

unibz

MASTER IN COMPUTATIONAL DATA SCIENCE

MARCO MONTALI
Freshers day, 2022

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...



10¹⁸



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

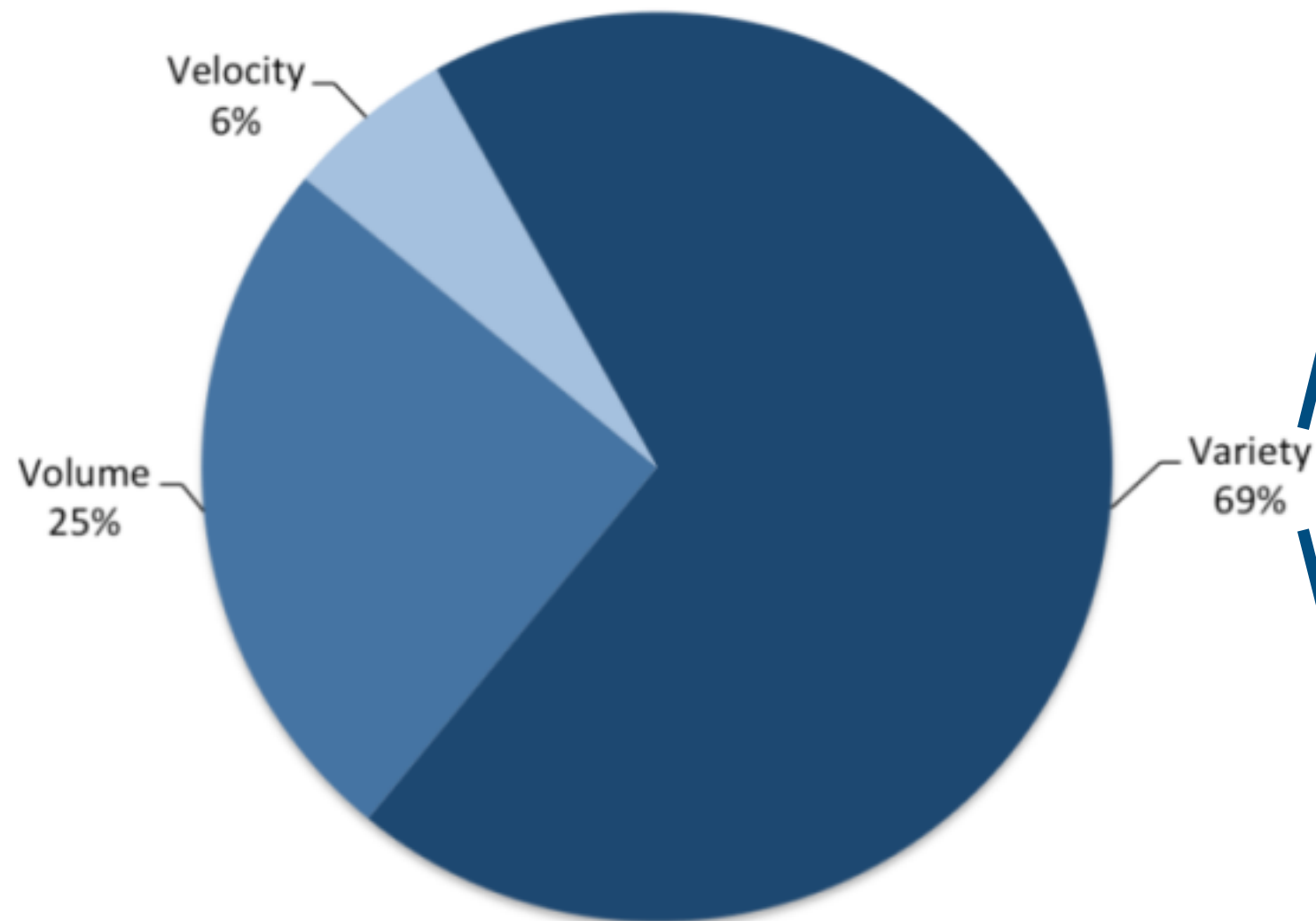
VARIETY

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

Why are data so important?

Designed
system



Why are data so important?

