



Datadinsdag 13 sept '22

Ervaringen uit de datacoalities

Doorpakken op digitalisering

Welkom!

Agenda

- Opening door Robert Leeftink
- Keynote Big Data Innovation Hub door Raymond Hoogendoorn
- Verhalen uit de coalities
- Discussietafels
- Afsluiting en borrel

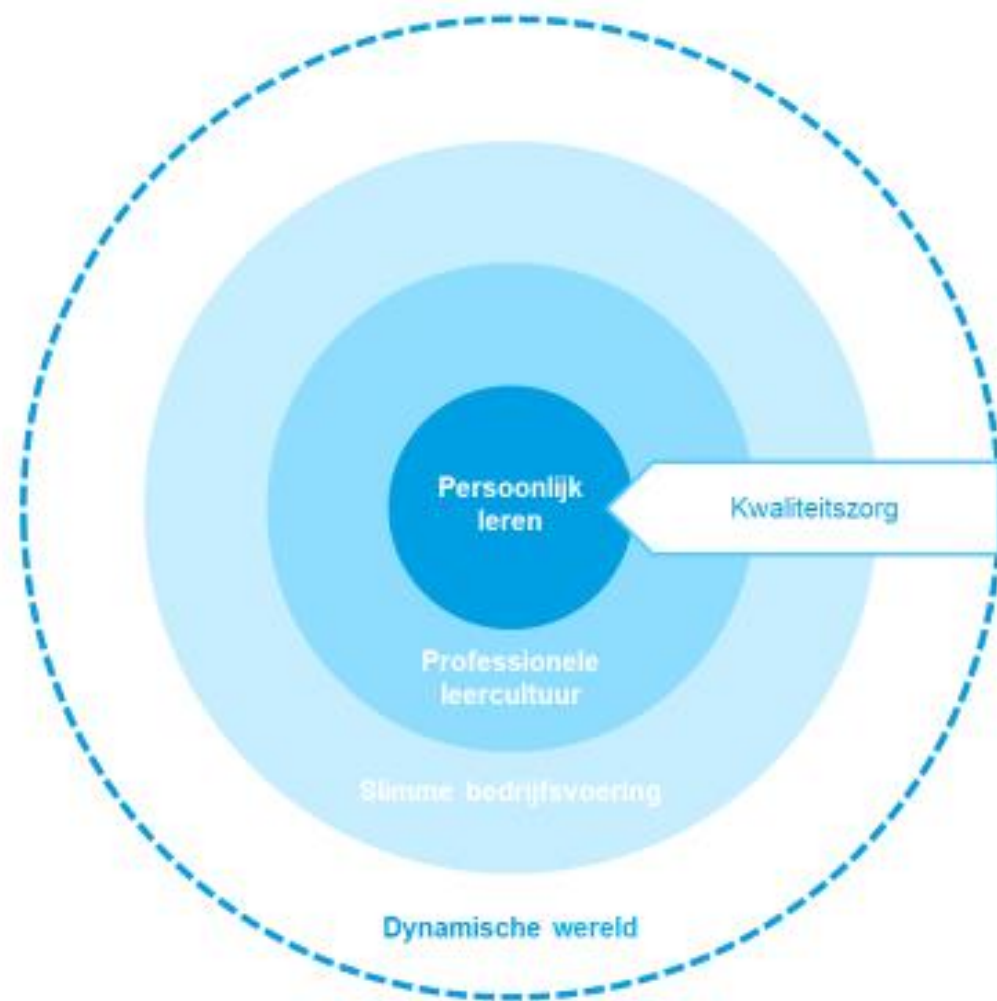




OPENING

Robert Leeftink

Bestuurder MBO Utrecht





BIG DATA INNOVATION HUB

Raymond Hoogendoorn

Professor of Applied Data Science and Logistics



Versie: 13 december 2021

BIG DATA INNOVATION HUB

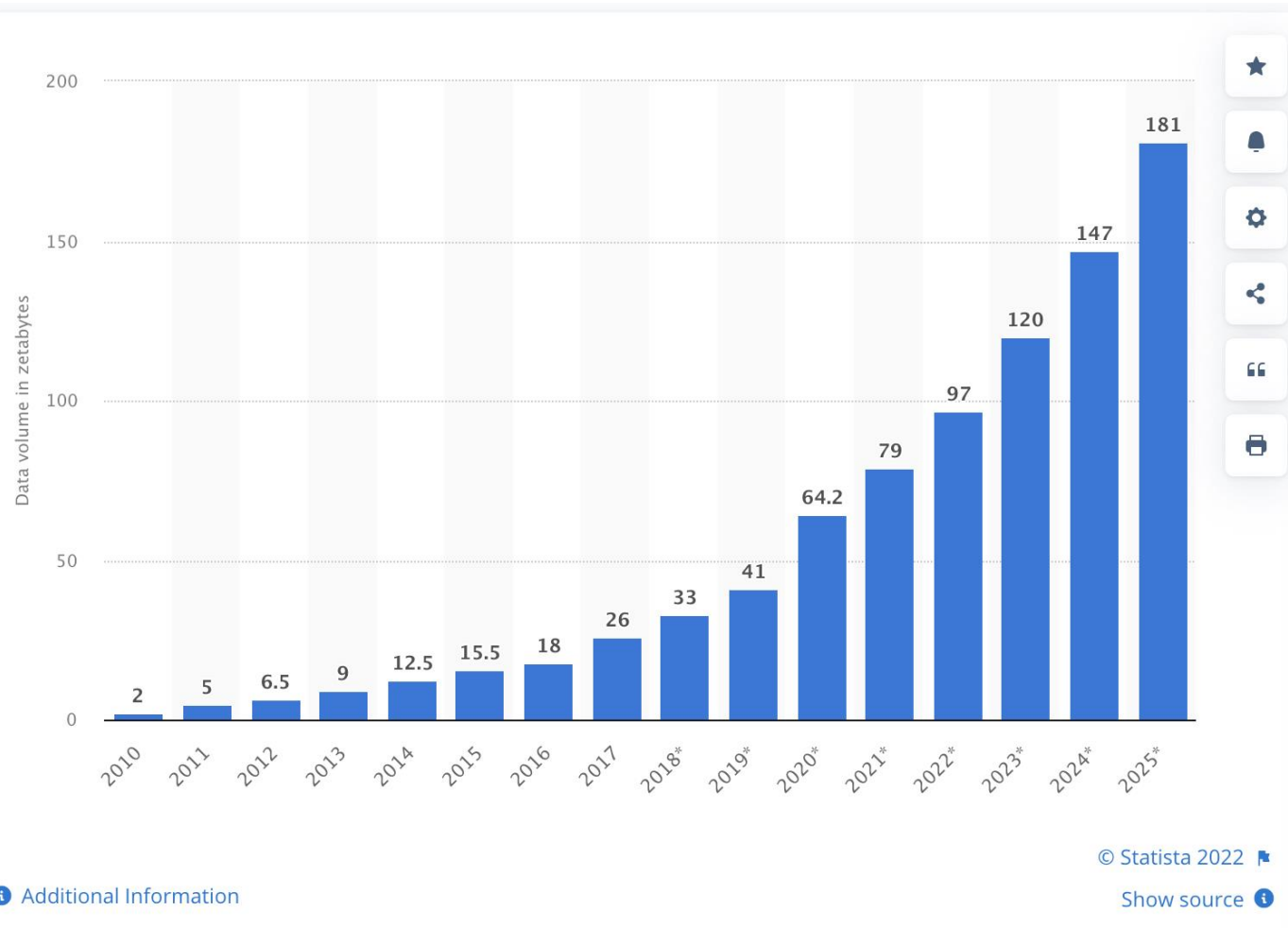
DE HAAGSE
HOGESCHOOL



Introduction

- Dr. Raymond Hoogendoorn
- Professor (Lector) Applied AI and Logistics @ RUAS
- Senior Lecturer Data Science @ THUAS
- Board Member Big Data Innovation Hub (BDIH)
- MSc Psychology
- PhD Civil Engineering
- Ass. Professor @ DUT
- Head of Data Science @ Transavia Airlines
- Strategic Data Consultant @ Ordina

Increase of data



Data scientist

- Attain insight and knowledge from data
- Perform prescriptive analytics
- Increasing demand

SPOTLIGHT ON BIG DATA

Spotlight

ARTWORK Tamar Cohen, Andrew J Buboltz
