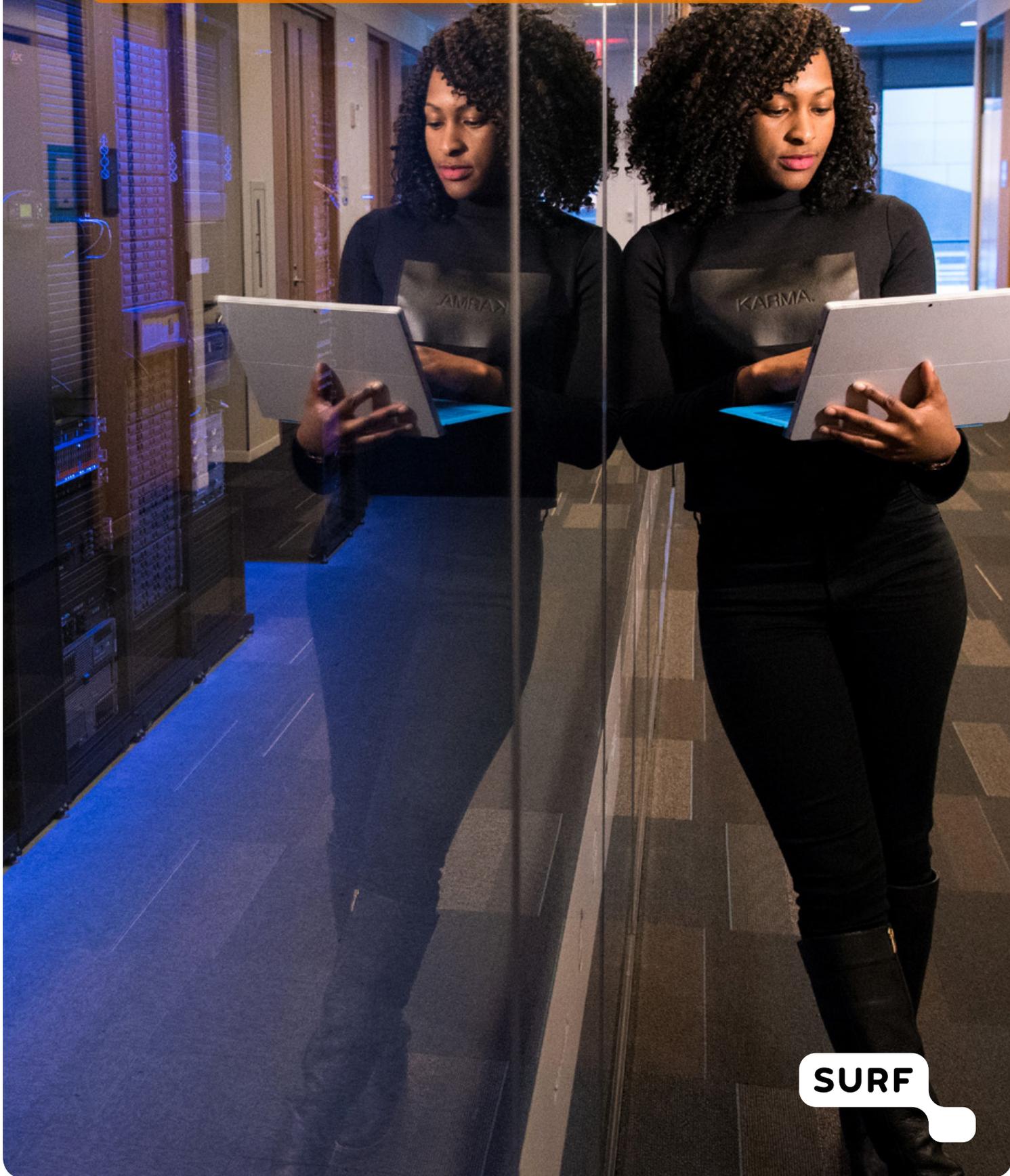
CC BY 4.0

Beschikbaar onder de licentie Creative Commons Naamsvermelding 4.0
Internationaal. <https://creativecommons.org/licenses/by/4.0/deed.nl>

ARTIKEL 7

LEARNING ANALYTICS IN
HET ONDERWIJS: EEN ONDER-
WIJSKUNDIG PERSPECTIEF

HOE DATA DE KWALITEIT VAN HET ONDERWIJS KUNNEN VERBETEREN



SURF

INHOUD

Inleiding	3
1. De potentie van learning analytics voor het Nederlandse hoger onderwijs	5
<i>Praktijkvoorbeeld: KU Leuven</i>	<i>7</i>
<i>Praktijkvoorbeeld: Vrije Universiteit</i>	<i>8</i>
<i>Praktijkvoorbeeld: SURF learning analytics experimenten</i>	<i>9</i>
2. Uitdagingen van learning analytics	10
<i>Praktijkvoorbeeld: Open Universiteit</i>	<i>14</i>
3. Beginnen met learning analytics: 10 tips	16
4. Opschalen van learning analytics: 15 tips	17

INLEIDING

De afgelopen jaren zijn er in het Nederlandse hoger onderwijs veel inzichten opgedaan over de mogelijkheden van datagebruik voor het onderwijs. Sommige van die inzichten zijn een direct resultaat van experimenten met learning analytics, andere komen voort uit onderzoek óver learning analytics.

Online onderwijs brengt een aantal grote veranderingen met zich mee. Eén ervan is dat docenten minder zicht hebben op wat studenten doen. In een lokaal zien docenten hoe het met de studenten gaat. Het is duidelijk of ze er zijn of niet, of ze meedoen en of ze hun opdrachten hebben gemaakt. In een online onderwijsomgeving ontbreekt dat overzicht grotendeels. Dat is lastig, want iedere docent vindt het belangrijk om de studenten goed te kunnen begeleiden. Het gebruik van data in het onderwijs kan hiervoor een oplossing bieden. De overkoepelende term hiervoor is learning analytics. Learning analytics is het verzamelen, analyseren en rapporteren van data van studenten en hun omgeving om zo het onderwijs en de omgeving waarin dat onderwijs plaatsvindt te begrijpen en te verbeteren.¹

Met learning analytics heeft de docent meer grip op online studiegedrag, zodat hij studenten kan ondersteunen en uitval te voorkomen. Daarnaast helpt learning analytics bij het verbeteren van online materiaal. Data over hoe vaak, hoe lang, wanneer en hoe materiaal wordt gebruikt, geven inzicht in de vraag of het materiaal wel doet waarvoor het is bedoeld. Eigenlijk is learning analytics een logisch gevolg van online onderwijs. Het is een middel om zicht te krijgen op wat er online gebeurt.