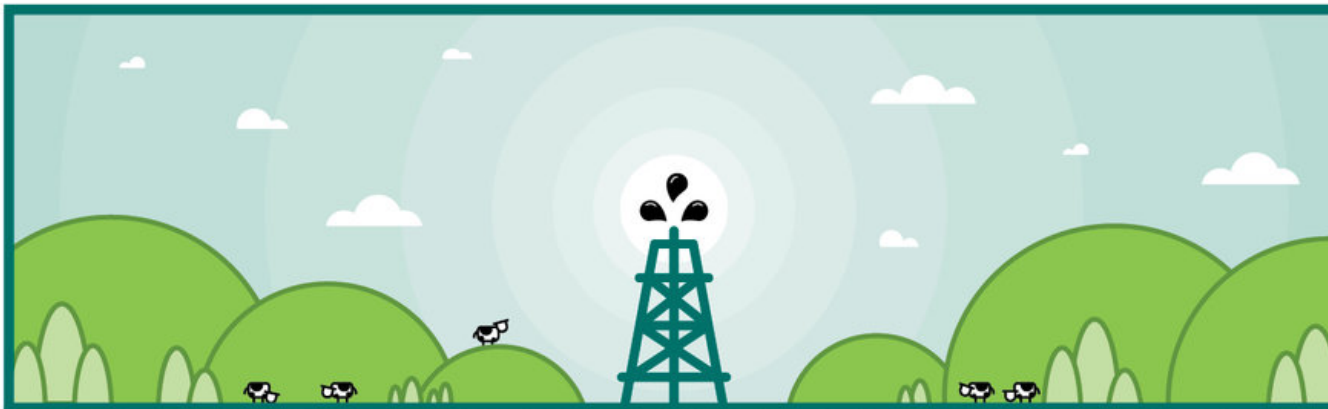


Used
system

Why are data so important?