2011, silk screen on a page from a high school
yearbook, 8.5" x 12"

Data Scientist: *The Sexiest Job of the 21st Century*

**Meet the people who
can coax treasure out of
messy, unstructured data.**

by Thomas H. Davenport
and D.J. Patil

70 Harvard Business Review October 2012





The Big Data Innovation Hub (BDIH)

We focus on **applied science** aimed at current societal and organizational issues in the field of for example finance, healthcare, smart living, smart mobility, etc.

The BDIH is a platform in which data research can be performed. It is a platform where **students, researchers, lecturers and professionals** from for example small and medium-sized businesses and (local) governments can work together on data projects.



Three objectives

- **state-of-the-art education**
- **high-quality research output**
- **external projects**

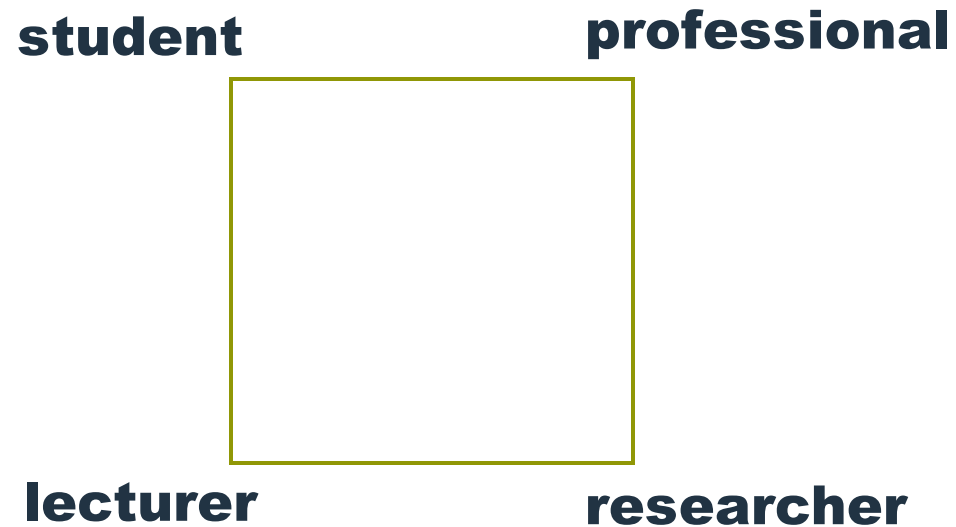


Three pillars

- grounded in practice
- co-creation
- build upon earlier acquired knowledge

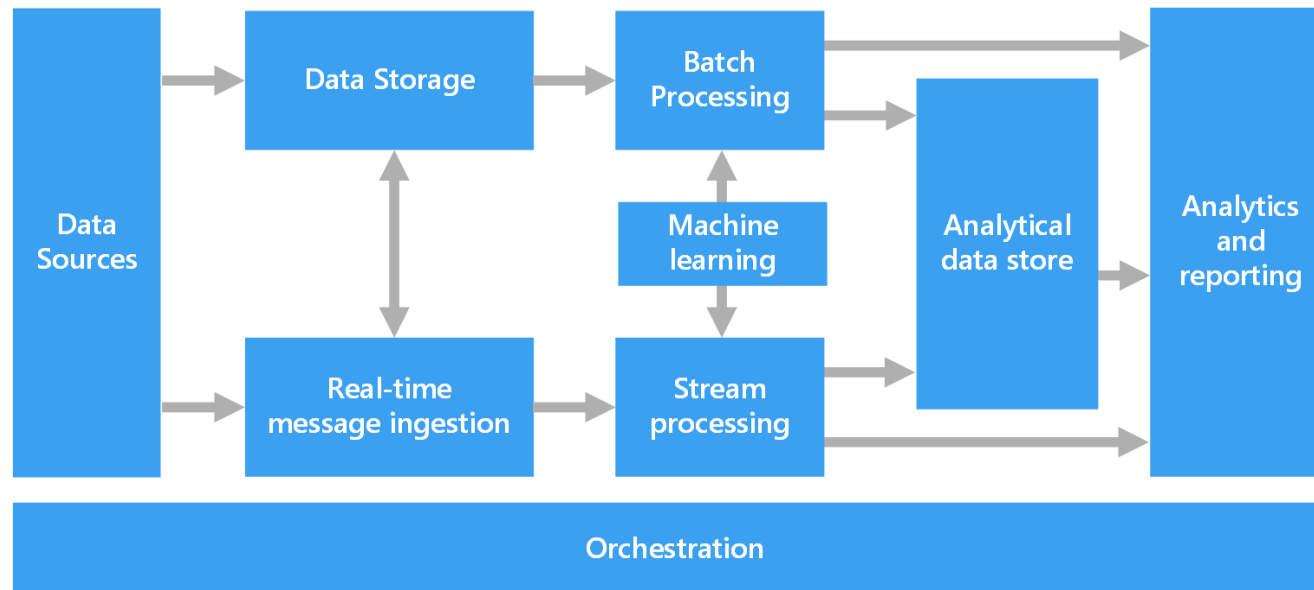
Education & projects

- Practice-based data projects
- focus on **repeatability** and **transferability**.



