Zürcher Hochschule
für Angewandte Wissenschaften

Colloquium of University of Mannheim's School of Business Informatics & Mathematics

CHOCOLATE FLAVOURED DATA SCIENCE

datalab
dlab.zhaw.ch

Zürcher Fachhochschule

Thilo Stadelmann
28.04.2014

UNIVERSITY OF MANNHEIM

Zurich University
of Applied Sciences

ZHAW Zurich University of Applied Sciences – School of Engineering

Switzerland’s biggest fully-featured university of applied sciences

- >10’000 students
- >2’600 employees
- >1’000 (associate) professors

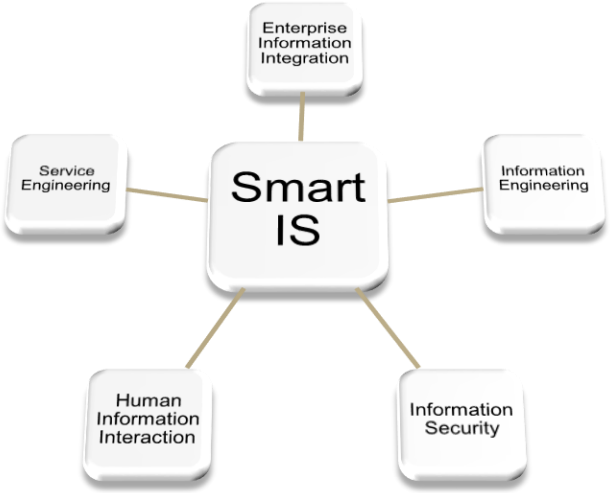
School of Engineering emanates from «Technikum Winterthur» (est. 1874)

Zurich University
of Applied Sciences

InIT Institute of Applied Information Technology

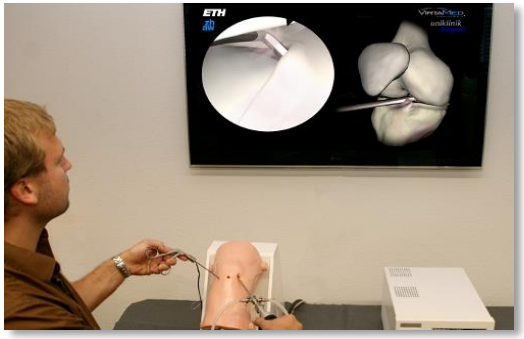
Smart Information Systems since 2005

- 5 Focus areas, one vision: Smart IS
- Undergraduate, graduate & continuing eductation in computer science
- Associated labs: Cloud Computing Lab, Data Science Lab, Visual Computing Lab, Services Lab, …



Facts & Figures

- >70 employees, >35 (associated) professors
- >6 MCHF business volume in 2013



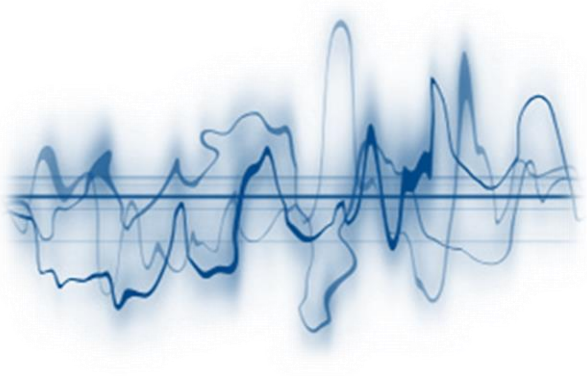
Information Engineering Research Group

Zurich University of Applied Sciences

zh aw

Combining structured and unstructured data analysis

- Information Retrieval and Data Warehousing
- Integration of heterogeneous and unstructured data
- Topic- and Trend-Detection (finding w/o search)
- Data- and Text-Mining
- Machine Learning and Artificial Intelligence
- Benchmarking of search engines
- Data Science within ZHAW Datalab



expert-match

promise

squirro

Information Engineering

Information Retrieval	Data Warehousing
Data Mining	Machine Learning

A Personal Story

- Fascinated by AI
- Studied computer science
- Applied ML & IR during Ph.D.
- Used DWH & DM professionally

A Personal Story

- Fascinated by AI
- Studied computer science
- Applied ML & IR during Ph.D.
- Used DWH & DM professionally

- Difficult to briefly explain professional interests

➔ Excited about term «Data Scientist»

A Personal Story

- Fascinated by AI
- Studied computer science
- Applied ML & IR during Ph.D.
- Used DWH & DM professionally

- Difficult to briefly explain professional interests

➔ Excited about term «Data Scientist»

Agenda: 1. Definition → 2. Projects → 3. State of the Union

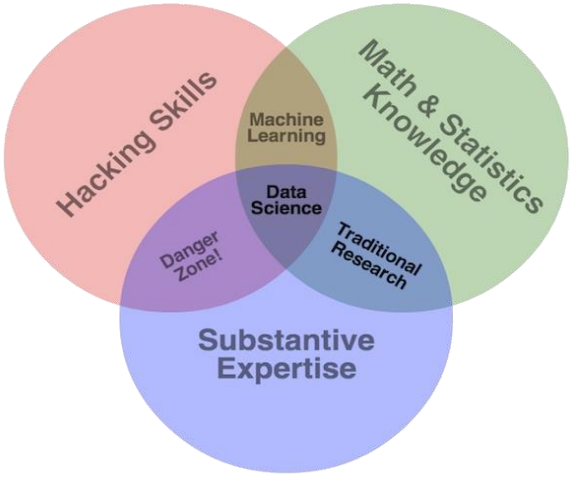
Zurich University
of Applied Sciences

What is a Data Scientist, what is Data Science?

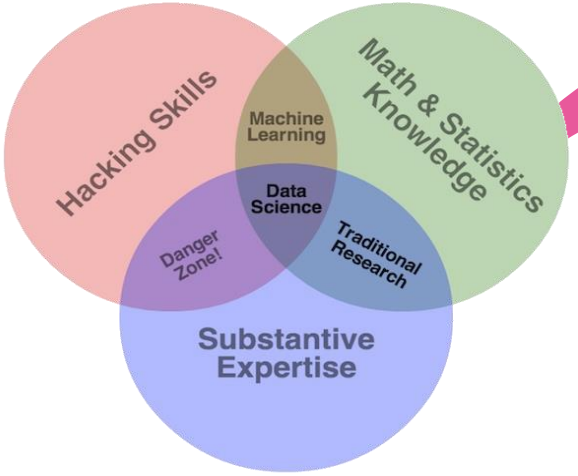
Definition & Classification

Zurich University
of Applied Sciences

Data Science?



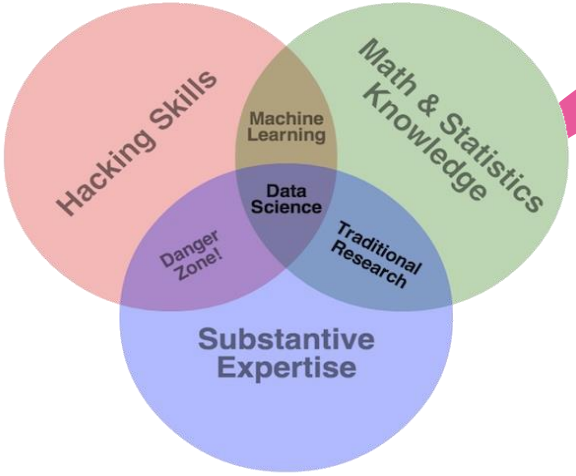
Data Science?



Whence?

amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

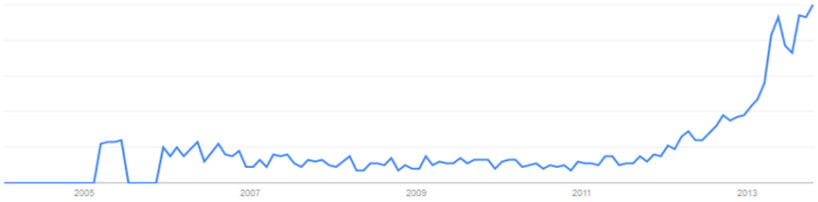
Data Science?



