Data Science in Practice:
A longterm healthcare case study

@Nederlandse Vereniging voor Medisch Onderwijs (NVMO) 18 May 2021

Leiden University Medical Center

Prof.dr. Marco Spruit
Professor Advanced Data Science in Population Health
Public Health and Primary Care

0031 6
m.r.spruit@lumc.nl
LUMC Campus Den Haag, Turfmarkt 99, 2511 DP THE HAGUE
LIACS, Niels Bohrweg 1, 2333 CA LEIDEN

About: Marco Spruit

Engineer

1993

- Information Retrieval programmer, ZyLAB Europe

1995

- Big Data system developer, Dutch Military Intelligence & Security Service (MIVD)

1997

- Product software developer/entrepreneur, Insertable Objects & Wizzer BV

Researcher

2003

- Ph.D researcher in Computational Linguistics, University of Amsterdam

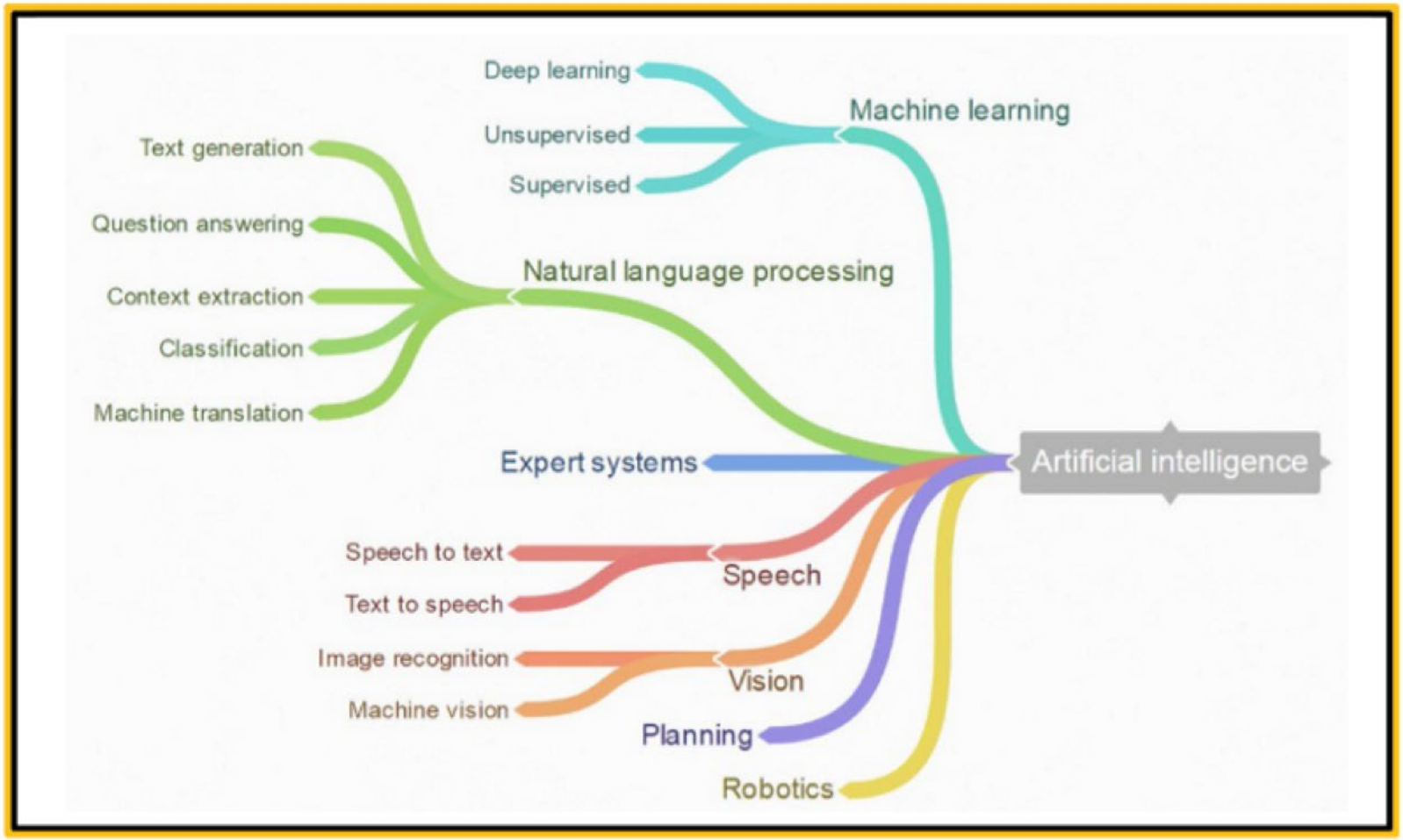
2007

- Assistant/Associate professor Information Science, Utrecht University
  - Applied Data Science Lab

2020

- Professor Advanced Data Science in Population Health, LUMC/Leiden University
  - Population Health Living Lab
  - CAIRELab

Data Science vs Artificial Intelligence

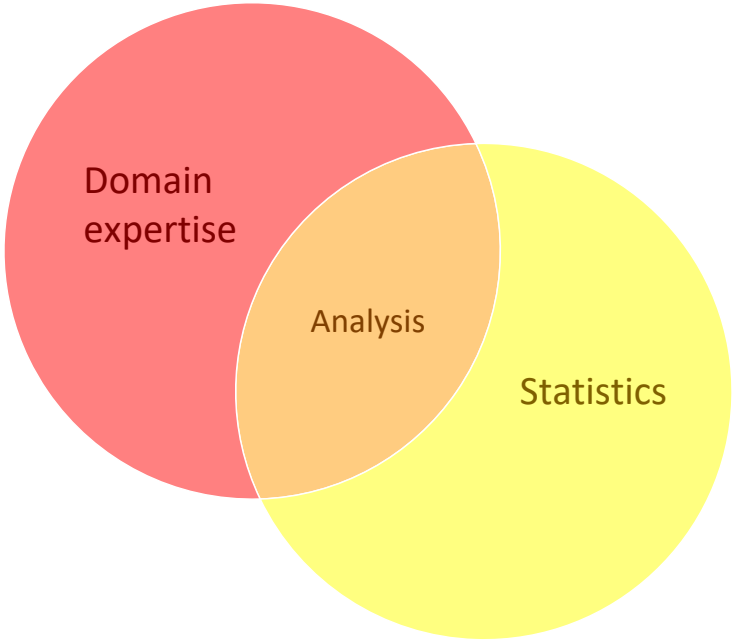


https://thenewstack.io/machine-learning-made-easy/

"Machine Learning is an approach to Achieve Artificial Intelligence"

What is... Science?