Bestaande data koppelen

Datagebruik kan ook breder worden ingezet. Hogeronderwijsinstellingen beschikken over heel veel verschillende databronnen: over de vooropleiding van de student, het oriëntatietraject dat hem naar de instelling bracht, zijn leeftijd, waar hij is opgegroeid, studieresultaten, et cetera. Mogelijk zijn er daarnaast WiFi-trackers binnen de instelling geïnstalleerd, of hangen er warmtesensoren in de collegezalen. Koppel een aantal van zulke data aan elkaar en je verkrijgt inzicht in het leergedrag van de studenten, de kwaliteit van het onderwijs en de effectiviteit van de instelling. Op instellingsniveau biedt learning analytics daarmee nieuwe mogelijkheden om trends te bestuderen, studieuitval te beperken en efficiënter te werken. Informatie die je toch al hebt, kun je gaan koppelen, omdat die informatie tegenwoordig digitaal beschikbaar is. Daardoor heb je bijvoorbeeld de mogelijkheid om risicostudenten te *nudgen*, een zetje te geven. Afhankelijk van het antwoord kun je al dan niet actie ondernemen.

Gericht zoeken naar verbanden

Nog maar enkele jaren geleden leek de belofte van big data dat het verzamelen van zo veel mogelijk data zou leiden tot het vinden van steeds meer verbanden. Nog afgezien van de vraag of dat klopt, is het een werkwijze die slecht samengaat met de AVG en de manier waarop hogeronderwijsinstellingen met hun studenten en medewerkers willen omgaan. Learning analytics kan beter onderwijs opleveren voor de student doordat het meer inzicht oplevert. Dat inzicht komt echter altijd voort uit een concrete vraag. De eerste vraag is dan ook wat je wil bereiken met het gebruik van data. Waar wil je precies achterkomen en waarom? De vervolgvraag is: welke data heb je beslist nodig om een antwoord te vinden

¹ Deze definitie is afkomstig van de Learning Analytics Community Exchange (LACE). Zie ook <http://www.laceproject.eu/faqs/learning-analytics/>

op je vraag? Niet onbegrensd grazen in data, maar gericht zoeken in een geminimaliseerde dataset, is anno 2019 het credo van learning analytics. Met een strak ingericht proces kun je uitstekend tot nieuwe inzichten komen, zo laat het praktijkvoorbeeld van de VU in deze uitgave zien.

Een andere ontwikkeling waarover een paar jaar geleden druk werd gespeculeerd, was dat instellingen met behulp van learning analytics konden gaan voorspellen welke eerstejaars weinig kans zouden maken om de studie te voltooien. In Europa bestaat vooralsnog geen animo om studenten te selecteren op basis van computergegenereerde voorspellingen. Het Nederlandse en Belgische hoger onderwijs is diep doordrongen van het belang van een ethische, transparante omgang met data van studenten en medewerkers. Statistieken kunnen wel worden ingezet om studenten een realistisch beeld te geven van hun studievoortgang, zoals de KU Leuven dat doet.

Leeswijzer

Wat is de potentie van learning analytics voor het Nederlandse hoger onderwijs? Waar hebben studenten en docenten daadwerkelijk wat aan? Daarover gaat hoofdstuk één. Aan de hand van een aantal praktijkvoorbeelden beschrijven we wat Nederlandse en Belgische hogeronderwijsinstellingen na een paar jaar van experimenteren zien als de belangrijkste meerwaarde. Hoofdstuk twee gaat over de didactische, juridische, ethische en technische uitdagingen en over de mogelijke impact op het onderwijs. In dit hoofdstuk worden de inzichten benoemd die voortkomen uit een eerder learning analyticsexperiment dat SURF met de instellingen uitvoerde. Ook laten we zien hoe deze inzichten leiden tot het inslaan van een nieuwe richting, met meer mogelijkheden voor ethisch datagebruik. De whitepaper bevat naast tips om te starten met het gebruik van data in het onderwijs ook tips voor instellingen die de experimentele fase zijn ontstegen en zich willen wagen aan het opschalen van learning analytics.

Verschillende vormen van analytics

De Learning Analytics Community Exchange (LACE) hanteert een brede definitie van *datagebruik in het onderwijs*: het verzamelen, analyseren en rapporteren van data van studenten en hun omgeving om zo het onderwijs en de omgeving waarin dat onderwijs plaatsvindt te begrijpen en te verbeteren. Dat zegt echter nog niets over het type data dat wordt gebruikt. Kijkt men vooruit, terug of kijkt men *real-time* mee?

In deze whitepaper hanteren we consequent de term 'learning analytics' voor alle vormen van analytics. Volgens sommigen zegt *learning analytics* echter alleen iets over wat er tijdens het leerproces plaatsvindt. Learning analytics biedt inzage in het studieproces, met als doel het beter kunnen helpen of bijsturen van de student. Volgens iets bredere definities kunnen voor learning analytics ook *multimodal data* worden gebruikt, zoals data afkomstig van fitbits, WiFi-trackers of detectiepoortjes binnen de instelling.

Student analytics maakt gebruik van bestaande data uit verschillende bronnen. De data worden toegankelijk gemaakt voor docenten en studiebegeleiders om bijvoorbeeld uitspraken te doen over mogelijke

risicostudenten of studenten die meer uitdaging kunnen gebruiken.

Academic analytics gaat om het analyseren van trends en ontwikkelingen via data op instellings- of instituutniveau. Denk aan rendementcijfers, maar ook bijvoorbeeld aan prognoses over de kans dat verschillende groepen studenten een opleiding afronden.

Machine learning is een ontwikkeling die zich nog veelal buiten het onderwijs afspeelt. Het komt nog weinig voor dat hogeronderwijsinstellingen een grote hoeveelheid data aan een algoritme voeren, in de hoop dat het algoritme zelfstandig nieuwe verbanden legt.

Over de precieze invulling van de termen bestaat discussie. Belangrijk is vooral dat er diverse manieren zijn om met data in het onderwijs om te gaan. Het is belangrijk om na te denken op welke manier je learning analytics wilt inzetten: wil je weten wat er gebeurt tijdens het leerproces van de studenten, of wil je trends analyseren aan de hand van bestaande bronnen? Hoe je die van vorm van datagebruik vervolgens noemt, is aan jou.

DE POTENTIE VAN LEARNING ANALYTICS VOOR HET NEDERLANDSE HOGER ONDERWIJS

LACE verzamelt wereldwijd bewijzen voor de effectiviteit van de inzet van learning analytics in het onderwijs. In de [LACE Evidence Hub](#) is een grote hoeveelheid mooie voorbeelden te vinden van instellingen die met learning analytics aan de slag zijn gegaan. Waar ligt de meerwaarde voor het Nederlandse hoger onderwijs? In interventies en onderwijs op maat? Of moet er op de korte termijn toch meer naar verbetering van het onderwijsmateriaal worden gekeken?

Justian Knobbout onderzoekt de effecten van learning analytics voor de Hogeschool Utrecht. Hij ziet een grote behoefte aan *low level analytics*. Studenten en docenten zijn vaak geholpen met eenvoudige dashboard, die inzicht geven in wat hun studenten doen in de digitale leeromgeving. “Zet learning analytics in voor het evalueren van het gebruik van onderwijsmateriaal,” zegt hij. “In het ideale geval krijgen de gebruikers daardoor veel beter inzicht in het leerproces dan nu het geval is. Dat kan zich vertalen in een situatie waarin de docent precies weet hoe hij verschillende studenten moet aansturen, of de student wat de volgende stap in het leerproces zou moeten zijn.”

