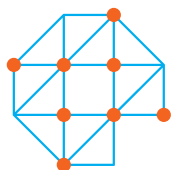


Whitepaper realtime studiedata



Versnellingsplan Onderwijsinnovatie met ICT

 veilig en betrouwbaar
benutten van studiedata



Whitepaper realtime studiedata

Versnellingsplan Onderwijsinnovatie met ICT
Zone Studiedata
www.versnellingsplan.nl



Versnellingsplan
Onderwijsinnovatie
met ICT

Auteur
Roland Ettema (informatiemanager Open Universiteit)

Augustus 2022



Op deze uitgave is de Creative Commons Naamsvermelding 4.0-licentie van toepassing. Maak bij gebruik van dit werk vermelding van de volgende referentie: Ettema, R. (2022). Whitepaper realtime studiedata. Utrecht: Versnellingsplan Onderwijsinnovatie met ICT.

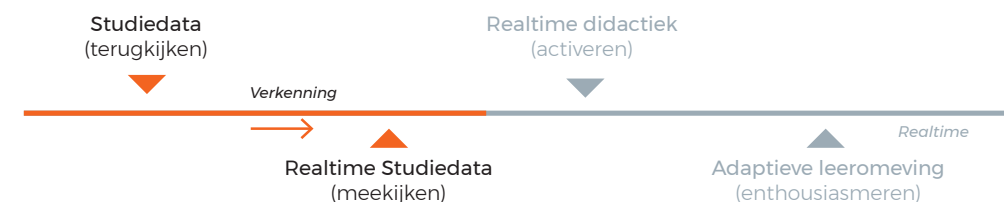
Inhoudsopgave

1 Inleiding	5
2 Waarom realtime studiedata?	9
2.1 Achtergronden	9
2.2 Belanghebbenden	11
2.3 Visie en keuzes	12
3 Randvoorwaarden	15
3.1 Privacyrichtlijnen: AVG	15
3.2 Ethisch kader	17
3.3 Betrokkenheid studenten en docenten	18
3.4 Eisen aan de infrastructuur	19
4 Technologische inrichting	23
4.1 Realtime dataverwerking	23
4.2 Dataopslag	24
4.3 Monitoring	26
5 Resultaten en inzichten	29
6 Samenvatting	33
7 Meer informatie	34

1 Inleiding, probleemstelling en doel

Over studiedata en het belang ervan voor het monitoren en verbeteren van studiegedrag is al veel geschreven. Met de opkomst van online modules en lesmateriaal komen er niet alleen steeds meer studiedata beschikbaar, het wordt bovendien mogelijk om die data realtime te raadplegen. Dat geeft docenten de mogelijkheid om direct en *just-in-time* feedback te geven op basis van waarnemingen in het studiegedrag van studenten. De COVID-19-pandemie en de bijbehorende lockdowns hebben het belang van realtime studiedata alleen maar vergroot, omdat veel opleidingen zijn overgeschakeld op (volledig) online onderwijs.

Voor veel onderwijsinstelling was de stap naar online onderwijs een noodmaatregel als reactie op de pandemie. Nu de ervaringen met online onderwijs toenemen, neemt ook de aandacht voor digitale didactiek toe. Bij toepassing van online onderwijs en digitale didactiek blijkt de wens van de student om de docent in nabijheid te ervaren een universeel probleem. Het is op dit punt waar het werken met realtime studiedata een verschil kan maken. Op basis van realtime studiedata kun je de digitale omgeving weer realtime laten reageren op het leerproces van de studerende terwijl het plaatsvindt. "Een op de juiste wijze ingerichte adaptieve leeromgeving kan enthousiasmeren, activeren en verkleint de afstand met de docent", zo is de gedachte bij de Open Universiteit (OU).



Figuur 1 Visie en routekaart op realtime studie data

Voor een zinvolle toepassing van realtime studiedata (waarbij de docent met de student kan meekijken) moet rekening worden gehouden met een groot aantal factoren. Allereerst is het van belang om goed na te denken over welke data men wil verzamelen en voor welk doel, met andere woorden: welke signaleringen mogelijk zijn op basis van welke data. De belangen en de visie van de verschillende belanghebbenden, vooral die van studenten en docenten, moeten daarbij goed worden afgewogen. Daarnaast impliceert werken met realtime data de beschikbaarheid van een toegankelijke ICT-infrastructuur, waarbij keuzes gemaakt moeten worden over onder meer dataopslag en architectuur.



Figuur 2 Gehanteerde begrippen in real time studiedata

In deze whitepaper gaan we in op de mogelijkheden, randvoorwaarden en eisen op het gebied van realtime studiedata. Centraal hierbij staat een recent project bij de Open Universiteit, gericht op de ontwikkeling van een realtime infrastructuur. Binnen dit project zijn drie pilots uitgevoerd die zich richten op realtime signalering van:

1. Meeliftgedrag door studenten;
2. Uitvalrisico van studenten in de eerste weken na start cursus;
3. Non-start studenten.

In dit project staan de opslag, technologie en architectuur centraal. In deze whitepaper is informatie hierover aangevuld met een aantal interviews, die meer ingaan op de context en het belang van realtime studiedata. We hebben vier (duo-)interviews¹ gehouden, waarin achtereenvolgens beleid en visie, (technische) uitvoering en het docentperspectief aan de orde zijn geweest.

¹ Interviews zijn gehouden met Nadine Roijackers (Stuurgroepvoorzitter, Directeur Expertisecentrum Onderwijs); Jan Jansen, (Stuurgroep lid, Directeur IT); Roland Ettema (Stuurgroep lid, Manager Informatie Expertisecentrum); Wilfred Rubens (Projectleider ECO, docent); Hubert Vogten (Senior Ontwikkelaar Expertisecentrum Onderwijs) en Martine Coun (docent OU).

Online leren zonder gebruik te maken van realtime studiedata is als het wegsturen van de astronauten in Apollo 11 zonder ondersteuning vanaf de aarde. De astronauten hadden waarschijnlijk de reis naar de ruimte wel kunnen maken, maar zonder realtime feedback vanuit ground control zou het lastig zijn geweest om koers te houden op de gewenste bestemming (de maan).



Afbeelding: NASA

De inhoud van deze whitepaper is als volgt. In hoofdstuk 2 gaan we in op het belang en de achtergrond van realtime studiedata. Hoofdstuk 3 behandelt de randvoorwaarden voor projecten gericht op het verzamelen en bewerken van studiedata, bijvoorbeeld op het gebied van privacy en technische infrastructuur. Vervolgens zoomen we in hoofdstuk 4 in op de technische infrastructuur die bij de OU is gehanteerd, met name op de dataopslag en de monitoring tool. Tot slot bespreken we in hoofdstuk 5 een aantal 'lessons learned'; dit onderdeel is in het bijzonder relevant voor andere instellingen die met (realtime) studiedata aan de slag willen gaan.

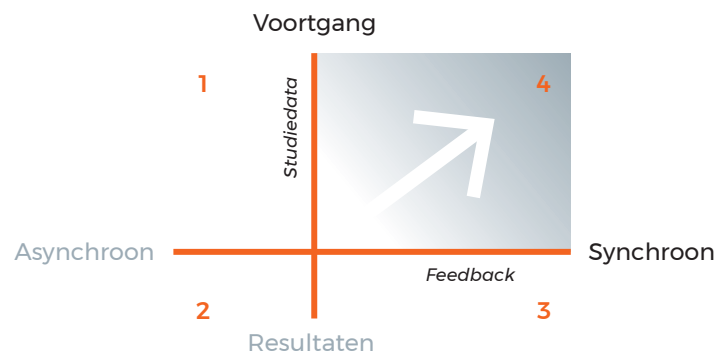
2 Waarom realtime studiedata?

