http://atlarge-research.com

Higher Education and Big Data: Vision, Ongoing Research, and Pragmatic Perspectives

Prof. dr. ir.
Alexandru
IOSUP

@Alosup

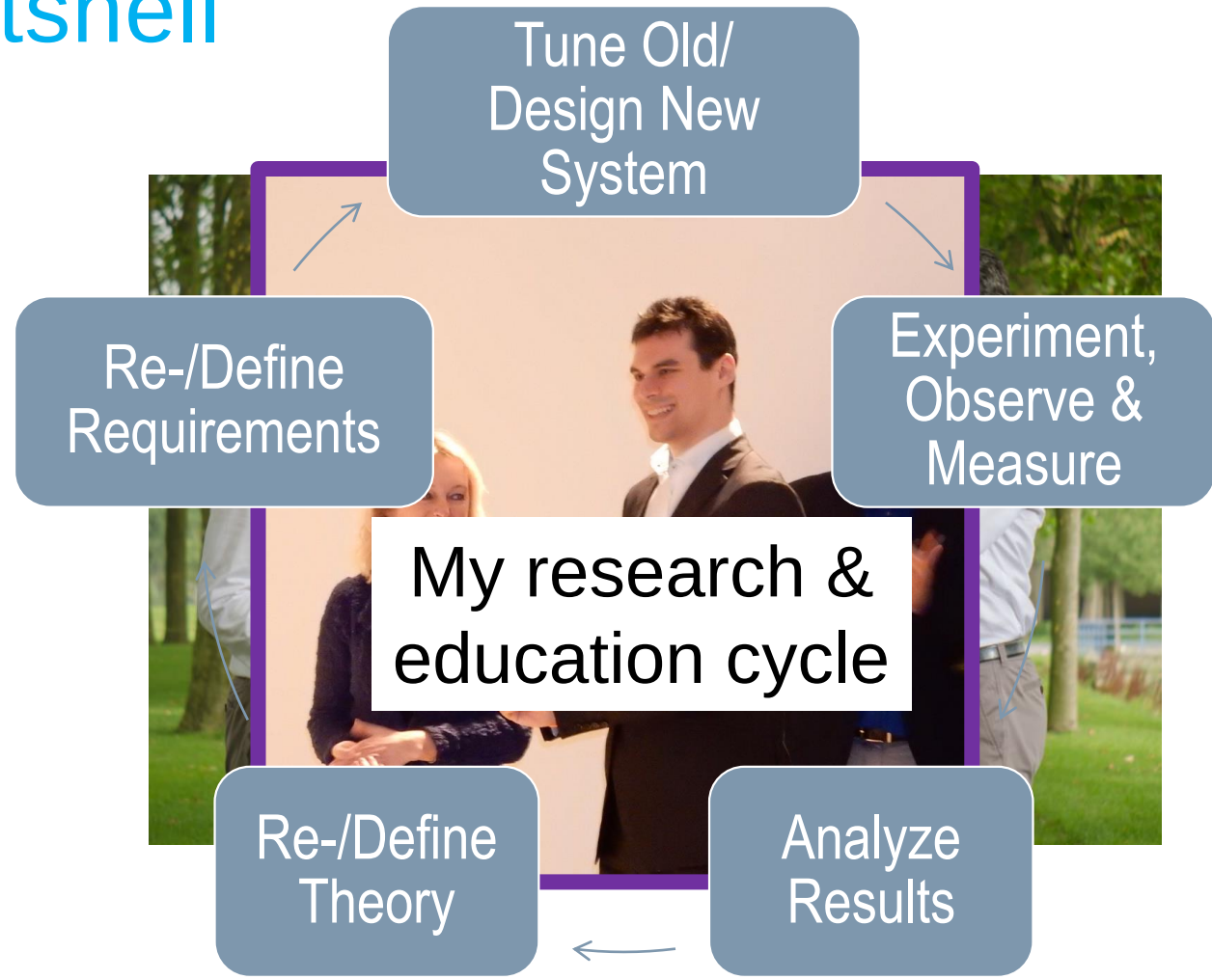
Background info, in a nutshell

I'm a Computer Scientist

- Large-scale computer systems
- Big data, cloud computing

I'm a Teacher

- Social gamification
- Young researchers
- Big Data in education



VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

Take-Home Message: What If Higher Education Processes Were Truly Informed?

Big Data = Progress in Higher Education

The Personal Academic File: Ethical Use of Student Data

- Data access and processing as basic service offered to all students
- Automated tools will inform and suggest course of action

Empowered = Engaged

- Student in control of own progress

Big Data = Automate + Enable

- Teachers focus on human activities

ICT challenges & Ethical Risks

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

21st Century Needs for Higher Education

Moving from a product-driven society to an information-/knowledge-driven society

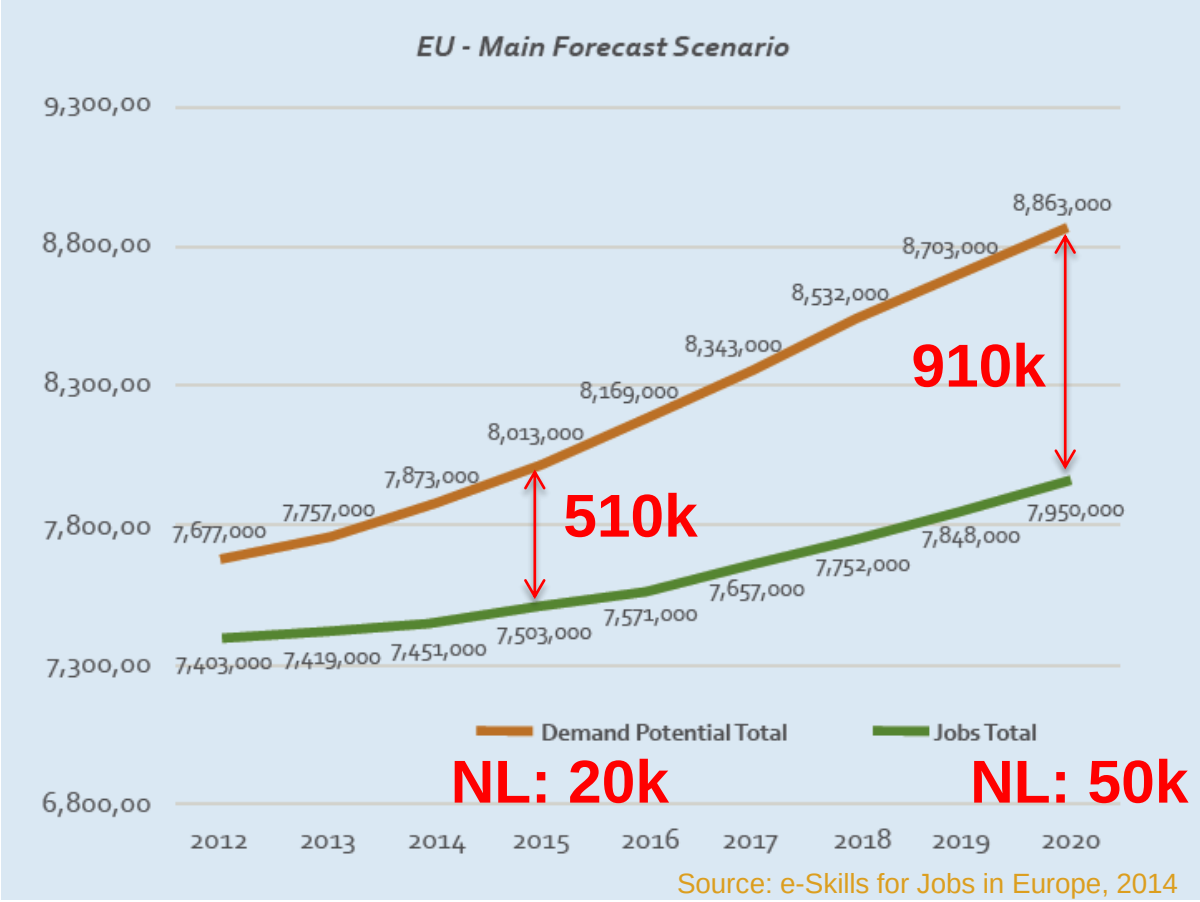
1. Students: finding flexibility, fun, self, and a job
2. Society: Massivizing and diversifying, behaving ethically
3. Industry: Requiring new and more complex skills
4. Academic System: Being accountable to stakeholders 1-3