Maar het ene vak is de andere niet; waar de ene docent bijvoorbeeld behoefte heeft aan een samenvatting van de meest besproken onderwerpen op het online forum, wil de andere docent vooral weten hoe zijn videomateriaal wordt bekeken. “Een systeem voor learning analytics moet flexibel zijn en in die verschillende wensen kunnen voorzien,” zegt Knobbout. Dat is nog niet altijd het geval.

Waar datagebruik in het onderwijs in de breedte aan een gestage opmars bezig zijn, blijft de praktijk in de diepte nog achter op de mogelijkheden, signaleert hij. Zo ziet hij dat instellingen bij het implementeren van een nieuwe digitale leeromgeving (DLO) learning analytics nog beschouwen als een ‘nice to have’. “Het is veel beter om het meteen uit te rollen, in plaats van achteraf te bedenken hoe je data wil onttrekken uit verschillende applicaties die in één omgeving samenwerken. Dat is namelijk bijna onmogelijk. Je moet vanaf de start een infrastructuur voor data-opslag hebben, met de mogelijkheid om die voorlopig nog uit te zetten zolang je er nog niet aan toe komt.”

Binnen het in 2016 opgerichte Amsterdam Center for learning analytics (ACLA), een onderzoeksgroep van de Vrije Universiteit, wordt onderzoek gedaan naar de effectiviteit van de inzet van learning analytics in het onderwijs. Directeur is Chris van Klaveren, tevens hoofddocent Onderwijswetenschappen & Economie aan de VU. Samen met Ilja Cornelisz en Martijn Meeter richtte hij ACLA op, omdat hij inzag dat learning analytics nauwe samenwerking vereist tussen verschillende disciplines. De onderzoeksgroep combineert inzichten uit de wetenschappelijke vakgebieden economie, computerwetenschappen, kunstmatige intelligentie, psychologie en onderwijskunde. Hij is enthousiast over de inzichten die uit de samenwerking voortkomt. “Je hebt niet zelf de kennis om alles op te lossen. Door de dialoog aan te gaan, erken je elkaars blinde vlekken en kun je die aanpakken.”

Learning analytics is in de eerste plaats mensenwerk. “Ik heb nog nooit data een vraag zien beantwoorden en algoritmes ontstaan niet vanzelf, maar worden bedacht en geprogrammeerd door mensen,” zegt Van Klaveren. “De meerwaarde van learning analytics verschilt niet van andere tools die het onderwijs individualiseren of personaliseren: het

heeft meerwaarde als het iets verbetert, of dat nu de leerbeleving, de leeromgeving of de leeruitkomst is." Veelbelovend is volgens hem het real time karakter ervan. "Je monitort de juiste informatie en de lerende kan die informatie op elk moment tot zich nemen. Dat kan in potentie positieve gevolgen hebben."

Vooraf studenten die al betrokken zijn bij het eigen leerproces, hebben veel baat bij learning analytics, blijkt uit onderzoek van Knobbout. "Zij reflecteren op hun leergedrag en daarbij komt een dashboard goed van pas. Maar eerstejaars die nog niet helemaal op hun plek zitten, zijn met andere zaken bezig dan hun leerproces."



PRAKTIJKVOORBEELD

ACTIONABLE FEEDBACK: DATA WAAR DE STUDENT IETS MEE KÁN

KU LEUVEN

De overgang van de middelbare school naar het hoger onderwijs is vaak groot. Daarom besloot de KU Leuven vanaf 2015 te onderzoeken hoe bestaande data kunnen worden ingezet om studenten tijdens deze overgang feedback te geven. Eerstejaars studenten ontvangen nu vier keer per jaar een dashboard dat inzicht geeft in hun leer- en studievaardigheden en academische resultaten. De dashboards geven de studenten ook concrete tips om zo nodig verbeteringen te treffen.

De data zijn afkomstig van twee dataspooren. Allereerst de punten van studenten, afkomstig van tussentijdse toetsen, de examens en de ijkingsstoets, de toets die voor de start van het academiejaar wordt afgenomen. Het tweede dataspoor zijn zelfgerapporteerde gegevens. Alle eerstejaars vullen een gevalideerde vragenlijst in over hun leer- en studievaardigheden, de zogeheten LASSI-vragenlijst (Learning and Study Strategies Inventory). Naar aanleiding van de vragenlijst ontvangen ze aan het begin van het studiejaar een eerste dashboard. De dashboards die ze later in het jaar ontvangen tonen onder meer hun positionering ten opzichte van studiegenoten. Ook krijgen ze te zien hoe studenten die in het verleden een vergelijkbaar aantal punten haalden het ervan af brachten. Tinne De Laet, hoofd van de dienst studentenbegeleiding van de faculteit ingenieurswetenschappen, vertelt: “We geven de studenten informatie die ze voorheen niet hadden, maar die eigenlijk heel gemakkelijk ontsloten wordt uit de databanken van een hogeronderwijsinstelling. Het is geen ingewikkeld algoritme, geen geavanceerde technologie. Iedereen kan dit gemakkelijk implementeren. De echte uitdaging ligt in het ontwerp en de integratie in de begeleiding van de student.” Want wat wil je de studenten precies tonen en hoe spoor je ze aan om actie te ondernemen als dat nodig is?

De Laet: “We proberen studenten advies te geven over iets dat *actionable* is, iets waar ze iets aan kunnen doen. Daarom nemen we het geslacht of de sociaal-economische status niet mee in het model, want daar kan de student niets aan veranderen. We koppelen de student analytics aan tips en verwij-

zingen waarmee ze zich kunnen verbeteren. Over leer- en studeervaardigheden biedt KU Leuven een heleboel trainingen aan.”

De aanpak blijkt succesvol. Naar aanleiding van het dashboard over de leer- en studeervaardigheden komen de studenten vaker op gesprek bij de studieloopbaanadviseur, nog vroeg in het semester. Ook geven ze aan dat ze de learning analytics en de tips nuttig vinden.

De grote ethische uitdaging is om studenten niet in de put te sturen door de statistieken. Daarvoor is intern een aantal richtlijnen opgesteld. Zo worden de studenten in de statistieken opgedeeld in drie groepen, waarbij voor elke groep toch nog studenten succesvol zijn in hun studie. De Laet: “Het kan altijd zijn dat de student er in slaagt om iets te realiseren dat nog nooit iemand hem of haar heeft voorgedaan, dus zullen we nooit communiceren dat dit niet mogelijk zou kunnen zijn. Het dashboard kan een hard signaal zijn, maar wel een genuanceerd signaal.” Het is eerlijk over de onzekerheid die op de voorspelling zit. En er wordt nooit gesproken over kansen. “Het is flauwekul om te zeggen dat iemand tachtig procent kans heeft om de bachelor in drie jaar te halen,” benadrukt De Laet. “We zeggen wel: van de studenten die dezelfde punten hadden als jij, heeft tachtig procent zijn bachelor in drie jaar gehaald. Maar dit is op basis van wat wij van jou weten en wij hebben geen compleet beeld van wie jij bent.”

