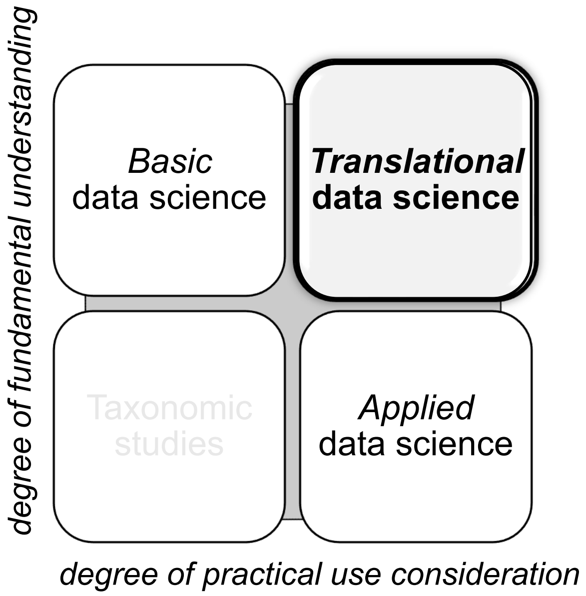
Health Campus Den Haag

Translational Data Science in Population Health

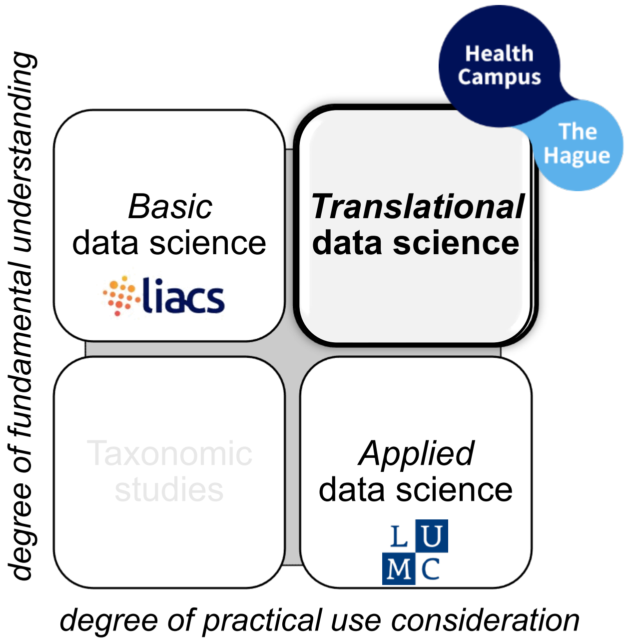
CRISP-DM Methodology in the TDS Lab

Amsterdam Public Health (APH) methodology kickoff workshop, 30 October 2023

Prof.dr. Marco Spruit

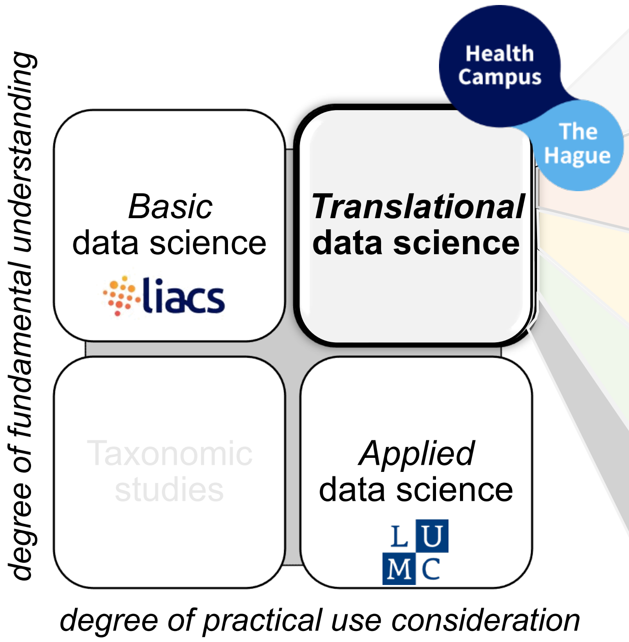


Basic data science
***Translational* data science**
Taxonomic studies
Applied data science
degree of fundamental understanding
degree of practical use consideration



degree of fundamental understanding
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of practical use consideration

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

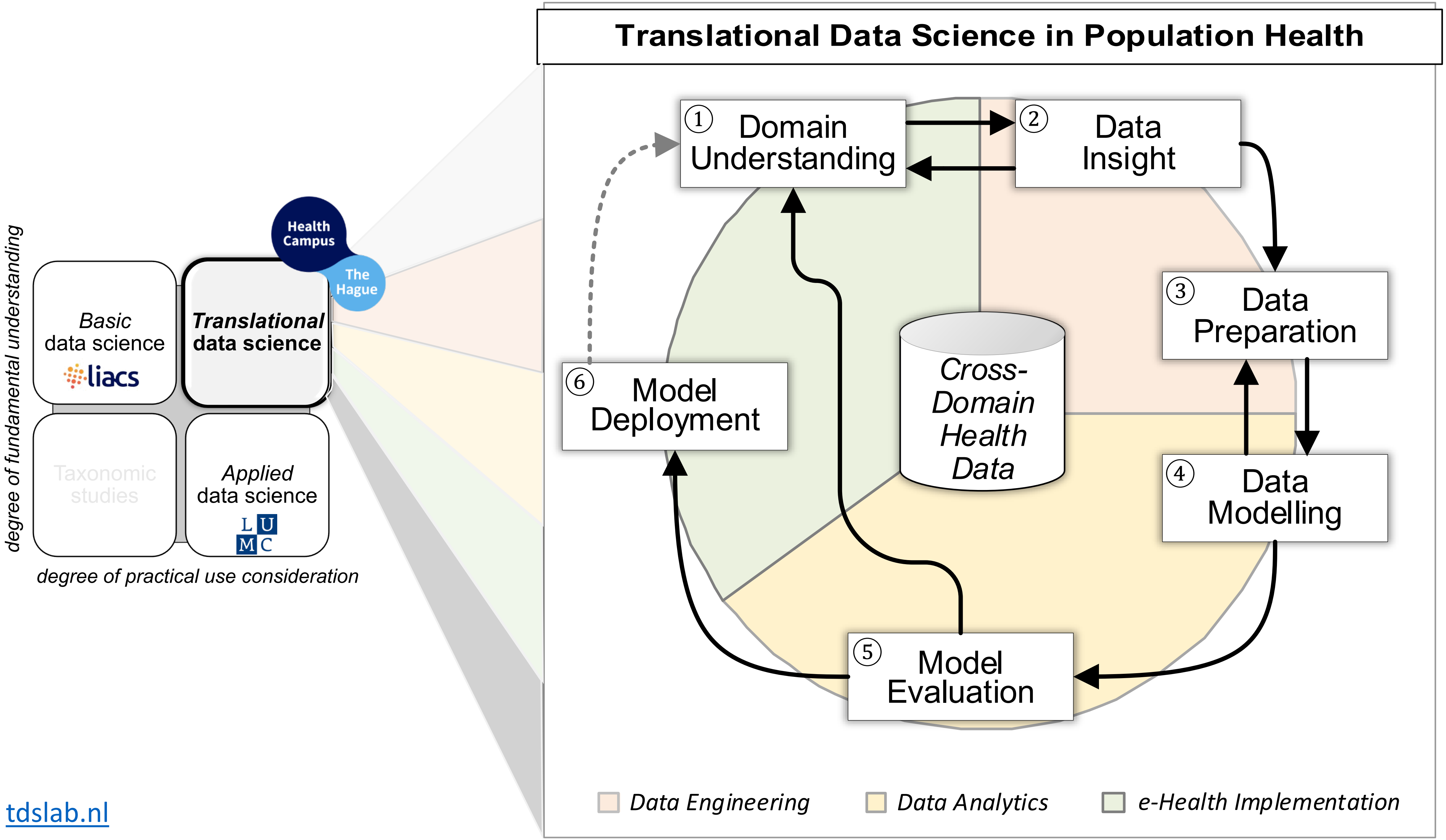
- Ph.D researcher in Computational Linguistics, University of Amsterdam

2007

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

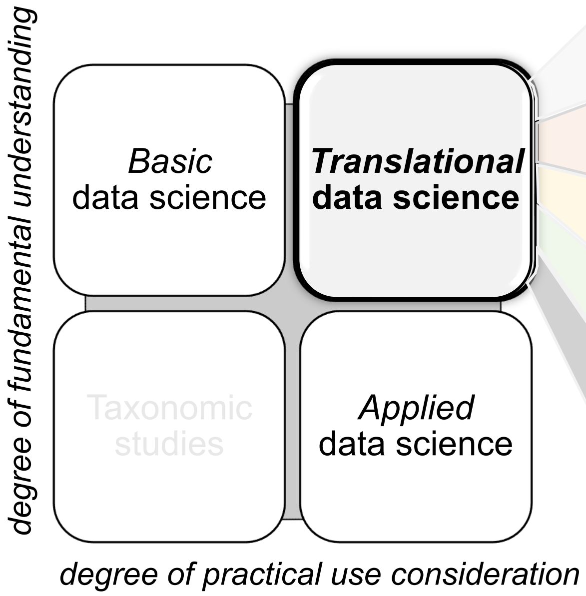
2020

- Professor Translational Data Science in Population Health, LUMC/Leiden University
  - **PH Living Lab**
  - **CAIRE Lab**
  - SIG Health Data Science



Translational Data Science in Population Health
degree of fundamental understanding
degree of practical use consideration
Basic data science
liacs
Translational data science
Health Campus The Hague
Taxonomic studies
Applied data science
LUMC
① Domain Understanding
② Data Insight
③ Data Preparation
④ Data Modelling
⑤ Model Evaluation
⑥ Model Deployment
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation

tdslab.nl

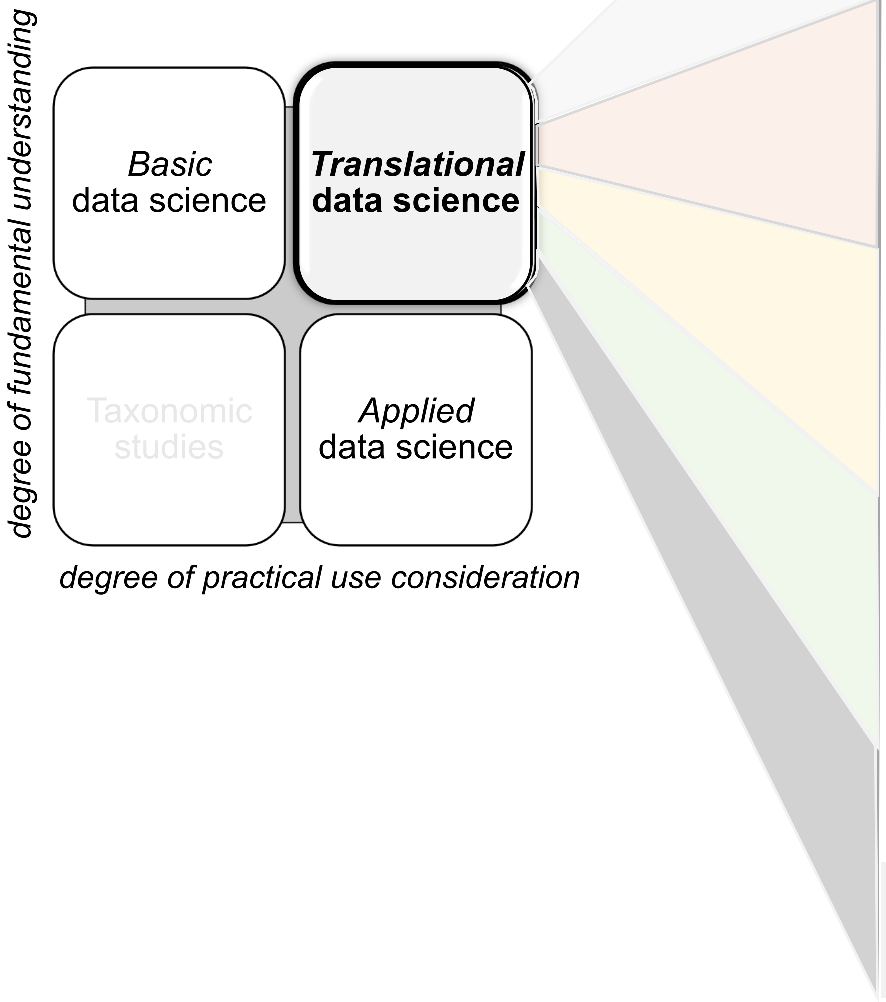


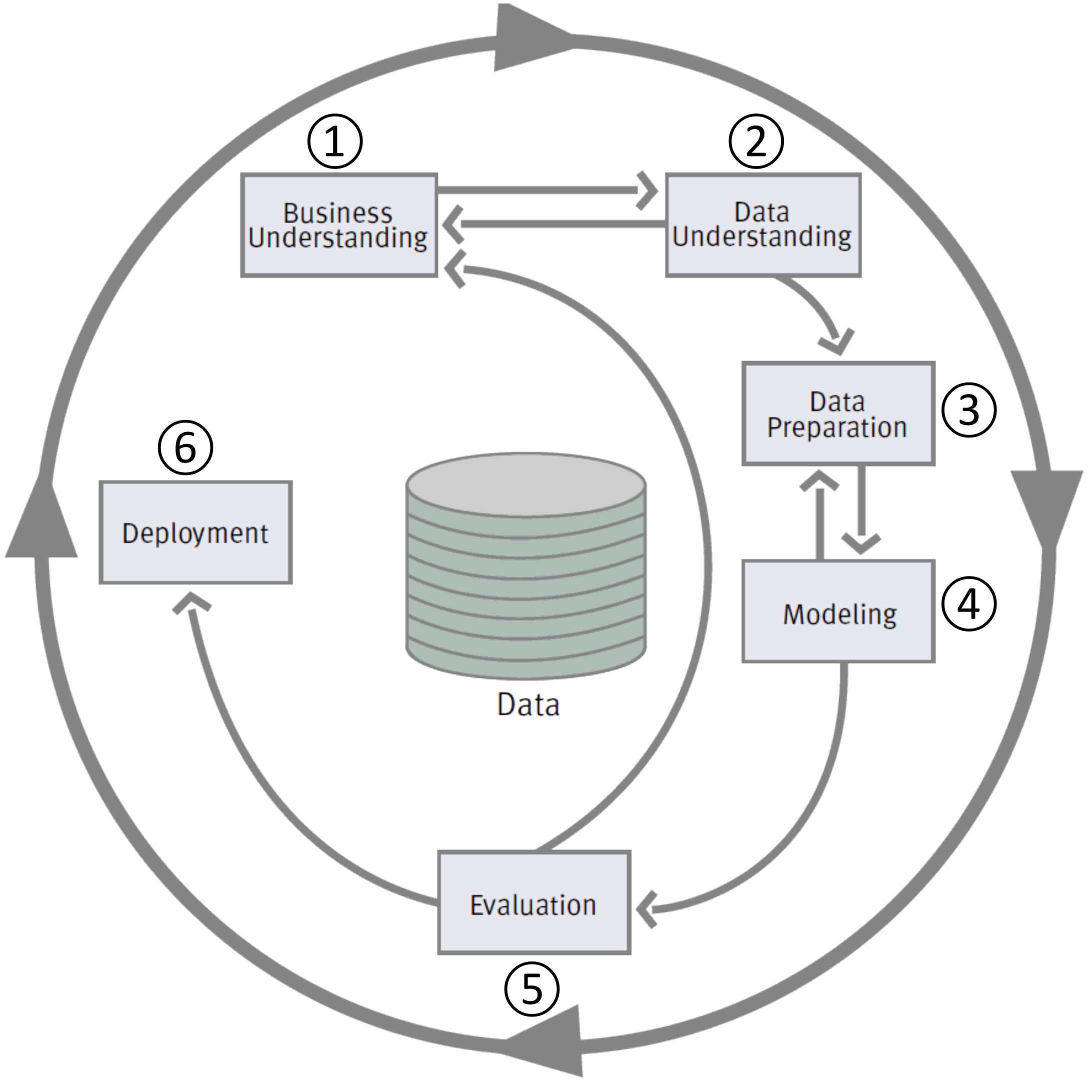
What main methodology are you using for your analytics, data mining, or data science projects ? [200 votes total]

2014 poll | 2007 poll

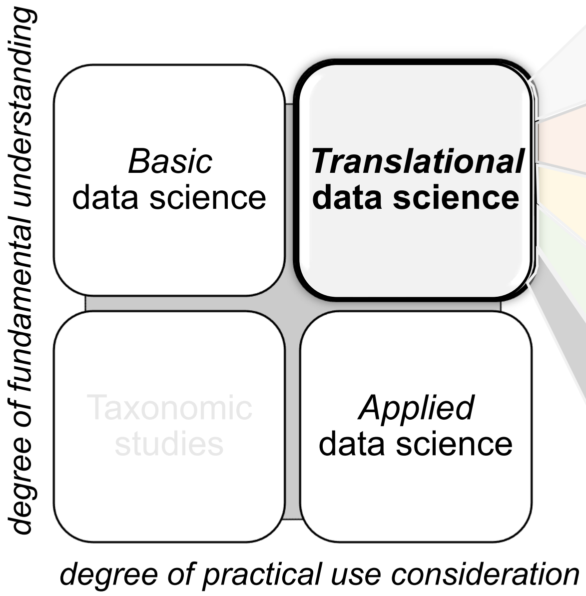
?

https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html



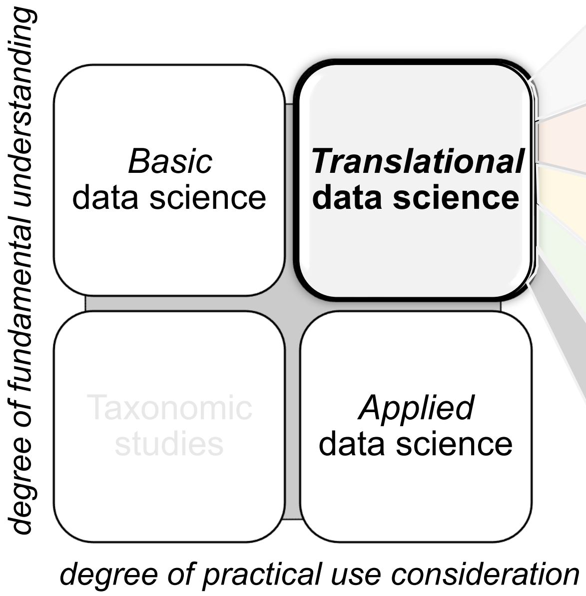


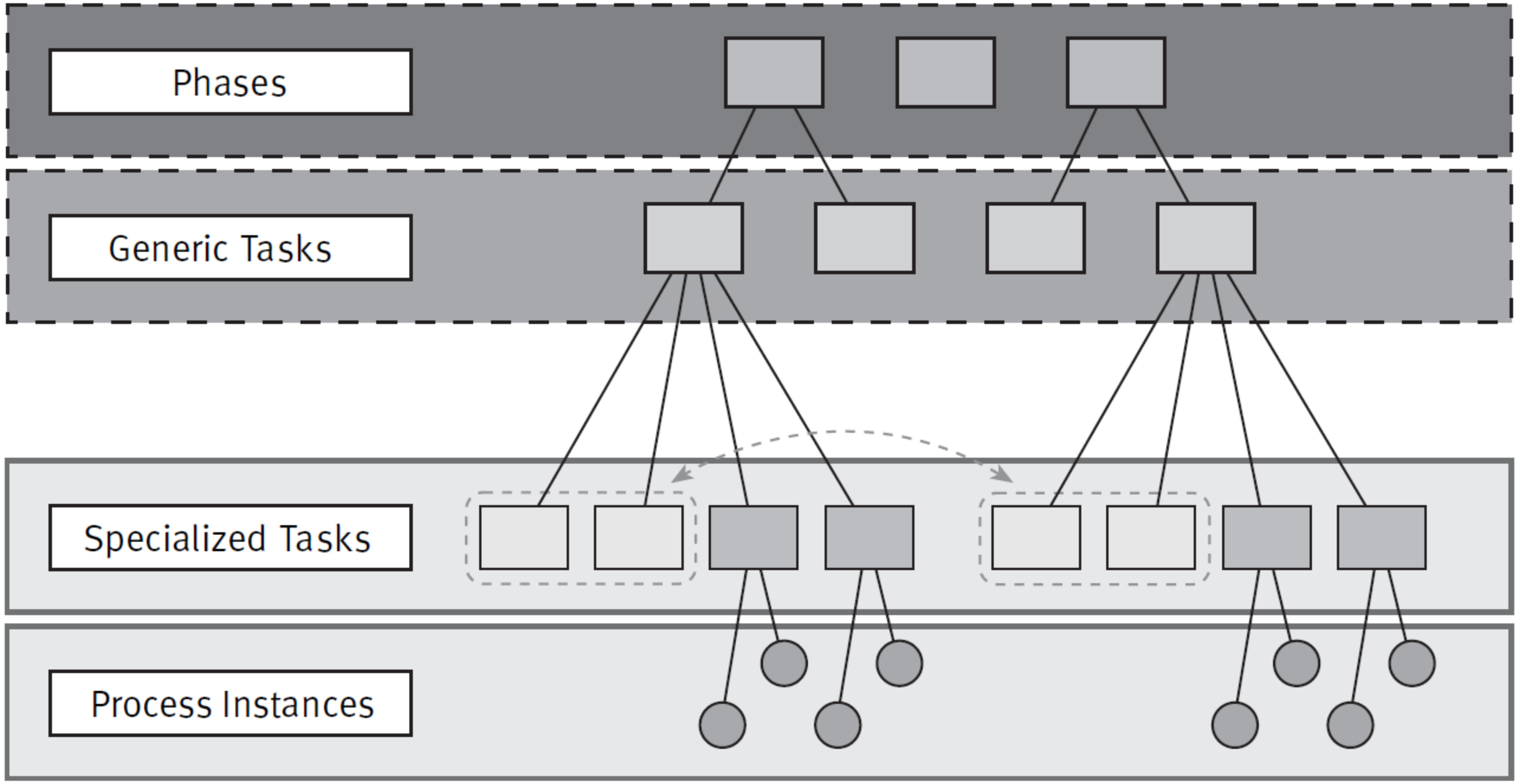
Chapman, P. Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). ***CRISP-DM** 1.0 Step-by-step Data Mining Guide (p.6).* [online]



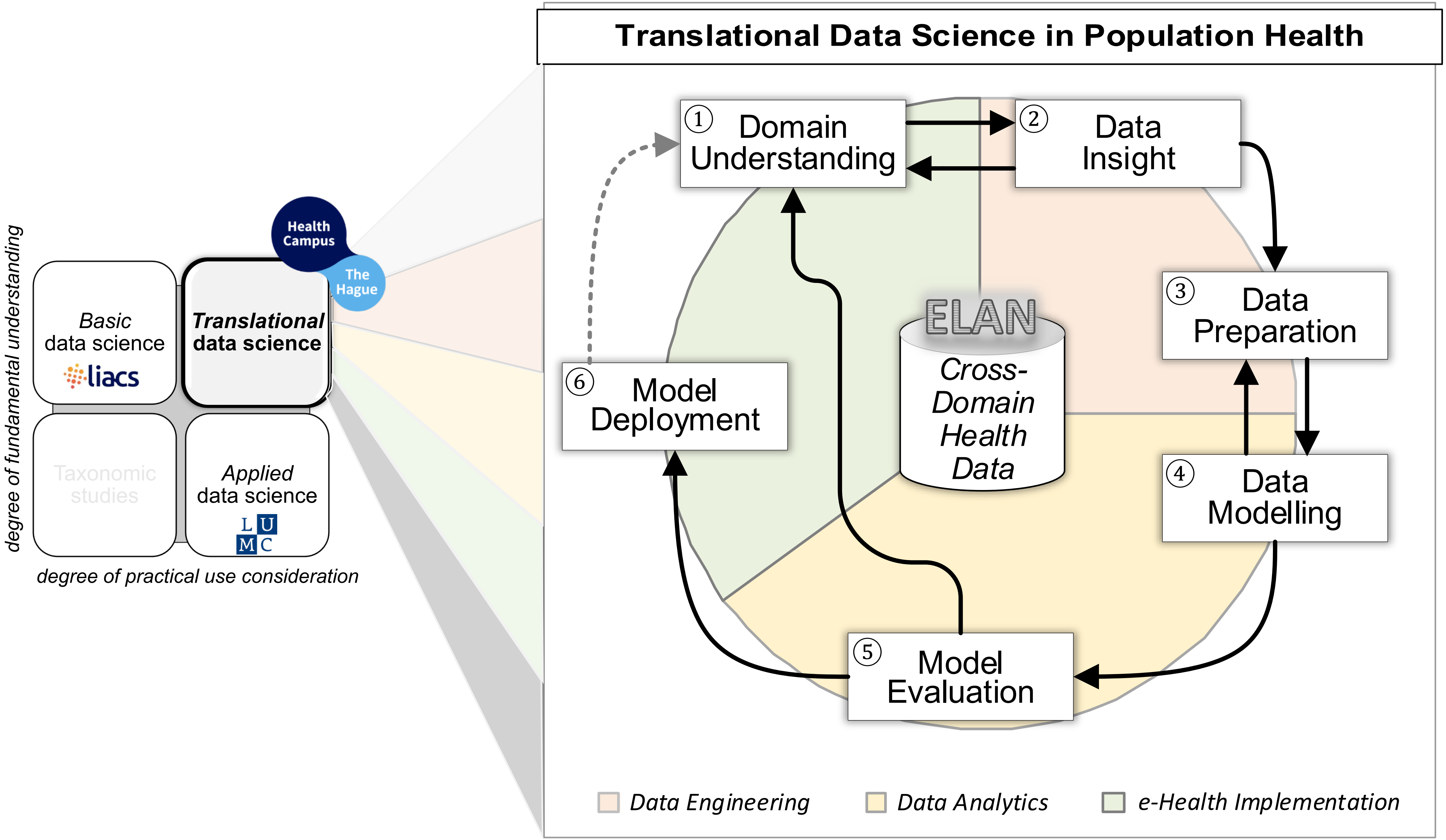
Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria* **Assess Situation** *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology*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

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 (p.6).* [online]