Data = digital footprint of reality



DATA IS THE NEW OIL

"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

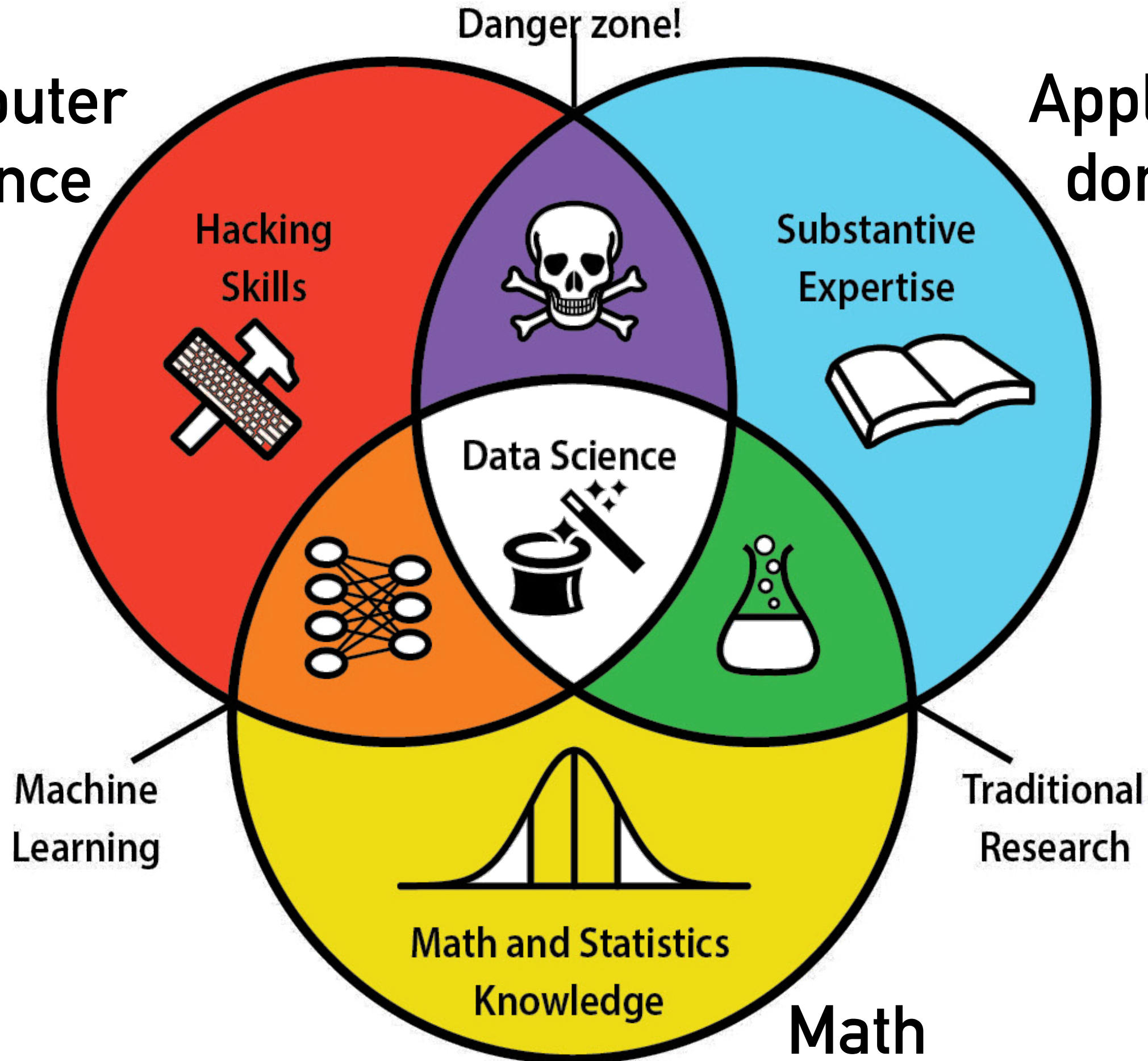


"Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, or chemicals to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."

- Clive Humby

**Computer
science**

**Application
domains**



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

CHRIS ANDERSON MAGAZINE 06.23.08 12:00 PM

SHARE



SHARE
75



TWEET



PIN
6



COMMENT



EMAIL

