We live in a digital society...

...generating 2.5 quintillions bytes of data every day

source: “10 Key Marketing Trends For 2017” by IBM Marketing Cloud. 2016
What is a “quintillion”? 
A hint...
To speak 1 quintillion words, humanity would need to continuously talk for \(~29\) years.
Not just volume…

BIG DATA

VOLUME
DATA SIZE

VELOCITY
SPEED OF CHANGE

VARIETY
DIFFERENT FORMS OF DATA SOURCES

VERACITY
UNCERTAINTY OF DATA
Classics never die!

Integration of legacy data sources
Access to unstructured data
Integration with social media and behavioral data

https://sloanreview.mit.edu/article/variety-not-volume-is-driving-big-data-initiatives/
Why are data so important?

Designed system
Why are data so important?
Why are data so important?

Data = digital footprint of reality
“Data is the new oil. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, or chemicals to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”

- Clive Humby
Computer science

Application domains

Data Science

Math and Statistics Knowledge

Math

Danger zone!

Hacking Skills

Substantive Expertise

Traditional Research

Machine Learning
COMPUTATIONAL DATA SCIENCE

Computer science
(the science of computing)
at the service of data science
Wait a minute...
Do we need models at all?

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE
IT WAS a $3.5 billion question: was the crashing of two aeroplanes into New York's twin towers in September 2001 one event or two? One, many insurers are relieved to know. On May 3rd a jury ruled that Swiss Re, the world's second-largest reinsurer, which wrote about a quarter of the coverage for the World Trade Centre, was bound by a form that classed such attacks as a single occurrence. Last week the same jury had reached a similar verdict for several Lloyd's of London syndicates and seven other insurers. The loser was Larry Silverstein, the centre's leaseholder. He had argued that another form was valid, in the hope of claiming around $7 billion for two events. Now he may get only half that.
A $3.5 BILLION QUESTION
was the crashing of two aeroplanes into New York’s twin towers in September 2001 one event or two?

Similar verdict for several Lloyd’s of London syndicates and seven other insurers. The loser was Larry Silverstein, the centre's leaseholder. He had argued that another form was valid, in the hope of claiming around $7 billion for two events. Now he may get only half that.
Spurious correlations

Japanese passenger cars sold in the US correlates with Suicides by crashing of motor vehicle

Correlation: 93.57% ($r=0.935701$)

Suicides by crashing

Japanese cars sold

600 thousand cars
800 thousand cars
1000 thousand cars
1200 thousand cars

1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009

80 suicides
100 suicides
120 suicides
140 suicides

Our consultant has been mining data all day.
The results are quite shocking.

According to the data, sales are always highest when I do this...
“it’s up to us to become more savvy about the models that govern our lives.”
“it’s up to us to become more savvy about the models that govern our lives.”

We need computer scientists able to combine model-driven and data-driven approaches at best!
From reality to insights and back
Computational Data Science
@ unibz

Our master programme will provide you with the key knowledge and skills needed to develop next-generation information systems used to describe and manage data, discover new facts and relations in the data, make predictions, and give advice to decision makers.
Computational Data Science @ unibz

- **COMPUTER SCIENCE AND IT METHODS** for managing and engineering big and complex data
- **MATHEMATICAL AND STATISTICAL TOOLS** for data interpretation and analysis
- **ARTIFICIAL INTELLIGENCE** to extract insights from data and reason for and together with decision makers
- **HUMAN-COMPUTER INTERACTION** to put humans in the loop
Future career

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE
Rising jobs in 2022 (LinkedIn)

- Italy: robotic engineer, machine learning engineer, cloud architect, data engineer, sustainability manager, data management consultant

- Sweden: talent acquisition specialist, business development representative, data engineer, career counsellor, machine learning engineer, content coordinator

- US: vaccine specialist, diversity and inclusion manager, customer marketing manager, machine learning engineer, process development scientist, business development representative
Our job profiles

**DATA ANALYST**

*Role*
Collects, processes and performs statistical data analyses

*Mindset*
Intuitive data junkie with high “figure-it-out” quotient

**DATA ENGINEER**

*Role*
Develops, constructs, tests and maintains architectures (such as databases and large-scale processing systems)

*Mindset*
All-purpose everyman

**DATA ARCHITECT**

*Role:*
Creates blueprints for data management systems to integrate, centralize, protect and maintain data sources

*Mindset:*
Inquiring ninja with a love for data architecture design patterns

---

**Languages**
- R, Python, HTML, Javascript, C/C++, SQL

**Skills & Talents**
- Spreadsheet tools (e.g. Excel)
- Database systems (SQL and NO SQL based)
- Communication & visualization
- Math, Stats, Machine Learning

**Languages**
- SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C++, Perl

**Skills & Talents**
- Database systems (SQL & NO SQL based)
- Data modeling & ETL tools
- Data APIs
- Data warehousing solutions

**Languages**
- SQL, XML, Hive, Pig, Spark

**Skills & Talents**
- Data warehousing solutions
- In-depth knowledge of database architecture
- Extraction Transformation and Load (ETL), spreadsheet and BI tools
- Data modeling
- Systems development
Initial directions...
Bird-Eye View of the Course

1st SEMESTER
- Data management technologies
- Math and statistics
- Programming and data visualisation
- Information systems design

