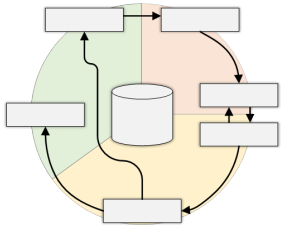


Translationele Data Science & AI

“Oók voor de Langdurige Zorg!”

Prof.dr. Marco Spruit & TDS team | Universiteit Leiden | AI 4 Topaz Inspiratiesessie | 11 nov 2025
https://www.universiteitleiden.nl/en/staffmembers/marco-spruit

Agenda

	Programma AI 4 Topaz inspiratie sessie 11/11/2025		
Tijd	**Onderdeel**		**Duur**
		Voorbeelden van AI Technologiëen in de Langdurige Zorg	
14:30	[Context]	Het *Translational Data Science & AI* Lab (Marco)	20
14:50	[Proces]	Van Informatiebehoeften naar Datadoelen in de Langdurige Zorg (Ma	20
15:10	[NLP]	Welzijn.AI voor Kwetsbare Ouderen met Demo (Bram)	30
	[Spraak]		
	[Robotica]		
15:40	[Hands-on]	Samen experimenteren met Welzijn.AI voor Topaz (Bram/Marco)	30
16:10		Discussie en vervolg	20
16.30			
			120

Over… Jullie! Topaz

Rein Souwerbren

Patrick Jansen

Maarten Ellenbroek

Marco & LUMC & LACE
Translational Data Science & AI Lab
Artificial Intelligence

Context

Introduction: Marco Spruit

Engineer

1993
- Information Retrieval programmer, ZyLAB Europe

1995
- Big Data system developer, Royal Netherlands Navy

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
  - TDS Lab: https://tdslab.nl
  - UNA Europa SSCs OH+DSAI
  - AI4Prevention Hub

LUMC Strategic Plan 2024-2028

Leiden University Medical Center

Strategy 2024-2028

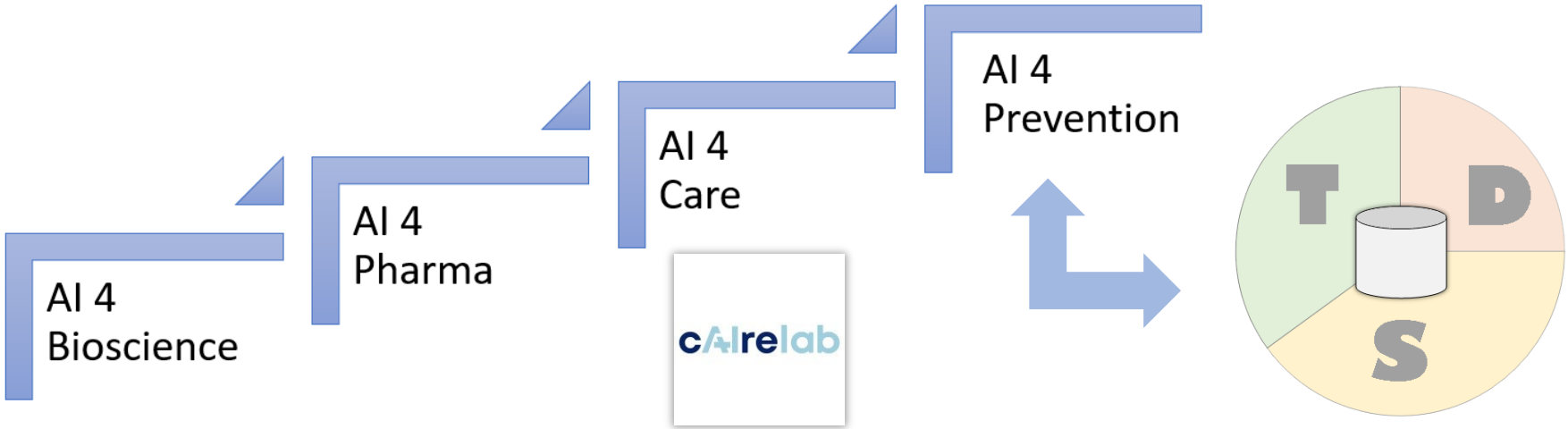
Together we work to improve healthcare and people's health. This is what drives us. Every single day. With *Driven by health* we jointly provide direction for a future ready LUMC.

- Our profile is **recognisable** and aligns with our scientific and clinical expertise. Our key areas are aligned with the focus areas of the Leiden Bio Science Park and Leiden University. We choose the following three key areas, also known as our societal outreach-themes:

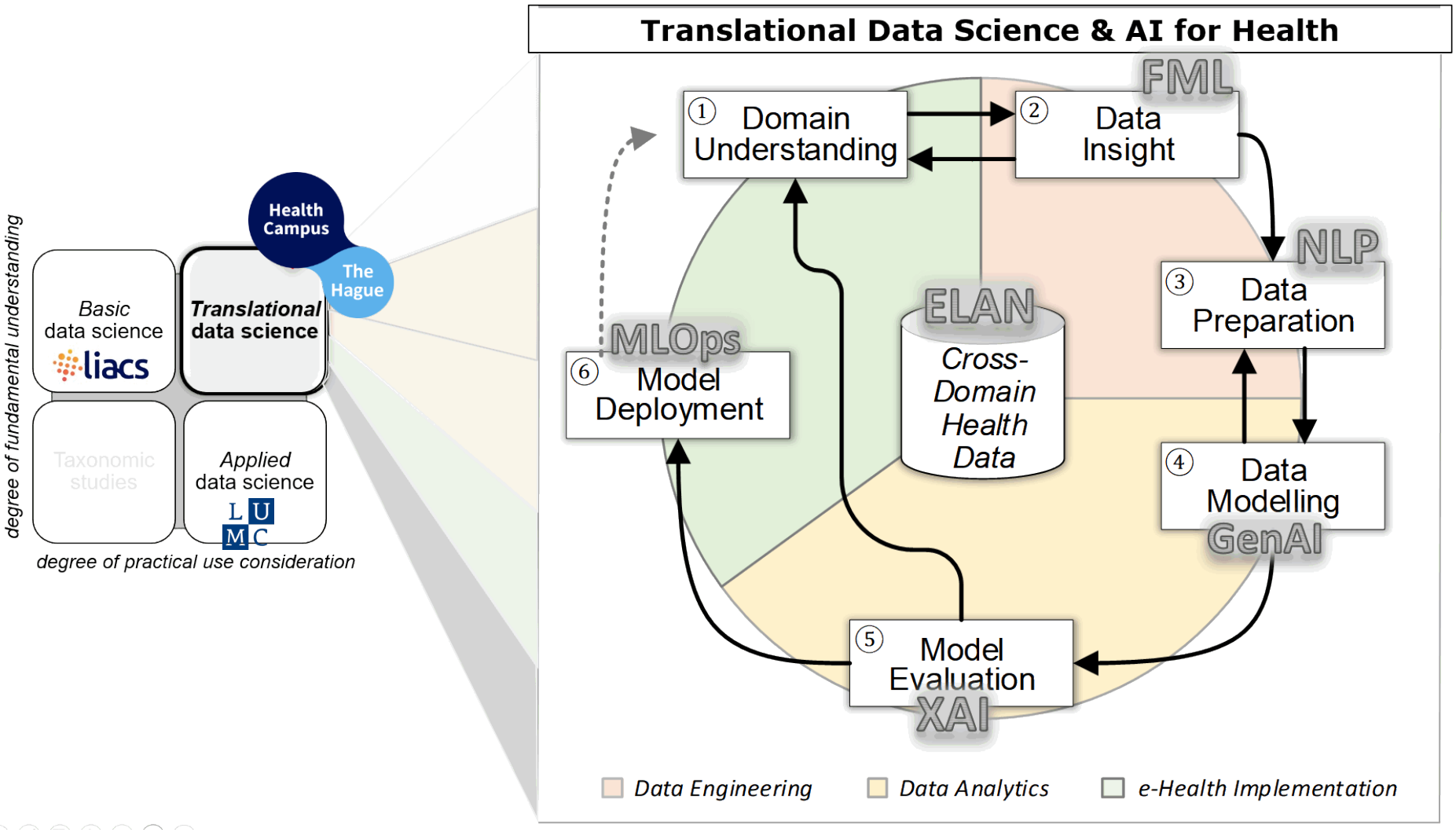
- Regenerative medicine
- Population health
- Data driven healthcare & artificial intelligence (AI)

AI4 Preventie,
Populatiegezondheid & Levensloop

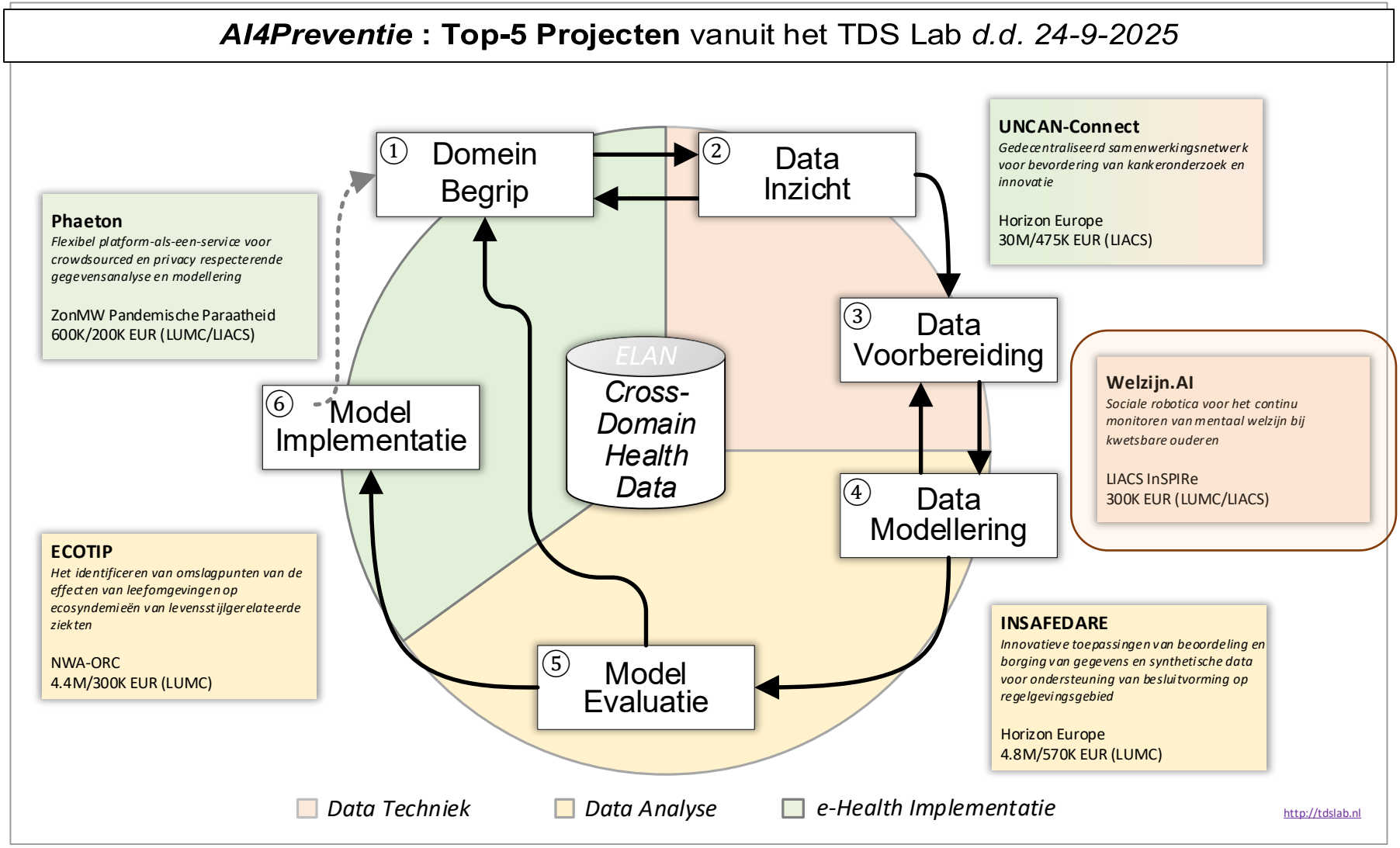
- Eén van de 4 thema's in het voorstel **LACE** (Leiden AI Center of Expertise for Health)
- In opdracht van CvB+RvB: Plan voor gemeenschap-pelijk UL/LUMC-expertise centrum op het gebied van AI



