

# *Data Science in Practice:* *A longterm healthcare case study*

@QConsult AI intervision

25 Jan 2022



Leiden University  
Medical Center

**Prof.dr. Marco Spruit**

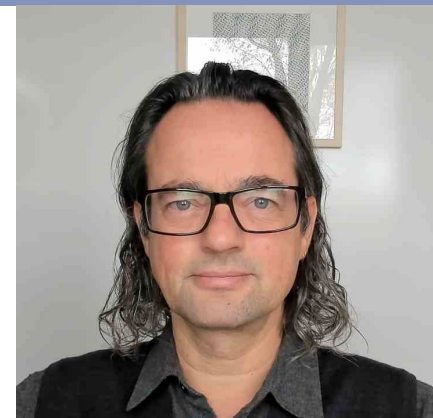
*Professor Advanced Data Science in  
Population Health  
Public Health and Primary Care*

0031 6

[m.r.spruit@lumc.nl](mailto: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



2003

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

2007

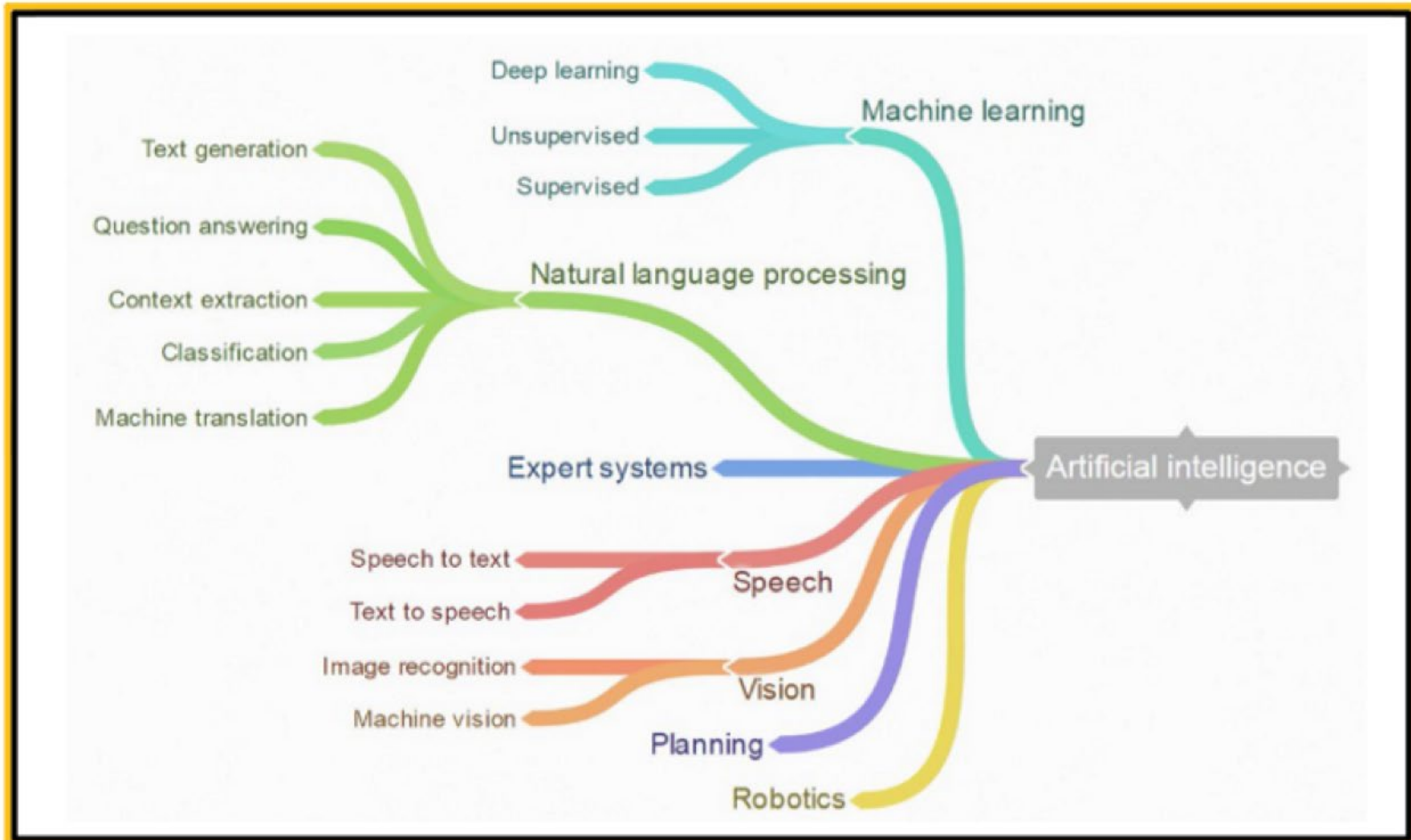
- Assistant/Associate professor Information Science, Utrecht University
  - Applied Data Science Lab