5. Educators: Receiving proper recognition, tasks, and time

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

The Workforce Skill-Gap in ICT, EU and NL

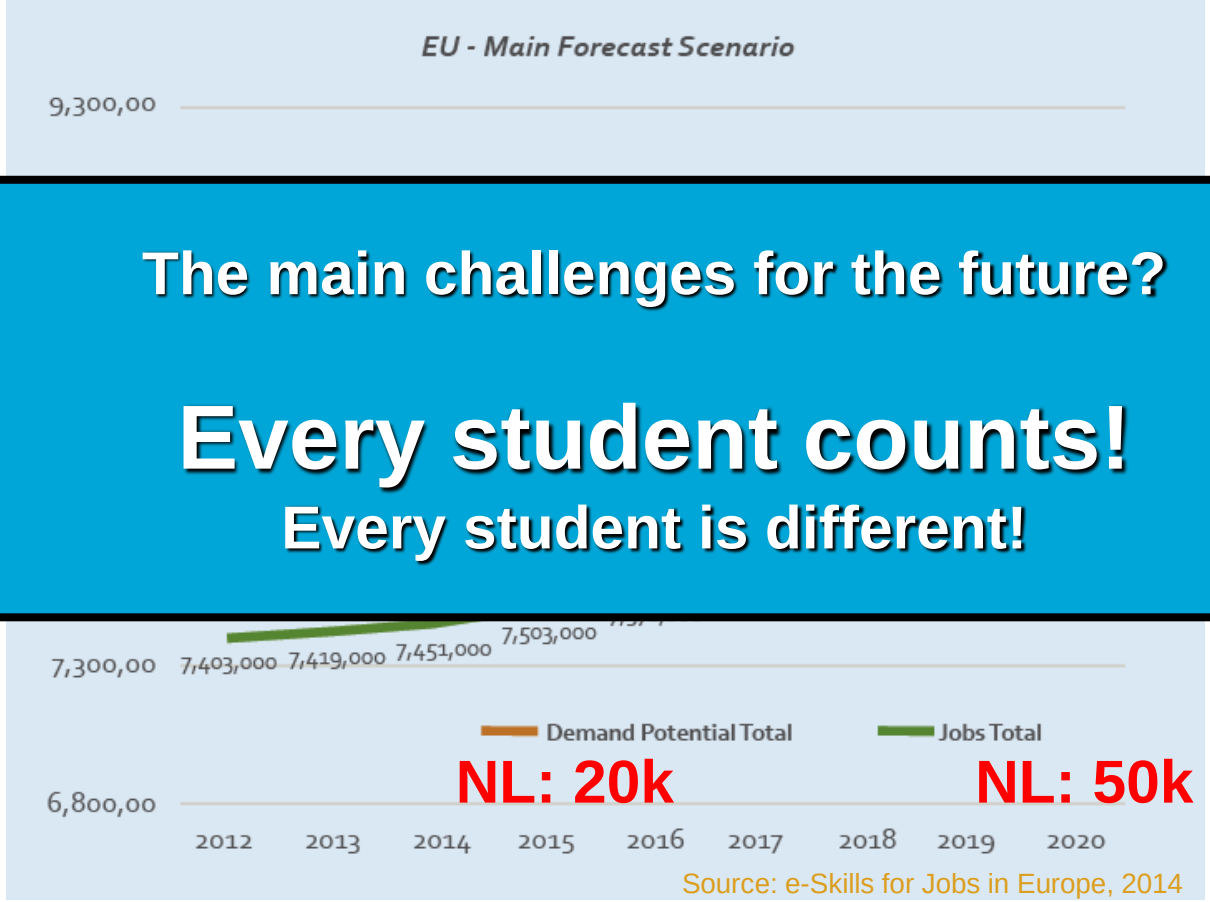


Source: e-Skills for Jobs in Europe, 2014

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

The Workforce Skill-Gap in ICT, EU and NL



Source: e-Skills for Jobs in Europe, 2014

The main challenges for the future?

Every student counts!

Every student is different!

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

The Workforce Skill-Gap in ICT, EU and NL

The main challenges for the future?

Every student counts!

Every student is different!

Every teacher counts!

Every teacher is different!

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

~35’

Higher Education + Big Data

Background info

2’ — Needs of 21st-Century Higher Education

5’ — Big Data: Myth vs. Reality

15’ — Vision for Big Data in Higher Education

1. Personalize Education
2. Increase accountability, reduce discrimination
3. Improve communication of student progress
4. New generation of learning design, didactics
5. Optimize processes, reduce bureaucracy

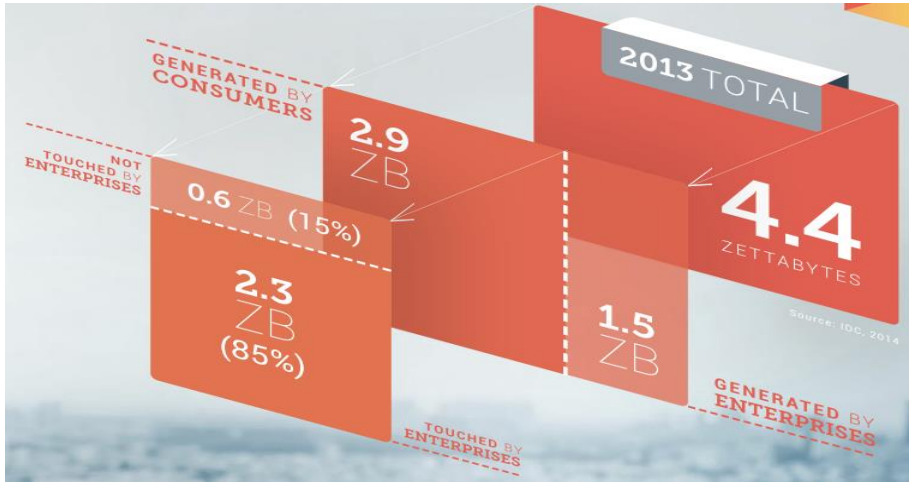
10’ — Big Data Challenges and Risks

2’ — A Prediction

2’ — Take-Home Message

How can Big Data help alleviate the challenges of modern higher education?