2.1 Achtergronden

Online onderwijs is inmiddels niet meer weg te denken uit het Nederlandse onderwijsstelsel. De verschuiving van regulier, klassikaal onderwijs naar online of blended varianten was al geruime tijd aan de gang, maar is ook nog eens versterkt door de komst van het coronavirus (COVID-19) en de beperkende maatregelen die daaruit voortvloeiden. Bij online leren hebben docenten minder zicht op de voortgang van studenten omdat ze hen minder zien. Maar ze kunnen wél beschikken over veel data over wat studenten leren en hoe ze dat doen.

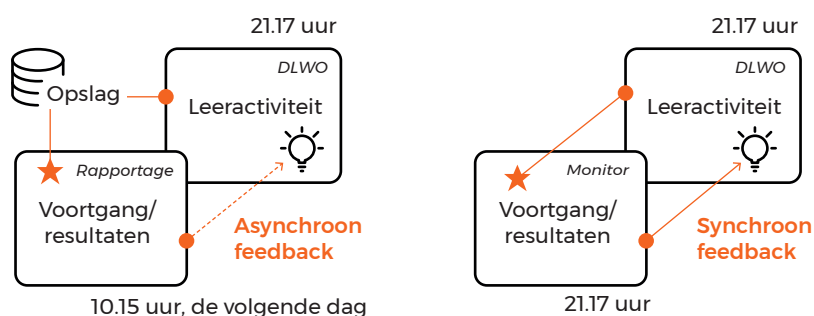
We kunnen studiedata op verschillende manieren indelen, maar een veel gemaakte indeling is die tussen studiedata die betrekking hebben op resultaten en studiedata die betrekking hebben op voortgang. Bij data die betrekking hebben op studieresultaten gaat het vaak om ondersteunende of secundaire processen, bijvoorbeeld data uit een Student Informatiesysteem (SIS) voor het vastleggen van inschrijvingen en studieresultaten. Bij data die betrekking hebben op de voortgang van studenten gaat het over het primaire proces: hieronder vallen bijvoorbeeld data uit Leermanagementsystemen (LMS), feedbacksystemen of toetsystemen.

We kunnen studiedata ook indelen op basis van het moment waarop ze gebruikt worden om feedback te geven. Bij asynchrone feedback analyseert een docent de studiedata achteraf en kiest hij zelf het moment waarop hij feedback geeft. De docent gebruikt dus zijn eigen expertise en ervaring om het feedback-moment te kiezen. De tegenhanger is synchrone feedback. Synchrone feedback is feedback die *tijdens* de studieactiviteit wordt gegeven, zodat een docent direct kan reageren op een leeractiviteit van een student. Bij asynchrone feedback wordt gebruik gemaakt van studiedata over plaatsgebonden leeractiviteiten, voor synchrone feedback is realtime informatie van belang. Een kanttekening daarbij is dat er bij menselijk ingrijpen altijd sprake is van enige vertraging tussen het gedrag van de student en de feedback van de docent; synchrone feedback is dus alleen mogelijk als deze feedback automatisch wordt gegenereerd.



Figuur 3 Realtime feedback-mogelijkheden o.b.v. realtime studiedata

In de pilots bij de OU zijn de beide perspectieven gecombineerd. Onderzoeken in kwadrant 1 en 2 (zie figuur 3, hierboven) richten zich bijvoorbeeld op studiedata die aangeven dat een student inefficiënt of langzaam studeert, dan wel uitstelgedrag vertoont (zie ook figuur 4, linkerhelft). In kwadrant 3 en 4 gaat het erom welke signalering over dit soort verschijnselen realtime mogelijk zijn. Daarbij komen veel verdiepende vragen naar boven, zoals: op basis van welke data ontstaat dat signaal dan? Hoe verzamel je deze data? En hoe transformeer je deze data in naar een (leesbaar) signaal op een monitor (zie figuur 4, rechts). Al deze vragen moeten ons inzicht geven in realtime dataverwerking en signalering *terwijl de student online aan het studeren is*.



Figuur 4 Verschil in synchrone en asynchrone feedback

Wilfred Rubens (projectleider bij de OU) over signalering:

“In dit project hebben wij meegekeken naar het inlevergedrag van studenten. In welke mate leveren ze hun opdrachten in en doen ze dat ook op tijd? Maken ze zelftesten? Wanneer loggen ze in, op welke onderdelen van een cursus klikken ze? We hebben in deze pilot veel meer realtime studiedata gemonitord dan je normaal gesproken doet in een digitale leeromgeving. De resultaten kunnen via de monitor realtime bekeken worden.”

Door actief gebruik te maken van realtime studiedata kunnen instellingen zien hoe leerprocessen zich bij studenten voltrekken. Door het identificeren van patronen in studiedata kan de digitale didactiek in een cursus worden verbeterd. De uitkomsten van dergelijke analyses vormen een goede basis om verbeteringen door te voeren in het ontwerp van de cursus in de digitale leer en werkomgeving (DLWO).

2.2 Belanghebbenden

Bij studiedata kunnen verschillende groepen stakeholders onderscheiden worden. De belangrijkste leveranciers van studiedata zijn de studenten, die studiedata genereren door hun activiteiten (bijvoorbeeld bij het inschrijven, bij het maken van oefeningen, bij het deelnemen aan virtuele bijeenkomsten enzovoort). Verder zijn ook de docenten leveranciers van studiedata door hun feedback, die ook wordt geregistreerd.

De belangrijkste gebruikers van studiedata zijn docenten, studieadviseurs en opleidingsdirecteuren. Deze laatste groep kan studiedata gebruiken om meer inzicht te krijgen in de kwaliteit van het aangeboden onderwijs, die deels afgelezen kan worden uit de resultaten en voortgang die studenten maken. Daarnaast vormen onderzoekers naar onderwijs(kwaliteit) een belangrijke doelgroep. Andere potentiële gebruikers zijn ondersteuners in het onderwijs, beleidsmakers en tot slot bestuurders.

De OU heeft bij de pilots die centraal staan in deze whitepaper goed gekeken naar manieren om de betrokkenheid van de belangrijkste partijen te borgen. Studenten moeten zich bewust zijn van de data die over hen worden verzameld en moeten er dan ook mee akkoord gaan dat de OU die data gebruikt om ‘mee te kijken’. Voor docenten zijn de belangrijkste vragen op welke manier ze kunnen meekijken en welke signaleringen realtime data mogelijk maken. Ook andere groepen betrokkenen hebben hun eigen belangen en prioriteiten kenbaar kunnen maken in de pilots.

Bij de OU hebben de docenten (ook buiten de pilots die in deze whitepaper centraal staan) de optie om te werken met studiedata. “We maken gebruik van een digitale leeromgeving, en daarin zit ook een voortgangsmonitor”, zegt Wilfred Rubens, docent en projectleider bij de pilot. “Daarop kun je bijvoorbeeld zien wanneer studenten voor het laatst zijn ingelogd en hoeveel taken ze hebben uitgevoerd. Dat is vrij beperkt en daar zitten ook wel een aantal tekortkomingen aan. Het systeem is niet zo gebruiksvriendelijk. Als je bijvoorbeeld veel studenten in een cursus hebt is het lastig om overzicht te houden. Vandaar dat we in onze pilots ook aandacht geven aan de wijze hoe je data presenteert.”

