# Data Science & Society

## Lecture 01: *Course introduction*

INFOMDSS 2018

http://www.cs.uu.nl/education/vak.php?stijl=2&vak=INFOMDSS

Dr. Marco Spruit & Dr. Matthieu Brinkhuis

ADS

# Agenda

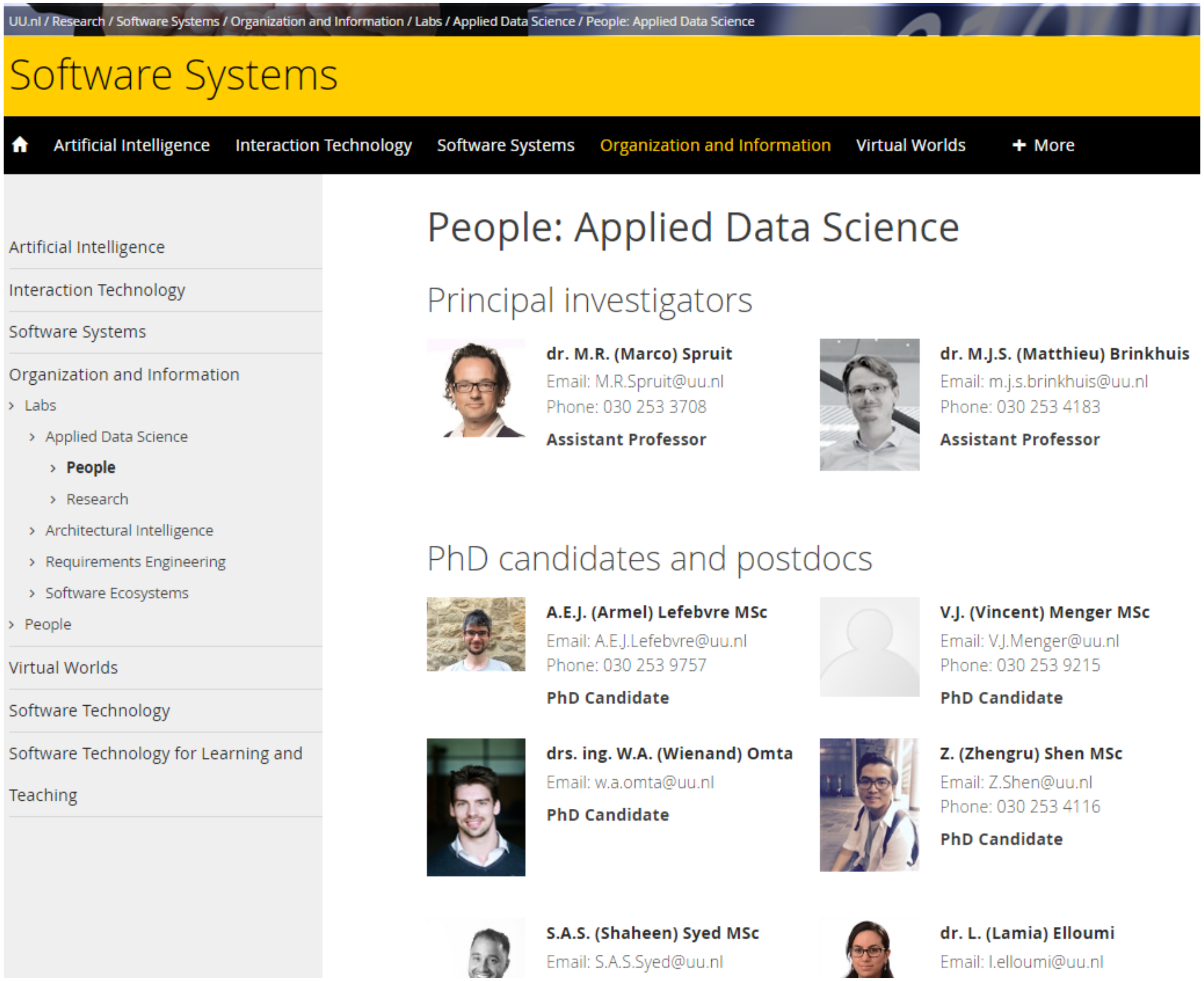
1. About us
2. About this course: Data Science & Society
   - Rationale
   - Learning objectives
   - Assignments & Workshops
   - Lectures, Topics and Guests
   - Your expectations
3. About this scientific field: ADS in context (KDP)
4. About Assignment 1: Book review

# 1. About us

ADS lab, MS, team, you

# About… Applied Data Science Lab

https://www.uu.nl/en/research/software-systems/organization-and-information/labs/applied-data-science



UU.nl / Research / Software Systems / Organization and Information / Labs / Applied Data Science / People: Applied Data Science

Software Systems

## People: Applied Data Science

### Principal investigators

**dr. M.R. (Marco) Spruit**
Email: M.R.Spruit@uu.nl
Phone: 030 253 3708
**Assistant Professor**

**dr. M.J.S. (Matthieu) Brinkhuis**
Email: m.j.s.brinkhuis@uu.nl
Phone: 030 253 4183
**Assistant Professor**

### PhD candidates and postdocs

**A.E.J. (Armel) Lefebvre MSc**
Email: A.E.J.Lefebvre@uu.nl
Phone: 030 253 9757
**PhD Candidate**

**V.J. (Vincent) Menger MSc**
Email: V.J.Menger@uu.nl
Phone: 030 253 9215
**PhD Candidate**

**drs. ing. W.A. (Wienand) Omta**
Email: w.a.omta@uu.nl
**PhD Candidate**

**Z. (Zhengru) Shen MSc**
Email: Z.Shen@uu.nl
Phone: 030 253 4116
**PhD Candidate**

**S.A.S. (Shaheen) Syed MSc**
Email: S.A.S.Syed@uu.nl

**dr. L. (Lamia) Elloumi**
Email: l.elloumi@uu.nl

Spruit,M., & Jagesar,R. (2016). Power to the People! Meta-algorithmic modelling in applied data science. In Fred,A., & Filipe,J. (Eds.), *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 29–35). November 11-13, 2016 in Porto, Portugal: SciTePress. [pdf]

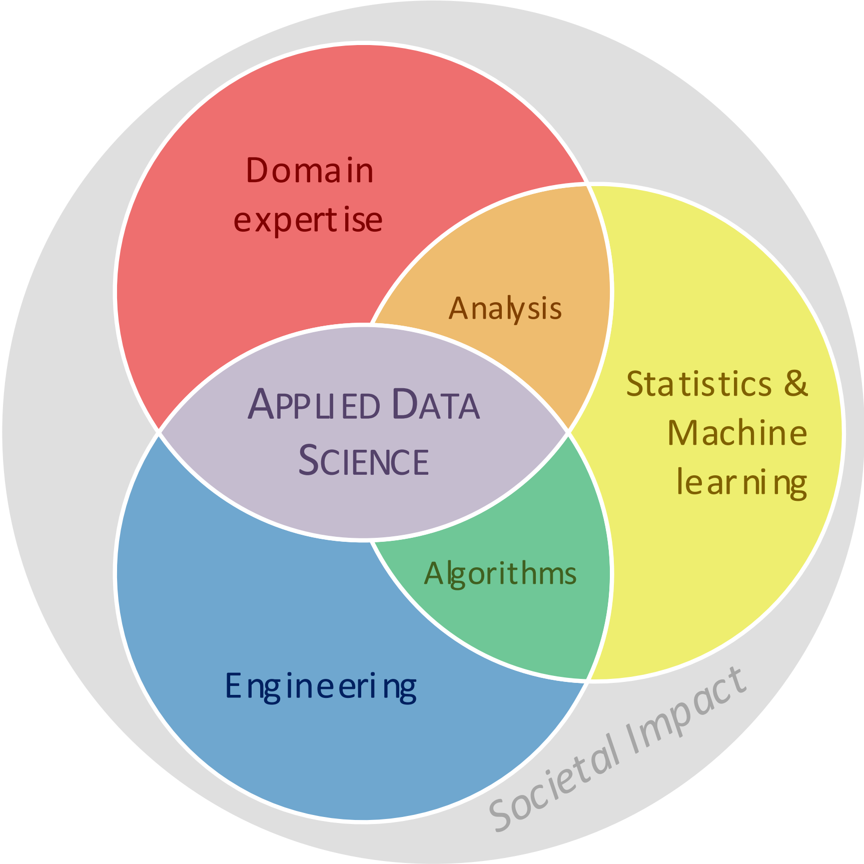
# About… Applied Data Science Lab

https://www.uu.nl/en/research/software-systems/organization-and-information/labs/applied-data-science

› **"Applied Data Science (ADS) is**

1. the knowledge discovery process in which
2. analytical applications are designed and evaluated to
3. improve the daily practices of domain experts."

› *Spruit & Jagesar (2016)*
*Spruit & Lytras (2018)*



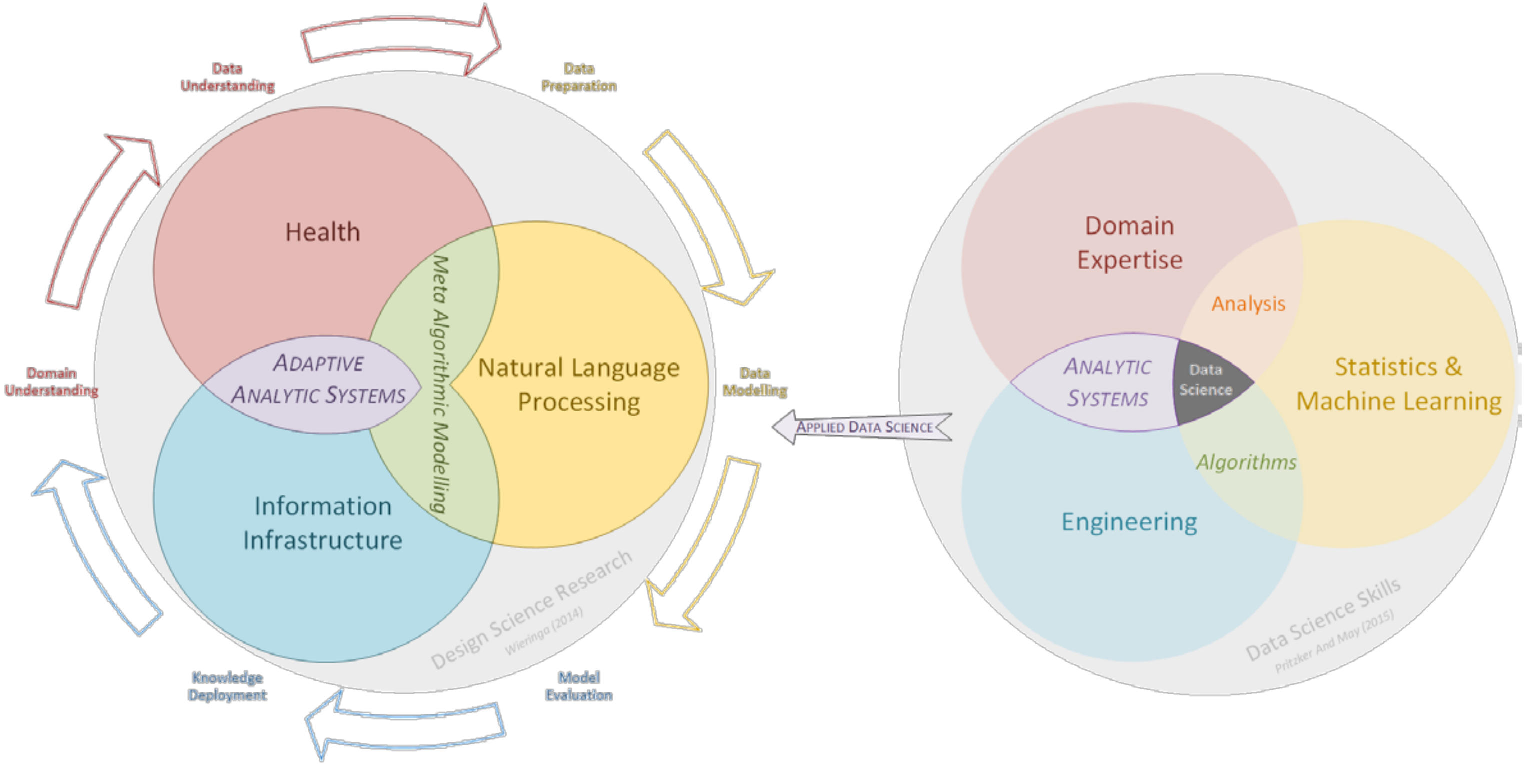
Spruit,M., & Jagesar,R. (2016). Power to the People! Meta-algorithmic modelling in applied data science. In Fred,A., & Filipe,J. (Eds.), *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 29–35). November 11-13, 2016 in Porto, Portugal. Porto, Portugal: SciTePress. [pdf]