BDIH Infrastructure

- Cloud-based (Azure)
- Governance
- Data Lineage

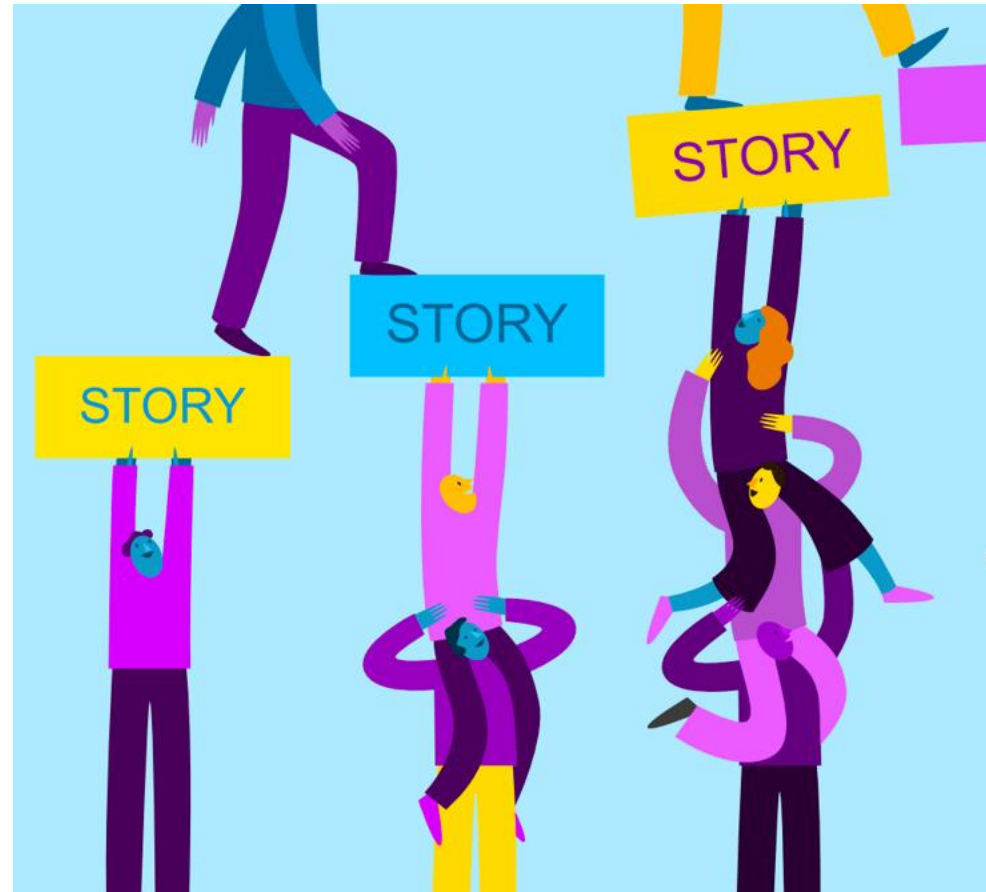


Questions



Verhalen uit de coalities

- Willem Jan Swiebel - WG2 DGO
- José Mulder- Studievoortgang
- Roel Klaver - WG3 DGO
- Leo de Wit - G4
- Tom Oltshoorn & Frank van Dijk - WG1 DGO
- Ronald Ferket - Student journey



14 mbo-instellingen onderzoeken samen met OCW, DUO en MBO Raad hoe data kan worden ingezet om de studieloopbaan van studenten te verbeteren en wat de toegevoegde waarde is om dit met een datagedreven aanpak te doen.

14 mbo-instellingen:



Noorderpoort

roc van
twente



deltion
college

ROC NIJMEGEN

ONDERWIJSGROEP
TILBURG

ZADKINE

albeda

KONING
WILLEM I
COLLEGE

curio

gildeopleidingen

Alfa -college

DUO:

OCW:

MBO Raad:

Nieuwe deelnemer:

Aventus ^A



Dienst Uitvoering Onderwijs
Ministerie van Onderwijs, Cultuur en
Wetenschap



Ministerie van Onderwijs, Cultuur en
Wetenschap



MBO Raad

Projectorganisatie:

hutspot

WG2: verbeteren databeschikbaarheid

Doel 1^e helft 2022:

1. Open databestanden van DUO rondom Werving, instroom en doorstroom beschikbaar volgens een verbeterde richtlijn een format
2. In gesprek met andere data-eigenaren (CBS, Cambo, SBB) over hun databestanden en de aansluiting op de instellingswensen

WG2: verbeteren databeschikbaarheid

De uitdaging

Constant de balans vinden tussen gewenst detailniveau en mogelijk detailniveau.

De complicatie:

Het mogelijke detailniveau is afhankelijk van de strenge richtlijn binnen DUO.

Onze aanpak

Samen met DUO gewerkt aan een betere passende richtlijn die nu voorgelegd is aan FG.

WG2: verbeteren databeschikbaarheid

Het vervolg

1. Het beschikbaar krijgen van de gewenste bestanden volgens de nieuwe richtlijn.
2. Organiseren van 3 inhoudelijke sessies hoe open databestanden (kunnen) worden gebruikt
 1. Hoe gebruiken we het nu in onze organisatie?
 2. Hoe kunnen we het makkelijker (technisch) ontsluiten en koppelen?
 3. Data storytelling met open databestanden



Fieldlab Studievoortgang mbo

José Mulder,
Kwartiermaker Fieldlab

The slide features several abstract, rounded green shapes scattered across the top and left side. These shapes vary in size and orientation, some appearing as thin horizontal bars and others as more substantial, rounded forms. They are set against a dark teal background.

Fieldlab Studievoortga ng Mbo

- Officieel van start: zomer 2022
- Deelnemers: 16 mbo-instellingen
- Begeleiding: Cinop/ecbo
- Eindigt: eind 2024



Achtergrond Fieldlab



Covid: geen zicht op
effect thuisonderwijs op
studievoortgang mbo-
studenten



NPO: Welk effect hebben
interventies op
studievoortgang?

Aanloop...