Docenten maken ook steeds meer een afweging tussen de moeite die het kost om met een bepaalde (digitale) tool te werken en het rendement dat het oplevert. Daarom is het belangrijk dat de tools die in de pilots ontwikkeld worden, anticiperen op het ongemak dat de docenten momenteel ervaren. Nu is dat nog niet altijd het geval, vertelt docente Martine Coun (betrokken bij de pilot): “Martine: “Als docent moet je goed weten wat je wilt zien. Je moet zelf actief scrollen en zoeken, terwijl je eigenlijk zou willen dat er automatisch al een aantal zaken aan je wordt gepresenteerd. Je kunt nu bijvoorbeeld niet snel zien welke studenten in de laatste drie weken hun werkstukken hebben afgerond. Dat zou je eigenlijk wel willen. Een andere wens is dat we als docent een bericht krijgen als een cursus niet goed loopt of als een groep studenten achterloopt. Daarmee zou het systeem de docent meer ondersteunen.”

Het huidige systeem maakt het al mogelijk om online feedback te geven, hoewel de mogelijkheden beperkt zijn. “Als studenten langer dan een week niet zijn ingelogd, dan attendeer ik ze erop”, zegt Wilfred Rubens. “Dan moet ik die groep handmatig selecteren, en dan stuur ik ze een berichtje. Ik kijk vooral of de studenten actief bezig zijn met de leerstof. Over het algemeen krijg ik positieve reacties, en soms geven ze ook een goede verklaring waarom ze niet met hun studie bezig waren. Maar dat werkt alleen bij relatief kleine groepen studenten.”

2.3 Visie en keuzes

Voor een instelling die aan de slag wil met studiedata (realtime of niet) is het van belang om na te denken welke onderwijskundige visie daaraan ten grondslag ligt. Dat is in de praktijk nog niet eenvoudig: onderzoekers en docenten laten zich namelijk nog veel leiden door nieuwe technische mogelijkheden. Nieuwe technologieën kunnen onderzoekers en studenten stimuleren om ermee aan de slag te gaan. Maar een visie helpt om te bepalen of het (onderwijskundige) doel in het oog wordt gehouden en of de inzet van technische middelen daartoe bijdraagt.

Om een visie te ontwikkelen, kunnen bestuurders, onderzoekers en docenten gebruik maken van het Discussiemodel Learning Analytics² dat SURF heeft ontwikkeld. In dat model wordt een aantal stappen doorlopen, zoals:

- Wat is het doel waarvoor je learning analytics wilt inzetten?
- Welke concrete vragen wil je met de data beantwoord hebben?
- Wie gaat daarvan welk voordeel ondervinden?
- Voor welke doelgroep wil je learning analytics inzetten?
- Welke kansen en risico's zijn er bij de keuze voor dit doel en deze doelgroep?

Bij de OU is de visie op studiedata nog in ontwikkeling, vertelt Nadine Roijackers, directeur van het Expertisecentrum Onderwijs (ECO): “Vanuit het perspectief van ons Expertisecentrum is het verzamelen van studiedata vooral interessant op het vlak van kwaliteitszorg. De kwaliteitszorg is bij ons altijd ingericht op basis van kwalitatieve informatie, zoals enquêtes onder de studenten. Maar het is de vraag of je daarmee altijd de waarheid boven tafel krijgt. Een voorbeeld: wij hanteren bij de OU de richtlijn dat een cursus drie maanden voor de start beschikbaar moet zijn, zodat studenten al kennis kunnen maken met het materiaal. Dan kun je ze achteraf vragen: wat is de waarde daarvan voor jou? Maar je kunt ook kijken in de leeromgeving naar het aantal clicks. Als je dan ziet dat er in de eerste maand helemaal geen clicks zijn geweest, heb je objectieve data die laten zien dat het niet zo veel zin heeft een cursus al zo lang van tevoren beschikbaar te stellen. Dat stelt je in staat om betere beslissingen te nemen. Als je in het onderwijs een vernieuwing doorvoert, kun je het effect daarvan monitoren. We blijven ook kwalitatieve informatie verzamelen, met behulp van surveys, interviews en focusgroepen. Maar we zijn ervan overtuigd dat het beeld over onderwijskwaliteit completer wordt als we deze kwalitatieve gegevens combineren met gedragsdata.”

De OU is tijdens de ontwikkeling van deze whitepaper (najaar 2021) bezig geweest met een grote vernieuwingsslag van alle systemen, zowel de onderwijskundige als de ondersteunende systemen. Die systemen maken het mogelijk uiteenlopende data te verzamelen, stelt Roijackers: “We krijgen data uit het leerplatform, uit het toetsstelsel, uit de inschrijvingen. Dat zijn allemaal realtime data, die in een groot stuwmeer van data worden opgeslagen. Alleen is de vraag wat je daar vervolgens mee gaat doen. Wat is nuttig? Daarover zijn we nu aan het nadenken in het kader van onze kwaliteitscultuur. Daarover moeten we in de komende tijd een visie ontwikkelen. Op dit moment zijn het vooral de onderwijs-

² ‘Discussiemodel: concreet aan de slag met learning analytics’: www.surf.nl/discussiemodel-concreet-aan-de-slag-met-learning-analytics.

wetenschappers die met deze data werken, maar in de pilot hebben de docenten er ook mee gewerkt. Het is wel het idee dat het gebruik van studiedata op korte termijn breder wordt geïmplementeerd.”

Jan Jansen (Directeur ITF bij de OU) over data huishouding:

“In onze aanpak voor vernieuwing van onze kernsystemen is een datastrategie opgenomen. Onze datastrategie begint met ethische uitgangspunten over waarom de OU data wil verzamelen. Verder kijken we naar de wettelijke kaders en welke ruimte die geven om uiteindelijk te bepalen wat we gaan verzamelen. Eén ding is ons duidelijk geworden: data-governance en -management is niet iets wat je er even bij doet. Daarom heeft de OU besloten om een Informatie Expertise Centrum (IEC) op te richten die deze handschoen moet oppakken. Inmiddels is het IEC een vast organisatieonderdeel geworden dat verantwoordelijk is voor onze instellingsbrede datastrategie en data-ontsluiting naar het onderwijs- en onderzoeksveld.”

3 Randvoorwaarden

3.1 Privacyrichtlijnen: AVG

Een belangrijke randvoorwaarde is dat de privacy van studenten niet in het geding mag komen bij het verzamelen van realtime studiedata. Bij het verzamelen en analyseren van studiedata worden in veel gevallen persoonsgegevens verwerkt. Als een onderwijsinstelling persoonsgegevens verwerkt, is daarop de Algemene Verordening Gegevensbescherming (AVG) van toepassing. Deze verordening stelt eisen aan de grondslagen voor het gebruik van studiedata, vraagt om een afweging van de risico's voor de studenten en om passende maatregelen. Een eenmalige toepassing en een privacyverklaring op de website zijn niet voldoende. SURF heeft een handreiking³ met een stappenplan gemaakt die instellingen kan helpen met het invullen van de privacy-verplichtingen die voortvloeien uit de AVG.

Bij het verzamelen van studentgegevens is het van belang duidelijkheid te verkrijgen op basis van welke grondslag dit kan gebeuren. Binnen onderwijsinstellingen bestaan verschillende opvattingen over het belang van learning analytics om begeleiding te geven en studievertraging en uitval te voorkomen. In principe zijn er twee mogelijkheden: het verzamelen en analyseren en gebruiken van studiedata kan vallen onder een 'gerechtvaardigd belang', of aan de studenten moet expliciet toestemming worden gevraagd.

Bij de Open Universiteit wordt een mengvorm gehanteerd. Uitgangspunt is dat het verzamelen en analyseren van realtime studiedata op zich valt onder het gerechtvaardigd belang; het draagt bij aan meer kennis over het onderwijsproces en maakt signaleringen mogelijk die in het belang zijn van de studenten. Maar op dit moment is voor het analyseren van studiedata bij de Open Universiteit toestemming nodig.

Voor de pilot is gebruik gemaakt van een optie in het LMS die de studenten in staat stelt aan te geven of ze mee willen werken aan onderzoek. Alleen van de studenten die deze optie hebben aangevinkt, zijn data verzameld.