2020

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

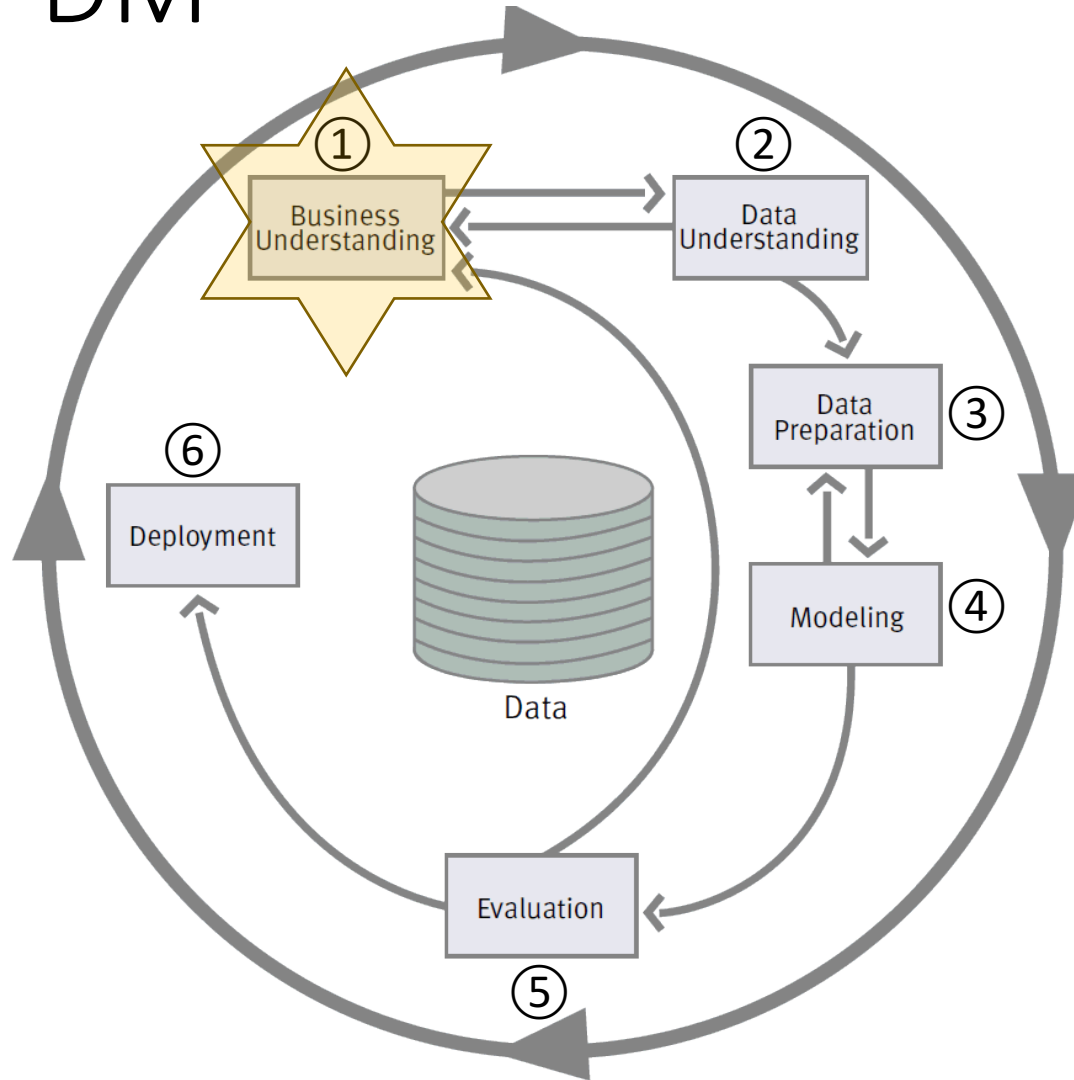
Researcher

# Data Science vs Artificial Intelligence



*“Machine Learning is an approach to Achieve Artificial Intelligence”*

# CRISP-DM



# CRISP-DM: Tasks & Outputs

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<p><b>Determine Business Objectives</b>  <i>Background            Business Objectives            Business Success            Criteria</i></p> <p><b>Assess Situation</b>  <i>Inventory of Resources            Requirements,            Assumptions, and            Constraints            Risks and            Contingencies            Terminology            Costs and Benefits</i></p> <p><b>Determine Data Mining Goals</b>  <i>Data Mining Goals            Data Mining Success            Criteria</i></p> <p><b>Produce Project Plan</b>  <i>Project Plan            Initial Assessment of            Tools and            Techniques</i></p>	<p><b>Collect Initial Data</b>  <i>Initial Data Collection            Report</i></p> <p><b>Describe Data</b>  <i>Data Description            Report</i></p> <p><b>Explore Data</b>  <i>Data Exploration            Report</i></p> <p><b>Verify Data Quality</b>  <i>Data Quality Report</i></p>	<p><b>Select Data</b>  <i>Rationale for Inclusion/            Exclusion</i></p> <p><b>Clean Data</b>  <i>Data Cleaning Report</i></p> <p><b>Construct Data</b>  <i>Derived Attributes            Generated Records</i></p> <p><b>Integrate Data</b>  <i>Merged Data</i></p> <p><b>Format Data</b>  <i>Reformatted Data</i></p> <p><i>Dataset            Dataset Description</i></p>	<p><b>Select Modeling Techniques</b>  <i>Modeling Technique            Modeling            Assumptions</i></p> <p><b>Generate Test Design</b>  <i>Test Design</i></p> <p><b>Build Model</b>  <i>Parameter Settings            Models            Model Descriptions</i></p> <p><b>Assess Model</b>  <i>Model Assessment            Revised Parameter            Settings</i></p>	<p><b>Evaluate Results</b>  <i>Assessment of Data            Mining Results w.r.t.            Business Success            Criteria            Approved Models</i></p> <p><b>Review Process</b>  <i>Review of Process</i></p> <p><b>Determine Next Steps</b>  <i>List of Possible Actions            Decision</i></p>	<p><b>Plan Deployment</b>  <i>Deployment Plan</i></p> <p><b>Plan Monitoring and Maintenance</b>  <i>Monitoring and            Maintenance Plan</i></p> <p><b>Produce Final Report</b>  <i>Final Report            Final Presentation</i></p> <p><b>Review Project</b>  <i>Experience            Documentation</i></p>

