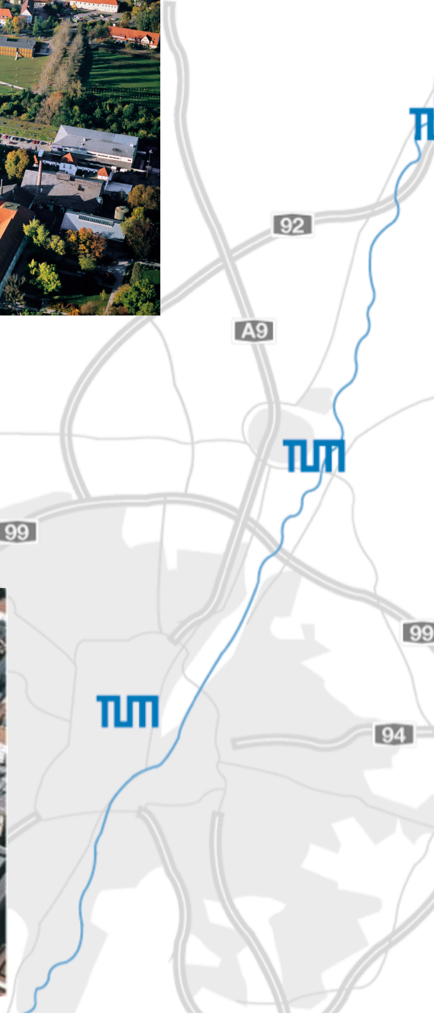
Predictive Analytics in Higher Education

Hans Pongratz
Senior Vice President & Chief Information Officer
pongratz@tum.de

EUNIS BITF Conference, March 7th 2014, Paris

TUM. Campus



TUM. Some numbers

13	Faculties	
~ 400	Buildings	
154	Studiengänge	
~ 36 000	Students	33% female Students 20% int. Students
~ 500	Professors	
~ 10 000	Staff members	

13	Nobel Prizes
15	Leibniz-Prizes
4	Humboldt Professors

#53 2013 Academic Ranking of World Universities

IT-Strategy: The Digital University

Leitmotif since 2002

Completed IT-Projects:

