

Markt- en technologieverkenning en de ambitie van SURF op het gebied van learning analytics

Gebruik van data in het hoger onderwijs vraagt om aandacht

Het hoger onderwijs heeft zich de afgelopen jaren snel ontwikkeld op het vlak van digitalisering en het inzetten van de daarbij beschikbaar komende data. Innovaties op het gebied van data bieden kansen om onderwijs en studentbegeleiding te verbeteren. Het verzamelen, beheer en gebruik van data kent echter ook risico's, waarvoor maatregelen genomen kunnen en moeten worden. Daarin zijn verschillende organisatorische en technische keuzes te maken.

Data als handelswaar

Tijdens de SURF Onderwijsdagen in 2019 gaf Christien Bok (impulsmanager Onderwijs bij SURF) een presentatie, waarin ze het hoger onderwijs opriep om snel afspraken te maken over open standaarden. De commerciële platformen verzamelen namelijk veel gebruikersdata. Niet alleen om hun dienstverlening te verbeteren, maar ook om te verkopen: die data bieden adverteerders namelijk waardevolle informatie waarmee ze gebruikers heel gericht kunnen benaderen.

Werken aan een veilige en verantwoorde digitale omgeving

Niet veel later kopte de Volkskrant "*Digitalisering bedreigt onze universiteit. Het is tijd om een grens te trekken*". In dit stuk riepen maar liefst 17 rectores magnifici op om als hoger onderwijs gezamenlijk grenzen en voorwaarden te stellen aan het gebruik van commerciële platformen door publieke organisaties. Ze benadrukten dat de sector verantwoordelijkheid moet nemen voor de publieke waarden en dat de onderwijssector zelf moet gaan werken aan een veilige en verantwoorde digitale omgeving.

De rol van SURF

"De grootste uitdaging van dit decennium is het innovatief, veilig en betrouwbaar ondersteunen van datastromen en de cyclus van data verzamelen, (selectief) opslaan, beschikbaar stellen, uitwisselen, selecteren, samenvoegen, verwerken en analyseren", geeft ook de Wetenschappelijke Technische Raad (WTR) van SURF aan. Daarbij spreekt de WTR het vertrouwen uit dat SURF een nieuwe datageoriënteerde infrastructuur gedurende de komende jaren kan realiseren, samen met leden en partners.

Bekeken vanuit studiedata

Het gebruik van (studie)data door platformen in het hoger onderwijs vraagt dus om aandacht. Het is van belang dat data vanuit het hoger onderwijs in een veilige en verantwoorde digitale omgeving worden verzameld, geanalyseerd en teruggekoppeld. Dat geldt ook voor learning analytics.

Het projectteam learning analytics van SURF heeft zich daarom beziggehouden met de vraag in hoeverre SURF volgens de instellingen oplossingen en aanbod zou moeten ontwikkelen als het gaat om studiedata en learning analytics.

Markt- en technologieverkenning learning analytics

Om te kunnen bepalen in hoeverre het aanbod vanuit SURF onderscheidend en aanvullend kan zijn, is het eerst van belang een goed beeld te krijgen van het huidige landschap van technologiecomponenten en aanbieders daarvan. Daarom voerde Verdonck, Klooster en Associates (VKA) een markt- en technologieverkenning uit. In dit onderzoek zijn de wensen van instellingen en het aanbod van de markt bekeken en afgewogen. De centrale vraag in

het onderzoek was: voldoet het bestaande aanbod aan normen en waarden op het vlak van privacy, gegevensbescherming en ethiek?

Het onderzoek is enerzijds bedoeld voor SURF. Het geeft ons input om te bepalen wat SURF zal ontwikkelen en aanbieden als het gaat om het gebruik van learning analytics. Anderzijds is het onderzoek bedoeld voor instellingen. Het laat hen zien welke ethische vraagstukken op welke plaats in de learning analyticsarchitectuur een rol hebben en waar privacy- en gegevensbescherming een plek heeft.

Richtlijnen

In het rapport worden richtlijnen voor een studiedata-infrastructuur behandeld en getoetst bij instellingen en marktpartijen. Het SURF-projectteam learning analytics heeft deze richtlijnen in eerste instantie opgesteld zodat we een duidelijk kader hebben voor wanneer wij zelf een product of dienst willen gaan ontwikkelen.

De richtlijnen zijn ook als aanbeveling te gebruiken door instellingen, die aan de slag gaan met een infrastructuur voor studiedata. De richtlijnen geven aan waar een instelling rekening mee kan houden en welke afwegingen van belang zijn bij het selecteren van een aanbieder of bij het zelf opzetten van een infrastructuur.

Belangrijke bevindingen

Enkele belangrijke bevindingen en conclusies geven aan SURF een potentiële rol kan spelen op het gebied van learning analytics in het hoger onderwijs. Bijvoorbeeld als het gaat om de transparantie van algoritmes, het (stimuleren van) gebruik van standaarden voor data-uitwisseling, maar ook als het gaat om ethische standaarden.

- *Leveranciers van oplossingen die gebruik maken van standaardalgoritmes zijn niet altijd transparant over de werking van het algoritme. Dit is een risico voor instellingen die gebruik willen maken van dergelijke oplossingen.*
- *In de markt bestaan diverse standaarden voor data-uitwisseling, waarbij voor onderwijsapplicaties de xAPI-standaard belangrijk is. De onderzochte oplossingen ondersteunen maar in beperkte mate de xAPI-standaard.*
- *Leveranciers hanteren uiteenlopende ethische standaarden.*

Visie SURF op het rapport

Deze technologie- en marktverkenning en het bijbehorende rapport vormen het startpunt van het peilen van de behoefte van instellingen en het verkennen van de mogelijkheden vanuit SURF om deze behoeften in te vullen. De uitgevoerde verkenning heeft ons handvatten gegeven voor het bepalen van de koers als het gaat om learning analytics. SURF heeft de ambitie om een veilige en betrouwbare infrastructuur voor learning analytics te ontwikkelen. We werken niet alleen aan een oplossing in technische zin, maar ook staan privacy, ethische aspecten, openheid over gebruikte algoritmes en standaarden centraal, - mede op basis van de bevindingen uit deze verkenning.

Vervolg

Vanaf eind juni 2020 bepalen we samen met de instellingen hoe we de behoeften van instellingen en de ambities die SURF op basis daarvan heeft ontwikkeld, bij elkaar kunnen brengen in een concreet aanbod vanuit SURF voor de instellingen.

**TECHNOLOGIE- EN MARKTVERKENNING
LEARNING ANALYTICS**

SURF

TECHNOLOGIE- EN MARKTVERKENNING LEARNING ANALYTICS

SURF

Gijs de Groot, Arjan van Venrooy, Laura Natrop, Joost van Lier

DATUM	1-4-2020
STATUS	Definitief
VERSIE	1.1
PROJECTNUMMER	20196809
INTERNE TOETS	Interne toets

MANAGEMENTSAMENVATTING

Doelstelling technologie- en marktverkenning

Door een technologie- en marktverkenning uit te laten voeren, wil SURF inzicht krijgen in hoeverre de oplossingen in de Learning Analytics markt voldoen aan haar richtlijnen. Dit inzicht moet SURF de benodigde handvatten bieden om zich te positioneren in het Learning Analytics veld.

Betrokkenheid instellingen bij verkenning

Instellingen zijn op verschillende momenten en manieren betrokken bij deze technologie- en marktverkenning. De concept richtlijnen en concept architectuur zijn besproken in een workshop met vertegenwoordigers van onderwijsinstellingen, SURF en landelijke studentenraden. Tijdens deze workshop zijn aan de hand van dilemma's de richtlijnen en architectuur verder aangescherpt. Verder zijn tijdens de gehele verkenning interviews gehouden met diverse vertegenwoordigers van SURF en verschillende hogescholen en universiteiten. De interviews hadden tot doel inzicht te krijgen in de belangrijkste ontwikkelingen bij instellingen in relatie tot Learning Analytics, hun visie op Learning Analytics architectuur, hun ervaring met leveranciers op dit gebied en de mogelijke rol(len) die zij zien weggelegd voor SURF.

Perspectieven op Learning Analytics

In dit onderzoek is onderscheid gemaakt in twee perspectieven op Learning Analytics: het onderwijskundig perspectief en big data perspectief. Een onderwijskundig perspectief betekent dat Learning Analytics zich richt op het verzamelen en analyseren van gegevens over het leergedrag en de leerresultaten van studenten met als doel interventies te plegen om de kwaliteit van het onderwijs(materiaal) te verbeteren en/of de student beter en meer op maat te kunnen ondersteunen/begeleiden. Bij een 'big-data perspectief' gaat het om inzichten op het niveau van een opleiding, faculteit of instelling als geheel en het analyseren en met elkaar in verband brengen van grote datasets over vooropleiding, achtergrond, instroom, doorstroom, uitval, uitstroom, etc. Beide perspectieven vallen in dit onderzoek onder het begrip Learning Analytics.

Learning Analytics architectuur

De oplossingen van leveranciers richten zich op één of meerdere componenten in de Learning Analytics architectuur. Deze architectuur kent drie lagen en vijf componenten:

- Registratielaag: bronsystemen zoals student informatiesystemen en learning management systemen
- Transformatielaag: opslag van data in een datawarehouse en een onderwijsspecifieke learning records store (LRS) en data-analyse in analyse- en visualisatiesystemen
- Informatielaag: presentatie van de verkregen inzichten

De onderzochte oplossingen richten zich op één architectuurcomponent (bijv. een LRS) of op meerdere componenten (bijv. een analysesysteem die tevens een stuk opslag biedt of een geïntegreerde Learning Analytics oplossing).

Richtlijnen infrastructuur

De oplossingen van leveranciers zijn getoetst aan de hand van 45 richtlijnen. Deze richtlijnen richten zich op de onderwerpen privacy, niet-functioneel (o.a. ethiek) en functioneel. Gelet op het

belang van de bescherming van de privacy van studenten en medewerkers ligt het zwaartepunt van de richtlijnen bij privacy en gegevensbescherming.

Conclusies leveranciersonderzoek

De primaire verantwoordelijkheid voor het borgen van privacy van studenten en medewerkers ligt bij instellingen. Leveranciers kunnen dit m.b.v. functionaliteit wel eenvoudiger maken voor instellingen.

Leveranciers van oplossingen die gebruik maken van standaardalgoritmes zijn niet altijd transparant over de werking van het algoritme. Dit is een risico voor instellingen die gebruik willen maken van dergelijke oplossingen.

De markt wordt op dit moment gedomineerd door Amerikaanse spelers. In hoofdstuk V van de AVG zijn verschillende waarborgen opgenomen op basis waarvan Europese organisaties hun data mogen opslaan bij Amerikaanse techbedrijven (EU US Privacy Shield raamwerk, door de EC goedgekeurde modelovereenkomst en bindende bedrijfsvoorschriften). De vraag is echter of deze waarborgen afdoende zijn. Instellingen dienen derhalve gedegen risico- en kosten-/batenafwegingen te maken voordat zij in zee gaan met Amerikaanse (sub)leveranciers.

De verantwoordelijkheid van het maken van ethische afwegingen in het gebruik van algoritmes ligt eveneens bij instellingen. Leveranciers hebben verschillende ethische gezichtspunten en het is aan instellingen om hier keuzes in te maken.

In de markt bestaan diverse standaarden voor data-uitwisseling, waarbij voor onderwijsapplicaties de xAPI-standaard belangrijk is. De onderzochte oplossingen ondersteunen maar in beperkte mate de xAPI-standaard.

Aanbevelingen

Gebaseerd op de conclusies van het leveranciersonderzoek en de gesprekken met instellingen, is er een substantiële rol weggelegd voor SURF in het Learning Analytics veld:

- Gebaseerd op conclusies leveranciersonderzoek:
 - I. Opstellen van een toolkit met praktische oplossingen voor toepassing van Learning Analytics met inachtneming van privacy. De toolkit bevat in ieder geval een praktische aanpak voor privacy by design principes i.r.t. de Learning Analytics architectuur, procesbeschrijvingen voor afhandelen van rechten van betrokkenen en een afwegingskader voor het gebruik van Amerikaanse software
 - II. Waarborgen van goede algoritmes door het opstellen en (laten) toepassen van een algoritme toetsingskader
 - III. Het ondersteunen van instellingen bij het maken van ethische afwegingen
 - IV. Stimuleren van standaardisatie en het maken van generieke afspraken daarover met leveranciers
 - V. Aanbieden van reken- en opslagcapaciteit om de afhankelijkheid van Amerikaanse partijen te verminderen door SURFSara diensten in te zetten of contracten af te sluiten met marktpartijen
- Gebaseerd op gesprekken met instellingen:
 - VI. Overzicht bieden van oplossingen en inkoop daarvan vergemakkelijken
 - VII. Het (blijven) stimuleren van kennisdeling

- VIII. Inzetten van consultants en projectmanagers bij instellingen voor praktische ondersteuning bij Learning Analytics projecten. SURF kan hiertoe bijv. een marktplaats voor consultants en projectmanagers van marktpartijen ontwikkelen, die instellingen vervolgens kunnen inhuren

INHOUDSOPGAVE

Managementsamenvatting	3
Inhoudsopgave	6
1 Inleiding	8
1.1 Aanleiding	8
1.2 Vraagstelling	8
1.3 Aanpak	9
1.4 Doel van dit rapport	10
1.5 Leeswijzer	10
2 Perspectieven op Learning Analytics	11
2.1 Begripsbepaling Learning Analytics	11
2.2 Perspectieven op Learning Analytics	12
2.3 Groeien in het gebruik van Learning Analytics	13
2.4 Ontwikkelingen	15
2.5 Belangrijkste bevindingen uit de interviews	20
3 Architectuur	23
3.1 Inleiding data-architectuur	23
3.2 xAPI en de Learning Records Store	24
3.3 Learning Analytics architectuur	24
3.4 Aansluiting met HORA	26
3.5 Architectuurinvullingen	26
3.6 Learning Analytics diensten	27
4 Richtlijnen infrastructuur	29
4.1 Richtlijnen infrastructuur	29
4.2 Toepassing richtlijnen op Learning Analytics architectuur	29
4.3 Gegevensverwerking in derde landen	30
5 Toepassing richtlijnen infrastructuur op oplossingen leveranciers	32
5.1 Selectie leveranciers	32
5.2 Bevindingen toets oplossingen aan richtlijnen infrastructuur	32
6 Conclusies en aanbevelingen	36
6.1 Conclusies oplossingen leveranciers	36
6.2 Aanbevelingen	37

A	Bronnen	41
B	Richtlijnen infrastructuur	43

1 INLEIDING

1.1 Aanleiding

Learning Analytics biedt tal van mogelijkheden voor instellingen voor wetenschappelijk onderwijs, hoger en middelbaar beroepsonderwijs. Door inzicht in de prestaties van de studenten en de wijze waarop studenten het liefste en het beste leren, ontstaan mogelijkheden om effectieve interventies in te zetten gericht op het vergroten van studiesucces en het welbevinden van studenten. Learning Analytics biedt tevens handvatten voor het optimaliseren van curricula van opleidingen, doordat duidelijk wordt welke onderdelen aansluiten bij behoeften van studenten en welke niet of minder. Op strategisch niveau helpt Learning Analytics instellingen om balans te vinden tussen toegankelijkheid en kwaliteit van onderwijs, doordat de inzet van middelen geoptimaliseerd kan worden.