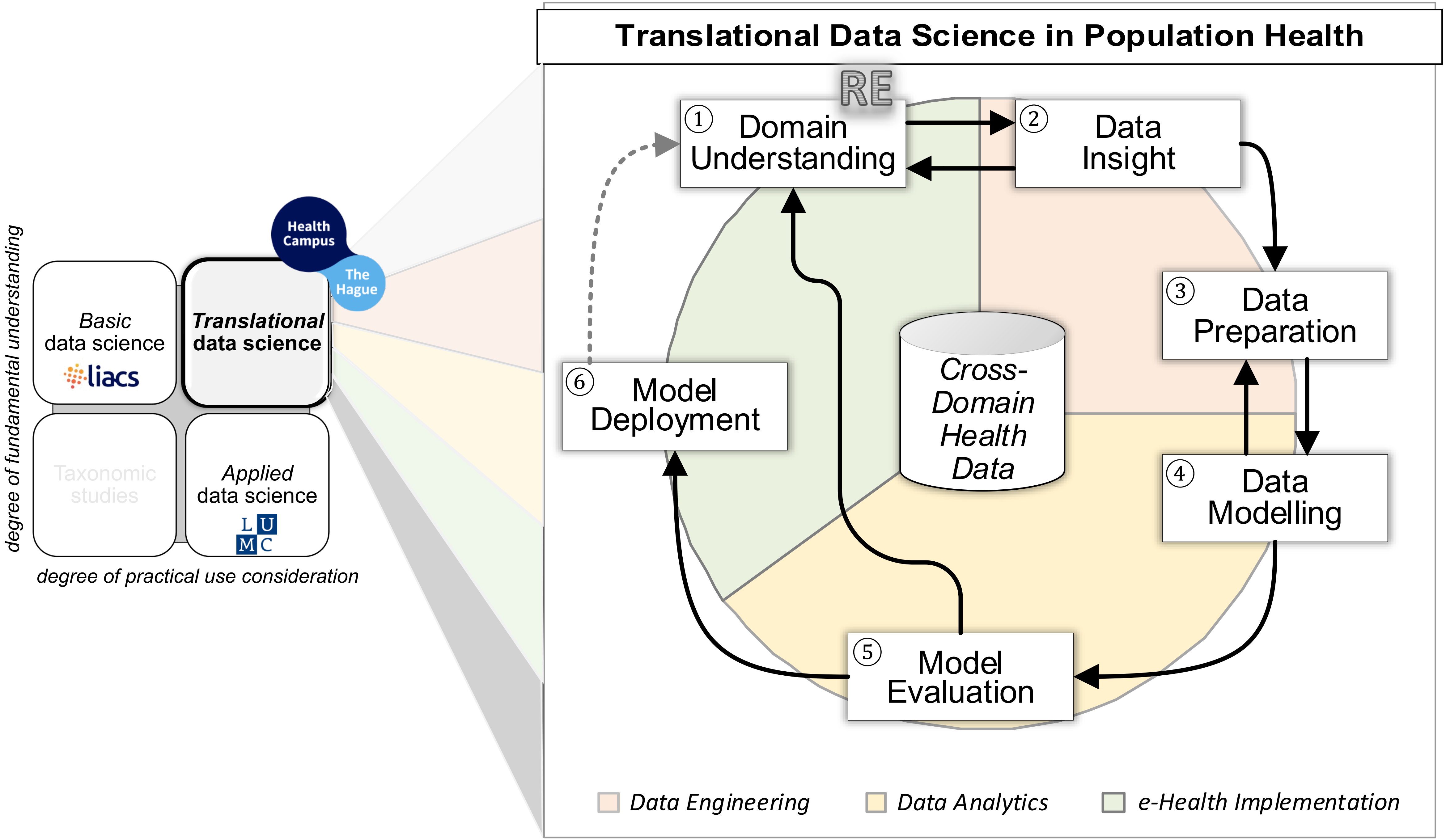




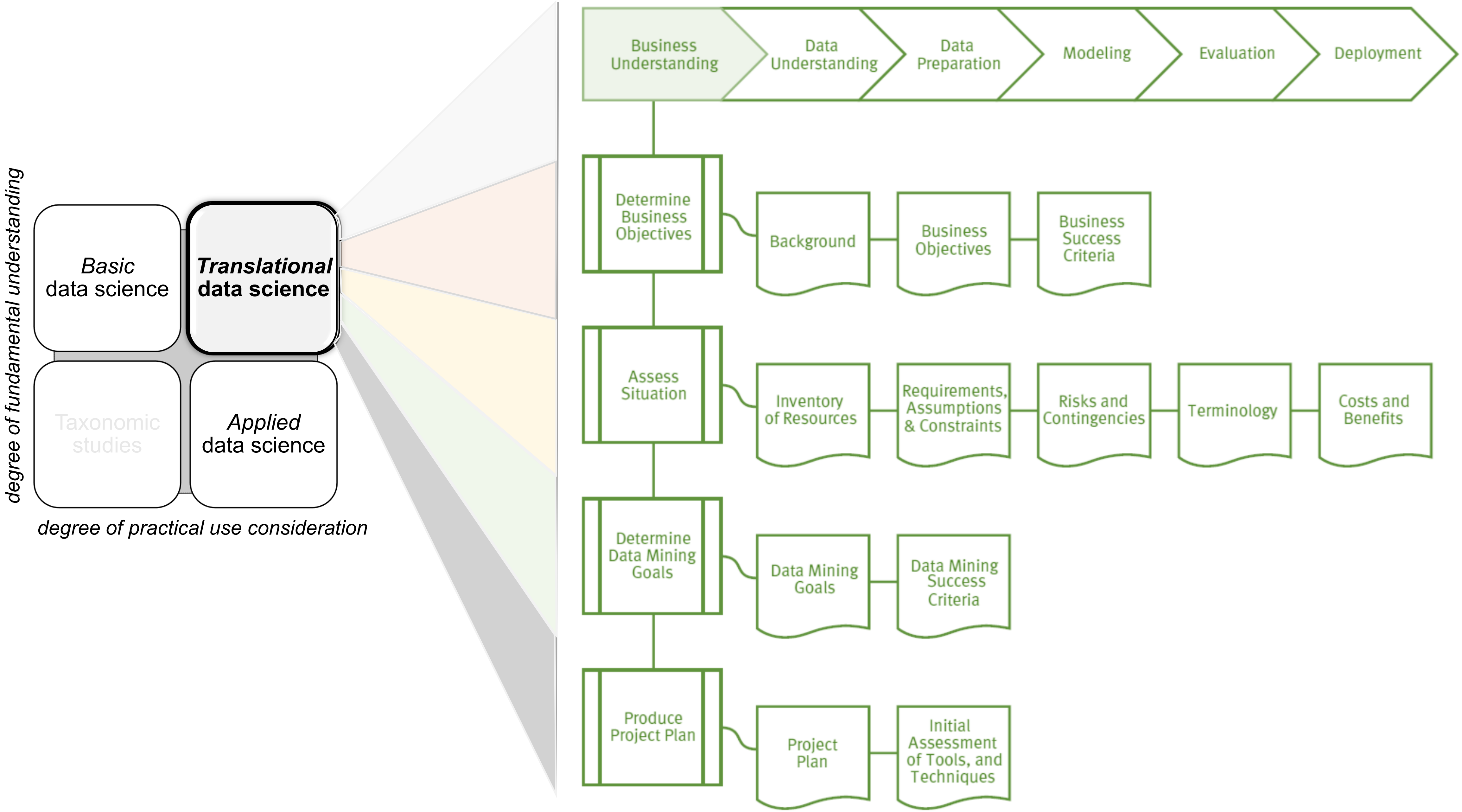
Chapman, P. Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). ***CRISP-DM** 1.0 Step-by-step Data Mining Guide (p.6).* [online]



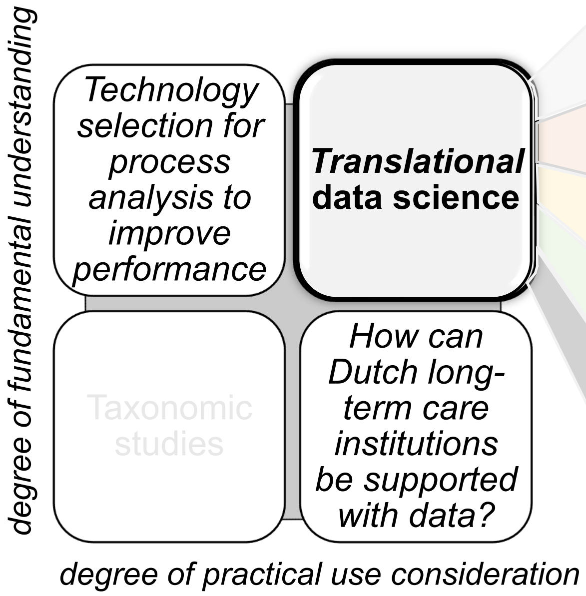
Translational Data Science in Population Health
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of fundamental understanding
degree of practical use consideration
① Domain Understanding
② Data Insight
③ Data Preparation
④ Data Modelling
⑤ Model Evaluation
⑥ Model Deployment
ELAN
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation



Translational Data Science in Population Health
RE
① Domain Understanding
② Data Insight
③ Data Preparation
④ Data Modelling
⑤ Model Evaluation
⑥ Model Deployment
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of fundamental understanding
degree of practical use consideration



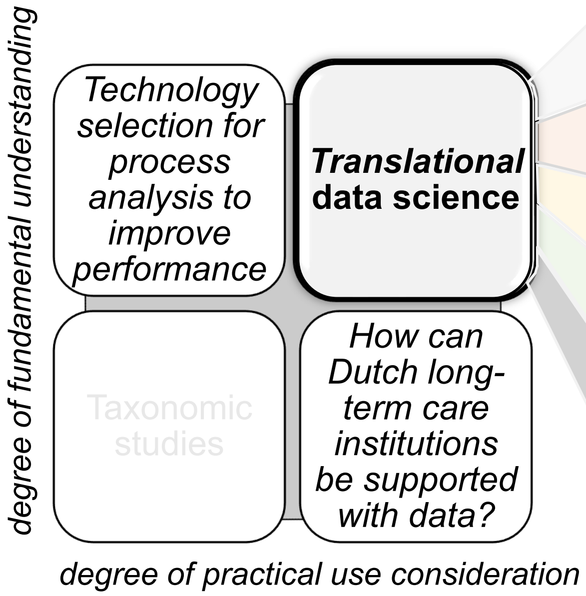
degree of fundamental understanding
Basic data science
Translational data science
Taxonomic studies
Applied data science
degree of practical use consideration
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Determine Business Objectives
Background
Business Objectives
Business Success Criteria
Assess Situation
Inventory of Resources
Requirements, Assumptions & Constraints
Risks and Contingencies
Terminology
Costs and Benefits
Determine Data Mining Goals
Data Mining Goals
Data Mining Success Criteria
Produce Project Plan
Project Plan
Initial Assessment of Tools, and Techniques



Example: Understanding Long-term Care

- Portion of Dutch healthcare budget: 38% = 34B!
- Care Intensity Package (ZZP)
  - *ZZP1:* Extramural living with some guidance
  - *ZZP8:* Intramural living under full surveillance and 24/7 care
- 18 unstructured in-depth interviews
  - 8 (board of) directors experts
  - 7 management experts
  - 7 experts from stakeholders perspective (MinVWS, IGZ, Care insurer)
- 56 information needs derived (33Q, 23F) from 18 unstructured in-depth interviews with 22 experts

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. [online]