- Juni '21: zo'n 20 scholen interesse in Datacoalitie/Fieldlab
- Najaar '21: notitie Fieldlab in relatie tot NPO
- Najaar '21: OCW en NRO akkoord begeleiding en onderzoek
 - Aanbesteding nodig
- Voorjaar '22: kwartiermakersfase

Kwartiermakersfase: wat willen scholen met Fieldlab?



Gesprekken met 21 mbo-instellingen die geïnteresseerd waren in Fieldlab



Medewerkers onderwijskwaliteit, managementinformatie, data-werkgroepen, etc.



Scholen zeer enthousiast, maar interesse ligt veel breder dan NPO.



Scholen willen zelf meer grip krijgen op/meer inzicht krijgen in studievoortgang eigen studenten.



Scholen willen van én met elkaar leren.

Fieldlab Studievoortgang mbo



Proeftuin. Iets nieuws uitproberen. Met en van elkaar leren zoals in een learning community.



Doel: Gezamenlijk komen tot concrete, op ervaring gebaseerde, handvatten om een actueler en completer beeld van de studievoortgang van mbo-studenten te kunnen realiseren en communiceren.



Eind Fieldlab. Scholen hebben 1) een helder beeld van wat zij als studievoortgang zien én 2) ervaring opgedaan met (data-gedreven) monitoring van studievoortgang.



Context: vanaf augustus 2023 dient mbo alle studenten (en ouders) inzicht te geven in studievoortgang.

Succesfactoren

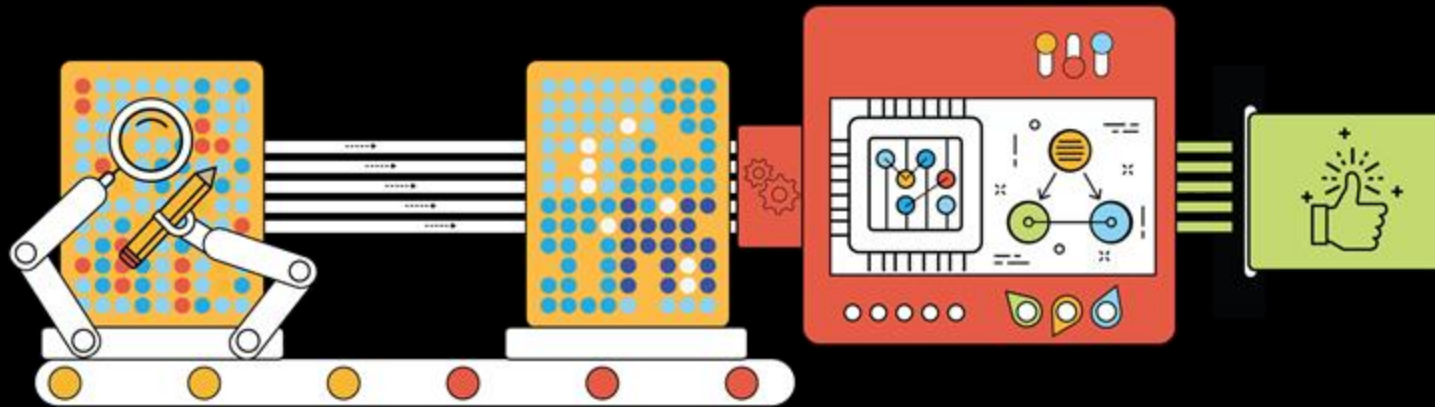
- Enthousiasme scholen
- Brede aanpak, blik op lange termijn
- Sluit aan bij bredere beweging van learning analytics, data-gedreven/ondersteund werken etc.

Mogelijke struikelblokken

- Concretisering doelstellingen
- Aantal deelnemers/betrokkenheid binnen scholen
- Gespreid opdrachtgeverschap