SURF is al een aantal jaren bezig om het onderwerp Learning Analytics onder de aandacht te brengen van haar leden en hen te ondersteunen bij de succesvolle toepassing ervan. Zo heeft SURF diverse onderzoeken gedaan naar Learning Analytics en heeft zij Learning Analytics experimenten bij diverse instellingen geïnitieerd en begeleid. In 2019 heeft SURF twee keer een studiereis georganiseerd naar het Verenigd Koninkrijk voor bestuurders en betrokkenen bij instellingen om te zien hoe daar met Learning Analytics wordt omgegaan. Dit heeft geresulteerd in een conceptvisie op Learning Analytics.

De volgende stap voor SURF is om te verkennen of er behoefte is bij onderwijsinstellingen om diensten aan te bieden op het gebied van Learning Analytics en welke mogelijkheden hiertoe zijn. Onderdeel van het reguliere proces van SURF om diensten te ontwikkelen is een technologie- en marktverkenning.

1.2 Vraagstelling

De technologie- en marktverkenning moet inzicht bieden in welke Learning Analytics componenten en diensten beschikbaar zijn in de markt en in hoeverre deze passen in een Learning Analytics architectuur en voldoen aan richtlijnen. Met dit inzicht wil SURF zichzelf positioneren in het Learning Analytics veld en keuzes maken over inkoop of het zelf ontwikkelen van Learning Analytics componenten en/of diensten ten behoeve van haar leden. Gelet op het belang dat door haar leden wordt gehecht aan privacy van studenten, dient in de verkenning in het bijzonder aandacht te worden besteed aan privacy.

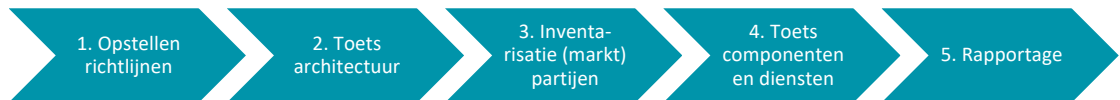
De vragen die SURF met de technologie- en marktverkenning wil beantwoorden zijn:

1. Uit welke (optionele en noodzakelijke) componenten en diensten bestaat een Learning Analytics architectuur?
2. Welke van deze componenten en diensten worden nu al door welke partijen aangeboden?
3. In hoeverre voldoen deze componenten en diensten aan de richtlijnen en kernwaarden van SURF?
4. Welke ruimte bieden de aanbieders van de componenten en diensten om – wanneer ze dat nog niet doen – te voldoen aan de beleidsprincipes en kernwaarden van SURF op het gebied van Learning Analytics?

5. Wat bieden de antwoorden op de eerste vier vragen voor mogelijkheden voor SURF ten aanzien van Learning Analytics componenten en diensten?

1.3 Aanpak

VKA heeft in opdracht van SURF onderzoek gedaan met als doel voornoemde onderzoeksvragen te beantwoorden. Hiervoor is een aanpak gehanteerd bestaande uit vijf stappen:



Figuur 1: Aanpak verkenning

De vijf stappen zijn hierna kort toegelicht:

1. In een samenwerking tussen betrokken medewerkers van SURF, de onderzoekers en privacy-experts van VKA is een set richtlijnen opgesteld. Deze “richtlijnen infrastructuur” vormen de meetlat waarlangs de oplossingen van (markt)partijen zijn gelegd. De richtlijnen hebben betrekking op privacy, niet-functionele en functionele aspecten van Learning Analytics oplossingen. Deze richtlijnen zijn opgenomen in bijlage B.
2. Stap 2 – die deels parallel is uitgevoerd aan stap 1 – stond in het teken van Learning Analytics architecturen. Er is een aantal Learning Analytics (data)architecturen ontwikkeld, waarvan er uiteindelijk één is gebruikt als kader voor het vervolg van het onderzoek.
3. De derde stap behelsde de inventarisatie van te onderzoeken (markt)partijen. Hiertoe hebben de onderzoekers op de SURF Onderwijsdagen de stands van relevante leveranciers bezocht. De lijst met leveranciers is compleet gemaakt door per architectuurcomponent de relevante leveranciers te selecteren.
4. Door middel van deskresearch en e-mail-/telefoonconversaties zijn de oplossingen van de partijen langs de richtlijnen gelegd. Er is – zo ver als mogelijk was – bepaald in welke mate de oplossingen voldoen aan de in stap 1 opgestelde richtlijnen. Dit is niet volledig gelukt voor alle oplossingen, omdat niet alle leveranciers gehoor hebben gegeven aan herhaalde oproepen om mee te doen aan het onderzoek. Tevens was soms de informatie die via de website van een leverancier te vinden was dusdanig beperkt, dat via e-mail/telefonisch contact niet alle onderwerpen aan de orde zijn gekomen.
5. Tot slot is deze rapportage opgesteld, waarin de resultaten, conclusies en aanbevelingen van het onderzoek uiteen zijn gezet.

Betrokkenheid instellingen bij verkenning

Instellingen zijn op verschillende momenten en manieren betrokken bij deze technologie- en marktverkenning. De concept richtlijnen en concept architectuur zijn besproken in een workshop met vertegenwoordigers van onderwijsinstellingen, SURF en landelijke studentenraden. Tijdens deze workshop zijn aan de hand van dilemma’s de richtlijnen en architectuur verder aangescherpt. De deelnemers aan deze workshop zijn in de bijlage vermeld. Verder zijn tijdens de gehele verkenning interviews gehouden met diverse vertegenwoordigers van SURF en verschillende hogescholen en universiteiten. De interviews hadden tot doel inzicht te krijgen in de belangrijkste

ontwikkelingen bij instellingen in relatie tot Learning Analytics, hun visie op Learning Analytics architectuur, hun ervaring met leveranciers op dit gebied en de mogelijke rol(len) die zij zien weggelegd voor SURF. De lijst met geïnterviewden is ook opgenomen in de bijlage.

1.4 Doel van dit rapport

Dit rapport heeft tot doel de resultaten, conclusies en aanbevelingen van de technologie- en marktverkenning Learning Analytics te beschrijven. Doelgroep van het rapport zijn de Learning Analytics experts van SURF en haar leden.

1.5 Leeswijzer

Dit rapport bestaat uit de volgende onderdelen:

1. De inleiding
2. Na de inleiding volgt een hoofdstuk waarin twee perspectieven op Learning Analytics worden beschreven en tot verdere begripsafbakening wordt gekomen. Hierin wordt tevens ingegaan op een aantal belangrijke ontwikkelingen in relatie tot Learning Analytics. Tevens beschrijven wij hier de belangrijkste punten die naar voren kwamen tijdens de interviews met vertegenwoordigers van onderwijsinstellingen.
3. In het derde hoofdstuk wordt de generieke Learning Analytics architectuur weergegeven en toegelicht. Hiermee wordt antwoord gegeven op onderzoeksvraag 1.
4. Hoofdstuk 4 staat in het teken van de richtlijnen die zijn opgesteld en waaraan de oplossingen van leveranciers zijn getoetst.
5. De uitkomsten van de toepassing van de richtlijnen op de oplossingen van leveranciers is beschreven in hoofdstuk 5. In dit hoofdstuk zijn de bevindingen ("rode draad") van het leveranciersonderzoek weergegeven. Hoofdstuk 5 geeft antwoord op onderzoeksvragen 2, 3 en 4.
6. Het zesde en laatste hoofdstuk vormt de conclusie van het onderzoek. Uit deze conclusie volgen de aanbevelingen aan SURF omtrent haar rol in het Learning Analytics veld. Hiermee wordt een antwoord geformuleerd op onderzoeksvraag 5.

In de bijlagen is een aantal zaken opgenomen:

- A. De geraadpleegde bronnen (documentatie, deelnemers workshop en geïnterviewde personen)
- B. Volledige set richtlijnen
- C. Het overzicht van de onderzochte leveranciers en oplossingen vormt een separate bijlage (dus geen onderdeel van dit rapport)
- D. Tevens in een separate bijlage zijn de detailbevindingen van het leveranciersonderzoek per oplossing weergegeven

2 PERSPECTIEVEN OP LEARNING ANALYTICS

In dit hoofdstuk introduceren wij een tweetal perspectieven op Learning Analytics en bakenen we het begrip af. Tevens schetsen wij een aantal belangrijke ontwikkelingen die van invloed zijn op de wijze waarop onderwijsinstellingen omgaan en bezig zijn met Learning Analytics. Tevens beschrijven wij hier de belangrijkste punten die naar voren kwamen tijdens de interviews met vertegenwoordigers van onderwijsinstellingen. Tegen deze achtergrond moet de Learning Analytics architectuur en richtlijnen infrastructuur uit volgende hoofdstukken begrepen worden.

2.1 Begripsbepaling Learning Analytics

Learning Analytics is een overkoepelende term voor 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 verbeteren¹. De Special Interest Group (SIG) Learning Analytics van SURF hanteert de volgende omschrijving²: “Learning Analytics richt zich op het verzamelen en analyseren van data uit leeromgevingen om het leerproces van studenten te verbeteren. Deze informatie wordt vervolgens beschikbaar gemaakt voor verschillende stakeholders, zoals de student zelf, de docent of opleidingsmanagement”. Wij voegen hier nog aan toe dat het niet alleen gaat om data uit leeromgevingen, maar ook om data uit andere systemen met studentdata, zoals student administratiesystemen. Kern is het verzamelen, analyseren en interpreteren van data over studenten en leergedrag om tot nieuwe inzichten te komen. Met deze inzichten worden passende interventies uitgevoerd om de kwaliteit van het onderwijs te verbeteren en om studenten beter te kunnen begeleiden in hun leerproces.

De zone “Veilig en betrouwbaar benutten van studiedata” uit het Versnellingsplan voegt in de begripsbepaling ‘Student Analytics’ toe. Wil je studiedata gebruiken om meer te weten te komen over ontwikkelingen en trends binnen de instelling? “Student Analytics gaat om data over groepen studenten over een langere termijn. Ze geven bijvoorbeeld inzicht in de invloed van de vooropleiding op het studiesucces, of over hobbels in de opleiding, die studenten verhinderen om de eindstreep te halen. Behalve voor analyses kunnen Student Analytics ook worden gebruikt om voorspellingen te doen, bijvoorbeeld over uitval”³.

Learning Analytics en Student Analytics gaan *niet* over het digitaliseren van bijvoorbeeld leermiddelen en toetsen. Dergelijke vormen van digitalisering maken de toepassing van Learning Analytics wel beter mogelijk. Door de inzet van digitale leermiddelen, een digitale leeromgeving en digitale toetsen komt meer data beschikbaar over het leergedrag en leerresultaten van studenten.

¹ Definitie van SURF. Behalve Learning Analytics zijn gebruikte synoniemen: Student Analytics/ Teaching Analytics / Course Analytics / Onderwijs Analytics

² Hoe data de kwaliteit van het hoger onderwijs kunnen verbeteren (2019), SURF, Utrecht

³ Folder Studiedata, 2019

Learning Analytics is daarnaast geen eenrichtingsweg van docent naar student, maar een wisselwerking vanuit een gezamenlijk doel om de onderwijskwaliteit te verbeteren. Leren komt tot stand in interactie tussen docent en student. Learning Analytics kan deze interactie ondersteunen met inzichten gebaseerd op studiedata.

Learning Analytics en Student Analytics ondersteunen medewerkers van onderwijsinstellingen bij het maken van betere beslissingen door de inzet van analyse van studiedata. Beslissingen worden genomen op basis van uit data verkregen inzichten, patronen en verbanden, in plaats van aannames en onderbuikgevoel. Hierdoor zijn onderwijsinstellingen in staat om m.b.v. (voorspellende) inzichten:

- Studenten beter te begeleiden
- Gerichter beleid te formuleren
- De kwaliteit van het leermateriaal continu te monitoren en te verbeteren

2.2 Perspectieven op Learning Analytics

In lijn met de begrippen Learning Analytics en Student Analytics onderscheiden we twee perspectieven op studiedata, een onderwijskundig perspectief en een ‘big data’ perspectief. Hieronder werken we beide perspectieven kort uit.

Een onderwijskundig perspectief op studiedata betekent dat Learning Analytics zich richt op het verzamelen en analyseren van gegevens over het leergedrag en de leerresultaten van studenten met als doel interventies te plegen om de kwaliteit van het onderwijs(materiaal) te verbeteren en/of de student beter en meer op maat te kunnen ondersteunen/begeleiden. Bij dit perspectief gaat het vaak over het gebruik en de analyse van leerresultaten die in het studentinformatiesysteem worden opgeslagen of informatie over het leergedrag uit de digitale leeromgeving. Op basis van deze gegevens kunnen interventies worden gepleegd richting individuele studenten of kan de inhoud van onderwijscontent worden aangepast. Een voorbeeld van een studentinformatiesysteem is Osiris en voorbeelden van digitale leeromgevingen zijn Canvas en Moodle. Zo kan een docent of tutor contact opnemen met een student bij tegenvallende leerresultaten of als blijkt dat een student al langere tijd geen leeractiviteiten heeft laten zien in de digitale leeromgeving. Een docent / tutor kan dan samen met de student mogelijke oorzaken bespreken en indien mogelijk de student gericht en persoonlijk ondersteunen. Ook kan via het leergedrag in de digitale leeromgeving de kwaliteit van de content worden gemonitord. Als blijkt dat sommige content nauwelijks wordt bekeken, kan deze worden aangepast.

Bij een ‘big-data perspectief’ op studiedata gaat het veel meer om Student Analytics en het analyseren en met elkaar in verband brengen van grote datasets over vooropleiding, achtergrond, instroom, doorstroom, uitval, uitstroom, etc. Daarbij worden (voor zover mogelijk binnen bestaande wet- en regelgeving) externe bronnen met demografische gegevens, arbeidsmarktgegevens, financiële gegevens (zoals studiekosten, studieschuld, (start)salaris) gecombineerd met interne bronnen. Op basis van deze gegevens worden statistische analyses uitgevoerd en voorspellende modellen opgesteld. Het doel is dan vooral om beleid en studiesucces

op het niveau van de opleidingen, faculteiten en de instelling als geheel te verbeteren. In dit perspectief gaat het over het voorspellen van leergedrag van groepen studenten, door het combineren van data uit verschillende databronnen en resultaten uit het verleden met het heden. Daarmee kan beter geanticipeerd worden op leergedrag en studiepaden. Gebruikte data is niet herleidbaar tot de individuele persoon. Veelal wordt hierbij onderzoek gedaan met als startpunt een hypothese.

Ook kunnen gegevens uit bijvoorbeeld ERP-systemen, facilitaire diensten, zoals klimaat (vocht, temperatuur), leerlocatie / reistijd en gebruik van data uit complexe leermiddelen, zoals serious games, worden toegevoegd. Dit inzicht leidt dan weer tot mogelijke interventies om de kwaliteit van het onderwijs en/of de onderwijsomgeving te verbeteren en om studenten beter en gericht te kunnen ondersteunen. Het perspectief gaat over het verbinden van bronnen, zowel uit de leerketen als uit een totaal andere context. Om deze vorm van analytics mogelijk te maken is een complexere data-infrastructuur nodig. Daarnaast is ook de impact op organisatie, docenten en studenten complexer.

Als we in het vervolg van deze verkenning spreken over Learning Analytics bedoelen we zowel het onderwijskundig als het 'big data' perspectief.