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE



Bad forms

After a rancorous trial, relief for many insurers of the twin towers

May 6th 2004



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.

AP



Bad forms

The
Economist

After a rancorous trial, relief for many insurers of the twin towers

May 6th 2004



Like 25

Tweet

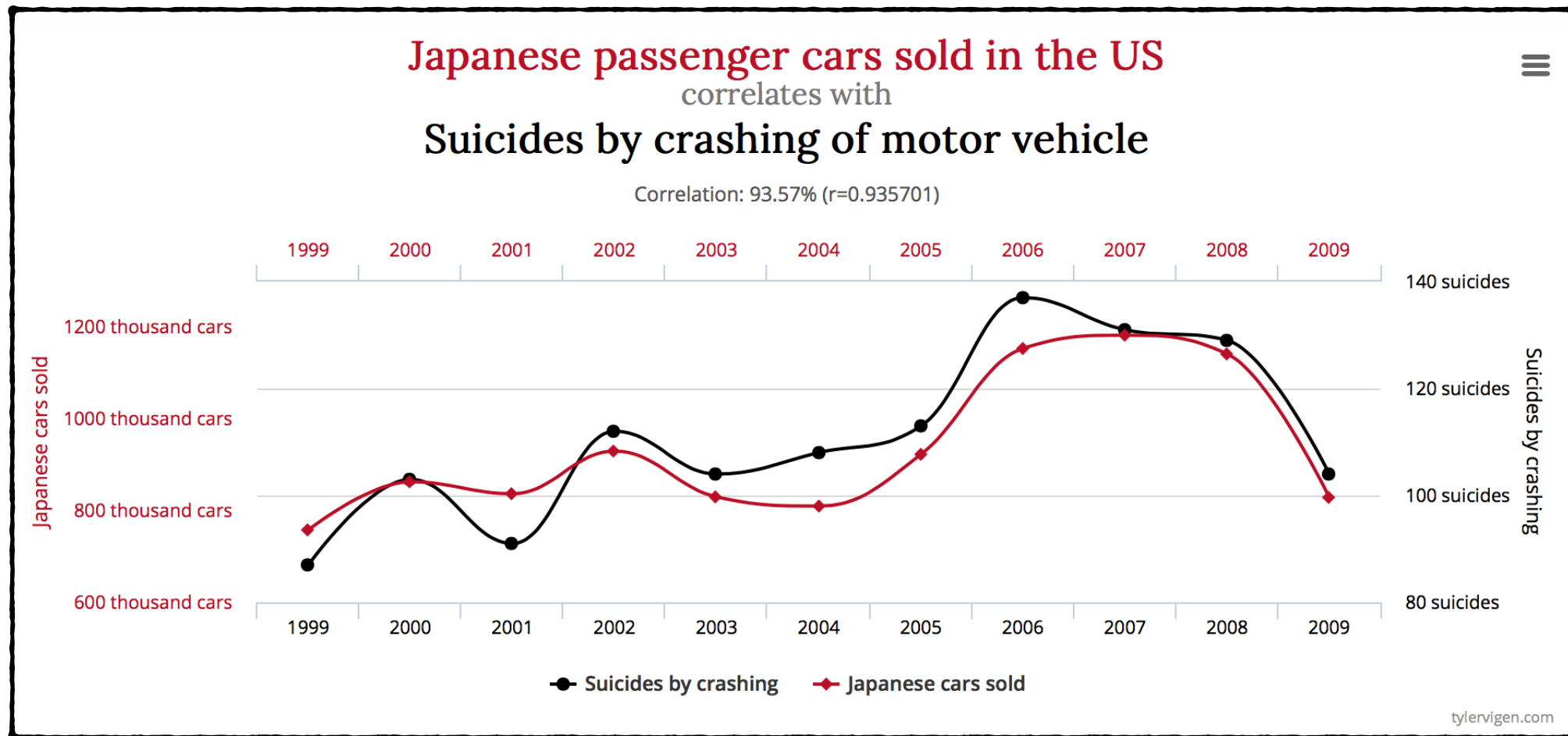
A \$3.5 BILLION QUESTION

was the crashing of two aeroplanes into New York's twin towers in September 2001 **one event or two?**