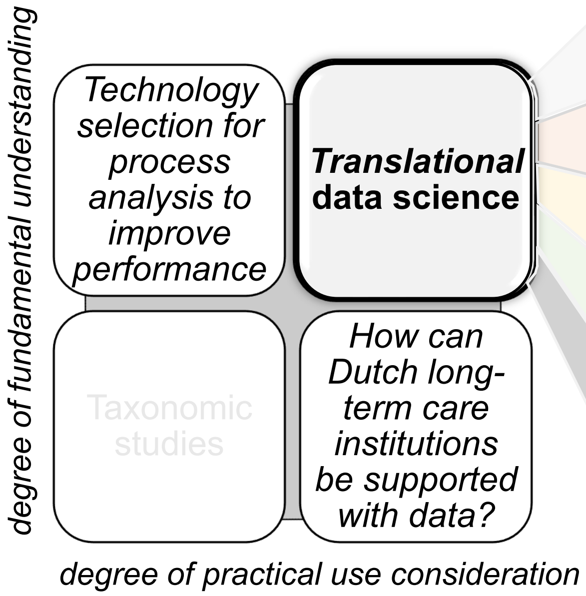


Example: Understanding Long-term Care

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

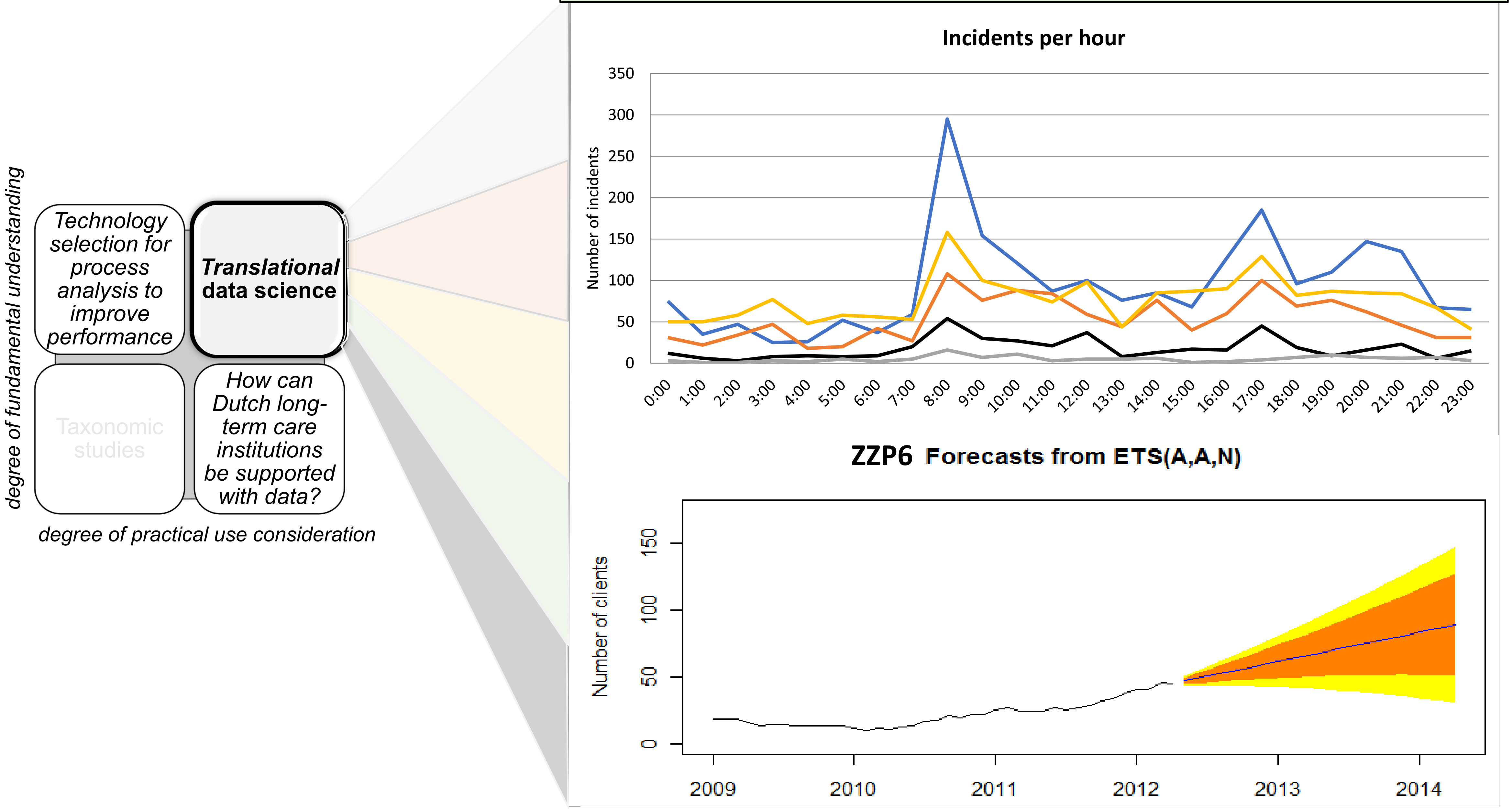
#	Type	Information need	Score
1	Q	Customer experience	16.6
2	F	Staffing with respect to ZZP-mix	14.8
3	F	ZZP-mix per business unit	13.6
4	F	ZZP-mix prognoses	13.6
5	F	Staffing with respect to operations	13.5
6	Q	Number of incidents occured	13.5
7	Q	Types of incidents occured	13.5
8	Q	Causes of occured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

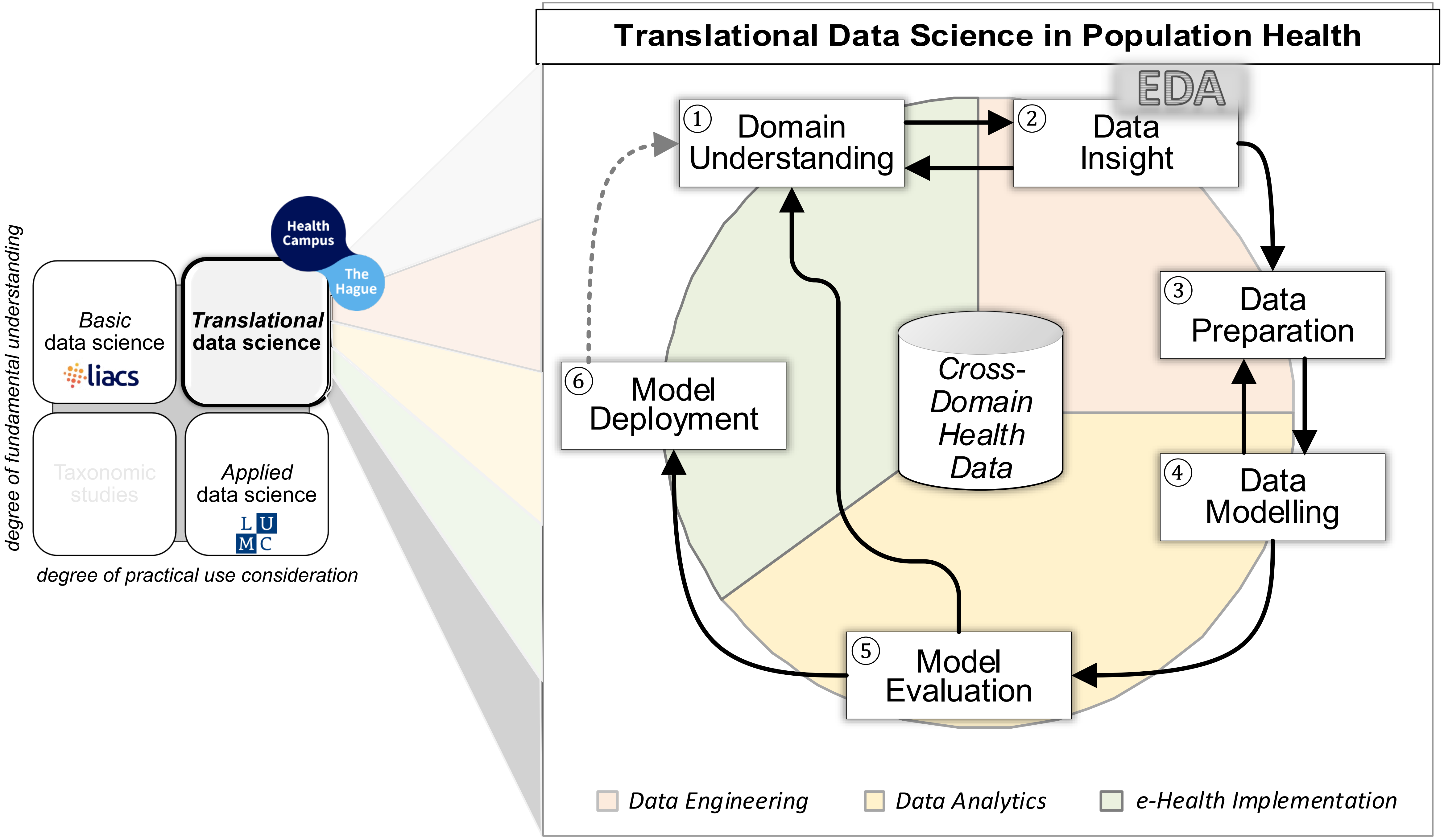
Example: Understanding Long-term Care



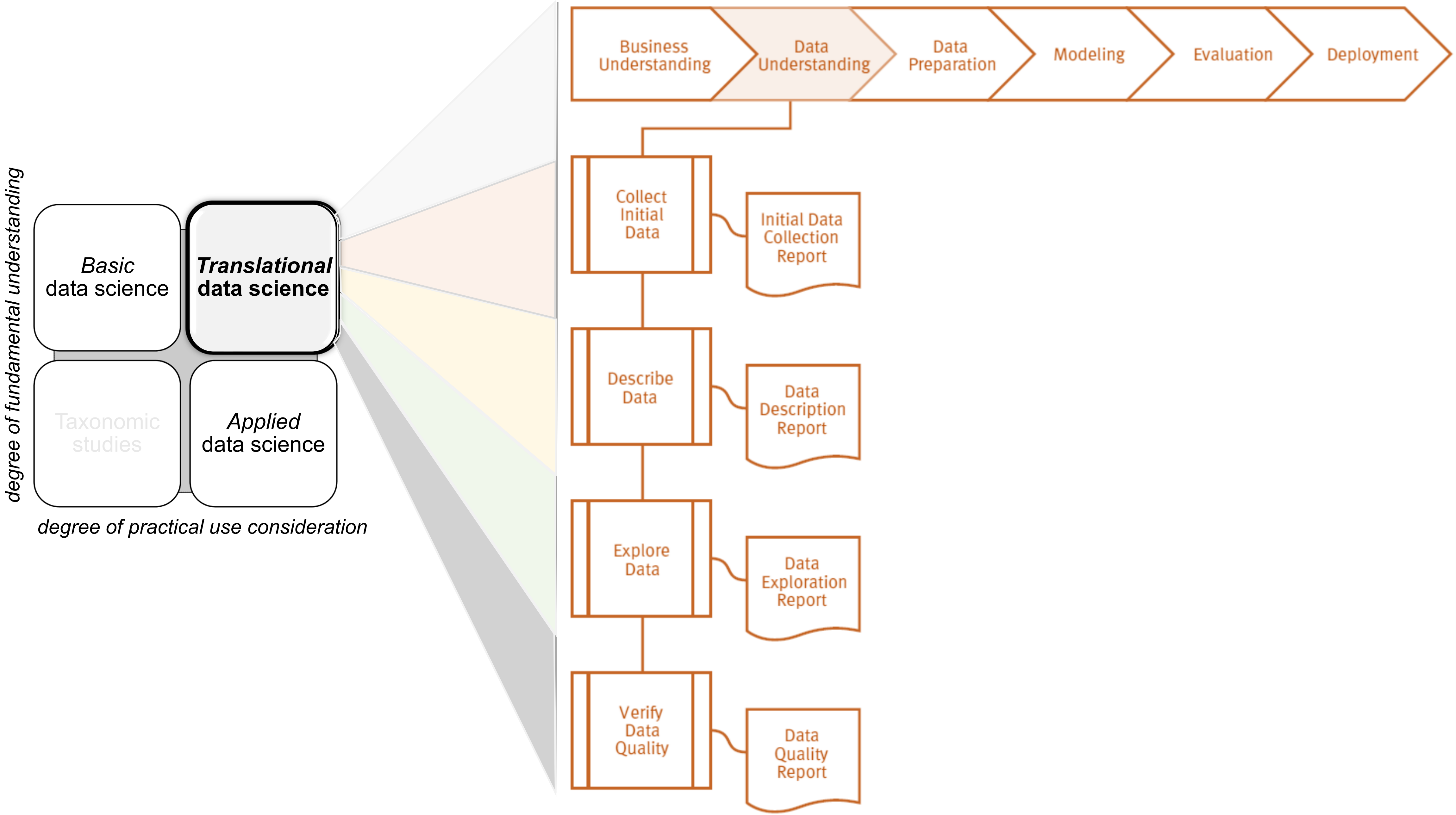
Information needs	*Data mining goals*
• Number of occured incidents • Types of occured incidents • Causes of the occured incidents • Patterns in occured incidents	• Identify patterns in incidents [1]
• Number of clients at an increased risk • Types of risk the clients run	• Identify relationships in 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 [2]

Example: Understanding Long-term Care

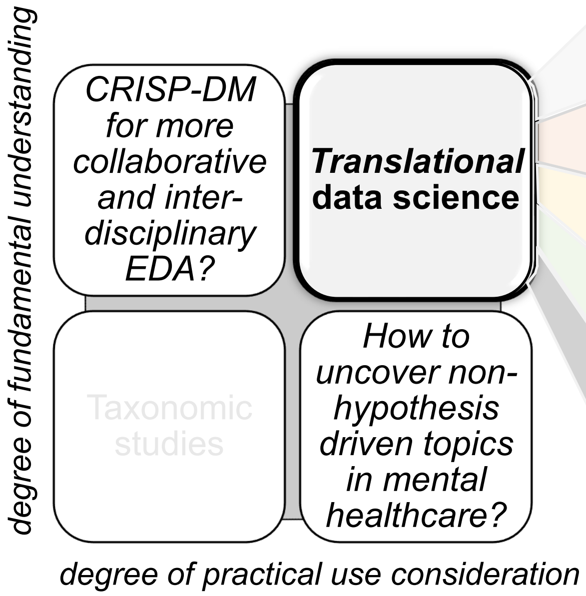




Translational Data Science in Population Health
EDA
1 Domain Understanding
2 Data Insight
3 Data Preparation
4 Data Modelling
5 Model Evaluation
6 Model Deployment
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of fundamental understanding
degree of practical use consideration


degree of fundamental understanding
Basic data science
Translational data science
Taxonomic studies
Applied data science
degree of practical use consideration
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Collect Initial Data
Initial Data Collection Report
Describe Data
Data Description Report
Explore Data
Data Exploration Report
Verify Data Quality
Data Quality Report

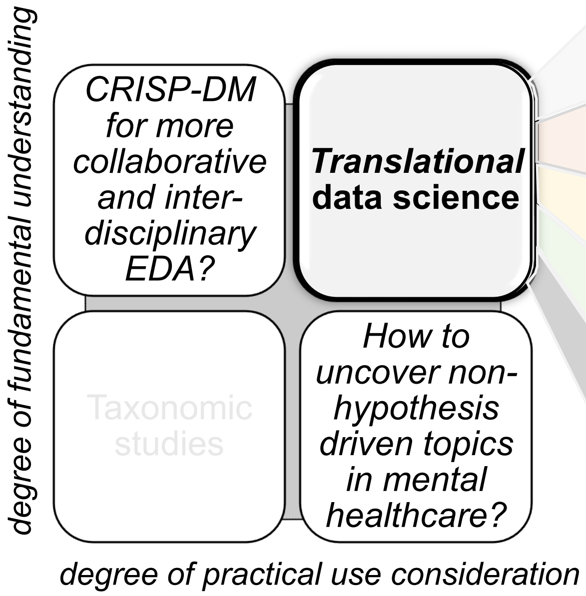
Example: Exploring Mental Healthcare



- CRISP-IDM: Cross Industry Standard Process for *Interactive* Data Mining
- *"Big data"* approach: *not* hypothesis-driven