Persoonlijke versus geanonimiseerde data

Het onderscheid in perspectieven (onderwijskundig perspectief en 'big data' perspectief) kent ook een onderscheid in het gebruik tussen persoonsgegevens en geanonimiseerde data. In het onderwijskundig perspectief worden tot de persoon herleidbare data in kaart gebracht over iemands studiegedrag in de digitale leeromgeving of iemands studieresultaten die in het studentinformatiesysteem zijn opgeslagen⁴. Op basis van deze aan een persoon gekoppelde data kunnen gerichte interventies worden gepleegd om de desbetreffende student beter te begeleiden.

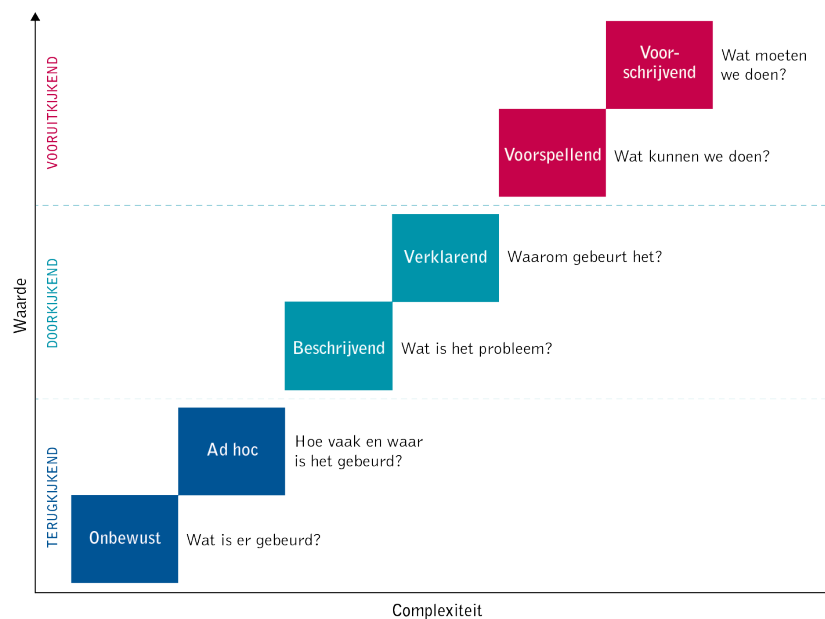
In het 'big data' perspectief worden geanonimiseerde data gebruikt, zoals het analyseren van grote hoeveelheden data om te bepalen welke factoren het meest bepalend zijn voor studiesucces en een goede doorstroom van studenten. Deze informatie kan bijvoorbeeld worden gebruikt voor een betere voorlichting van (aankomende) studenten.

Het verschil in gebruik van tot de persoon herleidbare gegevens en geanonimiseerde of gepseudonimiseerde gegevens heeft ook weer verschillende consequenties voor privacy-aspecten van de toepassing van Learning Analytics.

2.3 Groeien in het gebruik van Learning Analytics

Zoals uit bovenstaande perspectieven blijkt, zijn er verschillende vormen van Learning Analytics, die verschillen in de aard van de toepassing, complexiteit en toegevoegde waarde. In onderstaande figuur zijn verschillende niveaus van de toepassing van Learning Analytics weergegeven op basis van complexiteit en toegevoegde waarde.

⁴ Dit zou overigens ook met gepseudonimiseerde gegevens kunnen.



Figuur 2: Niveaus gebruik data en Learning Analytics perspectieven (gebaseerd op Gartner)

In het toepassen van Learning Analytics zijn de volgende fasen te onderscheiden⁵:

- **Onbewust:** Informatie speelt geen structurele rol in de onderwijsinstelling en wordt incidenteel gebruikt. Intuïtie en onderbuik zijn leidend. Het streven naar eenduidigheid van duiding van inzichten is onbekend.
- **Ad Hoc:** Proces van dataverzameling is niet gestructureerd. De onderwijsinstelling maakt zeer beperkt gebruik van informatie. Bij het uit de pas lopen van uitkomsten op de verwachtingen, wordt inzicht opgebouwd uit bestaande datasets.
- **Beschrijvend:** Interne informatie wordt gestructureerd verzameld in de eigen onderwijsinstelling. Informatie geeft antwoord op wat het probleem is. De onderwijsinstelling probeert grip te krijgen op dat een bepaald risico zich niet nog een keer voordoet.
- **Verklarend:** Interne en externe informatie worden gestructureerd verzameld en geanalyseerd. Informatie heeft een sturende rol. Doelstellingen worden m.b.v. (vaste) indicatoren gemeten.
- **Voorspellend:** Informatie wordt uit uiteenlopende databronnen verzameld. Ongestructureerde data wordt toegevoegd en geanalyseerd. Nieuwe modellen en structureringen worden ontwikkeld om tot voorspellingen en andere kans- en risico-inzichten te komen.
- **Voorschrijvend:** Informatie is leidend bij beslissingen. Op basis van analyses van data wordt een optimaal scenario gekozen. In bepaalde gevallen wordt deze keuze automatisch gemaakt. Vertrouwen in de voorspelling is groot.

Volwassen worden in het toepassen van Learning Analytics is niet eenvoudig. Hogere volwassenheidsniveaus zijn ook niet per definitie beter. De wenselijkheid van een bepaald

⁵ Zie afbeelding 2 Niveaus gebruik data en Learning Analytics perspectieven

volwassenheidsniveau is ook sterk afhankelijk van de context. Soms zijn beschrijvende en verklarende modellen voldoende. Zeker bij de stap naar de niveaus voorspellend en voorschrijvend nemen ook de ethische vraagstukken toe, zoals 'hoe transparant en integer is het algoritme dat voor de voorspellingen wordt gebruikt?'. Het volwassenheidsniveau van Learning Analytics dat een onderwijsinstelling nastreeft, moet daarom een bewuste keuze zijn waarbij dit soort vraagstukken in de keuze worden meegewogen en kan verschillen per toepassing. Om als onderwijsinstelling Learning Analytics tot ontwikkeling te brengen, moet de groei ingezet worden over meerdere pijlers. In onderstaande figuur zijn de pijlers weergegeven, waar een onderwijsinstelling aandacht aan moet besteden om te groeien in het gebruik van Learning Analytics en om ook echt (de kwaliteit van) het onderwijs met Learning Analytics te verbeteren.



Figuur 3: Pijlers waarover de toepassing van Learning Analytics groeit (model VKA)

Het groeien in het vastleggen van data en maken van keuzes op basis van deze data, gaat dus niet over data en ICT alleen (pijlers 'Data Management & Analytics' en 'Informatievoorziening & Technologie'), maar juist over groeien als organisatie in totaliteit (Pijlers 'Strategie & Besturing', 'Structuur & Processen' en 'Mensen & Gedrag'). Gefaseerd groeien als onderwijsinstelling met aandacht voor alle pijlers is randvoorwaardelijk om door te groeien en groei te bestendigen. Anders valt de onderwijsinstelling terug in de oude situatie.

2.4 Ontwikkelingen

Ontwikkelingen in het onderwijs

Ontwikkelingen in het onderwijs als meer maatwerk en gepersonaliseerd onderwijs leiden tot meer aandacht voor Learning Analytics. Flexibilisering en personalisatie zijn belangrijke trends in het

onderwijs. Bij gepersonaliseerd en flexibel onderwijs zijn vaak de persoonlijke leervragen en keuzes van de lerende het uitgangspunt, en niet het vaste opleidingsprogramma. Bij veel instellingen krijgen studenten binnen bepaalde kaders, steeds meer invloed op de inhoud, de werkvorm, de leermiddelen, de toetsing, de tijd, de plaats en het tempo van het leren. Het gebruik van Learning Analytics vanuit vooral het onderwijskundig perspectief kan leiden tot waardevolle informatie om meer personalisatie te bereiken. Learning Analytics kan leiden tot beter inzicht in het leergedrag van studenten en mogelijke gerichte interventies (op maat) om hen beter in hun leerproces te begeleiden.

Een andere ontwikkeling is dat studenten niet alleen leren binnen de deuren van de eigen instelling maar ook steeds meer daarbuiten. Alle informatie op internet draagt bij aan het leerproces van een student. Een geraadpleegde Wikipedia-pagina of een YouTube-video over een bepaald onderwerp kunnen inzichten bieden die het leerproces significant bevorderen. Daarnaast kunnen via andere onderwijsinstellingen en commerciële leerplatformen, nationaal en internationaal, ook onderwijsmodules worden gevolgd. Ook de toename van het aantal MOOCS draagt bij aan deze ontwikkeling. De student leert dus op verschillende plekken en in verschillende systemen⁶. Als we een goed beeld willen krijgen van het leerproces van studenten is het noodzakelijk om data uit deze verschillende systemen bij elkaar te brengen en daar analyses op uit te voeren.

Aansluitend bij het voorgaande en de trend van meer flexibilisering en maatwerk is dat de visie in het hoger onderwijs op de digitale leeromgeving van de toekomst steeds meer uitgaat van een samenstel van componenten/applicaties die flexibel en persoonlijk ingezet kunnen worden ter ondersteuning van het leerproces. De leeromgeving sluit aan bij behoeften en wensen van studenten en docenten, die, los van tijd en locatie, snel en gemakkelijk toegang tot informatie en materialen willen hebben. Een digitale leeromgeving van de toekomst is een samenhangend geheel van diensten en applicaties die studenten en docenten ondersteunen bij het leerproces en het onderwijs. Dit vraagt om nieuwe architecturen waarbij integratie van deze losse onderdelen een grote uitdaging vormt⁷.

Technologische ontwikkelingen

Learning Analytics is niet nieuw, maar de mate waarin data kunnen worden verzameld en geanalyseerd is de laatste jaren sterk gestegen. Dit heeft onder andere te maken met *technologische ontwikkelingen*, zoals digitalisering van het onderwijs, waardoor meer data verzameld en ontsloten kunnen worden. Doordat het onderwijsproces en de leeractiviteiten van studenten meer digitaal plaatsvinden, ontstaan steeds meer 'digitale sporen' van het leergedrag van studenten. Voorbeelden hiervan zijn digitale leeromgevingen, digitale formatieve toetsen, digitale summatieve toetsen, digitale feedback systemen, digitale discussiegroepen en chatfuncties, etc. Bovendien zijn door toenemende computerkracht en het breed beschikbaar

⁶ Learning Analytics, SURFConext & de Open Onderwijs API. Een technologie verkenning (rapport is op te vragen bij SURFnet).

⁷ Notitie "Een flexibele en persoonlijke leeromgeving", SURF.

komen daarvan via de cloud meer mogelijkheden om deze data te analyseren. Voor de analyse van deze grote hoeveelheden data wordt steeds vaker gebruik gemaakt van Artificiële Intelligentie (AI) en algoritmes. Via deze algoritmes en AI-toepassingen wordt de inzet van adaptieve leertechnologieën mogelijk, waarbij een geautomatiseerd systeem studenten coacht en suggesties geeft die specifiek zijn toegespitst op de individuele behoefte van de betreffende student. Deze adaptieve oplossingen kunnen met wetenschappelijke en data-gedreven benaderingen (analytics) de personalisatie ondersteunen.

De ontwikkeling in het onderwijs van een digitale leeromgeving van de toekomst die bestaat uit een samenhangend geheel van losse componenten dat schijnbaar als één systeem werkt, kan alleen als technische afspraken zijn gemaakt over de standaarden waaraan deze componenten moeten voldoen. Door standaarden te gebruiken, kunnen gegevens veilig, betrouwbaar en eenvoudig uitgewisseld worden. Ook het afstemmen van begrippenkaders is daarbij van belang, omdat er in het onderwijs vaak verschillende begrippen worden gebruikt voor dezelfde informatie. Diensten en applicaties zijn in de praktijk nog niet zover gestandaardiseerd dat ze moeiteloos op elkaar aansluiten. Om losse systemen als één systeem te laten functioneren is een integratie-infrastructuur noodzakelijk. Een integratie-infrastructuur zorgt ervoor dat gegevens tussen applicaties kunnen worden uitgewisseld.

Ontwikkelingen in Learning Analytics

Eerder hebben we onderscheid gemaakt tussen Learning Analytics en Student Analytics, waarbij voor beide vormen van de analyse van studiedata geldt dat data uit verschillende bronnen wordt samengebracht in een datawarehouse of Learning Records Store. Vervolgens worden analyses uitgevoerd op deze samengebrachte data.

Een andere vorm van analytics is ‘process mining’. Dit is het analyseren van gebeurtenissen in de systemen in de volgorde waarin ze zich voordoen. Process mining visualiseert het proces, doordat het rekening houdt met de volgordelijkheid van de data. In het bedrijfsleven wordt process mining al veel toegepast om processen te verbeteren. In het Nederlandse hoger onderwijs is de Open Universiteit vooralsnog de enige instelling die process mining inzet⁸. Deze vorm van analytics werken we in dit rapport niet verder uit. Het is onderdeel van de zone “Veilig en betrouwbaar benutten van studiedata” van het Versnellingsplan.

Een andere ontwikkeling in Learning Analytics is de behoefte om real-time inzichten te kunnen genereren uit real-time studie data in plaats van dat data batchgewijs wordt verwerkt met een bepaalde mate van vertraging. De Open Universiteit (OU) is in dat kader aan het experimenteren met haar zogenaamde ‘Dataspons’ architectuur⁹. De Dataspons architectuur is gebaseerd op de zogenaamde Kafka architectuur. Voor een uitgebreide toelichting verwijzen wij hier naar het

⁸ Folder Studiedata, 2019

⁹ Vogten, H., & Koper, R. (2018). Towards Big Data in Education: the case at the Open University of the Netherlands. In M. Spector, V. Kumar, A. Essa, Y-M. Huang, R. Koper, R. Tortorella, T-W. Chang, Y. Li, ... Z. Zhang (Eds.), *Frontiers of Cyberlearning: Emerging Technologies for Teaching and Learning* (pp. 125-143).

hiervoor genoemde artikel van Vogten en Koper (2018). Aangezien de toepassing van deze architectuur binnen het hoger onderwijs ook nog in de onderzoeksfase zit, wordt deze architectuur in dit rapport ook niet verder uitgewerkt.

Maatschappelijke ontwikkelingen

Er zijn ook *maatschappelijke ontwikkelingen* die doorwerken in het onderwijs. De samenleving digitaliseert en dataficeert razendsnel. Het gebruik van data en innovatieve technologieën op het gebied van Artificial Intelligence, Robotisering en Virtual Augmented Reality wordt steeds meer gemeengoed. De algemene verwachting is dat de impact van deze technologische ontwikkelingen en de mogelijkheden van het gebruik van data de komende jaren exponentieel zullen groeien. Dit komt doordat er simpelweg meer data beschikbaar komt (bijvoorbeeld door het verzamelen van data met slimme sensoren en het verder digitaliseren van het onderwijs) en doordat de mogelijkheden van geavanceerde data-analyse groeien (als gevolg van toename van rekenkracht en het beschikbaar komen daarvan via de cloud). Maar het is niet alleen de aanbodkant die deze ontwikkeling drijft. De maatschappij vraagt en verwacht meer van de dienstverlening van organisaties. Deze wordt persoonlijker en tijd- en plaatsonafhankelijk. Denk bijvoorbeeld aan Netflix die het mogelijk maakt overal en op elk moment een film te kijken, maar die ook persoonlijke aanbevelingen doet op basis van slimme algoritmen. Het World Economic Forum duidt deze periode aan als de Vierde Industriële Revolutie¹⁰, een tijd van disruptie waarin de manier waarop wij werken, leven en leren radicaal aan het veranderen is. Het is moeilijk te voorspellen waar dit naartoe zal gaan, maar duidelijk is dat het slim inzetten van nieuwe technologie leidt tot kansen voor het creëren van maatschappelijke waarde op allerlei terreinen. Daarbij moet ook voldoende aandacht zijn voor de keerzijden van de inzet van nieuwe technologieën en analyses, zoals inbreuk op de privacy, algoritmes met vooroordelen en interventies op basis van foutieve analyses.