TDS Lab: Thema & Methodologie



TDS Lab: Top-5 Projecten



TDS Lab: Huidig Team

Translational Data Science & AI LAB : current members *d.d. 24-9-2025*

Friso van Dijk
PhD candidate (UU)
Privacy Governance

Els Roorda
PhD candidate (LUMC)
Population Health Information Systems

Dr. Armel Lefebvre
Postdoc (LUMC/LIACS)
Research Data Management

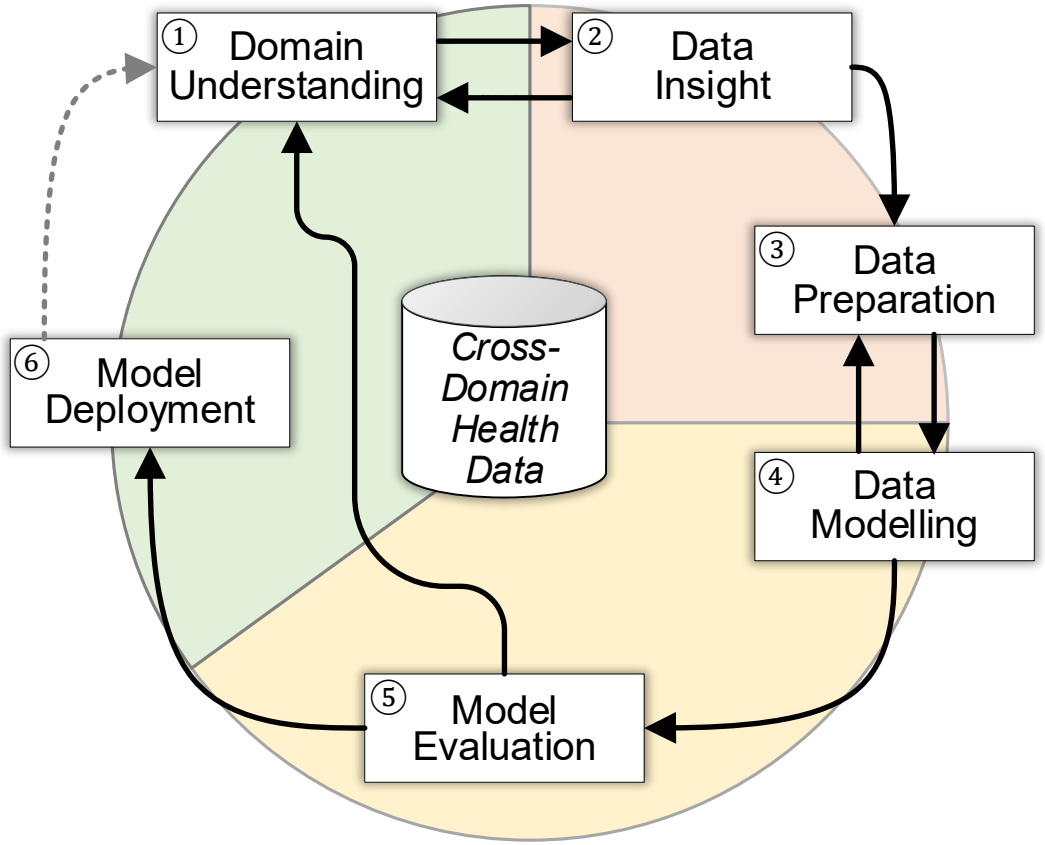
Dr. Marieke Vinkenoog
Postdoc (LUMC)
Machine learning for Pandemic preparedness

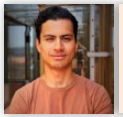
Kiana Shahrasbi
PhD candidate (LIACS)
NLP and MLOps for Cancer Data Research

Sukainah Alfaraj
PhD candidate (LUMC)
Prediction of Diabetes Risk progression

Dr. Marcel Haas
Assistant prof. (LUMC)
Health Data Science

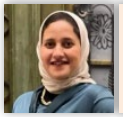
Prof.dr. Marco Spruit
Full Professor
Translational Data Science & AI (LUMC/LIACS)



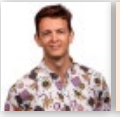


Bram van Dijk
Postdoc (LUMC)
LLMs for Health and ML for Synthetic data

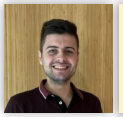
Ian Shen
PhD candidate (LIACS)
Healthcare and Open Science Engineering



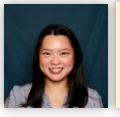
Samar Samir
PhD candidate (LIACS)
Federated NLP in Mental health



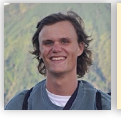
Emil Rijcken
PhD candidate (TUE)
Topic Modelling in NLP



Hielke Muizelaar
PhD candidate (LUMC)
NLP/ML for Patient Segmentation Modelling



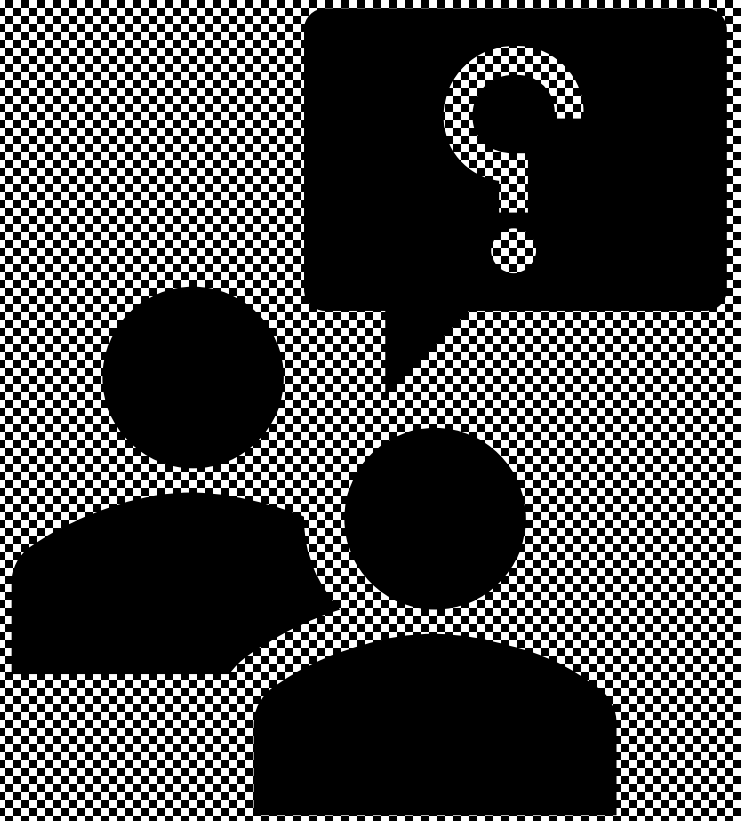
Yee Man Ng
PhD candidate (LIACS)
Multimodal Language Markers for QoL



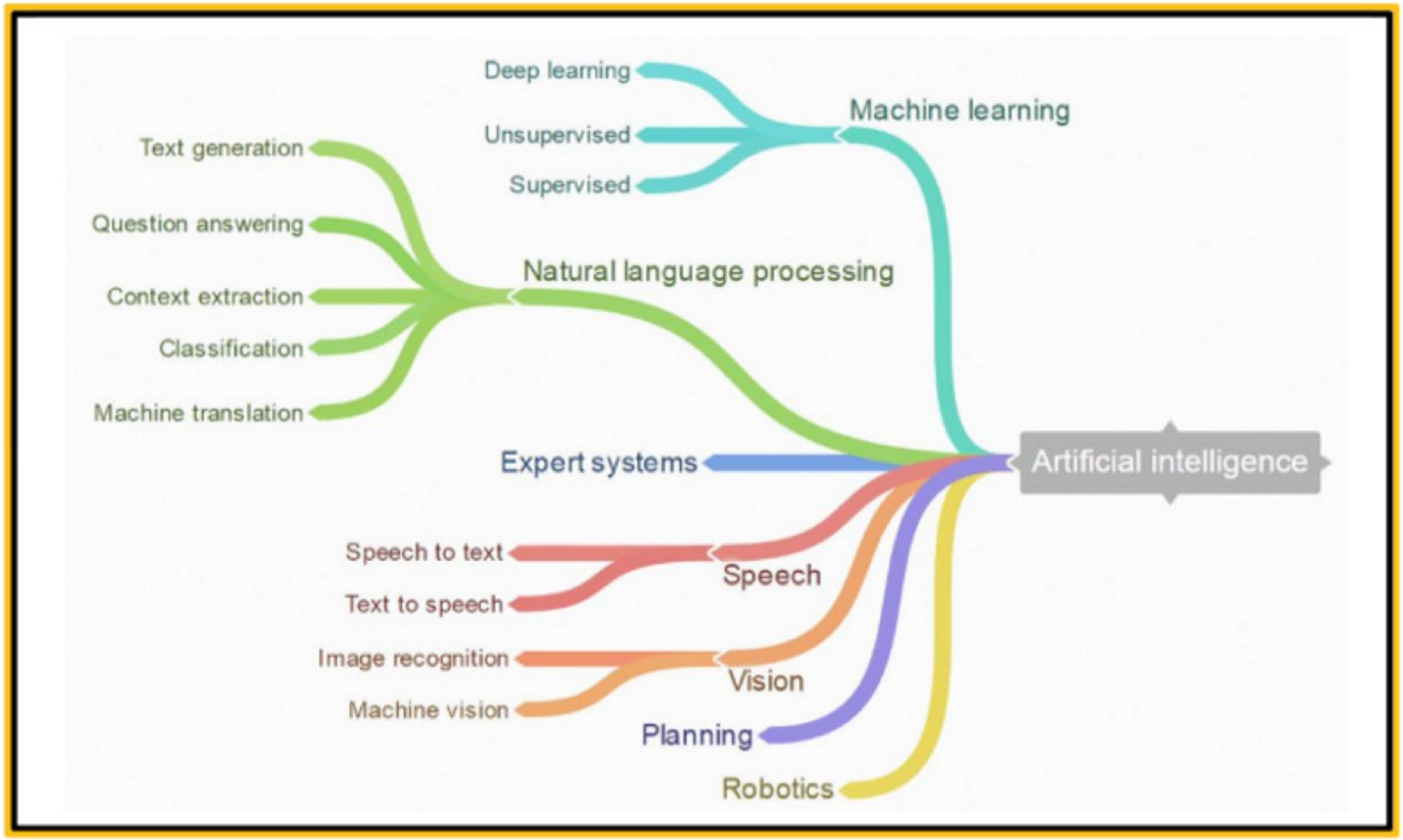
Jim Achterberg
PhD candidate (LUMC)
Synthetic HTA data Generation & Validation

https://tdslab.nl

What is de relatie tussen *Data Science* & *Artificial Intelligence*?