- Big Data seems to work in so many domains!



Sources: IDC, EMC.

- Specifically, Big Data helps to:
  - Engineer knowledge processes
  - Answer complex questions
  - Make decisions and personalize content
  - Become data-driven, fact-driven, knowledge-driven

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

Big Data, the Why (An Anecdotal Example)

The Overwhelming Growth of Knowledge

"When 12 men founded the Royal Society in 1660, it

Professionals already know they don't know [it all]

In the last 50 years, such has been the pace of scientific advance that even the best scientists cannot keep up with discoveries at frontiers outside their own field."

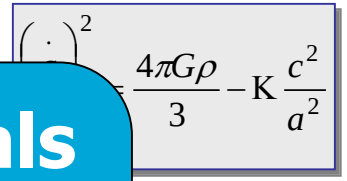
Tony Blair,
PM Speech, May 2002

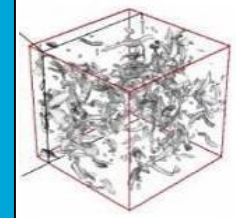
Number of Publications	1993 1997	1997 2001
		08
		5
Japan	289,751	336,858
France	203,814	232,058
Canada	168,331	166,216
Italy	122,398	147,023
Switzerland	57,664	66,761
Netherlands	83,600	92,526

Data: King,The scientific impact of nations,Nature'04.

From Hypothesis to Data Exploration (and Back)

1. Thousand years ago:
   science was **empirical** describing natural phenomena
2. Last few hundred years:





The Fourth Paradigm is suitable for professionals who already know they don't know [enough to formulate good hypotheses], yet need to deliver quickly

4. **data exploration**

unify theory, experiment, and simulation

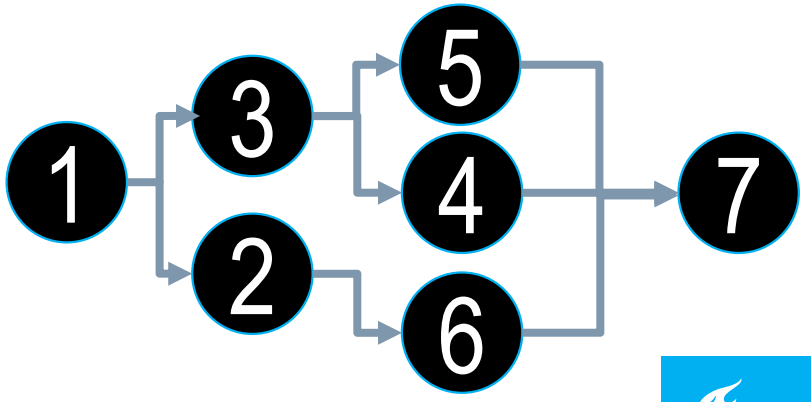
- Data captured by instruments or generated by simulator
- Processed by software
- Information/Knowledge stored in computer
- Scientist analyzes results using data management and statistics

Source: Jim Gray and "The Fourth Paradigm", http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Big Data, the How: from Data to Knowledge

1. Raw Data (that's what most hear about)
2. Datasets
3. Data Lakes
4. Data Primitives (pre-processing)
5. Computer Algorithms
6. Computer Tasks
7. Computer Workflows