³ 'Learning analytics in 5 stappen: een handreiking voor de AVG': www.surf.nl/learning-analytics-in-5-stappen-een-handreiking-voor-de-avg.

Wilfred Rubens, projectleider:

“We moeten studenten om toestemming vragen en de studenten die niet willen, doen niet mee. Maar er is wel een discussie over gerechtvaardigd belang: het belang van studiedata voor het geven van begeleiding is groot. Dat zou een reden zijn om misschien geen toestemming te hoeven vragen, maar daarover lopen de meningen uiteen. De OU staat nu nog op het standpunt dat we de studenten toestemming moeten vragen. Je kunt ook zeggen: dit zijn de voorwaarden die erbij horen.”

Nadine Roijackers, Directeur ECO:

“Je kunt er natuurlijk altijd bezwaar tegen maken dat je data gebruikt worden, dat ligt besloten in de privacywetgeving. Maar wij gebruiken de data vooral om de dienstverlening en het onderwijs te verbeteren, dus het is wel in het belang van de studenten dat we die kunnen gebruiken. We gaan de data niet verkopen of ermee naar buiten treden. Dus het zou wel een optie zijn om dat bij inschrijving af te dekken: als je hier komt studeren, maken we gebruik van een aantal studiedata. En we kunnen studenten aangeven: als we gebruik kunnen maken van studiedata, krijg je een eigen toolkit ter beschikking waarmee je je leerproces kunt bijsturen, zodat het rendement hoger wordt. Dat is alleen maar positief voor de studenten.”

3.2 Ethisch kader

Wettelijke regelingen zoals de AVG stellen grenzen aan en bieden ruimte voor de toepassing van studiedata. Maar in de visie van de OU bieden deze wettelijke kaders nog onvoldoende zekerheden. In december 2019 spraken de rectoren van verschillende universiteiten in een ingezonden brief in de Volkskrant hun zorgen uit over de groeiende machtspositie van tech-bedrijven, en het belang van publieke waarden bij de inzet van studiedata. Begin november deed de Autoriteit Persoonsgegevens in een brief⁴ aan de Vaste Kamercommissie OCW nog een aantal aanbevelingen om persoonsgegevens van studenten beter te beschermen. Het ontgaat de OU dan ook niet dat er al een tijd zorgen zijn over een verantwoorde inzet van studiedata in het hoger onderwijs. We zien dat onze compliance aan de AVG een aantal zorgen niet wegneemt.

⁴ ‘AP waarschuwt voor privacyrisico’s in het onderwijs’: autoriteitpersoonsgegevens.nl/nl/nieuws/ap-waarschuwt-voor-privacyrisicos-het-onderwijs.

Parallel aan de uitvoering van de pilots nam de OU deel aan het deelproject *Referentiekader privacy en ethiek voor studiedata*⁵ in de zone *Veilig en betrouwbaar studiedata benutten* binnen het nationale programma *Versnellingsplan Onderwijsinnovatie met ICT*. Door in de sector gezamenlijk helderheid te scheppen over ethische uitgangspunten en het waarborgen van privacy dragen we publiekelijk uit hoe de sector verantwoord om wil gaan met studiedata. Het referentiekader is een dynamisch document dat continu wordt bijgesteld op grond van maatschappelijke, wetenschappelijke en juridische veranderingen. De resultaten die in de pilot zijn behaald, hebben daarbij als waardevolle cases gediend.

Roland Ettema, Manager IEC:

“Bij de ontwikkeling van het referentiekader studiedata bleek al snel dat de wettelijke kaders geen garantie bieden om ‘onethisch’ gebruik van data te voorkomen. Tot dusver zijn er vier ethische richtlijnen waaraan de sector zich wil houden: 1) Instellingen zijn aanspreekbaar, transparant en leggen rekenschap af; 2) Instellingen wegen alle belangen van betrokkenen en belanghebbenden eerlijk af; 3) Instellingen zorgen voor zorgvuldigheid in hun analyses; en 4) Er is altijd plek voor de menselijke maat, ook wanneer de instelling gebruik maakt voor automatische processen.”

3.3 Betrokkenheid studenten en docenten

Het verzamelen, analyseren en opslaan van studiedata is uiteindelijk bedoeld om docenten in staat te stellen hun studenten beter te begeleiden en meer inzicht te krijgen in efficiency en rendement van het onderwijs. Maar het belang van realtime studiedata is voor veel docenten en studenten niet vanzelfsprekend. Voor docenten kan het lijken alsof studiedata een omslachtige manier zijn om inzicht te krijgen in het studiegedrag van hun studenten. Studenten kunnen bezwaar hebben tegen het idee dat ze gemonitord worden of dat hun online gedrag wordt vergeleken met dat van andere studenten.

Het is dus van belang dat studenten en docenten nauw betrokken worden bij elk initiatief dat gericht is op realtime studiedata. Wij beogen om nut en de waarde van real time signaleringen onder hun aandacht te brengen om vervolgens samen te kijken naar de balans tussen de toegevoegde waarde en de investeringen die van hen gevraagd worden. “Je hebt altijd voorlopers en mensen die wat traditioneler zijn ingesteld”, zegt Nadine Roijackers.

⁵ ‘Over het ‘Referentiekader privacy en ethiek voor studiedata’: doe-meer-met-studiedata.nl/article/referentiekader-privacy-en-ethiek-voor-studiedata.

“Bij onderwijswetenschappen is realtime studiedata al een belangrijk thema, en in de pilot hebben we wel veel belangstelling gezien, ook vanuit andere faculteiten. Je moet beginnen met de groep die er de waarde van inziet en die er positief tegenover staat. Hieruit kunnen weer inspirerende voorbeelden ontstaan die anderen weer kunnen inspireren”

Bij de pilot bij de OU zijn verschillende docenten betrokken geweest en hun feedback is meegenomen in de evaluatie. “Tijdens de pilot hadden we de beschikking over een monitor met een aantal didactische views”, zegt Wilfred Rubens. “Daarmee konden we veel sneller zien welke studenten betrokkenheid vertonen en welke studenten risico lopen op uitval. Bovendien werden docenten en ook de studiecoördinator automatisch geattendeerd op studenten die zich wel hadden ingeschreven, maar bijvoorbeeld twee weken niets hadden gedaan. Het bleek dat maar weinig studenten dat gedrag vertoonden, dus het nut was beperkt.”

De betrokken docenten hebben verschillende suggesties gedaan voor verbeteringen die in toekomstige projecten doorgevoerd zouden kunnen worden. Notificaties scoren daarbij hoog, aldus Wilfred Rubens: “We wilden eigenlijk dat de studenten zelf automatisch een bericht zouden krijgen als ze bijvoorbeeld te lang niet ingelogd waren. Maar het beleid bij de OU is dat geautomatiseerde feedback (nog) niet is toegestaan. Dus die optie konden we niet gebruiken.” Ook Martine Coun is een voorstander van automatische berichten: “Als je notificaties kunt laten verzenden naar docenten en studenten toe, kan de technologie je al redelijk helpen in het monitoren van een aantal zaken. Dan is technologie ondersteunend aan het onderwijs.”

Martine Coun, docent OU, over signalering:

“Ik vergelijk het een beetje met bol.com en booking.com. Stel dat je daar bezig bent met een bestelling en je hebt je zoektocht niet afgemaakt, dan krijg je een berichtje als: ‘Je hebt nog artikelen in je winkelmandje’. Dat zou je eigenlijk ook binnen de leeromgeving willen, als studenten bijvoorbeeld zijn bezig geweest met een opdracht, maar ze hebben die niet afgemaakt. Dat moet natuurlijk niet al te dwingend zijn, de toonzetting is heel belangrijk. De computer moet je niets gaan voorschrijven, maar een suggestie kan misschien wel.”