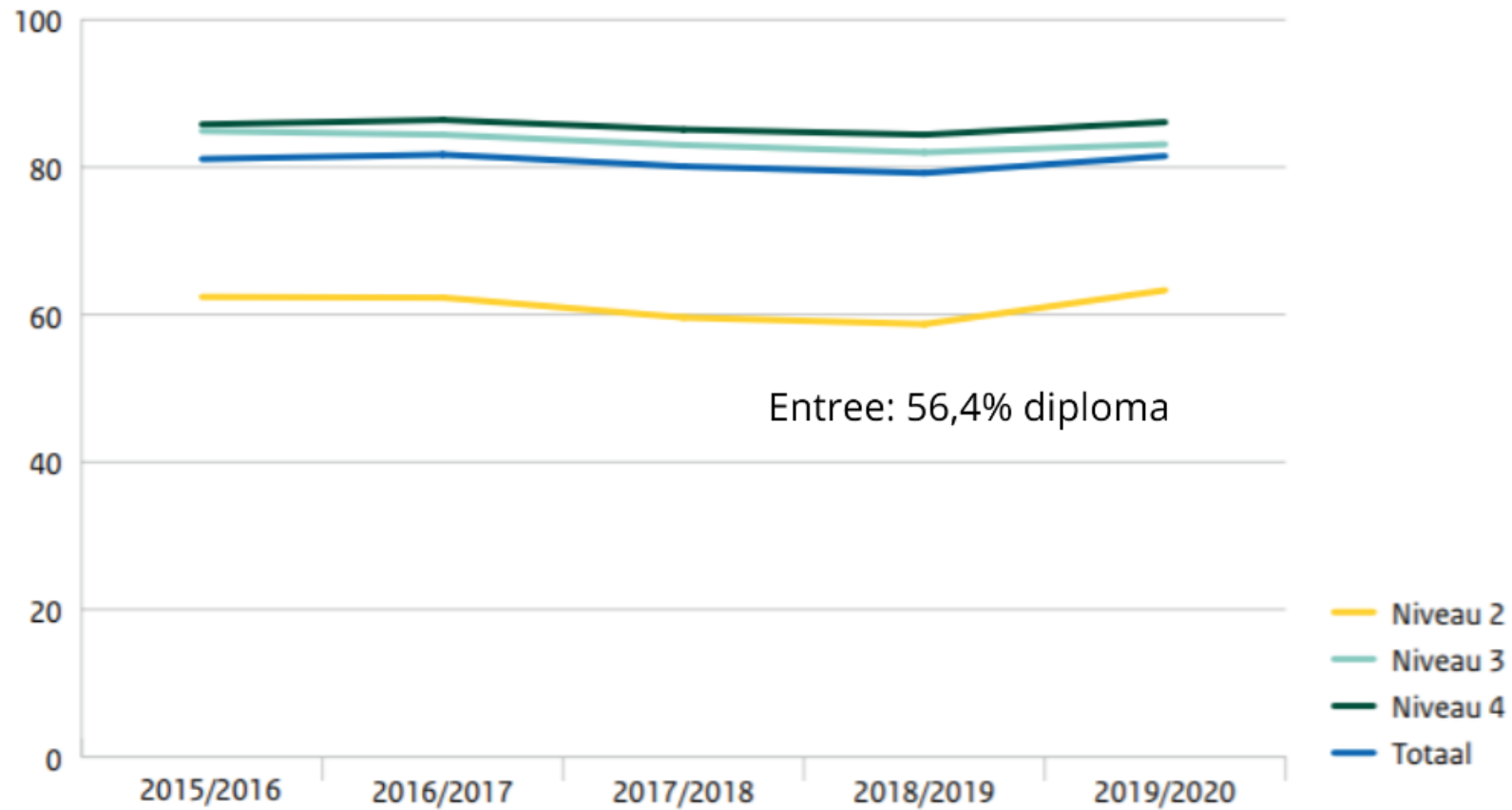


Voorspelmodel uitval mbo niveau 2
“Vroegtijdige identificatie tbv uitval preventie”



Vroegtijdige identificatie tbv uitval preventie

Figuur 5.1c Percentage studenten dat gediplomeerd het mbo verlaat in de jaren 2015/2016-2019/2020
(n 2019/2020*=147.921)



zi *voorlopige cijfers

Bron: Inspectie van het Onderwijs, 2021c

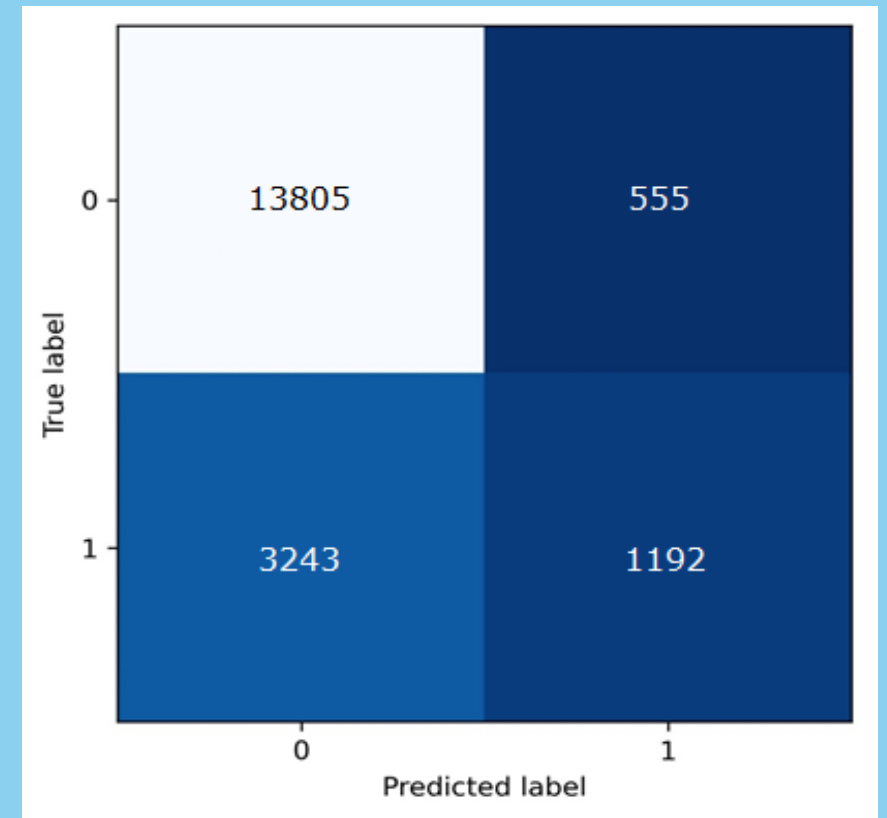
Vroegtijdige identificatie tbv uitval preventie

- Ontwikkeling voorspelmodel door data-scientist bij DUO
- Netwerkanalyse dat gebruik maakt van gradient-boosting techniek (machine learning).
- Niveau 2 studenten gelabeld als “uitval” en “geen uitval”.
- Model getraind met historische data van duo vanaf cohort 2015 t/m 2019
- Model bevat per student de persoonsgegevens, onderwijsverleden en omgevingskenmerken
- Populatie: mbo niveau 2 studenten die nieuw instromen in het mbo

Het “gradient boosting” model

Totale dataset: 187.944 leerlingen in de jaren 2016 – 2019 (77% niet-uitval, 23% uitval)
Testing set: 18.795 leerlingen (10% van de totale dataset)

Resultaat: een AUC = 0,82



Figuur 1: confusion matrix testing set gradient boosting

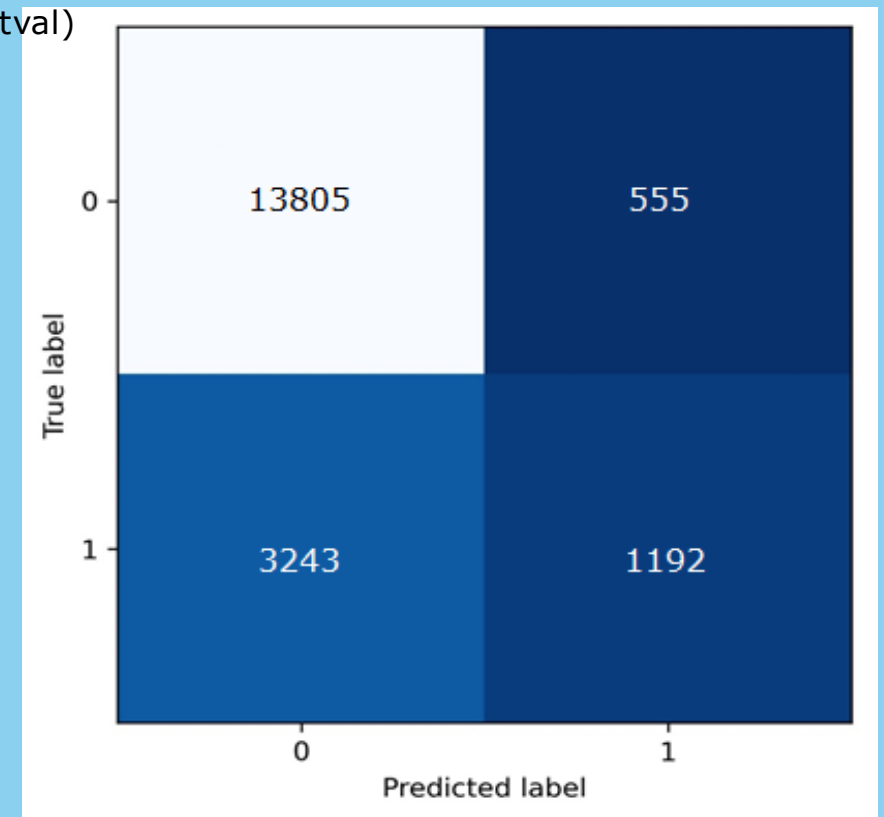
N.B. AUC is maat voor de mate waarin het model de uitvallers en succes studenten kan onderscheiden. 32