Een andere maatschappelijke ontwikkeling die in het kader van deze verkenning relevant is, is de toename van macht van grote techbedrijven in het maatschappelijk leven en verschillende sectoren, waaronder het onderwijs. Grote techbedrijven domineren de markt. Door de populariteit van hun platformen kunnen zij relatief eenvoudig grote hoeveelheden gebruikersdata en gebruiksdata verzamelen, waardoor zij weer nieuwe diensten kunnen ontwikkelen voor hun klanten, dat weer leidt tot meer en betere data, etc. De afhankelijkheid van dergelijke platformen neemt toe en daarmee de afhankelijkheid van (Amerikaanse) techbedrijven. In een op 23 december 2019 ingezonden artikel in de Volkskrant stellen de rectores magnifici van een groot aantal Nederlandse universiteiten dat het ‘tijd is om een grens te trekken’¹¹. Eén van hun belangrijkste voorstellen is dat “de sector verantwoordelijkheid moet nemen voor de publieke waarden. Onderwijs is een publiek goed, het dient breed toegankelijk te blijven. De onafhankelijkheid mag niet in het geding komen, de rechten en vrijheden van studenten en

¹⁰ Zie onder andere: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>

¹¹ <https://www.volkskrant.nl/columns-opinie/digitalisering-bedreigt-onze-universiteit-het-is-tijd-om-een-grens-te-trekken~bff87dc9/>

medewerkers – zoals privacy en non-discriminatie – moet gewaarborgd zijn. En het publiek toezicht moet zijn gegarandeerd. Die waarden dienen voorop te staan, ook bij digitalisering en het gebruik van commerciële platformen”. Volgens de rectores moet om dat te bewaken door de sector een kader worden ontwikkeld. De sector moet gezamenlijk grenzen stellen en voorwaarden bepalen voor het gebruik van commerciële platformen door publieke organisaties.

Juridische en ethische ontwikkelingen

De *juridische ontwikkelingen* die relevant zijn voor Learning Analytics liggen met name op het gebied van privacy. Het van kracht worden van de AVG heeft voor hernieuwde aandacht voor privacy gezorgd en veel discussie op gang gebracht, zeker bij toepassingen op het gebied van Learning Analytics. De uit de AVG voortkomende eisen ten aanzien van privacy moeten uiteraard onderdeel zijn van de kaders voor Learning Analytics, maar ook moeten de zorgen over privacy geadresseerd en besproken worden, zoals geuit door o.a. studenten.

Naast de juridische vraagstukken voortkomend uit de AVG is er steeds meer aandacht voor ethische vraagstukken rond de toepassing van Learning Analytics. Er leven binnen onderwijsinstellingen verschillende zorgen over onder andere de impact van Learning Analytics op het onderwijs. Uit interviews is gebleken dat deze zorgen met name gaan over hoe de verzamelde en geanalyseerde data wordt geïnterpreteerd en op welke wijze deze analyses leiden tot conclusies en interventies. Uit de pilots en trajecten op het gebied van Learning Analytics die door SURF zijn begeleid, volgt een vergelijkbaar beeld. Ook hier blijken vooral ook studenten zorgen te hebben over interpretaties van data, mogelijk verkeerde conclusies en averechtse interventies. Voor de wijze waarop met data moet worden omgegaan, moet continu aandacht zijn. Kritische reflectie op handelingswijze en toepassing zijn belangrijk.

Ontwikkelingen in de EdTech sector

De EdTech sector is sterk in ontwikkeling. Grote techbedrijven met geïntegreerde platformen, maar ook kleine start-ups met innovatieve onderwijsapps of specifieke AI-toepassingen vinden steeds meer hun weg in het onderwijs. Vanuit het perspectief van Learning Analytics zijn de volgende ontwikkelingen interessant.

Ten eerste richten de eerdergenoemde grote techbedrijven zich steeds meer op de onderwijssector. Zoals eerdergenoemd leidt dit tot platformisering en een grote afhankelijkheid van een aantal grote (Amerikaanse) techbedrijven. Deze techbedrijven verzamelen grote hoeveelheden data, waarop zij continu analyses uitvoeren. Studenten ‘betalen’ steeds vaker met hun data om de diensten van deze techbedrijven te kunnen gebruiken.

Ten tweede zijn er de leveranciers van geïntegreerde Elektronische Leeromgevingen (ELO's). Deze ELO's hebben steeds vaker ook Learning Analytics functionaliteit om de data van het gebruik van de ELO door studenten te kunnen analyseren. De Learning Analytics functionaliteit is gericht op individuele studenten op basis waarvan zij beter begeleid kunnen worden en op basis waarvan

individuele leerinterventies kunnen worden gedaan. Voorbeelden van dergelijke ELO's zijn Canvas, Blackboard, Moodle, Itslearning en Brightspace.

Ten derde worden er digitale adaptieve producten / systemen door educatieve uitgevers ontwikkeld. In dit adaptief digitaal leermateriaal vindt de (bij)sturing van het onderwijsleerproces plaats door analyse van het leergedrag. Op basis van die analyse worden patronen gevonden (profielen) die een voorspellende waarde hebben voor het leersucces van de studenten. Deze informatie wordt door de systemen gebruikt om zich aan te passen aan de student, bijvoorbeeld in de vorm van 'aanbevelingen' voor vervolgactiviteiten of specifieke formatieve toetsen. Educatieve uitgeverijen als bijvoorbeeld Pearson en McGraw-Hill zijn hier heel nadrukkelijk mee bezig en hebben inmiddels vele producten en diensten op dit gebied.

Ten vierde komen er steeds meer kleinere EdTech bedrijven als start-ups op de onderwijsmarkt. Zij leveren innovatieve onderwijsapps of hele specifieke ICT-toepassingen. Zij leveren één of een aantal specifieke componenten voor de eerder beschreven digitale leeromgeving van de toekomst die bestaat uit een samenhangend geheel van losse componenten.

2.5 Belangrijkste bevindingen uit de interviews

Hieronder beschrijven wij puntsgewijs de belangrijkste bevindingen die naar voren kwamen tijdens de interviews met vertegenwoordigers van onderwijsinstellingen. Deels zijn deze punten overlappend met eerder beschreven ontwikkelingen.

- Er gebeurt veel op het gebied van Learning Analytics binnen onderwijsinstellingen. Vaak zijn de initiatieven nog vrij ongecoördineerd en zijn onderwijsinstellingen bezig om te inventariseren wat er allemaal loopt bij faculteiten en diensten om vervolgens te bepalen hoe zij hier meer lijn en structuur in kunnen aanbrengen. Vaak worden er vier verschillende gebruikersgroepen van Learning Analytics onderscheiden: studenten, docenten, beleidsmedewerkers en onderzoekers. Beleidsmedewerkers en onderzoekers zijn daarbij gericht op analyses van grote hoeveelheden geanonimiseerde data en studenten en docenten zijn vaak gericht op data en inzichten over het leergedrag van studenten om deze op basis daarvan beter te kunnen begeleiden en/of het leermateriaal te verbeteren.
- De meeste onderwijsinstellingen die nu actief met Learning Analytics bezig zijn, denken na over het opstellen van een zogenaamde 'Code of Conduct' in lijn met de AVG of hebben deze al opgesteld. Binnen de zone "Veilig en betrouwbaar benutten van studiedata" van het Versnellingsplan, waarin vijf HBO-instellingen en vijf universiteiten zijn vertegenwoordigd, wordt een dergelijke Code of Conduct ontwikkeld (op basis van een voorbeeld van de VU) die vervolgens door alle onderwijsinstellingen kan worden gebruikt.
- In lijn met het vorige punt wordt door onderwijsinstellingen veel waarde gehecht aan ethische dilemma's als meer gebruik wordt gemaakt van Learning Analytics. Naast de vraag wat de instelling mag doen vanuit de AVG gezien, is een belangrijke vraag tot hoever de instelling wil gaan met Learning Analytics vanuit een meer ethisch perspectief. Instellingen onderkennen het belang om het 'ethische gesprek' hierover te voeren met docenten en studenten.
- Learning Analytics wordt ingezet voor verschillende doelen: beter beleid, betere individuele begeleiding van studenten en verbeteren van leermateriaal.
- De meeste onderwijsinstellingen onderschrijven het concept van een component gebaseerde DLO, maar in de praktijk hebben de meeste onderwijsinstellingen nog steeds een brede en geïntegreerde digitale leeromgeving (standaardpakketten, zoals Moodle, Blackboard, Canvas en Itslearning).

- De meeste onderwijsinstellingen hebben een eigen BI-afdeling, waarin rapportages worden gemaakt ten behoeve van de bestuurlijke informatievoorziening op basis van met name bedrijfsvoeringgegevens die worden verzameld in een datawarehouse. Deze BI-afdeling is niet per definitie de afdeling waar ook de Student Analytics en Learning Analytics activiteiten worden belegd. Onderwijsinstellingen worstelen daarmee, aangezien zij het ook als een gemiste kans zien als deze vormen van analytics binnen een instelling gescheiden werelden zijn. Bij sommige onderwijsinstellingen wordt er daarom voor gekozen om deze vormen van analytics wel bij elkaar te brengen, bijvoorbeeld door de BI-afdeling door te ontwikkelen tot een BI Competence Center.
- Wat een lastig vraagstuk is, is hoe onderwijsinstellingen aan de gebruiksdata komen van studenten die digitale methodes/digitaal leermateriaal gebruiken van uitgevers. Het streven is om een 360-graden beeld van de student te kunnen creëren, maar de data in de systemen/toepassingen van de uitgevers is daar lastig uit te krijgen. Aan de kant van de uitgevers vinden wel veel ontwikkelingen plaats als het gaat om de inzet van ICT en Learning Analytics om de behoefte aan meer maatwerk en flexibiliteit te kunnen faciliteren.
- Een ander punt dat wordt aangegeven door onderwijsinstellingen is dat zij verwachten dat de grote techleveranciers zich niet zullen houden aan onderwijsspecifieke standaarden, zoals bijvoorbeeld de standaarden die ten grondslag liggen aan een Learning Record Store. Hierdoor wordt het lastig om de data van deze partijen en hun platforms in Learning Analytics te betrekken.
- Sommige instellingen geven aan dat er spanning zit tussen de AVG en de Zorgplicht die instellingen hebben om goed onderwijs te geven en studenten zo goed mogelijk te begeleiden.
- In het kader van de ontwikkeling van meer flexibel onderwijs en het makkelijker kunnen volgen van verschillende vakken bij verschillende onderwijsinstellingen voorzien sommige instellingen een behoefte om de student op landelijk niveau in beeld te kunnen brengen. Om dit te kunnen doen, zou er ook een landelijke Learning Analytics infrastructuur nodig zijn.
- Er bestaan binnen instellingen grote verschillen over wie in de 'lead' is als het gaat om Learning Analytics. Bij sommige instellingen zijn dat meer de onderzoekers, bij andere instellingen de BI-afdeling, de Informatiemanagement afdeling en soms het onderwijs.
- Een belangrijk aandachtspunt bij instellingen is de competentieontwikkeling van docenten op het gebied van Learning Analytics. Het gaat daarbij om het kunnen omgaan met de 'tooling', het kunnen interpreteren van analyses en het op basis van nieuwe inzichten kunnen toepassen van interventies.
- Instellingen die met BI en Learning Analytics bezig zijn, onderkennen het belang van het hebben van een canonic datamodel. Zij geven aan dat het ontwikkelen daarvan veel tijd en energie kost en dat het nut ervan lastig is over te brengen op docenten en bestuurders.
- Onderwijsinstellingen hebben duidelijk behoefte aan ondersteuning op het gebied van Learning Analytics en zien daar ook verschillende rollen voor SURF. Dit werken we verder uit in hoofdstuk 6.

In dit hoofdstuk hebben we een tweetal perspectieven op Learning Analytics beschreven en hebben we het begrip verder afgebakend. Tevens zijn we nader ingegaan op een aantal belangrijke ontwikkelingen die van invloed zijn op de wijze waarop onderwijsinstellingen omgaan en bezig zijn met Learning Analytics. We zijn het hoofdstuk geëindigd met een puntsgewijze opsomming van de belangrijkste bevindingen die naar voren kwamen tijdens de interviews met vertegenwoordigers van onderwijsinstellingen.

Tegen de achtergrond van de hier beschreven ontwikkelingen moeten de Learning Analytics architectuur en richtlijnen infrastructuur uit de volgende hoofdstukken begrepen worden. In de

architectuur en principes hebben wij een aantal keuzes gemaakt, die zijn terug te voeren op beschreven ontwikkelingen in dit hoofdstuk.

3 ARCHITECTUUR

Dit hoofdstuk start met de introductie van een aantal begrippen die samenhangen met een Learning Analytics architectuur. Daarna wordt beschreven wat een Learning Analytics architectuur is en uit welke componenten deze bestaat. Verderop in het hoofdstuk wordt aansluiting gezocht met de HORA en worden mogelijke architectuurinvullingen geschetst. Tot slot wordt een aantal diensten beschreven die nodig zijn om met Learning Analytics (architectuur) aan de slag te gaan. Met dit hoofdstuk wordt deelvraag 1 beantwoord (*uit welke (optionele en noodzakelijke) componenten en diensten bestaat een Learning Analytics architectuur?*).

3.1 Inleiding data-architectuur

In een (moderne) data-architectuur staat de stroom van data naar (bruikbare) informatie centraal. Dit in tegenstelling tot de bekendere informatiearchitectuur, waarin de organisatieprocessen, de daarin gebruikte informatie en applicaties die de processen ondersteunen centraal staan. Een data-architectuur heeft dus een “data-centrische” kijk op applicaties en de organisatie.

Een data-architectuur bestaat uit drie lagen:

- Registratielaag. In deze laag bevinden zich de bronsystemen en componenten van de digitale leeromgeving. Dit zijn systemen waar bijv. studenten, docenten en ondersteunend personeel gegevens vastleggen ten behoeve van hun werkzaamheden. De data die hier wordt vastgelegd vormt de basis voor de bewerkingen en analyses die met behulp van de applicaties in de andere lagen in de architectuur worden gedaan. Het vastleggen van data kan heel bewust gedaan worden (denk aan de invoer van een nieuwe student door de studentenadministratie) óf onbewust (denk aan de data die wordt vastgelegd als een student het LMS gebruikt voor zijn leerproces, waarbij de gebruiksduur van de applicatie wordt vastgelegd).
- Transformatielaag. De ruwe data die uit de bronsystemen komt heeft bewerking voordat deze geanalyseerd kan worden. De data uit verschillende systemen/componenten heeft vaak (subtiel) verschillende betekenis. Betekenissen moeten gelijk worden getrokken (bijv. applicaties die verschillende coderingen gebruiken voor opleidingen). In de transformatielaag worden tevens koppelingen gelegd tussen de data (bijv. gebruiksgegevens uit het LMS worden gekoppeld aan toetscijfers uit het digitale toetsstelsel). In de transformatielaag worden de benodigde datamodellen gemaakt en wordt de getransformeerde data opgeslagen (vaak in een datawarehouse).
- Informatielaag. De analyses en rapportages die nodig zijn om de gevraagde inzichten te creëren worden in de informatielaag gemaakt. Het kan hier gaan om data-science toepassingen, waarmee voorspellende inzichten worden gecreëerd of machine learning wordt toegepast. Ook wordt op deze laag de meer traditionele business intelligence verzorgd, waarmee voor de organisatie belangrijke indicatoren in dashboards en rapporten worden gepresenteerd.