# About… Marco Spruit

- 1988-1995: Computational linguistics
- 1993-1997: Business intelligence engineer
- 1997-2006: Independent software vendor
- 2003-2015: Applied data scientist

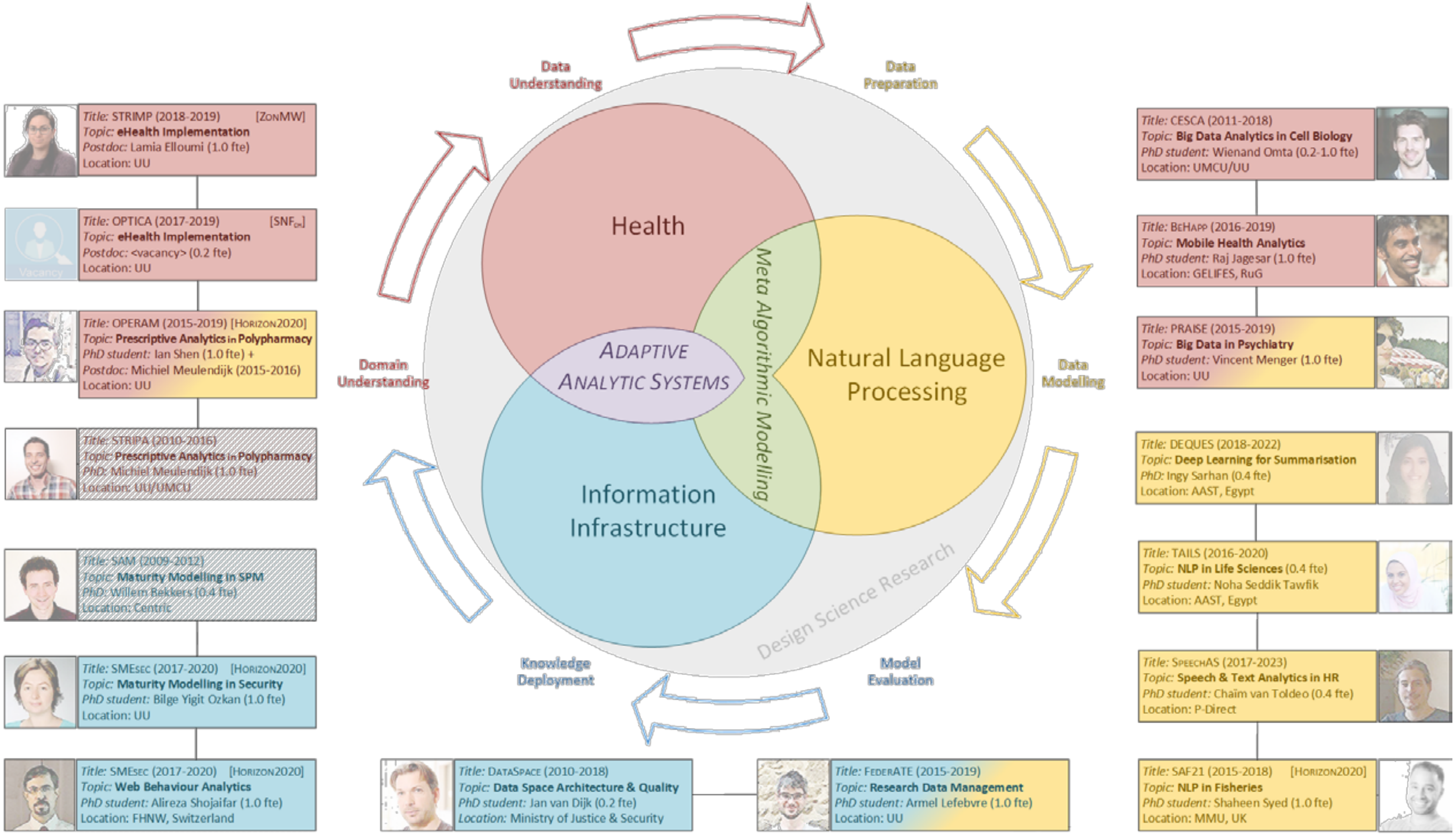
- http://www.linkedin.com/in/spruit

# Research: Adaptive Analytic systems



- http://m.spru.it/res

# Projects: Analytic systems



*Title:* STRIMP (2018-2019) [ZonMW]
*Topic:* **eHealth Implementation**
*Postdoc:* Lamia Elloumi (1.0 fte)
Location: UU

*Title:* OPTICA (2017-2019) [SNF$_{CH}$]
*Topic:* **eHealth Implementation**
*Postdoc:* <vacancy> (0.2 fte)
Location: UU

*Title:* OPERAM (2015-2019) [Horizon2020]
*Topic:* **Prescriptive Analytics in Polypharmacy**
*PhD student:* Ian Shen (1.0 fte) +
*Postdoc:* Michiel Meulendijk (2015-2016)
Location: UU

*Title:* STRIPA (2010-2016)
*Topic:* **Prescriptive Analytics in Polypharmacy**
*PhD:* Michiel Meulendijk (1.0 fte)
Location: UU/UMCU

*Title:* SAM (2009-2012)
*Topic:* **Maturity Modelling in SPM**
*PhD:* Willem Bekkers (0.4 fte)
Location: Centric

*Title:* SMEsec (2017-2020) [Horizon2020]
*Topic:* **Maturity Modelling in Security**
*PhD student:* Bilge Yigit Ozkan (1.0 fte)
Location: UU

*Title:* SMEsec (2017-2020) [Horizon2020]
*Topic:* **Web Behaviour Analytics**
*PhD student:* Alireza Shojaifar (1.0 fte)
Location: FHNW, Switzerland

*Title:* DataSpace (2010-2018)
*Topic:* **Data Space Architecture & Quality**
*PhD student:* Jan van Dijk (0.2 fte)
*Location:* Ministry of Justice & Security

*Title:* FederATE (2015-2019)
*Topic:* **Research Data Management**
*PhD student:* Armel Lefebvre (1.0 fte)
Location: UU

*Title:* CESCA (2011-2018)
*Topic:* **Big Data Analytics in Cell Biology**
*PhD student:* Wienand Omta (0.2-1.0 fte)
Location: UMCU/UU

*Title:* BeHapp (2016-2019)
*Topic:* **Mobile Health Analytics**
*PhD student:* Raj Jagesar (1.0 fte)
Location: GELIFES, RuG

*Title:* PRAISE (2015-2019)
*Topic:* **Big Data in Psychiatry**
*PhD student:* Vincent Menger (1.0 fte)
Location: UU

*Title:* DEQUES (2018-2022)
*Topic:* **Deep Learning for Summarisation**
*PhD:* Ingy Sarhan (0.4 fte)
Location: AAST, Egypt

*Title:* TAILS (2016-2020)
*Topic:* **NLP in Life Sciences** (0.4 fte)
*PhD student:* Noha Seddik Tawfik
Location: AAST, Egypt

*Title:* SpeechAS (2017-2023)
*Topic:* **Speech & Text Analytics in HR**
*PhD student:* Chaïm van Toldeo (0.4 fte)
Location: P-Direct

*Title:* SAF21 (2015-2018) [Horizon2020]
*Topic:* **NLP in Fisheries**
*PhD student:* Shaheen Syed (1.0 fte)
Location: MMU, UK

# Course team

- Dr. Marco Spruit

Melchior Keijdener

- Dr. Matthieu Brinkhuis

Laurens Müter

Vincent Menger, MSc

# About you

- Check your name on the Attendance list

- MBI?
- ADS profile?
- Not ICS dept?
- GSLS?

# 2. About this course

A bird's eye view of Data Science & Society

# Big Ideas vs Big Diversity vs Big Numbers

- Big Ideas
  - Trigger your enthousiasm for applied data science
  - Inspire you to aim for societal impact through data science
- Feasibility Constraints
  - Provide you with data/information science essentials
  - Account for your diversity
  - Balance classroom scalability limitations
- Background Diversity
  - Walk the fine line between Big Ideas and their Feasibility
- *“Third time’s the charm”*

# Learning objectives

1. Understand the role of data science and its societal impact
2. Recognise the knowledge discovery processes in applied data science
3. Identify trends and developments in big data technologies
4. Apply selected big data technologies to solve real-world problems

# Learning objectives in course components

| | Book review | Assign-ments | Work-shops | Guest talks | Regular lectures |
|---|---|---|---|---|---|
| Understand the role of data science and its societal impact | X | | | X | X |
| Recognise the knowledge discovery processes in applied data science | | X | X | | X |
| Identify trends and developments in big data engineering & analytics | X | | X | | X |
| Apply selected big data technologies to solve real-world problems | | X | X | | |

# Types of assignments

1. Explore data science and its societal impact
   - Read, review and (optionally) present book of your own choice
     - Individual assignment, assessed by staff
2. Study selected scientific literature
   - Read selected scientific papers that are being covered during the course
     - Individual assignments, assessed in written exam
   - Read parts of the two books that are being discussed during the course
     - Individual assignment, assessed in written exam
3. Practice with big data tools
   - Perform various tutorials and assignments to familiarize yourself
     - Individual assignments, assessed in written exam
   - Perform case studies as mid-term and final projects
     - Individual assignments, assessed in written exam

# Grading

A. Book review (2-pager + optional pitch)

B. Mid-term exam (remindo)

C. End-term exam (remindo)

D. Optional bonus for extraordinary participation/performance

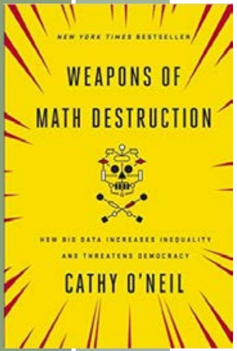
- Grade = A*0.10 + B*0.40 + C*0.50 + D

- **NB1:** To qualify for the second chance exam, all grading components need to be at least 4.0, and component A needs to have been submitted within the allotted time.
- **NB2:** The 2nd chance exam is an extensive market survey report assignment.
- **NB3:** You will *not* be graded on your weekly assignments.