Het “gradient boosting” model

Totale dataset: 187.944 leerlingen in de jaren 2016 – 2019 (77% niet-uitval, 23% uitval)
Testing set: 18.795 leerlingen (10% van de totale dataset)

| | Training set | Testing set |
|---------------|--------------|-------------|
| Sensitiviteit | 29,2% | 26,9% |
| Specificiteit | 73,4% | 68,2% |

Wat opvalt is de lage sensitiviteit → 3/4 van de uitvallers wordt voorspeld als niet-uitvaller



Figuur 1: confusion matrix testing set gradient boosting

Conclusie

- Niet mogelijk studiesucces in het eerste studiejaar te voorspellen mbv DUO data.
- Er is geen verband tussen de gebruikte data en studiesucces in het eerste studiejaar
- Aanname dat de achtergrond van de student (o.a. apcg) voorspellend is voor studiesucces is niet juist

Discussie

- Data die het gedrag en prestaties van de studenten beschrijven zijn vermoedelijk noodzakelijk om studiesucces te kunnen inschatten. Eegdeman et al. 2022
- Onderzoek repliceren bij andere instellingen, om robuustheid van het model te onderzoeken

Lessons Learned:

- Neem de tijd en voorkom tunnelvisie: controleren van de resultaten door verschillende experts (4 ogen principe)
- Start de ethische discussie op de werkvloer
- Betrek onderwijsteams in het onderzoek (bruikbaarheid resultaten en draagvlak)



VERHALEN UIT DE COALITIES

Leo de Wit – G4 onderzoek met CBS – geen slides

WG1: Integraal Teamdashboard

Doel 1^e helft 2022:

“Onderzoeken van 3 best practices van datagedreven werken door onderwijsteams met Teamdashboards”

- Albeda college
- Curio
- Gilde opleidingen
- Koning Willem I college
- ROC Friese Poort
- ROC Nijmegen
- ROC Tilburg
- ROC van Amsterdam
- ROC van Twente

WG1: Integraal Teamdashboard

Aanpak:

- Korte presentatie van de status van je dashboard aan WG1 (iedere instelling)
- Selectie van 3 instellingen voor verder onderzoek
- Opstellen format met onderzoekscriteria
- Locatiebezoeken WG1 aan 3 instellingen
 - ROC Tilburg (8 april)
 - Curio (24 juni)
 - Albeda college (30 juni)
- Onderzoeksverslag per instelling

WG1: Integraal Teamdashboard

Status Quo:

Verschillen

- huidige ontwikkelingsfase
- focus en prioritering
- organisatorische inbedding
- ontwikkelings- en ontwerpstrategie
- gebruikte dashboard-technologie
- technische expertise (in- of extern)
- gebruikte bronsystemen (applicatielandschap)

Overeenkomsten

- informatiebehoefte grotendeels overeenkomstig
- informatiebehoefte neemt overal toe

Creatieve ideeën en oplossingen!

WG1: Integraal Teamdashboard

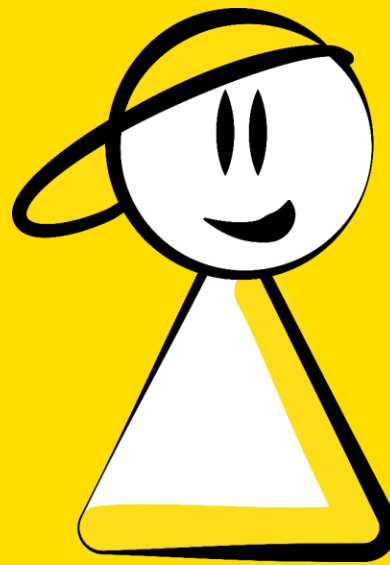
Lessons Learned:

- Organisatie commitment
 - Dashboard moet prominente positie krijgen
 - Goed luisteren
 - Aandacht voor het verhaal, niet alléén de cijfers
- Niet te groot willen starten
 - Prioritering
 - Iteratieve aanpak
 - Korte implementatie-cycli
- Voldoende personele capaciteit
 - Technische expertise intern
- Gemeenschappelijke taal
 - Eenduidige definities
 - Consistent datamodel
- Stel onderbouwende data van KPI's beschikbaar
 - Bewijslast (draagvlak)
 - Analyse

WG1: Integraal Teamdashboard

Vervolg:

- Twee aanvullende case studies:
 - ROC van Twente
 - ROC Mondriaan
- Twee use case uitwerkingen tot PvE
 - 'Datagedreven inzicht in de instroompopulatie'
 - 'Datagedreven besluitvorming over het BSA (voorlopig en definitief)'



Datacoalitie: Student Journey

Een ambitie om studie, leer en onderwijsdata effectiever in te zetten.

Agenda

- Introductie: datacoalitie Student Journey
- Student Journey & Albeda College
- Databronnen
- Struikelblokken

Datacoalitie



Uitgangspunten:

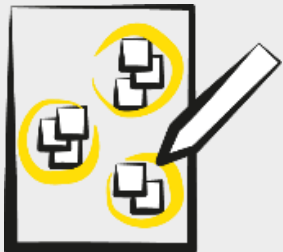
- Open-source oplossingen!
- Gezamenlijke AVG aanpak



Uitgangspunten:

- Werken met data dicht bij de student.
 - Onderwijsteamniveau.
- Reconstrueren van de student journey.
- Inzetten van Data Science.