SAP@TUM

IntegraTUM

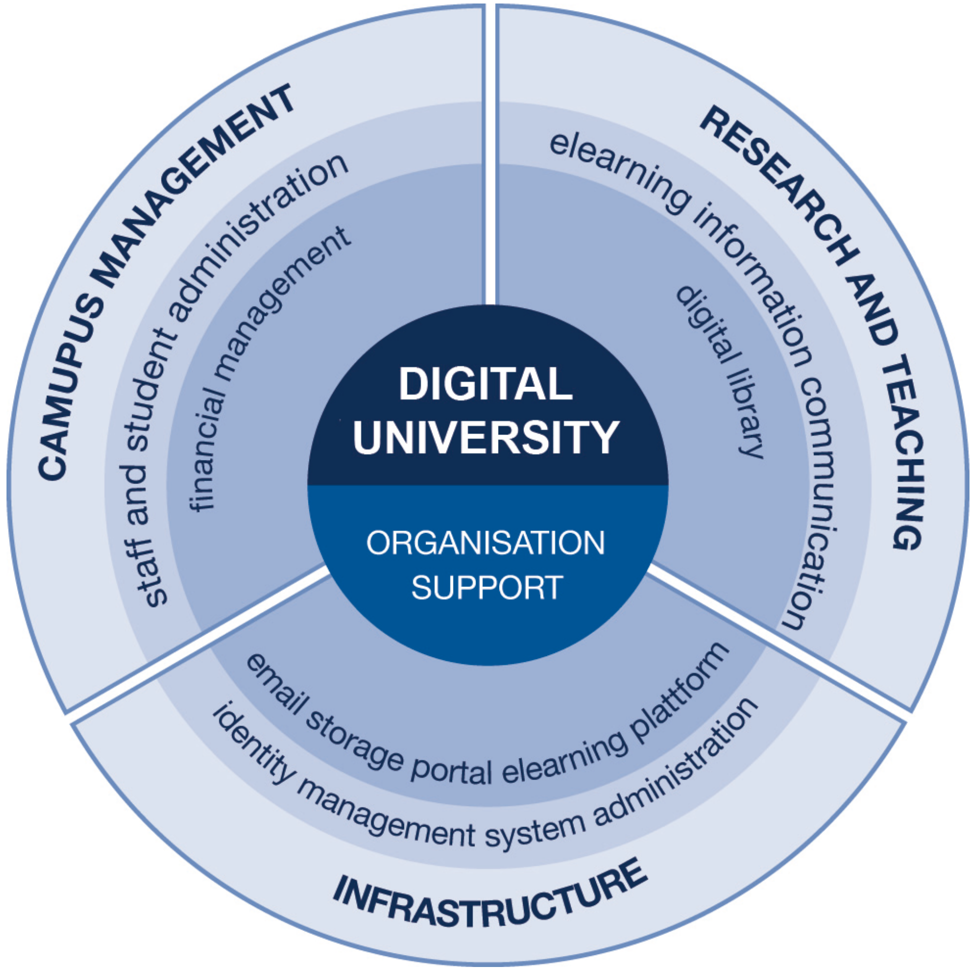
elecTUM

mediaTUM

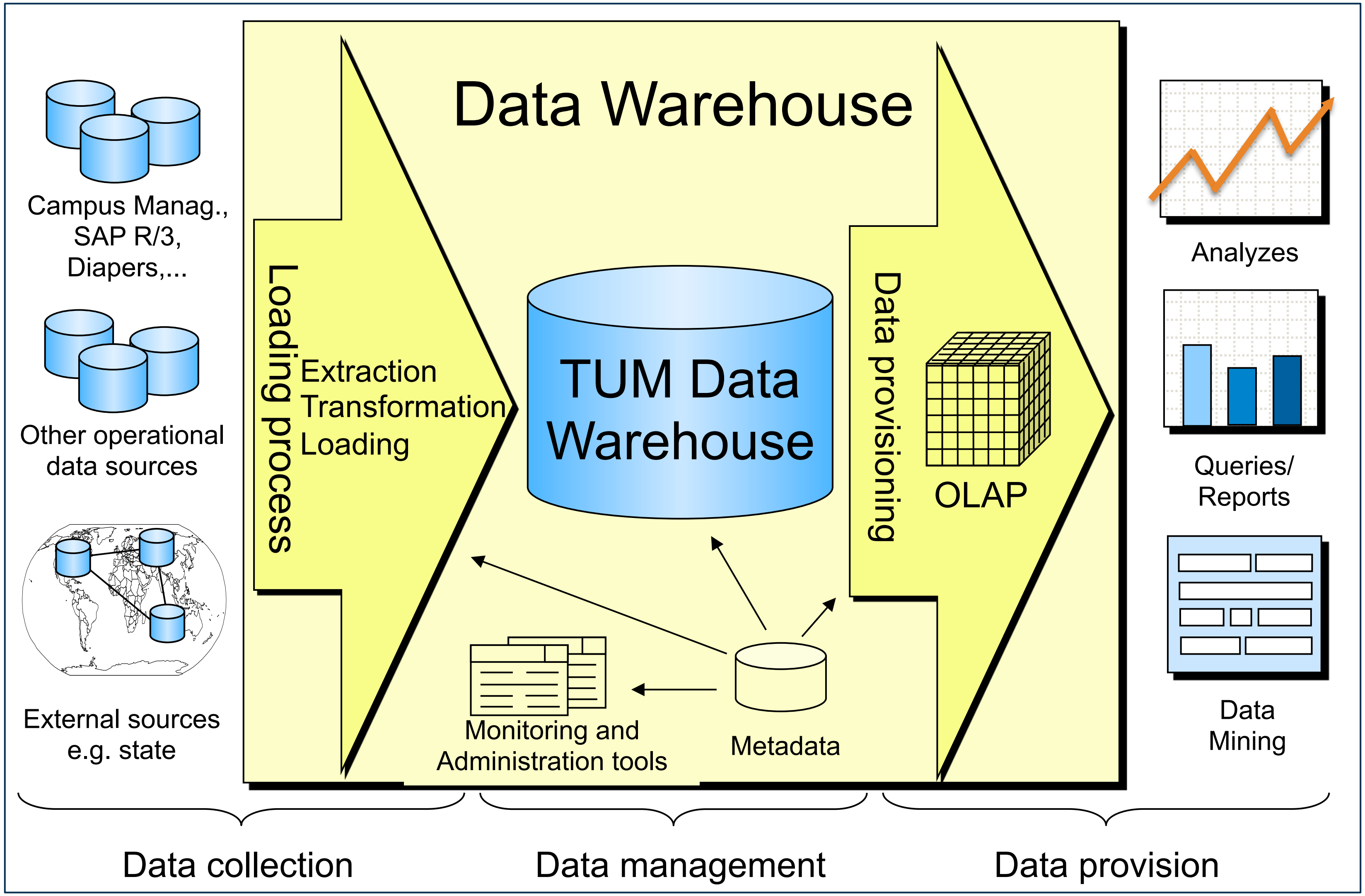
Data Warehouse

Corporate Design

CM@TUM



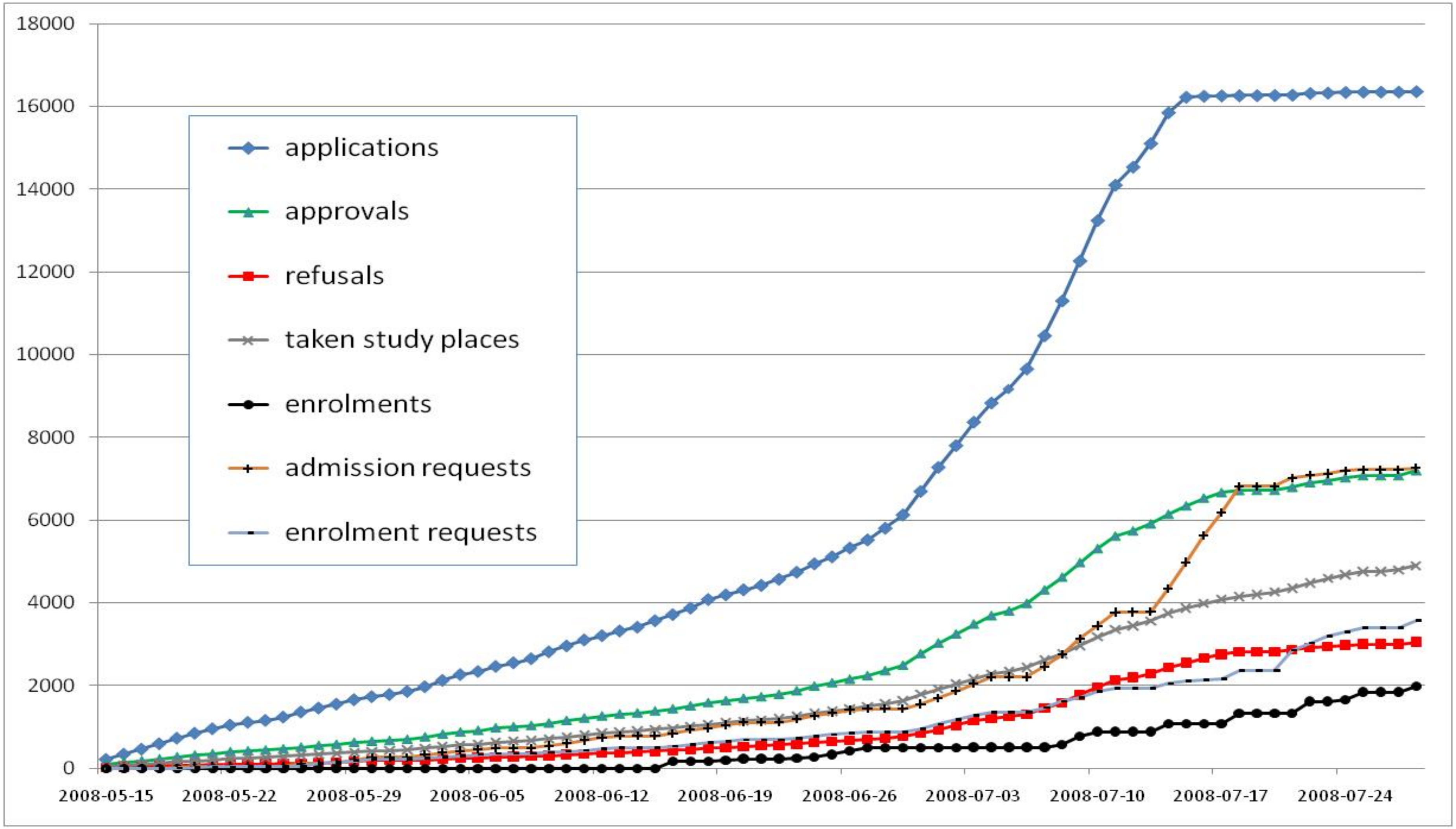
Data Warehouse: BW@TUM



Data sources at TUM

Domain	InfoProvider	Operatives System
Staff	Personnel Administration	SAP R/3 HR
	Organization Management	SAP R/3 HR
Student & Exam Data	Students	TUMonline
	Tests / Exams	TUMonline
	Applicants	TUMonline
Accounting	Financial Planing	SAP R/3 FI
	Funds Management	SAP R/3 FM
	Controlling	SAP R/3 CO
Integrationdomain	various	various

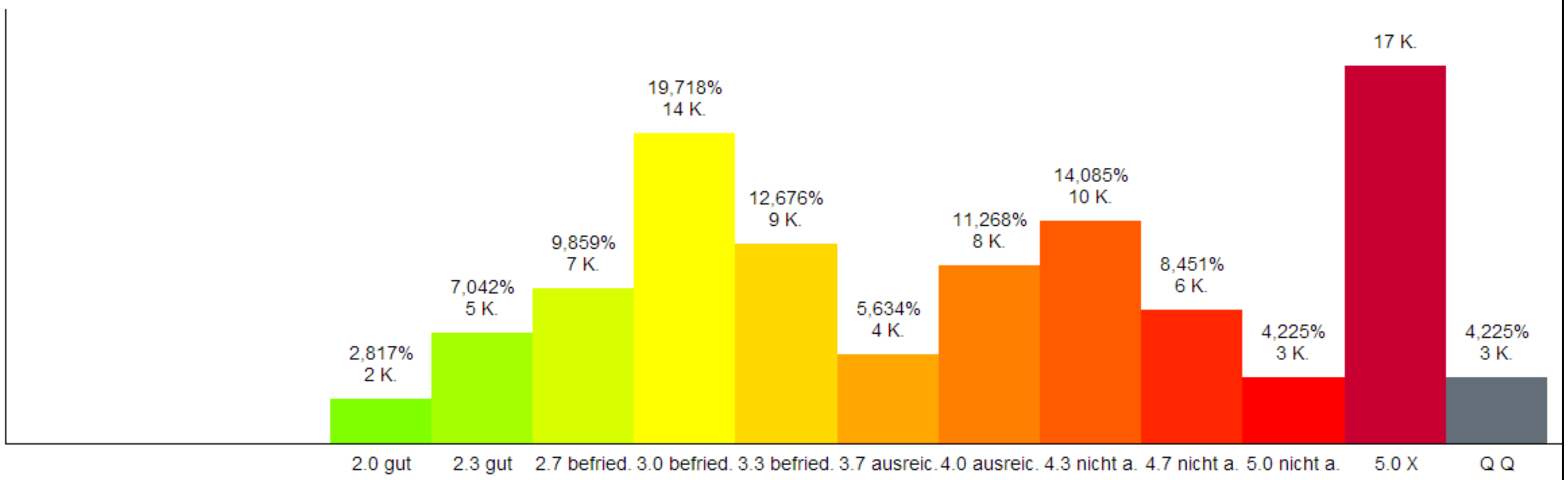
2008: Bachelor admission and enrolment statistic