Studenten zijn over het algemeen enthousiast over het project. Dat blijkt wel uit het feit dat de dashboards ondertussen op hun verzoek zijn uitgerold over de gehele bachelor. Sowieso zet de opmars van learning analytics binnen de KU Leuven gestaag door. Vanuit de faculteit ingenieurswetenschap is het project uitgerold naar de onderwijsgroep Wetenschap en Technologie. Afgelopen academiejaar was het de beurt aan 26 opleidingen binnen KU Leuven. Er zijn 12.000 studenten bereikt. De volgende stap is een instellingsbreed project.

PRAKTIJKVOORBEELD

ONVERWACHT PROFIJT OP INSTELLINGSNIVEAU

VRIJE UNIVERSITEIT

Als onderdeel van een groot project om de gehele onderwijsadministratie te digitaliseren, startte Theo Bakker, destijds werkzaam bij Deloitte, een deelproject met als doel om met gecombineerde data studenten meer inzicht te bieden in hun studievoortgang. Met Deloitte ontwikkelde de VU een model dat een prognose gaf van de kans dat eerstejaars hun eerste studiejaar zouden halen. Het model werd ingezet in een pilot om studieadviseurs en tutores een extra handvat te geven bij de studiebegeleiding. De VU ontwikkelde hiervoor een uitgebreide Code of Practice voor studenten. De pilot was geen onverdeeld succes, maar de dataset levert nog wekelijks interessante inzichten op.

In totaal heeft de VU een dataset van ruim 1500 kenmerken per student, afkomstig uit 80 databronnen. Bakker is inmiddels strategisch beleidsmedewerker en projectleider Student Analytics bij de VU en aanvoerder van de zone 'Veilig en betrouwbaar benutten van studiedata' van het Versnellingsplan Onderwijsinnovatie met ICT. Hij vertelt: "Wij kijken naar transactionele data, dus data die al aanwezig zijn in administraties. Van welke scholen komen de studenten, wanneer hebben ze zich aangemeld voor hun studie, zijn ze bij open dagen of meeloopdagen geweest, hebben ze matching gedaan? Ook gebruiken we data van de taaltoets die we al onze studenten afnemen. En we gebruiken achtergrondkenmerken, zoals vooropleiding, leeftijd, geslacht. Vervolgens bekijken we welke resultaten ze halen aan de VU."

Afhankelijk van de bron kan Bakker elke week nieuwe gegevens leveren. Die snelheid noemt hij een belangrijke meerwaarde van learning analytics. "Studieadviseurs komen in actie als een student langskomt. Dit soort instrumenten heeft de belofte dat je proactief

kunt werken en bijsturen." Voor de pilot bleek dat niet op te gaan; voor de studieadviseurs voegden de data weinig toe aan wat de student zelf al meldde. Bovendien waren ze bang om de studenten te ontmoedigen met de prognoses. Mede daarom is het begeleiden van individuele studenten op basis van prognoses voorlopig van de baan. Er zijn wel plannen om in 2020 in nauw overleg met studenten, studieadviseurs en tutores een nieuw project op studenten-niveau uit te voeren.

Maar de pilot had een belangrijke nevenvangst. In grote hoeveelheden bleken de data veel inzichten te bieden over de instelling. Bakker: "Opeens ontdekten wij dat van al onze eerstejaars studenten een groot deel al een studie had gedaan en helemaal geen VWO-student was. Dat gaf ons een argument om tegen Den Haag te zeggen dat onze bekostiging wezenlijk lager is dan die van andere universiteiten." Dergelijke inzichten wekten de interesse van het College van Bestuur. Inmiddels vinden er volop beleidsaanpassingen plaats op basis van data. "Bijna elke week doen we nieuwe ontdekkingen," zegt Bakker. "Ik was verrast hoe slecht de VU zichzelf eigenlijk kende. Drie jaar geleden werd veel beleid gemaakt op ideeën. Nu is dat zo veel mogelijk evidence-based."

In 2019 wordt VU Analytics een formeel team. VU Analytics richt zich dan ook op nieuwe beleidsterreinen, zoals werving en de arbeidsmarkt. Ook wil het team meer ad hoc vragen van docenten kunnen beantwoorden. Bakker: "Ik hoop dat er steeds meer onderzoeksgroepen komen die deze data zelf gaan onderzoeken."

PRAKTIJKVOORBEELD

SURF LEARNING ANALYTICS EXPERIMENTEN

EXPERIMENTEN GEVEN DOCENTEN HANDS-ON ERVARING

SURF ontwikkelde samen met enkele onderwijsinstellingen een experimenteeromgeving waarbinnen een aantal instellingen hun eerste schreden op het gebied van learning analytics hebben gezet. In 2016 vonden de eerste experimenten plaats, met docenten van de Vrije Universiteit en Hogeschool Utrecht. In 2017 deden de Vrije Universiteit, Hogeschool Windesheim, Technische Universiteit Eindhoven, Zuyd Hogeschool en ROC Noorderpoort mee. In totaal ging het om twaalf docenten van zes instellingen. De ervaringen van de instellingen staan beschreven in de [geleerde lessen](#).

VU

Ilja Cornelisz, mede-oprichter van Amsterdam Center for learning analytics (ACLA) en docent Statistiek aan de VU, was betrokken bij de ontwikkeling van de experimenteeromgeving voor learning analytics van SURF. Als docent nam hij twee keer deel aan het experiment, om meer inzicht te krijgen in het leerproces en het onderwijsaanbod beter te laten aansluiten op de behoeften van studenten. Hij benadrukt dat learning analytics signaleert, maar niet direct een oplossing biedt. Het is een hulpmiddel om het gesprek aan te gaan met studenten en docenten.

HU

De Hogeschool Utrecht onderzoekt hoe learning analytics kan bijdragen aan het verbeteren van het onderwijsproces. Het experiment was voor drie docenten een manier om learning analytics op een makkelijke manier in de praktijk te brengen. Zij zien het als een mooie manier om het onderwijs te monitoren.

Windesheim

Yvonne van Vooren, docent bij Hogeschool Windesheim, gebruikte learning analytics om meer inzicht te krijgen in het leergedrag en de leervaardigheden van haar studenten. Op basis daarvan ging ze met hen in gesprek over verbeterpunten aan de kant van de student en mogelijke aanpassingen aan de lessen. Ze concludeert dat learning analytics hiervoor een mooie tool biedt.

TU/e

De Technische Universiteit Eindhoven werkte aan een beleidskader over learning analytics en dat maakte deelname aan het experiment interessant. De deelnemende docenten wilden meer inzicht krijgen in het gebruik van het online leer materiaal binnen het vak Calculus, door het verzamelen van zo veel mogelijk data. Dat bleek een arbeidsintensieve en niet erg vruchtbare methode.

Zuyd

Het doel van het experiment van Zuyd Hogeschool was om docenten te laten kennismaken met learning analytics. Deze insteek was wat te vrijblijvend, concludeert de hogeschool, die naar aanleiding hiervan aanbeveelt dat learning analytics vraaggericht moet worden ingezet.