# Assignment 1: Book review - Explore data science and its societal impact

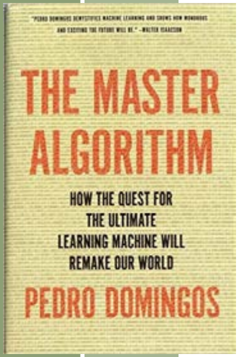
- Individual assignment
- Select a popular book from our longlist that may inspire or provoke yourself and us
  - Review the longlist of allowed books: http://bit.ly/infomdss-books
  - Submit your Top 3 selection: http://bit.ly/infomdss-form1
    - within 24 hours from now
- Read the assigned book
  - Order book if necessary, get it from the library, borrow it, etc...
- Submit the book review form before DEADLINE-02
  - Your structured review and 2-pager pitch:
    - http://bit.ly/infomdss-assignment1
- For each of the 20 books, the highest graded 2-Pager will be presented as a 3-minute-pitch in the lecture slot on Tue Oct 2

# Here's a random selection of the 20 books…



- Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy [amazon]
  - "A former Wall Street quant sounds an alarm on the mathematical models that pervade modern life — and threaten to rip apart our social fabric"

- Life 3.0 Being Human in the Age of Artificial Intelligence [amazon]
  - "How will Artificial Intelligence affect crime, war, justice, jobs, society and our very sense of being human?"



- The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World [amazon]
  - "Pedro Domingos lifts the veil to give us a peek inside the learning machines that power Google, Amazon, and your smartphone. He assembles a blueprint for the future universal learner-the Master Algorithm-and discusses what it will mean for business, science, and society. If data-ism is today's philosophy, this book is its bible."

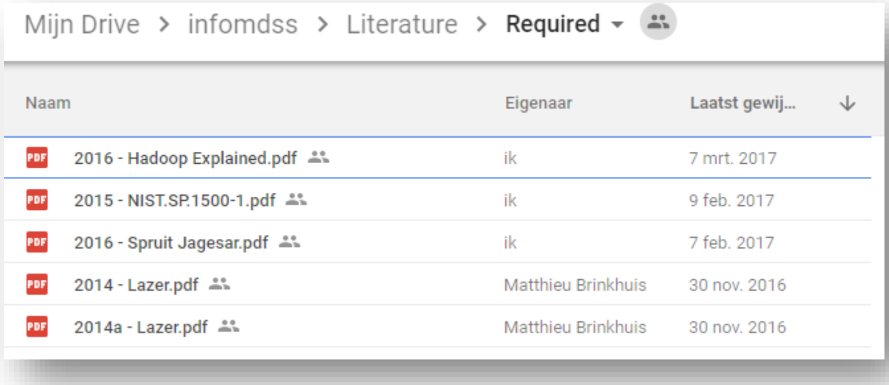
# Some of the course literature

See http://www.cs.uu.nl/education/vak.php?stijl=2&vak=INFOMDSS

- White, J. (2016). *Hadoop: The Definitive Guide.* Third edition. O'Reilly.
- Chambers, B., & Zaharia, M. (2018). *Apache Spark - The Definitive Guide.* O'Reilly.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science, 343(*6176), 1203-1205.
- *Fundamentals of information systems (IS.pdf)*

See **Required** folder in:

- http://bit.ly/infomdss-lit (...)

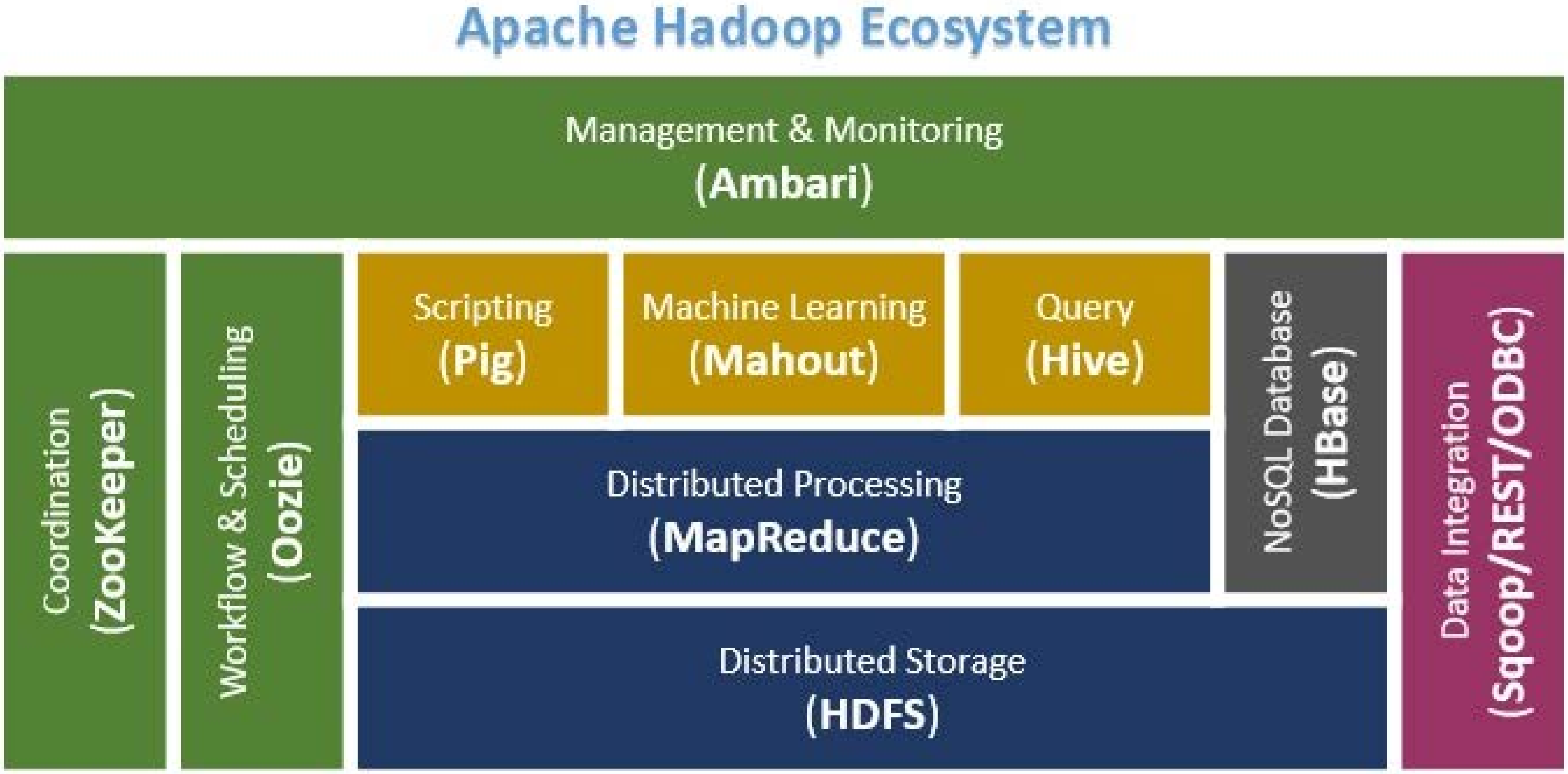


Mijn Drive › infomdss › Literature › Required

| Naam | Eigenaar | Laatst gewij... |
|---|---|---|
| 2016 - Hadoop Explained.pdf | ik | 7 mrt. 2017 |
| 2015 - NIST.SP.1500-1.pdf | ik | 9 feb. 2017 |
| 2016 - Spruit Jagesar.pdf | ik | 7 feb. 2017 |
| 2014 - Lazer.pdf | Matthieu Brinkhuis | 30 nov. 2016 |
| 2014a - Lazer.pdf | Matthieu Brinkhuis | 30 nov. 2016 |

# Assignment(s) : Practice with big data tools

- Python, R
  - on Azure in Ubuntu VM
- Hadoop
- Spark

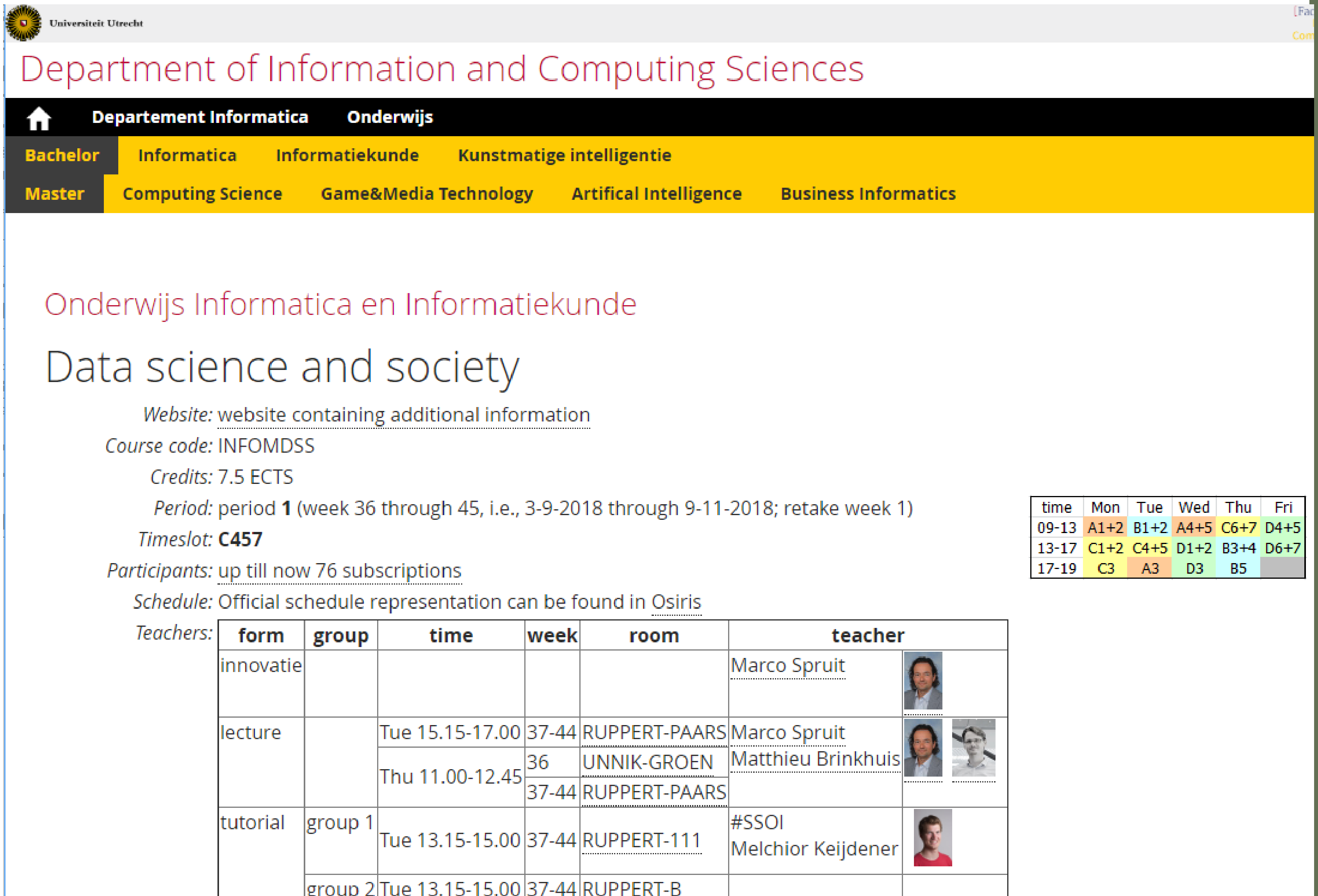


# Guest lectures (-1)

- Neonatology (WKZ/UMCU)
  - Prof. Manon Benders MD
  - Dr. Daniel Vijlbrief MD
- Epidemiology (Julius/UMCU)
  - Dr. Charlotte Onland-Moret MD
- Big Data in Psychiatry (UMCU)
  - Prof. Floortje Schepers, with Vincent Menger MSc
- Geospatial information systems
  - ESRI NL
- Ethics, Privacy, Regulations in Big Data
  - Menno Mostert, MSc (UMCU)