Self-service rankings for students per test

Exam date	WZ0152 13S 4SWS L Technology and Utilization of Non-Wood Biogenic Resources on 02.10.2013 for Schieder
Registered:	88
Angetreten:	71
Not present:	17
Withdrawal with approved reasons:	3
Not valid/cheating:	0
Rejection:	0
Percent. of exams assessed as failed:	26,761 %
Average total:	3,516
Average (assessed as passed):	3,12

Grade distribution: Percent % / grade
K. = Number of candidates



How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

334 comments, 173 called-out + Comment Now + Follow Comments

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

Charles Duhigg outlines in the New York Times how Target tries to hook parents-to-be at that crucial moment before they turn into rampant — and loyal — buyers of all things pastel, plastic, and miniature. He talked to Target

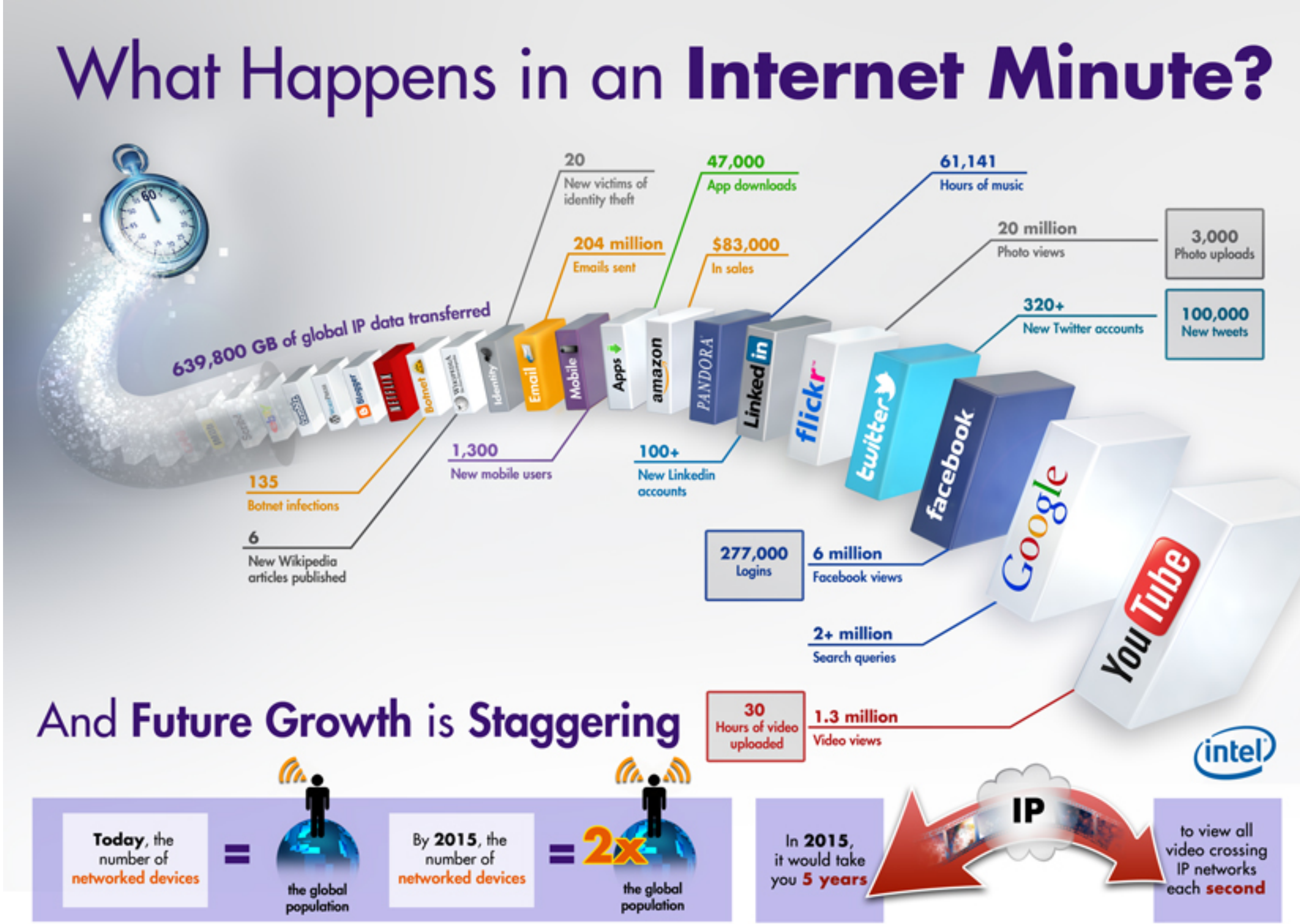
TARGET

> [Pole] ran test after test, analyzing the data, and before long some useful patterns emerged. Lotions, for example. Lots of people buy lotion, but one of Pole's colleagues noticed that women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium and zinc. Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signals they could be getting close to their delivery date.

Source: http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

New idea? No, e.g. former corner shops

Source: http://commons.wikimedia.org/wiki/File:Lebensmitteleinzelhandel_in_den_1950er.jpg