3.2 xAPI en de Learning Records Store

De data-architectuur in onderwijsinstellingen wijkt op beperkte onderdelen af van die in andere organisaties¹². Voor e-learning applicaties is namelijk een aparte API-standaard ontwikkeld: de Experience API of xAPI-standaard. Deze standaard maakt het mogelijk voor onderwijsapplicaties, zoals learning management systems en digitale toetssystemen, om geautomatiseerd uniforme gebruiksdata (“statements”) te genereren. Op deze wijze wordt het zogenaamde gedistribueerd leren mogelijk gemaakt: studenten leren via diverse kanalen en met de xAPI-standaard komen deze kanalen samen.

Een xAPI statement bestaat uit een aantal onderdelen:

- Actor: het subject van de ervaring (bijv. student X)
- Werkwoord: datgene dat er gebeurd is (bijv. heeft afgerond)
- Object: de activiteit waarop de ervaring van de actor betrekking heeft (bijv. toets Y)
- Context: de context waarin de ervaring zich heeft afgespeeld. Bijvoorbeeld een toets is voor een bepaald vak, dat vak heeft een docent en de student zit in een bepaalde groep.
- Resultaat: de ervaring kan bepaalde uitkomsten hebben. Bijvoorbeeld een cijfer voor een toets en hoe lang de actor over de toets heeft gedaan.
- Extensies: de hiervoor genoemde onderdelen van een xAPI statement kunnen worden uitgebreid met extensies. Extensies geven extra informatie bij de onderdelen. Bijv. het gemiddelde cijfer van de toets of het aantal studenten dat in de groep bij de student zit.

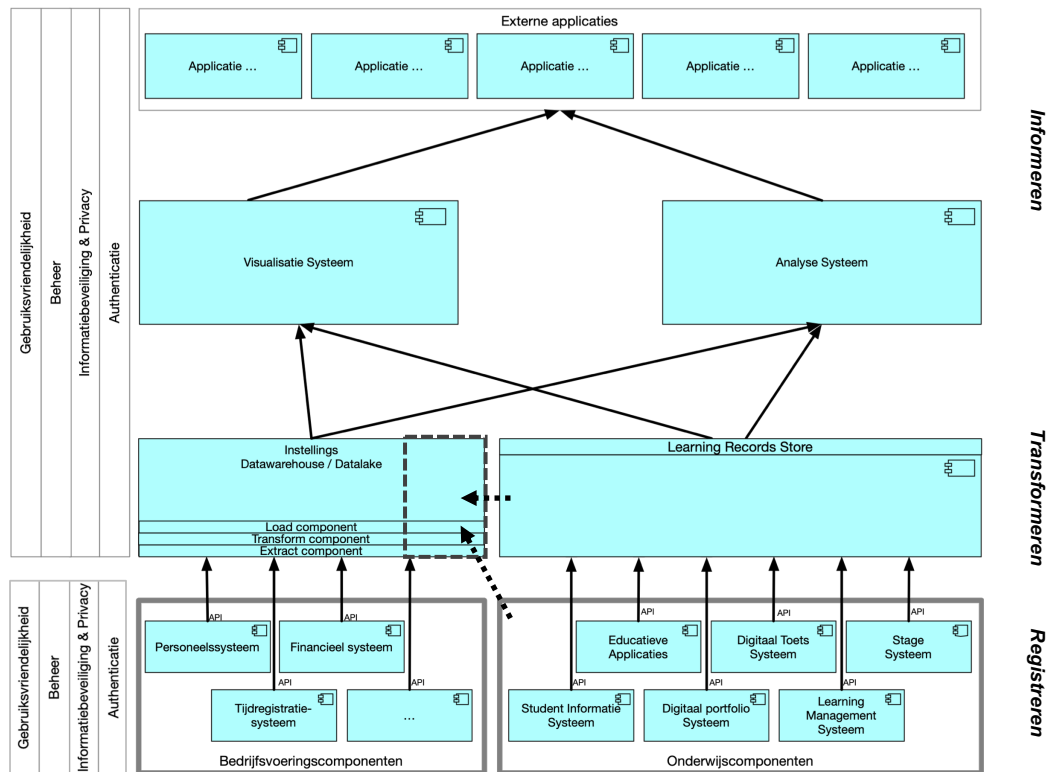
De xAPI statements worden verzameld in een Learning Records Store (LRS). Het doel van een LRS is het verkrijgen van xAPI statements (uit verschillende bronsystemen), het opslaan ervan en het beschikbaar stellen van de data. Een LRS is vaak een NoSQL document database. NoSQL databases hebben als kenmerk dat ze schaalbaar en snel zijn. Gelet op het grote aantal xAPI statements dat in een grote onderwijsinstelling dagelijks wordt gegenereerd, zijn schaalbaarheid en snelheid zeer belangrijk. Een LRS kan data beschikbaar stellen aan andere systemen, zoals datawarehouses, analysesystemen en rapportage-/visualisatiesystemen. Een LRS bevindt zich dus op de transformatielaag van de data-architectuur en is essentieel voor het bruikbaar maken van alle gebruiksdata (xAPI statements) die door bronsystemen worden gegenereerd¹³.

3.3 Learning Analytics architectuur

De data-architectuur voor onderwijsinstellingen, de zogenaamde Learning Analytics architectuur, die gebruikt wordt voor dit onderzoek kent een registratielaag, transformatielaag – met een LRS waarin xAPI statements worden opgeslagen en verwerkt – en een informatielaag. Schematisch ziet deze architectuur er als volgt uit:

¹² Uiteraard is het gebruiken van onderwijsapplicaties niet beperkt tot onderwijsinstellingen. Ook (grote) bedrijven en overheidsorganisaties kunnen gebruik maken van onderwijsapplicaties voor in-house trainingen en cursussen.

¹³ Lang niet alle onderwijsapplicaties maken overigens gebruik van de xAPI-standaard. Zie hiervoor tevens hoofdstukken 5 en 6 van dit rapport.



Figuur 4: Learning Analytics architectuur

De registratielaag bestaat uit twee subonderdelen. De bedrijfsvoeringsapplicaties staan los van de educatieve applicaties. Via API's of batches wordt data uit de bedrijfsvoeringsapplicaties direct geladen in het datawarehouse van de instelling. Educatieve applicaties die gebruik maken van de xAPI-standaard gebruiken een LRS voor opslag en transformatie, zoals beschreven in voorgaande paragraaf. De gestippelde pijl van de onderwijsapplicaties naar het datawarehouse duidt erop dat in de praktijk niet alle educatieve applicaties gebruik maken van de xAPI-standaard. De data uit deze applicaties wordt getransformeerd en opgeslagen in (een apart deel van) het instellings-datawarehouse.

Zoals beschreven in voorgaande paragraaf bestaat de transformatielaag uit een LRS en een "regulier" datawarehouse. Om data uit het LRS te koppelen met data uit educatieve applicaties die geen gebruik maken van de xAPI-standaard of data uit bedrijfsvoeringsapplicaties, dient er een verbinding te zijn tussen het LRS en (een apart deel van) het datawarehouse. Dit is weergegeven met de gestippelde pijl.

Tot slot zijn er op de informatielaag drie componenten te vinden. Dit zijn visualisatiesystemen (denk aan dashboard- en rapportagetools), analysesystemen (bijv. R- of Pythonstudio's of geïntegreerde data science-applicaties) en externe applicaties. De analyses, dashboards en rapportages kunnen gedistribueerd worden in de applicaties waarin ze zijn ontwikkeld (vooral dashboardtools bieden hier uitgebreide mogelijkheden voor), maar ook via externe applicaties.

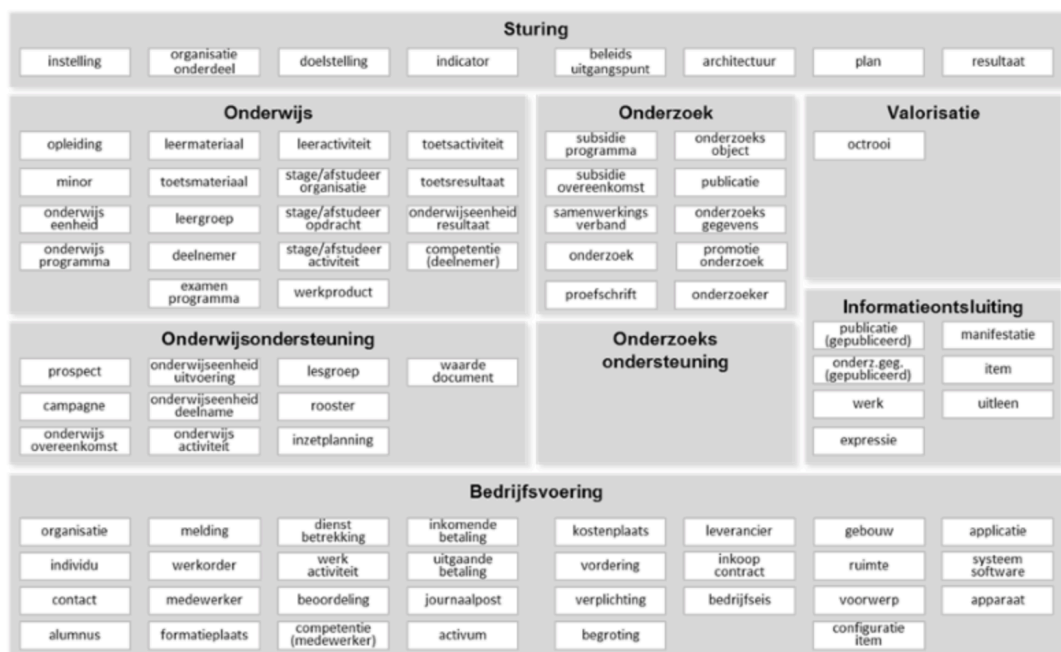
Een voorbeeld is een app die studenten op basis van een in een analysesysteem ontwikkeld algoritme ondersteunt bij bepaalde zaken.

In het vervolg van dit rapport worden de componenten aangeduid met letters:

- A. Learning records store
- B. Analysesysteem
- C. Visualisatiesysteem
- D. Onderwijscomponenten (bronsystemen)
- E. (Instellingen) datawarehouse/datalake

3.4 Aansluiting met HORA

De Hoger Onderwijs Referentiearchitectuur (HORA) is een verzameling van instrumenten voor het inrichten van de organisatie en informatiehuishouding van Nederlandse instellingen voor Hoger Onderwijs. Het inrichten en beheersen van de informatiehuishouding wordt steeds complexer door technologische ontwikkelingen en steeds groter wordende verantwoordingseisen vanuit de overheid. De HORA helpt instellingen bij het creëren van de noodzakelijke samenhang en kwaliteit en bij het beheersen van risico's. In de HORA wordt geen expliciete aandacht besteed aan analytics. De Learning Analytics architectuur zoals geschetst in vorige paragraaf raakt primair de HORA-onderdelen "onderwijs", "onderwijsondersteuning", "informatieontsluiting" en "sturing". Hierbij vormen de onderdelen onderwijs en onderwijsondersteuning bronnen. Analyse en rapportage vindt plaats bij informatieontsluiting. Sturing is noodzakelijk voor vraagarticulatie en het kiezen van indicatoren. In de figuur hierna is de HORA weergegeven.

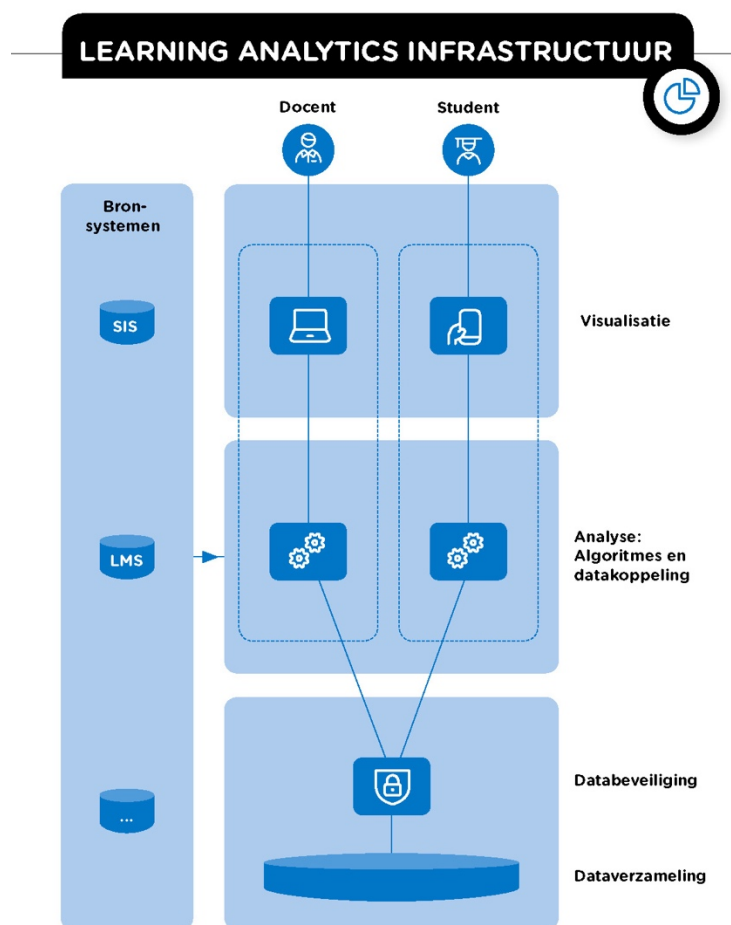


Figuur 5: HORA

3.5 Architectuurinvullingen

De Learning Analytics architectuur zoals beschreven in paragraaf 3.3 kan op verschillende manieren worden ingevuld. Zo kunnen de verschillende bronsystemen ("onderwijscomponenten") met behulp van xAPI-statements data pushen naar één LRS. Met analyse- of visualisatiesystemen

wordt een deel van de data uit het LRS, eventueel na een verrijking met andere data, geanalyseerd en gepresenteerd. Op deze wijze ontstaan er allerlei “services”. Het gaat bijv. om een service die inzicht geeft in het gemiddelde cijfer van de toetsen binnen een bepaald vak of om een complexere service die voorspelt of studenten een vak gaan halen o.b.v. fysieke en digitale activiteit. In de figuur hieronder is dit weergegeven.



Figuur 6: service gerichte Learning Analytics architectuur

Sommige Learning Management Systemen (LMS) beschikken over een eigen LRS. In de praktijk komt het voor dat data uit andere onderwijscomponenten aan dit LRS wordt toegevoegd. Het LRS van de LMS dient dan als hét LRS van de instelling.

3.6 Learning Analytics diensten

In dit hoofdstuk heeft tot nu toe de focus gelegen op techniek. Om succesvol te zijn met Learning Analytics is ook een aantal diensten nodig, als aanvulling op de geschetste architectuur:

- Privacy en ethische vraagstukken rondom Learning Analytics zijn complex. Het vergt de nodige expertise om deze vraagstukken te adresseren. Er is advies en begeleiding nodig om Learning Analytics toe te passen zonder privacy van studenten en medewerkers te schenden of ethische grenzen te overschrijden.