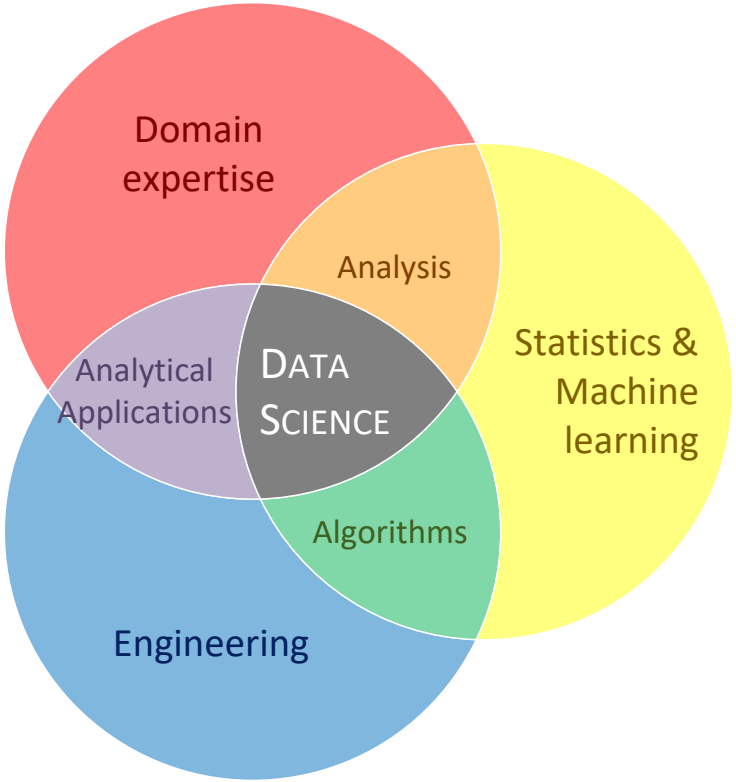
1. Domain expertise
2. Statistics



Source: NIST SP 1500-1 NIST Big Data interoperability Framework (NBDIF): Volume 1: Definitions, September 2015 [online]

What is... Data Science?

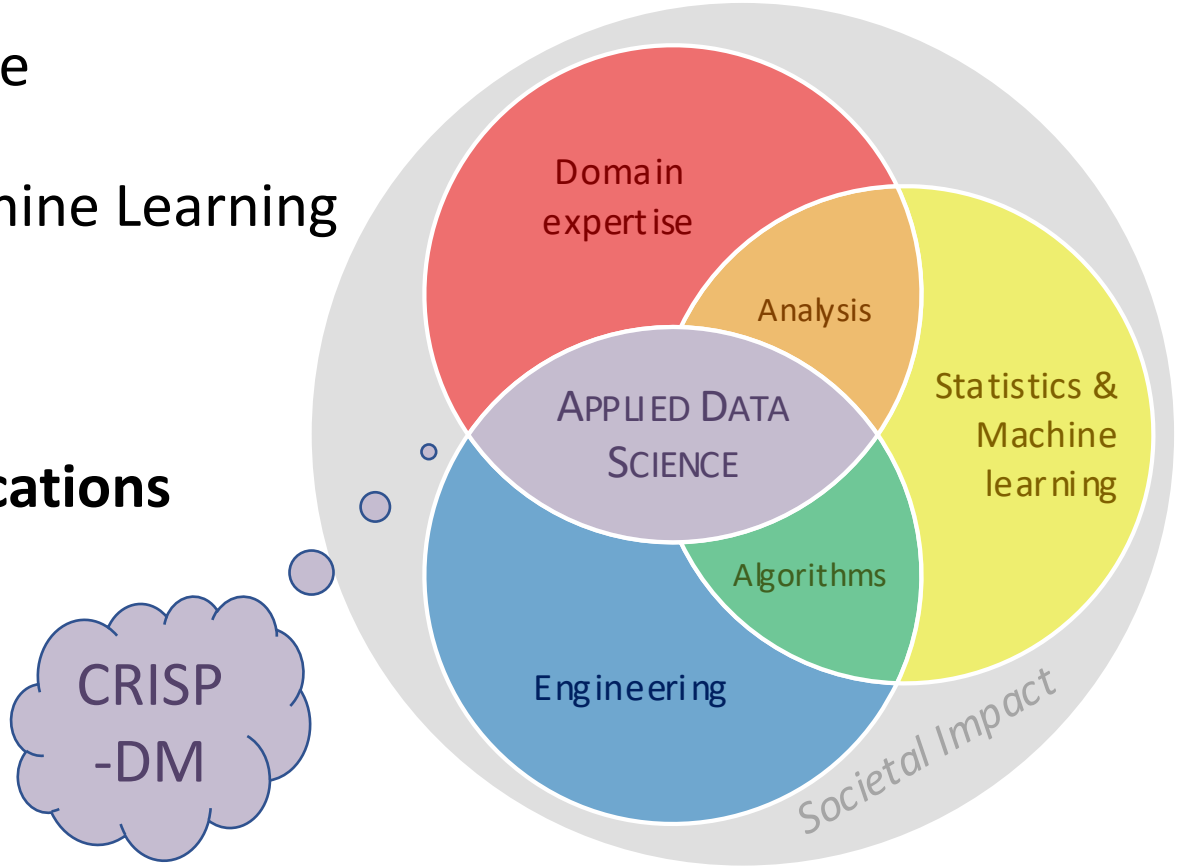
1. Domain expertise
2. Statistics & Machine Learning
3. Engineering
4. **Analytical applications**



Source: NIST SP 1500-1 NIST Big Data interoperability Framework (NBDIF): Volume 1: Definitions, September 2015 [online]