De docenten kunnen desgevraagd ook een aantal randvoorwaarden noemen waaraan toekomstige oplossingen voor realtime studiedata zouden moeten voldoen:

- **Laagdrempeligheid:** het moet niet te veel moeite kosten om met een applicatie te werken. De toegankelijkheid en gebruiksvriendelijkheid van een applicatie staan op de eerste

plaats. Het is geen bezwaar als de studiedata via een aparte applicatie beschikbaar worden gesteld, maar die moet bij voorkeur wel zijn geïntegreerd in de digitale leer- en werkomgeving. Het is niet handig als je daarvoor apart moet inloggen.

- **Overzichtelijkheid:** je moet snel patronen kunnen herkennen, je moet in één oogopslag kunnen zien of er iets aan de hand is. Verder is het handig als je snel kunt wisselen tussen verschillende cursussen. Voor een studieadviseur of coördinator kan het handig zijn over cursussen heen te kunnen kijken, om te zien hoe studenten scoren.
- **Relevantie:** het moet het gaan over belangrijke onderwijsvraagstukken, zoals betrokkenheid, risico op uitval en studiesucces.
- **Proactief:** het helpt als je als docent actief gewezen wordt op eventuele problemen, bijvoorbeeld door middel van notificaties.

Nadine Roijackers, Directeur ECO:

“Waar we naartoe gaan is steeds meer autonomie voor docenten en studenten in de co-creatie van de leerervaring. We willen de docenten in staat stellen om het onderwijs zelf aan te passen, zonder dat ze afhankelijk zijn van bijvoorbeeld onderwijstechnologen. In dat kader zie ik de toepassing van studiedata. Het gaat dan om de vraag: hoe stel je docenten (en ook studenten) in staat om zo goed mogelijk aan de knoppen van het leerproces te draaien? We kunnen bijvoorbeeld samen met de faculteit Onderwijswetenschappen kijken: als we dit op een monitor zien, wat zou dan goede feedback zijn? En hoe zou je die feedback willen vormgeven? Daar zou je dan een aantal mogelijkheden kunnen aanbieden die een docent gemakkelijk zelf kan inzetten.”

3.4 Eisen aan de infrastructuur

De hoeveelheid beschikbare data neemt exponentieel toe. Denk bijvoorbeeld aan de groei van activerende online elementen in een DLWO (Digitale Leerweg Omgeving). Daarbij gebruiken docenten en studenten niet één leermanagementsysteem, maar een samenhangend geheel van met elkaar geïntegreerde applicaties. Voorbeelden zijn een peer feedback-toepassing, een serious game, discussieforums en virtual classroom-bijeenkomsten. De diversiteit in online activerende elementen neemt een steeds grotere vlucht, mede dankzij alle nieuwe technologische mogelijkheden. Instellingen genereren op die wijze steeds grotere hoeveelheden aan gestructureerde en ongestructureerde data. De OU ziet dat voor de analyse en het gebruik van data afscheid wordt genomen van de traditionele datamanagement-benaderingen, die veelal uitgaan van gestructureerde dataopslag.

Bovendien komen er in de nabije toekomst nieuwe en slimmere diensten bij, de zogenoemde 'smart services'. Denk aan de mogelijkheid om schrijfvaardigheid te volgen en studenten tips te geven over hun vordering in het academisch schrijven. Zulke smart services baseren zich op data en genereren zelf ook weer data. De infrastructuur moet het mogelijk maken om de werking van deze smart services transparant te maken. Immers, we willen rekenschap over inzet van middelen kunnen afleggen aan derden. Via al deze data ontstaat een 'holistisch beeld' van de student. Bij de OU heeft deze ontwikkelingen eisen gesteld aan integratie en opslag die weer zijn vertaald in een 'data lake'-aanpak waaraan verschillende analysetools gekoppeld kunnen worden.

Een belangrijke eis in de pilots bij de OU was de integratie van data om deze in onderlinge samenhang realtime te kunnen verwerken en analyseren. Denk aan data die afkomstig zijn uit de DLWO, de toetsapplicaties of het SIS. Veel data worden nu vaak nog niet met elkaar in verband gebracht, laat staan dat dat realtime gebeurt. Daarnaast stelde de OU de eis dat de data worden opgeslagen en beschikbaar zijn om er offline analyses op te kunnen uitvoeren.

Een blauwdruk voor het realtime verwerken van data is de zogenoemde 'Lambda-architectuur' geïntroduceerd door Nathan Marz. Deze IT-architectuur-blauwdruk beschrijft een samenhangend geheel aan logische IT-componenten met als functie om grote hoeveelheden data te verwerken door gebruik te maken van zowel batch- als streaming-technieken. Die technieken zijn nodig omdat we uitgaan van een groot aantal studenten in een leeromgeving die heel veel leeractiviteiten uitvoeren die data opleveren. De Lambda-architectuur combineert batch processing en real-time processing, waardoor een actueel beeld van de data ontstaat.

De Lambda-architectuur-blauwdruk is op vele manieren te realiseren. Bij het maken van een keuze hebben we aanvullende eisen gesteld op het gebied van schaalbaarheid, beschikbaarheid, betrouwbaarheid en flexibiliteit:

1. Bij **schaalbaarheid** gaat het erom in hoeverre een systeem kan omgaan met actuele toenames van load of vraag. Daarbij is het met name van belang hoeveel data binnen een bepaalde tijd verwerkt kan worden en hoe snel de gevraagde data geleverd wordt (de zogeheten latentie).
2. De data moet altijd **beschikbaar** zijn, anders kunnen smart services niet direct reageren. De architectuur moet zo worden ontworpen dat hardware-problemen geen invloed hebben op de beschikbaarheid.
3. Een **betrouwbare** infrastructuur zorgt ervoor dat data correct en up-to-date zijn. Een gebruiker moet bijvoorbeeld wijzigingen kunnen aanbrengen zonder dat data verloren gaan of incorrect worden. Als er fouten worden geconstateerd, moeten er signaleringen beschikbaar zijn om adequaat op te treden.

4. De infrastructuur moet **flexibel** kunnen omgaan met veranderingen, bijvoorbeeld als de hoeveelheid geproduceerde data toeneemt. De infrastructuur moet ook kunnen omgaan met veranderingen in het type of de structuur van de data, zonder dat de beschikbaarheid en de betrouwbaarheid in het geding komen.

Uiteindelijk is binnen de pilots gekozen voor de Kafka-architectuur, die de Lambda-architectuur overbodig maakt. De Kafka-aanpak wordt ook wel Kappa-architectuur genoemd. In het volgende hoofdstuk gaan we hier nader op in.

4 Technologische inrichting

Ooit ging het binnen de OU om een paar bestanden, naderhand werd met interfaces gewerkt, de webservices maakten een intrede en de servicebus kwam er. Echter, met realtime-toepassingen komen er nieuwe eisen op het gebied van (streaming) data. Oude inrichtingen kunnen dat niet meer aan, dat kan je niet hebben in het onderwijs waar het operationele proces zo sterk data afhankelijk is. Niet alleen de omvang van de data is toegenomen, maar ook het belang ervan.

4.1 Realtime dataverwerking

Bij de Open Universiteit is ervoor gekozen om de Lambda-architectuur te implementeren met Apache Kafka⁶ (open source). De Kafka-architectuur omvat een platform waarin een gedistribueerd berichtensysteem, een realtime verwerker van datastromen en gedistribueerde dataopslag zijn opgenomen. Het voordeel van Apache Kafka is dat het minder complex is dan andere implementaties van de Lambda-architectuur. Bij de Kafka- (of Kappa-) architectuur komt het verschil tussen batch- en realtime-verwerking te vervallen. Hierdoor zijn er minder technologische componenten, wat de complexiteit reduceert. Verder voorziet de Apache Kafka-architectuur in load balancing, korte latentietijden, hoge doorvoer en grote fouttolerantie. Dat maakt het mogelijk om omvangrijke sets aan studiedata te verwerken.