Whence?

amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

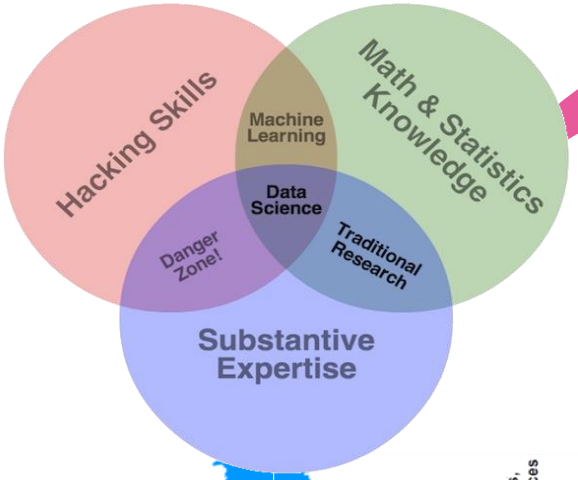
Why?



Zurich University of Applied Sciences

zh aw

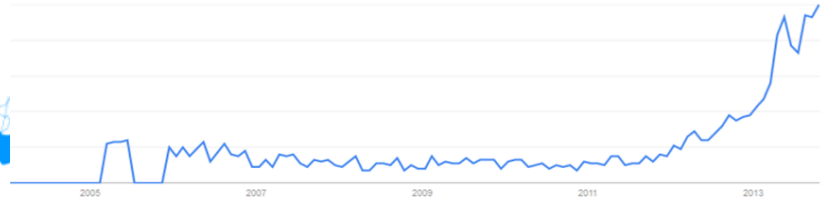
Data Science?



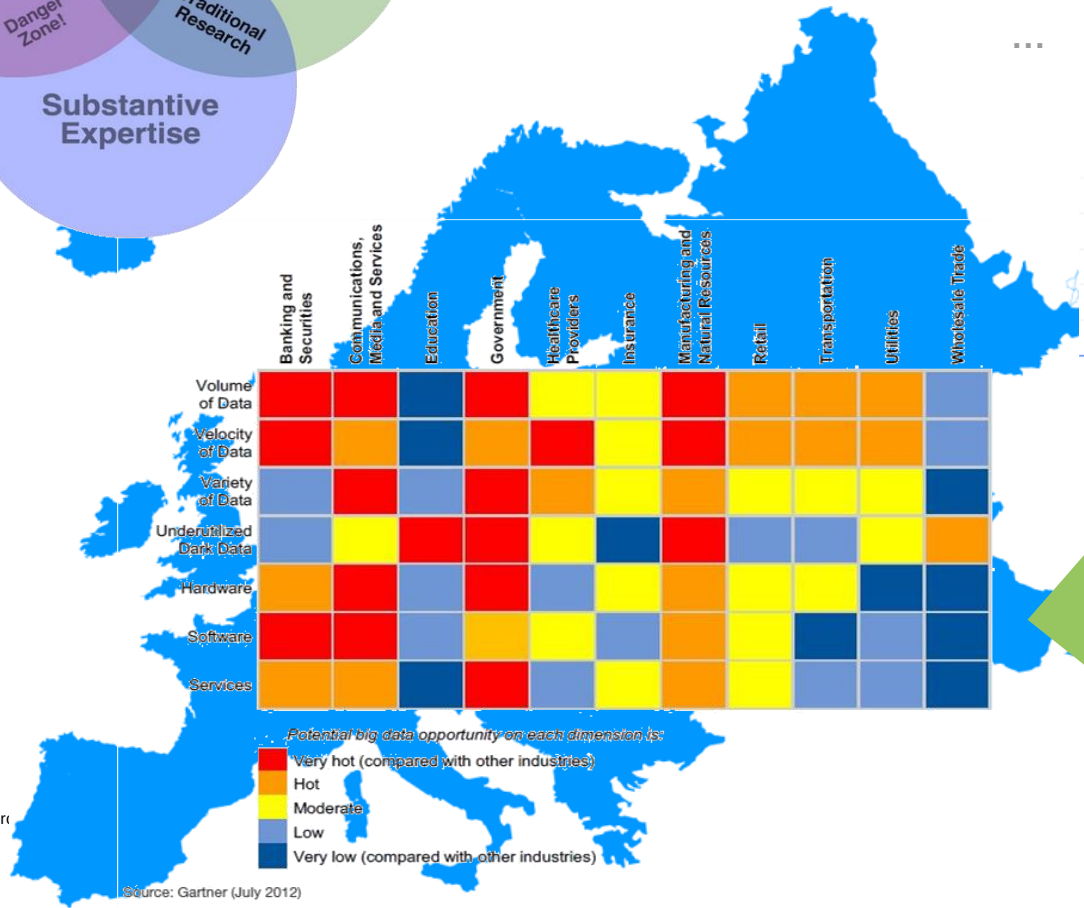
Whence?

amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

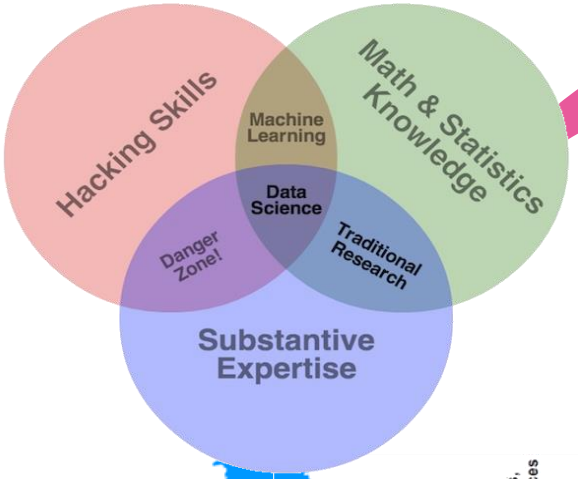
Why?



Where to?



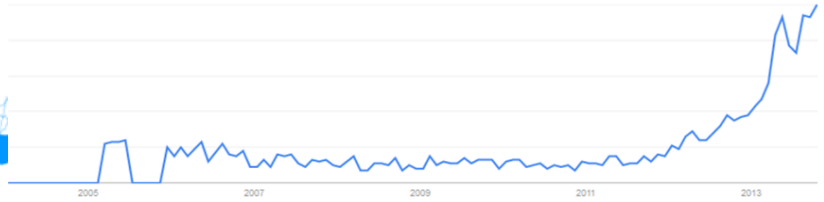
Data Science?



Whence?

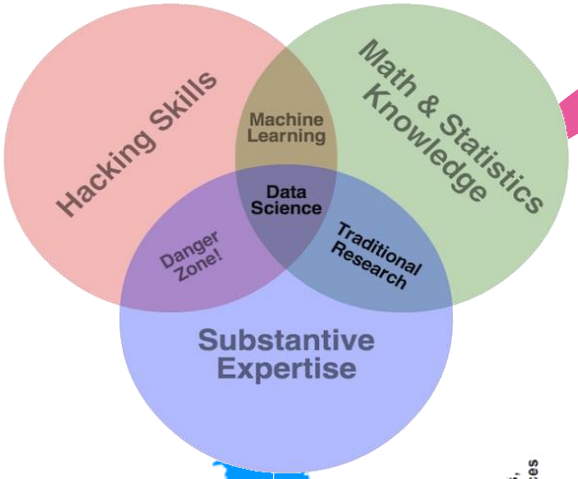
amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

Why?



Where to?

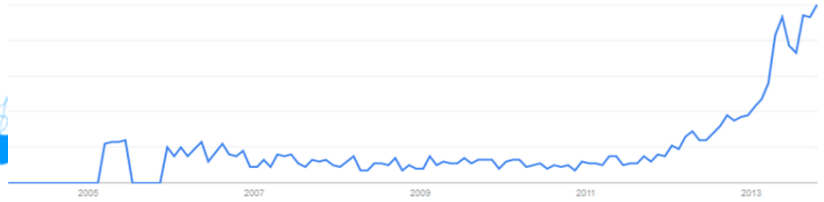
Data Science?