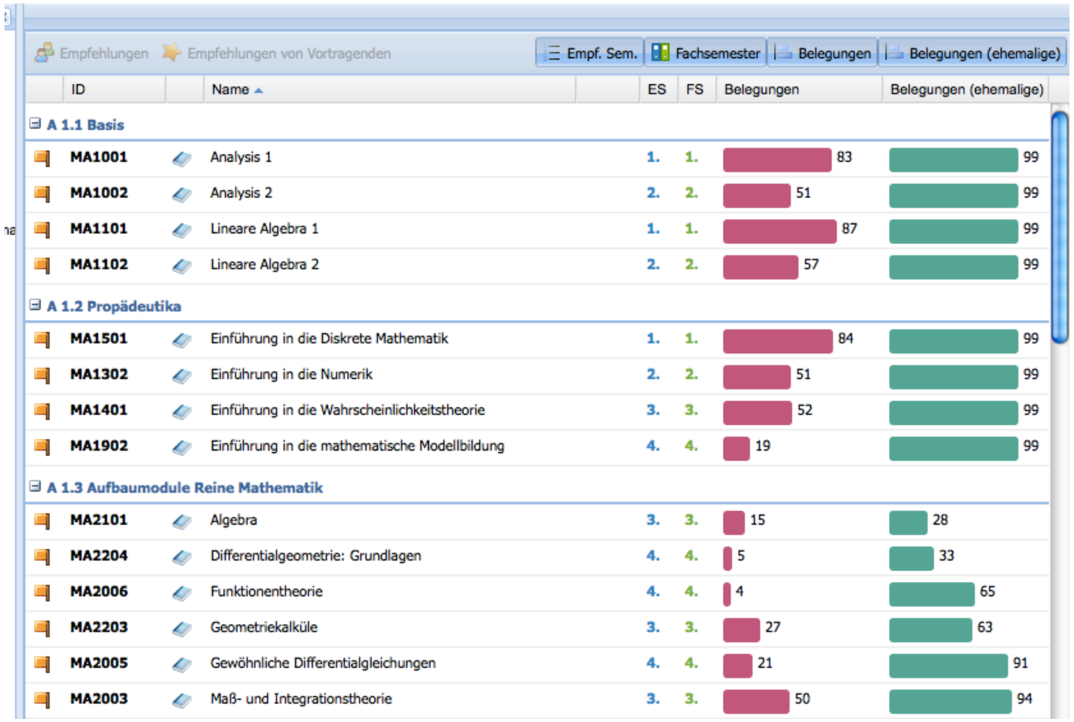


Source: Intel, www.intel.com

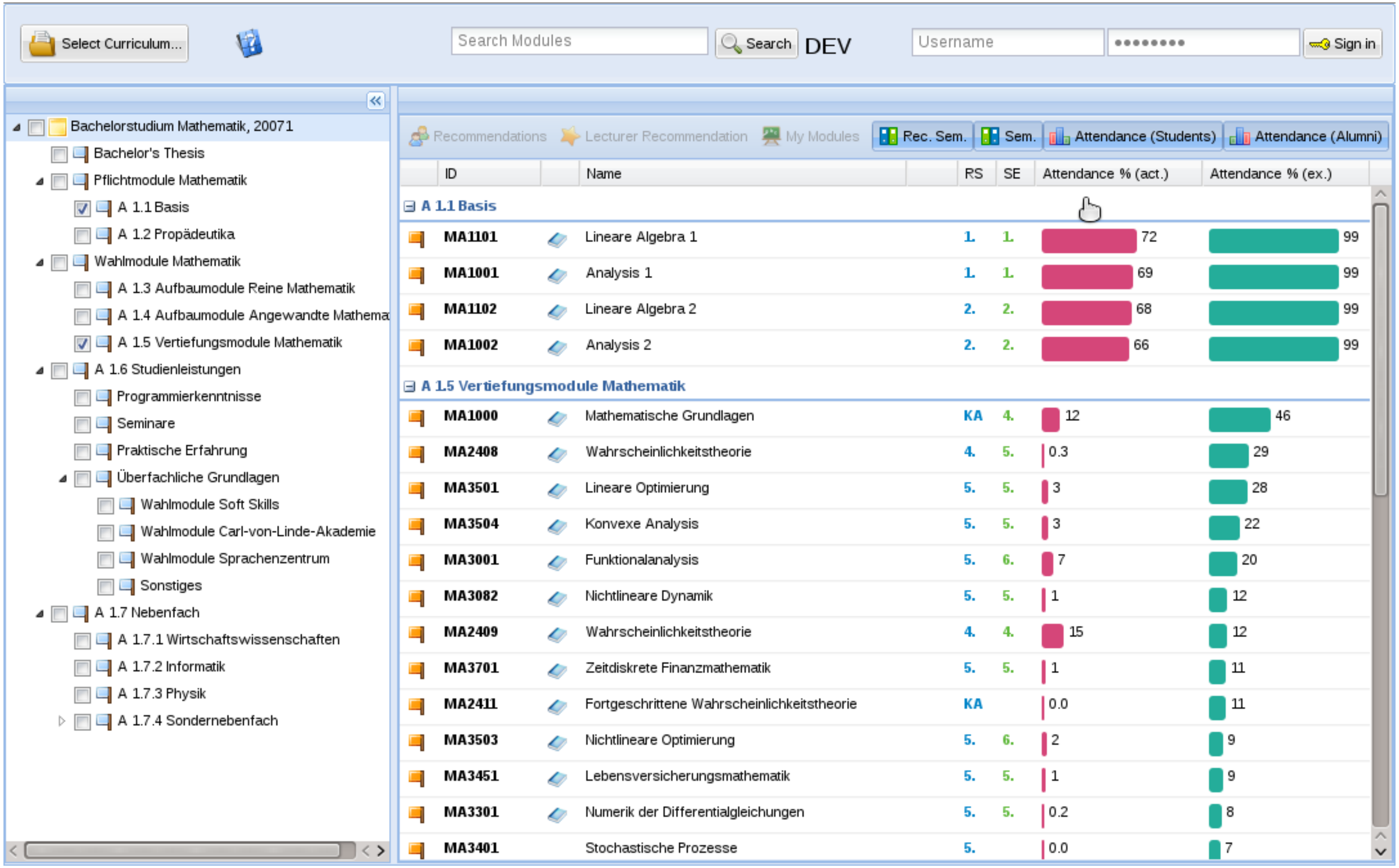
Academic & Learning Analytics

- Business Analytics for HEIs
- Goals, e.g. predicition of room allocation, study success, enrollment, …

Anzahl Belegung pro Wahlfach-modul	Anzahl Module MSc. Mech. Eng.	Anzahl Module BSc. Mech. Eng.
> 401	0	6
201 - 400	0	0
51-200	0	8
10-50	4	11
2-9	90	18
1	110	4



First experiences. Modul assignments (1/2)



Source: A. Baumann, TUM

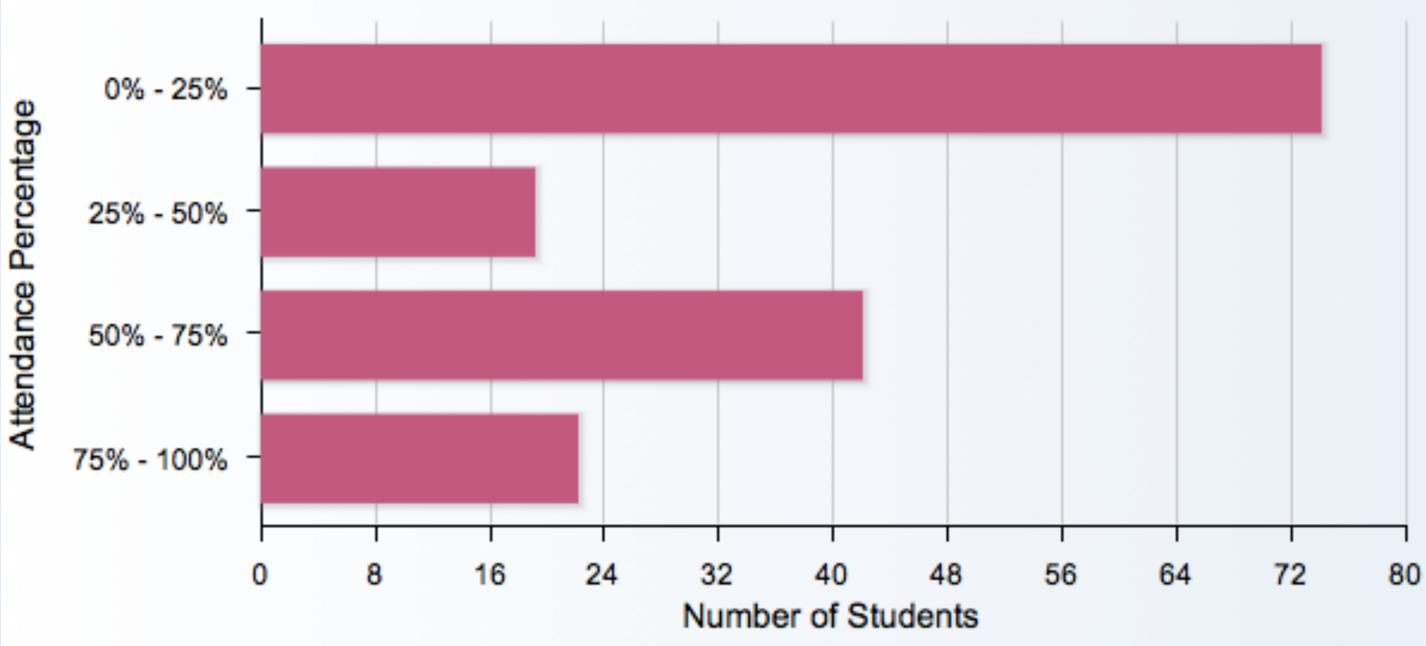
First experiences. Modul assignments (2/2)