Noorderpoort

ROC Noorderpoort is de eerste mbo-instelling die ervaring opdeed met learning analytics in de experimenteeromgeving van SURF. Het inzetten van learning analytics bij het vak Digitale Vaardigheden leverde docent Jacob Poortstra een schat aan informatie op en hij hoopt op een mogelijk vervolg.



UITDAGINGEN VAN LEARNING ANALYTICS

Didactische uitdagingen

De toepassing van learning analytics begint met een goede onderzoeksvraag, maar dat is niet nieuw. De vraag ‘wat is het beste onderwijs dat ik wil aanbieden met behulp van de mij beschikbare hulpmiddelen?’ is tijdloos. Omdat de hulpmiddelen steeds geavanceerder worden, moet je echter beter nadenken hoe je ze wil inzetten. Begin daarmee lang voordat je onderwijs start, anders blijken data bijvoorbeeld niet beschikbaar, of niet te interpreteren.

Vervolgens is er een aantal randvoorwaarden nodig om didactische meerwaarde te bereiken. De Schot Niall Sclater, als consultant verbonden aan Jisc, de Britse zusterorganisatie van SURF, somt op: “Heb je mensen in huis die de data begrijpen, die de databronnen kennen en ze kunnen interpreteren, die voorspellende modellen kunnen maken of statistische analyses kunnen doen, en nog belangrijker, als al het andere achter de rug is, interventies kunnen plegen met de studenten? Je kunt statistieken produceren tot je een ons weegt, maar als je niet in staat bent om interventies te plegen, dan kun je er net zo goed niet eens aan beginnen.”

De docenten hoeven zelf niet over ál deze vaardigheden te beschikken, maar een bepaalde handigheid met ICT is, naast kritisch denkvermogen en het vermogen om analyses te interpreteren, wel een vereiste. Didactische uitdagingen zijn niet los te zien van de technologische uitdagingen. Beginnen met learning analytics betekent allereerst dat de digitale leeromgeving naar behoeven functioneert en dat docenten beschikken over voldoende ICT-vaardigheden. Veel instellingen zijn nog niet zo ver.

Juridische uitdagingen

De komst van de Algemene Verordening Gegevensbescherming (AVG) heeft veel instellingen huiverig gemaakt voor het gebruik van learning analytics. Toch is er strikt genomen niet zo veel veranderd in de mogelijkheden op het gebied van datagebruik. SURF probeert op verschillende manieren om de onzekerheden van instellingen weg te nemen, bijvoorbeeld door middel van een onderzoek naar passende maatregelen bij de toepassing van learning analytics, dat in het voorjaar van 2019 wordt afgerond.

In vergelijking met de Wet bescherming persoonsgegevens (Wbp) moeten instellingen die aan learning analytics doen onder de AVG iets meer aan interne en externe verantwoording doen, zegt Niels Westerlaken, IT- en privacy-advocaat bij Project Moore Advocaten, dat het onderzoek uitvoert. “De andere stappen zijn vrijwel hetzelfde gebleven, maar juist van deze stap schrikken projectverantwoordelijken nog wel eens, omdat je daadwerkelijk moet laten zien wat je doet.” De vraag welke maatregelen vervolgens passend zijn, is niet eenduidig te beantwoorden, omdat de impact van een onderzoek waarin learning analytics worden ingezet flink kan verschillen. Westerlaken: “Er is een aantal algemene beginselen waaraan je altijd moet voldoen: genoeg informatie verstrekken, veel controle geven over de data, duidelijk maken wat je wil gaan doen. Bij projecten die impact hebben op individuen, waarin je misschien ook data gaat samenvoegen om profielen te maken, zou het vragen van toestemming een passende maatregel kunnen zijn.”

Een instelling zou kunnen redeneren dat het gebruik van data het onderwijs altijd ten goede komt en daarmee past binnen de wettelijke plicht. Met die redenering maakt hij korte metten. “Instellingen zullen een goed doel hebben met de onderzoeken die ze doen,

maar dat is geen vrijbrief om learning analytics direct toe te passen. Er zal namelijk een belangenafweging moeten plaatsvinden tussen de doelen van de instellingen en de privacy van individuen.”

Voor juristen zal een eerste uitspraak van een rechtbank op dit onderwijsterrein heel interessant zijn, omdat het meer duidelijkheid biedt over hoe de rechtelijke macht de wet interpreteert. Of een maatregel ‘passend’ is, is namelijk geen wet van meden en perzen. “De verwachtingen van degene over wie het gaat moeten aansluiten op wat er daadwerkelijk gebeurt,” legt Westerlaken uit. “Over vijf of tien jaar verwachten studenten wellicht al dat er learning analytics worden toegepast. Dan zul je een andere schaal van maatregelen nemen dan in de fase waarin het redelijk nieuw is. Nu zul je echt goed je best moeten doen om het voordeel ervan duidelijk te krijgen. Niet om jezelf in te dekken, maar vooral om het vertrouwen te krijgen van degene van wie je de gegevens verwerkt.”

De instelling heeft de verplichting om verantwoording af te leggen over welke gegevens worden verzameld en met welk doel dat gebeurt. Op dit moment komen de verwachtingen van studenten niet altijd overeen met de realiteit, met name als het gaat om het verzamelen van data. Dat men bijvoorbeeld het onderwijs verbetert aan de hand van de proeftoetsen in de digitale leeromgeving, dat verwachten veel studenten wel. Maar dat daarvoor bijvoorbeeld op individueel niveau wordt gemonitord welke fouten er worden gemaakt, is hen vaak veel minder duidelijk.

Westerlaken raadt instellingen aan om experimenten met learning analytics goed te documenteren. Voor verantwoording naar buiten en binnen, maar ook om de resultaten te delen buiten de eigen instelling. Hoe pas je de grondslagen toe? Welke doelen zijn vooraf vastgelegd? “Als je dat met andere instellingen kan delen, werk je toe naar een standaard. Dat is voor iedereen leerzaam.” Denk bijvoorbeeld aan uitkomsten van [Data Protection Impact Assessments](#), de [Code of conduct](#) van Jisc en de [Code of Practice](#) van de VU.

Privacy en andere ethische uitdagingen

Learning analytics zal een vanzelfsprekend onderdeel worden van de leeromgeving van een hogeronderwijsinstelling. Dat is de verwachting van Hendrik Drachslers, professor aan het Welten-instituut van de Open Universiteit en lid van het bestuur van de [Society of Learning Analytics Research \(SoLAR\)](#). In 2015 zei hij: “Ik denk dat we over vijf à tien jaar niet meer over learning analytics praten, maar het gewoon gebruiken. De docent haalt dan een rapportage uit het systeem om te gebruiken in het vak, zonder er verder bij stil te staan dat hij aan learning analytics doet.” In 2018 gelooft hij dat nog steeds, “het duurt alleen allemaal iets langer dan ik had verwacht vanwege de nieuwe privacywetgeving, maar wij zien meer en meer learning analytics in het dagelijks gebruik.”