Topic	*Theme*	*Priority*
What are relations between the different ROM scores, and can they predict treatment length?	ROM	1
Do medication prescription and change in medication influence the length of admission and the likeliness of readmission?	Medication	2
Can aggression incidents in inpatients be predicted?	Aggression	3
In what way are patients referred between, for example, general practitioners, secondary care institutions, and the UMCU?	Patient referrals	4

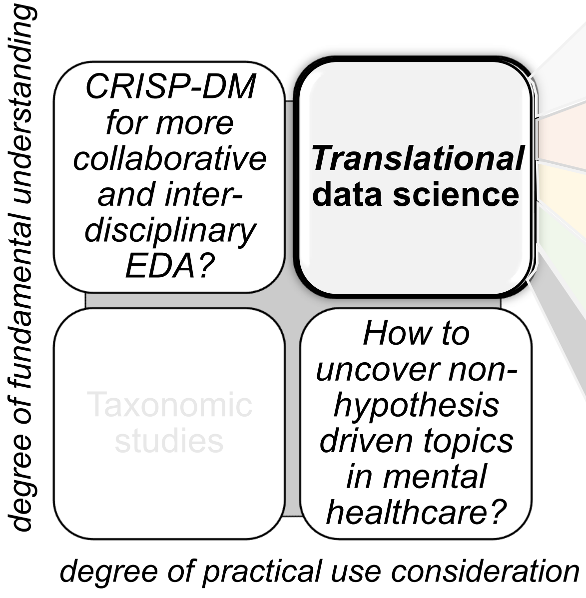
Menger,V., Spruit,M., Hagoort,K., & Scheepers,F. (2016). Transitioning to a data driven mental health practice: collaborative expert sessions for knowledge and hypothesis finding. *Computational and Mathematical Methods in Medicine*, Article ID 9089321, 11. [online]



Example: Exploring Mental Healthcare

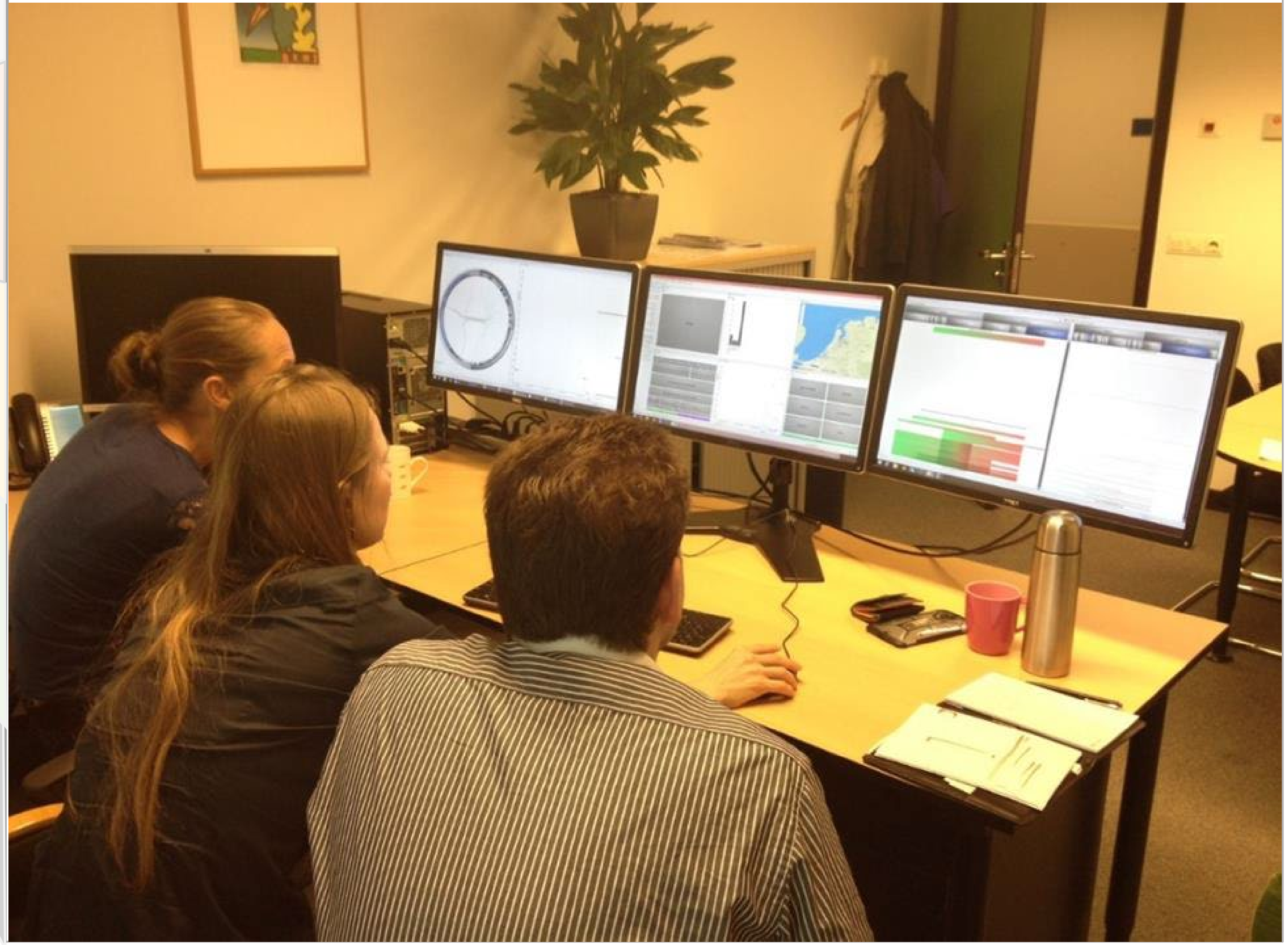
- *Data descr.:* EHR, Incident report system, External

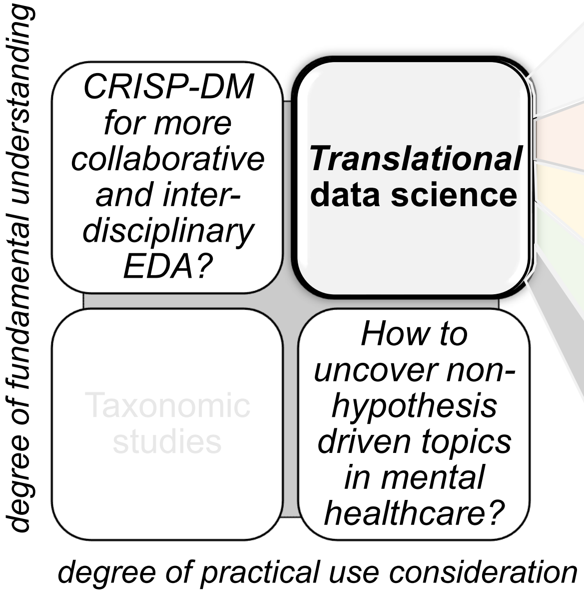
			Records
(1) Diagnosis	Categoric	Structured	5,800
(2) Treatment plan	Categoric, textual	Both	6,500
(3) Medication prescriptions	Categoric, numeric	Structured	22,000
(4) Routine Outcome Monitoring	Numeric, textual	Both	13,000
(5) Admission information	Categoric	Structured	5,400
(6) Daily reports	Textual	Unstructured	150,000
(7) Aggression incident reports	Categoric, textual	Both	1,200
(8) Census data	Numeric	Structured	21,000
(9) Geographic data	Numeric	Structured	5,000



Example: Exploring Mental Healthcare

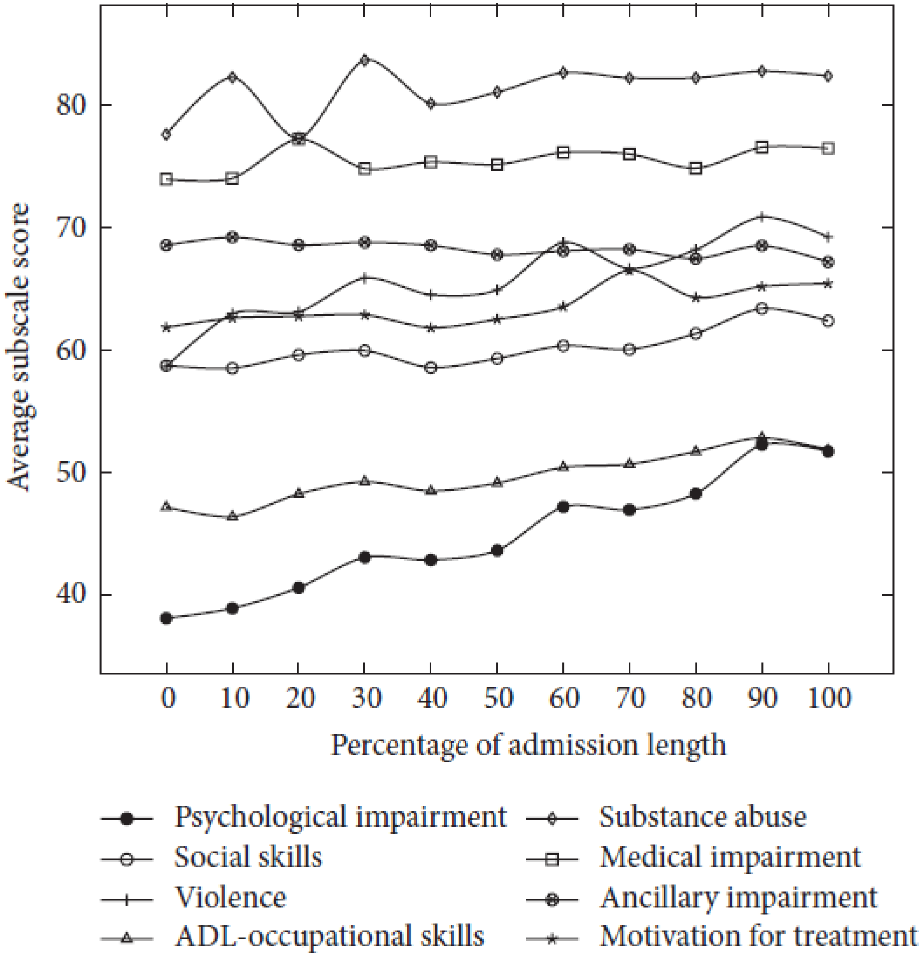
- 26 Weekly interactive data visualization explorations!

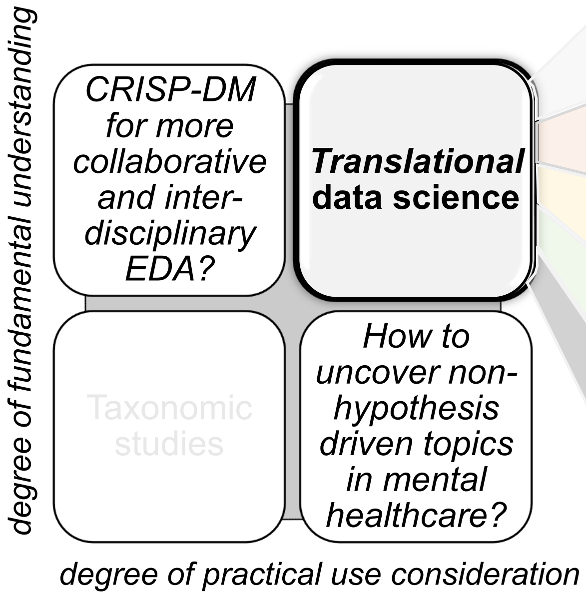




Example: Exploring Mental Healthcare

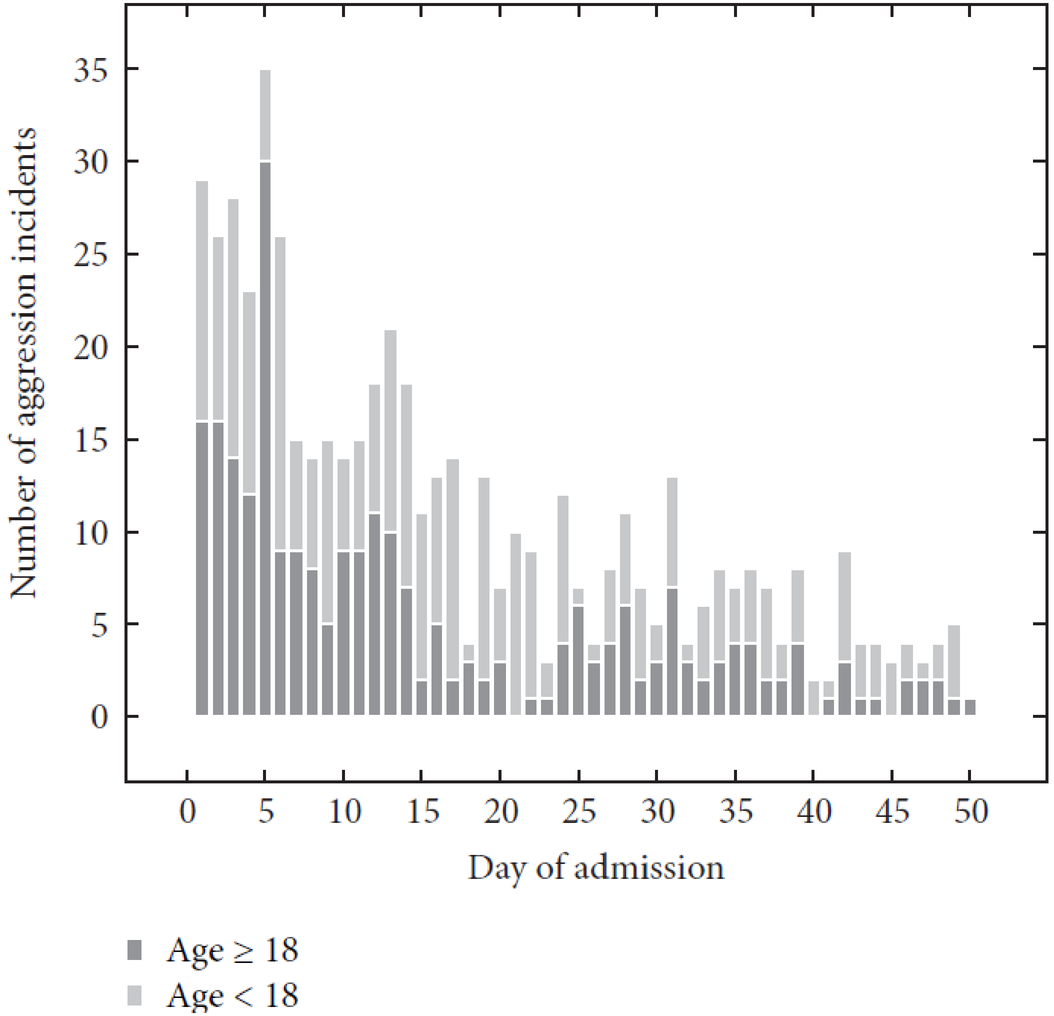
- *Finding:* Domain experts indicated that the lack of variation does not justify scoring the Kennedy Axis V (a ROM on Well-being) on a regular basis.