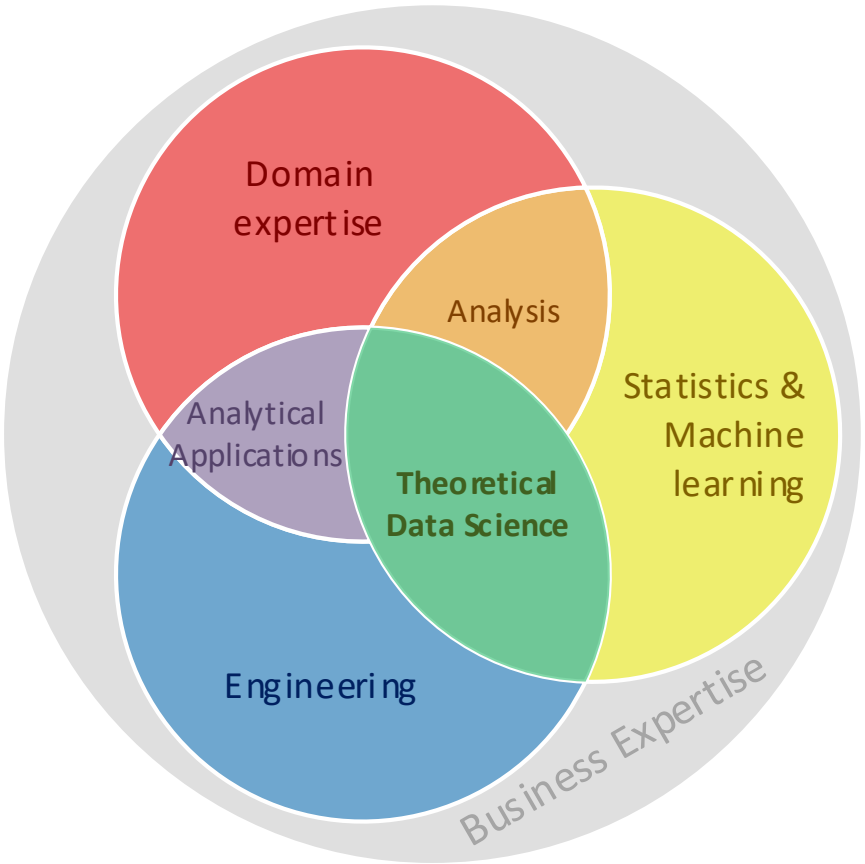
What is... Applied Data Science?

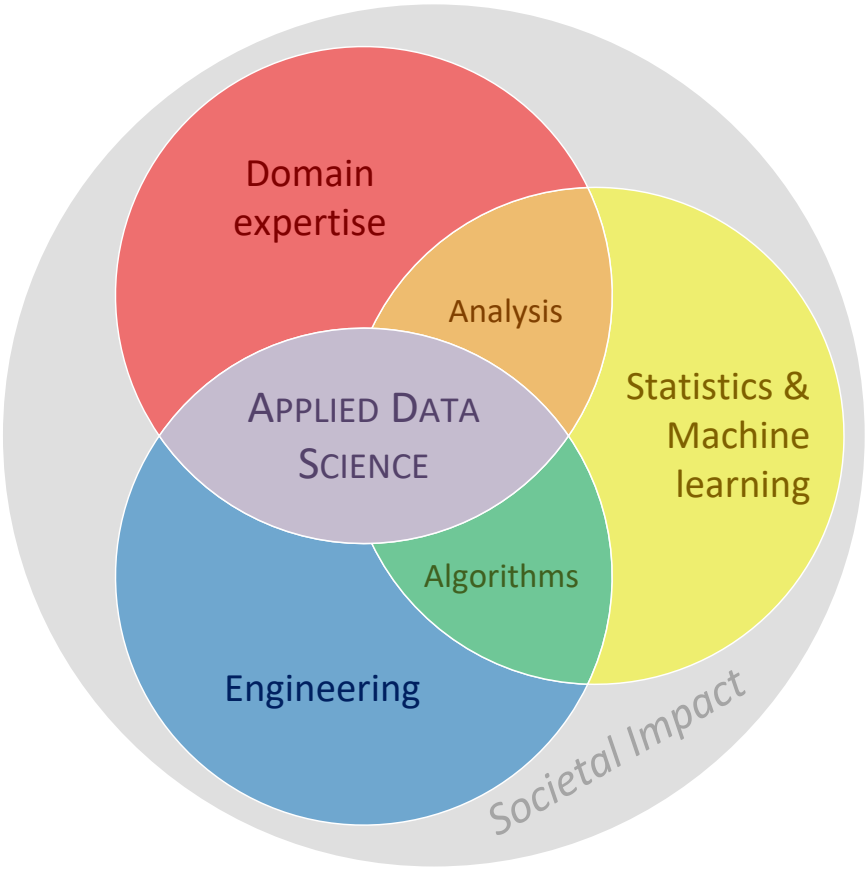
1. Domain expertise
2. Statistics & Machine Learning
3. Engineering
4. **Analytical applications**
5. *Societal impact*



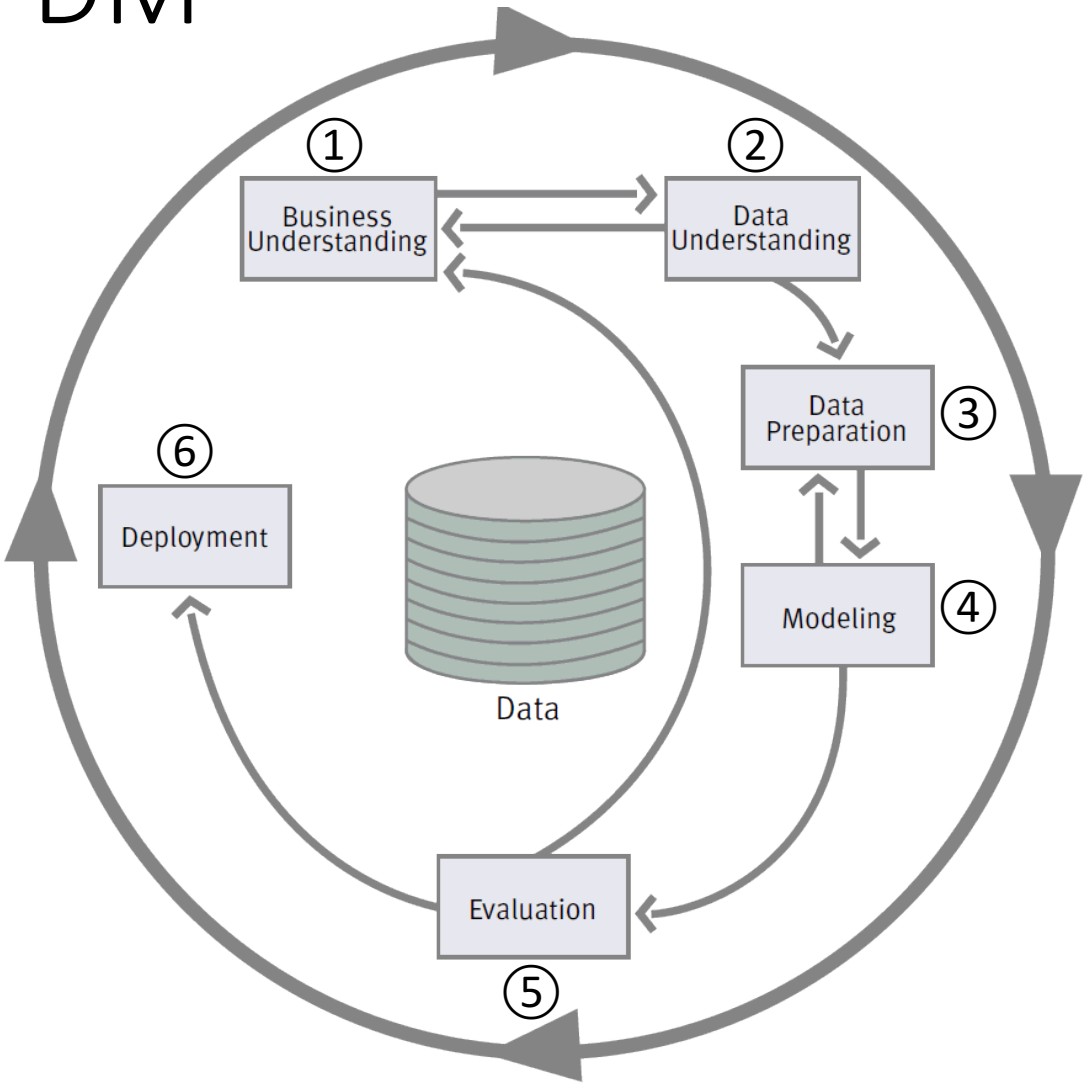
Spruit,M., & Jagesar,R. (2016). Power to the People! Meta-algorithmic modelling in applied data science. In Fred,A. et al. (Ed.), *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 400–406). KDIR 2016, November 11-13, 2016, Porto, Portugal: ScitePress. [pdf] [online]