Similar verdict for several Lloyds 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



WEAPONS OF MATH DESTRUCTION



HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY

CATHY O'NEIL

**“it’s up to us to
become more savvy
about the models
that govern our
lives.”**

WEAPONS OF MATH DESTRUCTION



HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY

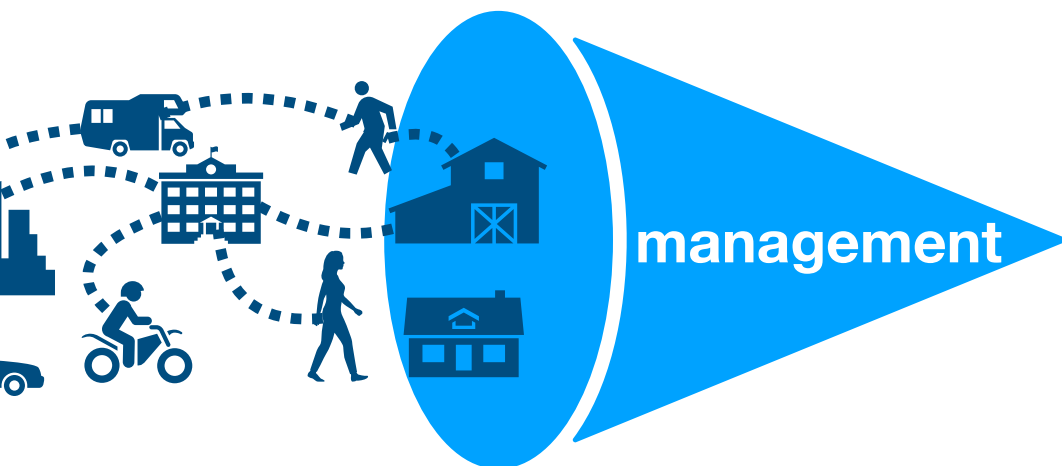
CATHY O'NEIL

**“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



reality

AGGIORNAMENTO 04/06/2020 ORE 17.00										
Regione	POSITIVI AL nCoV				DIMESSI/ GUARITI	DECEDUTI	CASI TOTALI	INCREMENTO CASI TOTALI (rispetto al giorno precedente)	TAMPONI	CASI TESTATI
	Ricoverati (per sistemi)	Terapia intensiva	Isolamento domiciliare	Totale attualmente positivi						
Lombardia	2.954	125	17.145	20.224	23.101	18.205	89.536	+ 84	780.887	460.658
Piemonte	808	43	1.705	4.556	22.292	1.910	30.758	+ 24	331.241	214.551
Emilia Romagna	320	49	2.310	2.688	21.016	8.154	27.860	+ 18	341.710	204.614
Veneto	97	2	1.220	1.319	15.955	1.914	19.188	+ 4	708.475	344.495
Toscana	65	24	794	883	9.180	1.058	10.122	+ 1	262.762	186.622
Liguria	149	6	239	394	7.899	1.479	9.722	+ 21	111.587	61.688
Lazio	488	51	2.215	2.754	4.260	752	7.764	+ 11	265.380	214.051
Marche	52	5	1.246	1.303	4.447	88	6.788	+ 3	110.877	67.429
Campania	180	5	665	850	3.572	420	4.822	+ 2	210.001	102.305
Puglia	131	9	827	967	5.012	514	4.503	+ 4	125.049	83.870
Trento	14	1	192	207	3.762	484	4.433	0	51.631	49.920
Sicilia	57	6	816	879	2.282	276	3.447	0	157.868	134.035
Friuli V.G.	14	2	177	213	2.790	110	3.279	+ 3	134.942	81.854
Abruzzo	109	3	192	204	2.116	413	3.355	+ 3	79.720	54.527
Bolzano	7	2	103	112	2.184	292	2.598	0	68.513	52.135
Umbria	15	2	14	31	1.164	26	1.431	0	79.394	52.229
Sardegna	12	1	85	108	1.118	131	1.357	0	58.556	50.865
Valle d'Aosta	9	0	1	10	1.034	143	1.187	0	15.512	12.156
Calabria	19	0	83	102	253	87	1.158	0	73.872	29.813
Molise	0	2	123	125	288	21	436	0	15.241	15.241
Basilicata	3	0	17	20	352	27	399	0	31.096	30.706
TOTALE	5.503	318	32.569	38.429	264.859	28.666	244.913	+ 177	4.049.544	2.524.788



insights

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

Harvard
Business
Review



DATA

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 'DATA DETECTIVE'