Reconstrueren: Student Journey



Doel:

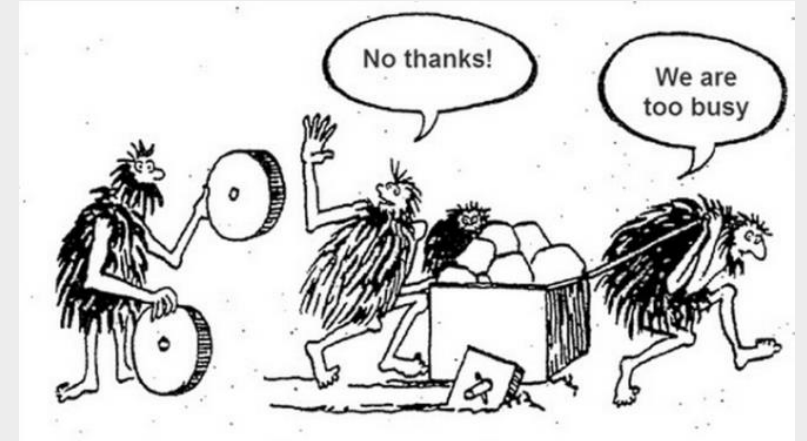
- Effectiever inzetten van onderwijs data.
- Aanpak voor werken met ongestructureerde data.
- Aan de slag met data!

saMBO-ICT

- Kennisdeling borgen.
- Communicatie.
- Gezamenlijke lobby.
- Olievlek werking.

Waarom samen?

- Kennisdeling
 - Technische barrières wegnemen:
 - Data Science expertise breed beschikbaar
 - Hoe kunnen we komen tot meer eenduidige bronnen
-
- 7 mbo instellingen & universiteit (JADS)



Wie doen er mee?

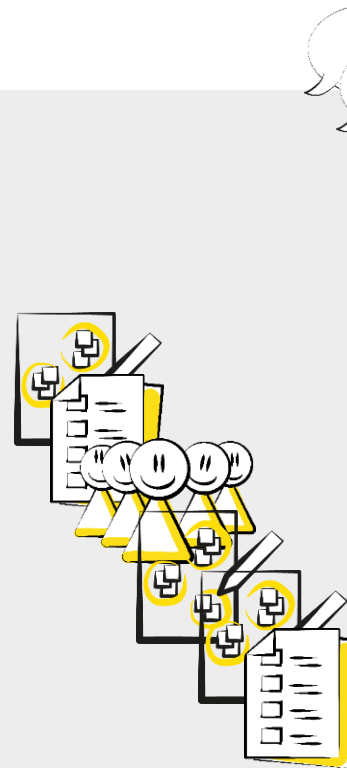
Noorderpoort



CINOP

Waarom Data Science?

- Complexiteit van de data
 - Verschillende structuren
 - Ontzettend veel dimensies
 - Verschillende vormen (entiteiten)
 - Geen eenduidige aanpak



| Eigenschap | Waarden |
|------------|----------------|
| Categorie | Opdracht |
| Ingeleverd | Ja |
| Resultaat | Voldoende |
| Op tijd | -7000 seconden |
| Feedback | Goed gedaan! |

- Traditionele onderzoeksmethoden komen tekort

Doel: reconstrueren Student Journey

- Inzetten en combineren van tot heden vaak onbenutten data, zoals:
 - Elektronische leeromgeving
 - Microsoft Teams
 - Aan/afwezigheid
 - Tussentijdse resultaten
- Samenvoegen tot een compleet beeld van de Student Journey:
 - Overstijgende factoren, zoals motivatie, discipline, activiteit etc.
 - Het leergedrag van student
 - Zicht op factoren uit de Student Journey die leiden tot studentsucces.

Programma: Studentsucces

Pilot



Van studiesucces naar student succes

De kwaliteit van het onderwijs is op orde, maar het studiesucces niet.

- Sturen op studiesucces verhoogd de kwaliteit van het onderwijs niet genoeg.
- Sturen op *studentsucces* geeft handvaten.

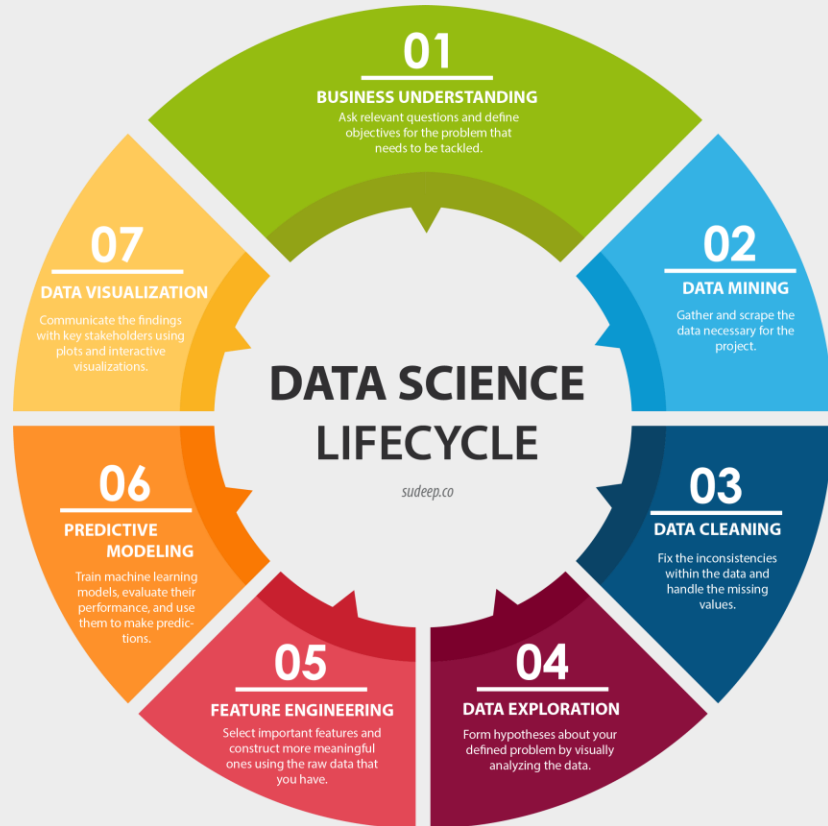
Data gedreven

- Wat is er aan de hand? (data nodig voor diagnose)
- Wat werkt? (data nodig voor analyse)
- Wat was het effect (data nodig voor evaluatie)

Van twee jaar geleden naar nu

- Vele apps die niet gebruikt werden.
- Teams zijn voorzien van uitgebreide datasets uit de apps
- Deze zijn besproken met teams
- We hebben opgehaald wat de teams missen, willen, nodig hebben:
 - *Wie is mijn student, heeft hij een kansenachterstand, wat deed hij voor dat hij bij ons kwam, wat zijn de kenmerken van uitvallers, wanneer vallen ze uit, wonen ze ver, wat gebeurt er met mijn student als hij bij mij weg is, toelatingsrecht, effect van bsa? Relatie tevredenheid, instroom, uitval, relatie verzuim en uitval. Binnen houden of juist vroeg doorverwijzen? Etcetera.*
- We hebben aannames bevestigd of ontkracht.
- We gingen bouwen!
- De Student Journey brengt de hele keten in beeld: van voor het Albeda naar tijdens het Albeda tot na het Albeda.