Statistics

Module Attendance by Subjecttype

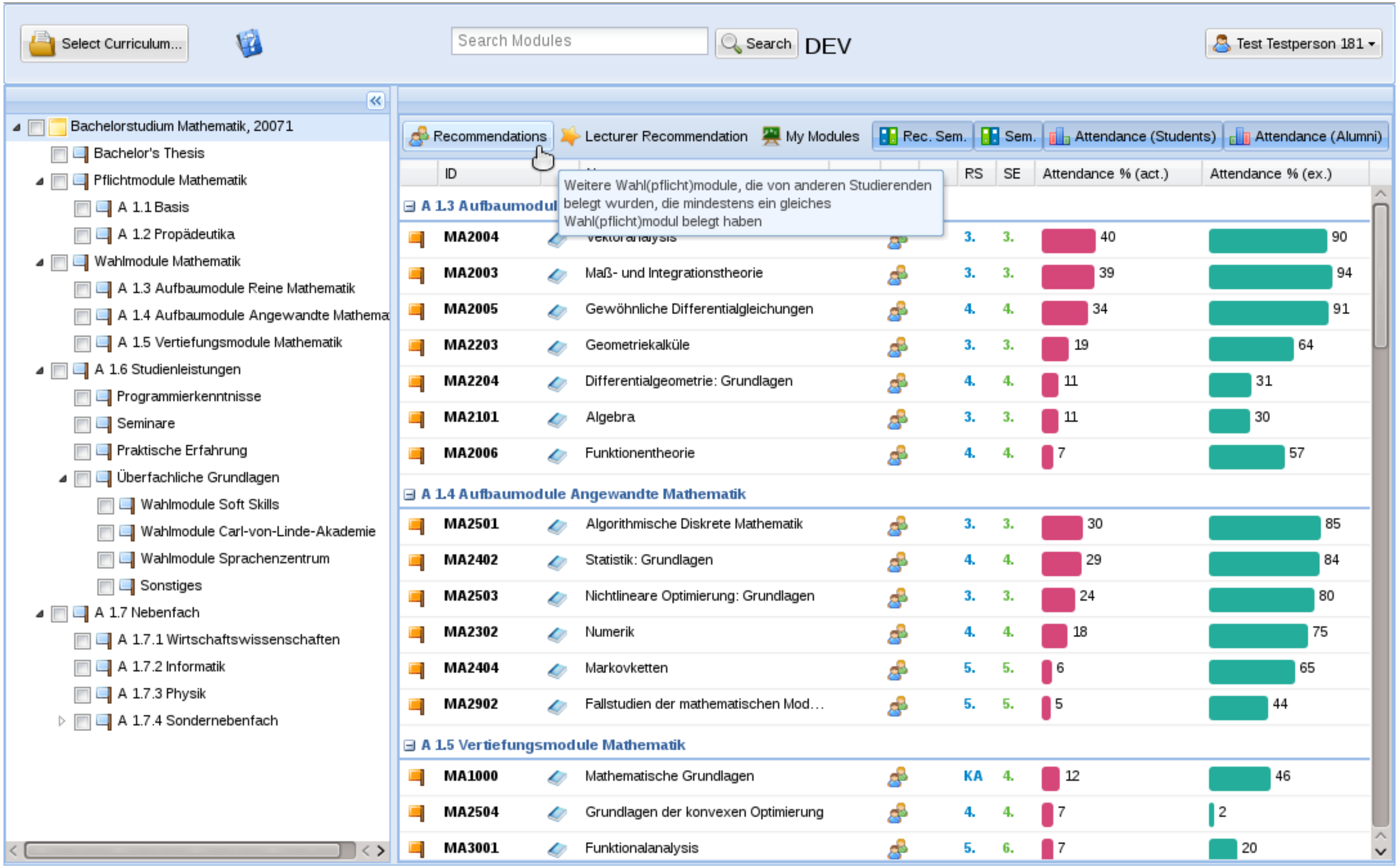
Subject Type	Number of ...	MIN (Attendance)	AVG (Attendance)	MAX (Attendance)
None	60	1	16	302
Compulsory subjects	12	18	325	518
Elective subjects	73	1	63	347