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

# CRISP-DM: Method & Task Levels

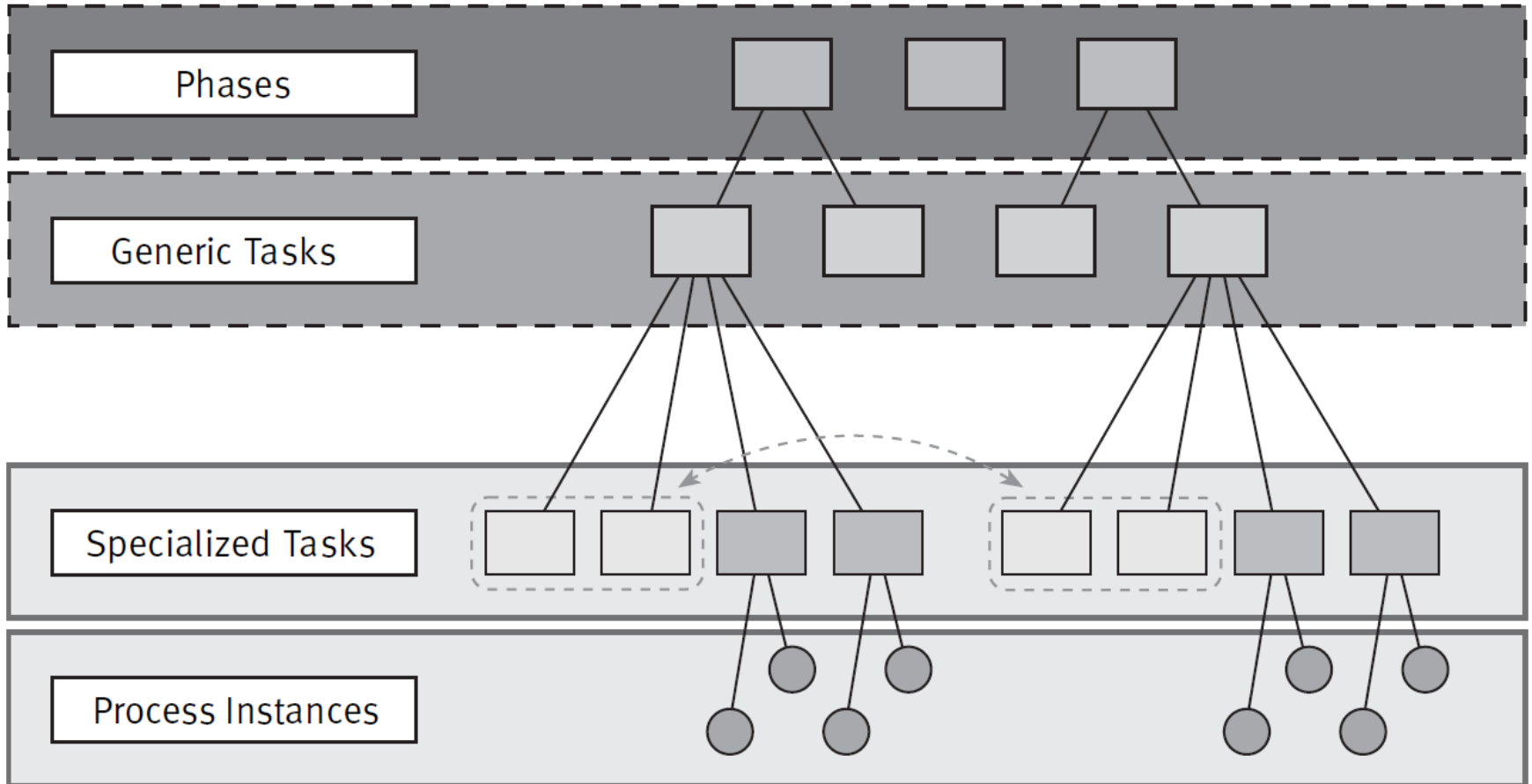


Figure 1: Four level breakdown of the CRISP-DM methodology



# 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'

**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.



26-02-2020

46

# 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  $\Rightarrow$  *Wlz/VV Zorgprofiel*
- Electronic Client Record data
  - Large valuable dataset, but
  - Not fully exploited at this moment



« [Naar zoeken](#)

## Regeling langdurige zorg

Geldig van 03-09-2020 t/m heden

Alles openklappen



Alles dichtklappen



### Inhoudsopgave

Opschrift



Aanhef



⊕ Hoofdstuk 1



Algemene bepalingen

(Artikel 1.1)