Applied Data Science compared





CRISP-DM



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. [ftp]

From information needs to data mining goals

Spruit,M., Vroon,R., & Batenburg,R. (2014). Towards healthcare business intelligence in long-term care: an explorative case study in the Netherlands. *Computers in Human Behavior, 30*, 698–707. [pdf] [online]

Why focus on Understanding?

'Massale opslag metadata telefoongegevens Amerikanen leverde vrijwel niets op'

26-02-2020

46

De vier jaar durende massale opslag van metadata van telefoongesprekken die Amerikanen hebben gevoerd, heeft uiteindelijk vrijwel niets opgeleverd. Dat blijkt uit een rapport van een overheidsinstantie over het project van inlichtingendienst NSA.

Naar aanleiding van informatie uit de massale opslag van metadata van telefoontjes is er slechts één onderzoek gestart door een andere inlichtingendienst. Dat schrijft The New York Times naar aanleiding van een rapport van een overheidsinstantie dat naar het Amerikaanse Congres is gegaan. Een andere keer leverde het programma ook bruikbare informatie op, maar dat leidde uiteindelijk niet tot een onderzoek. Alle andere keren kregen inlichtingendiensten alleen informatie die ze al hadden.

De NSA stopte vorig jaar al met het verzamelen van de metadata. Toen al zei de inlichtingendienst dat de dataverzameling te veel geld kostte en te weinig bruikbare informatie opleverde. In totaal heeft het project honderd miljoen dollar gekost, schrijft The New York Times.

NSA heeft vier jaar lang metadata van telefoontjes van Amerikanen bewaard. De bedoeling is om de wet later dit jaar aan te passen om de dataverzameling niet langer mogelijk te maken. Het rapport van de overheidsinstantie is niet openbaar.

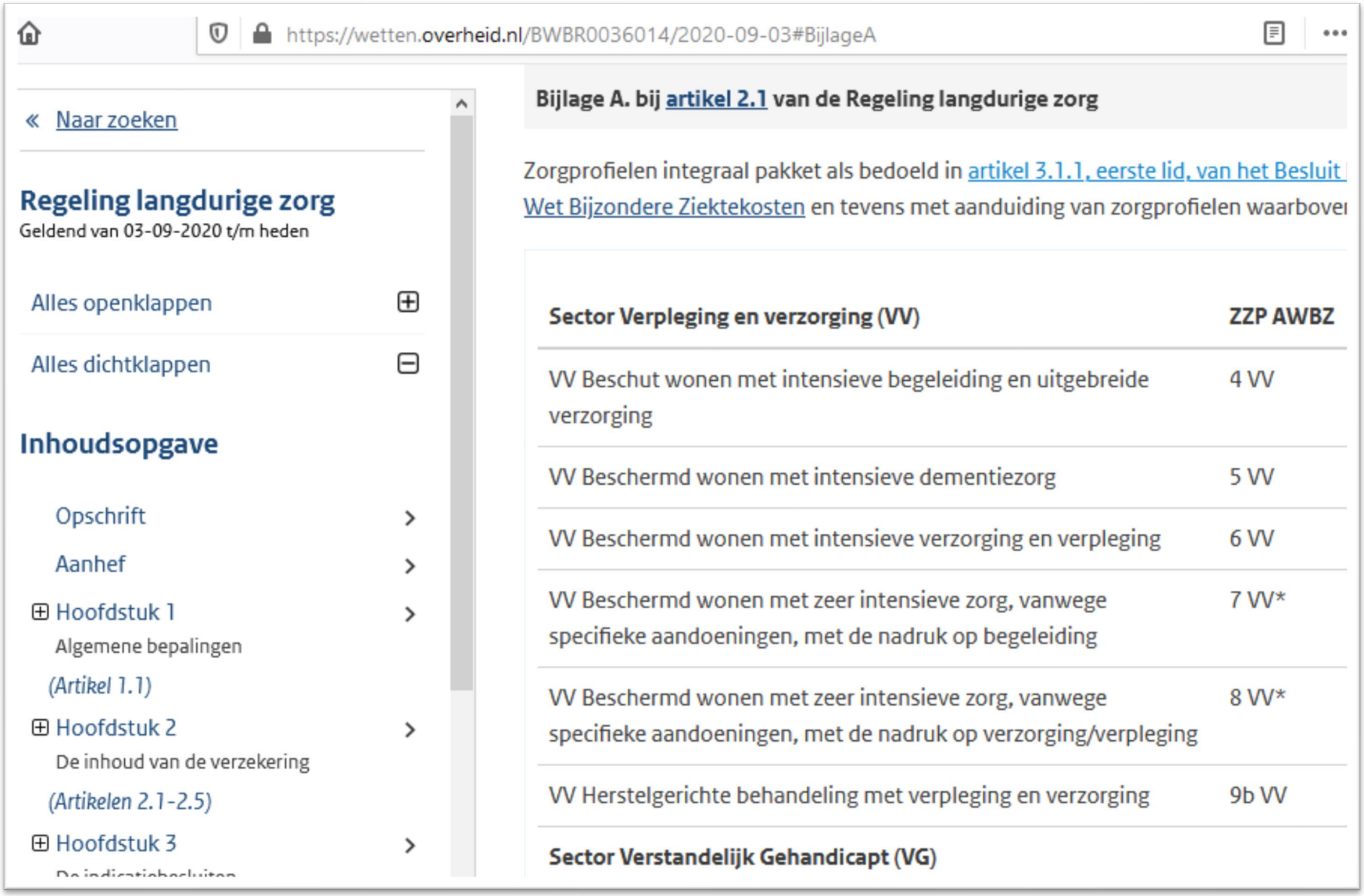
https://tweakers.net/nieuws/163922/massale-opslag-metadata-telefoongegevens-amerikanen-leverde-vrijwel-niets-op.html