Role

Collects, processes and performs statistical data analyses

Mindset

Intuitive data junkie with high "figure-it-out" quotient



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

DATA ENGINEER SOFTWARE ENGINEERS BY TRADE

Role

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

Mindset

All-purpose everyman



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

DATA ARCHITECT THE CONTEMPORARY DATA MODELLER

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

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



HIRED BY

VISA Coca-Cola logitech



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
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

DATA ANALYTICS

DATA MANAGEMENT

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

Course structure



Data
Analytics

Data
Management

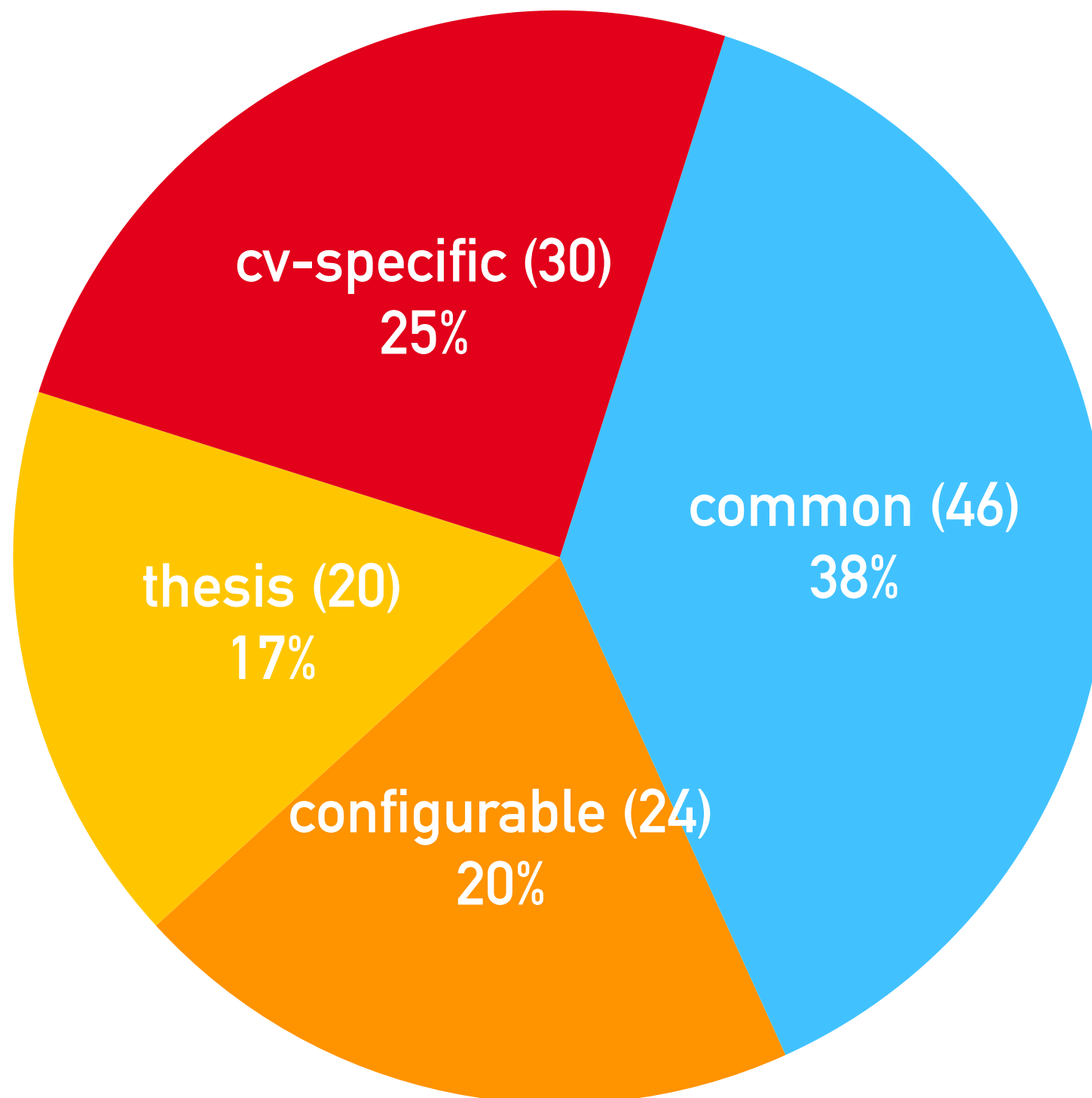


- You choose curriculum at enrolment time (including “no curriculum”)
- You can change within the 1st year
- Curricula interact quite a lot
- You can personalise your studies and get your unique mix of competencies
- Balance between foundations and practice
- Capstone project + possible internship