⊕ Hoofdstuk 2



De inhoud van de verzekering

(Artikelen 2.1-2.5)

⊕ Hoofdstuk 3



De indicatiebesluiten

### Bijlage A. bij [artikel 2.1](#) van de Regeling langdurige zorg

Zorgprofielen integraal pakket als bedoeld in [artikel 3.1.1, eerste lid, van het Besluit Wet Bijzondere Ziektekosten](#) en tevens met aanduiding van zorgprofielen waarboven

#### Sector Verpleging en verzorging (VV)

#### ZZP AWBZ

VV Beschut wonen met intensieve begeleiding en uitgebreide verzorging

4 VV

VV Beschermd wonen met intensieve dementiezorg

5 VV

VV Beschermd wonen met intensieve verzorging en verpleging

6 VV

VV Beschermd wonen met zeer intensieve zorg, vanwege specifieke aandoeningen, met de nadruk op begeleiding

7 VV\*

VV Beschermd wonen met zeer intensieve zorg, vanwege specifieke aandoeningen, met de nadruk op verzorging/verpleging

8 VV\*

VV Herstelgerichte behandeling met verpleging en verzorging

9b VV

#### Sector Verstandelijk Gehandicapt (VG)

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

# Research Method: CRISP-DM

- Cross Industry Standard Process 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

- $$\text{Score} = \sum_{\text{Expert level}} \frac{\text{Times mentioned}}{\text{Number of interviews}} \times \text{Valuation}$$

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

# 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 occurred	6	4	2	13.5
7	Q	Types of incidents occurred	6	4	2	13.5
8	Q	Causes of occurred 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...*

Spruit, M., Kais, M., & Menger, V. (2021). Automated Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding. *Future Internet*, 13(10), Trends of Data Science and Knowledge Discovery, 243. [[pdf](#)] [[online](#)]

---

*Verb keywords*

*Noun keywords*



*future internet*



---

*Article*

## **Automated Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding**

Marco Spruit <sup>1,2,3,\*</sup> , Marcin Kais <sup>3</sup> and Vincent Menger <sup>3</sup>

<sup>1</sup> Leiden University Medical Center (LUMC) Campus The Hague, Leiden University, Turfmarkt 99, 2511 DC The Hague, The Netherlands

<sup>2</sup> Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

# *Sidenote: Alternatively...*

Spruit, M., Kais, M., & Menger, V. (2021). Automated Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding. *Future Internet*, 13(10), Trends of Data Science and Knowledge Discovery, 243. [[pdf](#)] [[online](#)]