Attendance Percentage



Source: A. Baumann, TUM

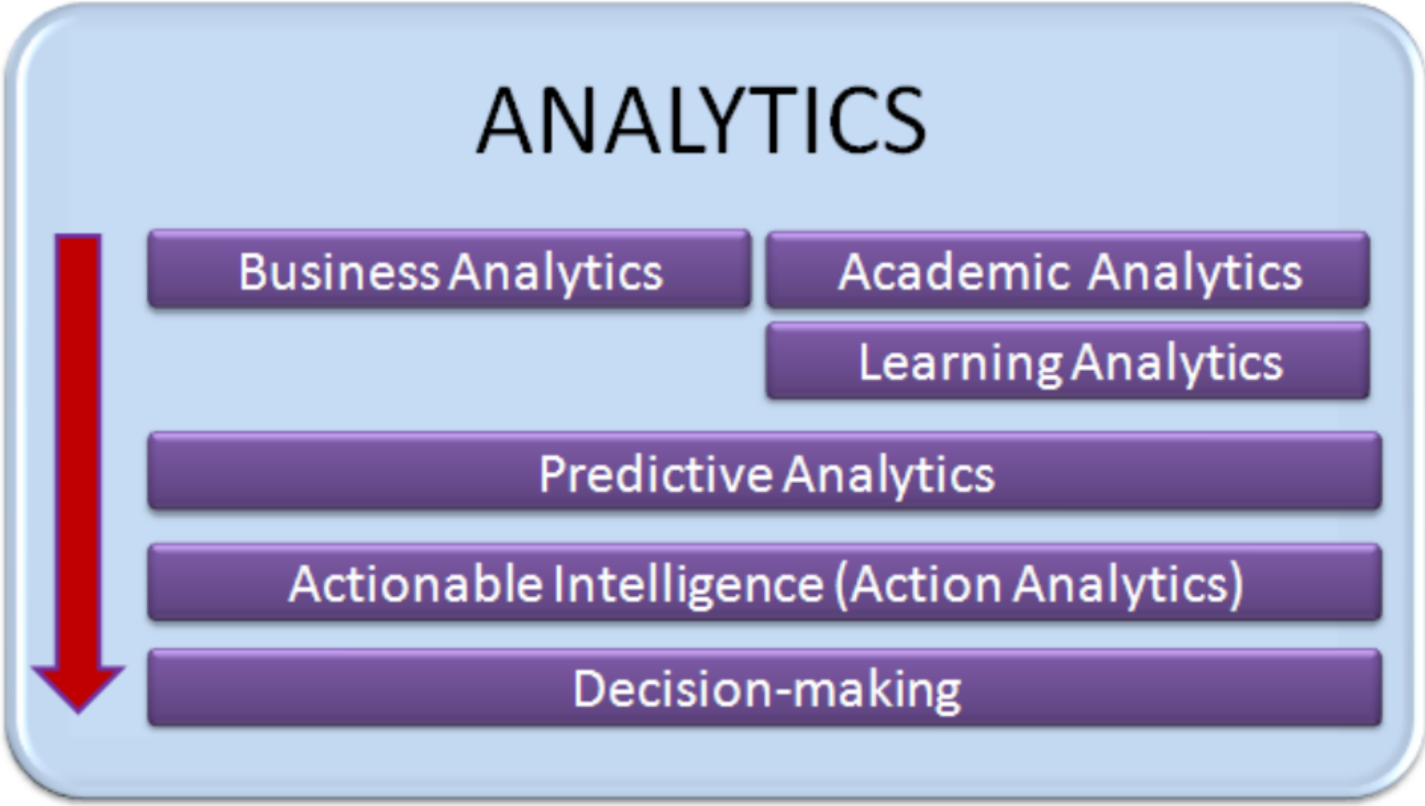
Frist experiences. Elective Modules



Source: A. Baumann, TUM

Conceptual Framework

Proposal by EDUCAUSE



Source: EDUCAUSE, http://net.educause.edu/ir/library/pdf/ELI3026.pdf

EDUCAUSE: Proposed Definitions

Term	Proposed Definition	Proposed Level of Focus	Sample Projects (see below for links)
Analytics	An overarching concept that is defined as data-driven decision making (from Ravishanker).	All levels	• M-Reports Dashboard • Learning and Career Outcomes
Academic Analytics	A process for providing higher education institutions with the data necessary to support operational and financial decision making (adapted from Goldstein and Katz).	Institution	• Effectiveness Sources Portal (ESP) • Sponsored Project Excellence Achieved through Redesign (SPEAR)
Learning Analytics	The use of analytic techniques to help target instructional, curricular, and support resources to support the achievement of specific learning goals (adapted from Bach).	Department/ learner	• Course Signals • Check My Activity
Predictive Analytics	An area of statistical analysis that deals with extracting information using various technologies to uncover relationships and patterns within large volumes of data that can be used to predict behavior and events (adapted from Eckersen).	All levels	• Student Success Plan • Student Readiness Inventory

Source: EDUCAUSE, http://net.educause.edu/ir/library/pdf/ELI3026.pdf

Purdue University’s Course Signals project

Examples & Projects (1/5)

Based on the dissertation of Dr. John Campbell a tool analyzes the learning behavior of students and can identify students who are not expected to successfully complete on the basis of their learning activity the course.

2009: first release

2010: „Users scored up to 26 % more As or Bs. Earned up to 12 % fewer Cs and up to 17 % fewer Ds & Fs”.

2011: 7k of 20k student used tool

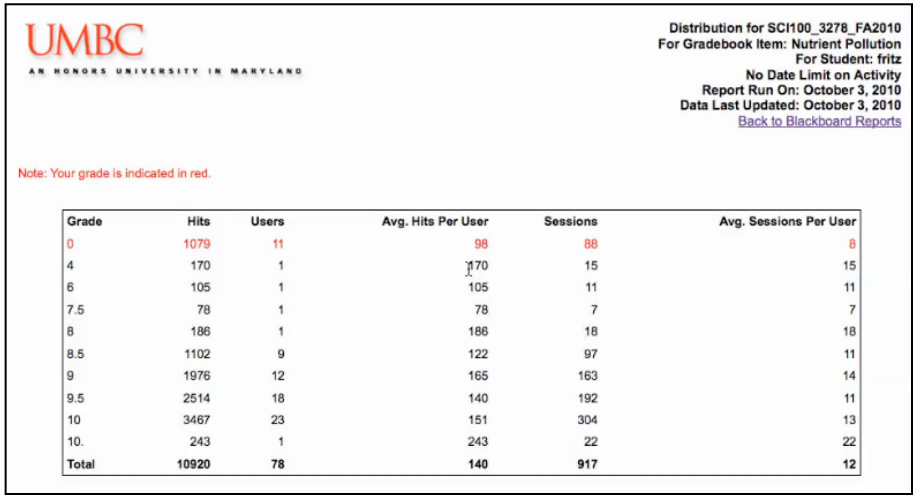
now: added value, e.g. Workshops, Consultation, Advisors, …

Further details: www.itap.purdue.edu/studio/signals/

University of Maryland, Baltimore County (UMBC)

Examples & Projects (2/5)

„Check My Activity (CMA)“ allows self-comparison between students with regard to their activity within the LMS and achieved course score.



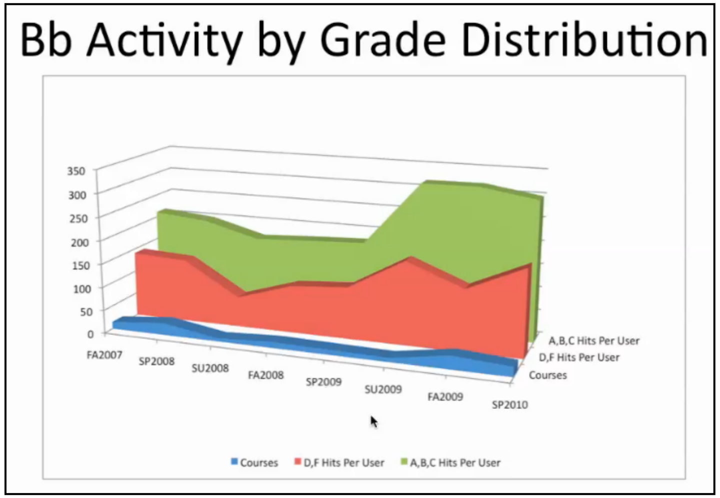
UMBC

AN HONORS UNIVERSITY IN MARYLAND

Distribution for SCI100_3278_FA2010
For Gradebook Item: Nutrient Pollution
For Student: fritz
No Date Limit on Activity
Report Run On: October 3, 2010
Data Last Updated: October 3, 2010
Back to Blackboard Reports

Note: Your grade is indicated in red.

Grade	Hits	Users	Avg. Hits Per User	Sessions	Avg. Sessions Per User
0	1079	11	98	88	8
4	170	1	170	15	15
6	105	1	105	11	11
7.5	78	1	78	7	7
8	186	1	186	18	18
8.5	1102	9	122	97	11
9	1976	12	165	163	14
9.5	2514	18	140	192	11
10	3467	23	151	304	13
10.	243	1	243	22	22
Total	10920	78	140	917	12

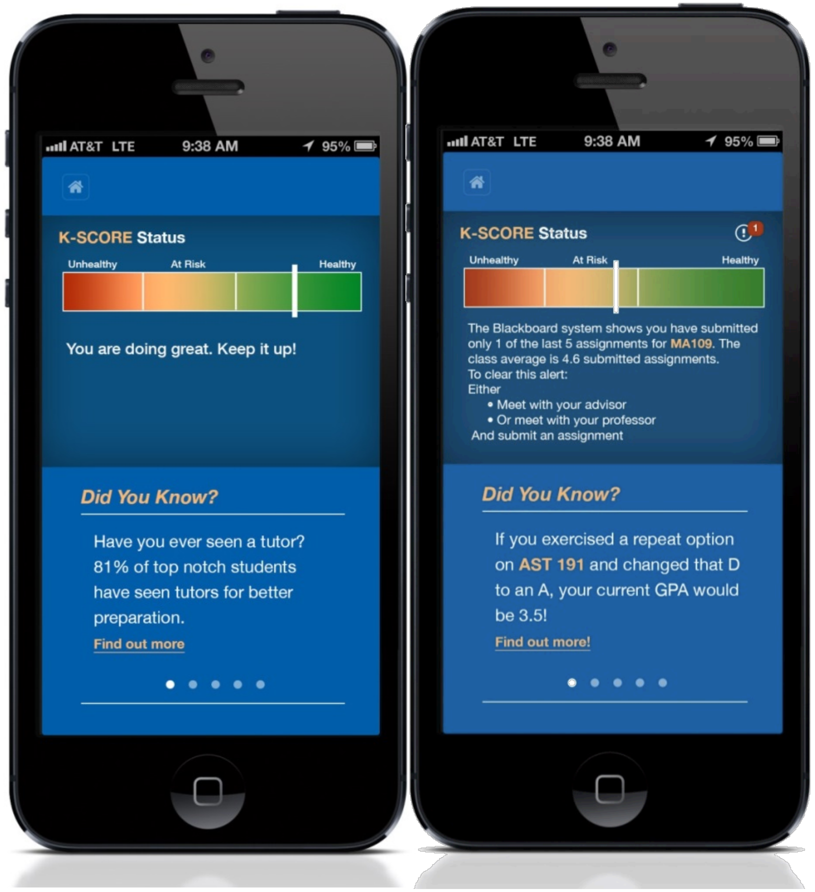


Report Code is open, see „Get The Code“

Further details: www.umbc.edu/blackboard/reports

University of Kentucky: Academic Health student app

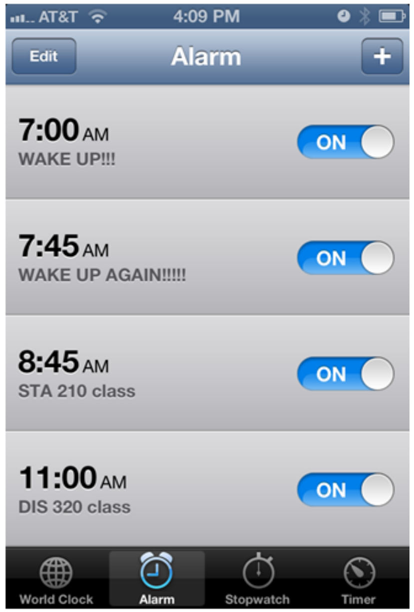
Examples & Projects (3/5)