Met Apache Kafka kan voor dataverwerking gebruik worden gemaakt van twee paden, het koude en het dynamische pad:

- Via het koude pad worden alle binnenkomende gegevens opgeslagen in onbewerkte vorm. Deze gegevens worden in batches verwerkt. Het resultaat van deze verwerking wordt opgeslagen als een batchweergave.
- Via het dynamische pad kunnen inkomende data in realtime worden geanalyseerd. Deze laag is ontworpen voor een lage latentie. Dit gaat dan ten koste van de nauwkeurigheid.

De gegevens vanuit beide paden kunnen zichtbaar worden gemaakt op een monitor. Deze monitor kan gegevens vrijwel realtime weergeven (waarbij ze iets minder nauwkeurig mogen zijn) door een koppeling met het dynamische pad. In aanvulling daarop kunnen

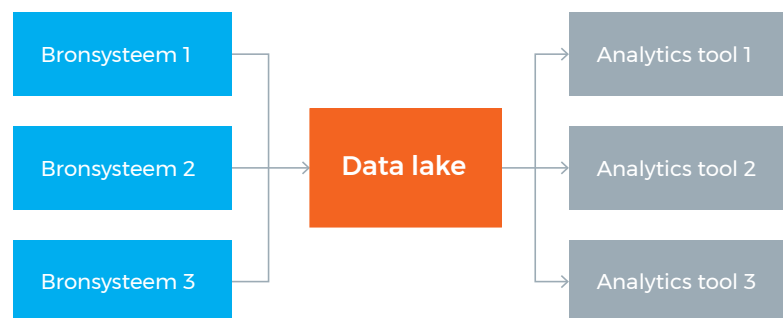
⁶ Zie kafka.apache.org.

we andere data op de monitor weergeven, door gebruik te maken van het koude pad (dat nauwkeuriger is, maar minder actueel). Met andere woorden: het dynamische pad bevat gegevens voor een relatief klein tijdvenster, waarna de resultaten kunnen worden bijgewerkt met nauwkeurigere gegevens uit het koude pad.

4.2 Dataopslag

Achter batch- of realtime dataverwerking is een oplossing voor dataopslag noodzakelijk. Apache Kafka zorgt voor een permante opslag van data, waardoor een zogenoemd data lake ontstaat. Een data lake is een opslagplaats van gegevens die zijn opgeslagen in hun oorspronkelijke format. Het gaat daarbij om kopieën van gegevens uit diverse bronsystemen (zoals een SIS, een LMS en andere applicaties). Een data lake kan verschillende soorten gegevens bevatten (gestructureerd, semigestructureerd en ongestructureerd). Deze gegevens moeten wel beperkt in omvang zijn.

Aan het data lake kunnen verschillende applicaties worden gekoppeld om analyses uit te voeren, patronen zichtbaar te maken, managementinformatie te genereren of onderzoek te doen:



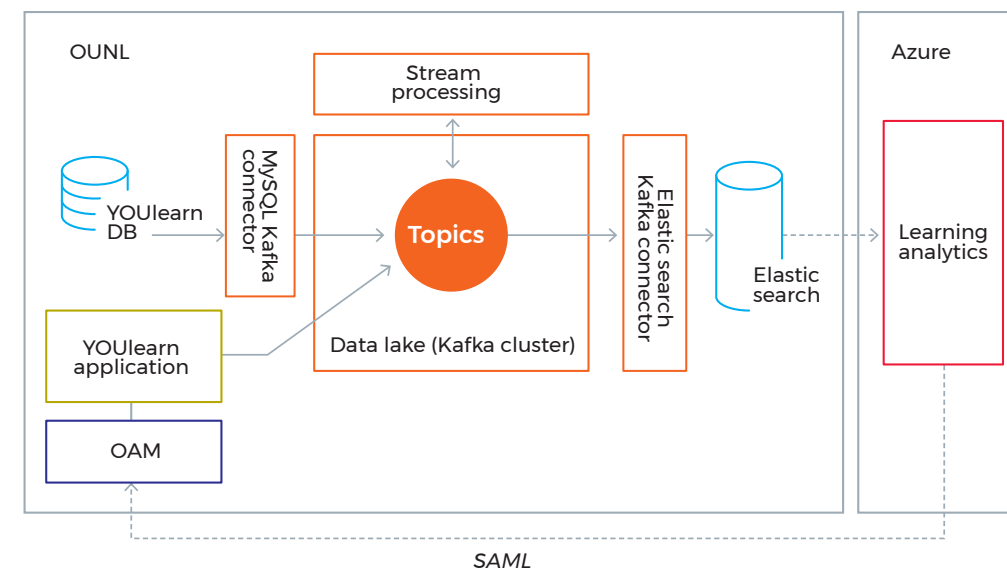
Figuur 5 Schematische opzet van een data lake

De keuze voor een data lake is voor een groot deel ingegeven door de beperkingen van traditionele technologieën. Daarbij gaat het om relationele databases, al dan niet in combinatie met applicaties en webservers. Het probleem hierbij is vooral dat deze technologieën maar beperkt kunnen opschalen. Opschalen kan in de praktijk leiden tot performanceproblemen van de infrastructuur en tot verlies van data.

Hubert Vogten, Senior Ontwikkelaar ECO:

“Het data lake maakt het mogelijk om data uit verschillende bronnen live te omvatten. Daarbij gaat het om verschillende soorten data, bijvoorbeeld een database event of een change event. Ieder record wordt opgeslagen in een persistent topic. Het verschil met een database is dat een database vaak maar één waarde van een record vastlegt. Het data lake is dynamischer en kan de volledige historie vastleggen, waarmee je bijvoorbeeld kunt volgen hoe een student door een cursus is heengegaan.”

Tijdens een pilot zijn een aantal applicaties gekoppeld aan dit data lake; in eerste instantie ging het daarbij om de database van de digitale leeromgeving (zie onderstaande afbeelding). Daarbinnen kwamen gegevens beschikbaar zoals persoonsgegevens van studenten, inloggedrag, aangeklikte items in de cursussen, ingeleverde opdrachten en data van gemaakte zelftoetsen. Deze gegevens werden realtime gestreamd naar het data lake. De gegevens zijn beschikbaar gemaakt via een monitor.



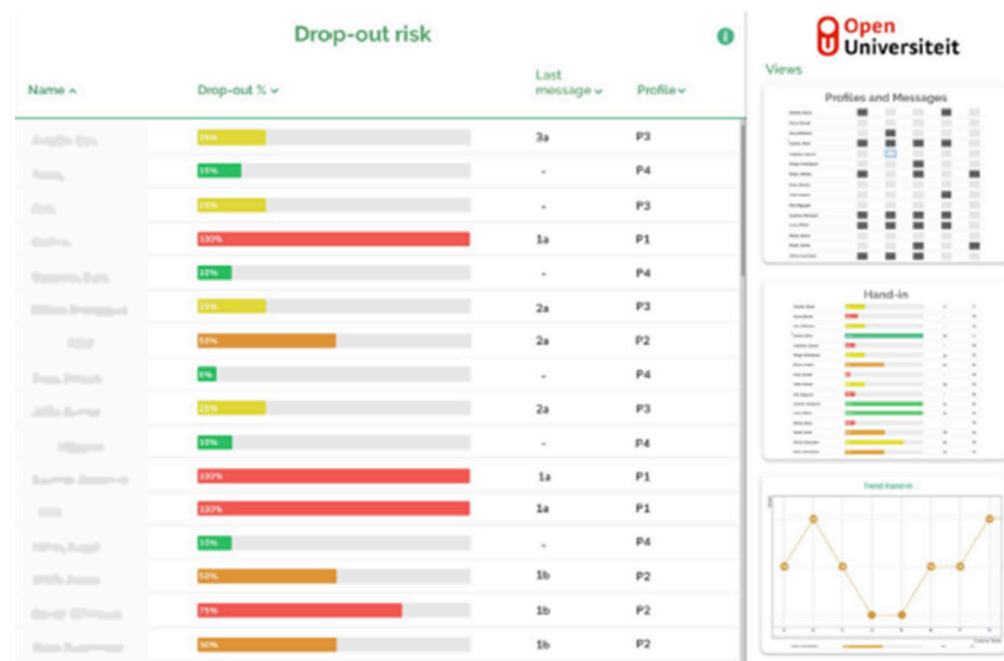
Figuur 6 Data lake met verbonden applicaties