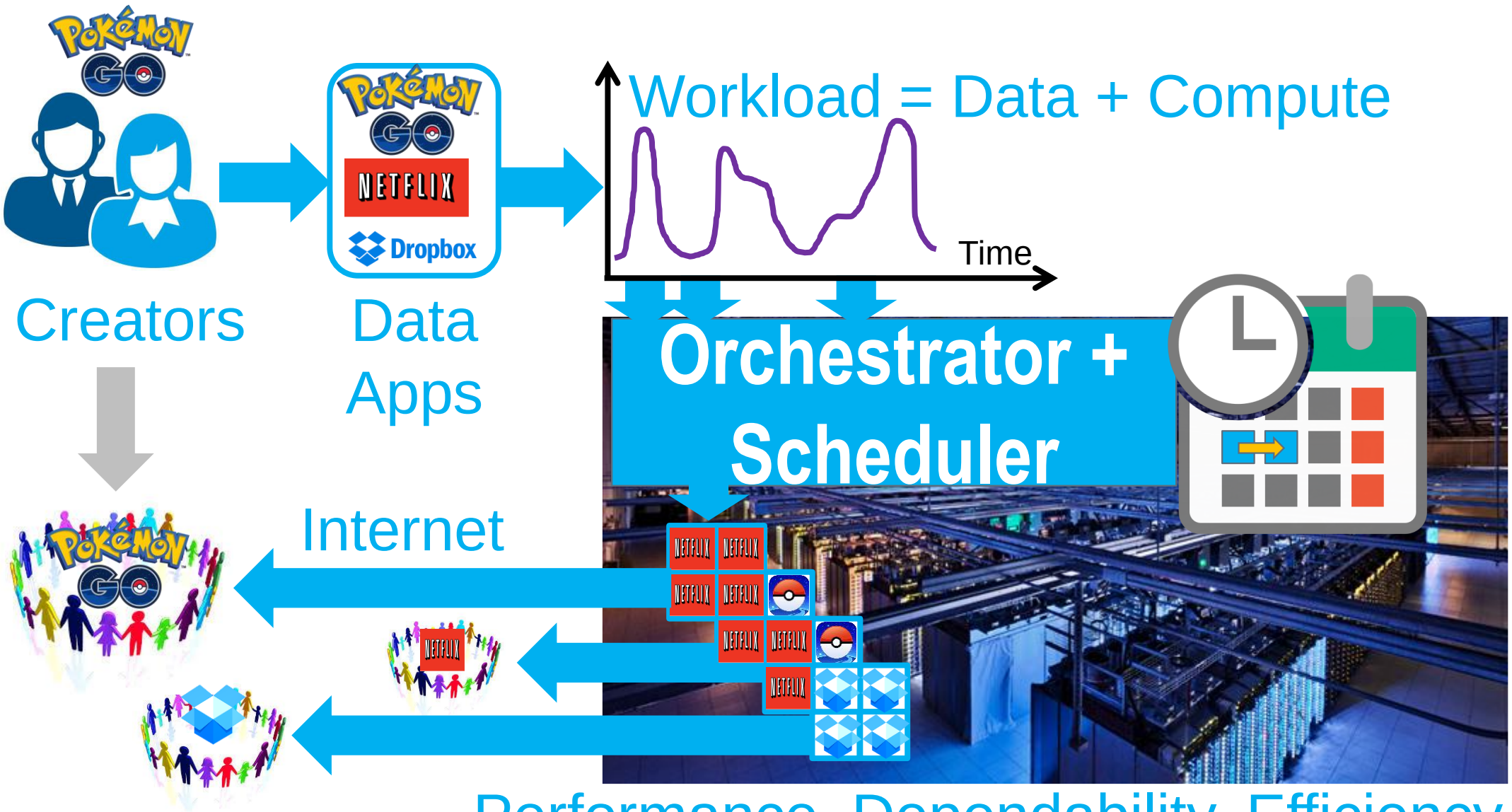
Source: ComputerWeekly.com



VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

Big Data, the How: the Current Technology



~35’ Higher Education + Big Data

Background info

2’ — Needs of 21st-Century Higher Education

5’ — Big Data: Myth vs. Reality

15’ — Vision for Big Data in Higher Education

1. Personalize Education
2. Increase accountability, reduce discrimination
3. Improve communication of student progress
4. New generation of learning design, didactics
5. Optimize processes, reduce bureaucracy

10’ — Big Data Challenges and Risks

2’ — A Prediction

2’ — Take-Home Message

So What Can Big Data Do for Higher Education?

- As any technology:
  1. Improve existing processes, typically through automation
  2. Enable new processes, typically by augmenting human abilities

- Big Data could help Higher Education to:

  "Significantly improve and scale higher education, assisted by digital means and other methods to improve the efficiency and the quality of education, working ethically across different learning cultures, orienting students towards science, industry, governance, and society at large."—TU Delft pilot-project started by A. Iosup (2015—ongoing)

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

Vision for Big Data Use in Higher Education

The Personal Academic File: Ethical Use of Student Data

- Enable data access and processing to help students
- Develop automated tools to inform and, if possible, to suggest course of action
- Use data ethically: opt-in, etc.

Empowered = Engaged

- Student in control of own progress
- Detailed feedback becomes possible
- Student advisor can give better advice

Big Data = Automate + Enable

- Make Higher Education efficient
- Let teachers focus on human activities
- Enable a new generation of design, didactics

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

1/ Personalizing Education (Guiding Students to Achieve More)

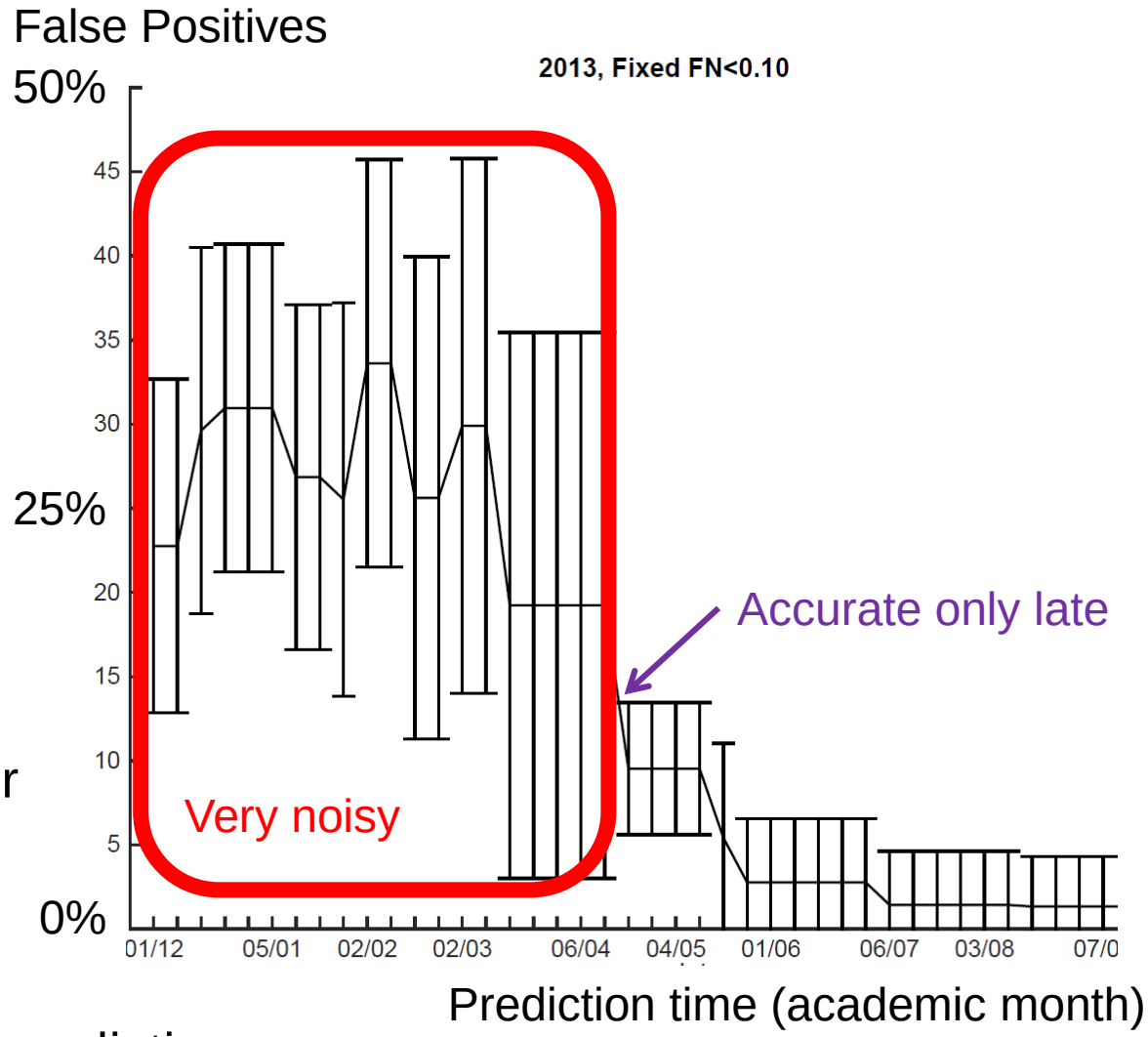
- Problem: internationalization, diversification poses numerous problems, study advisors have difficulty coping meaningfully
- Vision: big data personalizes education, guides students or helps study advisors and educators
- A few ideas on how to use big data:
  1. Learning analytics (educational data mining) automate analyzing learning paths
  2. Big data workflows use learning analytics as basis for recommendations
  3. Students receive personalized analysis, cues, advice, coaching