Experimenteren: mockup



Welke databronnen?

- Elektronische leeromgevingen
- Webshop (studieboeken)
- Aan/afwezigheid
- Tussentijdse resultaten/leeractiviteiten
- Intake gegevens (eerste aanbevelingen)
- Geldende examenresultaten

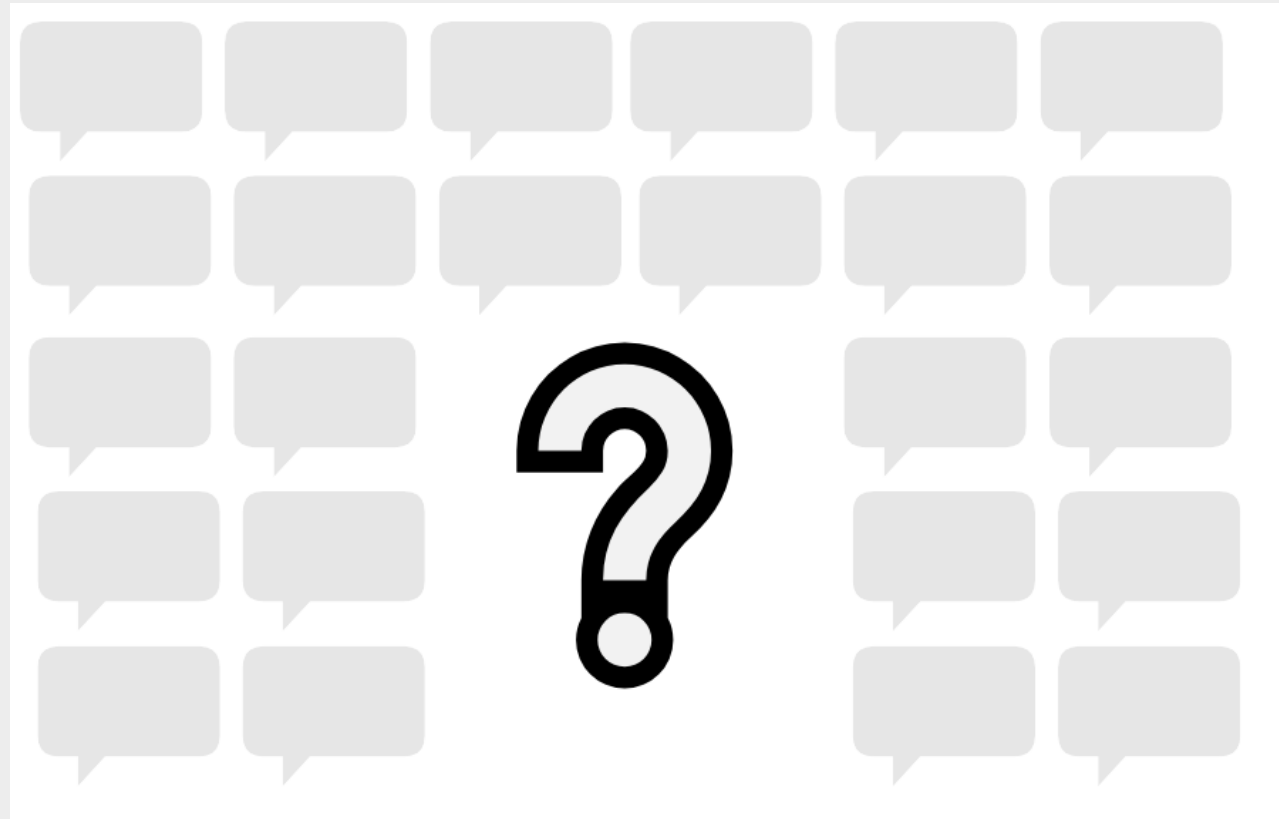
Struikelblokken (1)

- Verwerkersovereenkomst
- PIA → Privacy Impact Assessment
 - Geen goed zicht op de exacte variabelen
 - Herstructureren van variabelen in:
 - Controlerende
 - Structurerende
 - Onderzoek factoren
- Exploratief karakter, geen 'concrete' onderzoeksvraag
- Individueel afstemmen met de instelling
 - Meer vragen dan antwoorden
- Volledige anonieme data
 - Koppelen van data?

Struikelblokken (2)

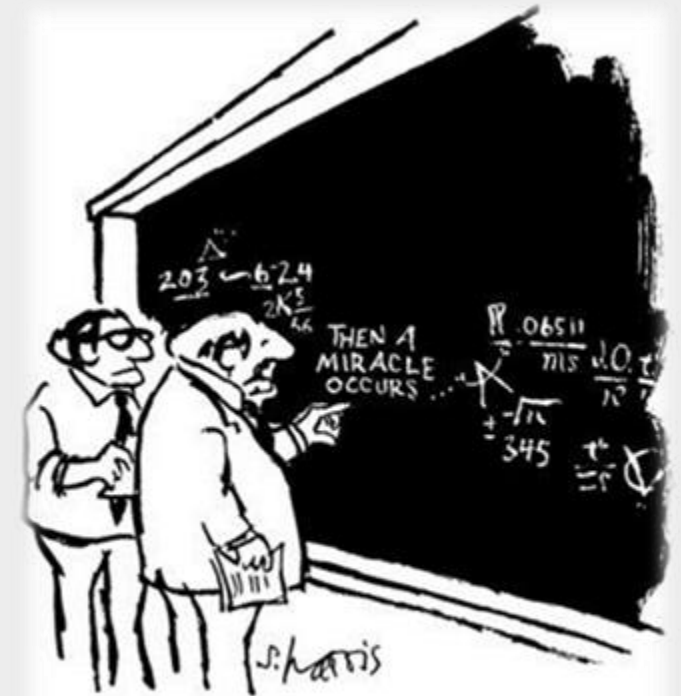
- Is exploratief onderzoek mogelijk met (anonieme) onderwijsdata?
- Wanneer is de data anoniem?
 - Ook als we geen zicht hebben op welke data er exact wordt verzameld?
- Wat kan/mag wel en niet met onderwijsdata van de student?
- Technische kennis van de FG/PO.

Vragen



Contactinformatie

- **Ronald Ferket**
- rferket@cinop.nl
- +31 6-10970963



"I think you should be more explicit here in step two."

Bedankt!



mbo^odigitaal



SURF



Ministerie van Onderwijs, Cultuur en
Wetenschap

Kennisnet

beroepsonderwijs  bedrijfsleven

Ga naar mbodigitaal.nl/doorpakken voor het volledige programma

Doorpakken op digitalisering