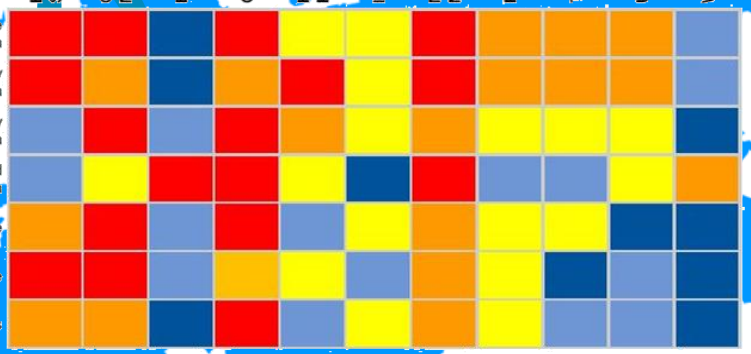


Whence?

amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

Why?

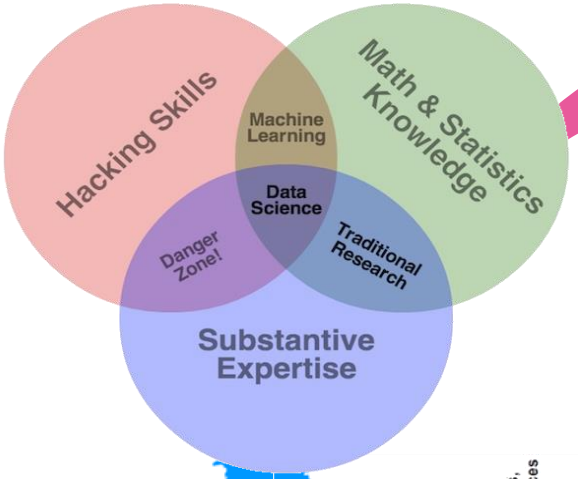




Source: Gartner (July 2012)

Where to?

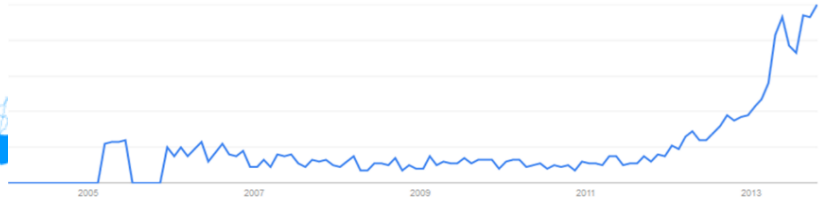
Data Science?



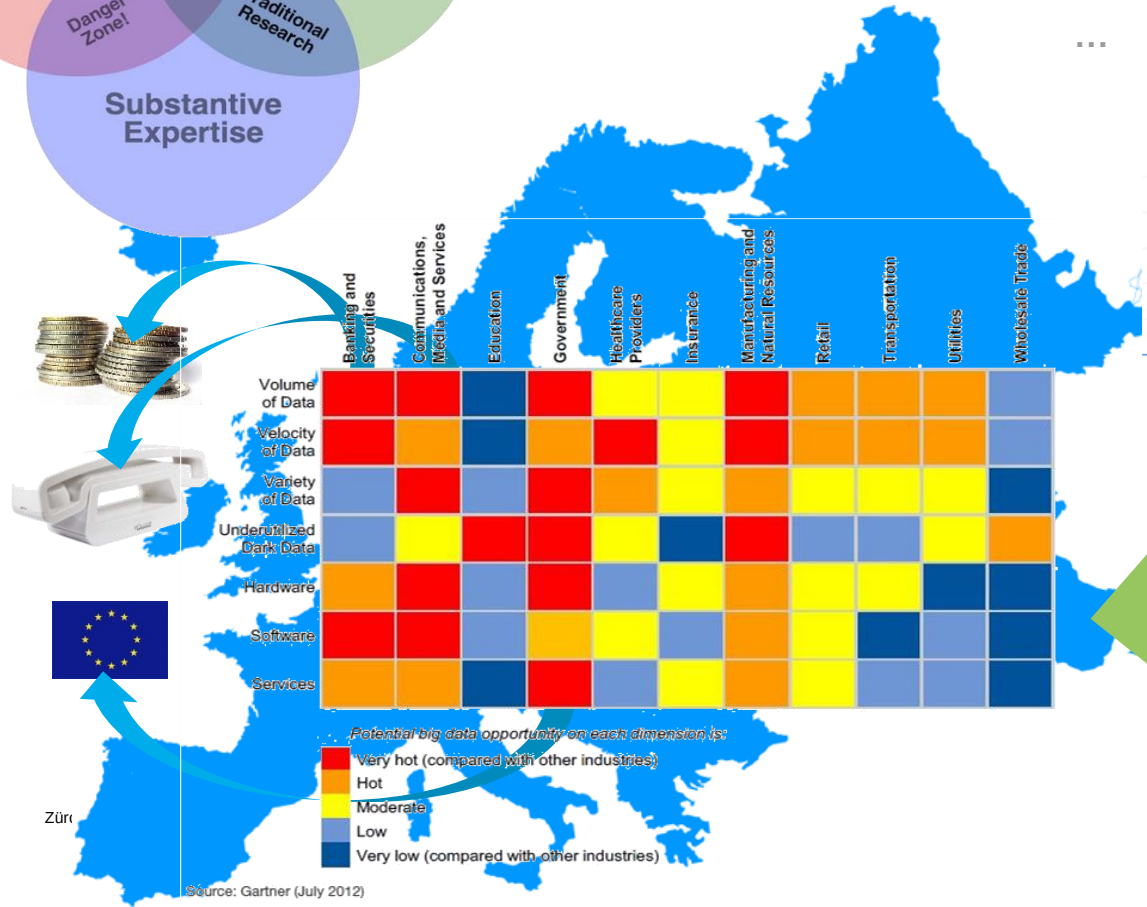
Whence?

amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

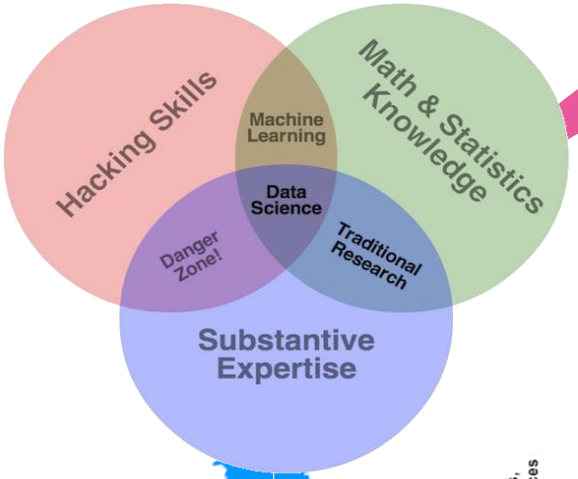
Why?



Where to?



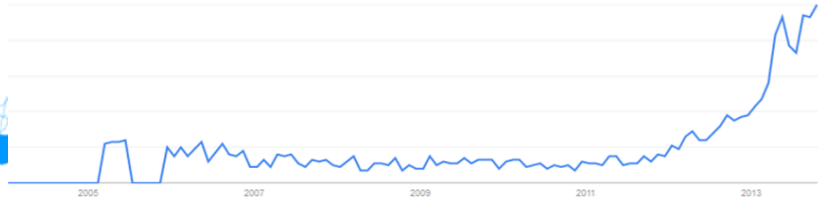
Data Science?