1 Predicting BSA Pass/Fail

TU Delft, Faculty EEMCS, the Netherlands

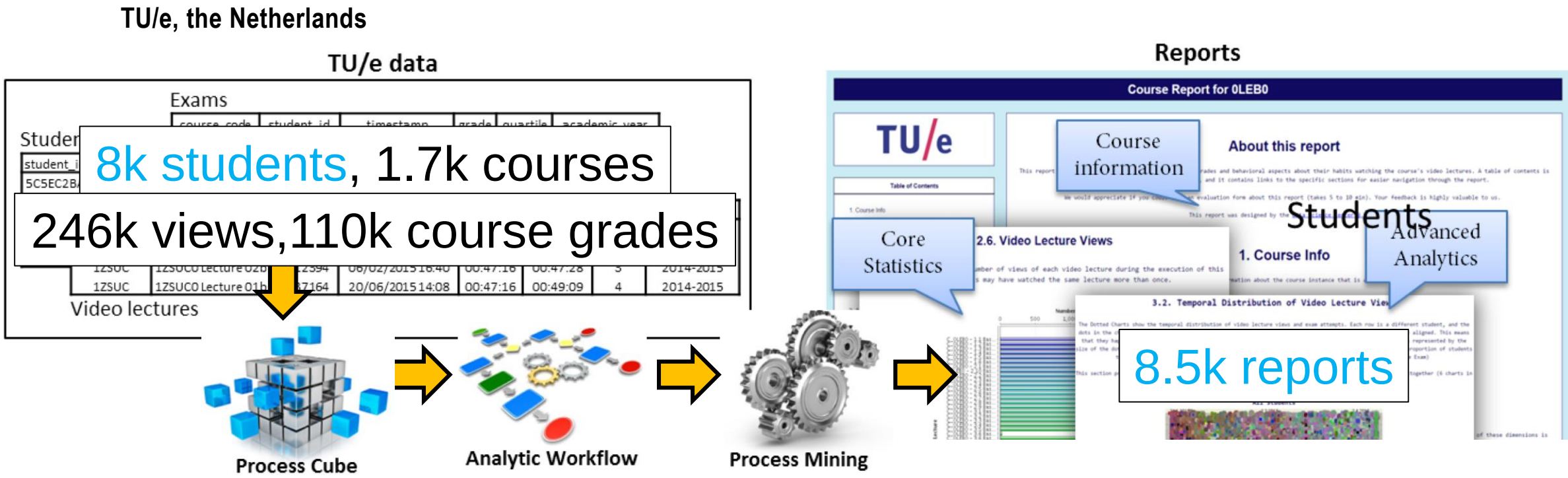
- Tried various features
- Tried 7 classifiers (+ config)

- Found about predicting pass/fail:
  - (confirmation) Wiskunde D good predictor
  - Exam pass/fail better than exam grades
  - Prediction is **noisy**, accurate only **late**
  - Higher number of students means better predictive power



D. Tax, S. Verwer, M. de Weerdt, O. Visser, A. Iosup, E. Hendriks: Using early achievements of first-year students for counselling. (draft, unpublished)

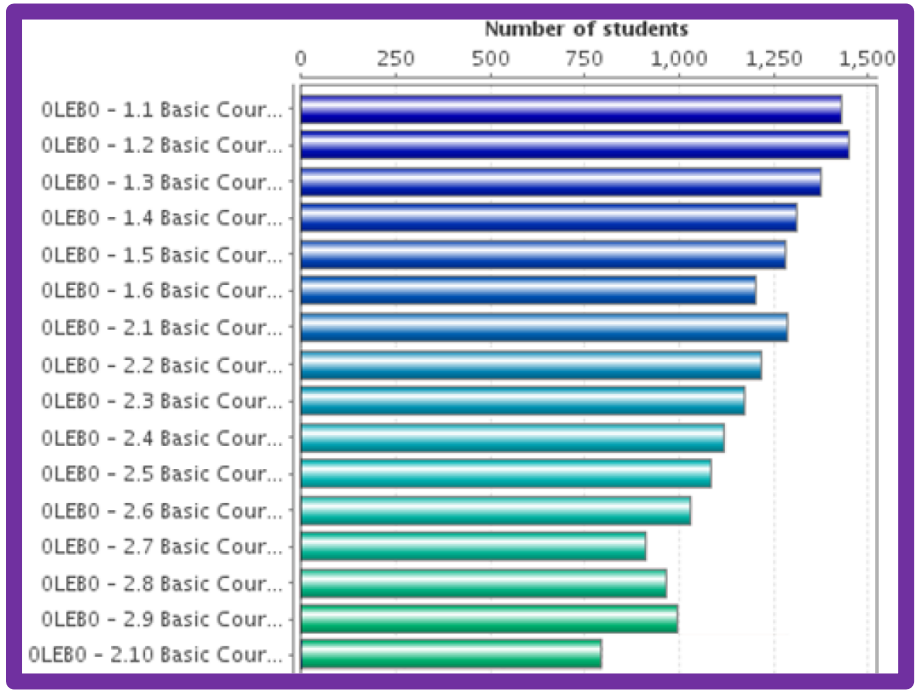
❷ Process Cubes and Mining, Analytic Workflow