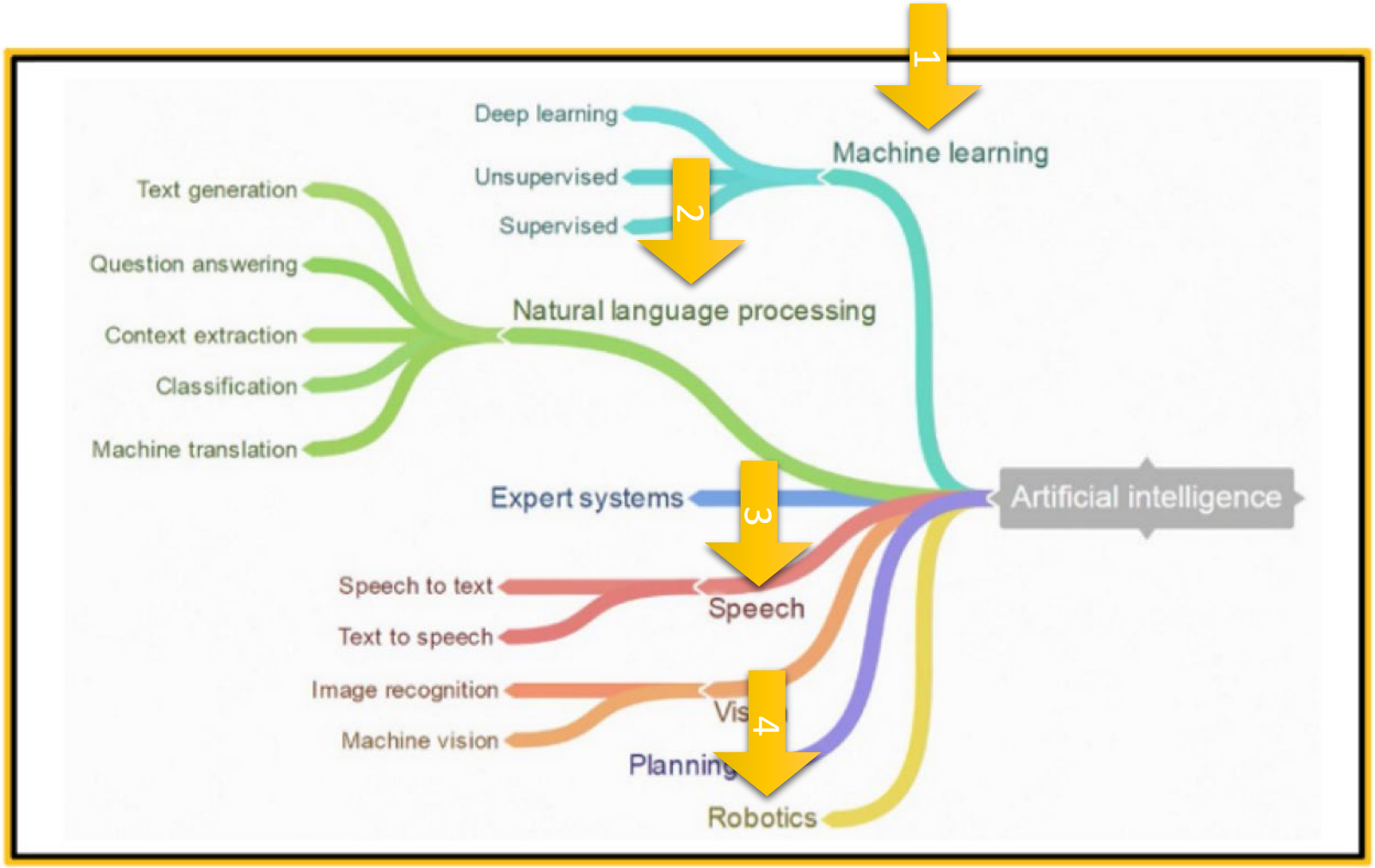
Artificial Intelligence & Data Science both encompass:



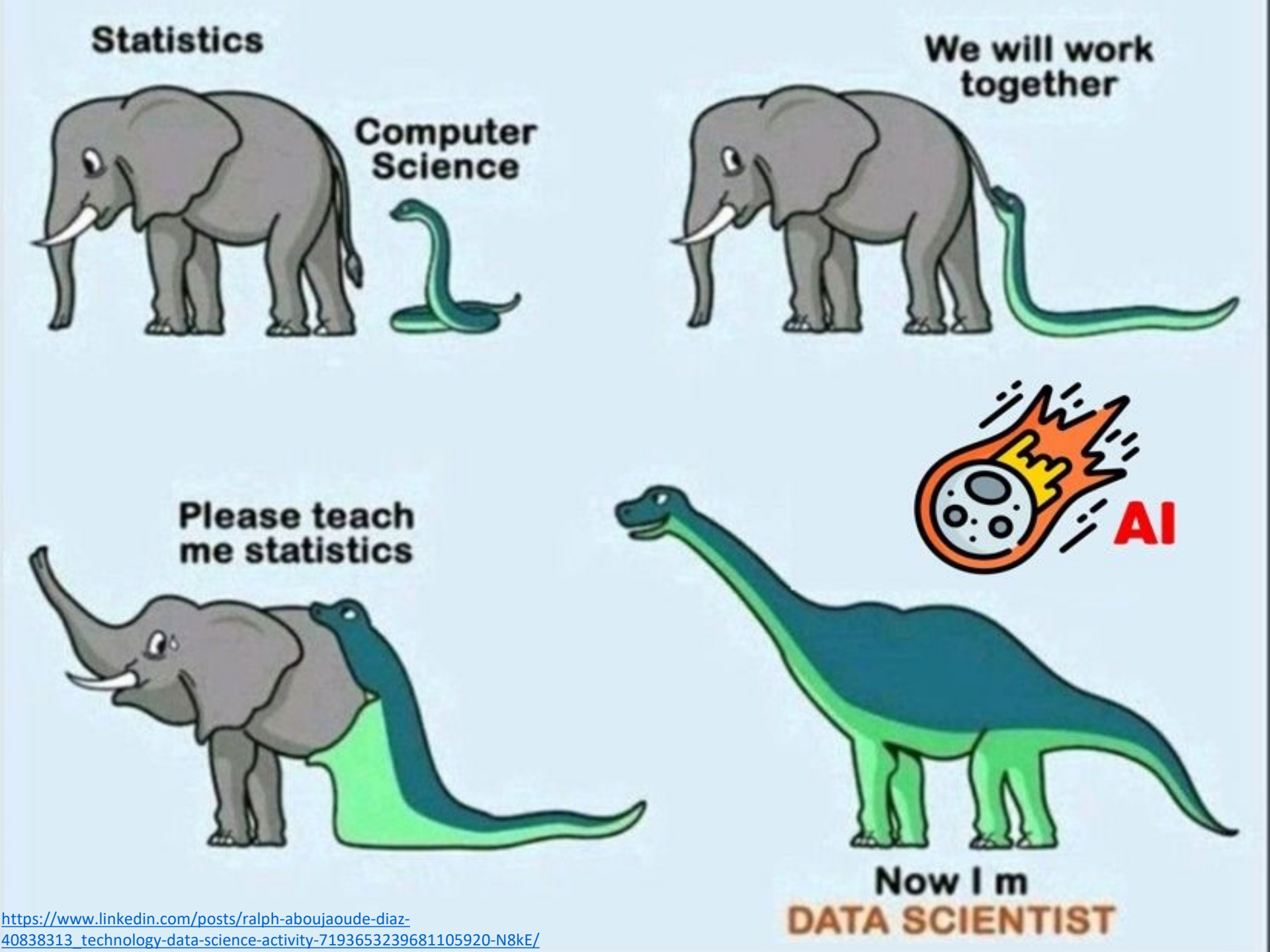
https://thenewstack.io/machine-learning-made-easy/

"Machine Learning is an approach to Achieve Artificial Intelligence"

Onze voorbeelden van AI Vandaag



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



https://www.linkedin.com/posts/ralph-aboujaoude-diaz-40838313_technology-data-science-activity-7193653239681105920-N8kE/

AI Slotnoot


Reciprocal Tariffs
Country
Tariffs Charged to the U.S.A. Including Currency Manipulation and Trade Barriers
U.S.A. Discounted Reciprocal Tariffs
China 67% 34%
European Union 39% 20%
Vietnam 90% 46%
Taiwan 64% 32%
Japan 46% 24%
India 52% 26%
South Korea 50% 25%
Thailand 72% 36%
Switzerland 61% 31%
Indonesia 64% 32%
Malaysia 47% 24%
Cambodia 97% 49%
nited Kingdom 10% 10%
uth Africa 60% 30%
zil 10% 10%
gladesh 74% 37%
gapore

— Opinion

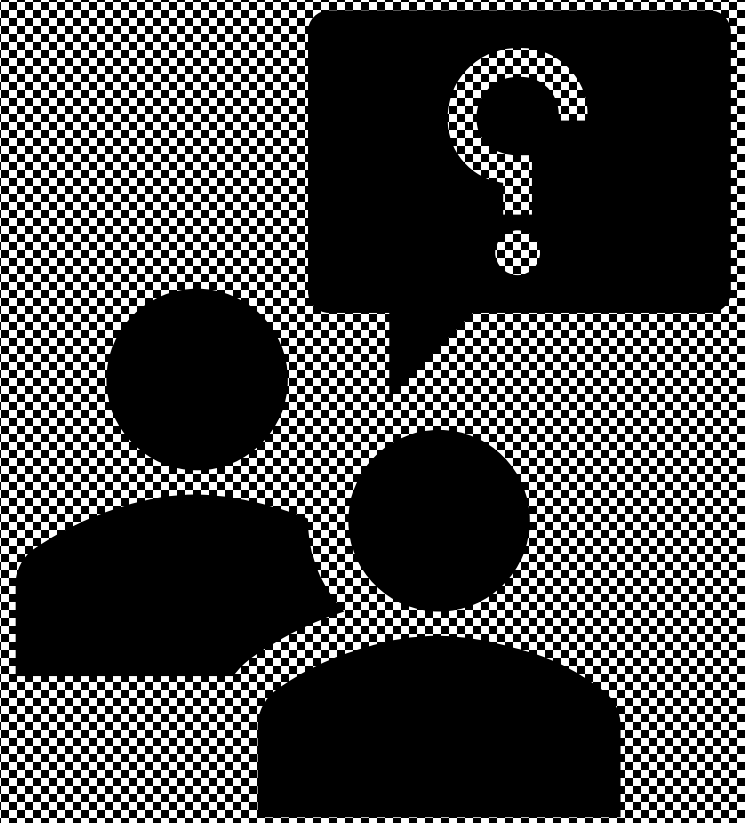
ChatGPT told Trump how to set tariffs – and now the world's paying for it

"Vibe-Governing" suggests policymakers are letting AI-generated content guide major economic or political decisions, leading us into surreal, potentially dangerous territory.