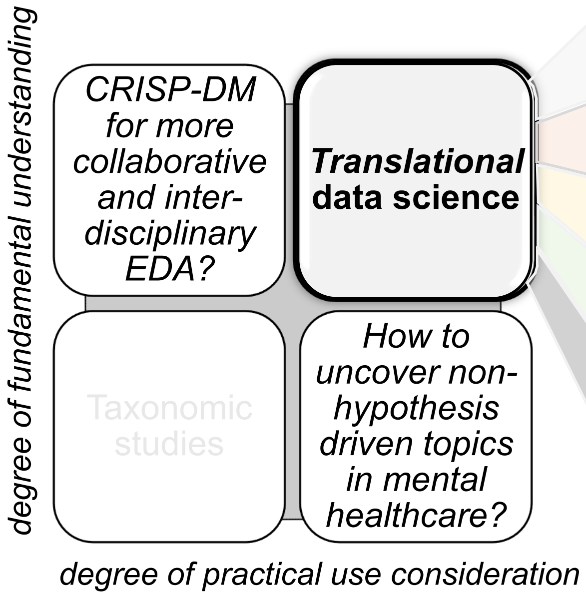




Example: Exploring Mental Healthcare

- *Finding:* A peak in aggression incidents occurs at day five, esp. in adult patients (dark)?

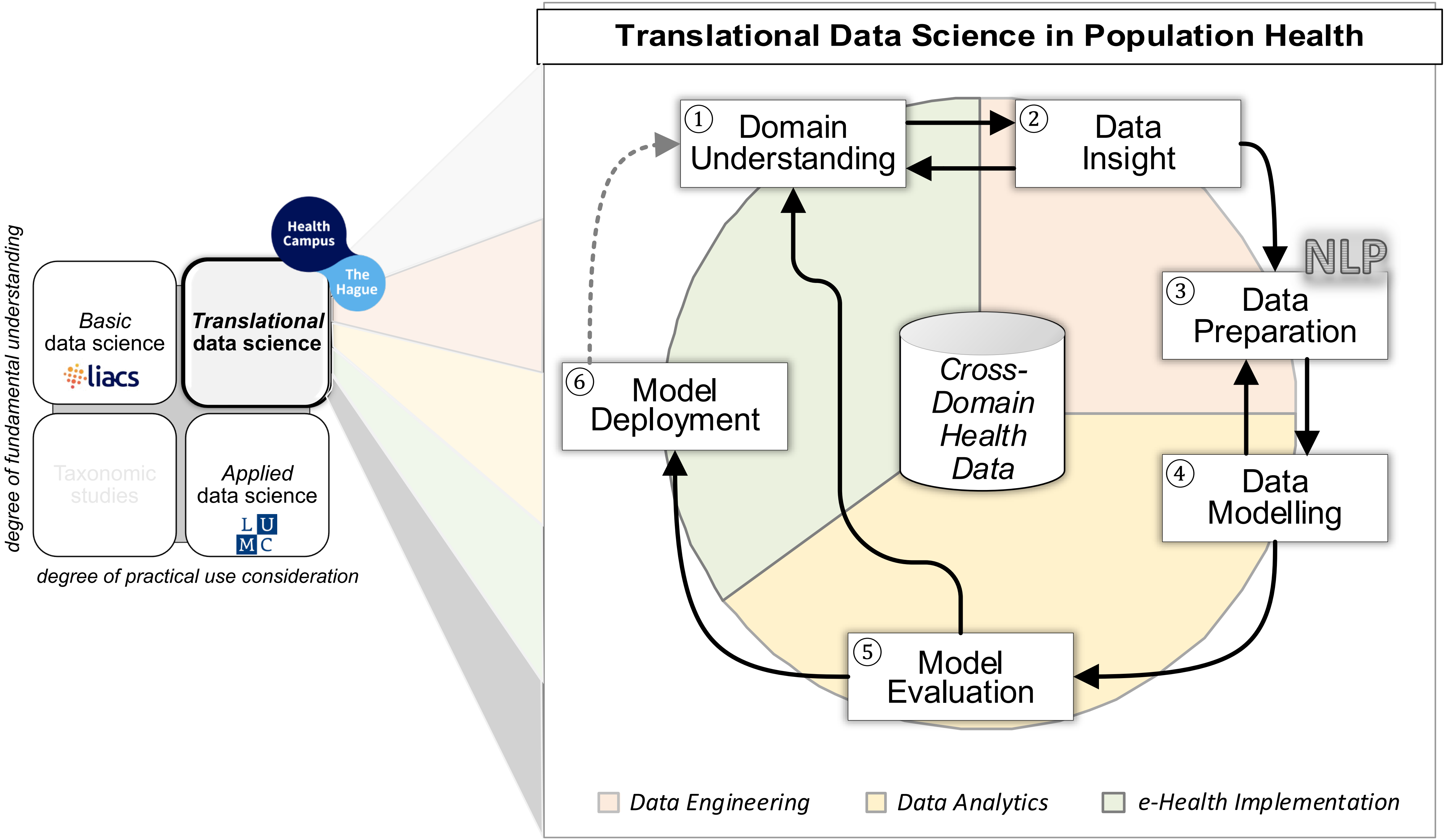




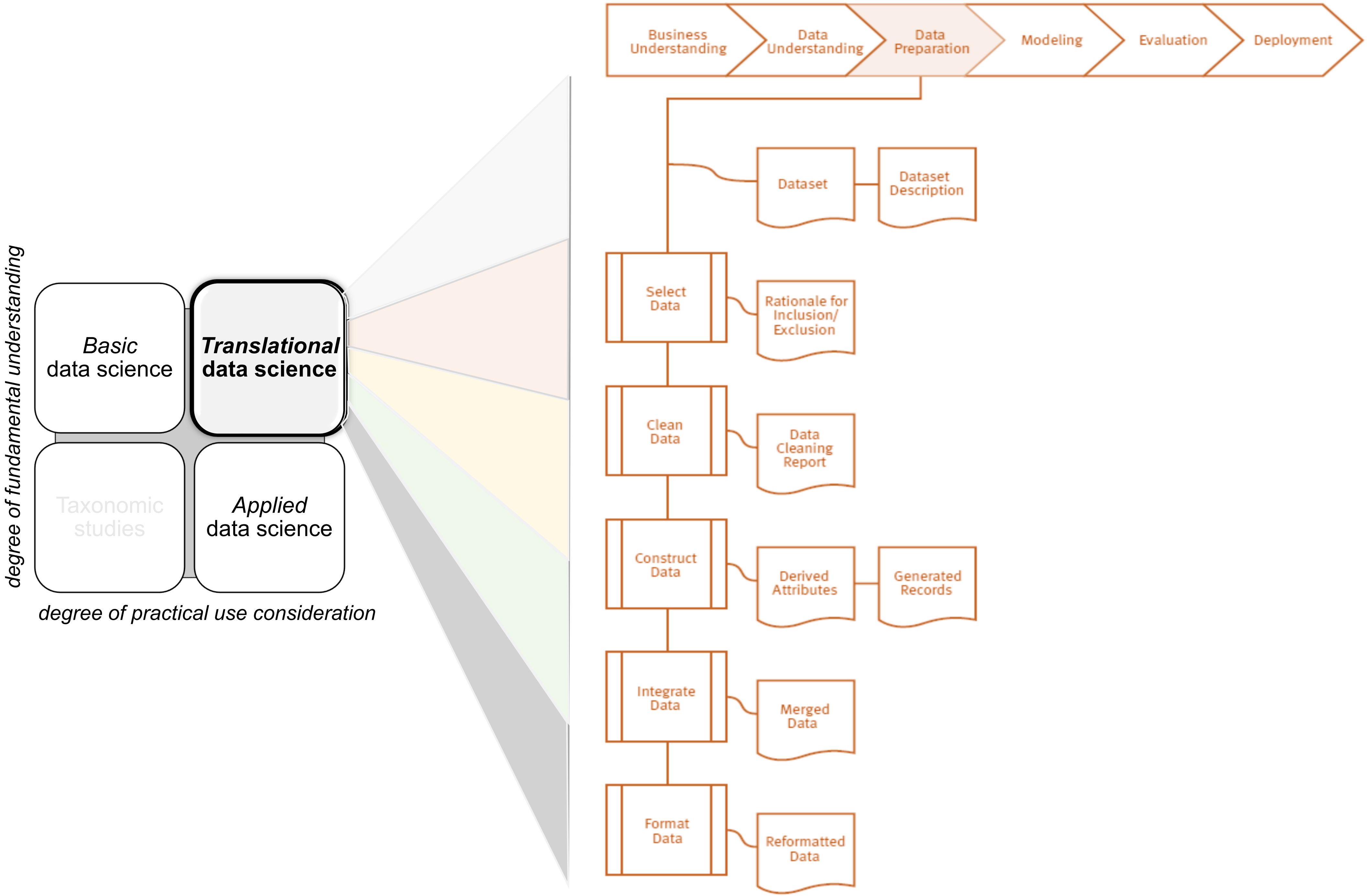
Example: Exploring Mental Healthcare

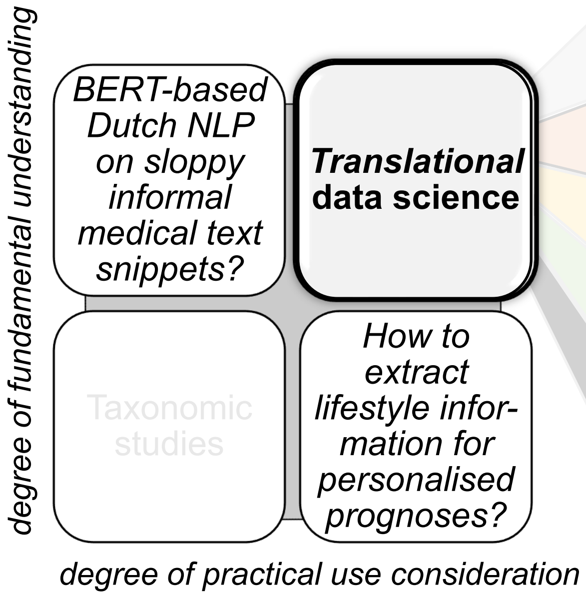
- *Finding:* 24/29 hypotheses are new due to CRISP-IDM!

Top-5 hypotheses	***Theme***
There exists a positive relation between season of admission and length of admission (longer admissions during winter)	Admission
A peak in aggression incidents occurs on the fifth day of admission	Aggression
There exists a relation between aggression incidents and wearing of medication effects in patients diagnosed with ADHD	Aggression, medication
There is an absence of a relation between amount of green space in patient environment and likelihood of developing a disorder	Context factors
There is a negative relation between economic status of living environment and length of admission	Admission, context factors



Translational Data Science in Population Health
degree of fundamental understanding
degree of practical use consideration
Basic data science
liacs
Translational data science
Health Campus
The Hague
Taxonomic studies
Applied data science
LUMC
1 Domain Understanding
2 Data Insight
3 Data Preparation
NLP
4 Data Modelling
5 Model Evaluation
6 Model Deployment
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation


degree of fundamental understanding
Basic data science
Translational data science
Taxonomic studies
Applied data science
degree of practical use consideration
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Dataset
Dataset Description
Select Data
Rationale for Inclusion/ Exclusion
Clean Data
Data Cleaning Report
Construct Data
Derived Attributes
Generated Records
Integrate Data
Merged Data
Format Data
Reformatted Data



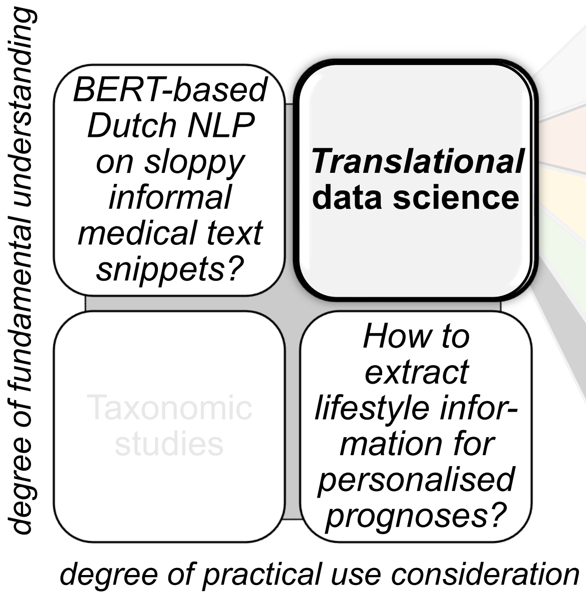
Example: Extracting Lifestyle Characteristics with NLP

Muizelaar,H., Haas,M., Putten,P. v.d., & Spruit,M. (submitted).