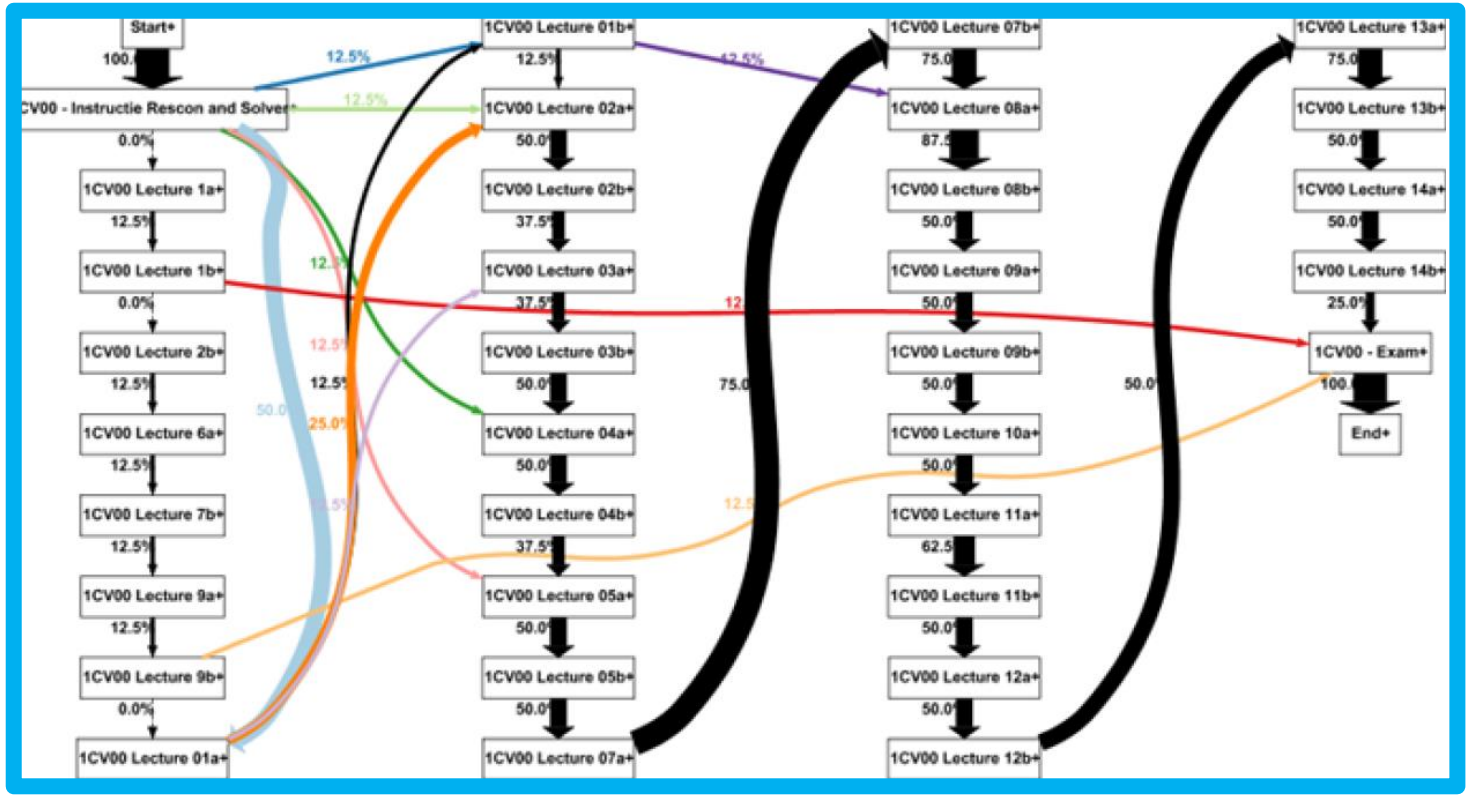
- Specialized process cube and analytic workflow
- Generic process mining tools

A. Bolt, M. de Leoni, W. van der Aalst, P. Gorissen: Exploiting Process Cubes, Analytic Workflows and Process Mining for Business Process Reporting: A Case Study in Education. SIMPDA 2015: 33-47

❷ Process Cubes & Mining, Analytic Workflows



Dropout: Complete data



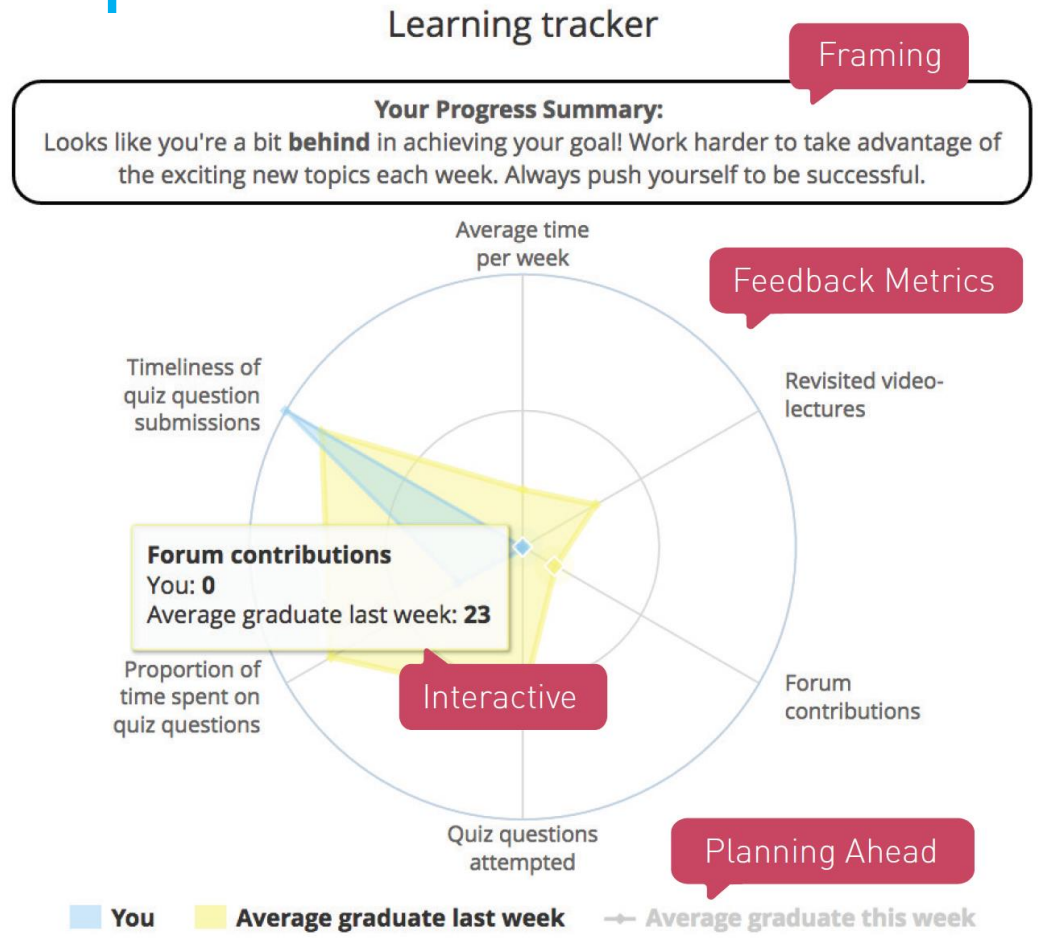
Path of advancement: model vs. actual progress

A. Bolt, M. de Leoni, W. van der Aalst, P. Gorissen: Exploiting Process Cubes, Analytic Workflows and Process Mining for Business Process Reporting: A Case Study in Education. SIMPDA 2015: 33-47

3 Giving Cues and Social Comparison

TU Delft, MOOCs, the Netherlands

- Theory: We establish worth socially
- Experiments
  - 4 randomized controlled trials
  - 34k students
- Findings:
  - Personalized social-comparison feedback correlated with completion
  - Only highly educated learners benefit
  - Cultural context manipulated, found significant (ethical?)