UNIVERSITY OF KENTUCKY see blue.

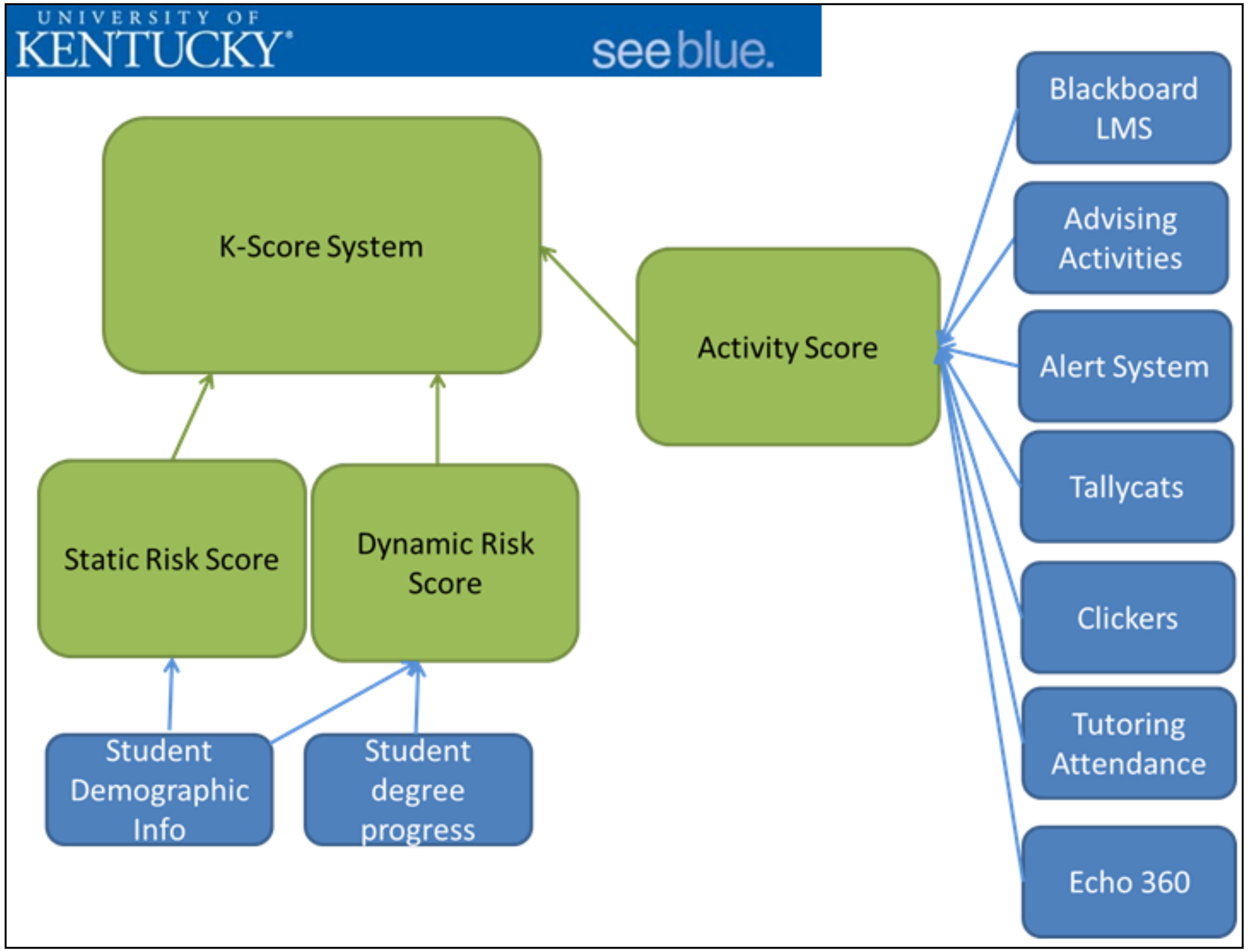
WAKE UP! GET TO CLASS!

- Who sets alarms for themselves?
- Why not automatically set alarms for students around their schedule?
- Why not have automated wake-up calls?
- Why not suggest wake up times based on class attendance?
- Why not consider manipulation of reminders as a form of engagement?
- Can we ascertain student prospective memory capability and personalize based on it?



Source:
http://www.educause.edu/annual-conference/2013/improving-student-success-using-groundbreaking-analytics-and-fast-data-improve-student-retention

University of Kentucky: Framework



Source:
http://www.educause.edu/annual-conference/2013/improving-student-success-using-groundbreaking-analytics-and-fast-data-improve-student-retention

Overview: EDUCAUSE review Academic Analytics

Examples & Projects (4/5)

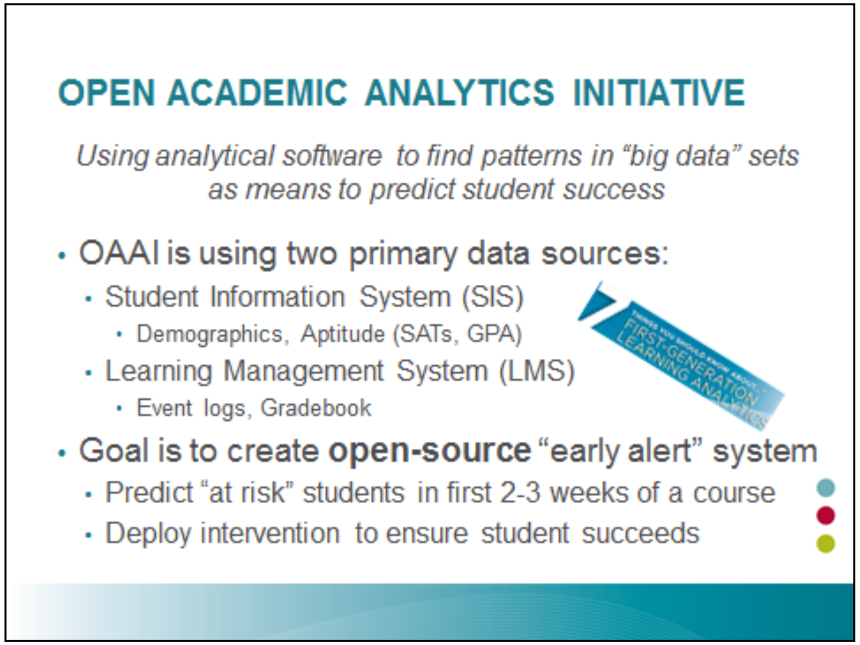
- Enrollment Predictive Modeling at Baylor University
- Predicting and Improving Student Retention at the University of Alabama
- Developing a Student Success Plan and Early Alert System at Sinclair Community College
- Connecting Resource Utilization, Risk Level, and Outcomes at Northern Arizona University