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



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 facultiesLectures 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

Advisory service	general info, foreign students, disabled students, accommodation
Student secretariat	enrolment, tuition fees
Language center	language courses, enrolment to tests
Ufficio per diritto allo studio universitario	grants, accommodation in student dormitories
sh.asus	general information, support in filling out forms for grants, peer-to-peer support
Faculty of CS	administration, academic matters

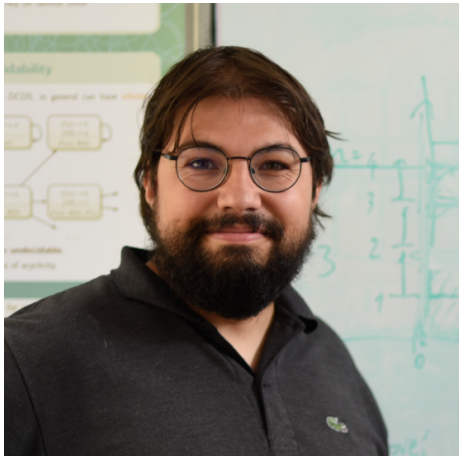
Contact points



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

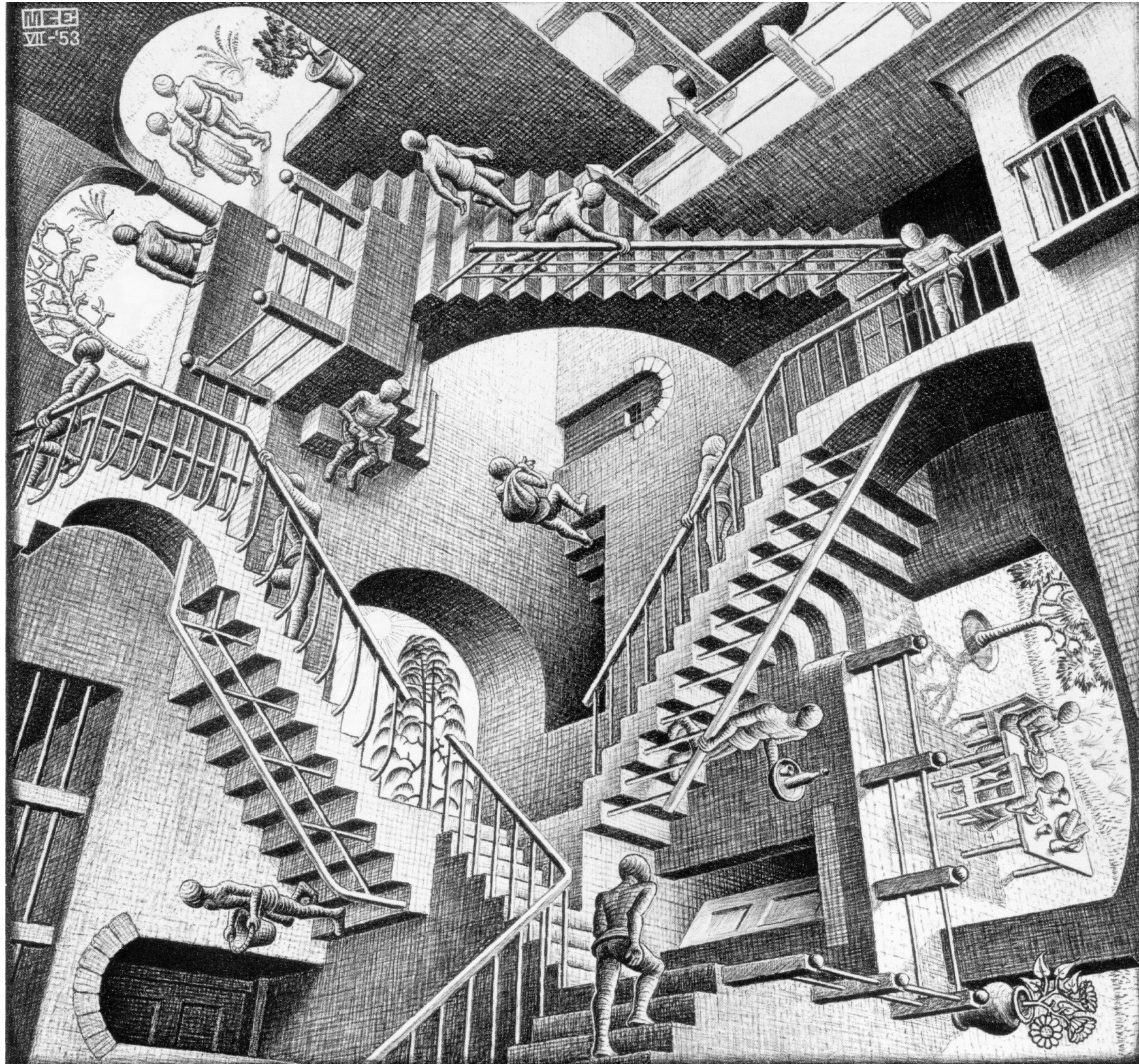
- **CORE: Conceptual and Cognitive Modelling Research Group**
- **FDT: Foundations of Database Technologies**
- **In2Data: Intelligent Integration and Access to Data**
- **PRISM: PRocess-aware Information Systems Management**
- **CoM: Computational Mathematics**
- **CoVIP: Computer Vision and Image Processing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