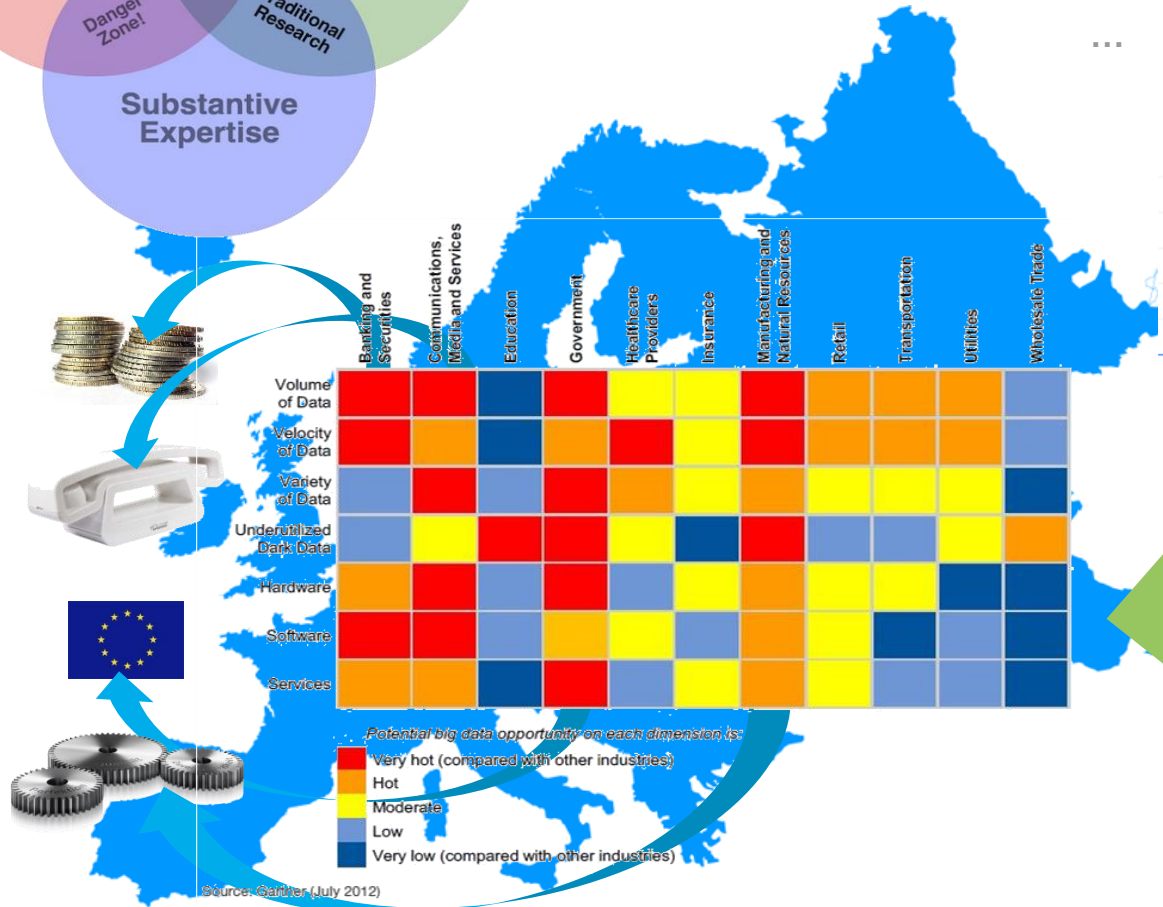


Whence?

amazon
facebook
Google
IBM
LinkedIn
NETFLIX
...

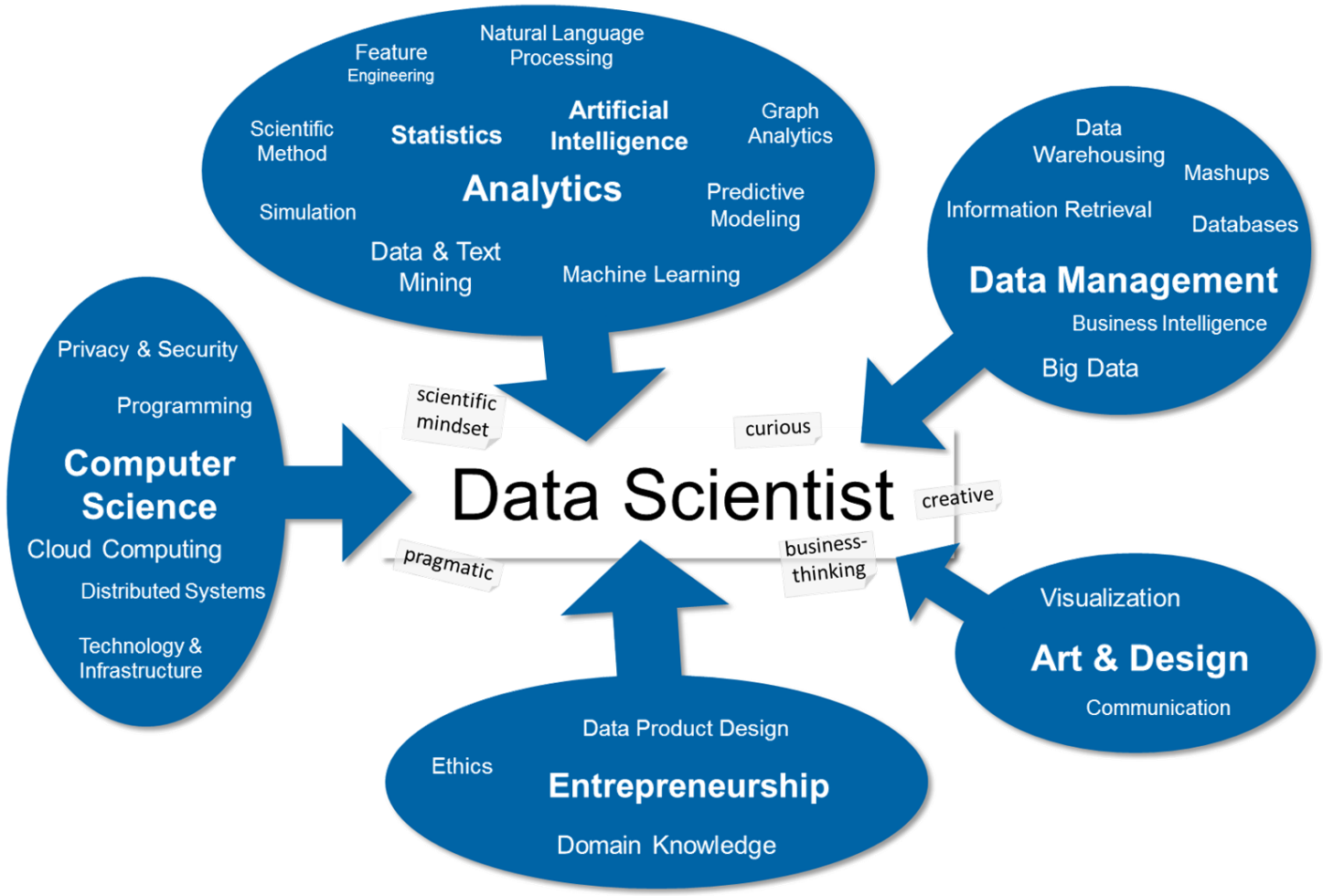
Why?