Rebecca Johnson *Contributor*

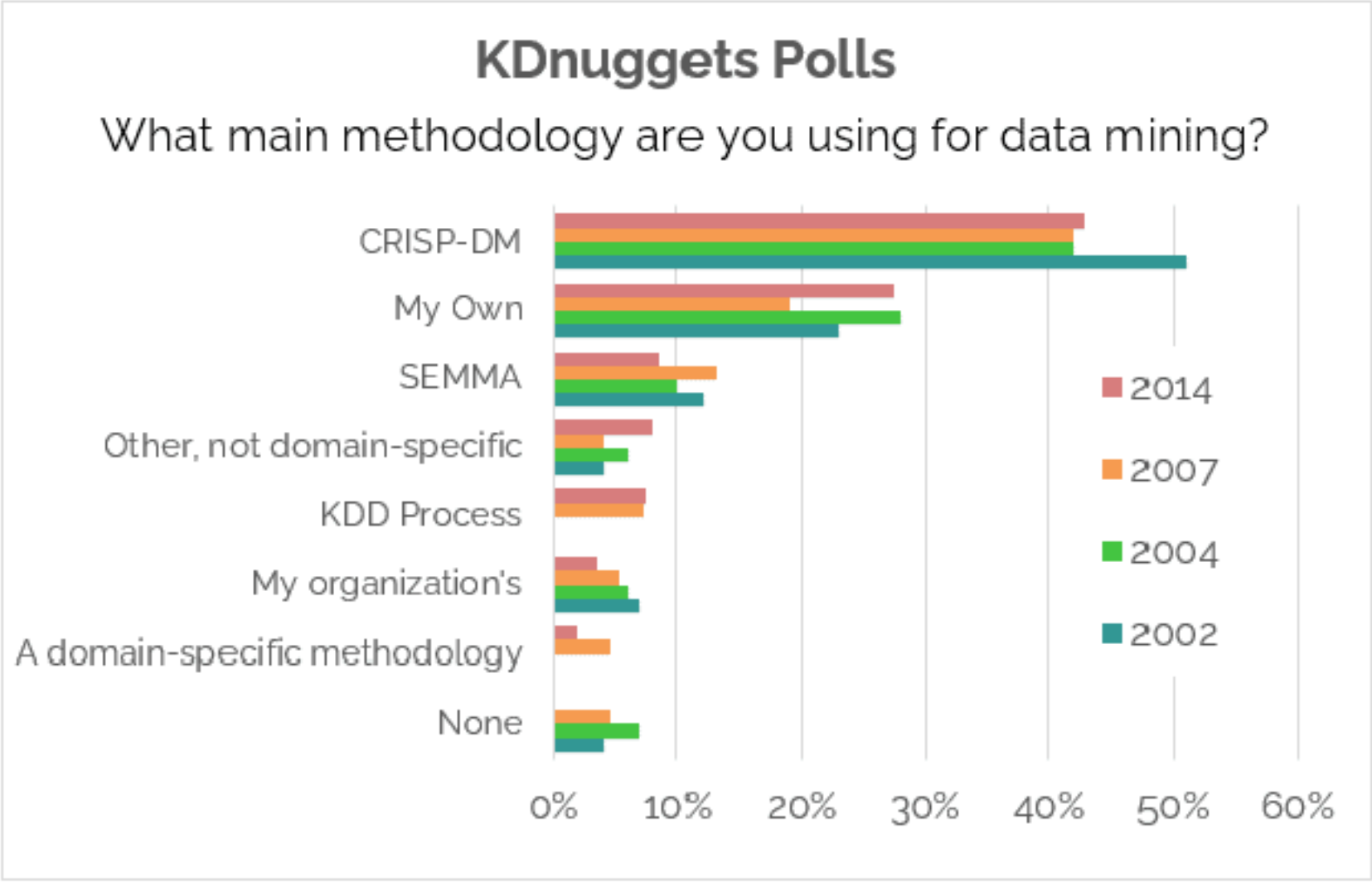
Apr 7, 2025 – 1.10pm

What is volgens experts dé *best practice* methode om Data Science & AI projecten betrouwbaar uit te voeren?

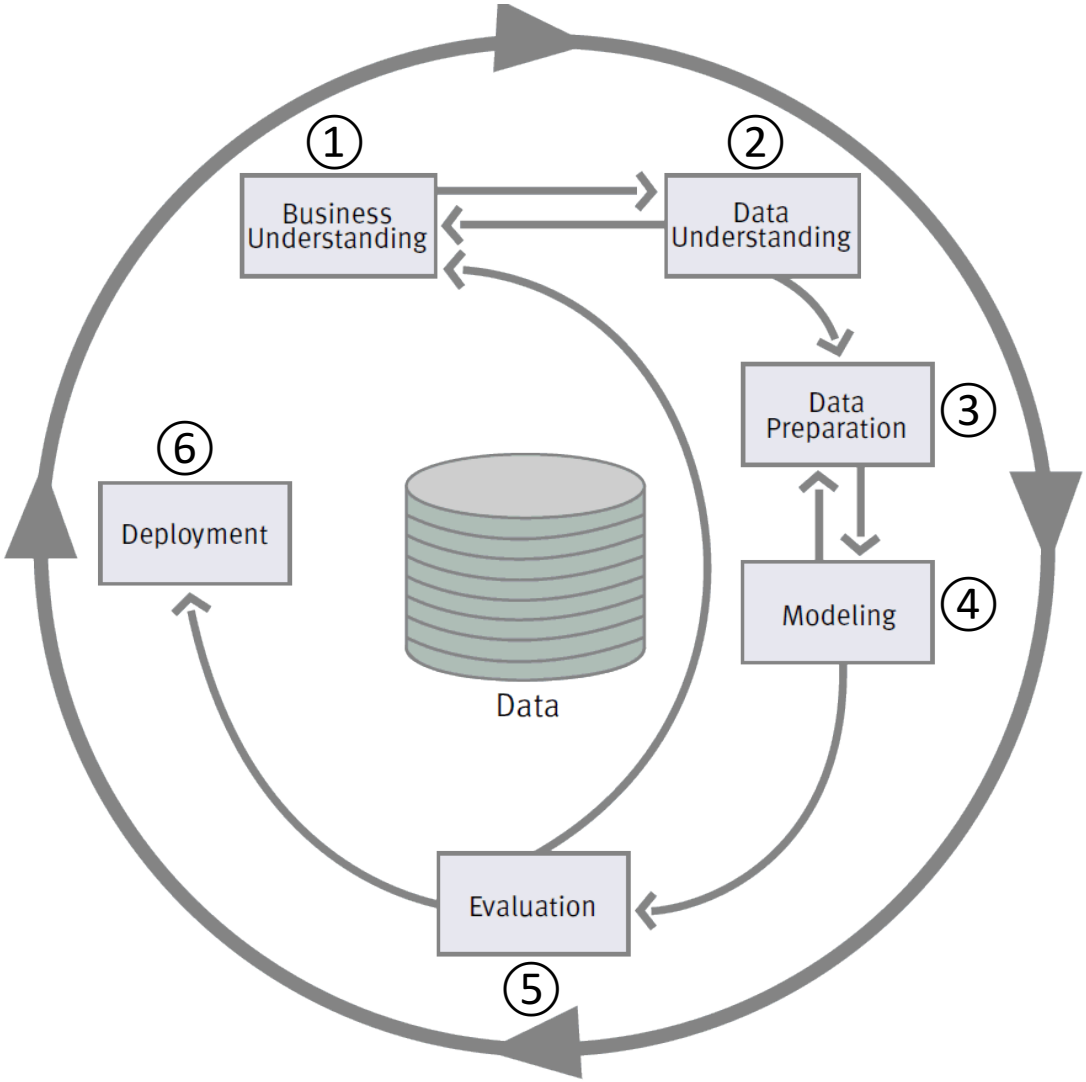


CRISP-DM

CRoss-Industry Standard Process for Data Science & AI



CRISP-DM: Cross-Industry Standard Process for Data Science



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]

CRISP-DM: Tasks & Outputs

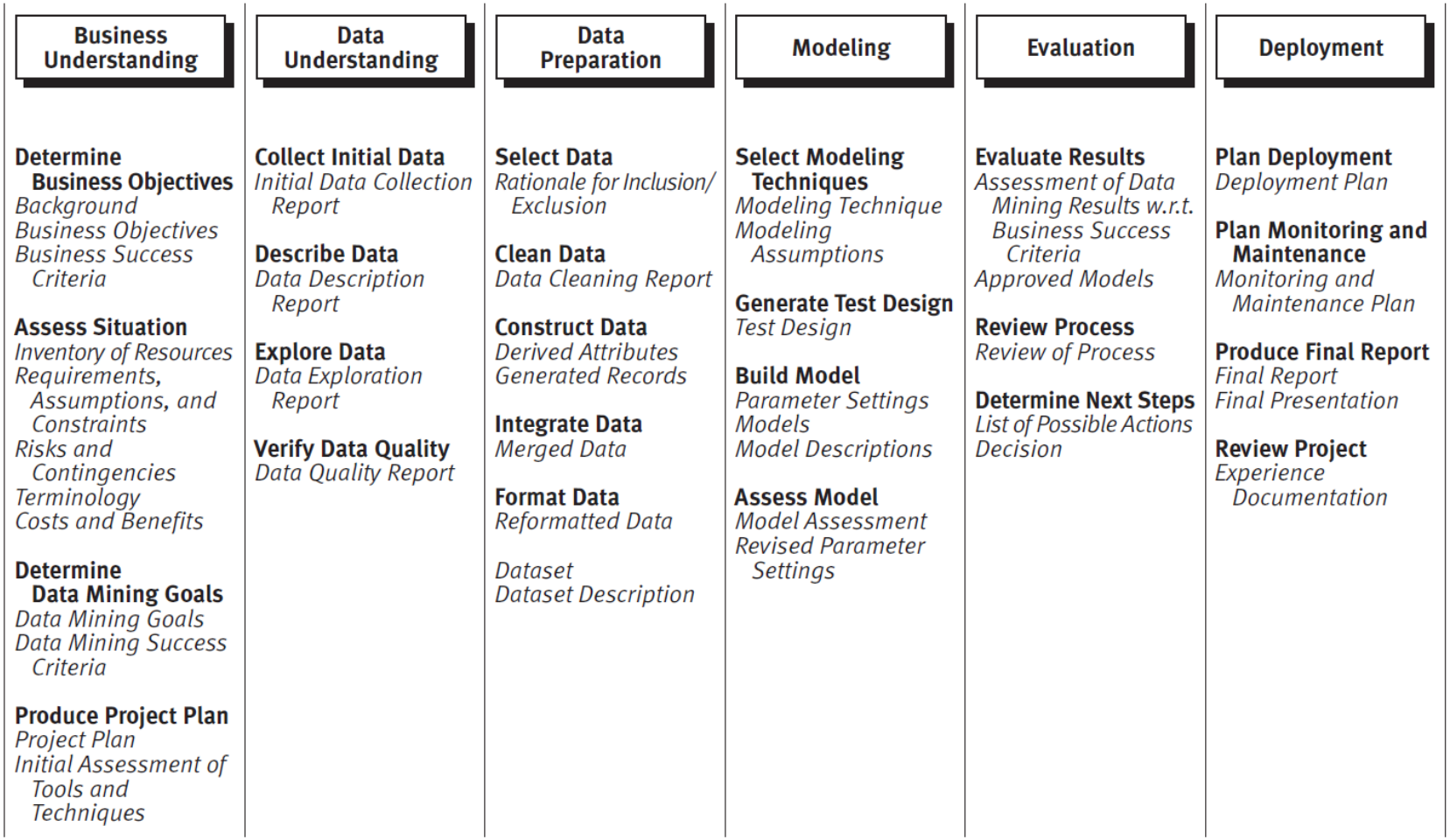


Figure 3: Generic tasks (bold) and outputs (italic) of the CRISP-DM reference model