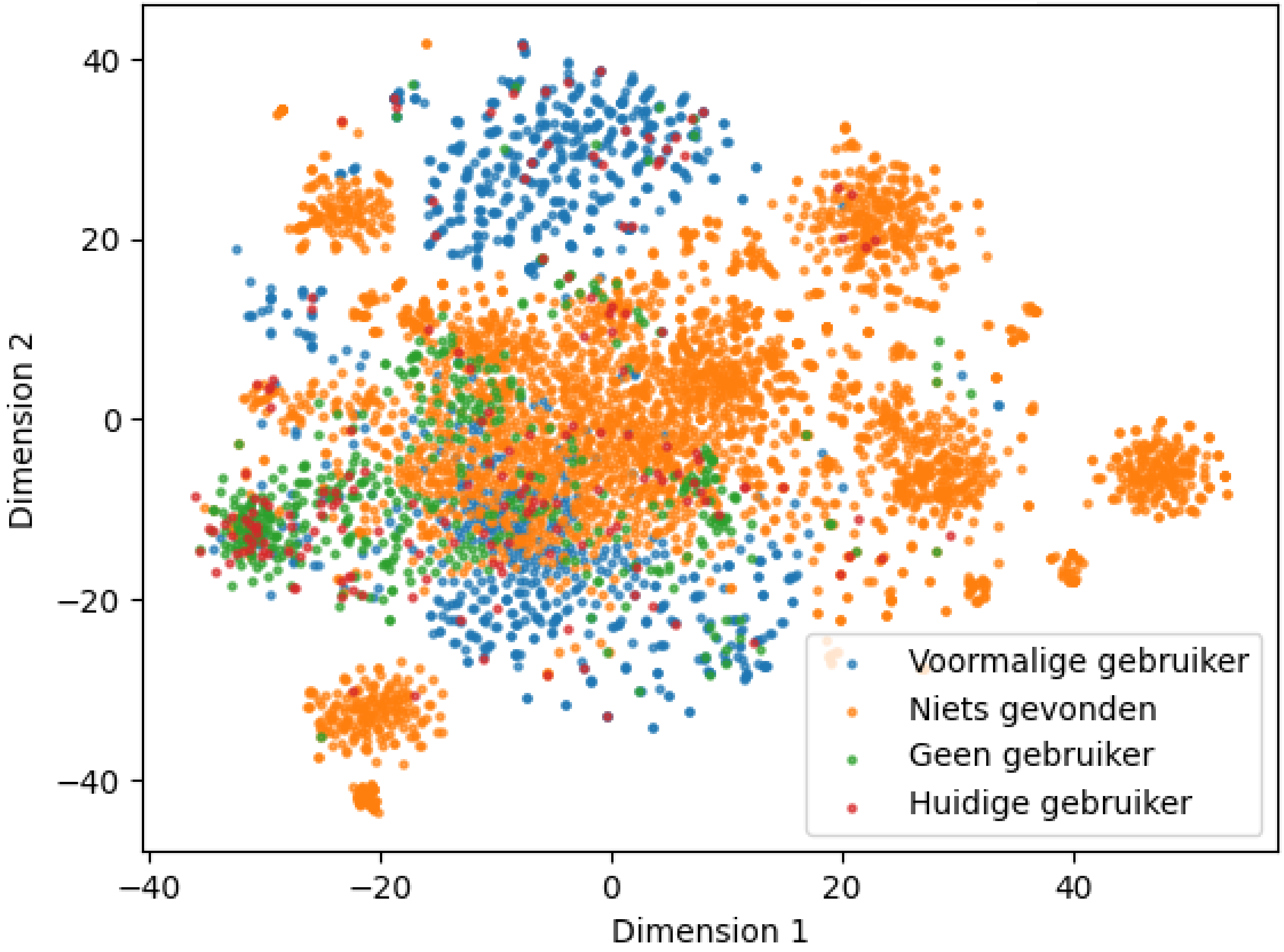
Example text data	Smoking	Alcohol	Drugs
Patient smokes, does not drink or use drugs	Current user	Non-user	Non-user
Patient used to smoke, drinks 1 beer a day	Former user	Current user	Unknown
Patient used to smoke, uses marihuana daily	Former user	Unknown	Current user

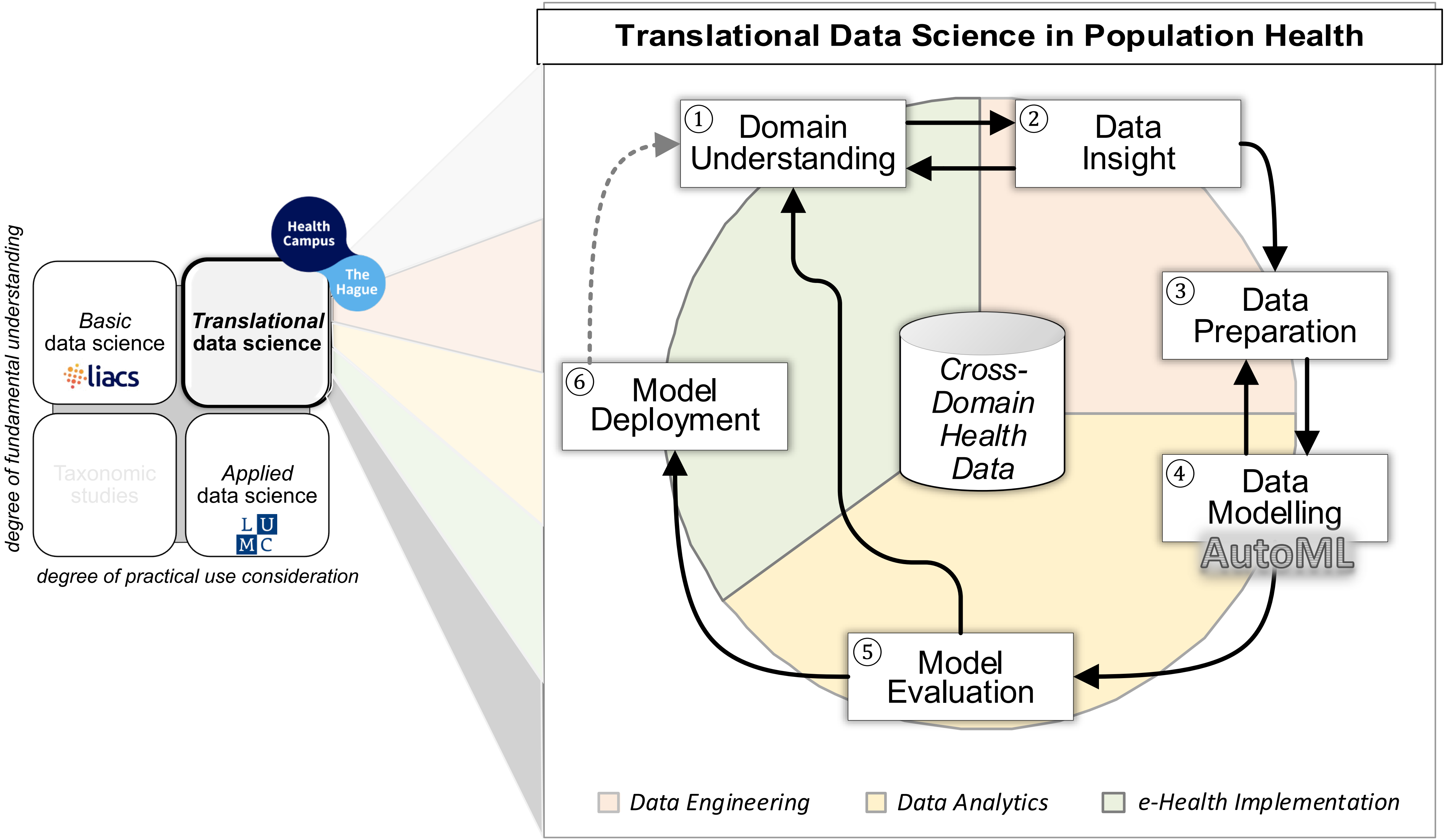
Model	***Smoking***	***Alcohol***	***Drugs***
String Matching	0.84	0.74	0.68
Machine Learning (SGD)	0.85	0.71	0.60
HAGALBERT	0.66	0.54	0.43
RobBERT-HAGA	0.87	0.71	0.63
belabBERT-HAGA	0.48	0.64	0.57
MedRoBERTa.nl-HAGA	0.93	0.79	0.77
BioBERT (translated)	0.91	0.72	0.52
ClinicalBERT (translated)	0.92	0.80	0.61

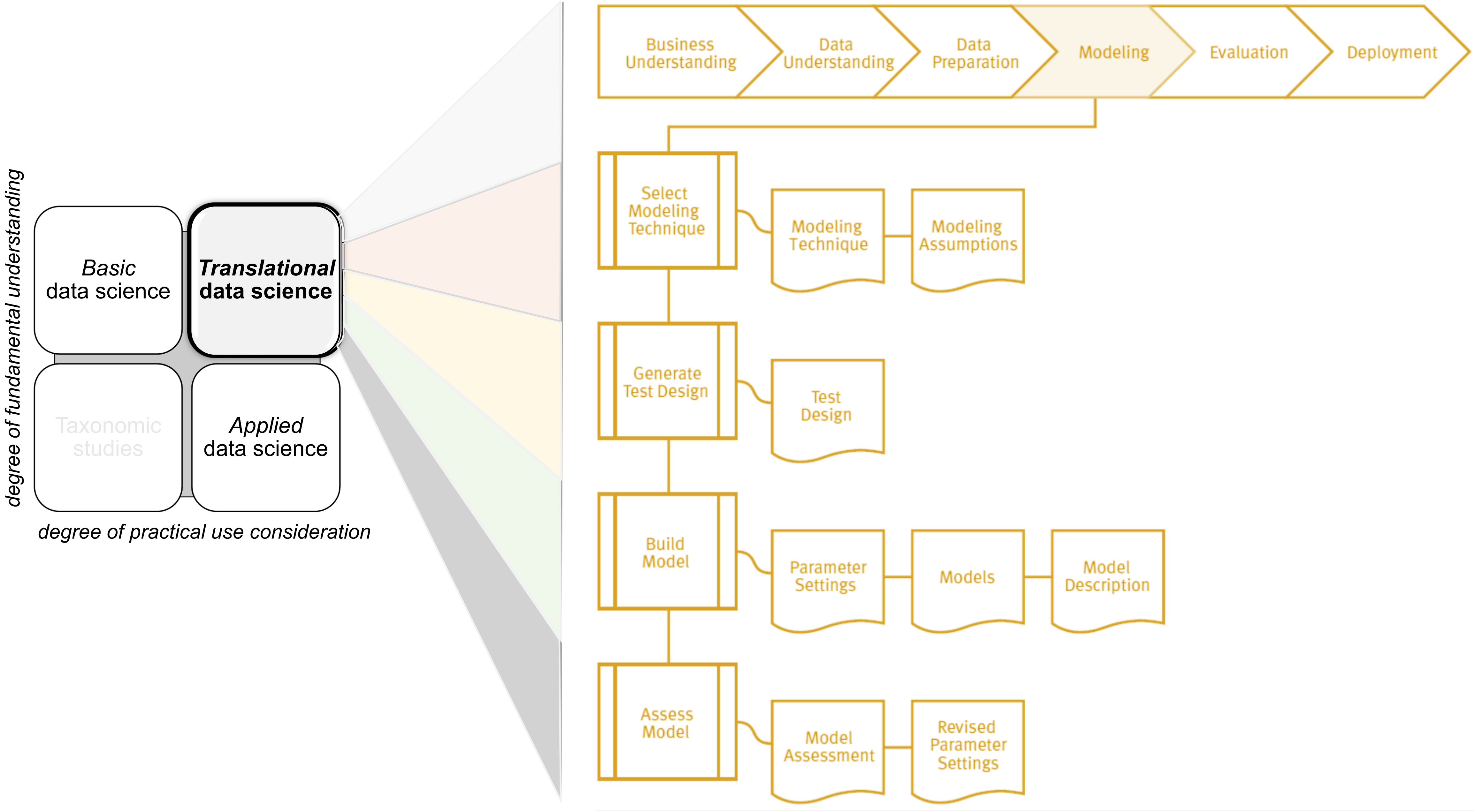
Example: Extracting Lifestyle Characteristics with NLP