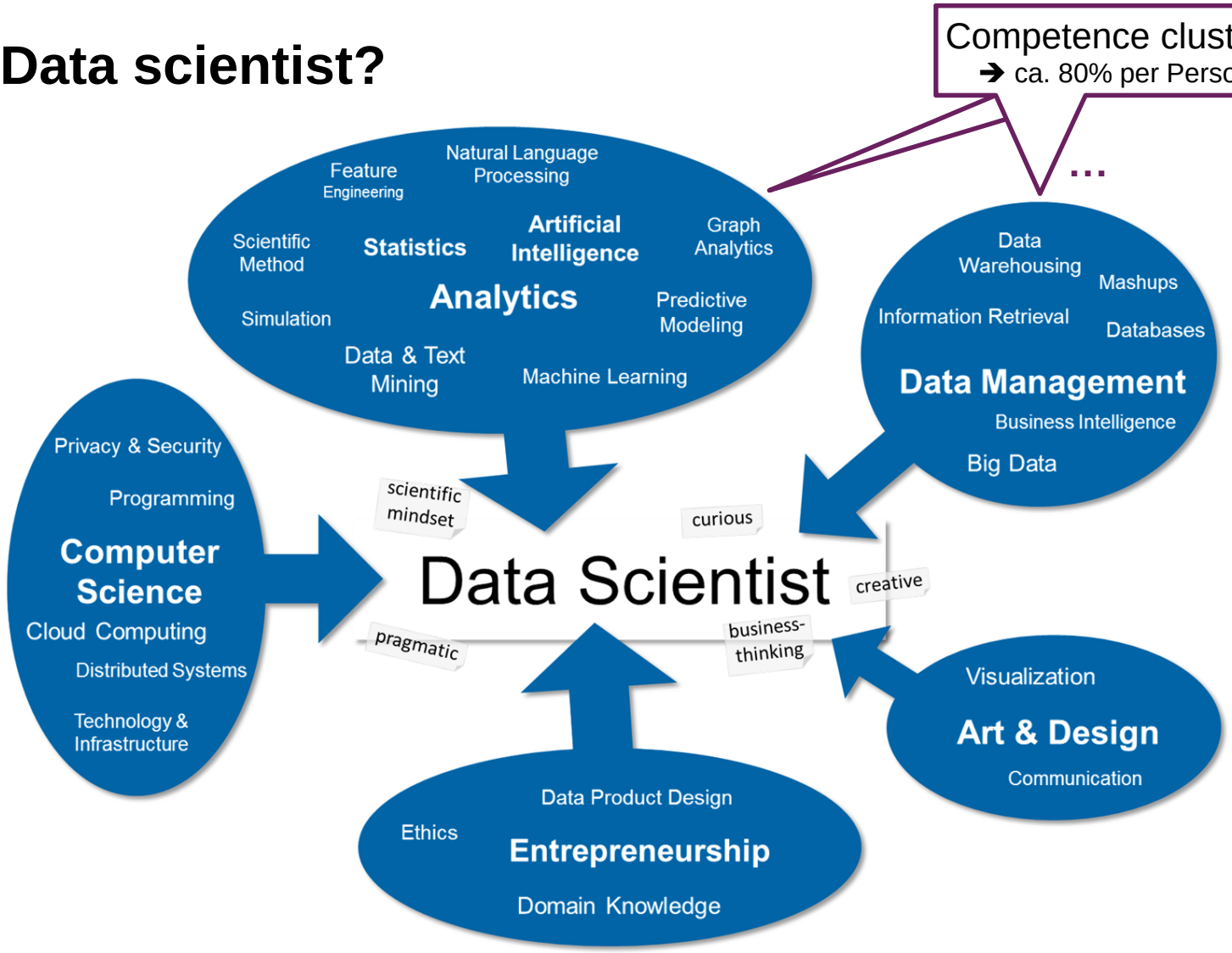


Where to?

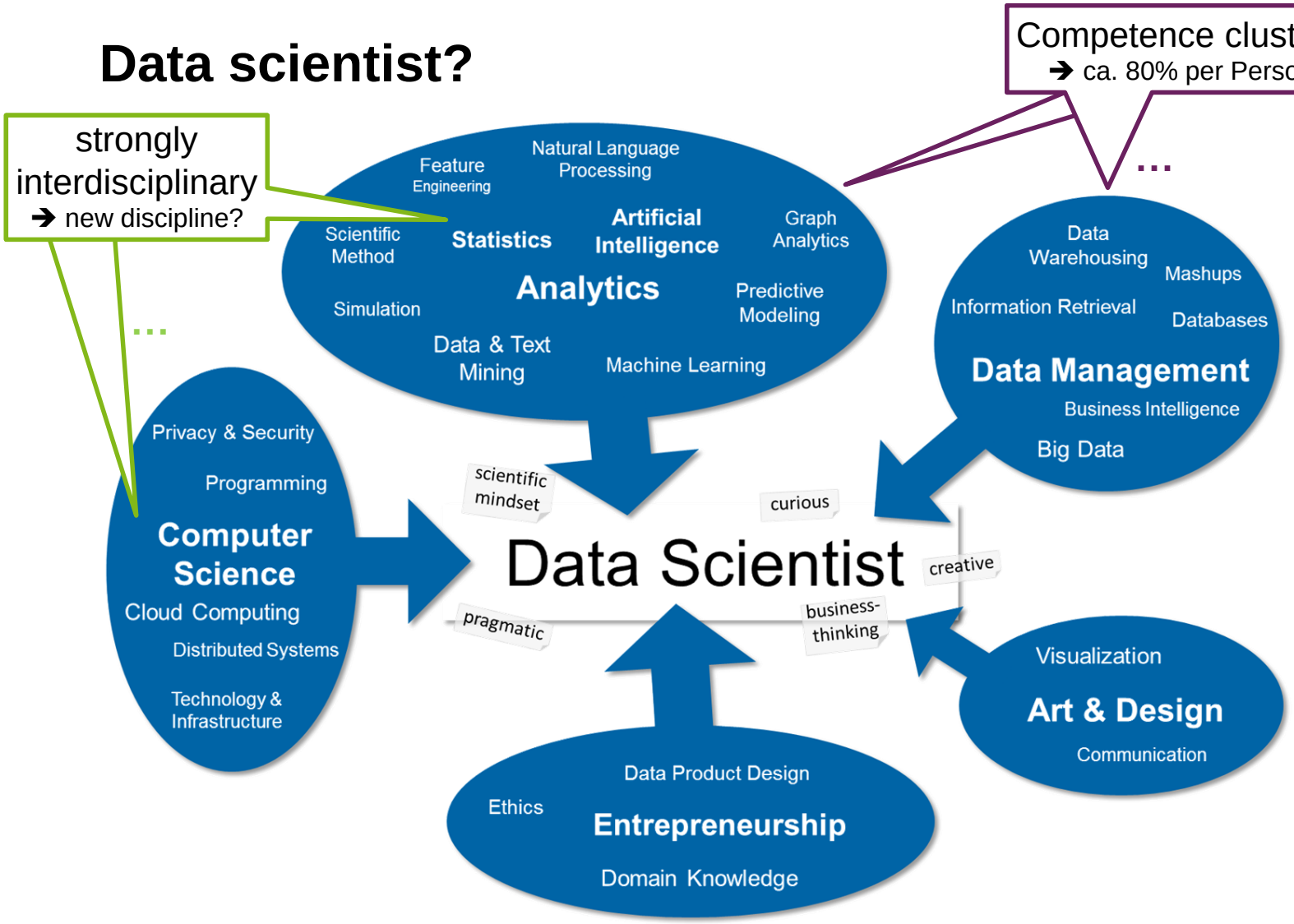
Data scientist?



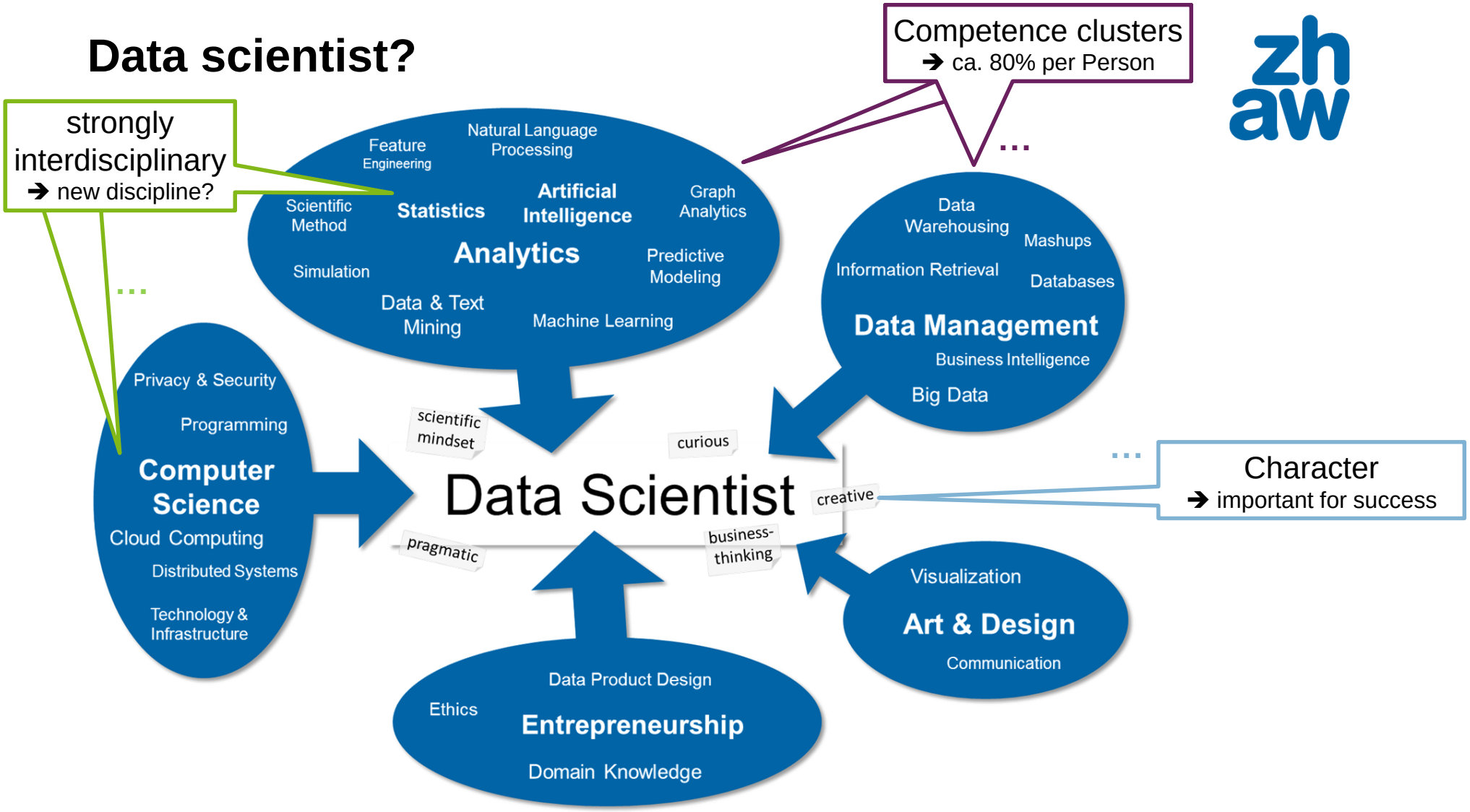
Data scientist?