---

## *Verb 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

---

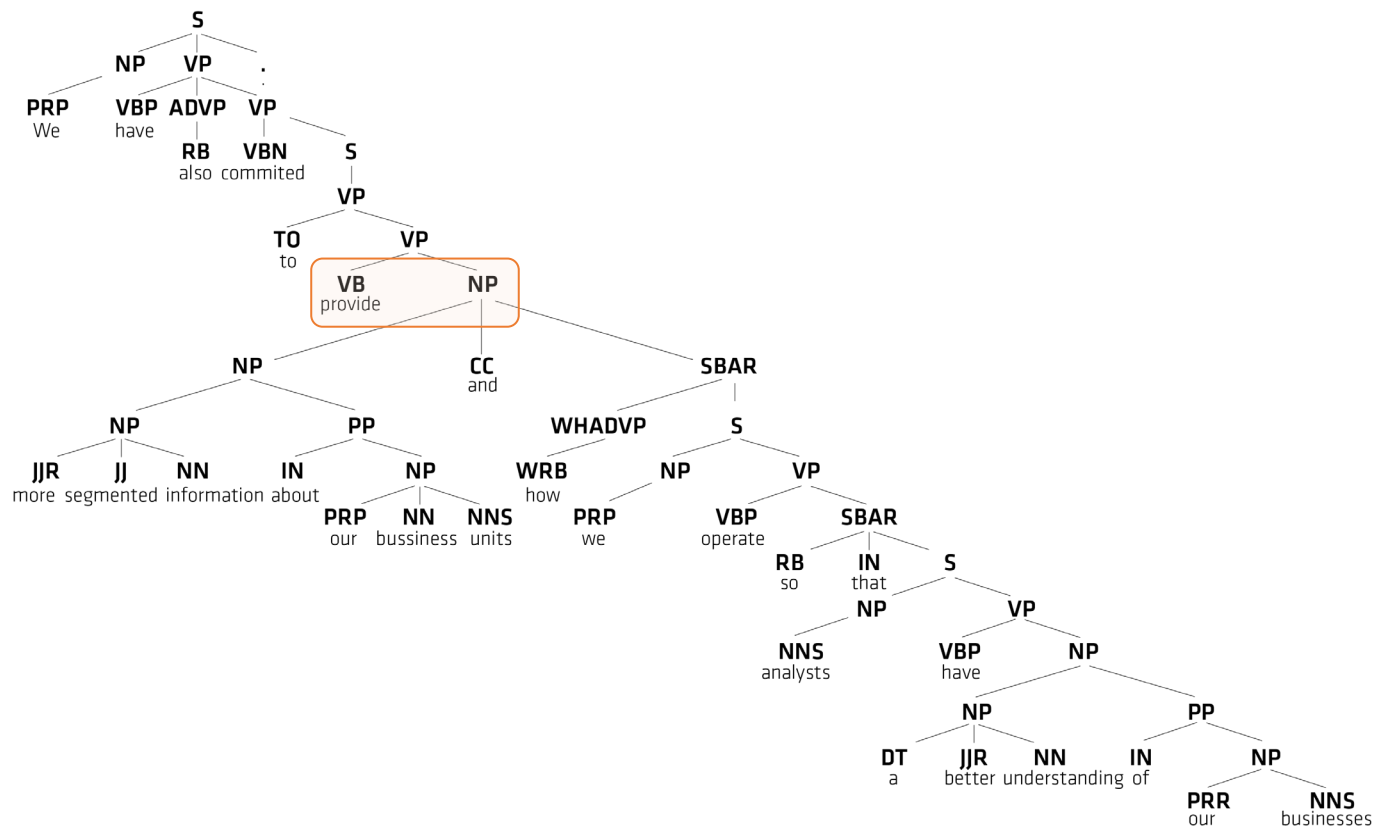
## *Noun keywords*

objective, aim, purpose, goal

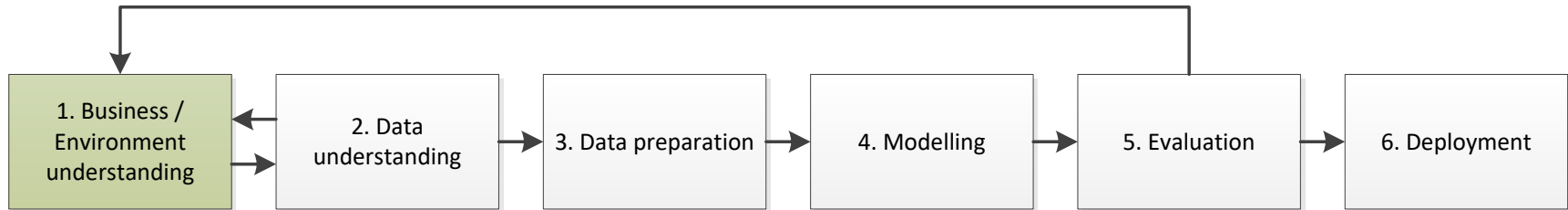
---

# Sidenote: Alternatively...

Spruit, M., Kais, M., & Menger, V. (2021). Automated Business Goal Extraction from E-mail Repositories to Bootstrap Business Understanding. *Future Internet*, 13(10), Trends of Data Science and Knowledge Discovery, 243. [[pdf](#)] [[online](#)]