4.3 Monitoring

In de pilots is gebruik gemaakt van een monitor van een extern bedrijf. Deze monitor is met een beveiligde VPN-verbinding gekoppeld aan het data lake. Via deze koppeling kan de monitoring tool realtime data zoeken binnen het data lake. De OU heeft geconfigureerd welke data gestreamd mogen worden naar de monitoring tool. Dat is vastgelegd in een Verwerkersovereenkomst.

Tutoren en docenten kunnen op basis van algoritmes en dankzij het combineren van data binnen de monitoring tool zicht krijgen op een aantal onderwijskundige thema's: betrokkenheid van studenten, risico op voortijdige uitval, meeliftgedrag bij samenwerkend leren en inlevergedrag. Hiervoor zijn verschillende 'views' ontwikkeld, zoals lijsten of trendlijnen. Tutoren kunnen studenten ook met elkaar vergelijken of vergelijken met het groeps-gemiddelde. Deze informatie kan worden gebruikt voor de begeleiding van studenten.

Hieronder is een voorbeeld opgenomen van onze monitor op uitvalrisico:



Figuur 7 Voorbeeld van een monitor

Dit scherm geeft een indicatie van het uitvalrisico, waarbij de kleur en de lengte van de balk de hoogte van het risico aangeeft. Het risicopercentage wordt berekend op basis van betrokkenheid van de studenten in de eerste drie weken van een cursus. Die betrokkenheid is een functie van login data, uitgevoerde taken en ingeleverde opdrachten.

De drie kleine schermen aan de rechterkant geven de andere beschikbare views aan. Zo geeft de bovenste optie, Profiles and Messages, inzicht in de verschillende student-profielen binnen deze view: 'non starters', 'late staters', 'assignment 1 overdue' en 'assignment 1 on time'.

Tot slot hebben ook studenten tijdens de pilot beschikking gekregen over een monitor. Daarmee konden zij hun eigen voortgang vergelijken met de gemiddelde voortgang van de groep.

Wilfred Rubens, projectleider:

"De views zijn bepaald in samenspraak met een aantal onderwijsonderzoekers. We hebben na afloop van de pilot gevraagd wat de docenten nog meer zouden willen. Daaruit blijkt dat ze graag genotificeerd willen worden, ze willen bijvoorbeeld een mailtje ontvangen als een bepaald percentage studenten het risico loopt om uit te vallen. Dat is dan voor hen een trigger om de monitor te raadplegen. We weten bijvoorbeeld dat als studenten zich inschrijven voor een cursus maar niet gelijk beginnen, de kans op uitval groter is. De studietoelbehoefte kan dan gericht actie ondernemen. In de pilot is in verband met de AVG nog geen gebruik gemaakt van automatische notificaties."

5 Resultaten en inzichten

Tijdens de pilot bij de OU is veel ervaring opgedaan met het realtime verzamelen, verwerken, opslaan en beschikbaar maken van studiedata. Deze ervaringen kunnen nuttig zijn voor instellingen die zelf aan de slag willen met realtime studiedata. Wij geven belangstellende instellingen de volgende tips, do's en dont's mee op het gebied van achtereenvolgens de organisatie, de privacy, de techniek en de didactiek.

Organisatie, visie en beleid

- Een instelling moet een duidelijke visie hebben op het verzamelen, analyseren en gebruiken van studiedata plus een bijpassend beleid. Zonder visie en beleid is de kans groot dat er technologisch van alles geprobeerd wordt zonder dat het waarde toevoegt in het onderwijs.
- De ontwikkeling en implementatie van een oplossing voor realtime studiedata is een omvangrijk project dat veel vraagt van de organisatie. Bij de OU is dit project door een aantal medewerkers uitgevoerd naast de lopende werkzaamheden, wat soms veel vergde van de betrokkenen. Beter is om in een vroeg stadium een realistische inschatting te maken van de benodigde capaciteit en middelen. Als er te weinig in zo'n project wordt geïnvesteerd, behaal je niet het maximale rendement.
- Het gaat bij realtime studiedata en de techniek daaromheen veelal om nieuwe kennis. Dat betekent dat de betrokken medewerkers in het begin een hoge leercurve doormaken. Daar is voldoende support voor nodig. "Zo'n project vraagt veel expertise, die moet je opbouwen", zegt Wilfred Rubens. "Dat moet dus ook een projectdoel zijn: dat je al doende de nodige expertise ontwikkelt. Maar voor de eindgebruikers kan dat wel frustrerend zijn."
- Voor het realiseren van een oplossing voor het analyseren, verzamelen en opslaan van studiedata is expertise nodig, in huis of via inhuur. Bij de OU ging het om expertise om een data lake op te zetten, in te richten en te koppelen met systemen en applicaties aan de input- en outputkant.
- Begin klein vanuit een visie, maar blijf ook weer niet te lang in de planfase steken. Door te experimenteren en in de praktijk aan de slag te gaan, leer je heel veel.

Ethiek, privacy en AVG

- Bepaal in een vroeg stadium wat de consequenties van een project voor de AVG (kunnen) zijn. "Wij hadden al een voorziening in de digitale leeromgeving, waarmee studenten kunnen aangeven of ze wel of niet willen meedoen aan een onderzoek", zegt Wilfred Rubens. "We hebben ervoor gezorgd dat we in onze monitoring tool alleen de data krijgen van studenten die hiervoor toestemming hebben gegeven. Dat is ook doorgesproken met de security officer. Dat is wel een laagdrempelige manier om AVG-conform te werken."

- Als de implicaties van een project voor de AVG nog niet duidelijk zijn, kan dat vertraging opleveren. “Heel belangrijk was dat we de richtlijnen van de AVG konden respecteren door de manier waarop we de data hebben verzameld”, zegt Hubert Vogten. “In wezen maak je namelijk een kopie van de data en je gaat die gebruiken voor andere doeleinden. Maar wij hebben een muur gebouwd tussen de data en het gebruik ervan, daardoor kun je vast beginnen met de infrastructuur en komen de AVG-issues pas bij het gebruik van de data aan de orde.”
- Tijdens de uitvoering was er nog geen ethisch kader, maar met terugwerkende kracht kunnen we vaststellen dat de AVG- en privacyrichtlijnen te beperkt zijn om te beoordelen of je ‘ethisch’ handelt. Het is dus aan te raden om als instelling zelf een ethisch kader te ontwikkelen.