CRISP-DM: Method & Task Levels

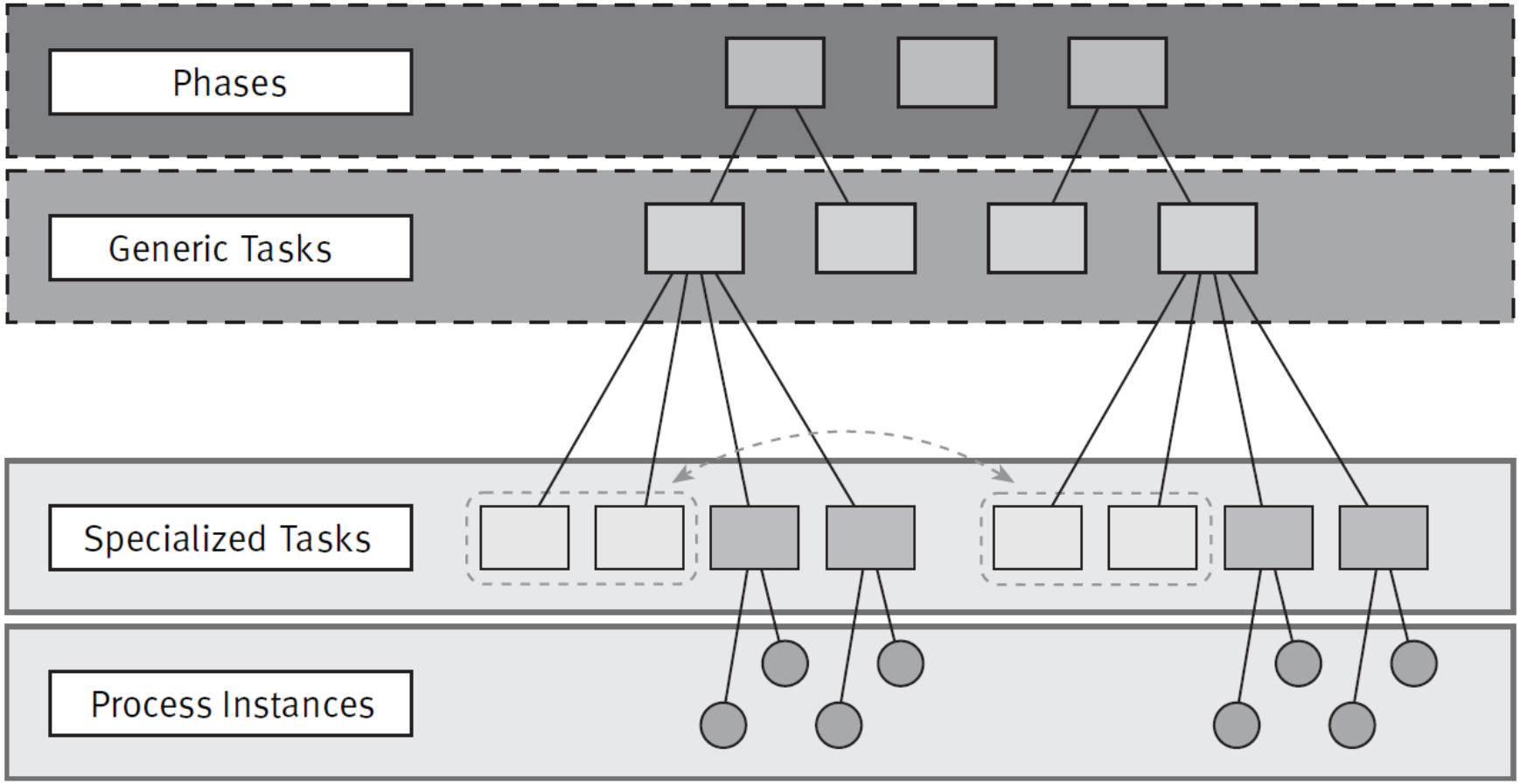
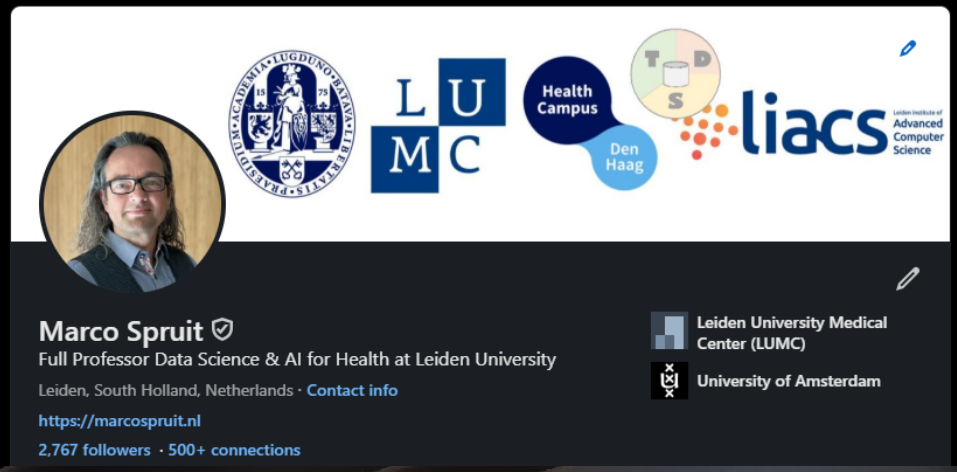


Figure 1: Four level breakdown of the CRISP-DM methodology

1



Data Science
Machine Learning

CRISP-DM in Action

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. https://doi.org/10.1016/j.chb.2013.07.038 *(80+ citations)*

Introduction (1/2)

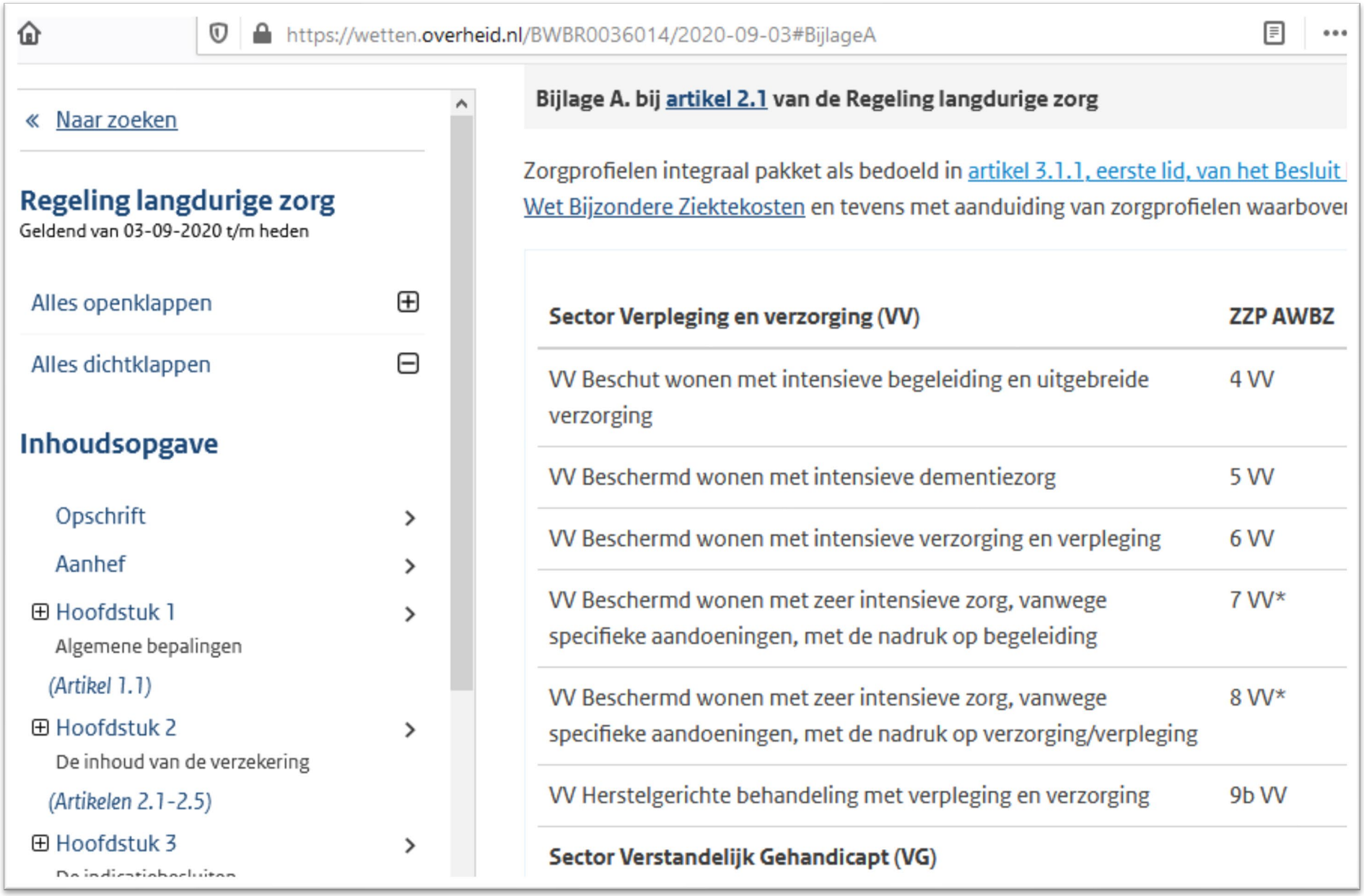
- Long-term care
  - Care for people with a long-term or chronic disorder
  - Relatively unexplored
- Main goals for long-term care
  - Care of good quality
  - Acceptable cost level
- One of the biggest expenses of Dutch government
  - 38% of the total healthcare budget
  - €14 billion in 2000 → €27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

Totaal 2024: ~36 Miljard!

https://www.nza.nl/zorgsectoren/langdurige-zorg/kerncijfers-langdurige-zorg

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)

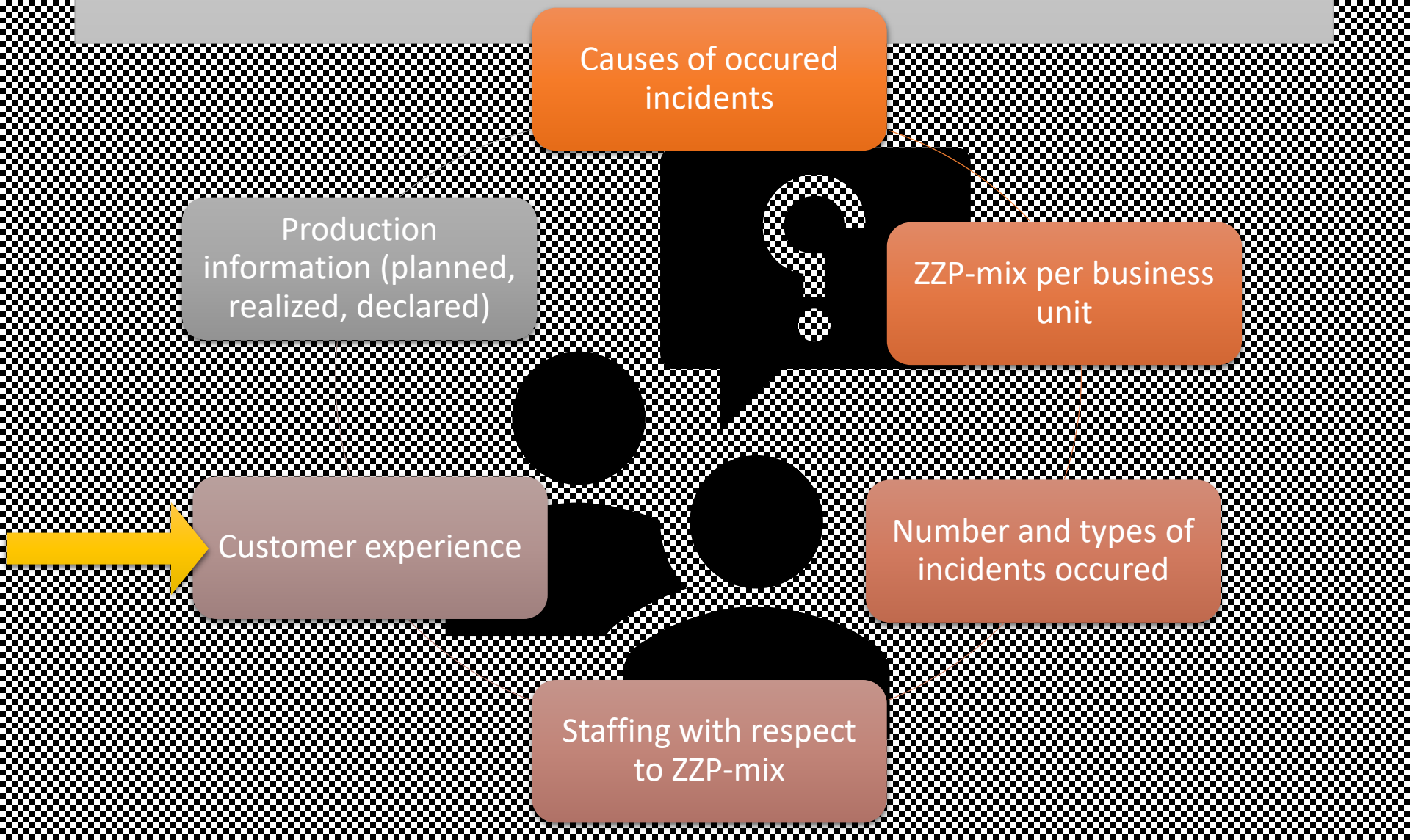
Research Method: CRISP-DM

- 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
  - 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 = $\left(\frac{8}{8} \times 10\right) + \left(\frac{4}{5} \times 6\right) + \left(\frac{3}{5} \times 3\right) = 16.6$

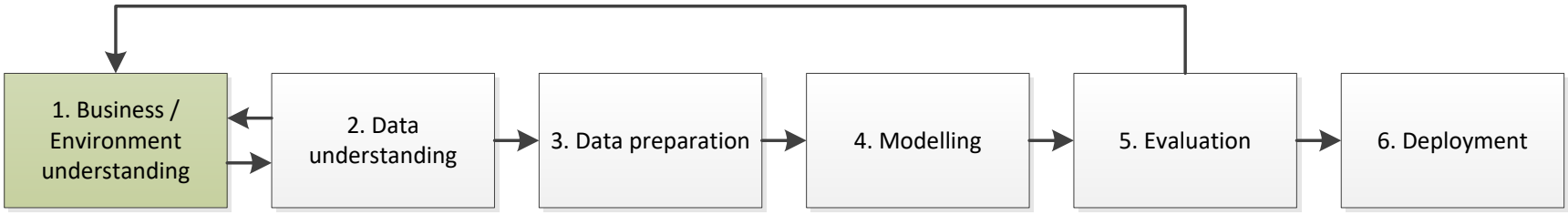
What is the #1 Information Need?