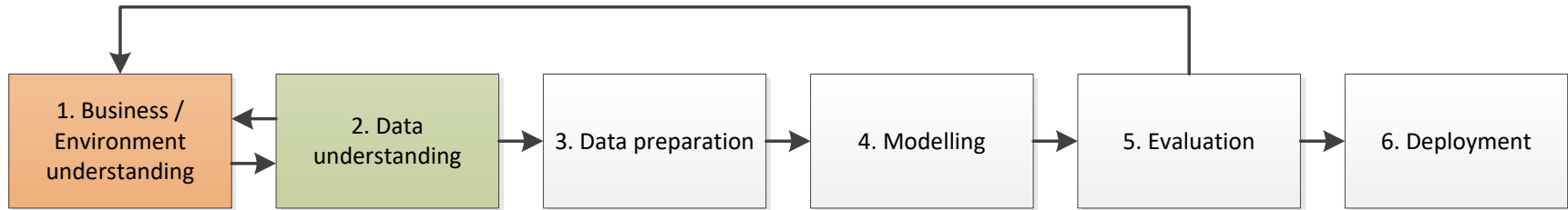


# Findings: Business understanding



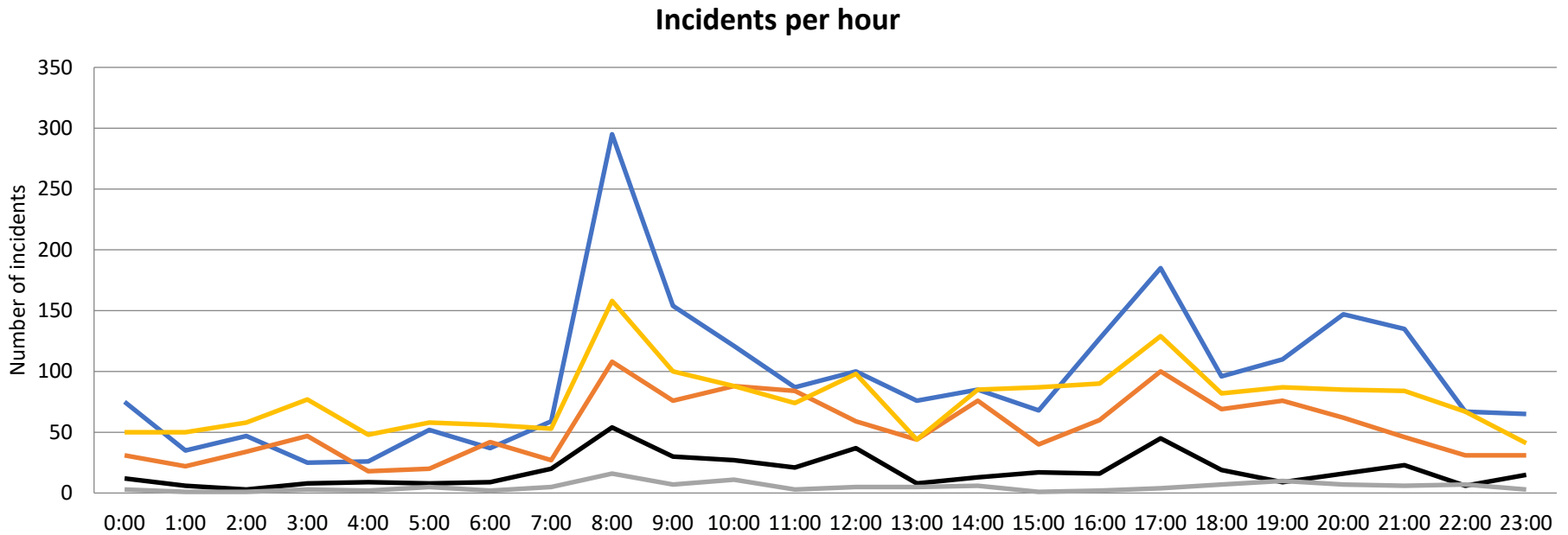
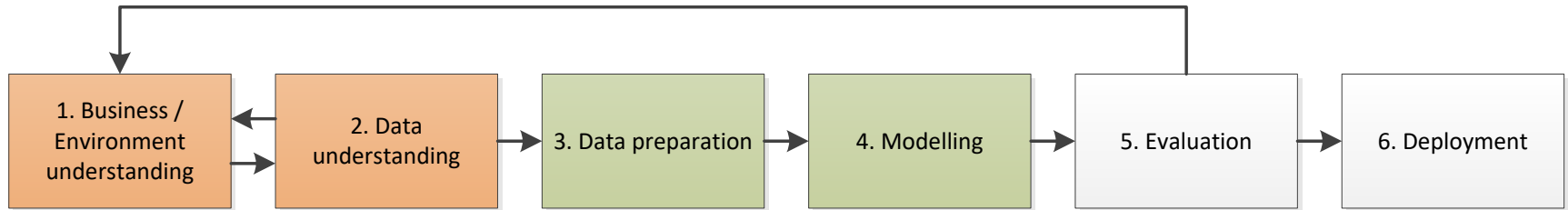
Information needs	Data mining goals
<ul style="list-style-type: none"> <li>• Number of occurred incidents</li> <li>• Types of occurred incidents</li> <li>• Causes of the occurred incidents</li> <li>• Patterns in occurred incidents</li> </ul>	<ul style="list-style-type: none"> <li>• Identify patterns in incidents</li> </ul>
<ul style="list-style-type: none"> <li>• Number of clients at an increased risk</li> <li>• Types of risk the clients run</li> </ul>	<ul style="list-style-type: none"> <li>• Identify relationships in risk assessment</li> </ul>
<ul style="list-style-type: none"> <li>• <b>Progress of care-related measures</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Identify patterns in care-related measures</b></li> </ul>
<ul style="list-style-type: none"> <li>• <b>Treatment goals (obtained &amp; not-obtained)</b></li> <li>• <b>Care plan information</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>Identify patterns in obtained and not-obtained treatment goals</b></li> </ul>
<ul style="list-style-type: none"> <li>• Number of clients per demand for care</li> <li>• ZZP-mix</li> <li>• ZZP-mix prognosis</li> </ul>	<ul style="list-style-type: none"> <li>• Identify &amp; predict the ZZP mix</li> </ul>