2nd SEMESTER
- Artificial intelligence
- Machine learning
- Semantic technologies

3rd SEMESTER
- Data curation and integration
- Capstone project

4th SEMESTER
- Advanced English for scientific communication
- Final thesis

Advanced statistics
Recommender systems
Deep learning
Information retrieval

Real-time big data processing
Systems security

Data and process modelling
Algorithms
Human-computer interaction
Enterprise digital transformation
Two curricula

DATA ANALYTICS

• mathematics, statistics, and machine learning for data interpretation and analysis
• advanced IT applications to retrieve and explore data, extract insights from data, and visualise the obtained results

DATA MANAGEMENT

• computer science and artificial intelligence for the management of complex, heterogeneous data
• advanced information systems for describing, integrating, and processing data, and for supporting the execution of business and decision processes.

student choices + soft-skills + capstone project
• You choose curriculum at enrolment time (including “no curriculum”)
• You can change within the 1st year
• Curricula interact quite a lot
• You can personalise your studies and get your unique mix of competencies
• Balance between foundations and practice
• Capstone project + possible internship
Capstone project

• Application of the studied techniques to a real data set coming from a concrete application domain

• Ongoing/future collaborations with local territory (unibz, research centres, institutions, companies)

• Plethora of application domains: business information systems, travel and tourism, bioinformatics, healthcare, internet of things, industry 4.0, automation, agriculture, economics, …
Course structure

- cv-specific (30) 25%
- common (46) 38%
- configurable (24) 20%
- thesis (20) 17%
- cv-specific (30) 25%
- common (46) 38%
- configurable (24) 20%
- thesis (20) 17%
How to configure your path

Part 1/2: Optional courses

• Can be selected within a list of optional courses

• The list contains…
  
  • some courses from the other curriculum (cross-fertilization)
  
  • pure optional courses (activated year-by-year)

• Check the semester: considering the correspondence of 1st/3rd and 2nd/4th semesters!
How to configure your path

Part 2/2: Free choice credits

- Any lecture at the CS faculty, including
  - mandatory and optional courses from both curricula!!!
  - courses taken from our BSc (to fill some gaps in your cv) - if not already covered in your previous studies

- Lectures at the other faculties
  - Lectures at other Euregio universities (Trento, Innsbruck) + whatever Italian institution

- Internships (using 6-12 credits)

- 2x capstone (using 6 credits)

- Some choices: automatically accepted
  Others: to be approved by the course council
## Who is doing what

<table>
<thead>
<tr>
<th>Department</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advisory service</td>
<td>general info, foreign students, disabled students, accommodation</td>
</tr>
<tr>
<td>Student secretariat</td>
<td>enrolment, tuition fees</td>
</tr>
<tr>
<td>Language center</td>
<td>language courses, enrolment to tests</td>
</tr>
<tr>
<td>Ufficio per diritto allo studio universitario</td>
<td>grants, accommodation in student dormitories</td>
</tr>
<tr>
<td>sh.asus</td>
<td>general information, support in filling out forms for grants, peer-to-peer support</td>
</tr>
<tr>
<td>Faculty of CS</td>
<td>administration, academic matters</td>
</tr>
</tbody>
</table>
Contact points

FEDERICA CUMER
Problem solver
federica.cumer@unibz.it

MARCO MONTALI
Degree course director
montali@inf.unibz.it

ANTON DIGNÖS
2nd member of the degree committee
dignoes@inf.unibz.it
Where do I find this stuff?

• MSc web page

• Student cockpit
  • In particular the “knowledge” area
  • The “info for current students” section of the unibz portal redirects to specific sections in cockpit
When and where

• Timetable on the unibz portal - personalised one available in cockpit

• Cockpit also for enrolling in the exams (check the deadlines)

• Room code: C2.01

Area  Floor  Room#
Real life tends to interfere...

Don’t be afraid to ask
• us
• people around you (we have some local students!)
• student representative

Remember that there are rules...
Working students

• We have minimised the courses with mandatory participation
• Carefully check the CPF of courses
• Get in touch with the lecturers
• Ask your study plan advisor
Why here?
You are at the right place!

• Very good teacher/student ratio

• Very international environment

• One of the top universities in Italy

• Small, but with worldwide recognition in key areas related to data science
Let’s put this in numbers

Worldwide…
- CS faculty: 201-250 worldwide

In Italy…
VQR (evaluation of research quality):
- 4th CS department in Italy

Censis Italian ranking
- 1st small non-public university, with best evaluation overall

8 of our professors among “top 2% scientists for long career impact” (Stanford study on normalised citation metrics)
Closer to research

- Especially during the capstone and then the thesis, you can get involved in our research

- Three main research centres at the faculty
  - IDSE: Information and Database Systems Engineering
  - KRDB: Knowledge and Data
  - SwSE: Software and Systems Engineering
Research lines/groups

- CORE: Conceptual and Cognitive Modelling Research Group
- FDT: Foundations of Database Technologies
- In2Data: Intelligent Integration and Access to Data
- PRISM: PRocess-aware Information Systems Management
- CoM: Computational Mathematics
- CoVIP: Computer Vision and Image Processing Group
- DBS: Database Systems Group
- DSPRS: Decision Support, Personalization and Recommender Systems
- HCC: Human Centred Computing
- SwSE: Software and Systems Engineering
Smart Data Factory

Questions?