Introduction (1/2)

- Long-term care
  - Care for people with a long-term or chronic disorder
  - Relatively unexplored
- Main goals for long-term care
  - Care of good quality
  - Acceptable cost level
- One of the biggest expenses of Dutch government
  - 38% of the total healthcare budget
  - €14 billion in 2000 → €27 billion (budgeted) in 2012

Introduction (2/2)

- Care Intensity Package (ZZP)
  - Introduced in 2009
  - Different levels of care intensity
    - ZZP1: Extramural living with some guidance
    - ZZP8: Intramural living under full surveillance and 24/7 care
  - Operational cost depends on ZZP level
  - 2015: *Wet langdurige zorg* (Wlz) : ZZP ⇨ Wlz/VV *Zorgprofiel*
- Electronic Client Record data
  - Large valuable dataset, but
  - Not fully exploited at this moment



2015: Wet langdurige zorg (WLZ) ← Algemene Wet Bijzondere Ziektekosten (AWBZ)

Research Method: CRISP-DM

- CRoss Industry Standard Proces for Data Mining
  - *Phase 1:* Business understanding
    - 18 unstructured in-depth interviews
      - 8 (board of) directors experts
      - 7 management experts
      - 7 experts from stakeholders perspective
        - (MinVWS, IGZ, Care insurer)
- Result
  - 56 information needs derived from the 18 unstructured in-depth interviews with 22 field experts

Calculation: Information Needs

- 56 information needs
  - 33 related to quality of care
  - 23 related to financial state
- Information needs scored based on a valuation equation
  - Board members: 10
  - Managers: 6
  - Stakeholders: 3
- Score = $\sum_{Expert\ level} \frac{Times\ mentioned}{Number\ of\ interviews} \times Valuation$
- Score = $\left(\frac{8}{8} \times 10\right) + \left(\frac{4}{5} \times 6\right) + \left(\frac{3}{5} \times 3\right) = 16.6$

What do you think?

What was by far the #1 information need in this long-term care sector?

Calculation: Top 10 Needs

#	Type	Information need	Board	Mgmt	Stakeh.	Score
1	Q	Customer experience	8	7	10	16.6
2	F	Staffing with respect to ZZP-mix	7	4	2	14.8
3	F	ZZP-mix per business unit	7	4	0	13.6
4	F	ZZP-mix prognoses	7	4	0	13.6
5	F	Staffing with respect to operations	6	4	2	13.5
6	Q	Number of incidents occured	6	4	2	13.5
7	Q	Types of incidents occured	6	4	2	13.5
8	Q	Causes of occured incidents	6	4	2	13.5
9	F	Operations per ZZP	7	3	1	13.0
10	F	Production information (planned, realized, declared)	7	3	1	13.0

Sidenote: Alternatively…

- Automating Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding

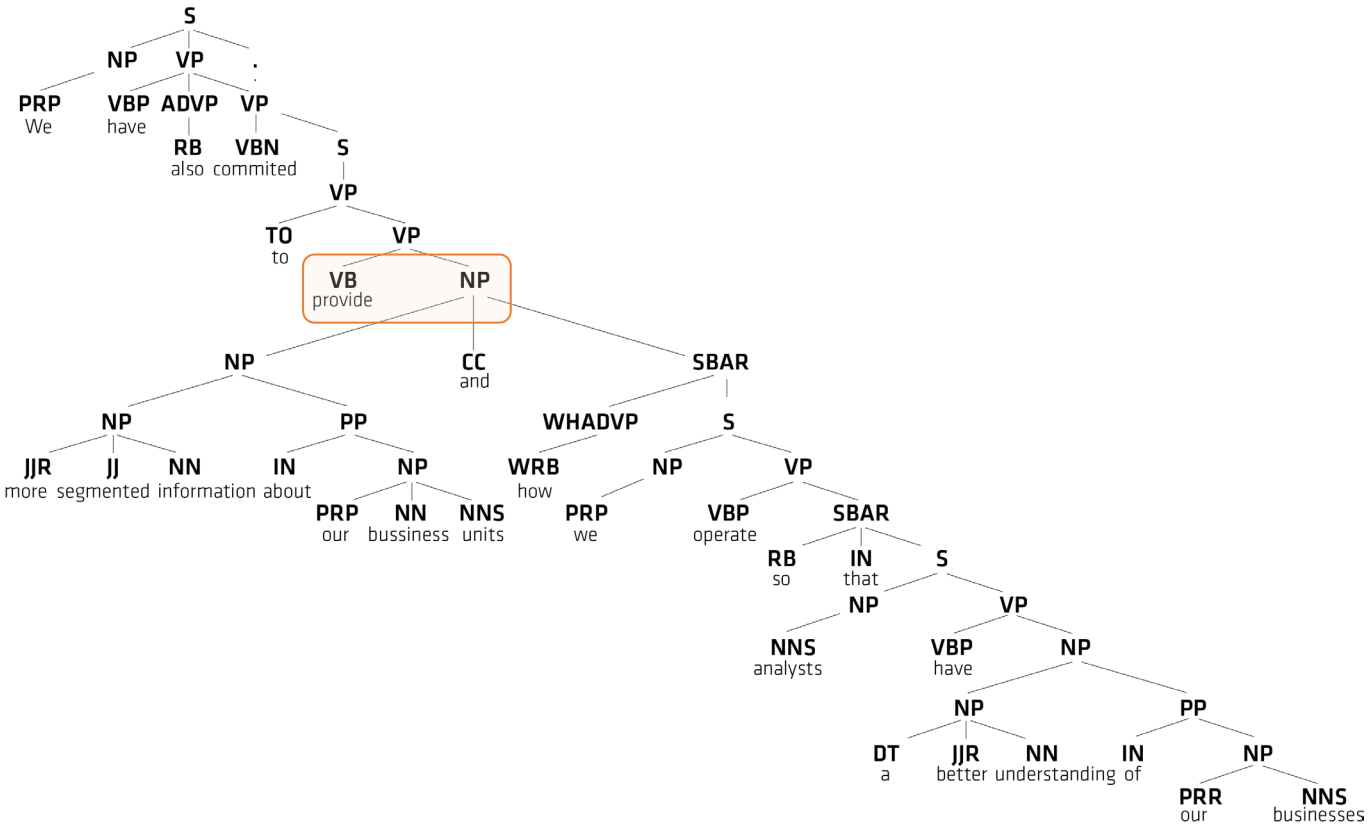
  (Spruit, Kais, Menger, under review)