D. Davis, I. Jivet, R. F. Kizilcec, G. Chen, C. Hauff, G.-J. Houben: Follow the successful crowd: raising MOOC completion rates through social comparison at scale. LAK 2017: 454-463

2/ Increase Accountability, Reduce Discrimination

- Problem: groups defined by socioeconomic status, race, ethnicity, or gender become under-represented or have achievement gaps
- Vision: big data complements and informs policy-based approaches.
- A few ideas on how to use big data:
  1. Automated feedback and dynamic recognition for student & educator achievement
  2. Data recorded at every stage in the process allows understanding, correcting bias
  3. Diversity of metrics fosters self-mastery, alleviates implicit bias in exam grading
  4. Data + statistics- and simulation-based approaches allow testing explanations

3/ Improve Communication of Student Progress

- Problem: agreements between students and educators are poorly recorded, student achievements are inconsistently reported to industry
- Vision: big data facilitates understanding student progress, enables hiring
- A few ideas on how to use big data:
  1. Student log-books, shared as code is shared right now (GitHub for study-work)
  2. Student portfolios as early business-card for students, and educators and courses
  3. Mining log-books and portfolios useful to improve achievement, completion

VU VRIJE UNIVERSITEIT AMSTERDAM

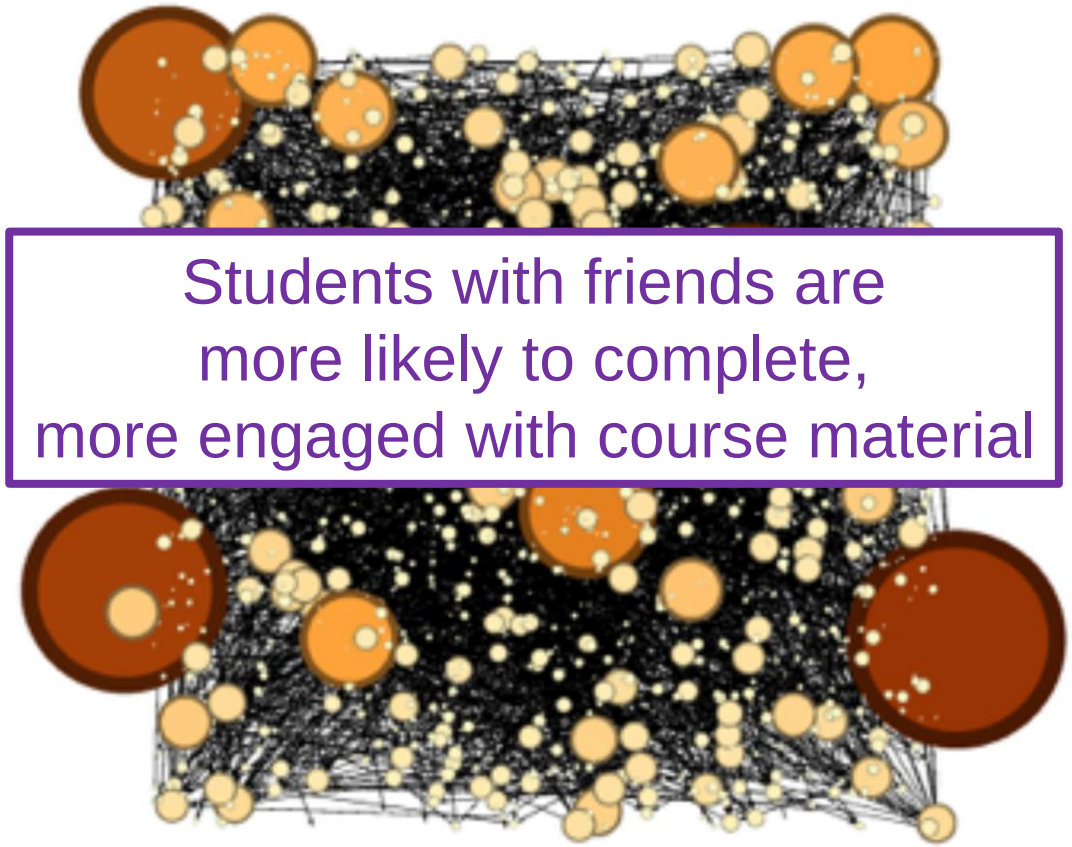
TU Delft

2

Student Portfolios across the Social Web

TU Delft, MOOCs, the Netherlands

- Identified ~4k student profiles
  - Gravatar, LinkedIn
  - Twitter
  - StackExchange, GitHub
- Found:
  - Can identify learners, socnet
  - Can infer skill, gender



Students with friends are more likely to complete, more engaged with course material

G. Chen et al.: Learning Transfer: Does It Take Place in MOOCs?, L@S 2016

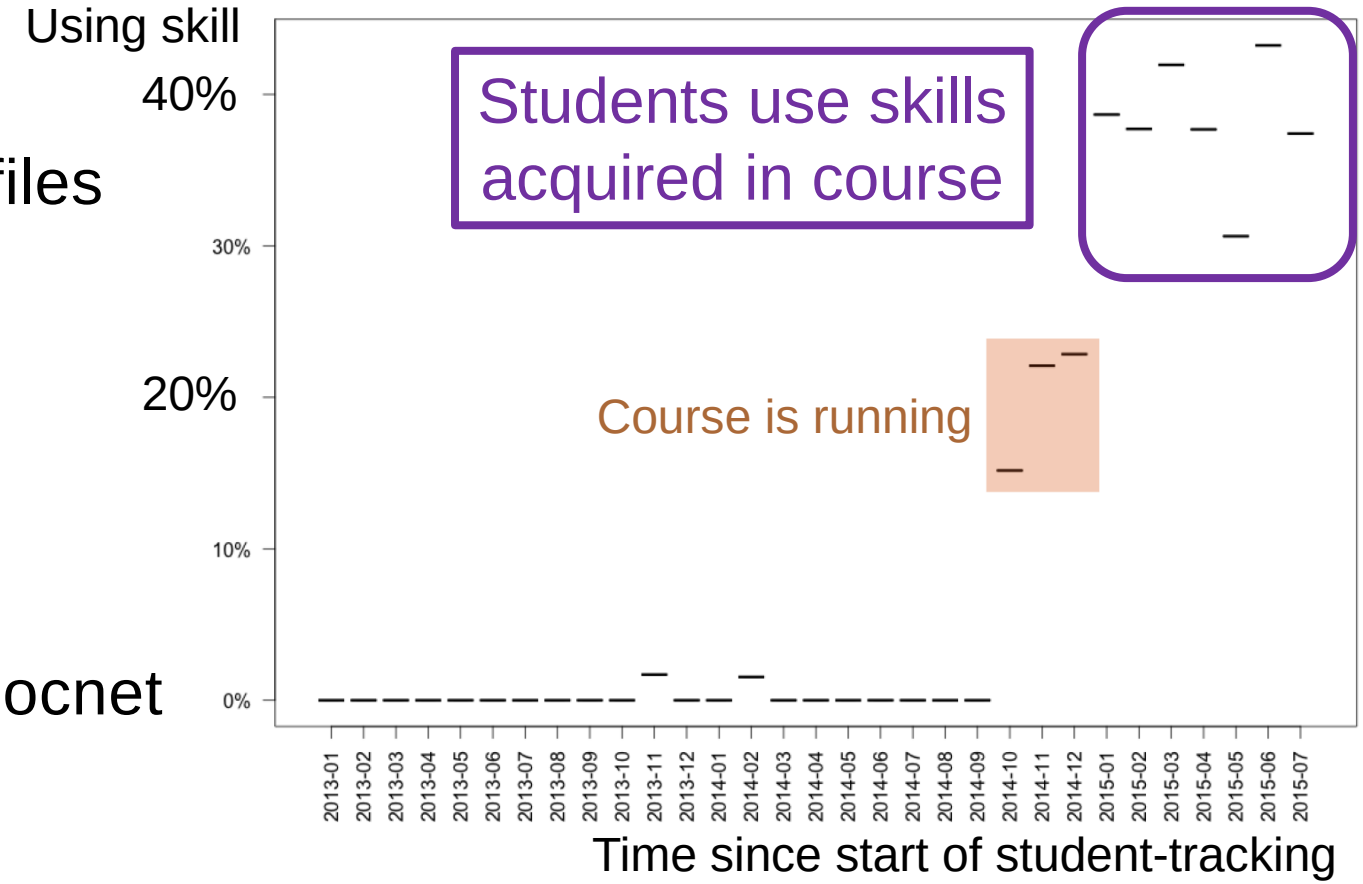
G. Chen et al.: Beyond the MOOC platform, WebSci 2016