Een van Drachslers grootste nachtmerries is dat een van zijn studenten naar de rechter stapt omdat hij bezwaar maakt tegen het gebruik van zijn gegevens voor learning analytics. Tegelijkertijd zou die student waarschijnlijk een hoog cijfer krijgen, want Drachslers ziet in de toepassing van learning analytics onder meer een geschikte manier om studenten kritische vragen te leren stellen over datagebruik in het algemeen. *Data literacy*, noemt hij dat. “In het Duits zou het *Daten-Mündigkeit* heten. Dat gaat nog wat verder dan literacy. Een data-mündig mens is iemand die zich heeft laten informeren over het gebruik van zijn data, die daartegen in opspraak komt en daar ook recht toe heeft. Ik zie *trusted learning analytics*, zoals ik het noem, als een mogelijkheid om jonge mensen daarin te trainen.”

Hogeronderwijsinstellingen die learning analytics willen toepassen, zullen allereerst een visiedocument moeten opstellen en beleid maken, zegt Drachslers. Transparantie staat daarbij voorop. Welke gegevens worden verzameld? Wie heeft er toegang tot de gegevens? Welke mogelijkheden heeft de eigenaar ervan? Drachslers: “Een visie vraagt om een innovatie-

programma om mensen bewust te maken van waar ze naartoe willen en de instrumenten waarmee ze dat willen bereiken. Vaak is daar verandermanagement voor nodig. Er moet nog veel training voor docenten plaatsvinden. Ook studenten hebben meer kennis nodig. Learning analytics vraagt om een langdurig organisatieproces.”

Er is een aantal hulpmiddelen beschikbaar die instellingen ondersteunen bij een transparante omgang met data. Zo is er de DELICATE checklist, door Drachsler en anderen ontworpen binnen LACE. Jisc onderscheidt maar liefst **86 ethische en juridische kwesties** rondom learning analytics. Niall Sclater, die de lijst opstelde, zegt: “De allergrootste uitdaging is de cultuurinstelling; zijn de mensen er klaar voor om beslissingen te nemen op basis van data?”

Technologische uitdagingen

De grootste technische uitdaging van learning analytics is het omgaan met grote hoeveelheden data uit verschillende bronnen. Data kunnen voortkomen uit een studentinformatiesysteem (cijfers en leerpaden), de digitale leeromgeving (leeractiviteiten), maar ook bijvoorbeeld aanwezigheidsdata kunnen een databron vormen. Data worden verzameld in een zogenaamde Learning Record Store. In deze store zijn de data gestructureerd opgeslagen zodat ze kunnen worden gebruikt voor analyse en visualisatie. Om de verschillende bronsystemen waaruit de data afkomstig zijn met de Learning Record Store te laten communiceren, is het gebruik van een protocol nodig. Twee protocollen vechten om de standaard in de markt te worden: xAPI en de IMS Caliper Analytics.

De technologie voor learning analytics heeft zichzelf redelijk uitgekristalliseerd, maar twee ontwikkelingen zorgen voor nieuwe vraagstukken. Enerzijds verbreedt de doelgroep van learning analytics zich: naast studenten, docenten en studiebegeleiders willen onderzoekers de onderwijsdata inzetten en ook op instellingsniveau neemt de interesse in data-analyse toe. Alle doelgroepen willen hun eigen vragen met behulp van learning analytics beantwoord zien. Ook de hoeveelheid databronnen wordt alleen maar groter. Studenten leven niet in een digitale leeromgeving, maar vooral ook daarbuiten. Door data uit de digitale leeromgeving te koppelen aan additionele databronnen, zoals gezondheidstrackers en aan bestaande gegevens, ontstaat een completer beeld en is het mogelijk om een beter antwoord te geven op een onderzoeksvraag.

Tegelijkertijd is er de verplichting om aan de AVG te voldoen. Dat vraagt onder meer om doelmatigheid: je verzamelt alleen data voor een vooraf vastgesteld doel. Probleem is echter dat er met de huidige technologie al snel te veel data worden opgeslagen.

Aanpassingen aan de learning analytics-architectuur van SURF

Deze ontwikkelingen maken het noodzakelijk om de learning analytics-architectuur van SURF te veranderen. Drie overwegingen leiden momenteel tot aanpassingen:

- 1. Eigenaarschap.** De gegevens worden verzameld van de student, voor de student. De student is in controle over zijn persoonlijke data en kan zijn persoonlijke data met een docent of studiebegeleider delen. Geanonimiseerde data, die niet te herleiden zijn tot de individuele student, kunnen worden gebruikt door anderen in de instelling om hun learning analytics-vragen te beantwoorden.
- 2. Dataminimalisatie.** De gegevens worden specifiek voor een bepaalde vraag verzameld. Stel dat een docent wil weten hoe vaak zijn video wordt bekeken. Is het dan relevant dat hij ook ziet dat een student er om 3 uur 's nachts en om 8 uur 's ochtends naar keek? Eigenlijk niet. Dus moet er alleen worden geteld en mogen er geen tijdstippen worden geregistreerd. De uitdaging is om precies genoeg data te verzamelen voor het goed kunnen beantwoorden van de vraag. Dit is de doelbinding², die het mogelijk maakt om learning analytics binnen de AVG uit te voeren. Dataminimalisatie is één van de belangrijkste

privacy-by-design-principes die SURF wil hanteren. Andere privacy-by-design-principes die SURF inzet, zijn het scheiden van de data, het zoveel mogelijk anonimiseren van de data en het abstraheren van de data, zoals in het voorbeeld hierboven.

3. Vertrouwen en transparantie. Door studenten controle te geven over hun data en alleen data te verzamelen die nodig is voor het beantwoorden van een specifieke vraag, is de basis gelegd voor een nieuw platform, dat is gebaseerd op vertrouwen. Daarnaast is er transparantie nodig over de te gebruiken vragen en de algoritmes die de visualisaties bij de vraag verzorgen (open algoritmes). Ook wil SURF volledig transparant zijn over de werking van het nieuwe platform (open architectuur). Is hieraan voldaan, dan verwacht SURF dat learning analytics een nieuwe impuls kan krijgen. Bij het ontwikkelen van de vragen en de open algoritmes worden experts in het veld betrokken, zodat voor iedere stakeholder duidelijk is waar de vraag en het antwoord in de visualisatie vandaan komen. Er komt een pilot met een platform waarop onderzoekers en ontwikkelaars zelf onderzoeksvragen formuleren en databronnen koppelen, al dan niet met ondersteuning van ICT-ers van de instelling. Het platform biedt verschillende services, specifiek gericht op de vragen die de verschillende stakeholders hebben. De services kunnen tussen instellingen worden gedeeld, waardoor iedere instelling baat heeft bij nieuwe ontwikkelingen.

Deze overwegingen hebben een grote impact op de architectuur. De belangrijkste aanpassing is een nieuwe beveiligingslaag voor gecontroleerd databeheer. Deze laag zorgt ervoor dat altijd aan de AVG wordt voldaan. Als een stakeholder een vraag formuleert en gebruikmaakt van een service op het platform, checkt deze laag of het gebruik valt binnen het doel. Voor vragen waarvoor alleen geanonimiseerde gegevens nodig zijn en voor sommige gepersonaliseerde vragen waar de instelling een gerechtvaardigd belang heeft, zal het systeem automatisch toestemming geven. Voor andere vragen met een specifiek doel of gebruik van een specifiek data type kan een consent-scherm voor de student nodig zijn. Er wordt dus niet globaal toestemming gevraagd voor het gebruik van gegevens, maar voor specifieke vragen op het moment dat het zich voordoet. In 2019 zullen de eerste pilots met de nieuwe architectuur van start gaan.