Verb keywords	*Noun keywords*
make, improve, increase, promote, develop, formulate, prepare, reduce, maintain, administer, guarantee, offer, prolong, endorse, manage, obtain, avoid, block, prevent, achieve, require, lack, ensure, motivate, decrease, reduce, enhance, enable, support, provide, aim	objective, aim, purpose, goal

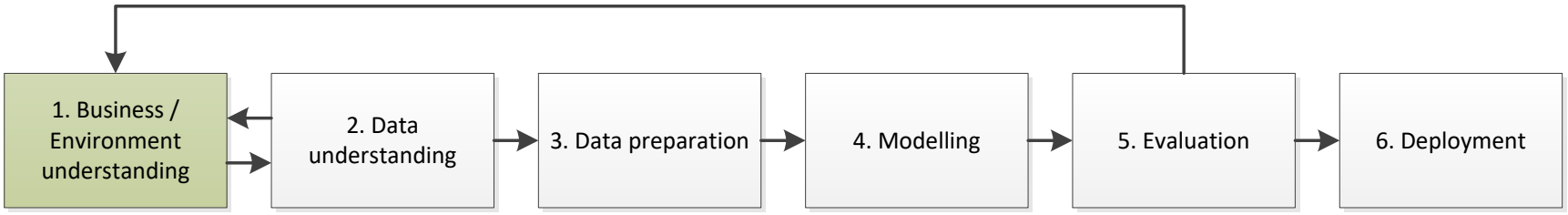
Sidenote: Alternatively…

- Automating Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding

  (Spruit, Kais, Menger, under review)

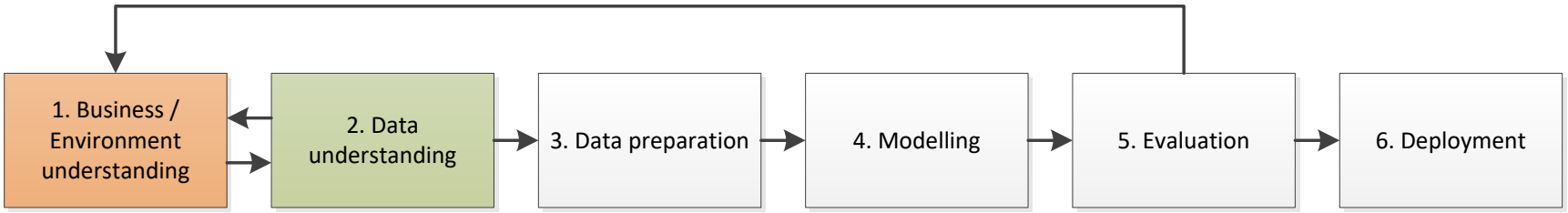


Findings: Business understanding



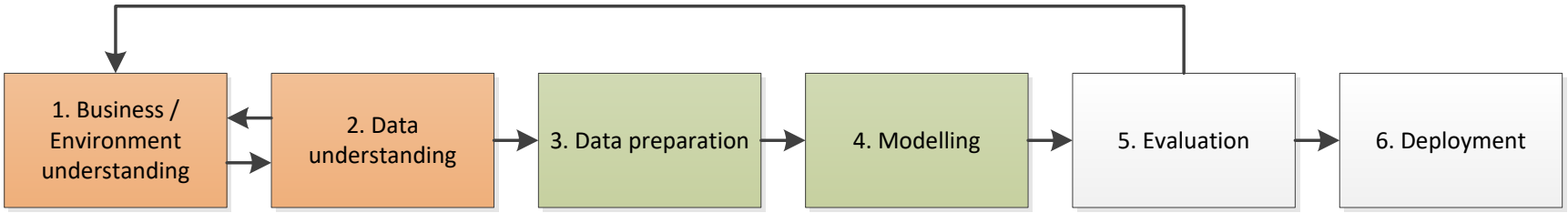
Information needs	Data mining goals
• Number of occured incidents • Types of occured incidents • Causes of the occured incidents • Patterns in occured incidents	• Identify patterns in incidents
• Number of clients at an increased risk • Types of risk the clients run	• Identify relationships in risk assessment
• Progress of care-related measures	• Identify patterns in care-related measures
• Treatment goals (obtained & not-obtained) • Care plan information	• Identify patterns in obtained and not-obtained treatment goals
• Number of clients per demand for care • ZZP-mix • ZZP-mix prognosis	• Identify & predict the ZZP mix