- Uit de interviews die in het kader van dit onderzoek zijn gehouden bleek dat het “data-volwassenheidsniveau” van medewerkers van instellingen over het algemeen laag is. Medewerkers dienen geholpen te worden bij het stellen van de juiste vragen die m.b.v. Learning Analytics beantwoord kunnen worden. Er is behoefte aan advies/begeleiding bij vraagarticulatie.
- Een Learning Analytics traject is technisch uitdagend. De benodigde data-science expertise dient de instelling in huis te hebben. Daarnaast hebben instellingen behoefte aan data-engineers die ervoor zorgen dat data-scientists over de benodigde data kunnen beschikken. Een projectmanager is nodig om aansluiting te realiseren bij de behoeften van de eindgebruikers en in het goede banen leiden van alle facetten van het project.
- Omdat het data-volwassenheidsniveau over het algemeen laag is, zullen veel medewerkers binnen instellingen moeite hebben met het juist interpreteren en toepassen van Learning Analytics inzichten. Een consultant of change manager kan hen helpen bij acceptatie en toepassing van de inzichten en het aanpassen van processen op basis daarvan.

4 RICHTLIJNEN INFRASTRUCTUUR

De richtlijnen infrastructuur – de “meetlat” waarlangs de oplossingen van leveranciers worden gelegd – worden in dit hoofdstuk geïntroduceerd. Tevens worden de richtlijnen gerelateerd aan de in het vorige hoofdstuk geïntroduceerde Learning Analytics architectuur en wordt dieper ingegaan op gegevensverwerking in derde (niet-EU) landen.

4.1 Richtlijnen infrastructuur

In samenwerking tussen betrokkenen vanuit SURF, de onderzoekers en privacy-experts van VKA is een set van 45 richtlijnen opgesteld waaraan de oplossingen van leveranciers zijn getoetst. Deze richtlijnen hebben betrekking op privacy, niet-functionele eisen en functionele eisen. De richtlijnen zijn gecategoriseerd rondom de volgende thema's (tussen haakjes staan de nummers van de richtlijnen):

Privacy

- a. Rechten van betrokkenen (1 – 4, 11 , 12)
- b. Locatie verwerking (5, 6)
- c. Transparantie algoritmes (7, 8)
- d. Informatiebeveiliging (9, 10, 19)
- e. Privacy by design (13 – 18)

Niet functioneel

- f. Toegankelijkheid (20)
- g. Open source (21)
- h. Metadata (22)
- i. Authenticatie (23)
- j. Gebruiksvriendelijkheid (24)
- k. Evidence based (25, 26)
- l. Ethiek (27)

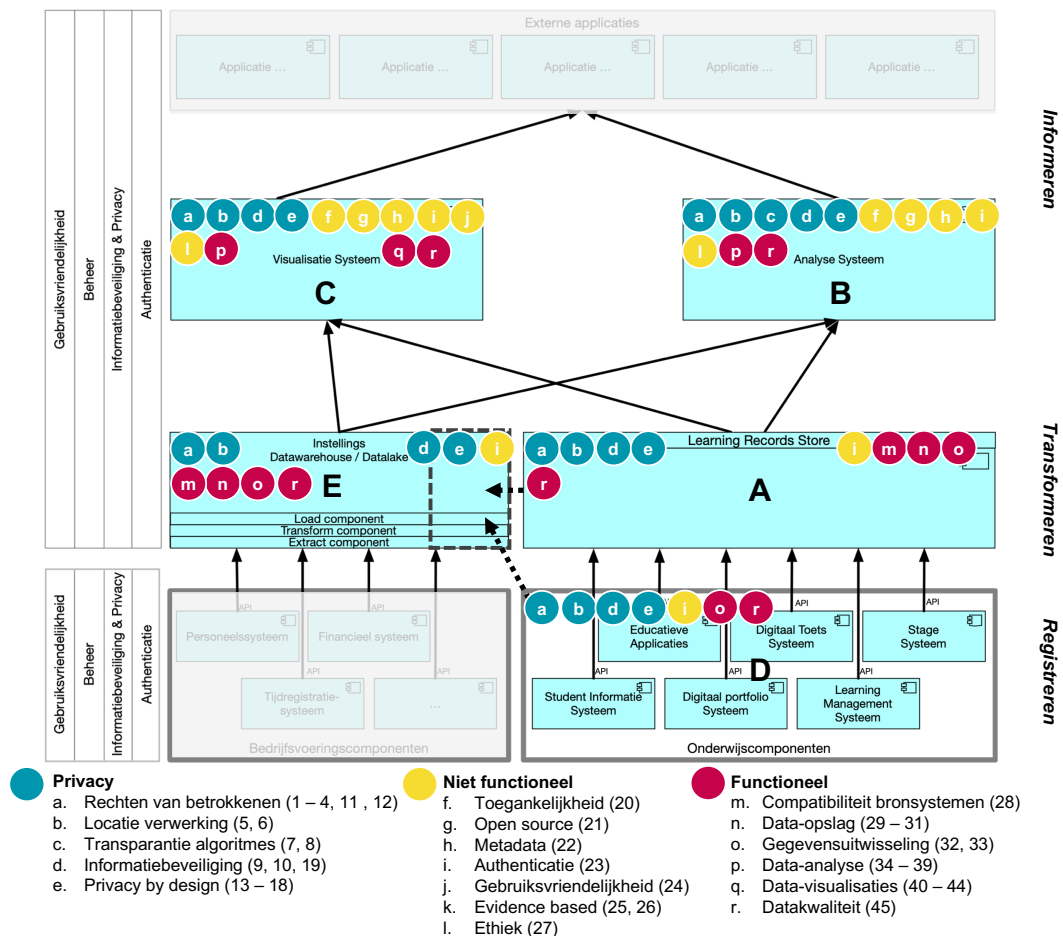
Functioneel

- m. Compatibiliteit bronsystemen (28)
- n. Data-opslag (29 – 31)
- o. Gegevensuitwisseling (32, 33)
- p. Data-analyse (34 – 39)
- q. Data-visualisaties (40 – 44)
- r. Datakwaliteit (45)

Gelet op het belang van privacy van studenten en medewerkers in het onderwijs, ligt het zwaartepunt van de richtlijnen op privacy en gegevensbescherming. De niet-functionele eisen waarborgen een aantal andere gerelateerde (publieke) waarden, o.a. door het maken van ethische afwegingen, zoals toegankelijkheid en gebruik van wetenschappelijk bewijs. De functionele richtlijnen toetsen of de oplossingen voldoende mogelijkheden bieden om de benodigde analyses uit te voeren en uitkomsten op een goede manier te presenteren. De volledige lijst richtlijnen is opgenomen in bijlage B.

4.2 Toepassing richtlijnen op Learning Analytics architectuur

Een aantal richtlijnen zijn van toepassing op alle componenten van de Learning Analytics architectuur. Daarentegen zijn er ook richtlijnen die van toepassing zijn op één of enkele componenten. In de volgende figuur zijn de thema's (en daarmee de onderliggende richtlijnen) toegepast op de Learning Analytics architectuur.



Figuur 7: Toepassing richtlijnen op Learning Analytics architectuur

4.3 Gegevensverwerking in derde landen

In principe mogen organisaties persoonsgegevens van EU-ingezetenen alleen verwerken in landen binnen de EU. Onder bepaalde voorwaarden mogen data in een zogenaamd derde (niet-EU) land worden verwerkt. Deze voorwaarden zijn vervat in hoofdstuk V van de AVG. Aangezien privacy en gegevensbescherming belangrijke thema's zijn in de richtlijnen, zijn deze voorwaarden hierna op hoofdlijnen uitgewerkt:

- Bepaling van de Europese Commissie (EC) dat een (sector in een) bepaald land een voldoende beschermingsniveau van persoonsgegevens biedt. Hieronder valt bijvoorbeeld het Privacy Shield raamwerk. Dit is een afsprakenstelsel tussen de EU en de VS waarin de bescherming van persoonsgegevens van EU-ingezetenen wordt gewaarborgd wanneer deze in de VS worden verwerkt. Amerikaanse organisaties kunnen zich (laten) certificeren onder dit raamwerk, waarmee zij aantonen over een voldoende gegevensbeschermingsniveau beschikken.
- De Europese data-exporteur en derde data-importeur sluiten een door de EC goedgekeurde modelovereenkomst of een overeenkomst die wordt goedgekeurd door de toezichthouder in een EU-lidstaat.

- Multinationals kunnen met bindende bedrijfsvoorschriften privacyregels vaststellen voor entiteiten buiten de EU. De toezichthouder in een EU-lidstaat toetst of de bindende bedrijfsvoorschriften voldoen aan de AVG en keurt deze al niet goed.

5 TOEPASSING RICHTLIJNEN INFRASTRUCTUUR OP OPLOSSINGEN LEVERANCIERS

Dit vijfde hoofdstuk staat in het teken van de bevindingen van het leveranciersonderzoek. Na een onderbouwing voor de selectie van leveranciers, worden per richtlijnthema, zoals beschreven in hoofdstuk 4, de bevindingen gepresenteerd. Dit hoofdstuk geeft antwoord op deelvragen 2, 3 en 4 (*welke van deze componenten en diensten worden nu al door welke partijen aangeboden?; in hoeverre voldoen deze componenten en diensten aan de richtlijnen en kernwaarden van SURF?; welke ruimte bieden de aanbieders van de componenten en diensten om – wanneer ze dat nog niet doen – te voldoen aan de beleidsprincipes en kernwaarden van SURF op het gebied van Learning Analytics?*).

5.1 Selectie leveranciers

In de Learning Analytics architectuur is onderscheid gemaakt tussen drie lagen met elk hun eigen onderdelen. Voor de onderdelen onderwijscomponenten, learning records store, analysesystemen (inclusief bijbehorend opslagcomponent, zie hierna) en visualisatiesystemen zijn 3 tot 5 oplossingen van leveranciers onderzocht. Het betreft de oplossingen van de grootste leveranciers per onderdeel. Voor de analysesystemen geldt dat deze onderdeel zijn van het cloud (analytics) aanbod van de leverancier en daar niet los van kunnen worden gezien. De benodigde opslag en rekenkracht worden geleverd vanuit de cloud. Bij de analysesystemen is dus ook gekeken naar de functionaliteit van de opslagcomponent (er is geen apart onderzoek gedaan naar datawarehouse/datalake oplossingen).

Tevens is een aantal geïntegreerde Learning Analytics oplossingen onderzocht. Deze oplossingen bieden meerdere onderdelen aan in één pakket (opslag en verwerking van data uit bronsystemen, uitvoeren van analyses a.d.h.v. standaardalgoritmes en distribueren van de uitkomsten in apps). De oplossingen worden als dienst aangeboden. Aangezien deze oplossingen verschillende architectuurcomponenten combineren in één pakket, worden deze als aparte “architectuurcomponent” gezien en aangeduid met een F om onderscheid te maken met de andere componenten. In bijlage C is een overzicht van de onderzochte oplossingen weergegeven.

5.2 Bevindingen toets oplossingen aan richtlijnen infrastructuur

In de volgende tabel zijn de bevindingen van het onderzoek naar de oplossingen in de markt beschreven. Hierin wordt onderscheid gemaakt in algemene bevindingen, geldend voor (bijna) alle oplossingen en specifieke bevindingen die bijzonder zijn voor één of een klein aantal oplossingen. In bijlage D zijn de bevindingen per leverancier opgenomen.

Thema	Toepassing	Algemene bevindingen	Specifieke bevindingen
a. Rechten van betrokkenen	A – F	In principe is het aan een Privacy Officer (uitgevoerd door een beheerder) van de instelling om verzoeken van gebruikers om inzage of verwijdering van persoonsgegevens te behandelen. Soms is hiervoor standaardfunctionaliteit ingebouwd in de oplossing (bijv. een knop voor studenten en een portal voor beheerders). Voor de oplossingen waarbij dit niet het geval is, dienen studenten contact op te nemen met een Privacy Officer van de instelling.	Een aantal oplossingen kan geconfigureerd worden om zonder persoonsgegevens te werken (bijv. door alleen gebruik te maken van sleutels). Sommige oplossingen bieden uitgebreide mogelijkheden voor

			Privacy Officers/beheerders om verzoeken van studenten in bulk in te zien en af te handelen.
b. Locatie verwerking	A – F	De onderzochte partijen zijn bijna allemaal Amerikaans of maken gebruik van Amerikaanse hosting partners. De Europese Commissie heeft bepaald dat de VS bij wet onvoldoende privacybescherming bieden. Europese organisaties hebben echter een aantal grondslagen waarop ze persoonsgegevens mogen delen met Amerikaanse partijen (hoofdstuk V van de AVG, zie paragraaf 4.3). Onder het EU-US Privacy Shield raamwerk hebben de EU en de VS afspraken gemaakt over gegevensbescherming. Praktisch alle leveranciers in het onderzoek zijn gecertificeerd onder het EU-US Privacy Shield. Wanneer een bedrijf gecertificeerd is onder het Privacy Shield biedt deze op papier voldoende waarborgen voor bescherming van de privacy van EU-ingezetenen. De vraag is echter of dit daadwerkelijk zo is. Er loopt op dit moment namelijk een rechtszaak bij het Europees Hof van Justitie ("Schrems II") waarin het Privacy Shield en de standaard contractbepalingen (een andere manier waarop niet-Europese bedrijven kunnen aantonen dat zij voldoende privacybescherming bieden) ter discussie staan. De klacht in deze rechtszaak is dat beide instrumenten niet genoeg bescherming bieden tegen inzage van (persoons)gegevens door Amerikaanse overheden. De meeste leveranciers bieden de mogelijkheid om een datacenter binnen de EER te kiezen voor opslag en verwerking van de data.	Een klein aantal leveranciers geeft aan dat data de EER kan verlaten met het oog op continuïteit van dienstverlening en/of het bieden van ondersteuning. Een klein aantal leveranciers geven door middel van bindende bedrijfsvoorschriften en/of standaard contractbepalingen aan over een adequaat beschermingsniveau te beschikken tegen mogelijke privacy-schendingen door Amerikaanse overheden.
c. Transparantie	B en F	Leveranciers van analysesystemen geven aan dat de verwerkingsroutes (pipelines) van de data zo transparant mogelijk zijn. Daarbij geven ze aan dat sommige machine learning toepassingen kenmerken van een 'black box' kunnen hebben. De onderzochte integrale oplossingen geven aan transparant te zijn over verwerkingsroutes, onduidelijk is echter hoe. Onduidelijk is in hoeverre algoritmes in zijn te zien.	
d. Informatie-beveiliging	A – F	Bijna alle leveranciers zijn gecertificeerd conform de ISO27001 standaard voor gegevensbescherming. Alle oplossingen waarvoor dit bekend is, bieden uitgebreide autorisatiemogelijkheden aan en/of een koppeling met organisatiebrede autorisatiemethoden. De meeste oplossingen loggen het gebruik.	Sommige applicaties dienen door een implementatiepartner of de instelling zelf te worden gehost. Die partner of de instelling dient dan informatiebeveiliging op het juiste niveau te realiseren
e. Privacy by design	A – F	De verantwoordelijkheid voor Privacy by Design ligt voor de meeste oplossingen bij de instelling. Het configureren van bronssystemen (xAPI) om zo min mogelijk of geen persoonsgegevens door te sturen dient een instelling zelf te (laten) doen.	Een enkele oplossing pseudonimiseert persoonsgegevens automatisch