# Course website:

[http://www.cs.uu.nl/education/vak.php?stijl=2&vak=INFOMDSS](http://www.cs.uu.nl/education/vak.php?stijl=2&vak=INFOMDSS)



Universiteit Utrecht

Department of Information and Computing Sciences

Onderwijs Informatica en Informatiekunde

Data science and society

*Website:* website containing additional information

*Course code:* INFOMDSS

*Credits:* 7.5 ECTS

*Period:* period **1** (week 36 through 45, i.e., 3-9-2018 through 9-11-2018; retake week 1)

*Timeslot:* **C457**

*Participants:* up till now 76 subscriptions

*Schedule:* Official schedule representation can be found in Osiris

| time | Mon | Tue | Wed | Thu | Fri |
|---|---|---|---|---|---|
| 09-13 | A1+2 | B1+2 | A4+5 | C6+7 | D4+5 |
| 13-17 | C1+2 | C4+5 | D1+2 | B3+4 | D6+7 |
| 17-19 | C3 | A3 | D3 | B5 | |

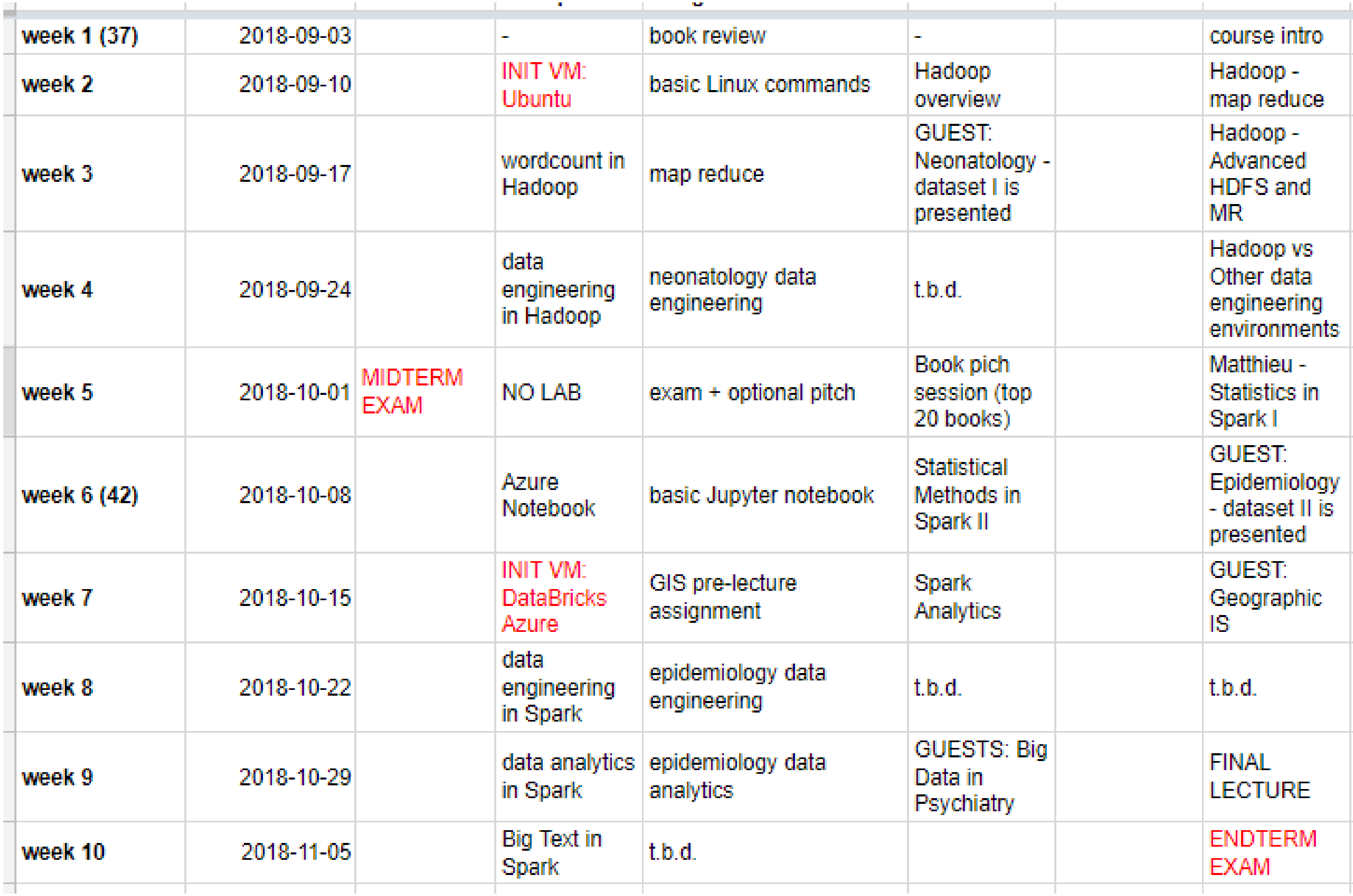
*Teachers:*

| form | group | time | week | room | teacher |
|---|---|---|---|---|---|
| innovatie | | | | | Marco Spruit |
| lecture | | Tue 15.15-17.00 | 37-44 | RUPPERT-PAARS | Marco Spruit<br>Matthieu Brinkhuis |
| | | Thu 11.00-12.45 | 36 | UNNIK-GROEN | |
| | | | 37-44 | RUPPERT-PAARS | |
| tutorial | group 1 | Tue 13.15-15.00 | 37-44 | RUPPERT-111 | #SSOI<br>Melchior Keijdener |
| | group 2 | Tue 13.15-15.00 | 37-44 | RUPPERT-B | |

# Course Communication: Teams

- Join http://bit.ly/infomdss-teams
- ... NOW; with your SolisID/UUemail

# Putting it together... Course Schema *(subject to updates)*



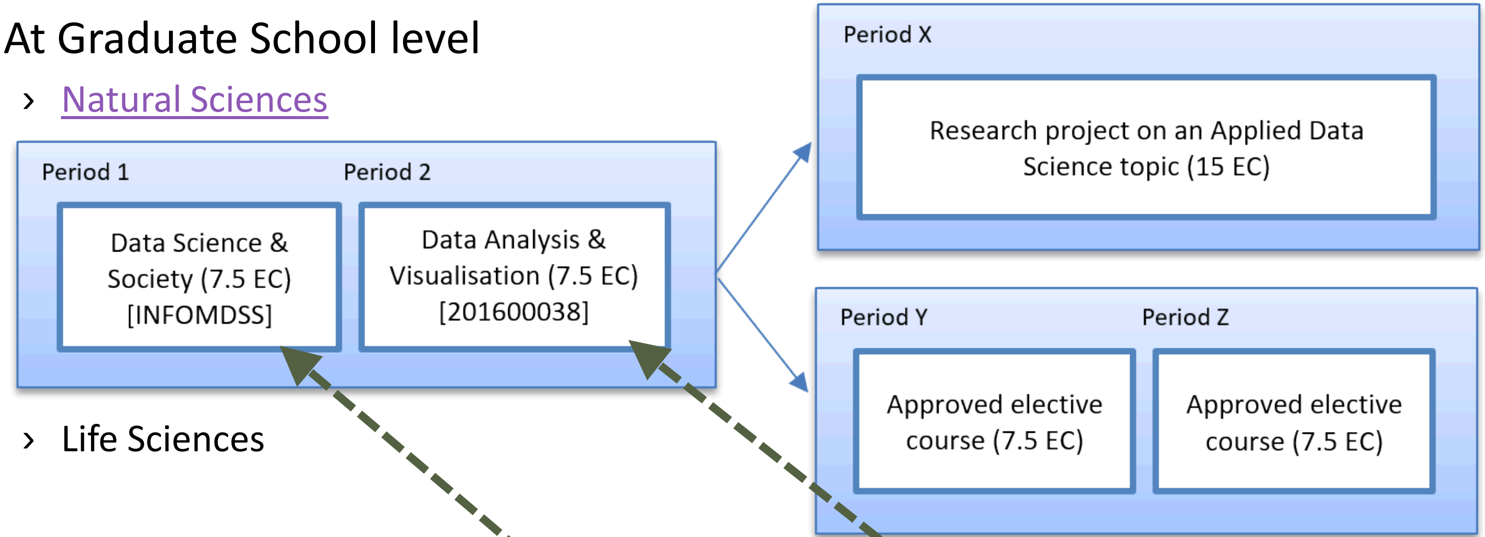
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| week 1 (37) | 2018-09-03 | | - | book review | - | | course intro |
| week 2 | 2018-09-10 | | INIT VM: Ubuntu | basic Linux commands | Hadoop overview | | Hadoop - map reduce |
| week 3 | 2018-09-17 | | wordcount in Hadoop | map reduce | GUEST: Neonatology - dataset I is presented | | Hadoop - Advanced HDFS and MR |
| week 4 | 2018-09-24 | | data engineering in Hadoop | neonatology data engineering | t.b.d. | | Hadoop vs Other data engineering environments |
| week 5 | 2018-10-01 | MIDTERM EXAM | NO LAB | exam + optional pitch | Book pich session (top 20 books) | | Matthieu - Statistics in Spark I |
| week 6 (42) | 2018-10-08 | | Azure Notebook | basic Jupyter notebook | Statistical Methods in Spark II | | GUEST: Epidemiology - dataset II is presented |
| week 7 | 2018-10-15 | | INIT VM: DataBricks Azure | GIS pre-lecture assignment | Spark Analytics | | GUEST: Geographic IS |
| week 8 | 2018-10-22 | | data engineering in Spark | epidemiology data engineering | t.b.d. | | t.b.d. |
| week 9 | 2018-10-29 | | data analytics in Spark | epidemiology data analytics | GUESTS: Big Data in Psychiatry | | FINAL LECTURE |
| week 10 | 2018-11-05 | | Big Text in Spark | t.b.d. | | | ENDTERM EXAM |

# 3. About this field

Positioning Data Science & Society

# Applied Data Science @UU

- Master’s profile Applied Data Science
  - At Graduate School level
    - Natural Sciences



Period 1

Data Science & Society (7.5 EC) [INFOMDSS]

Period 2

Data Analysis & Visualisation (7.5 EC) [201600038]

Period X

Research project on an Applied Data Science topic (15 EC)