# Findings: Data understanding

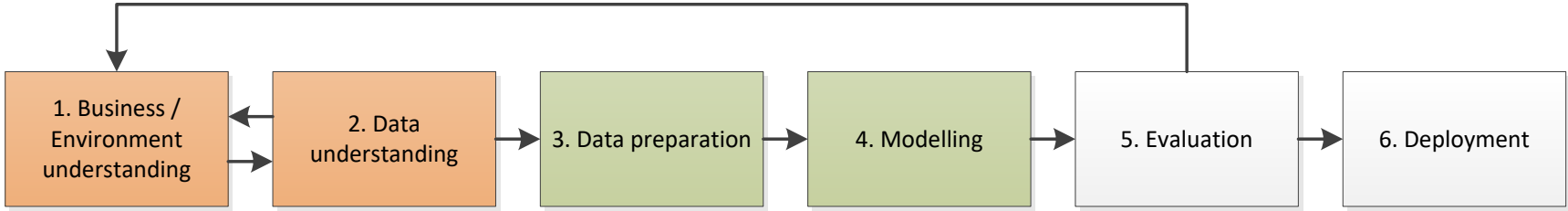


Data mining goals	Available data
<ul style="list-style-type: none"> <li>Identify patterns in incidents</li> </ul>	<ul style="list-style-type: none"> <li>5,692 records with incidents</li> <li>13 different incident types</li> </ul>
<ul style="list-style-type: none"> <li>Identify patterns in risk assessment</li> </ul>	<ul style="list-style-type: none"> <li>Depression risk: 2,129 records</li> <li>Falling risk: 889 records</li> <li>Incontinence risk: 877 records</li> <li>Medication risk: 806 records</li> <li>Problem behaviour risk: 0 records</li> <li>Weight risk: 567 records</li> </ul>
<ul style="list-style-type: none"> <li>Identify patterns in care-related measures</li> </ul>	<ul style="list-style-type: none"> <li>27,174 records with care-related measures</li> </ul>
<ul style="list-style-type: none"> <li>Identify patterns in obtained and not-obtained treatment goals</li> </ul>	<ul style="list-style-type: none"> <li>20,725 records with treatment goals</li> <li>14.59% obtained goals</li> <li>85.41% not obtained goals</li> </ul>
<ul style="list-style-type: none"> <li>Identify &amp; predict the ZPP mix</li> </ul>	<ul style="list-style-type: none"> <li>1,831 records with historical delivered ZPP's</li> </ul>

# Findings: Identify patterns in incidents

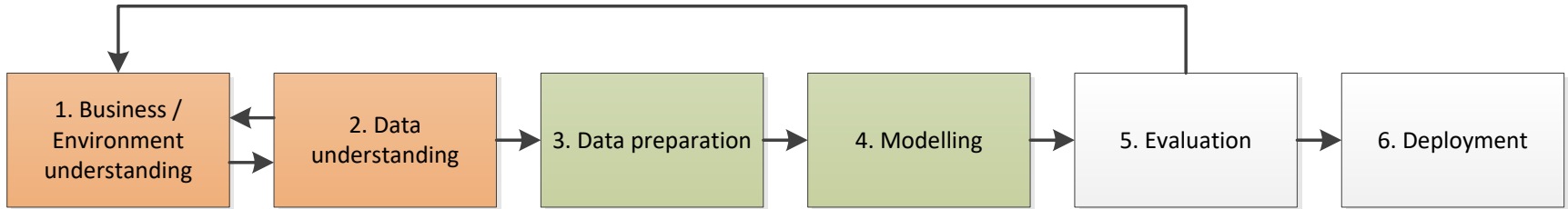


# Findings: Identify relationships in risk assessment

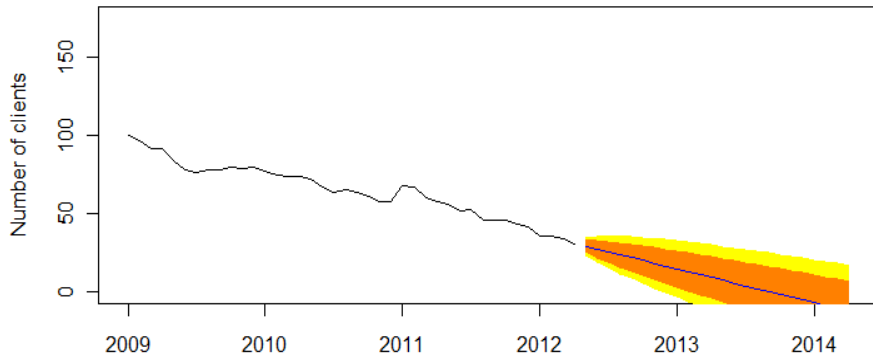


Rule	Support	Confidence	Lift
Incontinence, Medication → Falling	11.76%	90.63%	1.824
Falling, Medication → Incontinence	11.76%	95.08%	1.769
Incontinence, Weight intramural → Falling	11.56%	87.69%	1.765
Falling, Weight intramural → Incontinence	11.56%	93.44%	1.738
Depression, Incontinence → Falling	31.64%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

# Findings: Predict the ZZP mix

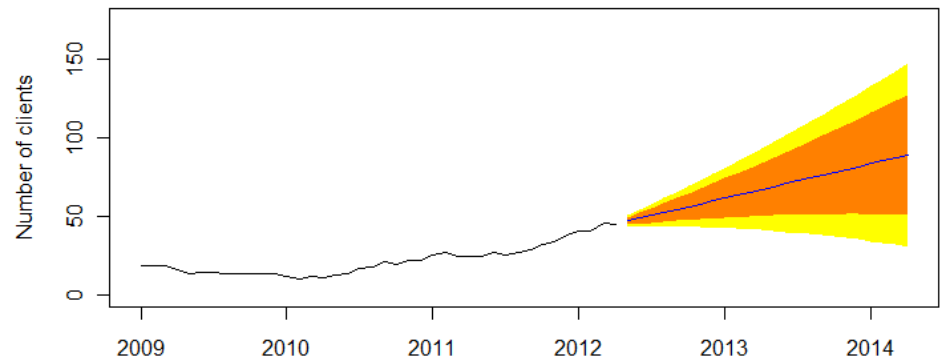


Forecasts from ETS(A,A,N)