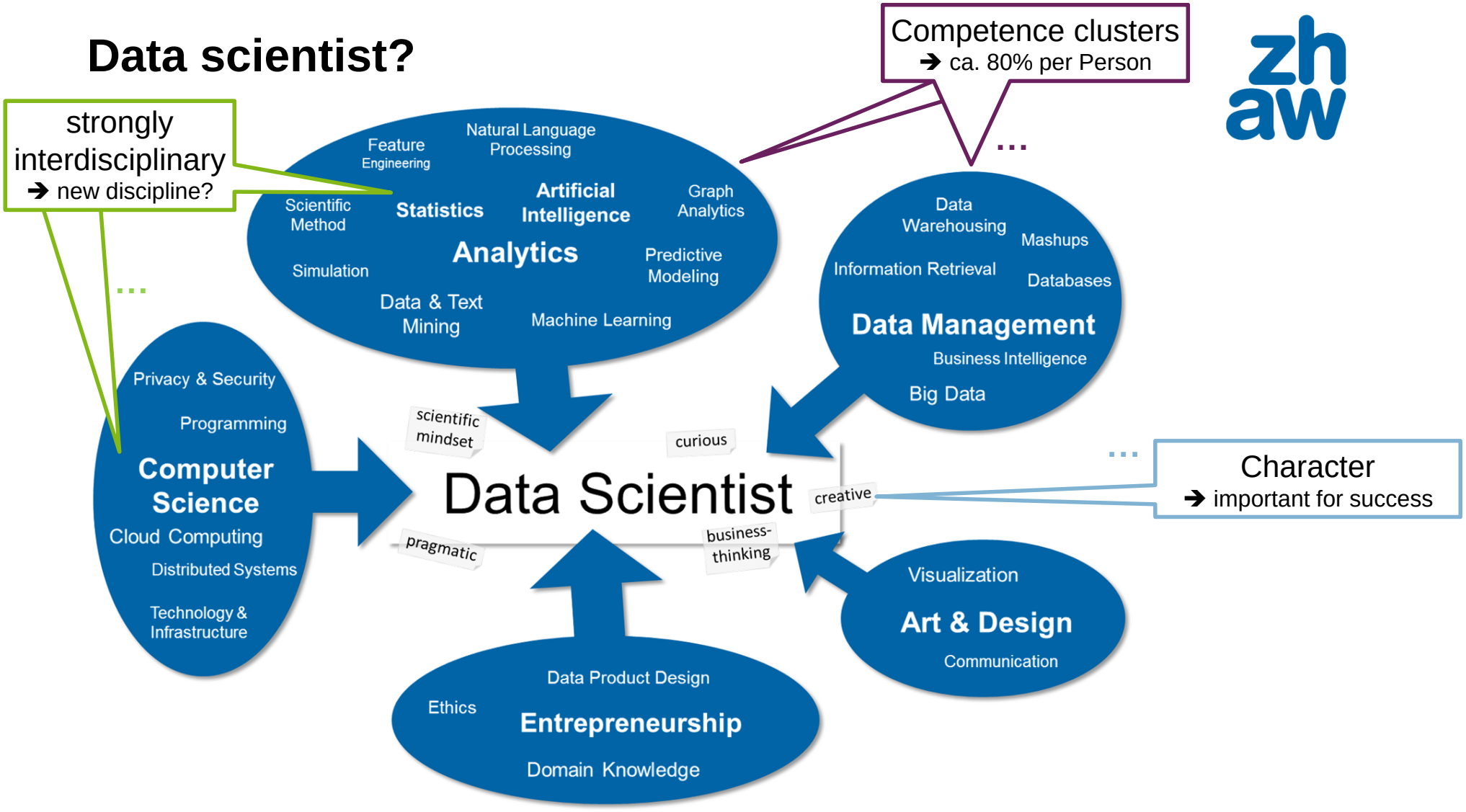
Data scientist?



Data scientist?



Data scientist?



➔ **Unique blend of skills** from AI, engineering & communication aiming at generating **value from the data**

Zurich University
of Applied Sciences

Project Examples

The ZHAW Data Science Laboratory

Zurich University of Applied Sciences

The ZHAW Data Science Laboratory

- One of the first European centers for Data Science R&D and teaching
- Interdisciplinary virtual organization spanning several instituts
- Core competency: Data Product design with structured and unstructured data

datalab
dlab.zhaw.ch

Projects

- Many years of successful collaborations between academia and business
- Focused on Swiss SMEs as well as European programmes

Teaching

- Undergraduate and Graduate Courses
- One of the first European professional education programmes for Data Scientists
- Seminars and Workshops for idea exchange with industry

MSE

Foundation Register 2.0

Zurich University of Applied Sciences

stiftungschweiz.ch

NonproCons
Neue Wege für Nonprofit-Organisationen

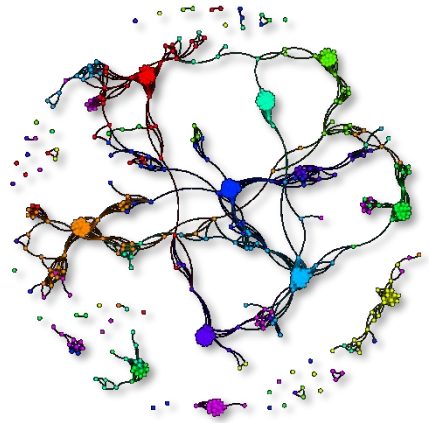
Situation: Ca. 7’500 foundations in Switzerland

Goal: Simple visual search by project proposal should yield most probable sponsor

Challenges: Quantify and visualize content-based similarity of foundation’s missions

Solution:

- Develop multilingual retrieval system
- Search on very small document collection (7.5k foundation’s mission statements)
  ➔ Extremely recall-oriented search
- High amount of data for intuitive visualizations
  ➔ “Forced Directed Layout” of similarities from term-document matrix and topic modeling