		Leveranciers geven aan wettelijke bewaartermijnen te respecteren. Oplossingen bieden vaak Privacy Officers/beheerders van instellingen de mogelijkheid om bewaartermijnen (soms zelf per gegeven) in te stellen. Het gescheiden opslaan van persoonsgegevens en niet-persoonsgegevens is een verantwoordelijkheid van de instelling.	
f. Toegankelijkheid	B, C en F	De informatie die leveranciers geven over toegankelijkheid is beperkt. Een aantal leveranciers geven aan toegankelijkheid te bevorderen door voorleesfunctie, vergrootglas en keyboard-only commando's.	
g. Open source	B, C en F	Analysesystemen en visualisatiesoftware bieden de mogelijkheid om de pipeline en queries (code) in te zien. Deze kunnen desgewenst gedeeld worden. Van de geïntegreerde oplossingen is het onduidelijk in hoeverre deze open source zijn. Het lijkt erop dat de meeste leveranciers de code van hun standaardalgoritmes niet delen.	
h. Metadata	B, C en F	Er is weinig informatie gevonden over metadata. Een aantal oplossingen biedt een directory met kolom- en tabeldefinities of heeft een openbaar logisch datamodel. Daarnaast bieden de xAPI-standaard en LRS handvatten voor goede metadatering.	
i. Authenticatie	A – F	Bijna alle oplossingen beschikken over een eigen authenticatiedienst (gebruik makend van standaarden) of kunnen aansluiten bij de authenticatiedienst van de instelling.	
j. Gebruiks-vriendelijkheid	C en F	Alle visualisatiesystemen bieden de mogelijkheid om dashboards en rapportages te ontwikkelen geschikt voor elk type device	Aan studenten worden soms apps aangeboden die alleen via mobiel zijn te raadplegen
k. Evidence based	F	Voor zover bekend geven de leveranciers aan dat hun standaardalgoritmes evidence based zijn. Dit is echter lastig te verifiëren. Leveranciers hebben vaak processen ingericht waarmee ze hun voorspelmodellen verbeteren gebaseerd op meer en/of gedetailleerdere data.	
l. Ethiek	B, C en F	De verantwoordelijkheid voor het maken van ethische afwegingen ligt primair bij de instelling.	Een leverancier geeft aan dat zij nadrukkelijk geen gebruik maakt van demografische kenmerken van studenten, omdat zij dit onethisch vindt. Voor een andere leverancier is demografie juist een kernelement in haar voorspelmodellen.
m. Compatibiliteit bronsystemen	A en E	De compatibiliteit met de bronsystemen ligt primair aan de functionaliteit van bronsystemen. Wanneer deze de xAPI-standaard ondersteunen, dan kunnen xAPI statements worden gepusht naar de LRS. Een aantal van de onderzochte bronsystemen ondersteunen de xAPI-standaard echter (nog) niet. De data uit deze systemen	Een van de analysesystemen biedt via haar marketplace een LRS aan die standaard gekoppeld kan worden aan het analysesysteem

		kan met REST API's of CSV-bestanden beschikbaar worden gesteld aan datawarehouses of datalakes.	
n. Dataopslag	A en E	De meeste learning record stores zijn NoSQL databases en slaan data ongestructureerd op. Met behulp van (externe) tools kunnen de data in de LRS'en omgezet worden in een gestructureerd formaat. De ruwe ongestructureerde data is eveneens te queryen met (externe) tools. Ook de onderzochte analysesystemen bieden opslagmogelijkheden voor gestructureerde en ongestructureerde data.	
o. Gegevens-uitwisseling	A, D, E en F	De onderzochte LRS'en maken per definitie gebruik van de xAPI-standaard om gegevens uit te wisselen met bronsystemen. M.b.v. (REST) API's of CSV-bestanden kan de data uit een LRS beschikbaar worden gesteld aan analyse- en visualisatiesystemen. Zoals hiervoor gesteld ondersteunen niet alle onderzochte bronsystemen de xAPI-standaard. De onderzochte analysesystemen kunnen op diverse wijzen (API's, CSV-bestanden) gegevens uitwisselen. Voor de geïntegreerde LA-oplossingen die zijn onderzocht is onduidelijk in hoeverre de data uitgewisseld kan worden met andere applicaties.	
p. Data-analyse	B en C	De onderzochte analysesystemen zijn van de belangrijkste spelers in de markt. Deze bieden allen mogelijkheden om gestructureerde en ongestructureerde data te analyseren en daar m.b.v. data science technieken (voorspellende) inzichten mee te creëren. De meeste onderzochte analysesystemen bieden een codearme "drag & drop" omgeving aan, naast of in combinatie met de mogelijkheid om R en Python code te schrijven. De analysemogelijkheden in de visualisatiesystemen zijn veel beperkter. Wel bieden de visualisatiesystemen de mogelijkheid om (m.b.v. plug-ins) R en Python scripts te schrijven en/of te draaien t.b.v. analyses.	
q. Data-visualisaties	C	De visualisatiesystemen die onderzocht zijn, zijn leidend in de markt. Zij bieden allen een "drag & drop" omgeving waarin zonder code te schrijven dashboards kunnen worden gemaakt. Dashboards kunnen worden gedeeld op diverse devices. Tevens is het mogelijk bij alle systemen om dashboards te embedden in derde applicaties (bijv. websites).	
r. Datakwaliteit	A – F	Het waarborgen van datakwaliteit is op alle plekken in de data-architectuur relevant. In het onderzoek is niet kunnen achterhalen welke functionaliteit leveranciers bieden rondom datakwaliteit.	Eén leverancier biedt een datavalidatie tool aan die instellingen helpt bij het waarborgen van voldoende datakwaliteit

6 CONCLUSIES EN AANBEVELINGEN

In dit laatste hoofdstuk worden de conclusies getrokken op basis van de bevindingen van het leveranciersonderzoek uit hoofdstuk 5. De conclusies vormen de aanzet voor de aanbevelingen voor SURF over positionering in het Learning Analytics veld. Er is onderscheid gemaakt tussen aanbevelingen gebaseerd op het leveranciersonderzoek en de gesprekken met instellingen. Hiermee wordt antwoord gegeven op deelvraag 5 (*wat bieden de antwoorden op de eerste vier vragen voor mogelijkheden voor SURF ten aanzien van Learning Analytics componenten en diensten?*).

6.1 Conclusies oplossingen leveranciers

Privacy

De eerste conclusie van het onderzoek onder marktpartijen is dat de verantwoordelijkheid van het waarborgen van privacy primair bij instellingen ligt. Zij dienen systemen dusdanig in te richten dat er geen of zo min mogelijk persoonsgegevens worden uitgewisseld tussen de componenten (privacy by design). Het anonimiseren of pseudonimiseren van persoonsgegevens en het apart opslaan van persoonsgegevens moet een instelling zelf regelen, eventueel met ondersteuning van leveranciers (bijv. door analyses uit te voeren op gepseudonimiseerde data en pas wanneer een interventie richting een persoon gewenst is de gegevens van die persoon erbij te halen).

Ook het borgen van de rechten van betrokkenen ligt bij de instelling. Zij dienen een proces in te richten waarmee zij verzoeken van studenten om inzage en verwijdering van persoonsgegevens kunnen afhandelen. Het is overigens ook niet wenselijk om dit op een andere manier te doen (bijv. door leverancier deze verzoeken af te laten handelen). Sommige leveranciers bieden functionaliteit waarmee studenten gemakkelijk verzoeken in kunnen dienen en Privacy Officers/beheerders deze gemakkelijk af kunnen handelen.

Hoewel het begrijpen en kunnen uitleggen van algoritmes geen eenvoudige opgave is, kan een instelling, wanneer deze zelf algoritmes ontwikkelt, transparantie van beslissingen genomen o.b.v. deze algoritmes borgen. Leveranciers van oplossingen die werken met standaardalgoritmes zullen de werking van die algoritmes echter niet zomaar prijsgeven. De werking van hun algoritmes is immers een belangrijk onderdeel van hun verdienmodel. Instellingen moeten zich bewust zijn van de risico's die het gebrek aan transparantie met zich meebrengt (discriminatie, uitsluiting, onterechte bevoor- of benadeling etc.).

De markt wordt gedomineerd door Amerikaanse partijen. Ook niet-Amerikaanse leveranciers maken vaak gebruik van Amerikaanse hosting partners. Het voordeel van deze grote (hosting)partijen is dat zij door de enorme schaal waarop zij werken een hoog niveau van bescherming kunnen bieden tegen hacks en andere beveiligingsdreigingen. De vraag is echter hoe deze Amerikaanse partijen omgaan met de (persoons)gegevens die zijn opgeslagen op hun servers. De onderzochte partijen zijn gecertificeerd onder het EU US Privacy Shield raamwerk. De vraag is echter of dit raamwerk voldoende waarborgen biedt, gelet op de rechtszaak die op dit moment loopt bij het Europees Hof van Justitie ("Schrems II"). Instellingen dienen een risico- en kosten-batenafweging te maken, voordat zij in zee gaan met een Amerikaanse (sub)leveranciers. Weegt het risico van mogelijke privacyschendingen door Amerikaanse overheden op tegen hogere kosten

van het alternatief (bijv. on-premise)? Daarbij is het de vraag of het überhaupt wenselijk is om zo afhankelijk te zijn van Amerikaanse bedrijven¹¹.

Niet-functioneel

In het verlengde van de conclusie omtrent transparantie, dienen instellingen te verifiëren in hoeverre de standaardalgoritmes van leveranciers evidence based zijn. In dit onderzoek is onvoldoende onderbouwing van algoritmes gevonden. Tevens ligt de verantwoordelijkheid voor het maken van ethische afwegingen bij gebruik van analyses en algoritmes bij de instelling. Leveranciers hanteren namelijk uiteenlopende ethische standaarden.

Functioneel

Gelet op de verschillende oplossingen in de markt zijn er primair twee manieren om Learning Analytics toe te passen:

- Gebruikmaken van een geïntegreerde oplossing die data ophaalt uit bronsystemen, opslaat, a.d.h.v. standaardalgoritmes analyseert en de uitkomsten rapporteert in een app, LMS of portal. Het voordeel van deze oplossingen is dat er snel en zonder grote voorinvesteringen inzichten kunnen worden gegenereerd. De flexibiliteit van deze systemen is vaak beperkt door het gebruik van standaardalgoritmes. Tevens zijn de algoritmes vaak niet of onvoldoende transparant, waardoor het lastig is om na te gaan op welke grond inzichten tot stand komen (en de acties die daaruit voortvloeien).
- Een instelling ontwikkelt zelf algoritmes door data uit een LRS te analyseren in een analysesysteem. Het voordeel is dat de instelling zelf controle heeft over het algoritme en de data die wordt gebruikt om het algoritme te ontwikkelen of trainen. Het algoritme is transparant en er kunnen waarborgen rondom privacy en ethiek worden ingebouwd. Wanneer een instelling een algoritme zelf heeft ontwikkeld, kan het deze eventueel delen met collega-instellingen. Het nadeel is dat de instelling investeringen moet doen in de verschillende onderdelen van een Learning Analytics architectuur. Tevens heeft de instelling de benodigde data-science expertise in huis nodig. Het bieden van zekerheid rond de ontwikkelde algoritmes is bovendien niet eenvoudig.

Er zijn standaarden in de markt, maar lang niet alle oplossingen ondersteunen deze standaarden. Een belangrijk voorbeeld is de xAPI-standaard die door een deel van de onderzochte bronsystemen niet of beperkt wordt ondersteund. Dit maakt het voor instellingen lastig om één generieke data-architectuur te kiezen.

6.2 Aanbevelingen

Gelet op de hiervoor beschreven conclusies en de behoeften die bij instellingen zijn geïnventariseerd op het gebied van Learning Analytics, zijn er diverse mogelijkheden voor SURF ten aanzien van Learning Analytics componenten en diensten:

- Gebaseerd op conclusies leveranciersonderzoek:
 - I. Opstellen van een toolkit voor toepassing van Learning Analytics met inachtneming van privacy
 - II. Waarborgen van goede algoritmes door opstellen/toepassing van een algoritme toetsingskader
 - III. Het ondersteunen van instellingen bij het maken van ethische afwegingen

- IV. Stimuleren van standaardisatie en het maken van generieke afspraken daarover met leveranciers
- V. Aanbieden van reken- en opslagcapaciteit om de afhankelijkheid van Amerikaanse partijen te verminderen
- Gebaseerd op gesprekken met instellingen:
 - VI. Overzicht bieden van oplossingen en inkoop daarvan vergemakkelijken
 - VII. Het (blijven) stimuleren van kennisdeling
 - VIII. Inzetten van consultants en projectmanagers bij instellingen voor praktische ondersteuning bij Learning Analytics projecten

Hieronder worden deze aanbevelingen verder uitgewerkt.

Gebaseerd op conclusies leveranciersonderzoek

I. Toolkit privacy en Learning Analytics

De verantwoordelijkheid van instellingen voor het borgen van privacy van studenten en werknemers is groot. SURF kan instellingen helpen door een toolkit met praktische kaders en modellen op te stellen waarmee instellingen op een privacybestendige manier Learning Analytics kunnen toepassen. In deze toolkit dient in ieder geval aandacht te worden besteed aan:

- Privacy by design: een model dat aangeeft welke privacy by design technieken waar in de Learning Analytics architectuur kunnen/moeten worden toegepast
- Proces rondom rechten van betrokkenen: modelprocesbeschrijvingen om rechten van betrokkenen af te handelen. Daarbij een set functionele eisen die een instelling kan stellen aan oplossingen om het voor betrokkenen en Privacy Officers/beheerders zo makkelijk mogelijk te maken om verzoeken in te dienen en af te handelen.
- Afwegingskader voor gebruik Amerikaanse producten: er zitten risico's aan in zee gaan met Amerikaanse leveranciers. Daarentegen is het alternatief ook niet zonder risico's en kosten. Een afwegingskader helpt instellingen om de benodigde risico- en kosten-/batenanalyses uit te voeren.

Naast de toolkit is er behoefte aan praktische ondersteuning "op locatie" bij instellingen om Learning Analytics op een verantwoorde manier toe te passen. SURF kan hiertoe via bijv. een marktplaats experts van marktpartijen inzetten (analoog aan de inzet van consultants en projectmanagers, zie laatste aanbeveling).

II. Algoritme assurance

Het verifiëren van een algoritme, zeker wanneer deze niet door de instelling is ontwikkeld, is uitdagend. SURF kan instellingen helpen door een toetsingskader op te stellen (of te selecteren uit het reeds bestaande aanbod) waarmee instellingen algoritmes kunnen toetsen. "Transparantie" en "evidence based" dienen in ieder geval onderwerpen in dit toetsingskader te zijn. Additioneel kan SURF ondersteuning bieden aan instellingen bij het toepassen van dit toetsingskader door marktpartijen met expertise op dit gebied aan zich te verbinden. SURF kan eveneens een trusted third party zoeken die de algoritmes van commerciële partijen toetst.

III. Ethiek

De verantwoordelijkheid voor het maken van ethische afwegingen ligt bij de instelling. Er zijn diverse ethische hulpmiddelen beschikbaar. Een voorbeeld is de DEDA (De Ethische Data Assistent) -methodiek ontwikkeld door de Universiteit Utrecht. SURF kan instellingen helpen door deze of

een andere methodiek toe te passen als dienst bij instellingen of door instellingen te laten trainen door experts in het gebruik van de methodiek.

In de Versnellingszone wordt door SURF en instellingen reeds gewerkt aan een 'Code of conduct' voor Learning Analytics waarin ethiek een belangrijk aspect is.