De Open Universiteit houdt zich met dezelfde overwegingen bezig als SURF, zo blijkt uit het praktijkvoorbeeld. De ontwikkeling van een systeem voor *trusted learning analytics* vertoont grote gelijkenissen met de architectuur waaraan SURF bouwt.



² Volgens de AVG moet de verwerking van persoonsgegevens gebonden zijn aan specifieke verzameldoelen. Dit heet doelbinding. Zie voor meer informatie de Handleiding Algemene verordening gegevensbescherming en Uitvoeringswet Algemene verordening gegevensbescherming: <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/handleidingalgemeneverordeninggegevensbescherming.pdf>

PRAKTIJKVOORBEELD

OP WEG NAAR TRUSTED LEARNING ANALYTICS

OPEN UNIVERSITEIT

De brede uitrol en adoptie van learning analytics in Europese hogeronderwijsinstellingen stagneert door ongerustheid over privacy en ethiek met betrekking tot persoonlijke data en de AVG. Die onzekerheid leidt ertoe dat innovatiemanagers twifelen over implementatie van learning analytics in hun instelling. Om de verworvenheden van tien jaar onderzoek naar learning analytics voor docenten en studenten op een veilige manier beschikbaar te maken, werkt de Open Universiteit (OU) samen met het DIPF - Leibniz Institute for Research and Information in Education in Duitsland - aan een zogenoemde Trusted Learning Analytics[1] (TLA) infrastructuur. Deze infrastructuur pakt het huidige gebrek aan vertrouwen en transparantie in learning analytics aan. Hendrik Drachsler, een van de initiatiefnemers voor TLA, zegt: "Het TLA consortium probeert het 'contract' tussen studenten en onderwijsaanbieders te vernieuwen, niet alleen door een hoog vertrouwensniveau te genereren, maar ook door de datawijsheid en betrokkenheid van studenten te vergroten. Het ontwerp van de TLA-infrastructuur en -tools is gebaseerd op ontwerpprocessen die hetzelfde belang toekennen aan ethische en privacy-onderwerpen als aan functionele eisen. Het resultaat is een systeem dat niet alleen technische, maar ook ethische en humane doelen bereikt."

Trusted learning analytics moet worden ontwikkeld vanuit een gebruikersperspectief. De gebruikers zouden volledige controle moeten hebben over hun data en geen enkele wantrouwen moeten voelen ten opzicht van het systeem. Om dit niveau van vertrouwen te bereiken, is het cruciaal om zo veel mogelijk 'black box' data en systemen te veranderen in een 'white box'. Het consortium probeert zo veel mogelijk open te zijn over de toegepaste algoritmes en transparant over de gebruikte methodieken. Een belangrijk doel van een TLA-systeem is het geven van feedback, in plaats van dat het wordt gebruikt voor

het maken van beslissingen. Het TLA-systeem zou volledige toegang moeten bieden aan de 'datasubjecten' - degenen wiens data worden geanalyseerd - en hen moeten toestaan om data te corrigeren, becommentarieer en verwijderen. Ook moeten zij zelf kunnen beslissen wie er toegang heeft tot persoonlijke data. In huidige learning analytics infrastructuur is dit vaak niet het geval. Drachsler: "Voor de OU legt de TLA-aanpak de basis voor een nieuwe manier van omgaan met onderwijsdata en het aanbieden van diensten en activiteiten voor stakeholders."

Een TLA-systeem vraagt van het datasubject, dus de student of de docent, om een hoog niveau van vaardigheden, waaronder kritisch denken, het analyseren van informatie en zelfregulering. Het open source TLA-systeem probeert een omgeving te creëren waar ze deze cruciale vaardigheden kunnen trainen.

³ Drachsler, H. (2018). *Trusted Learning Analytics*. *Synergie*, 06, *Fachmagazine für Digitalisierung der Lehre*. <https://synergie.blogs.uni-hamburg.de/ausgabe-06-beitrag-drachsler/>

Impact op het onderwijs

Data beantwoorden geen vragen als er geen goed onderzoeksdesign aanwezig is. Evenmin bieden ze kant-en-klare oplossingen. Het teruggeven van informatie over gedrag, wil nog niet zeggen dat mensen hun gedrag vervolgens gaan verbeteren. Chris van Klaveren van Amsterdam Center for learning analytics (ACLA) zegt: “In de huidige vorm ziet learning analytics er mooi uit en je kunt heel veel ‘voetstapjes’ terugzien, maar ik vermoed dat daarmee alleen nog geen effecten worden bereikt. Als we onderwijs op een effectieve manier willen personaliseren, is het van belang dat we niet alleen onderzoeken wat gemiddeld genomen de beste oplossing is. Wat dat betreft zijn Bayesiaanse Bandit modellen⁴ interessant, omdat deze vertrekken vanuit het idee dat we willen weten hoe jÓuw dynamiek werkt. Op deze manier kunnen we dus ook leren welke interventies niet werken en waarom niet.” Hij benadrukt de noodzaak hiervan. “Het is twee voor twaalf. De technologie wordt steeds mooier, met VR, dashboards, eye tracking, maar ik zeg oprecht dat het niet beter gaat werken in het verkrijgen van betere leeruitkomsten. Niet zolang het diagnostisch element ontbreekt.”

Hendrik Drachsler van de Open Universiteit sluit zich daar tot op zeker hoogte bij aan. “De meeste dashboards zijn onvoldoende geëvalueerd en voldoen niet aan educatieve processen, blijkt uit onderzoek. Het bewijs dat het dashboard daadwerkelijk impact heeft op het leerproces ontbreekt, want juist dat wordt niet onderzocht.” Hij ziet nog te veel instellingen die denken dat ze er met het aanbieden van een dashboard wel zijn. “Het is niet genoeg om studenten of docenten een dashboard voor te schotelen, het gaat om een permanent begeleidingsproces. Alleen zo verbeter je de kwaliteit van het onderwijs.” Het SHEILA-project (Supporting Higher Education to Integrate Learning Analytics) ondervroeg vierduizend Europese studenten, docenten en experts over hun verwachtingen van learning analytics. Dat leverde een framework op dat het implementatietraject ondersteunt.

Sommige zaken zijn ook gewoon nog onbekend. Onderzoek naar learning analytics was de eerste jaren vooral gericht op de techniek. Inmiddels is er meer aandacht voor interventies. Justian Knobbout van de Hogeschool Utrecht is benieuwd wat de volgende fase zal opleveren. Hij zegt: “Het learning analytics-proces is cyclisch. Er zijn nog weinig voorbeelden van onderzoek naar het opnieuw uitvoeren van de cyclus. Wat gebeurt er wanneer we studenten blijven volgen en oude analytics gebruiken om opnieuw te meten, zodat we gerichtere interventies kunnen doen?” Onderzoek naar datagebruik in het onderwijs is nog jong. Veel zal in de komende jaren moeten blijken.