Techniek

- De datahuishouding moet op orde zijn. Daarbij zijn heldere definities belangrijk. Denk aan vragen als: Wat is een student? Wat is een inschrijving? Ook moet er helderheid zijn over de relaties die gelegd worden op basis van studiedata, bijvoorbeeld over de vraag welk online gedrag een indicatie geeft van betrokkenheid. Deze relaties moeten gebaseerd zijn op onderzoek.
- Zorg ervoor dat je alle relevante data kunt gebruiken, ook al zijn die data afkomstig uit verschillende systemen. Beperk je niet tot de monitoring die deel uitmaakt van het LMS.
- Bij connectieproblemen zit de ‘devil in the details’. Check dan vooral de koppelvlakken tussen de bronsystemen en het platform waarop de analysetools zijn geïnstalleerd.
- Je moet het onmiddellijk doorhebben als een server eruit ligt of als er een probleem is met een koppeling, want zonder verbinding kunnen realtime data verloren gaan. Realtime koppelingen vergen een heel ander type beheer.
- Voor het Kafka-cluster gold dat het veel stappen heeft gekost om een goede infrastructuur neer te zetten, aldus Hubert Vogten: “Er zit veel redundantie in het systeem en die moet je op de goede manier inbouwen. Het systeem moet betrouwbaar zijn, je moet voorkomen dat wanneer er één node uitvalt, de data niet meer betrouwbaar zijn.”

Didactiek

- Wilfred Rubens: “Je moet de monitor daadwerkelijk inbedden in de begeleiding. Daar heeft de OU nog wel wat te leren. Dat moet je dan ook bekijken in de bredere context: nu moeten de docenten soms (heel) grote groepen begeleiden. Je kunt ze dan wel een prachtige tool geven, maar dan hebben ze nog niet de ruimte om daar veel mee te doen.”
- Martine Coun: “Je moet voorzichtig zijn met de conclusies die je trekt. We kunnen bijvoorbeeld monitoren of studenten actief zijn in de leeromgeving. Maar in cursussen die bijvoorbeeld zijn gebaseerd op een boek zijn er ook studenten die niet actief zijn maar wel het boek goed bestuderen en dan een 7 of een 8 halen.”

- Martine Coun: “Tot slot moet je je realiseren dat je nooit een volledig beeld zult krijgen, een deel van het studiegedrag blijft altijd buiten beeld.”

Wilfred Rubens: “Een ander punt is dat je heel goed en gedetailleerd moet kijken naar een aantal factoren die van invloed kunnen zijn op bijvoorbeeld de didactische views die we hebben ontwikkeld. Een voorbeeld is de vraag: wanneer start een student met een cursus? Is dat als hij het eerste item aanklikt? Maar er zijn studenten die zich helemaal niet houden aan de manier waarop wij denken dat er gestudeerd moet worden. Die analyse hebben wij tijdens het ontwikkelproces gemaakt, maar die zou je eigenlijk van tevoren moeten doen, dan verspil je minder tijd.”

Tot slot Nadine Roijackers: “Ik zie de verdere ontwikkeling van realtime studiedata als mogelijkheid om terug te gaan naar waar het in het onderwijs om draait: persoonlijke aandacht voor de studenten. Ik denk dat de trend die we nu zien gaat doorzetten, en dat je steeds meer hybride en online onderwijs gaat krijgen. Als je onderwijstechnologie goed gebruikt, kun je docenten in staat stellen meer tijd te besteden aan de persoonlijke begeleiding van studenten. Als je voldoende tools hebt die je ondersteunen en ontlasten, krijg je meer ruimte voor begeleiding. Je kunt het reguliere werk dan automatisch laten uitvoeren.”

6 Samenvatting

Deze whitepaper gaat in op het verzamelen en analyseren van realtime studiedata. De basis ervan wordt gevormd door een pilot die is uitgevoerd bij de Open Universiteit: de uitkomsten van deze pilot zijn aangevuld met inzichten uit vier interviews. Hierdoor ontstaat een totaalbeeld over de vele facetten die een rol spelen bij het werken met (realtime) studiedata. Instellingen die hiermee zelf aan de slag willen, kunnen hun voordeel doen met de inzichten, tips en lessons learned.

In de praktijk werken onderwijsinstellingen met verschillende systemen die studiedata opleveren. Maar deze data leveren vaak onvoldoende inzicht in op onder andere de studievoortgang. Meer inzicht is nodig om beter te kunnen sturen op de kwaliteit van het onderwijs, de studievoortgang te bewaken en studiesucces te vergroten. Verder zien we dat het ontwikkelen van een monitor een nieuw onderwerp is dat een aanvulling vormt op de traditionele manier van het werken met studiedata. Het ontwikkelen van een realtime infrastructuur voor studiedata kan bijdragen om dit "nieuwe" speelveld verder te verkennen.

De Open Universiteit vindt het belangrijk om studiedata, verzameld door verschillende organisatieonderdelen en binnen diverse systemen, in onderlinge samenhang te analyseren en te gebruiken. Daarom heeft de Open Universiteit een data lake ontwikkeld waaraan verschillende analysetools gekoppeld kunnen worden. De database van de digitale leeromgeving yOUlearn is via een zogenoemde MySQL Kafka-connector te koppelen aan het Kafka-cluster. Op deze wijze worden onder meer persoonsgegevens van studenten, inloggedrag, aangeklikte items in de cursusstructuur, ingeleverde opdrachten en data van gemaakte zelftoetsen realtime gestreamd naar het Kafka cluster. Via een VPN verbinding is een monitor tool gekoppeld aan dit cluster. De tool verschaft docenten en tutoeren zicht op diverse onderwijskundige thema's (bijvoorbeeld risico op voortijdige uitval en meeliftgedrag). Deze informatie is gebruikt voor de begeleiding van studenten en onderwijsverbetering.

Tijdens het project zijn de randvoorwaarden in kaart gebracht die vervuld moeten zijn om op een effectieve manier met (realtime) studiedata aan de slag te gaan. Daarnaast zijn ervaringen verzameld van onder meer docenten en IT-specialisten. Die ervaringen hebben betrekking op de onderliggende technologie, de organisatie, de privacy en de onderwijskundige aspecten. Zo blijken docenten behoefte te hebben aan notificaties, zoals berichten die aangeven dat een bepaalde groep studenten een uitvalrisico vertoont. Een belangrijke technische 'lesson learned' is dat alle relevante data gebruikt moeten worden, ook al zijn die afkomstig uit verschillende systemen en hebben ze een verschillend formaat.



7 Meer informatie

Over studiedata en learning analytics is al veel informatie beschikbaar. We geven hieronder een aantal handleidingen en whitepapers die zijn ontwikkeld binnen SURF:

Infrastructuur voor learning analytics en studiedata

www.surf.nl/files/2021-05/surf-aandachtspunten-voor-een-veilige-en-betrouwbare-infrastructuur-voor-learning-analytics-en-studiedata.pdf

Discussiemodel: concreet aan de slag met learning analytics

www.surf.nl/discussiemodel-concreet-aan-de-slag-met-learning-analytics

Learning analytics in 5 stappen: een handreiking voor de AVG

www.surf.nl/learning-analytics-in-5-stappen-een-handreiking-voor-de-avg

Experimenteeromgeving learning analytics

www.surf.nl/onderzoek-naar-infrastructuur-learning-analytics

Rapport onderwijskundig perspectief

www.surf.nl/files/2019-04/rapport-onderwijskundigperspectief.pdf

Instellingen experimenteren met learning analytics

www.surf.nl/files/2019-02/Instellingen-experimenteren-met-learning-analytics.pdf



Het Versnellingsplan Onderwijsinnovatie met ICT is een vierjarig programma van SURF, Vereniging Hogescholen en de VSNU dat inzet op het samenbrengen van initiatieven, kennis en ervaringen en snel en concreet aan de slag gaan met kansen voor het hoger onderwijs. Dit gebeurt in acht verschillende 'zones'. In de zone Studiedata werken 10 instellingen aan de hand van 16 deelprojecten aan het veilig en betrouwbaar benutten van studiedata in hoger onderwijs.



Meer informatie en onze publicaties vind je op
www.versnellingsplan.nl