IV. Standaardisatie

Lang niet alle leveranciers ondersteunen de bestaande standaarden. Standaarden zijn nodig om complexiteit (en dus kosten en risico's) van Learning Analytics architecturen bij instellingen te verminderen. SURF kan door gesprekken aan te gaan met leveranciers standaardisatie stimuleren en generieke afspraken maken over standaarden. Hierin kan zij waar nodig en mogelijk samen optrekken met zusterorganisaties in Europa. Tevens kan zij met behulp van een inkoopstelsel leveranciers die zich conformeren aan standaarden aanbevelen.

Een andere wijze waarop SURF standaardisatie kan stimuleren is door standaard datamodellen en metadata voor Learning Analytics te ontwikkelen of adopteren¹⁴. Dit maakt het voor leveranciers gemakkelijker zich te conformeren aan de gewenste standaardsemantiek.

Tot slot kan SURF landelijke open databronnen rondom onderwijs beter ontsluiten. Open databronnen worden nu nog versnipperd aangeboden (DUO, CBS etc.). SURF kan instellingen helpen veel meer waarde uit deze open data te halen door deze beter te ontsluiten en eventueel een aantal standaardrapportages te (laten) ontwikkelen.

V. Aanbieden opslag- en reken capaciteit

Eén van de conclusies is dat de afhankelijkheid van Amerikaanse partijen voor onderwijsinstellingen groot is. Om een alternatief te bieden, kan SURF een hosting- en rekendienst aanbieden specifiek voor Learning Analytics. Het zou gaan om invulling van de componenten analysesysteem (inclusief opslag), learning records store en visualisatiesysteem. Een mogelijkheid is dat SURF hiertoe haar SURFSara diensten inzet of contracten afsluit met marktpartijen.

Gebaseerd op gesprekken met instellingen

VI. Menukaart en inkoopstelsel

Zoals aangetoond met dit onderzoek zijn er diverse oplossingen in de markt met elk hun eigen voordelen, nadelen en risico's. Het helpt instellingen wanneer SURF overzicht biedt van oplossingen (inclusief eigenschappen) a.d.h.v. de Learning Analytics architectuur. Aanvullend kan SURF hier een inkoopstelsel aan koppelen waarmee instellingen gemakkelijk licenties kunnen aanschaffen. Zusterorganisatie van SURF in het VK JISC biedt een dergelijk inkoopstelsel aan waarin diverse marktpartijen Learning Analytics software en (consultancy)dienstverlening aanbieden. Een instelling kan een vraag stellen op het online inkoopplatform van JISC a.d.h.v. een programma van eisen. JISC handelt vervolgens de mini-competitie en het inkoopproces af¹⁵.

¹⁴ Zie bijvoorbeeld de Unified Data Definition standaard ontwikkeld door JISC:

<https://github.com/jiscdev/analytics-udd>

¹⁵ Voor een globale procesbeschrijving zie: <https://docs.analytics.alpha.jisc.ac.uk/docs/learning-analytics/Learning-Analytics-Purchasing-Service>

VII. Kennisdeling

SURF doet al veel aan kennisdeling rondom Learning Analytics. Het is aan te bevelen dat SURF hiermee door gaat. SURF zou dit uit kunnen breiden met het inrichten van een community waarin instellingen bewezen praktijken rondom Learning Analytics en ontwikkelde algoritmes met elkaar kunnen delen. Instellingen vinden vaak het wiel nog opnieuw uit. Deze community helpt om dit te voorkomen en stimuleert innovatie.

VIII. Inzet van consultants en projectmanagers

SURF kan met de inzet van consultants en projectmanagers gedegen toepassing van Learning Analytics stimuleren en ondersteunen bij de implementatie. Hiertoe kan SURF bijv. een marktplaats ontwikkelen waar geschikt bevonden consultants en projectmanagers van marktpartijen of ZZP'ers zich aan verbinden. Een instelling kan vervolgens via de marktplaats een aanvraag doen of een consultant/projectmanager direct benaderen.

A Bronnen

Interviews

Geïnterviewde	Organisatie	Datum
Iris Huis in 't Veld	SURF	15-10-2019
Herman van Dompsele	SURF	15-10-2019
Karianne Vermaas	SURF	15-10-2019
Marieke de Wit – van de Vecht	SURF	15-10-2019
Sebas Veeke	SURF	15-10-2019
Justian Knobbout	Hogeschool Utrecht	15-10-2019
Frank Benneker	Universiteit van Amsterdam	17-10-2019
Chris van Klaveren	Vrije Universiteit	17-10-2019
Ilja Cornelisz	Vrije Universiteit	17-10-2019
Theo Bakker	Vrije Universiteit	18-10-2019
Bram Enning	Hogeschool Leiden	19-12-2019
Kees de Brouwer	Hogeschool Leiden	19-12-2019
Alexander Dortland	MBO Rijnland	07-01-2020
Jan-Paul van Staalduinen	TU Delft	08-01-2020
Hubert Vogten	Open Universiteit	09-01-2020
Pieter Volleberg	Open Universiteit	09-01-2020
Wilco te Winkel	Erasmus Universiteit	20-01-2020
Alex Lodder	Universiteit Utrecht	23-01-2020
Lex Freund	Hogeschool Rotterdam	11-02-2020
E-mail-/telefooncontact met medewerkers van leveranciers	Divers	December 2019/ januari 2020

Workshopdeelnemers

Geïnterviewde	Organisatie	Datum
Alan Berg	Universiteit van Amsterdam	28-10-2019
Erik Bolhuis	Hogeschool Windesheim	28-10-2019
Tom Koning	Breda Hogeschool	28-10-2019
Frank Benneker	Universiteit van Amsterdam	28-10-2019
Jan Tjeerd Groenewoud	Rijksuniversiteit Groningen	28-10-2019
José Knoth	Tilburg University	28-10-2019
Eline van Hoven	ISO	28-10-2019
Roos van Leeuwen	LSVB	28-10-2019

Documenten

Omschrijving
Gartner (2016), Three important ways the analytics landscape is changing in higher education
Gartner (2016), Assessing the state of Learning Analytics in higher education
Gartner (2016), Getting started with Learning Analytics
Gartner (2016), Higher education analytics for data-driven insight
Gartner (2016), How to make the business case for Learning Analytics
Hogeschool van Amsterdam (2016), Learning Analytics
Kennisnet, Learning Analytics whitepaper
Sclater, N., Berg, A. & Webb, M. (2015), Developing an open architecture for Learning Analytics, EUNIS Journal of Higher Education, 2015/3.
SURF, LA-01: pijlers Learning Analytics platform
SURF (2018), Versnellingsplan onderwijsinnovatie met ICT, SURF
SURF, Richtlijnen Learning Analytics 0.1
SURF (2019), Ethisch kader – opzet
SURF (2018), LA experiment en AVG
SURF (2019), Hoe data de kwaliteit van het hoger onderwijs kunnen verbeteren
SURF (2019), Folder studiedata
SURF, Learning Analytics, SURFConext & de Open Onderwijs API. Een technologie verkenning
SURF, Notitie “Een flexibele en persoonlijke leeromgeving”
University of Seoul (2016), A Reference Model for Learning Analytics
Vogten, H., & Koper, R. (2018), Towards Big Data in Education: the case at the Open University of the Netherlands, in M. Spector, V. Kumar, A. Essa, Y-M. Huang, R. Koper, R. Tortorella, T-W. Chang, Y. Li, ... Z. Zhang (Eds.), Frontiers of Cyberlearning: Emerging Technologies for Teaching and Learning (pp. 125-143).
Volkskrant (2019), https://www.volkskrant.nl/columns-opinie/digitalisering-bedreigt-onze-universiteit-het-is-tijd-om-een-grens-te-trekken~bff87dc9/
World Economic Forum, https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/
<i>Websites van de onderzochte leveranciers/oplossingen</i>

B Richtlijnen infrastructuur

	Richtlijnen Privacy	Categorie	Vragen	Toepassing
1	a. Rechten van betrokkenen	Minimaal	Leverancier werkt mee om verzoeken van betrokkenen die voortvloeien uit de AVG en andere relevante wet- en regelgeving uit te kunnen voeren	alle
2	a. Rechten van betrokkenen	Wenselijk	Betrokkenen kunnen direct bij de leverancier op eenvoudige wijze verzoeken van betrokkenen die voortvloeien uit de AVG en andere relevante wet- en regelgeving indienen en de leverancier draagt er zorg voor dat de verzoeken tijdig bij verwerkingsverantwoordelijke terecht komen om beoordeeld te kunnen worden.	alle
3	a. Rechten van betrokkenen	Minimaal	Persoonsgegevens kunnen geanonimiseerd worden als er aan een verzoek tot vergetelheid gehoor moet worden gegeven	alle
4	a. Rechten van betrokkenen	Wenselijk	Gegevens kunnen direct verwijderd worden als er aan een verzoek tot vergetelheid gehoor moet worden gegeven	alle
5	b. Locatie verwerking	Minimaal	Persoonsgegevens worden enkel buiten de EER verwerkt indien aan een van de wettelijke bepalingen uit de AVG (hoofdstuk V) is voldaan	alle
6	b. Locatie verwerking	Wenselijk	Persoonsgegevens worden enkel verwerkt binnen de EER	alle
7	c. Transparantie	Minimaal	Leverancier kan uitleggen op basis van welke data welke keuzes, beoordelingen of resultaten worden gemaakt	B en F
8	c. Transparantie	Wenselijk	Leverancier kan alle verwerkingsroutes aantonen en gebruikte algoritmes volledig transparant uitleggen	B en F
9	d. Informatiebeveiliging	Minimaal	Leverancier heeft een procedure voor het op gezette tijdstippen testen, beoordelen en evalueren van de doeltreffendheid van de technische en organisatorische maatregelen ter beveiliging van de verwerking en ter bescherming van persoonsgegevens en werkt in ieder geval eens per jaar mee aan een onafhankelijke audit	alle
10	d. Informatiebeveiliging	Wenselijk	Leverancier is gecertificeerd voor informatiebeveiliging tegen de ISO27001	alle
Open vragen privacy				
			Vragen	
11	a. Rechten van betrokkenen		Welke functionaliteit biedt de leverancier om procesmatige aspecten uit de AVG te borgen (rechten van betrokkenen, transparantie)	alle
12	a. Rechten van betrokkenen		In hoeverre heeft de betrokkene regie over wat er met zijn/haar gegevens gebeurt?	alle
13	e. Privacy by design		In hoeverre wordt de verwerking van persoonsgegevens geminimaliseerd (geselecteerd, uitgesloten, verwijderd en vernietigd)?	alle
14	e. Privacy by design		Weke functionaliteit biedt de leverancier voor het handhaven van bewaartermijnen?	alle
15	e. Privacy by design		In hoeverre wordt de verwerking van persoonsgegevens gescheiden van andere datasets, danwel bijzondere persoonsgegevens gescheiden van reguliere persoonsgegevens?	alle
16	e. Privacy by design		In hoeverre wordt het detail waarin persoonsgegevens worden verwerkt beperkt?	alle

17	e. Privacy by design	In hoeverre worden persoonsgegevens beschermd, onherleidbaar of onobserveerbaar gemaakt?	alle
18	e. Privacy by design	Welke mogelijkheden voor automatische anonimisering/ pseudonimisering van gegevens biedt de leverancier?	alle
19	d. Informatiebeveiliging	Hoe is / kan autorisatie en logging (worden) ingericht?	alle
Richtlijnen non-functional		Vragen (Indien van toepassing)	
20	f. Toegankelijkheid	Het LA-systeem is toegankelijk voor mensen met een beperking	B, C en F
21	g. Open Source	Alle gebruikte code binnen het Learning Analyticssysteem is open source beschikbaar en herbruikbaar voor iedereen die dit wil	B, C en F
22	h. Metadata	Data en resultaten worden opgeleverd met bijbehorende metadata voor begripsvorming en traceerbaarheid (definities & herkomst)	B, C en F
23	i. Authenticatie	Applicaties zijn in staat gebruikers te authenticeren op basis van hedendaagse standaarden	alle
24	j. Gebruiksvriendelijkheid	Rapportages kunnen op alle typen devices (laptop, tablet en mobiele telefoon) geopend en weergegeven worden	C en F
Open vragen non-functional		Vragen (Indien van toepassing)	
25	k. Evidence based	Welke (wetenschappelijke) evidence ligt ten grondslag aan de gestandaardiseerde elementen/ algoritmen van de LA oplossing?	F
26	k. Evidence based	Is er een proces ingericht om op basis van voortschrijdende inzichten de gestandaardiseerde elementen/ algoritmen up te daten / te herzien?	F
27	l. Ethiek	Welke waarborgen heeft de leverancier om te voorkomen dat gebruikte algoritmes bevooroordeeld, discriminerend of willekeurig zijn?	B, C en F
Richtlijnen Functionaliteit		Vragen (Indien van toepassing)	
28	m. Compatibiliteit bronsystemen	Het LA-systeem ondersteunt het overnemen van gegevens uit de ondersteunende systemen die in de HORA als volgt genoemd worden: - Student Informatie Systeem - Educatieve applicaties - Digitaal Portfolio Systeem - Digitaal Toets Systeem - Stage Systeem	A en E
29	n. Data-opslag	Het LA-systeem beschikt over ETL-functionaliteit om data op te schonen en te transformeren	A en E
30	n. Data-opslag	Het LA-systeem is in staat data gestructureerd op te slaan	A en E
31	n. Data-opslag	Het LA-systeem is in staat om ruwe data aan te bieden ter analyse	A en E
32	o. Gegevensuitwisseling	De oplossing haalt informatie op uit de bronsystemen door middel van API's	A, D, E en F
33	o. Gegevensuitwisseling	De door de oplossing gehanteerde API's zijn opgebouwd op basis van open standaarden	A, D, E en F
34	p. Data-analyse	Het LA-systeem is in staat gestructureerde data te analyseren	B en C

35	p. Data-analyse	Het LA-systeem is in staat om ongestructureerde data te analyseren en om te zetten naar gestructureerde data door middel van Natural Language Processing (NLP)	B
36	p. Data-analyse	Het systeem biedt mogelijkheden voor:	B en C
		- Statistische analyses	
		- Onderwerp analyses	
		- Netwerk analyses	
		- Sociale analyses	
		- Patroon analyses	
		- Dynamische modellen	
		- Associatie analyses	
37	p. Data-analyse	Het LA-systeem biedt een data-analyse interface waarmee gebruikers de gevraagde analyses kunnen configureren	B en C
38	p. Data-analyse	Analyses kunnen door middel van blokken en voorgedefinieerde acties geconfigureerd worden	B en C
39	p. Data-analyse	Voor expert gebruikers is een code interface beschikbaar waarin analyses door middel van één of meerdere programmeertalen gedefinieerd kunnen worden	B en C
40	q. Data-visualisaties	Dashboards kunnen gepersonaliseerd worden	C
41	q. Data-visualisaties	Datavisualisaties kunnen gerealiseerd worden door middel van drag & drop zonder daarbij te hoeven programmeren	C
42	q. Data-visualisaties	Datavisualisaties kunnen gerealiseerd worden door middel van programmeren met veelgebruikte talen zoals R of Python	C
43	q. Data-visualisaties	Uitkomsten van visualisaties kunnen in 'derde'-applicaties gebruikt worden	C
44	q. Data-visualisaties	Uitkomsten van visualisaties kunnen in 'derde'-applicaties ingesloten worden door middel van code-snippets (te vergelijken met het inbedden van een Youtube filmpje op een website/PPT)	C
	Open vragen functionaliteit	Vragen (Indien van toepassing)	
45	r. Datakwaliteit	In hoeverre biedt de leverancier tooling of mogelijkheden/controls voor de verhoging van datakwaliteit?	Alle