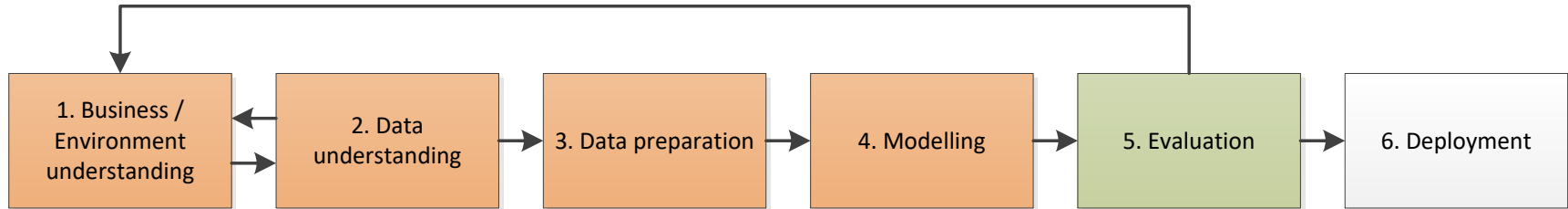
**ZZP1**

Forecasts from ETS(A,A,N)



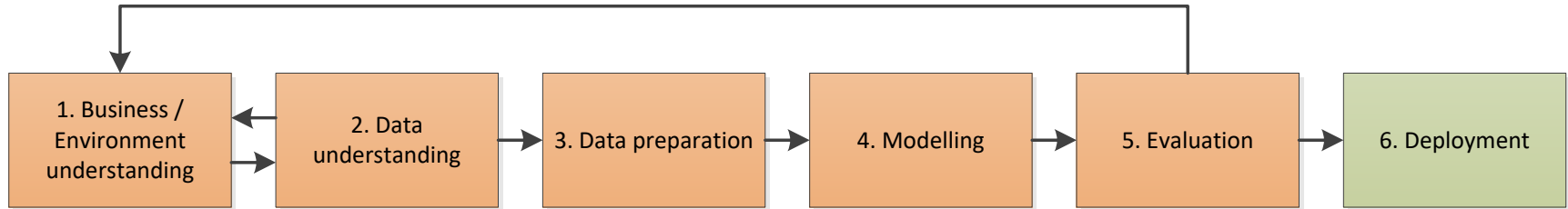
**ZZP6**

# Findings: Evaluation



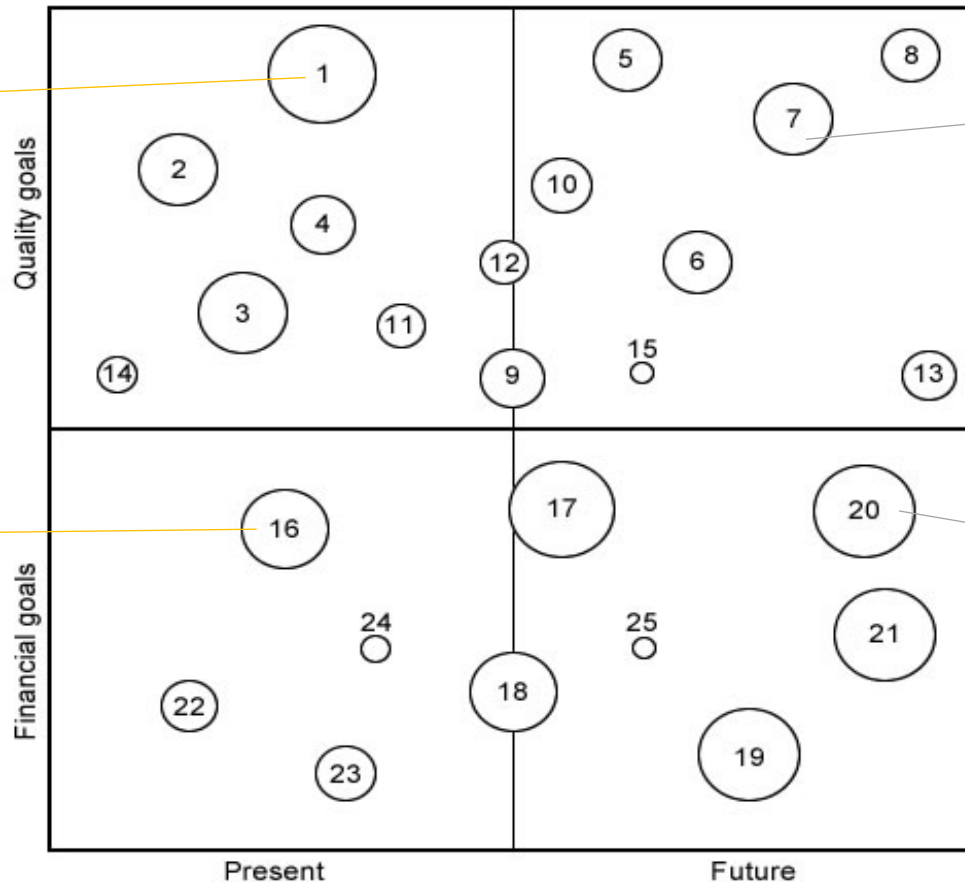
- Explorative research → no groundbreaking results
  - Information needs of *multiple* care institutions & stakeholders
- Lack of standardization
  - Heterogeneous data
- Predictions are too dependent on environmental factors
  - Limited historical data
  - Too many dependent factors for the forecasts 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)

# *Thank you*

[m.r.spruit@lumc.nl](mailto:m.r.spruit@lumc.nl)



**Leiden University**  
**Campus The Hague**