Om substantiële stappen te zetten met datagebruik in het Nederlandse hoger onderwijs is de zone ‘Veilig en betrouwbaar benutten van studiedata’ van het Versnellingsplan onderwijsinnovatie met ICT in het leven geroepen. Ook de SIG Learning analytics speelt een belangrijke rol bij de kennisdeling onder de instellingen. In de toekomst ziet SURF mogelijk een rol voor zichzelf weggelegd als de aanbieder van een basissysteem voor learning analytics dat voldoet aan de AVG.

⁴ Deze blogpost geeft een korte uitleg over de Bayesian Bandit: https://medium.com/@haydar_ai/learning-data-science-day-20-bayesian-bandit-problem-f7e87acfc2b2

BEGINNEN MET LEARNING ANALYTICS

Beginnen met learning analytics is vooral een kwestie van doen. 10 tips voor instellingen die willen experimenteren met learning analytics.

1. Top down of bottom up? Voor beide is iets te zeggen. Begin klein, raden een aantal de kundigen aan, met gratis software en kleine onderzoeksvragen. Maar, zeggen anderen, een groot project heeft meer impact en biedt meer mogelijkheden om learning analytics binnen de organisatie geaccepteerd te krijgen.
2. Begin met een vraag. Welk probleem zou je willen oplossen, of wat zou je willen weten over je onderwijs? Maak het proces interactief. Vraag aan docenten en studenten welke vragen ze met behulp van learning analytics opgelost willen zien. Tinne De Laet van KU Leuven zegt: “Wij vroegen de studieloopbaanbegeleiders wat we zouden kunnen maken. Daarop zijn we met iets gekomen dat ze niet verwacht hadden, maar waar ze zich wel heel erg thuis in voelen.”
3. Betrek de ICT-afdeling al vanaf het begin van je project.
4. Regel de zaken op het gebied van privacy en ethiek vooraf. Betrek juridische specialisten voor advies.
5. Vermijd jargon. Studenten van KU Leuven weten bijvoorbeeld niet dat ze aan learning analytics doen, of dat ze gebruikmaken van learning dashboards.
6. Wees duidelijk tegenover de studenten over welke data je analyseert en waarom. Laat ze merken dat de data van hen zijn. Zorg voor vertrouwen in het systeem.
7. Benoem een toezichthouder op centraal niveau. Leiderschap is van groot belang. Instellingen die met succes experimenteren, beschikken vaak over een sterke leidersfiguur, die de positie heeft om learning analytics binnen de gehele instelling te introduceren.
8. Maak bekend dat je met learning analytics bezig bent. Binnen instellingen zijn vaak al verschillende initiatieven waar lang niet altijd van iedereen van op de hoogte is en daardoor worden kansen op samenwerking gemist.
9. Neem de tijd. Het koppelen van alle verschillende databases, het goed regelen van privacyzaken en het vinden van technische oplossingen kost tijd.
10. Een belangrijke uitdaging is de cultuuromslag: is men klaar om beslissingen te nemen op basis van data? Geef de medewerkers tijd om aan het idee te wennen. Ondersteun ondertussen de early adopters.

OPSCHALEN VAN LEARNING ANALYTICS

We hebben experts gevraagd naar hun beste adviezen om learning analytics een goed verankerd onderdeel van de organisatie te maken. 15 tips voor het opschalen van learning analytics.

1. Zorg dat het management van de universiteit op het hoogste niveau aan boord is.
2. Neem daarin ook medezeggenschap mee: de studentenraden en OR.
3. Laat het project besturen door een stuurgroep met een brede vertegenwoordiging van stakeholders in een instelling: wetenschappers, studenten, studieadviseurs, IT-leveranciers, beleidsmedewerkers.
4. Maak een jaarkalender van strategisch projecten binnen de organisatie, zodat je op tijd aan kunt haken op actuele beleidsthema's.
5. Lever niet alleen inzichten op beleidsvragen, maar denk ook actief mee in werkgroepen over adviezen die voortvloeien uit beleid.
6. Laat leden van het team met elkaar bedenken waar ze aan willen werken en in welke volgorde, waarbij je ook de prioriteiten van de instelling meeneemt.
7. Leg in een Code of Practice vast voor welke doeleinden je analyses wel of niet zult gaan gebruiken, op welke grondslag dit gebeurt, wie toegang heeft tot de data, wie eindverantwoordelijk is, welke data je precies hebt verzameld, voor welke beleidsadviezen het al gebruikt is, hoe lang de data bewaard gaat worden, welke rechten studenten hebben, en hoe omgegaan met bijzondere persoonsgegevens. Koppel dit aan ethische uitgangspunten die de instelling heeft.
8. Houd er bij het eerste ontwerp al rekening mee dat het schaalbaar moet zijn.
9. Start met de data die er zijn. Die zijn intrinsiek heel schaalbaar.
10. Leg goed vast welke datasets gebruikt worden, wat daarin staat, wanneer ze geleverd worden, wie daarvoor verantwoordelijk is.
11. Maak gebruik van een stijlgids waarin precies wordt afgesproken hoe variabelen in datasets heten en hoe code wordt gedocumenteerd.
12. Automatiseer zo veel mogelijk uitvoerend werk waar voor medewerkers maar weinig plezier aan beleefd wordt, zoals het inlezen van bestanden.
13. Test elke stap die databewerking doorloopt met behulp van automatische testscripts. Daarmee blijft de kwaliteit van je data op orde en ook de betrouwbaarheid van analyses.
14. Blijf de stakeholders bij het project betrekken. De studieloopbaanbegeleiders van KU Leuven kunnen zelf de inhoud van de dashboards bepalen. In de praktijk passen ze de teksten niet veel aan, maar ze voelen hierdoor wel eigenaarschap over het project en dat zorgt voor acceptatie.
15. Investeer in communicatie. Alle stakeholders moeten worden voorzien van de resultaten.

COLOFON

Auteur

Marjolein van Trigt

Projectleiding

Jocelyn Manderveld

Vormgeving

Vrije Stijl, Utrecht

Fotografie cover

Christina Morillo / pexels.com

Januari 2019

Copyright

CC BY 4.0

Beschikbaar onder de licentie Creative Commons Naamsvermelding 4.0 Internationaal. <https://creativecommons.org/licenses/by/4.0/deed.nl>

Samen aanjagen van vernieuwing

Universiteiten, hogescholen, mbo-instellingen, onderzoeksinstellingen en universitaire medische centra werken binnen SURF aan ICT-voorzieningen en -innovaties. Met als doel: beter en flexibeler onderwijs en onderzoek. Dat doen we door de best mogelijke digitale diensten te leveren, kennisdeling en -uitwisseling te stimuleren en vooral door steeds te blijven innoveren! Hiermee dragen we bij aan een sterke en duurzame Nederlandse kenniseconomie.

The SURF logo consists of the word "SURF" in white, bold, uppercase letters inside a black rounded rectangle. A black line extends from the bottom right corner of the rectangle, ending in a small black circle.

SURF