Period Y

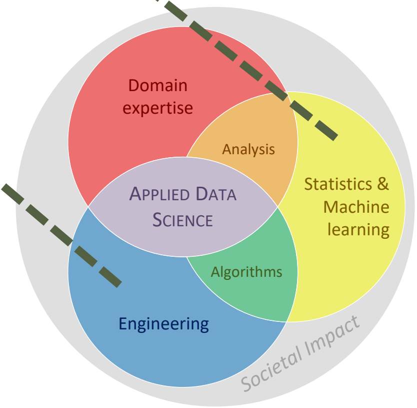
Approved elective course (7.5 EC)

Period Z

Approved elective course (7.5 EC)

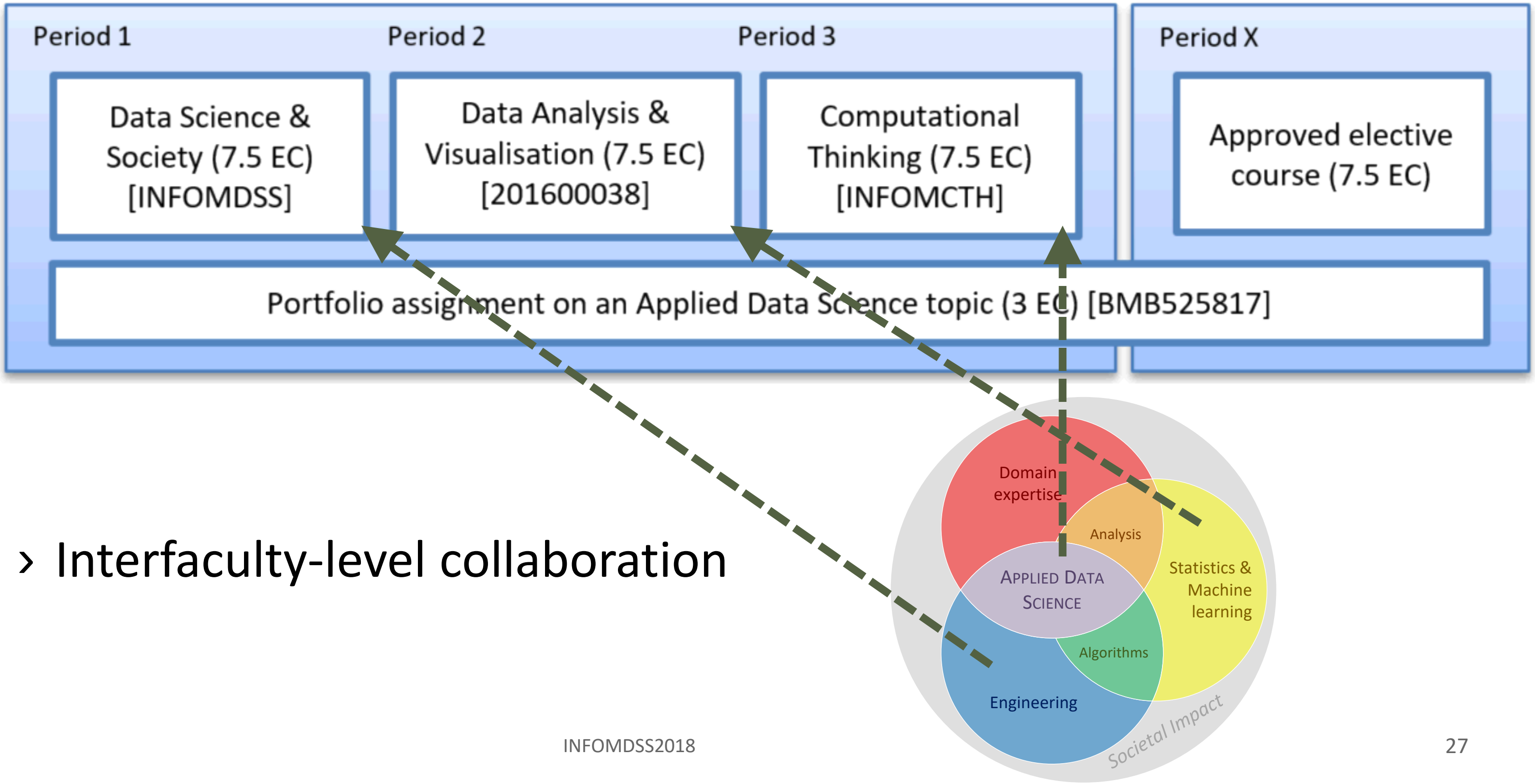
    - Life Sciences

- Interfaculty-level collaboration



# Applied Data Science @UU

- Master’s profile Applied Data Science
  - At Graduate School level
    - Natural Sciences
    - Life Sciences



Period 1

Period 2

Period 3

Period X

Data Science & Society (7.5 EC) [INFOMDSS]

Data Analysis & Visualisation (7.5 EC) [201600038]

Computational Thinking (7.5 EC) [INFOMCTH]

Approved elective course (7.5 EC)

Portfolio assignment on an Applied Data Science topic (3 EC) [BMB525817]

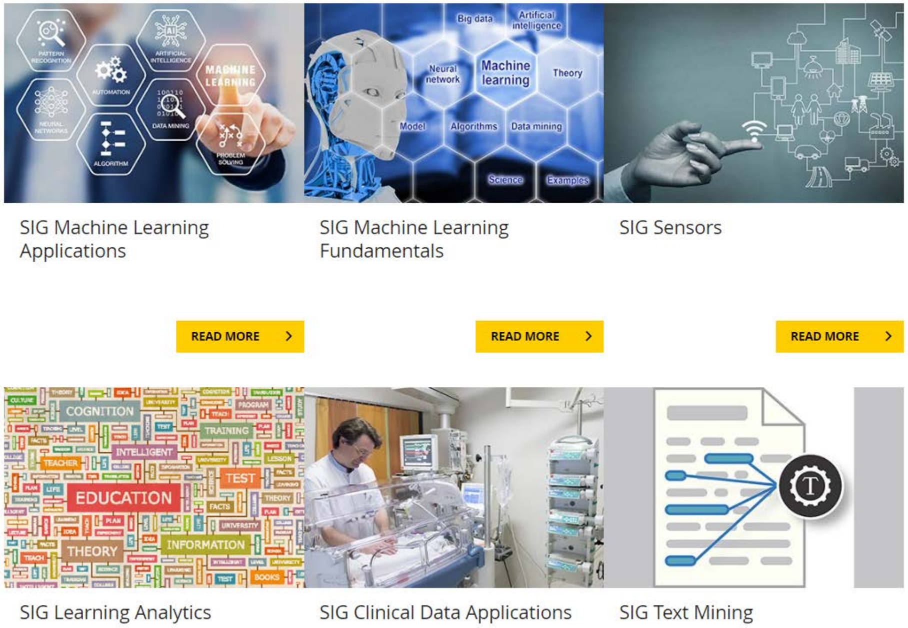
- Interfaculty-level collaboration

# Positioning DS&S course @UU

| Course | Comparison |
|---|---|
| INFOMDM (Data Mining) | - Master level<br>- Computer science →<br>  - algorithmics, complexity, etc |
| INFOMPR (Pattern Recognition) | - ditto |
| Big Data (INFOMBD) | - COSC Master level course on big data analytics/algorithmics |
| Business Intelligence (INFOMBIN) | - Business Analytics through Datawarehousing (structure++) |

# Applied Data Science @UU

- Master’s profile Applied Data Science: ✓
- Focus area Applied Data Science: ... ✓
  - – https://www.uu.nl/ads
- Postgraduate MSc programme Applied Data Science in Health: ... ✓
- Community: Applied data science SIGs: ... ✓