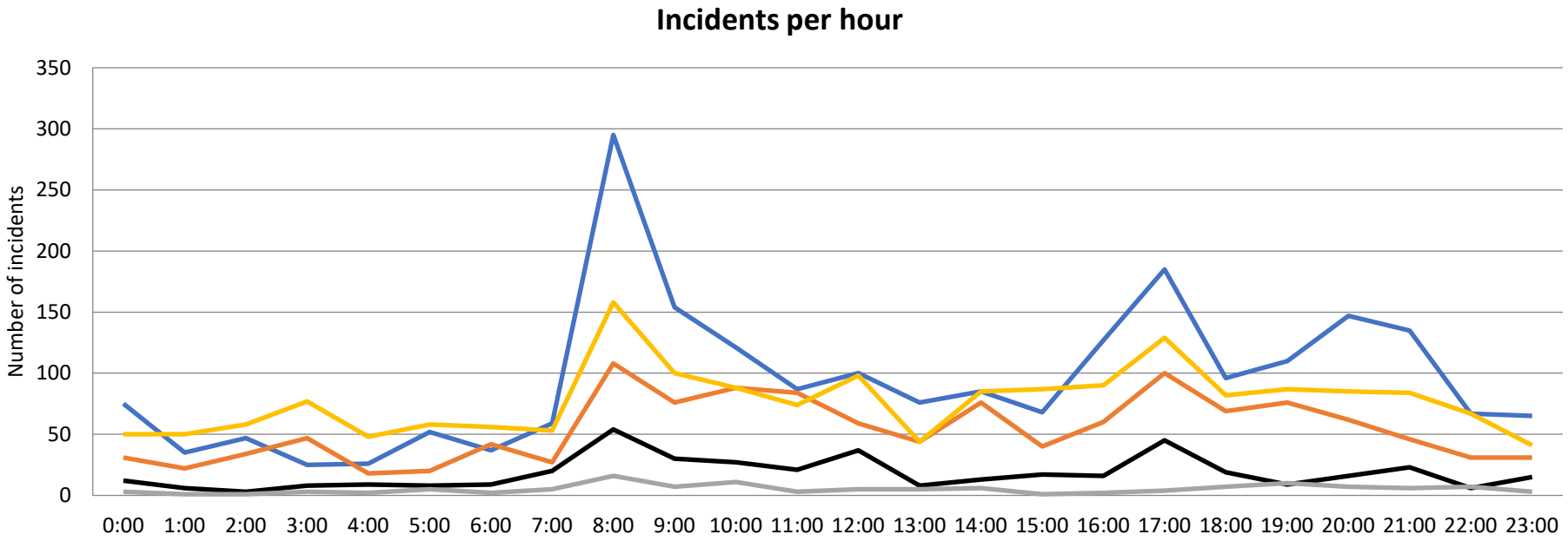
Findings: Data understanding



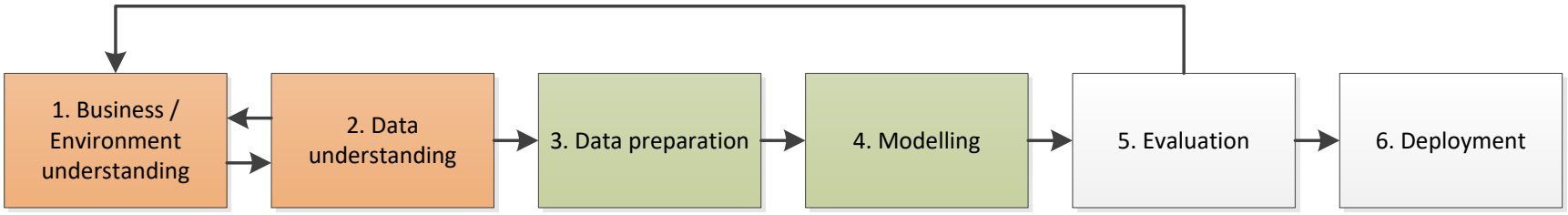
Data mining goals	Available data
• Identify patterns in incidents	• 5,692 records with incidents • 13 different incident types
• Identify patterns in risk assessment	• Depression risk: 2,129 records • Falling risk: 889 records • Incontinence risk: 877 records • Medication risk: 806 records • Problem behaviour risk: 0 records • Weight risk: 567 records
• Identify patterns in care-related measures	• 27,174 records with care-related measures
• Identify patterns in obtained and not-obtained treatment goals	• 20,725 records with treatment goals • 14.59% obtained goals • 85.41% not obtained goals
• Identify & predict the ZZP mix	• 1,831 records with historical delivered ZZP's

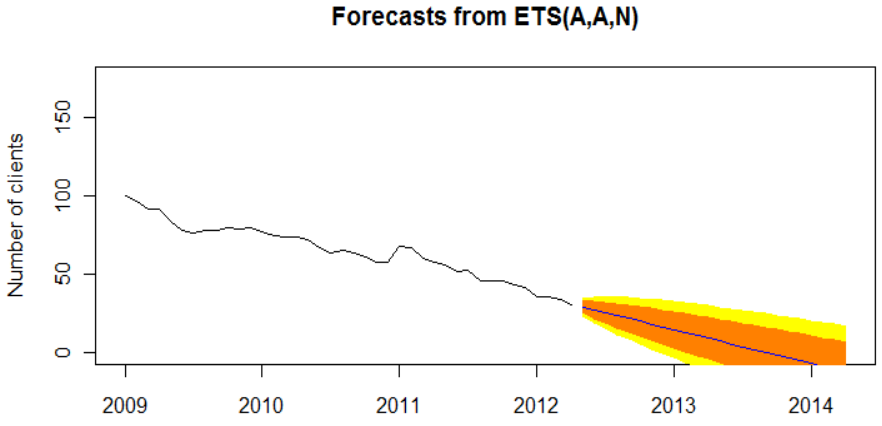
Findings: Identify patterns in incidents