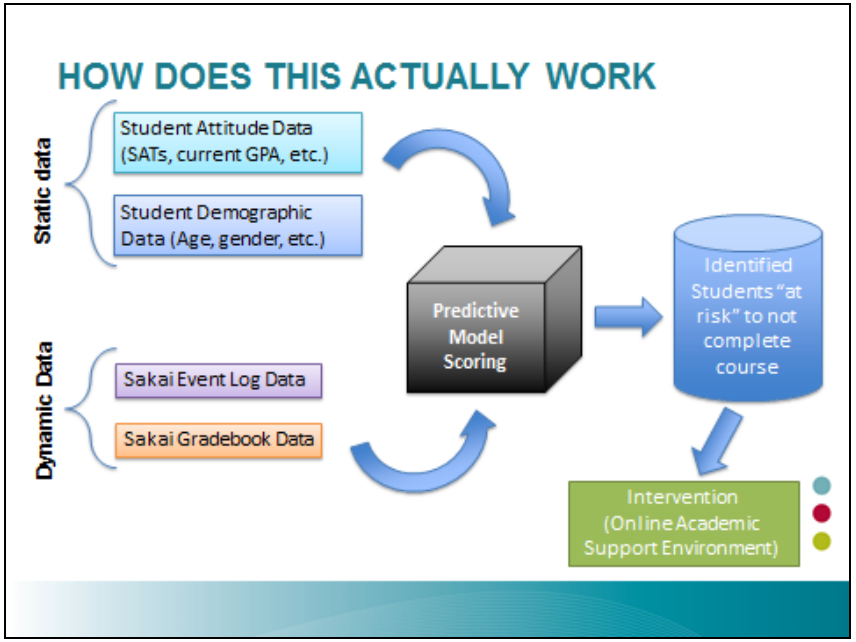
Further details: http://www.educause.edu/ero/article/academic-analytics-new-tool-new-era

Open Academic Analytics Initiative (OAAI)

Examples & Projects (5/5)

Goal: open-source "early alert" system, which predicts "at risk" students within the first 2-3 weeks of a course and provides assistance for them.





Further details: www.educause.edu/events/educause-learning-initiative-2012-annual-meeting/open-academic-analytics-initiative-leveraging-openness-improve-learne

Open Academic Analytics Initiative (OAAI)

Examples & Projects (5/5)

Finding #1: Imp
Our research shov ... nt i
To determine this ... ng t
the control group ... ard
also included stud
receive an interve
which had a signif

It is important to
between the two

	College	AAR run	# Students	Accuracy	FP Rate	Precision	Recall
Spring 2012	Savannah	AAR1	504	67.26%	35.36%	61.48%	70.54%
		AAR2	504	74.40%	32.50%	67.15%	83.04%
		AAR3	504	79.37%	18.21%	77.03%	76.34%
	Cerritos	AAR1	502	61.95%	43.69%	47.41%	72.32%
		AAR2	601	71.88%	27.49%	59.62%	70.78%
		AAR3	649	75.19%	25.12%	62.50%	75.76%
	Redwoods	AAR1	195	67.69%	40.48%	52.78%	82.61%
		AAR2	195	78.97%	13.49%	72.58%	65.22%
		AAR3	195	77.95%	14.29%	70.97%	63.77%

	College	AAR run	# Students	Accuracy	FP Rate	Precision	Recall
Fall 2012	Savannah	AAR1	425	68.47%	38.34%	58.19%	78.49%
		AAR2	425	72.59%	30.04%	65.17%	76.16%
		AAR3	425	73.41%	26.88%	65.13%	73.84%
	Cerritos	AAR1	502	65.38%	32.35%	49.49%	61.01%
		AAR2	601	70.75%	27.78%	55.96%	67.92%
		AAR3	649	73.98%	24.51%	60.11%	71.07%
	Redwoods	AAR1	182	83.63%	16.52%	71.21%	83.93%
		AAR2	182	83.82%	16.52%	72.06%	84.48%
		AAR3	182	85.63%	13.04%	76.56%	83.05%
	NCAT	AAR1	719	64.12%	31.25%	26.53%	45.45%
		AAR2	719	71.07%	24.83%	35.29%	54.55%
		AAR3	719	75.10%	20.14%	40.82%	55.94%

Table 3 - Prediction analysis on spring and fall pilot data

Table 1 - Correlations between course grades and CMS

Undergraduate CMS event frequencies		Course Grade: Marist Fall 2010 N=18968	Course Grade: Campbell (2007) N=27276
Sessions Opened	Correlation	0.147	(no values reported)
	Significance	0.000(**)	
	N	11195	
Content Viewed	Correlation	0.098	0.112
	Significance	0.000(**)	0.000(**)
	N	7651	19205
Discussions Read	Correlation	0.133	0.068
	Significance	0.000(**)	0.000(**)
	N	1552	7667
Discussions Posted	Correlation	0.233	0.061
	Significance	0.000(**)	0.000(**)
	N	1507	7292
Assign. Submitted	Correlation	0.146	0.163
	Significance	0.000(**)	0.000(**)
	N	3245	4309
Asssmts Submitted	Correlation	0.161	0.238
	Significance	0.000(**)	0.000(**)
	N	1423	4085

(**) Significant at the 0.01 level (2-tailed)
Marist data uses ratios over course mean instead of frequencies

First results s

See:
https://confluence.sakaiproject.org/download/attachments/75671025/OAAI%20Final%20Progress%20Report.pdf?version=1&modificationDate=1391705397000&api=v2

Talking about laws (1/2)

- EU Directive 95/46/EC on data protection
- EU Directive requires member states to achieve result by not dictating
- Terms and conditions

Personal data: any information concerning the personal or material circumstances of identified or identifiable natural person (concerned). Under personal data thus fall details of name, student number, degree, address, affiliations associations, email, etc. (http://www.bfdi.bund.de/bfdi_wiki/index.php/3_BDSG_Kommentar_Absatz_1_Beispiele)

Anonymized in the sense of the Bavarian Laws (quite similar to German & EU): Personal data will be considered anonymous if the data has been modified so that the reference to individuals cannot or only under extremely difficult conditions are restored (see also Article 4, Section 8 BayDSG, http://byds.juris.de/byds/009_1.1_DSG_BY_1993_Art4.html)

Talking about laws (2/2)

"*Online as soon as it happens*", ENISA, p. 28, Whitepaper of the European Network and Information Security Agency, ISBN-13: 978-92-9204-036-9, February 2010

Member States shall provide that personal data must be:

- ✓ Processed fairly and lawfully.
- ✓ Collected for specified, explicit and legitimate purposes and used accordingly.
- ✓ Appropriate and relevant in relation to the purpose for which they are processed.
- ✓ Accurate and kept up to date.
- ✓ Kept no longer then the time necessary for the purpose for which they are processed ([50]).

Personal data can be processed if:

- ✓ The data subject has been adequately informed and has given unambiguously his consent for the collection and further use of his data.
- ✓ Processing is necessary to perform a contract having as a party the data subject or to enter into a contract requested by the data subject.
- ✓ A legal obligation requires the processing of personal data.
- ✓ Processing data is necessary in order to ensure the essential interests of the data subject;
- ✓ Processing is necessary to perform tasks of public interests or carried out by an official authority.
- ✓ The data controller has a legitimate interest in processing the personal data of the data subject; this interest, however, has to be necessary balanced with the right to privacy of the data subject ([51]).

Questions to address & action items

- Root Questions vs. Research Approach
- Data sources
- Tools to use
- Anonymization of data
- Information about stored data
- Deleting stored data after X month/years
- Policy for data analytics?
- How gets access to data at which level?

Get in touch with data protection officer of your organization!

Exchange ideas, tools, results, approaches! -> pongratz@tum.de ☺