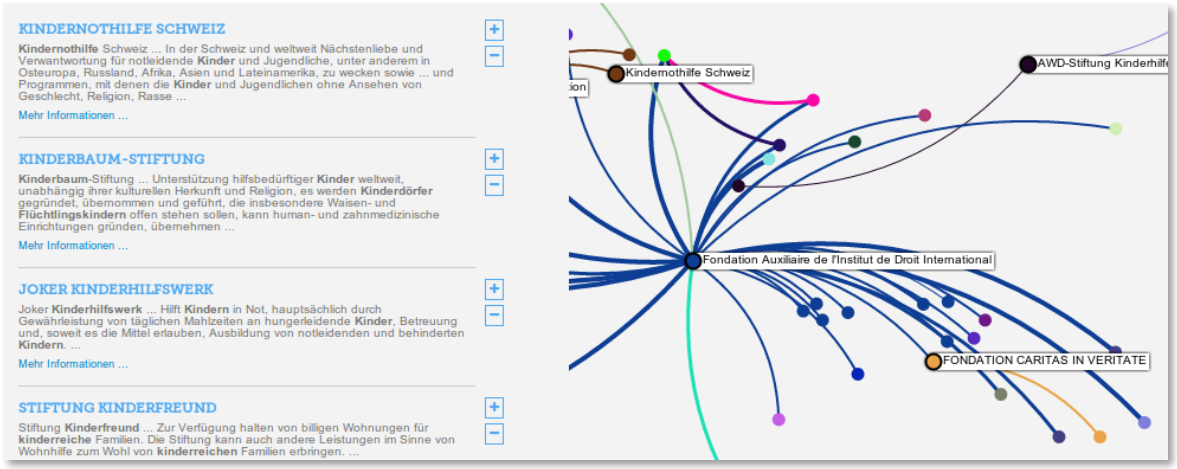


Foundation Register 2.0
Technical Aspect: Multilingual IR System

Zurich University of Applied Sciences

Solution:

- Different index fields for search (DE, FR, IT) and network similarity calculation (CROSS)
- Dual iterative search process:
  - search in own language → relevance feedback on visual neighborhood → foreign language docs appear in search results
- CROSS field construction:
  - Aggressive stop word elimination (outlier removal according to Zipf distribution)
  - No stemming (to retain precision)
  - Machine translation to main language ➔ drop words that couldn't be translated
- Visualization generation:
  - Build similarity matrix of docs in translated CROSS field for network generation
  - Use QuadTree algorithm for large collections



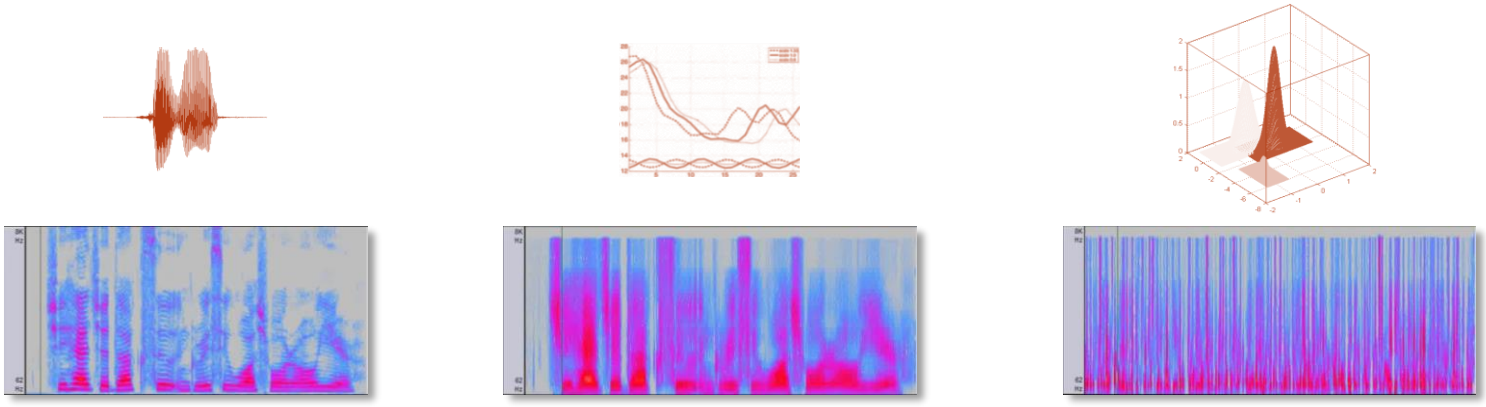
Talkalyzer

Goal: Speaker Recognition in meetings on mobile devices

Challenge: Build reliable speaker models

Approach:

- Loosen i.i.d. assumption on feature vectors
- Use Viola&Jones' face detection approach on audio features
  ➔ find typical sounds of a speaker in a spectrogram



Enterprise Knowledge Curation

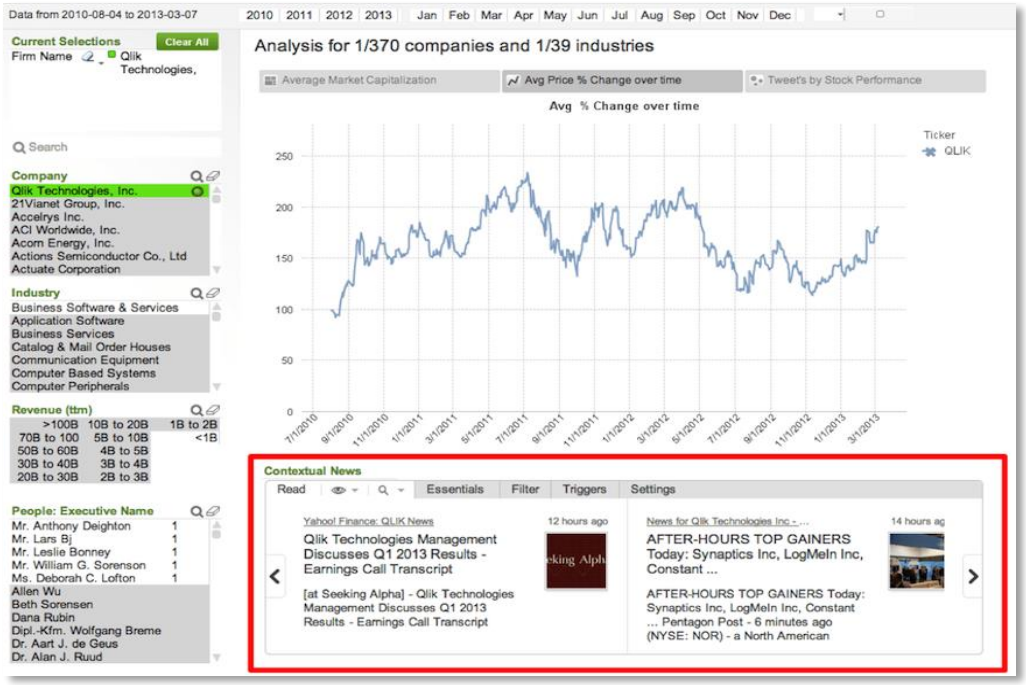
Zurich University of Applied Sciences

Goal: Contextual News for Structured Data
Challenge: Bridging the sematic gap

squirro

Solution:

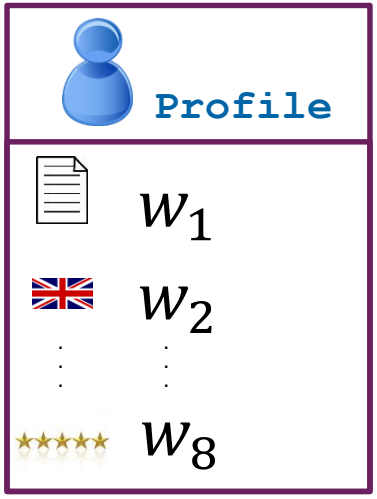
- IR Pipeline with implicit relevance feedback
- Learning user profiles from structured and unstructured data



Enterprise Knowledge Curation (contd.)
Technical Aspect: Data Fusion

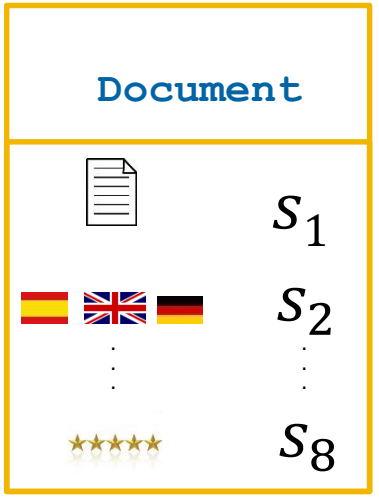
Subgoal: Find relevant news documents per user (document = text & metadata)
Challenge: How to combine information on text and metadata