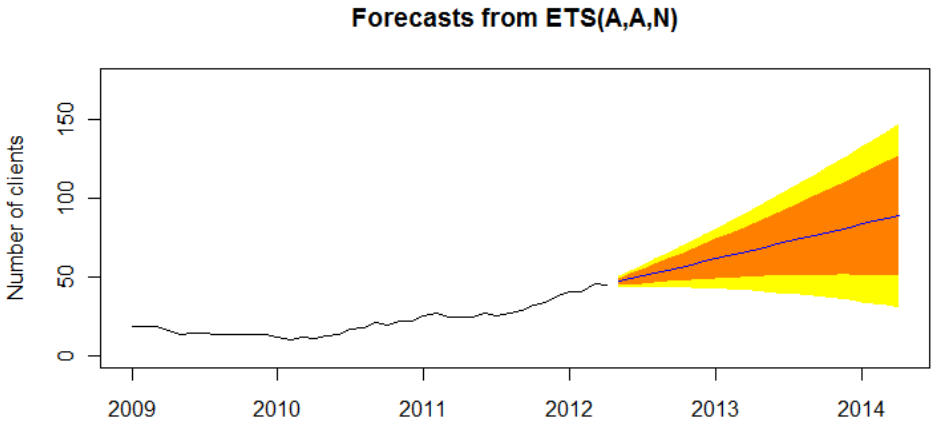


Findings: Predict the ZZP mix



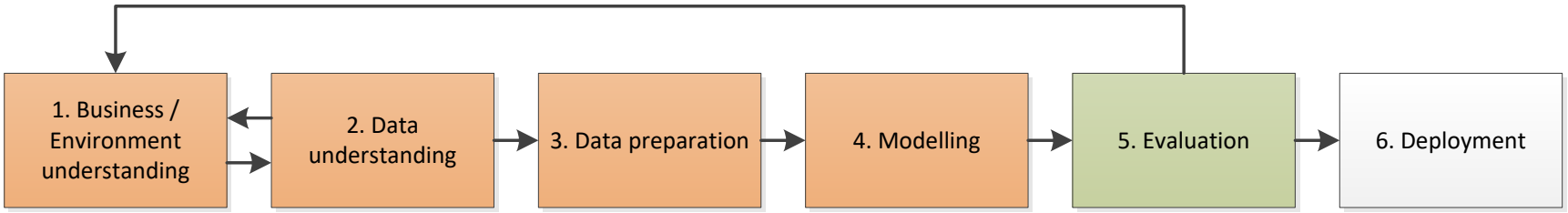


ZZP1



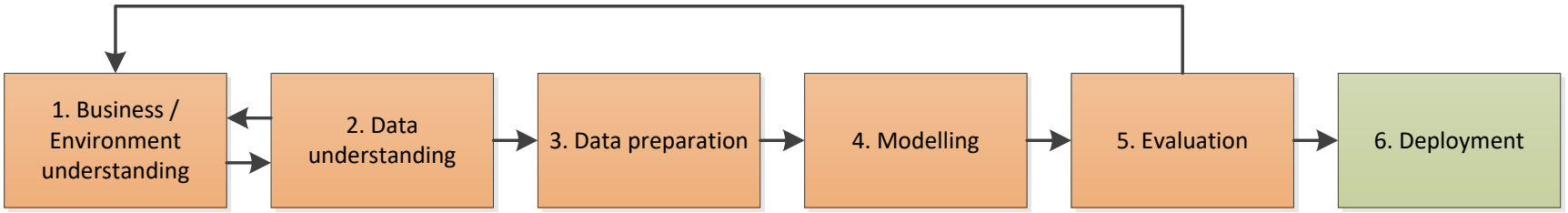
ZZP6

Findings: Evaluation



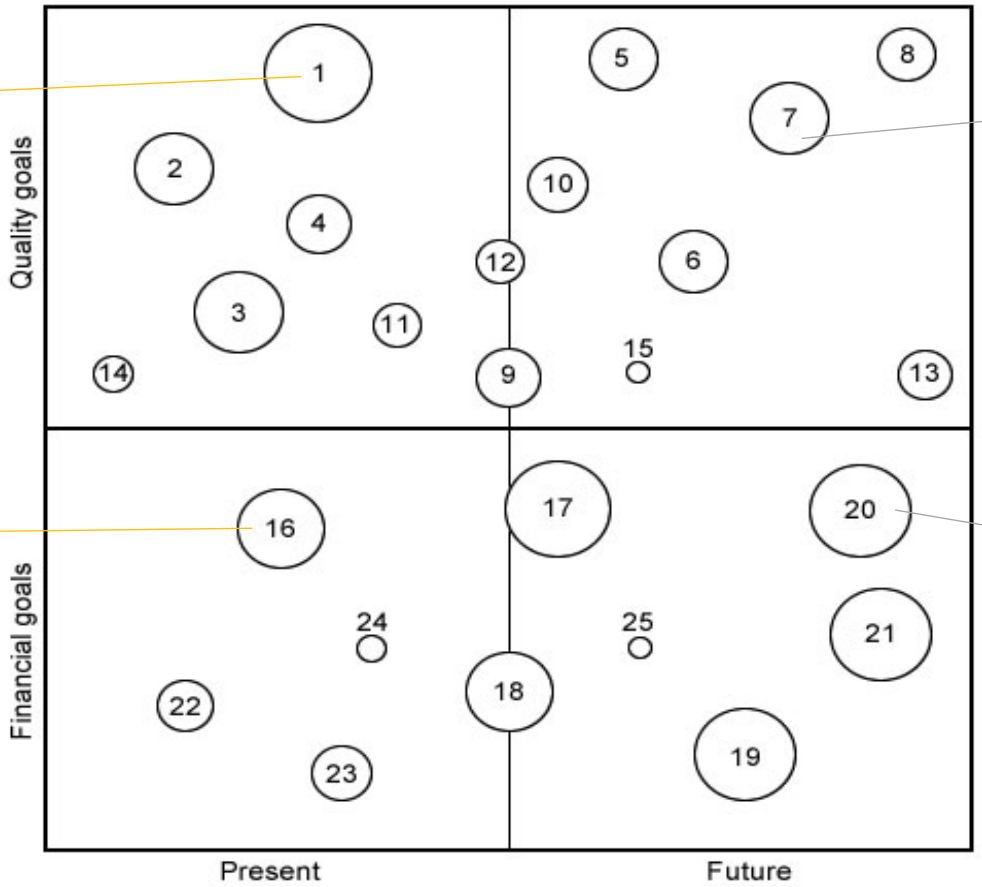
- Explorative research → no groundbreaking results
  - Information needs of *multiple* care institutions & stakeholders
- Lack of standardization
  - Heterogeneous data
- Predictions are too dependent on environmental factors
  - Limited historical data
  - Too many dependent factors for the forecasts to be practically useful
    - Changes in laws and regulations have direct effect on analysis
- Qualitative data (*e.g.* free text input)
  - Complex to analyse

Findings: Deployment



Identify the patterns in incidents

Identify & predict the ZZP-mix



Relationship between care-related measures and incidents

Identify care within & outside ZZP indication (planned, realized)

Universiteit Leiden

Search for subject or person | All categories | Nederlands | English

Research | Education | Academic staff | About us | Collaboration | Faculties | Campus The Hague | Alumni | Library

Home > Education > Study Programmes > Population Health Management (MSc)

Wíth a Data Science track

- Population Health Management (MSc)
  - Why Population Health Management at LUMC-Campus The Hague?
  - About the programme
  - Career prospects
  - Admission and application
  - Student life
  - Information activities
  - Contact

Population Health Management (MSc)

Today's health care system faces many challenges. The new interdisciplinary two years master PHM will train you to contribute to a more integrated health care system.

What does this master's programme entail?

Health care faces ageing populations, rising health care costs, fragmented health care supply and advancing medical technologies and IT systems. Health care professionals will require new competences to meet these challenges in the current health care system. Population Health Management (PHM) is a broadly based response to the challenges and has emerged worldwide as an important strategy for health care.

Interested?

Facts & Figures

Title
MASTER OF SCIENCE

Language
ENGLISH

City
THE HAGUE

Number of first-year students
25

Faculty
MEDICINE/LUMC

Duration
TWO YEARS

https://www.universiteitleiden.nl/en/education/study-programmes/master/population-health-management