Calculation: Top 10 Needs

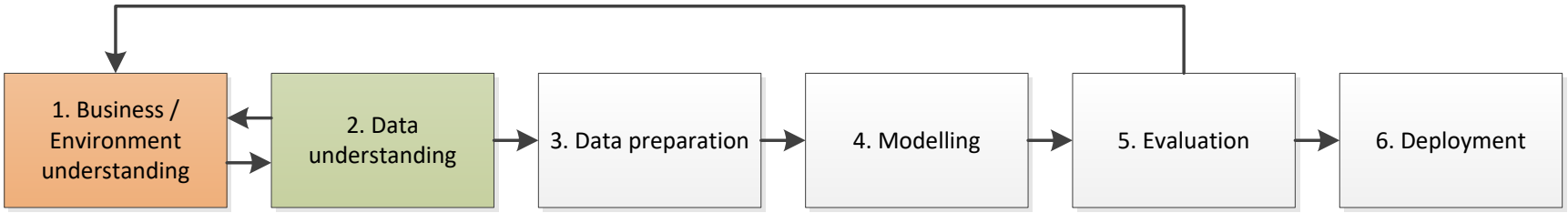
#	Type	Information need	Board	Mgmt	Stakeh.	Score
1	Q	Customer experience	8	7	10	16.6
2	F	Staffing with respect to ZZP-mix	7	4	2	14.8
3	F	ZZP-mix per business unit	7	4	0	13.6
4	F	ZZP-mix prognoses	7	4	0	13.6
5	F	Staffing with respect to operations	6	4	2	13.5
6	Q	Number of incidents occured	6	4	2	13.5
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

Findings: Business understanding



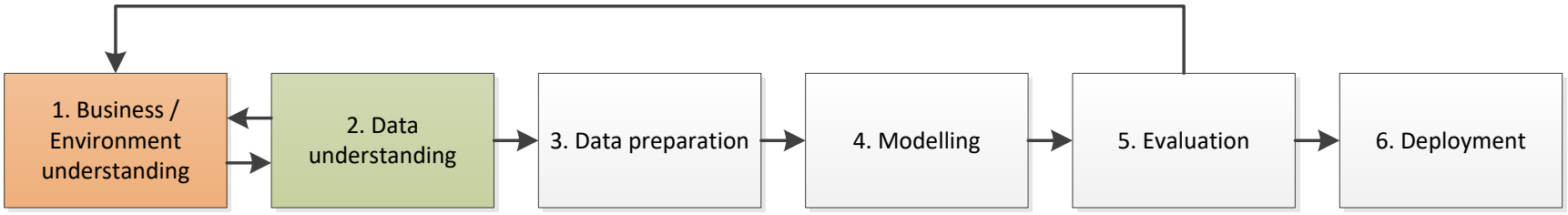
Information needs	Data mining goals
• Number of occured incidents • Types of occured incidents • Causes of the occured incidents • Patterns in occured incidents	• Identify patterns in incidents
• Number of clients at an increased risk • Types of risk the clients run	• Identify relationships in risk assessment
• 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

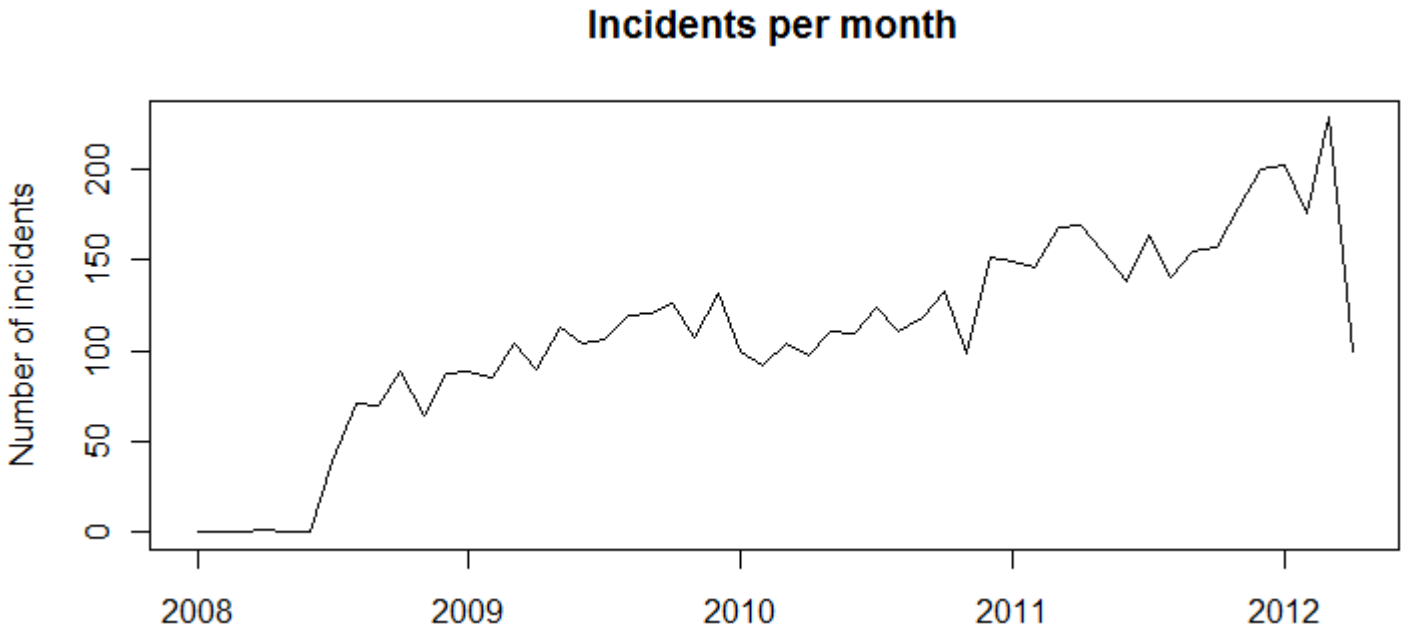
Findings: Data understanding



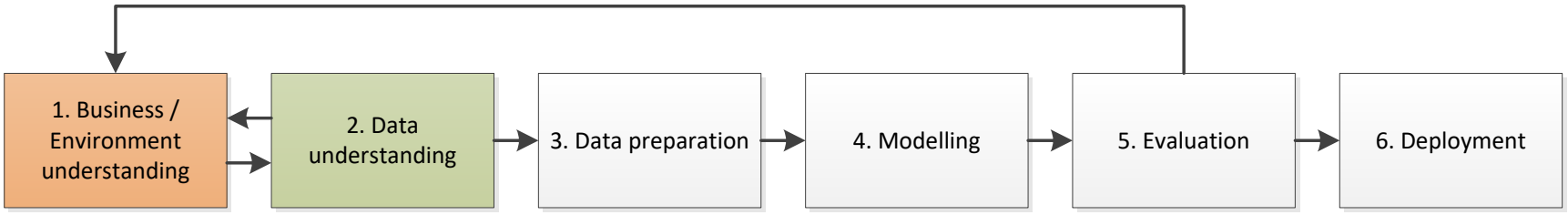
Data mining goals	Available data
• Identify patterns in incidents	• 5,692 records with incidents • 13 different incident types
• Identify patterns in risk assessment	• Depression risk: 2,129 records • Falling risk: 889 records • Incontinence risk: 877 records • Medication risk: 806 records • Problem 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

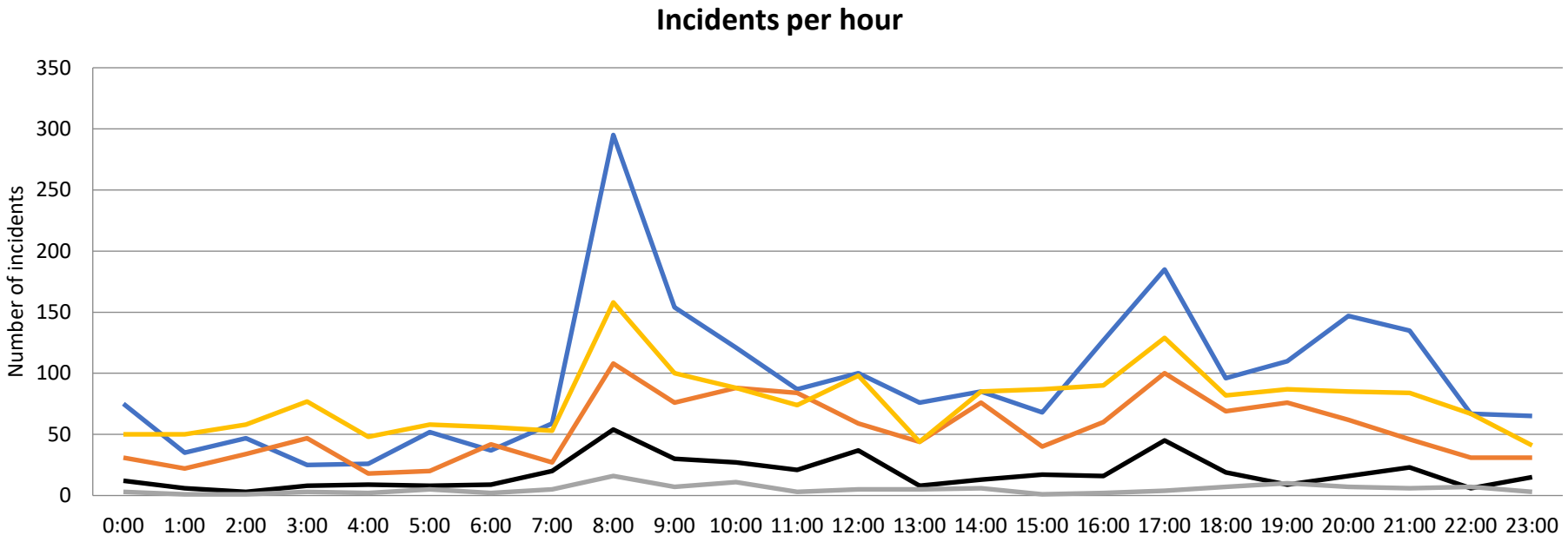
Findings: Identify patterns in incidents