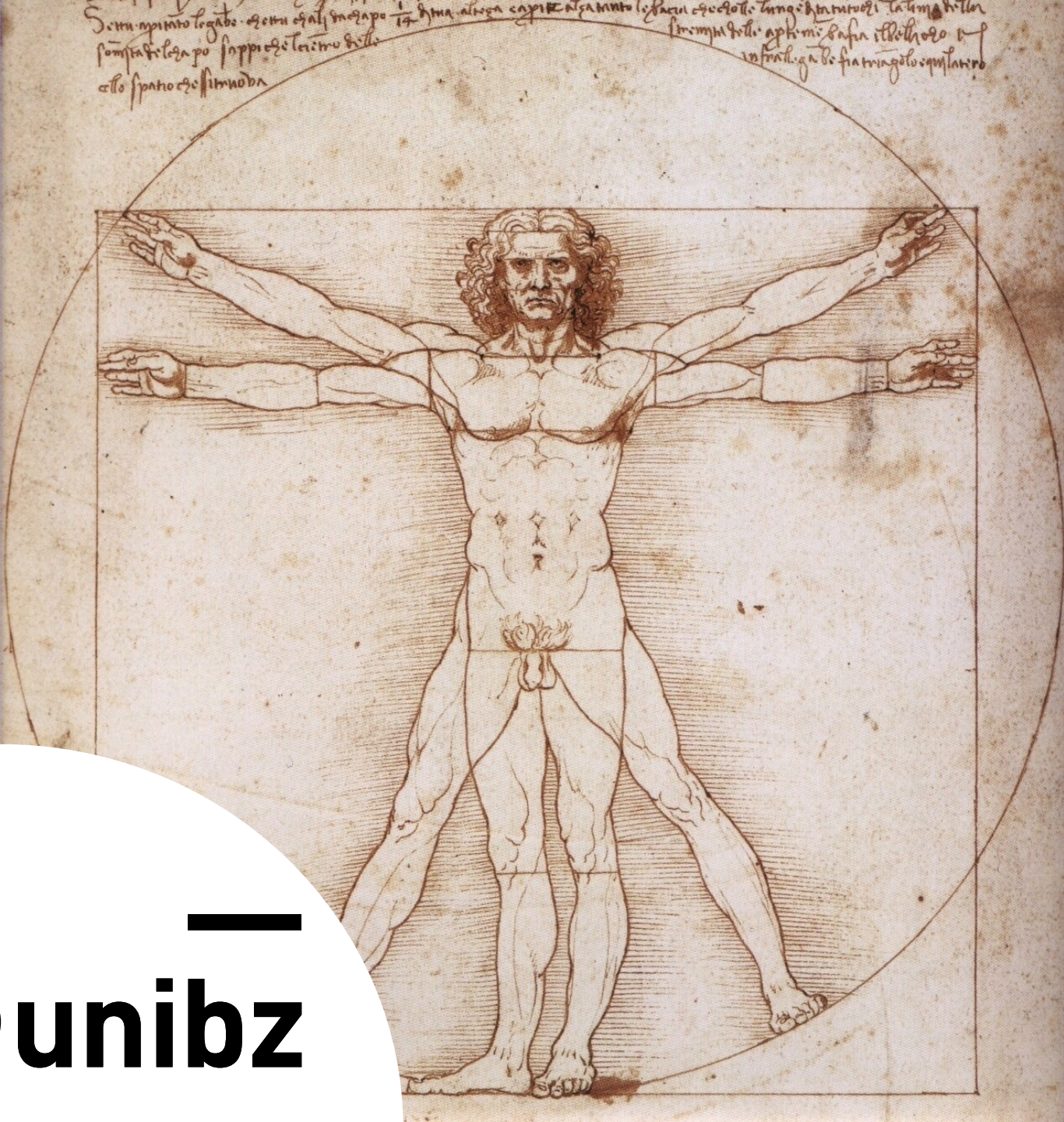
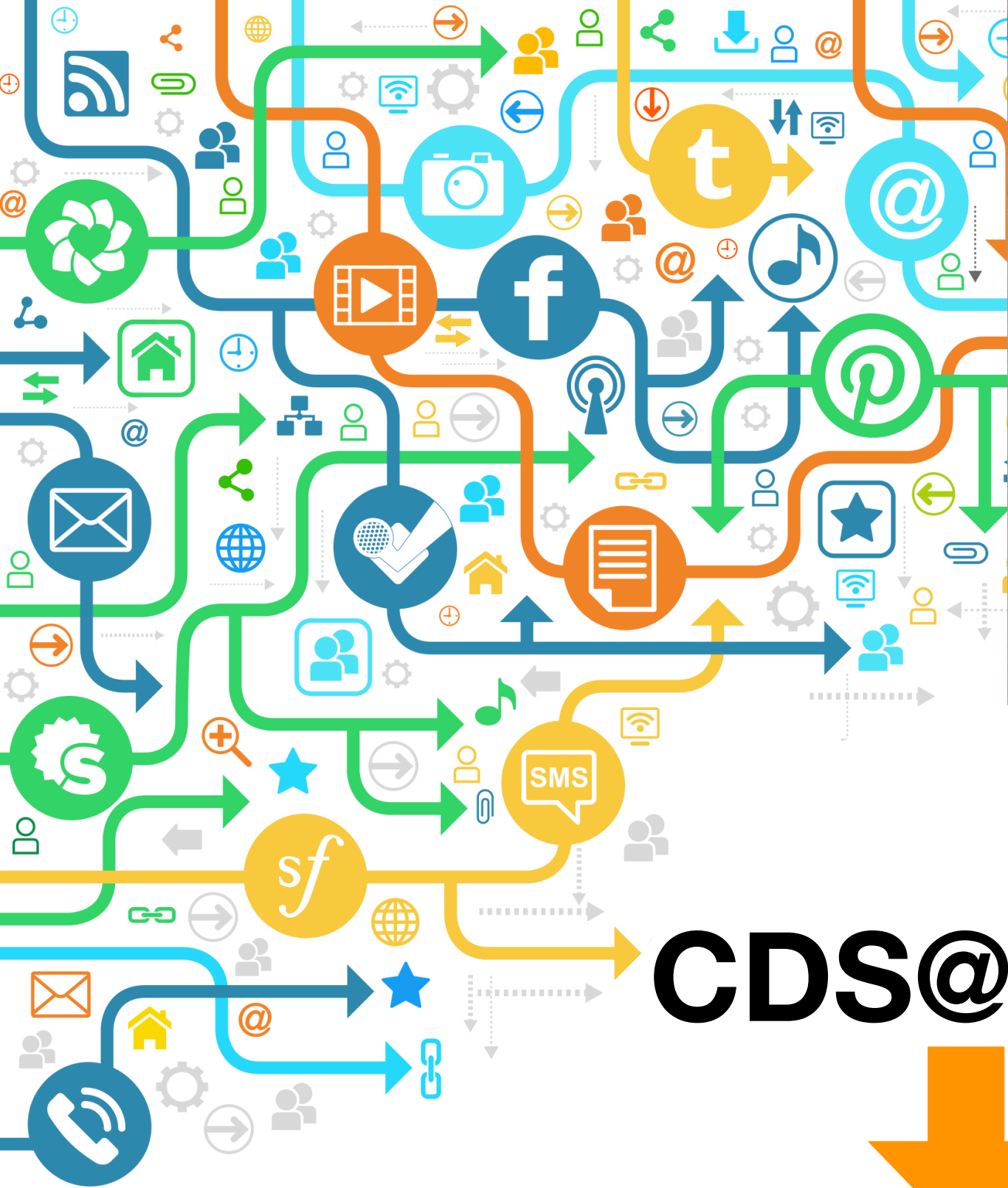
Smart Data Factory



<https://noi.bz.it/en/infrastructure/labs/smart-data-factory>

Questions?





CDS@unibz