SIG Machine Learning Applications

SIG Machine Learning Fundamentals

SIG Sensors

SIG Learning Analytics

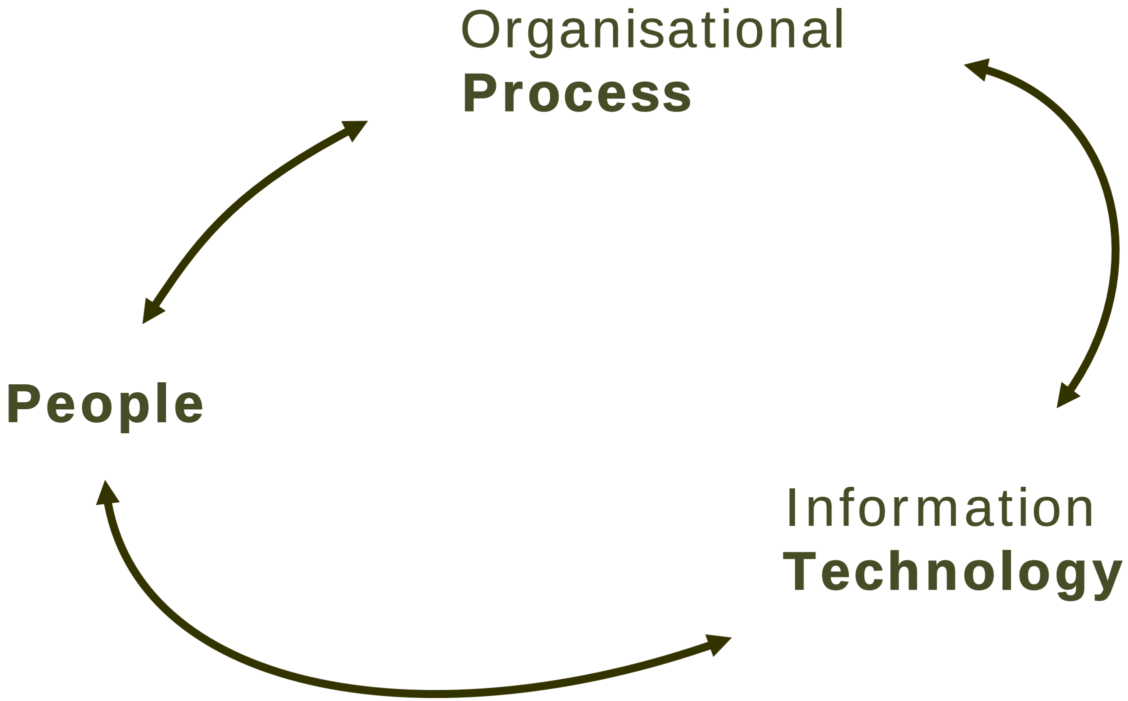
SIG Clinical Data Applications

SIG Text Mining

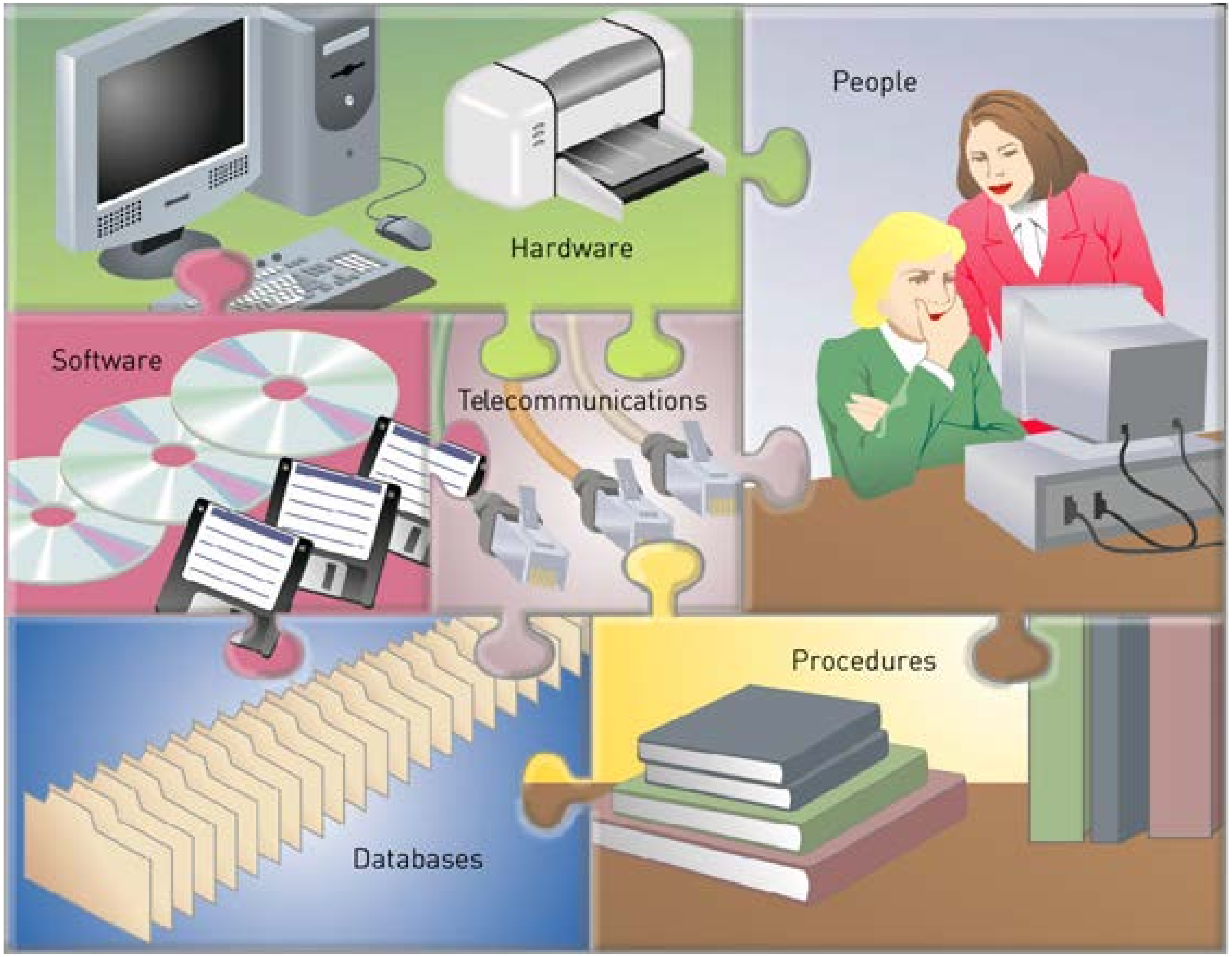
# Builds on topics from bachelor Informatiekunde...

| | *Information Science CORE* | | |
|---|---|---|---|
| ***Start:*** | Data Modelling | Mobile Programming | Information Science Project |
| ***Basis:*** | Scientific Research Methods | | Modelling and System Development |
| | *Information Science PATHS* | | |
| ***Study path:*** | *Organisations & Society* | *Games & Interaction* | *Life Sciences & Health* |
| ***Start:*** | Organisations & ICT | Design of Interactive Systems | People, Society & ICT |
| | Information Systems | Information Exchange | |
| ***Basis:*** | e-Business | Cognition & Emotion | Data Analytics |
| | Product Software | Usability Engineering | Knowledge Systems |
| ***Deepening:*** | Information Security | Game Design | Persuasive Technologies |
| ***Synthesis:*** | Strategic Management & ICT | Applied Games | Life Sciences & Health Informatics |
| ***Completion:*** | Research Project | | |

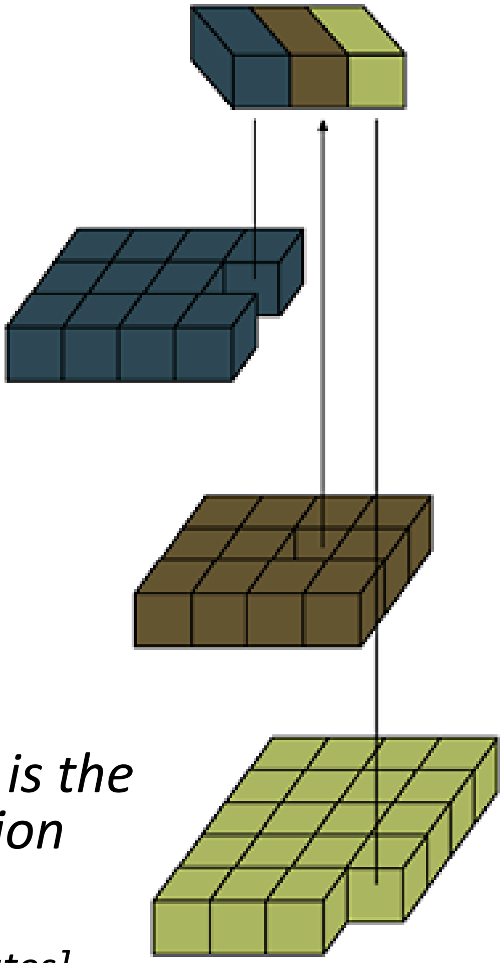
# Applied Data Science vs. Information systems

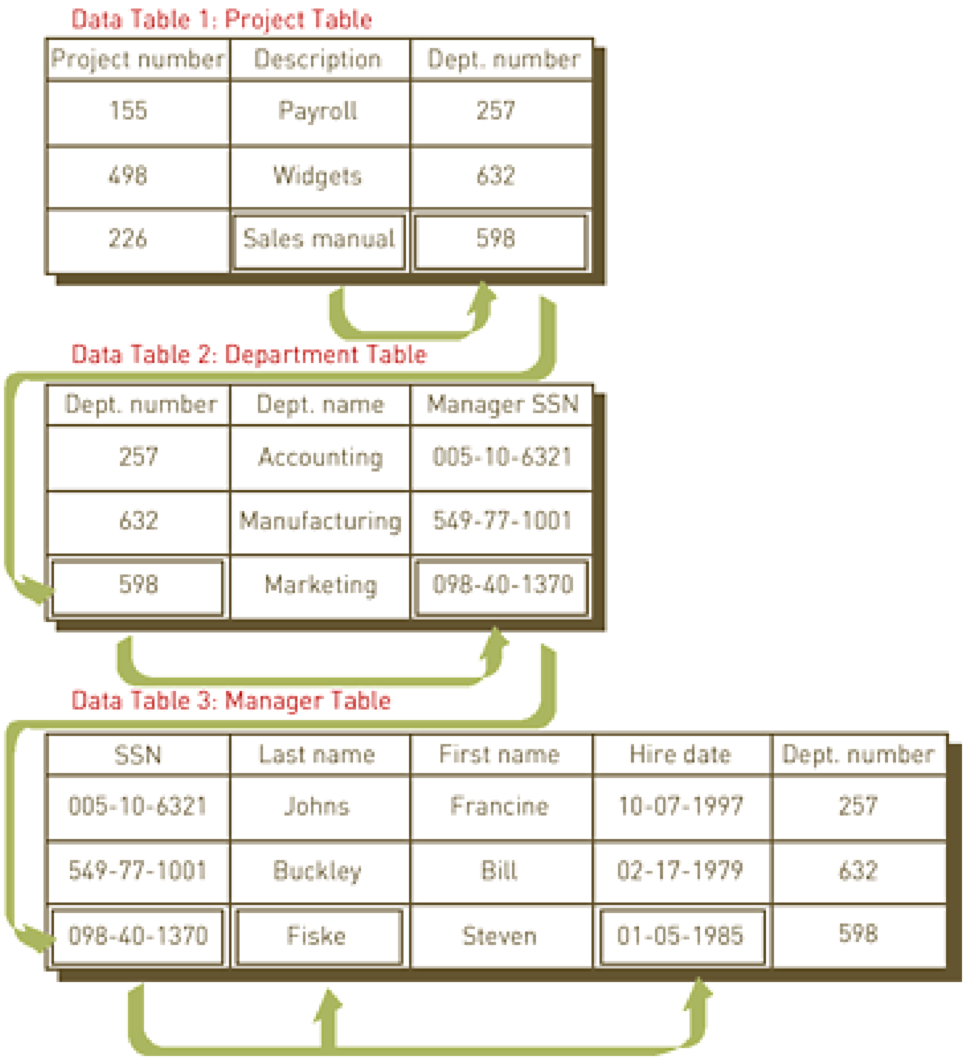


# CBIS: Computer-Based Information System



# Applied Data Science vs. Databases & SQL, in Jeopardy





Data Table 1: Project Table

| Project number | Description | Dept. number |
|---|---|---|
| 155 | Payroll | 257 |
| 498 | Widgets | 632 |
| 226 | Sales manual | 598 |

Data Table 2: Department Table

| Dept. number | Dept. name | Manager SSN |
|---|---|---|
| 257 | Accounting | 005-10-6321 |
| 632 | Manufacturing | 549-77-1001 |
| 598 | Marketing | 098-40-1370 |

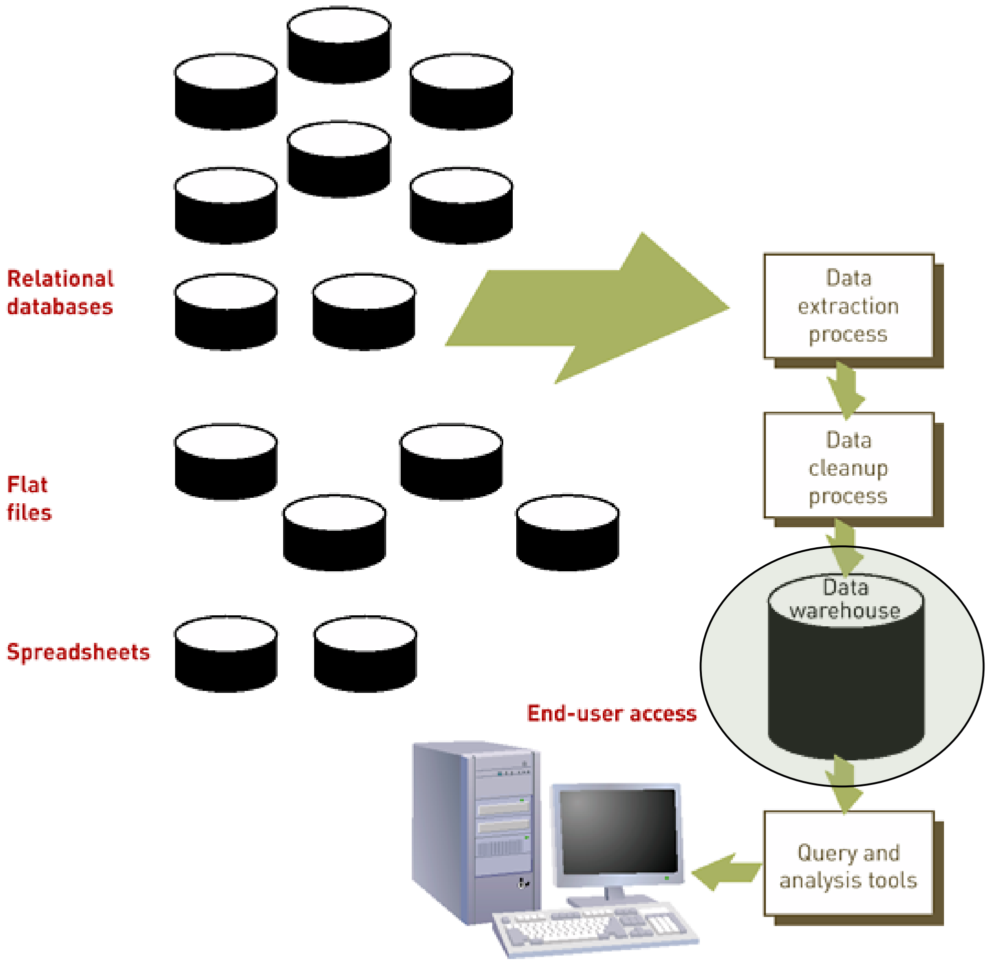
Data Table 3: Manager Table

| SSN | Last name | First name | Hire date | Dept. number |
|---|---|---|---|---|
| 005-10-6321 | Johns | Francine | 10-07-1997 | 257 |
| 549-77-1001 | Buckley | Bill | 02-17-1979 | 632 |
| 098-40-1370 | Fiske | Steven | 01-05-1985 | 598 |