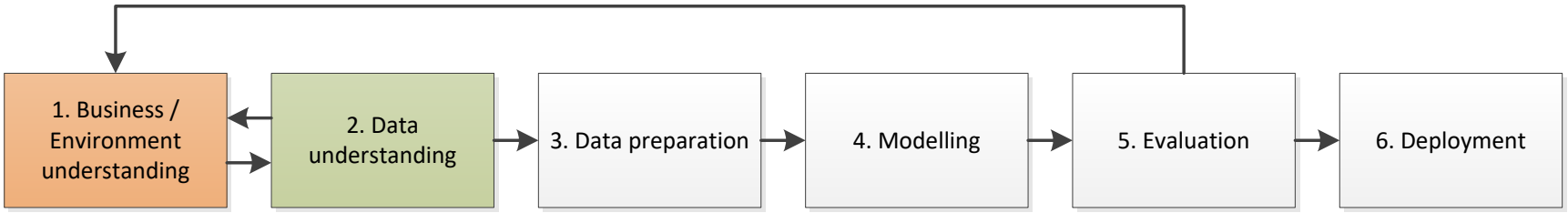


Findings: Identify patterns in incidents



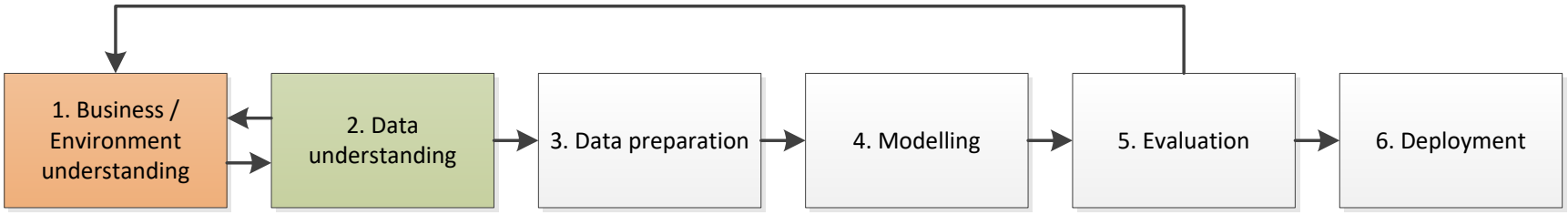


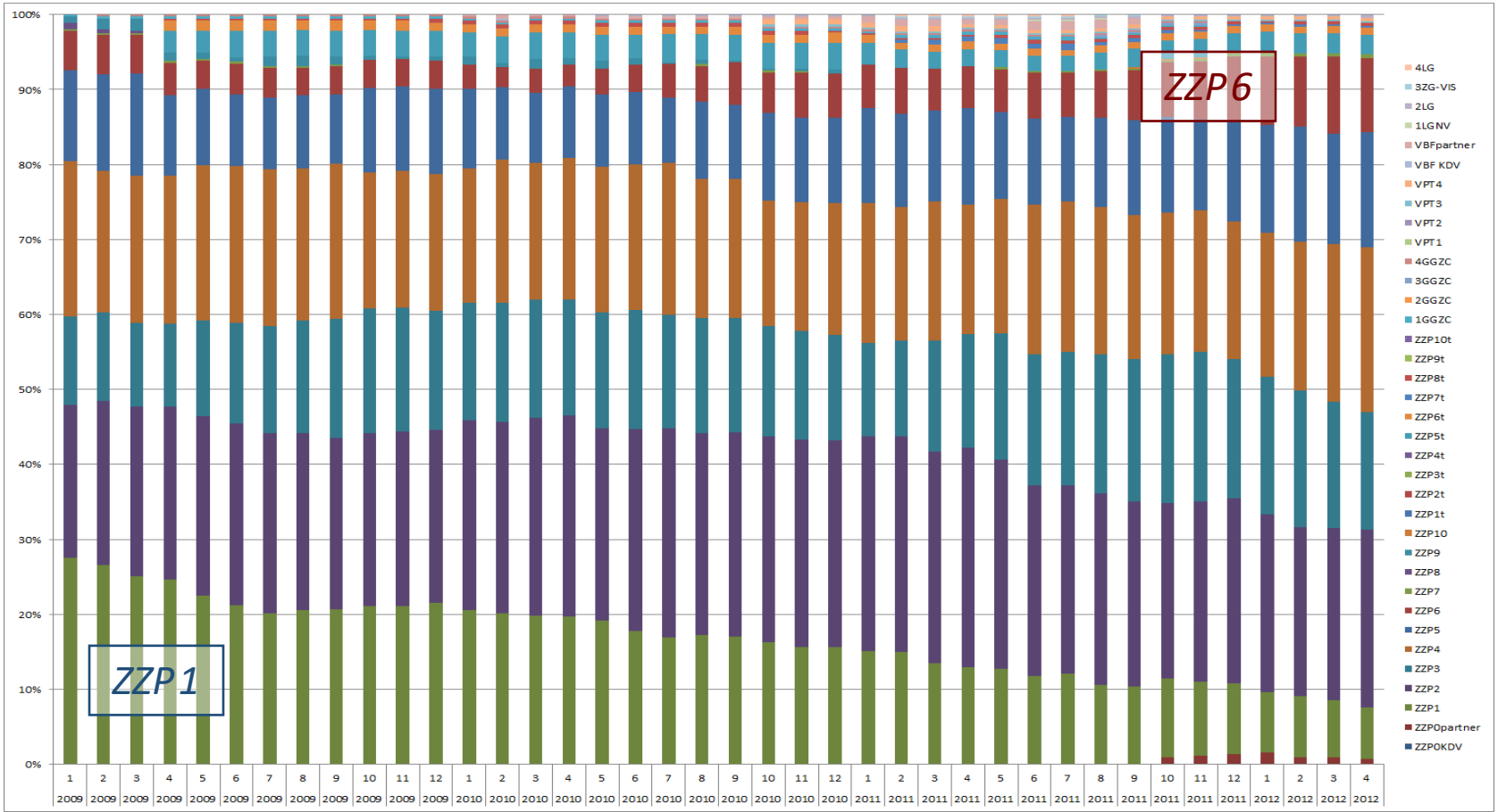
Findings: Identify patterns in incidents



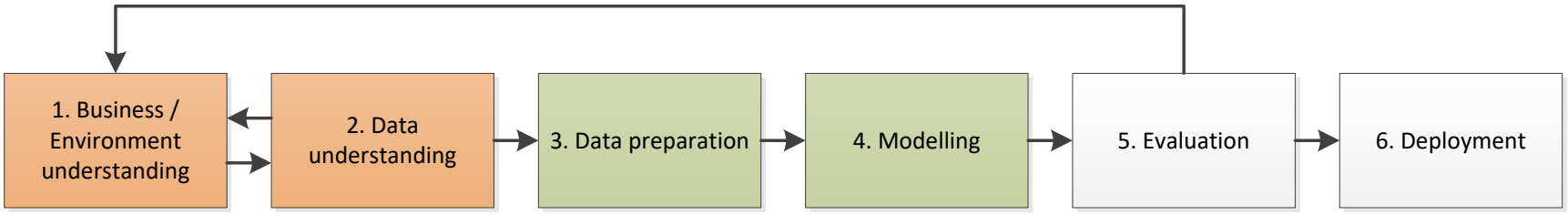
Incident location	*Loc1*	*Loc2*	*Loc3*	*Loc4*	*Loc5*	*Total*	*%*
Activities room	11	0	1	10	0	**22**	**0.4%**
Bathroom	126	77	152	43	17	**415**	**7.3%**
Corridor	101	72	92	9	3	**277**	**4.9%**
Kitchen	225	154	110	46	10	**545**	**9.6%**
Bedroom	524	278	575	38	31	**1,446**	**25.4%**
Toilet	62	45	54	5	8	**174**	**3.1%**
Living room	895	500	587	231	40	**2,253**	**39.6%**
Totals	**2,160**	**1,251**	**1,738**	**422**	**121**	**5,692**	**100%**

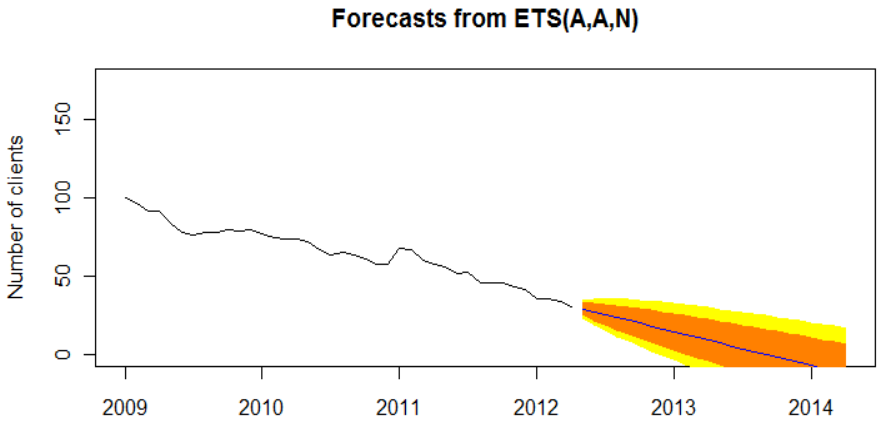
Findings: Predict the ZZP mix



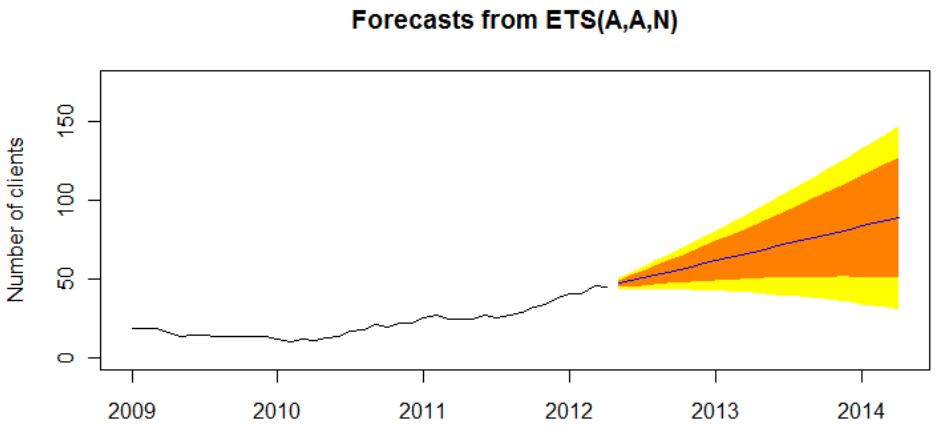


Findings: Predict the ZZP mix



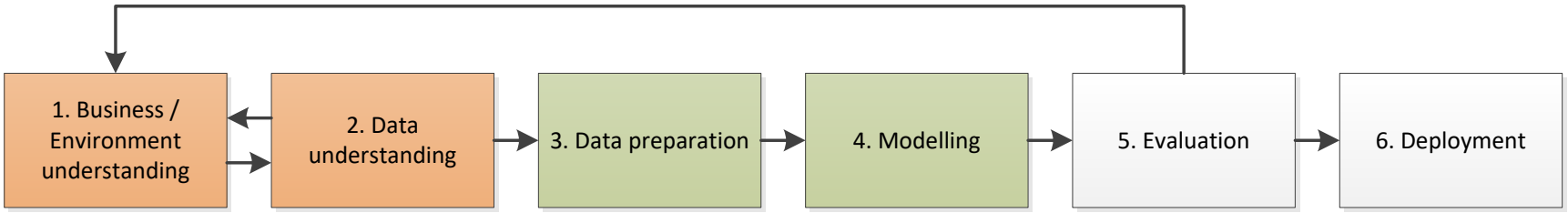


ZZP1



ZZP6

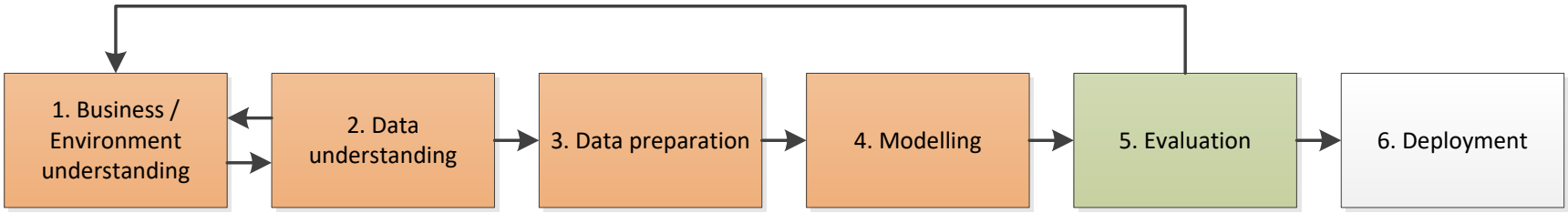
Findings: Identify relationships in risk assessment