Solution:
- Score Fusion: Fusion of separate score using a weighted average
- Learning to Rank: Learn weights with logistic regression
- Training data: Click-based implicit user feedback



$$s(d,q) = \sum_i w_i \cdot s_i$$

➔ Next challenge: overcome cold-start by avoiding learning on training data

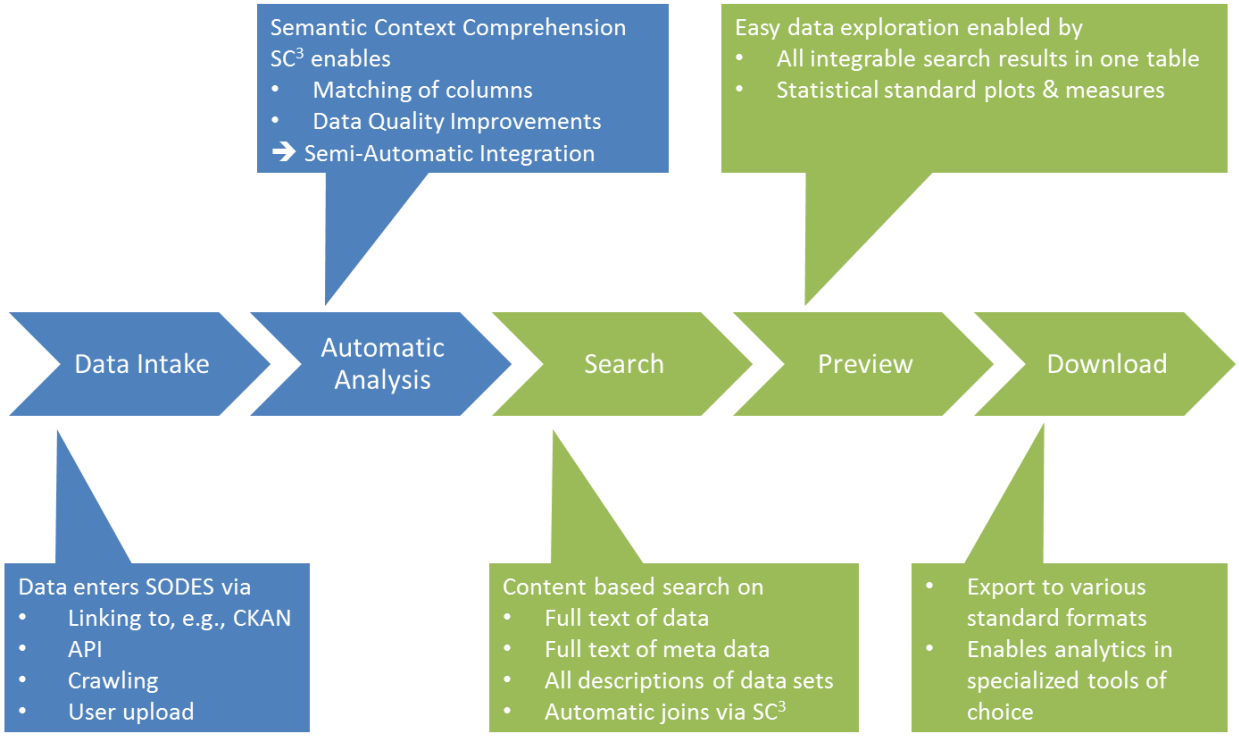


SODES – Swiss Open Data Exploration System

Challenge: Open Data promises to be a gold mine – but accessing and combining data from different data sources turns out to be non-trivial and very time consuming

Goal: A platform that enables easy and intuitive access, integration and exploration of different data sources

Solution:



The State of the Union

Zurich University
of Applied Sciences

Summary of challenges faced so far

Zurich University of Applied Sciences

Transition US → Switzerland

- We don't have internet behemoths
- But we have banks/telcos, industry, retailers, smart*, societal challenges

Uphold quality standard

- It's Data *Science!*
- ➔ Develop curricula

«Evaluation»: DAS in Data Science

Zurich University of Applied Sciences

Diploma of Advanced Studies (DAS) professional education programme

- Start: this fall
- Three modules (part time, one afternoon + evening per week)

CAS Data Science Applications

Machine Learning, Big Data Visualization, Design & Development of Data Products, Data Protection & Security

CAS Information Engineering

Scripting inPython, Information Retrieval & Text Analytics, Databases & SQL, Data Warehousing, Big Data

CAS Data Analytics

Data Description & Visualization, Statistical Foundations of Analytics, Multiple Regression, Time Series & Forecasting, Clustering & Classification

➔ Strong demand from industry (CAS Data Analytics already overbooked)

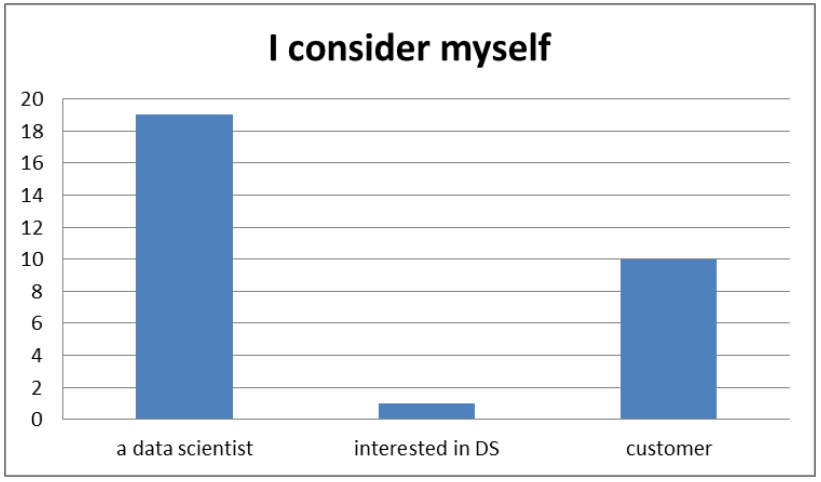
Zurich University
of Applied Sciences

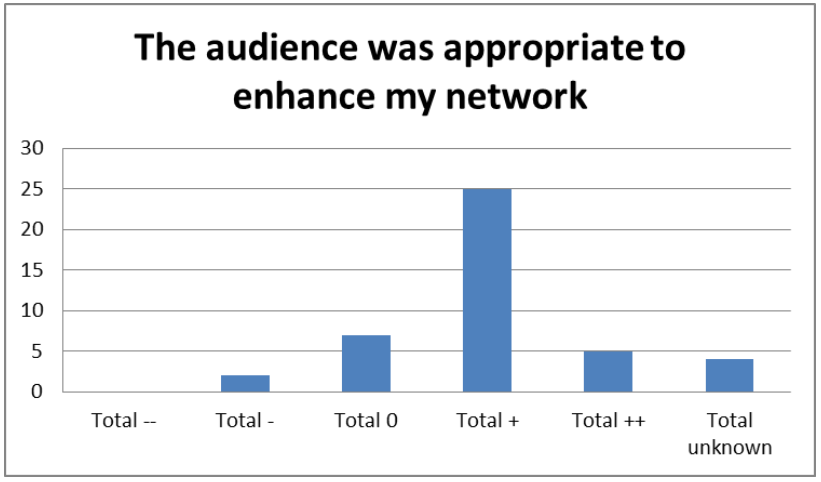
Evaluation: 1st Swiss Workshop on Data Science

Impressions

Evaluation: 1st Swiss Workshop on Data Science

- Expected 60 participants → got 120 plus 7 sponsors
- Ca. ¾ opted in to build a community of Swiss Data Scientists
- Two groups of participants: Data Scientists / Managers needing Data Scientists





Conclusions

- Big **demand** from **industry** regarding Data Science related topics
  - Partly due to routing of requests for formerly diverse topics to central institution
  - More inquiries than actual projects…
  - …but typically more projects than researchers
- Data Science has **different Application** areas in Switzerland
  - But marketing / customer analytics is still a big one
- Chance for academia
  - Bring together diverse expertise from many disciplines
  - **Coin a field**