- *What is the question here?*
- *[2 minutes]*

*Who is the manager of the Sales Manual and since when is he on the company's payroll?*

# Big Data vs. Data Warehouses & Data Marts



# KDD: Knowledge Discovery in Databases

- “The KDD process is interactive and iterative, involving numerous steps with many decisions made by the user.”

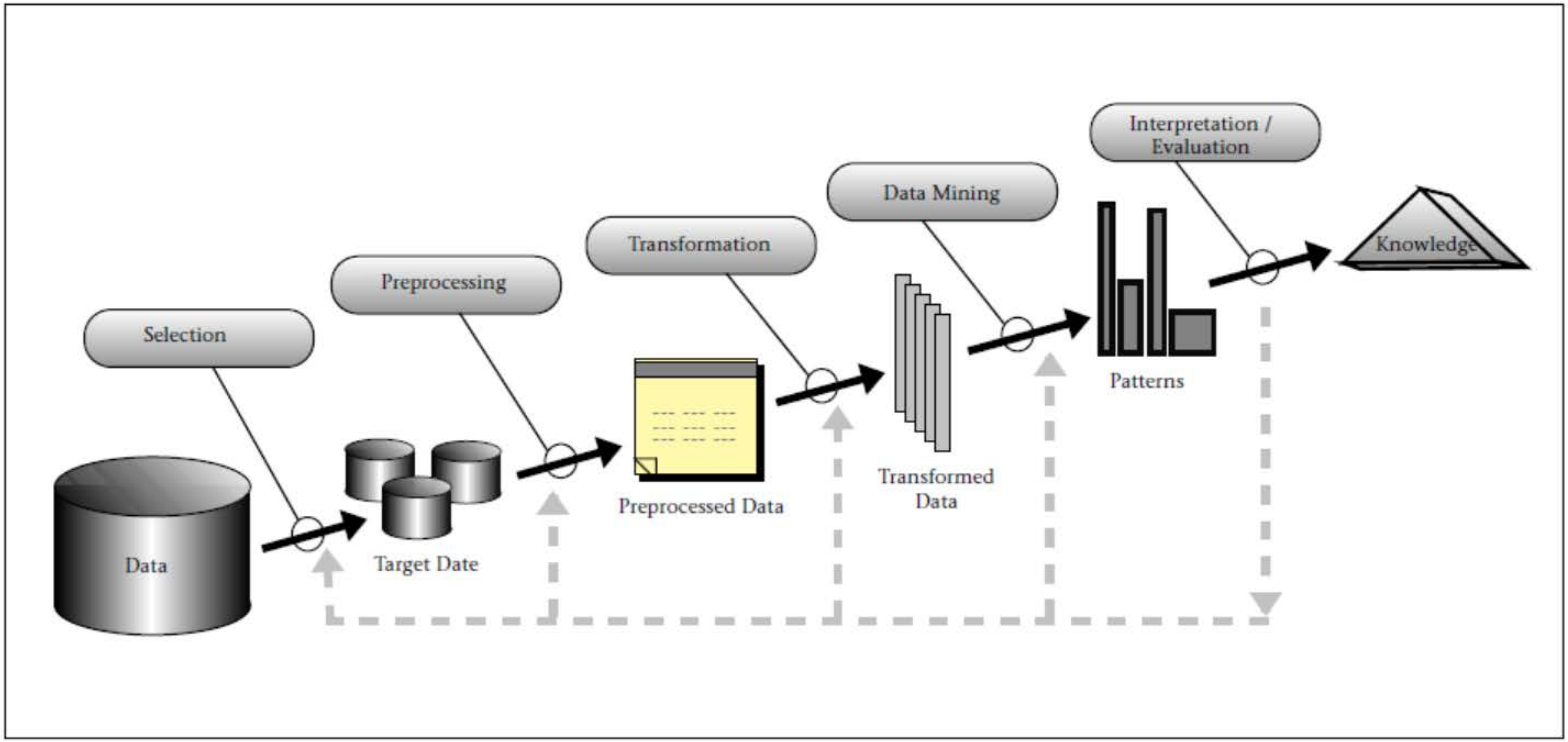
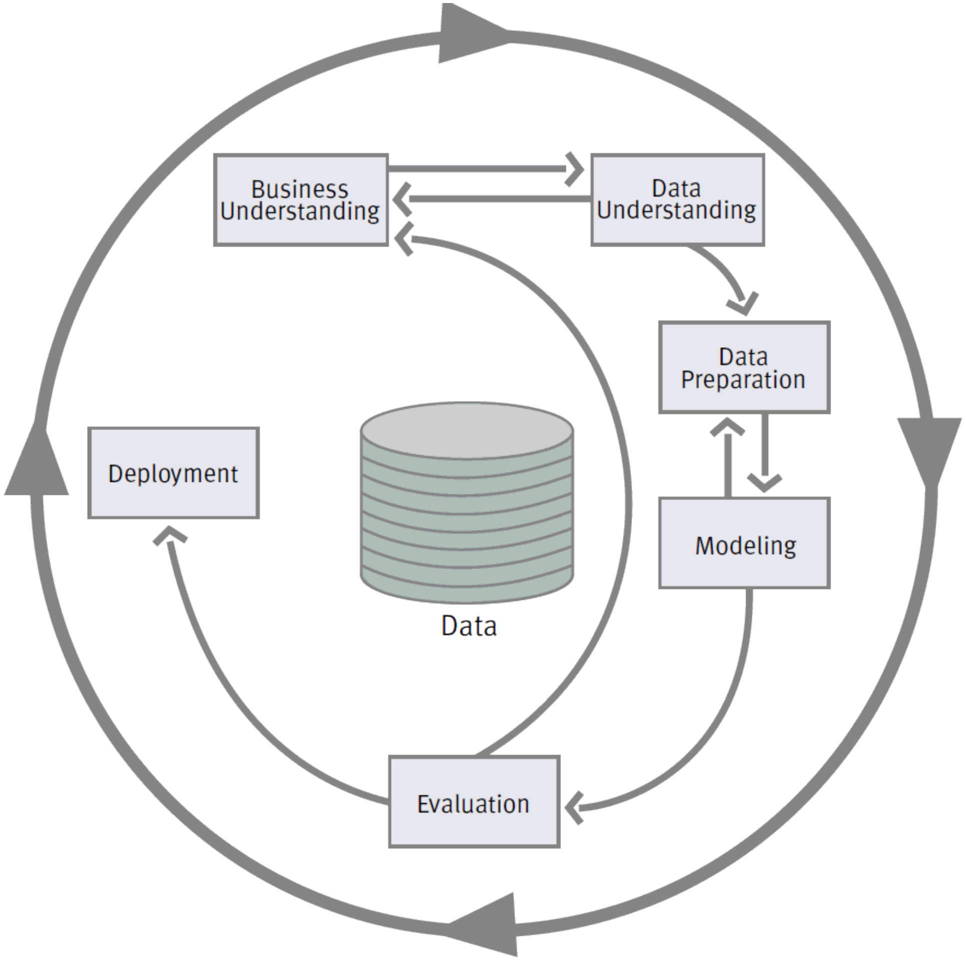