Rule	Support	Confidence	Lift
Incontinence, Medication → Falling	11.76%	90.63%	1.824
Falling, Medication → Incontinence	11.76%	95.08%	1.769
Incontinence, Weight intramural → Falling	11.56%	87.69%	1.765
Falling, Weight intramural → Incontinence	11.56%	93.44%	1.738
Depression, Incontinence → Falling	31.64%	82.11%	1.652

$$\text{Support}(X) = \frac{\text{Number of transactions containing the X}}{\text{Total number of transactions}} \quad \text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cap Y)}{\text{Support}(X)} \quad \text{Lift}(X \rightarrow Y) = \frac{\text{Support}(X \cap Y)}{\text{Support}(X) \times \text{Support}(Y)}$$

Findings: Evaluation

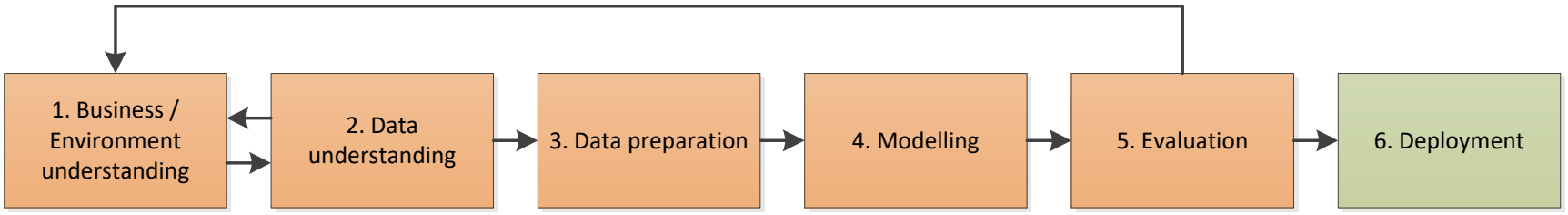


- Explorative research → no groundbreaking results
  - Information needs of *multiple* care institutions & stakeholders
- Lack of standardization
  - Heterogeneous data

NLP

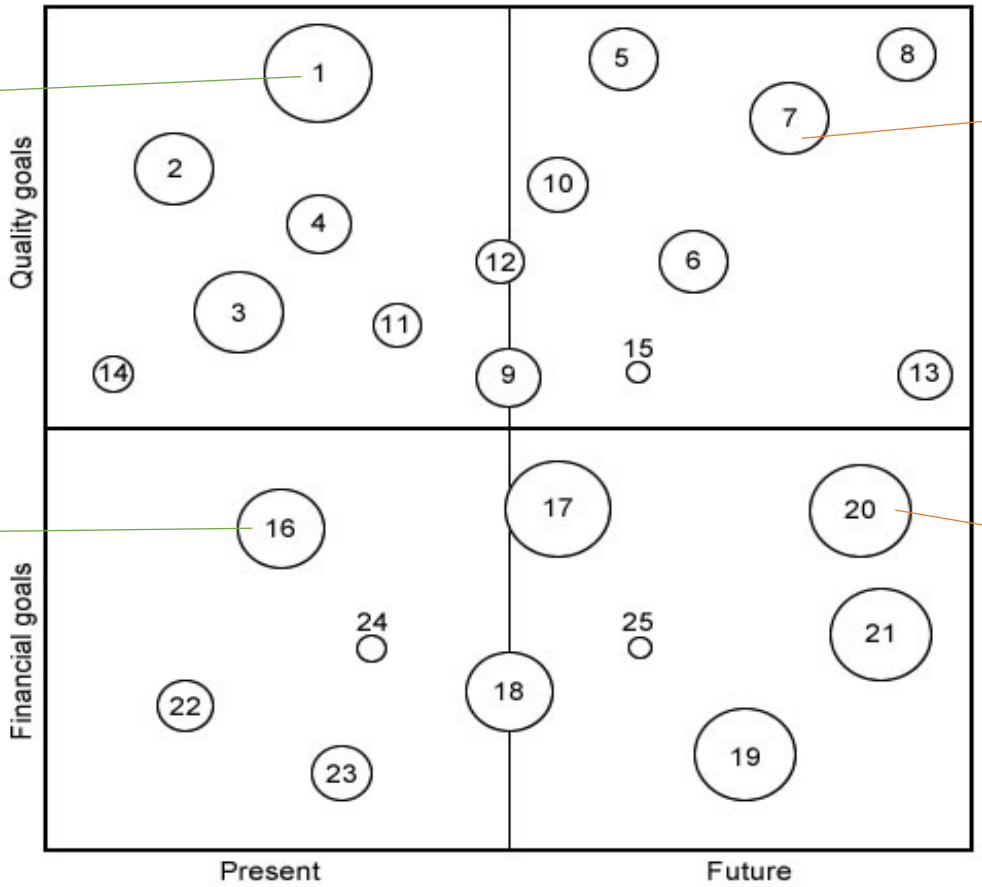
- Predictions are too dependent on environmental factors
  - Limited historical data
  - Too many dependent factors for the forecasts to be practically useful
    - Changes in laws and regulations have direct effect on analysis
- Qualitative data (*e.g.* free text input)
  - Complex to analyse

Findings: Deployment



Identify the patterns in incidents

Identify & predict the ZZP-mix

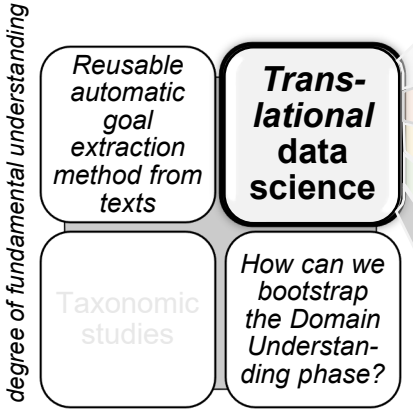


Relationship between care-related measures and incidents

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

One-pager on…

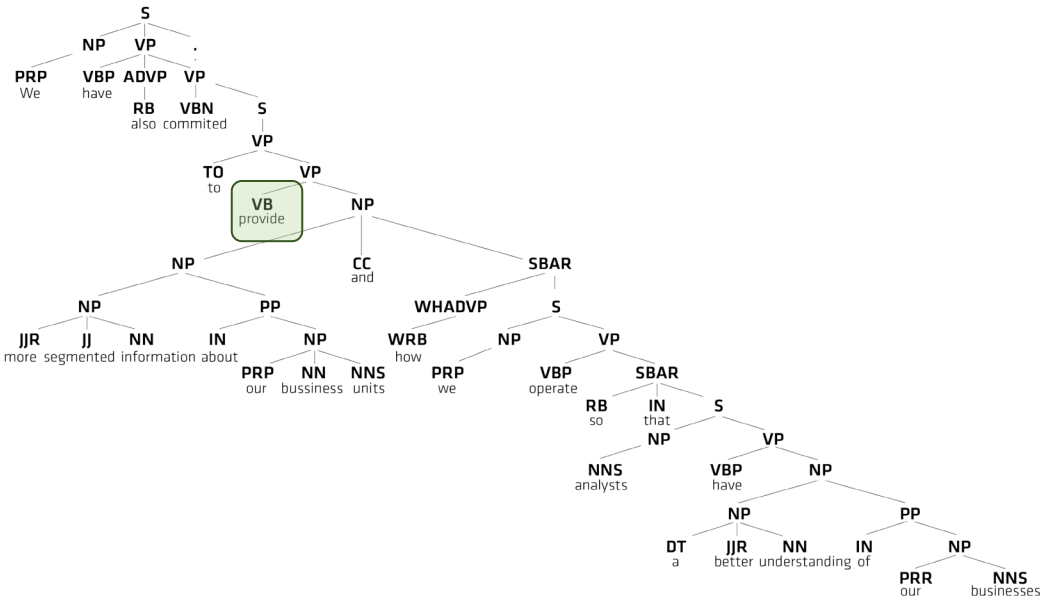
An automated approach to gather information needs



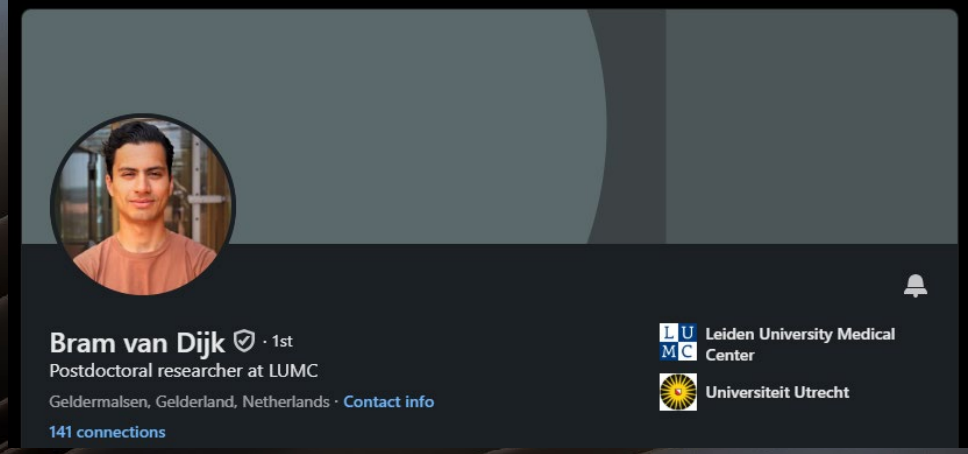
Example: NLP-Automated Domain Understanding (NAMBU)

Table 1. Goal-related keywords.

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



2



AI Deelgebieden: NLP & Spraak & Robotica

Welzijn.AI

Op een passend vervolg!

Marco Spruit
Full Professor Data Scie
Leiden, South Holland, Net
https://marcospruit.nl
2,767 followers · 500+ co

Bram van Dijk
Postdoctoral researcher
Geldermalsen, Gelderland,
141 connections

Translational Data Science & AI L/

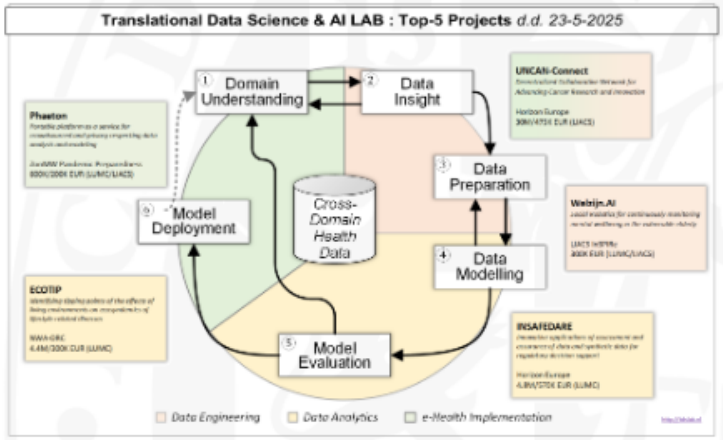
https://tdslab.nl/index.php

Home | About | News | Team | Systems | Projects | Publications | Open Topics | tdslab.nl | marcospruit.nl

The *Translational* Data Science & AI Laboratory

The **Translational Data Science & AI Lab** (a.k.a. *TDS Lab*)'s mission is to connect practical problems in healthcare practices to fundamental challenges in data science and to subsequently address both simultaneously. This is our encompassing *Translational Data Science (TDS)* research theme, which bridges the best of both worlds. We pursue a better fundamental understanding of the world around us *through* data science & AI innovations *by being* societally inspired, demand-driven and solution-oriented. See our About page for more information.

Top-5 TDS Lab projects



Current research grants (Total: 25)

2025: **GenAI4EU**, 9K EUR (LUMC).

Recent #tdslab News on Mastodon

Marco Spruit 04 Sept 2025
@marcospruit@mastodon.social

Congrats to #tdslab's Marieke who wins the Krijn Rietveld Doctoral Societal Impact Award for her innovative research on blood donation and colorectal cancer!
universiteitleiden.nl/en/news/...

Leiden University

Two young researchers win Krijn Rietveld Award for innovative research on blood donation and colorectal cancer

One used 3D cell models to study colorectal cancer, while the other used machine learning for more efficient