2

Student Portfolios across the Social Web

TU Delft, MOOCs, the Netherlands

- Identified ~4k student profiles
  - Gravatar, LinkedIn
  - Twitter
  - StackExchange, GitHub
- Found:
  - Can identify learners, socnet
  - Can infer skill, gender
  - Can track skill use in professional networks



G. Chen et al.: Learning Transfer: Does It Take Place in MOOCs?, L@S 2016

G. Chen et al.: Beyond the MOOC platform, WebSci 2016

4/ New Generation of Learning Design, Didactics

- Problem: designing new learning objects, learning approach, and didactics is difficult to test or tune, esp. before obsolescence
- Vision: big data will help testing and tuning designs, early
- A few ideas on how to use big data:
  1. Tracking use of designs, creating stats and correlations
  2. Proposing hypotheses, A/B and similar hypothesis testing
  3. Dynamic generation, configuration, and use of learning objects
  4. Dynamic selection and configuration of learning designs

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

5/ Optimize Processes, Eliminate Bureaucracy

- Problem: education processes are heavy on bureaucracy, incur much setup, regular, and exceptional work per student
- Vision: big data workflows optimize education processes and reduce the burden of bureaucracy
- A few ideas on how to use big data:
  1. Complex workflows inform multiple stakeholders while efficiently using human time and infrastructure
  2. Interoperable big data systems allow drawing conclusions unavailable at smaller levels, including institutional, regional, and national

~35’ Higher Education + Big Data

Background info

2’ — Needs of 21st-Century Higher Education

5’ — Big Data: Myth vs. Reality

15’ — Vision for Big Data in Higher Education

1. Personalize Education
2. Increase accountability, reduce discrimination
3. Improve communication of student progress
4. New generation of learning design, didactics
5. Optimize processes, reduce bureaucracy

10’ — Big Data Challenges and Risks

2’ — A Prediction

2’ — Take-Home Message

Technological Risks: Mastering the Big Data "V"s

- Initial Vs: (why traditional databases can't cope)

Volume	Velocity	Variety
Bytes or Records	Batch for tomorrow	(Semi-)Structured
Files or Transactions	Interactive a-la-minute	Unstructured
Producers/Consumers	Streams for now	Multi-factor/-variate

Adapted from: Doug Laney, "3D data management", META Group/Gartner report, Feb 2001. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Technological Risks: Mastering the Big Data "V"s

- Initial Vs: (why traditional databases can't cope)

Technology-driven

How to use data, algorithms, and computing infrastructure for new-generation data processing?

- The later Vs:

Veracity	**Viability & Value**
Trustworthy? Authentic?	Causal link? Actionable?

Adapted from: Doug Laney, "3D data management", META Group/Gartner report, Feb 2001. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Technological Risks: Mastering the Big Data "V"s

- Initial Vs: (why traditional databases can't cope)

Technology-driven
How to use data, algorithms, and computing infrastructure for new-generation data processing?

- The later Vs:

Question-driven
How to answer complex questions with Big Data?

- Our own V:

Vicissitude (Tech- and Question-driven)
How to cope with changing Vs?

NL winners IEEE Scale Challenge 2014

Adapted from: Doug Laney, "3D data management", META Group/Gartner report, Feb 2001. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Ethical Risks that Require Governance, Societal, and Academic Checks and Balances

1. Data risks
2. Computing risks

Who controls data/compute? Who collects data? Who develops algos? Are all processes humane? Is there a right to recourse? A right to forgiveness? Who is excluded and why? Etc.

3. Scientific risks

Gaining data-driven (loose inference-based) knowledge must not mean losing deep (model-based) or deductive (principle-based) knowledge

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

~35’ Higher Education + Big Data

Background info

2’ — Needs of 21st-Century Higher Education

5’ — Big Data: Myth vs. Reality

15’ — Vision for Big Data in Higher Education

1. Personalize Education
2. Increase accountability, reduce discrimination
3. Improve communication of student progress
4. New generation of learning design, didactics
5. Optimize processes, reduce bureaucracy

10’ — Big Data Challenges and Risks

2’ — A Prediction

2’ — Take-Home Message

Predictions, Pred…, Pr…: A Decade from Now,

Big Data in Higher Education

- Everyday tool for all stakeholders (diversity a massive challenge in society)
- Does not solve: adequate financing, recognition of teachers, research/education

The Personal Academic File: Ethical Use of Student Data

- Data access and processing as basic service offered to all students (digital literacy)
- Automated tools will inform and suggest course of action (ethical issues)

ICT challenges

- Effective and efficient platform (technology evolution)
- Automated tools coping with big data (vicissitude)

Many ethical risks

- Many solvable problems
- Tough and wicked problems

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

Take-Home Message: What If Higher Education Processes Were Truly Informed?

Big Data = Progress in Higher Education

The Personal Academic File: Ethical Use of Student Data

- Data access and processing as basic service offered to all students
- Automated tools will inform and suggest course of action

Empowered = Engaged

- Student in control of own progress

Big Data = Automate + Enable

- Teachers focus on human activities

ICT challenges & Ethical Risks

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft

Disclaimer: images used in this presentation obtained via Google Images.

- Images used in this lecture courtesy to many anonymous contributors to Google Images, and to Google Image Search.
- Many thanks!

VU VRIJE UNIVERSITEIT AMSTERDAM

TU Delft