*Figure 1. An Overview of the Steps That Compose the KDD Process.*

# CRISP-DM:
# CRoss-Industry Standard Process for Data Mining

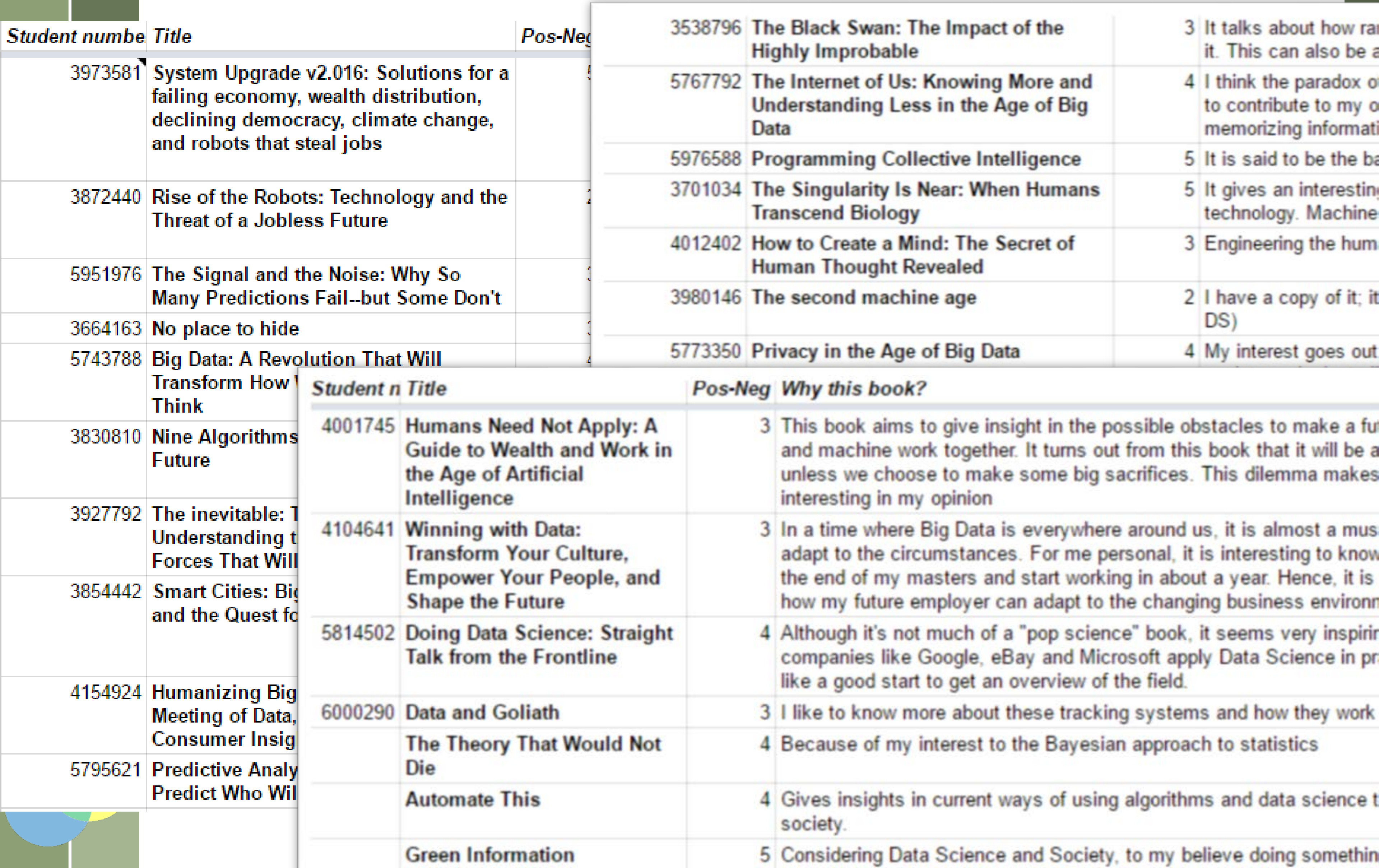


1. Pitch-on-paper assignment
2. Pitch-in-event assignment
3. Tips & tricks
4. Example

# 4. More about Assignment 1

Book review

# Assignment 1: Previously selected books...



| Student numbe | Title | Pos-Neg |
|---|---|---|
| 3973581 | System Upgrade v2.016: Solutions for a failing economy, wealth distribution, declining democracy, climate change, and robots that steal jobs | 5 |
| 3872440 | Rise of the Robots: Technology and the Threat of a Jobless Future | 2 |
| 5951976 | The Signal and the Noise: Why So Many Predictions Fail--but Some Don't | 3 |
| 3664163 | No place to hide | 3 |
| 5743788 | Big Data: A Revolution That Will Transform How Think | 4 |
| 3830810 | Nine Algorithms Future | |
| 3927792 | The inevitable: Understanding t Forces That Will | |
| 3854442 | Smart Cities: Bi and the Quest fo | |
| 4154924 | Humanizing Big Meeting of Data, Consumer Insig | |
| 5795621 | Predictive Analy Predict Who Wil | |

| | | | |
|---|---|---|---|
| 3538796 | The Black Swan: The Impact of the Highly Improbable | 3 | It talks about how ra it. This can also be a |
| 5767792 | The Internet of Us: Knowing More and Understanding Less in the Age of Big Data | 4 | I think the paradox o to contribute to my o memorizing informati |
| 5976588 | Programming Collective Intelligence | 5 | It is said to be the ba |
| 3701034 | The Singularity Is Near: When Humans Transcend Biology | 5 | It gives an interestin technology. Machine |
| 4012402 | How to Create a Mind: The Secret of Human Thought Revealed | 3 | Engineering the hum |
| 3980146 | The second machine age | 2 | I have a copy of it; it DS) |
| 5773350 | Privacy in the Age of Big Data | 4 | My interest goes out |

| Student n | Title | Pos-Neg | Why this book? |
|---|---|---|---|
| 4001745 | Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence | 3 | This book aims to give insight in the possible obstacles to make a fu and machine work together. It turns out from this book that it will be a unless we choose to make some big sacrifices. This dilemma makes interesting in my opinion |
| 4104641 | Winning with Data: Transform Your Culture, Empower Your People, and Shape the Future | 3 | In a time where Big Data is everywhere around us, it is almost a mus adapt to the circumstances. For me personal, it is interesting to know the end of my masters and start working in about a year. Hence, it is how my future employer can adapt to the changing business environ |
| 5814502 | Doing Data Science: Straight Talk from the Frontline | 4 | Although it's not much of a "pop science" book, it seems very inspirir companies like Google, eBay and Microsoft apply Data Science in pr like a good start to get an overview of the field. |
| 6000290 | Data and Goliath | 3 | I like to know more about these tracking systems and how they work |
| | The Theory That Would Not Die | 4 | Because of my interest to the Bayesian approach to statistics |
| | Automate This | 4 | Gives insights in current ways of using algorithms and data science t society. |
| | Green Information | 5 | Considering Data Science and Society, to my believe doing somethin |

# Assignment 1a: Pitch-on-paper

- Submit Top-3 books through Google Form:
  - http://bit.ly/infomdss-form1
- Submit pitch “review” through: http://bit.ly/infomdss-assignment1
  - four open questions, each with a 1-5 rating on “perceived quality”
  - also upload “2-pager pitch” PDF
- Explaining your book’s:

1. Key problem related to data science
2. Described societal impact of this problem
3. Application field of focus
4. Estimated scientific feasibility

- Overall

a) recommendation score

b) set of 10 ordered thematic keywords

# Assignment 1b: Pitch-in-event

- After submission, all 2-pagers will be graded
- The best pitch for each book present a 3-minute pitch in the Pitching lecture on Tue Oct 2
- 0.5 bonus on grading component A if pitch was satisfactory

- *A good pitch satisfies at least the following criteria:*

1. Your pitch provides instant insight into the key message of the book.
2. You *show* (explicitly) what the Big Problem is that this book tackles and what Big Solution it envisions or aims to deliver.

# Pitching tips & tricks

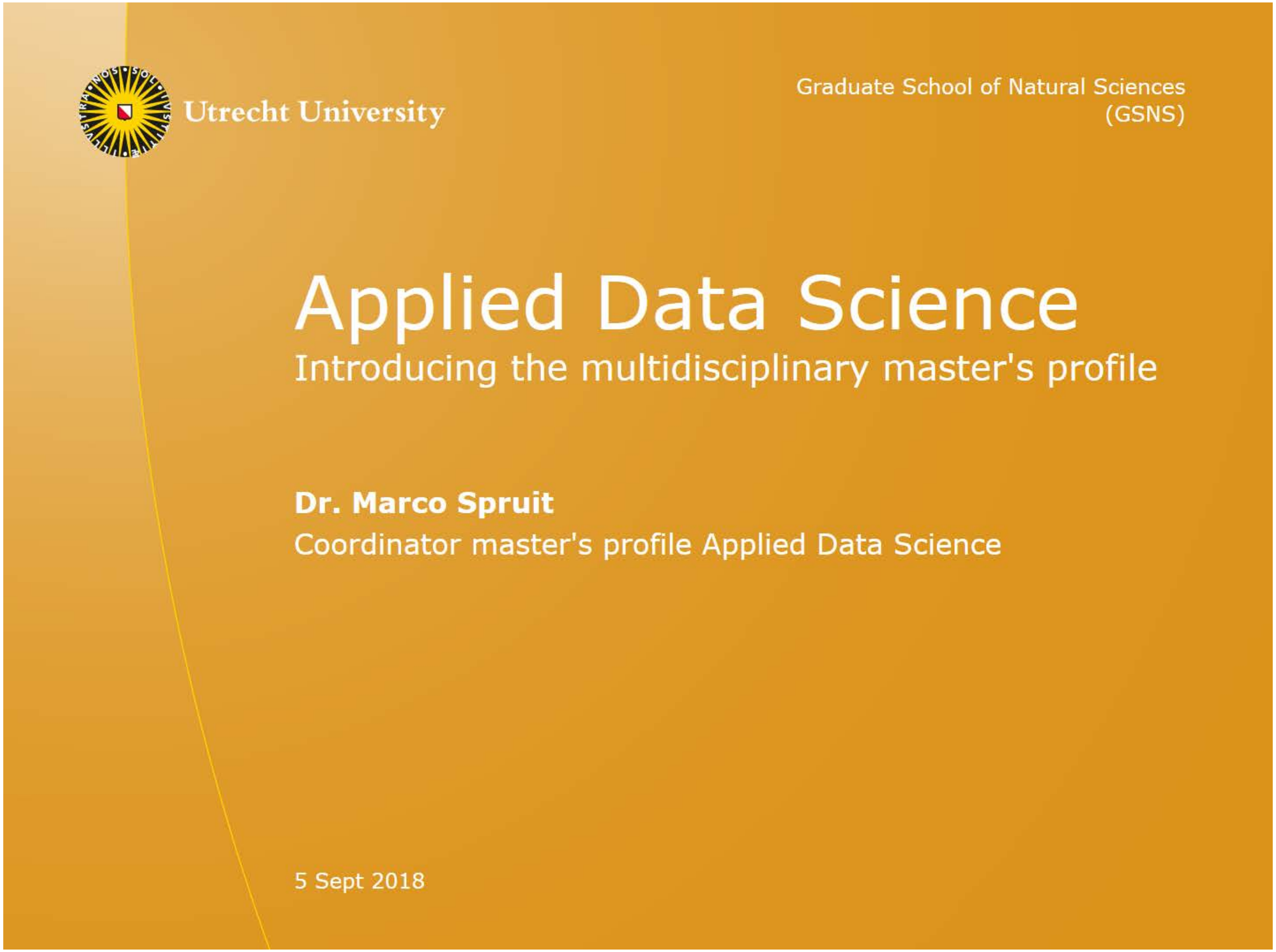
› *Next, you enhance the pitch by PLUS-ing it:*

1. **“Pulling demand”:** your pitch needs to leave the listeners behind with a desire to read the book themselves.
2. **“Lasting”:** your pitch needs to be remembered by your audience. You only have max. 3 minutes to make a good impression. Start with a catchy opening (e.g. a question or anecdote), and also introduce yourself very briefly, and why you chose this book.
3. **“Undeniable”:** Be very clear. Be ‘to-the-point’.
4. **“Simple”:** The pitch should be very light on technical language or jargon. This pitch should be understandable by your grandparents at one of your family parties!

# Pitching tips & tricks

- *Presenting without pictures – The perfect presentation in eight steps*

1. The Law of the three Ps:
    - Prepare, prepare, prepare!
2. Your objective: Touch your audience
3. Trigger emotions
4. Use your body language
5. The magic of the number three
6. Come up with one-liners
7. Deviate from the standard
8. Conform to best practices



Utrecht University

Graduate School of Natural Sciences
(GSNS)

# Applied Data Science

## Introducing the multidisciplinary master's profile

**Dr. Marco Spruit**

Coordinator master's profile Applied Data Science

5 Sept 2018

See http://bit.ly/infomdss-lit

# Week 01: Literature

- Davenport, T. H., & Patil, D. J. (2012). Data scientist: The Sexiest Job of the 21st Century. *Harvard business review, 90*(5), 70-76.
- Stair, R. & Reynolds, G. (2012). *Fundamentals of Information Systems.* Sixth Edition.
  **NOTE: Chapters 1 and 3 ONLY.** Cengage: Boston, MA. ISBN-13: 978-0-8400-6218-5. *(other more recent editions are also fine).*
- Chapman, P. Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step Data Mining Guide*. [@IBM]
- Pritzker, P., and May, W. (2015). *NIST Big Data interoperability Framework (NBDIF): Volume 1: Definitions.* NIST Special Publication 1500-1. Final Version 1. National Institute of Standards and Technology.
- Spruit, M., & Lytras, M. (2018). Applied Data Science in Patient-centric Healthcare: Adaptive Analytic Systems for Empowering Physicians and Patients. *Telematics and Informatics, 35*(4), 643–653.