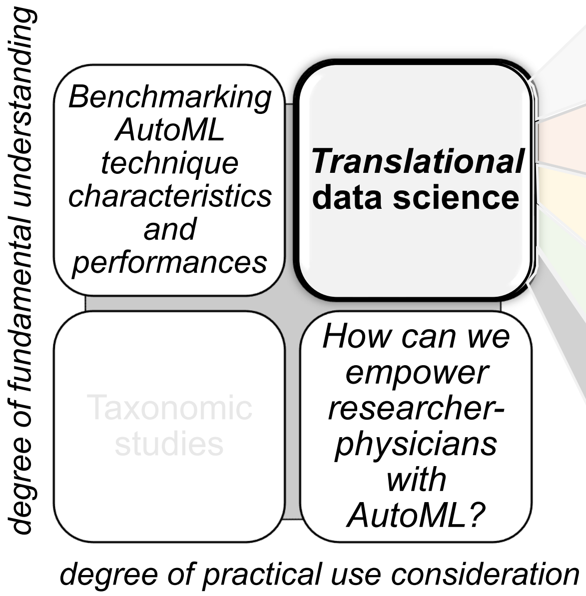


- tSNE visualisation of MedRoBERTa.nl-HAGA sentence embeddings



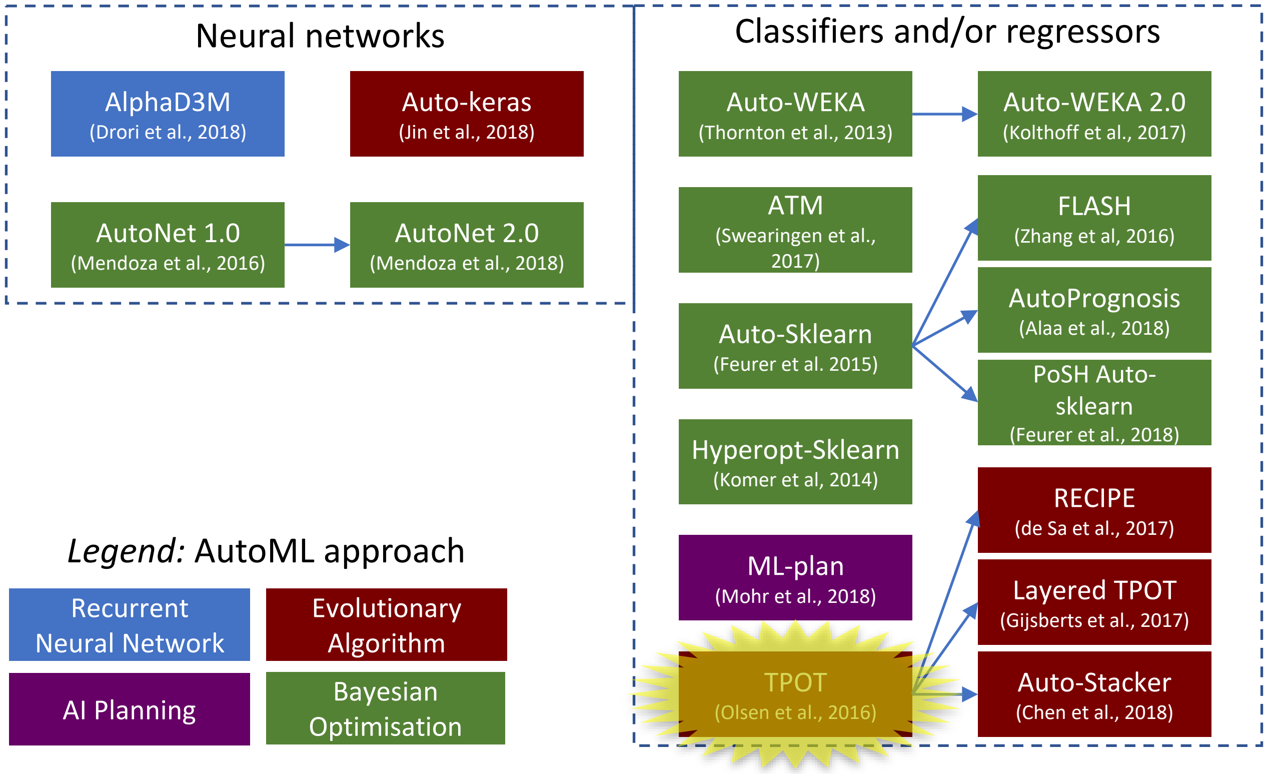

Translational Data Science in Population Health
degree of fundamental understanding
degree of practical use consideration
Basic data science
liacs
Translational data science
Health Campus The Hague
Taxonomic studies
Applied data science
LUMC
1 Domain Understanding
2 Data Insight
3 Data Preparation
4 Data Modelling
AutoML
5 Model Evaluation
6 Model Deployment
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation


degree of fundamental understanding
Basic
data science
Translational
data science
Taxonomic
studies
Applied
data science
degree of practical use consideration
Business
Understanding
Data
Understanding
Data
Preparation
Modeling
Evaluation
Deployment
Select
Modeling
Technique
Modeling
Technique
Modeling
Assumptions
Generate
Test Design
Test
Design
Build
Model
Parameter
Settings
Models
Model
Description
Assess
Model
Model
Assessment
Revised
Parameter
Settings

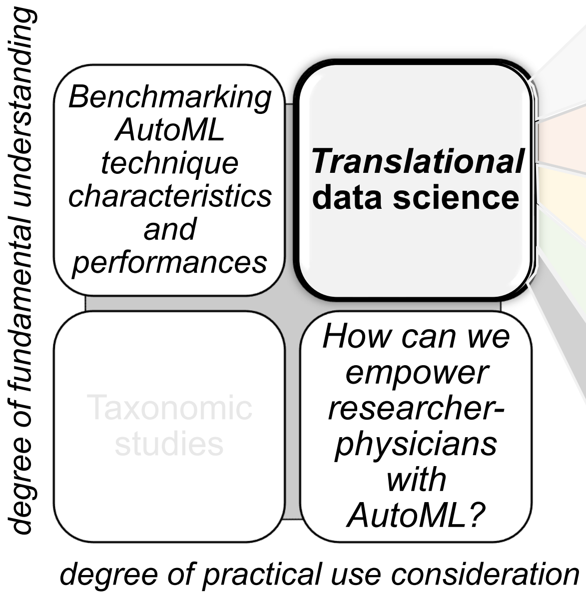


Example 1: Automated Machine Learning in Healthcare

- *Question:* "How can we support the knowledge discovery process of domain experts in healthcare using automated machine learning?"



Ooms,R., & Spruit,M. (2020). Self-Service Data Science in Healthcare with Automated Machine Learning. *Applied Sciences, 10*(9), Medical Artificial Intelligence, 2992. [online]

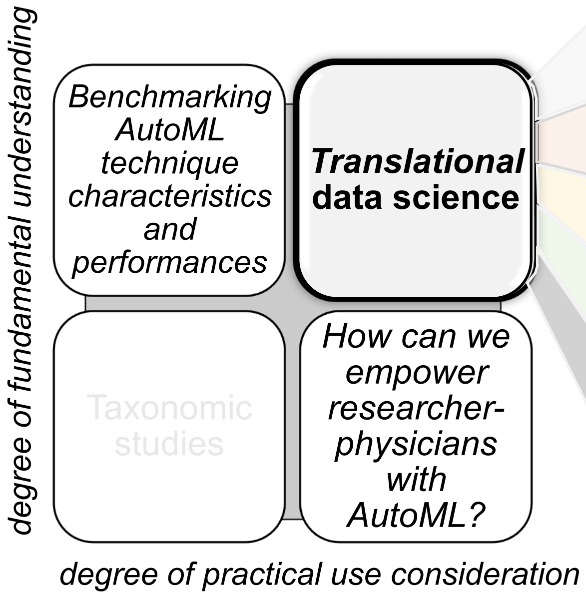


Example: Automated Machine Learning in Healthcare

- *Data:* All medical datasets suited for binary classification problems (4) in OpenML-CC18 open-source benchmark suite (Gijsbers *et al.*, 2019)

Dataset	***Data points***	***Missing data***	***Predictive features***	***Class variable***
Breast cancer	699	-	9	458/241
Diabetes	768	-	8	500/268
Indian Liver Patients	583	-	10	416/167
Sick	3772	6064	29	3541/231

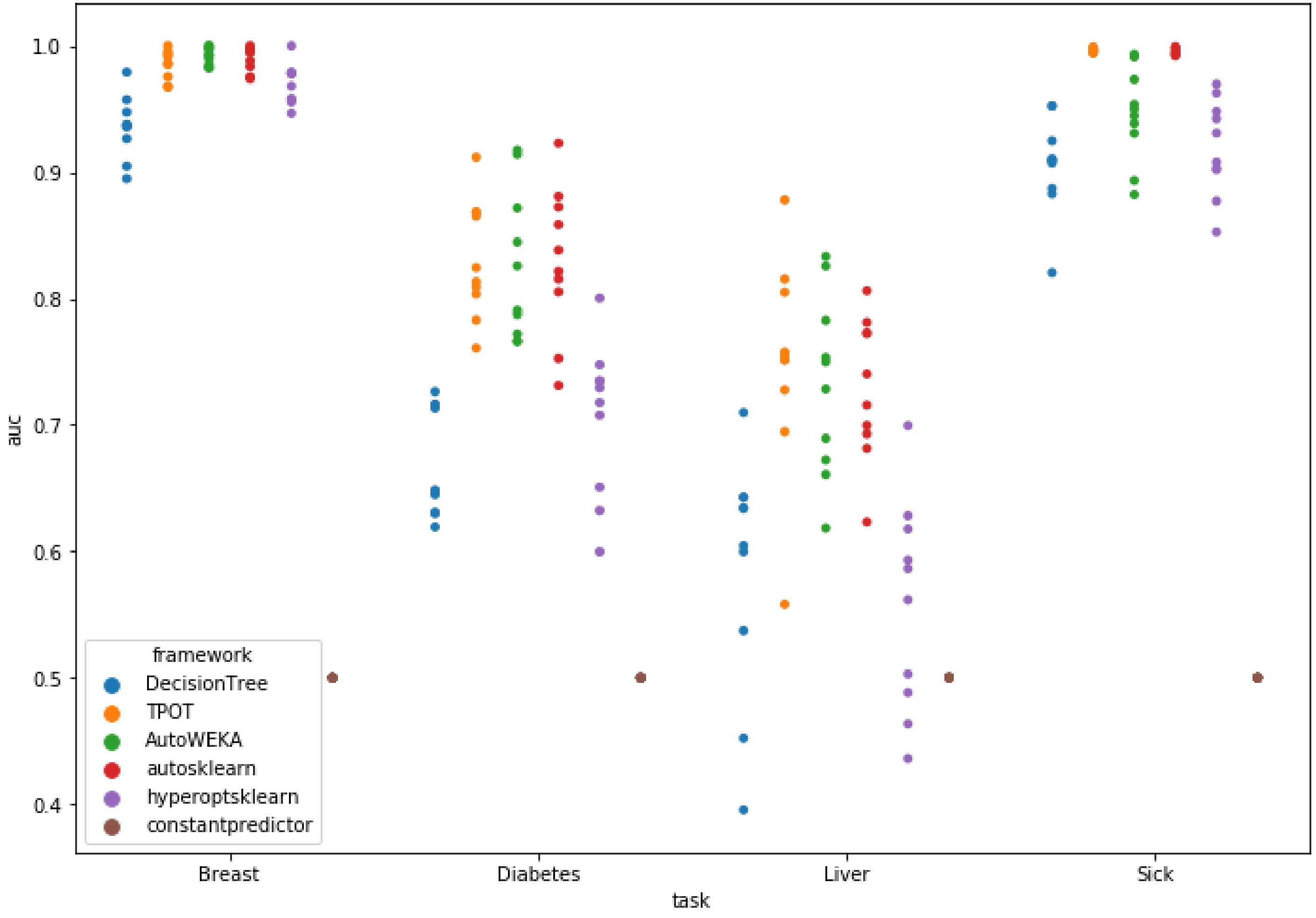
- All AutoML methods receive 1 hour in a 10-fold cross-validation set-up to create the best pipeline on these datasets



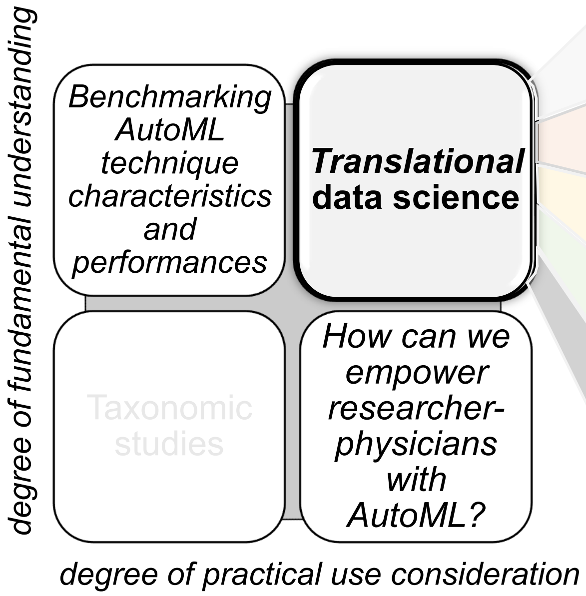
Example: Automated Machine Learning in Healthcare

- Decision tree and constant predictor as baseline
- Hyperopt performs worst; TPOT and Auto-Sklearn best

<1-hour budget>

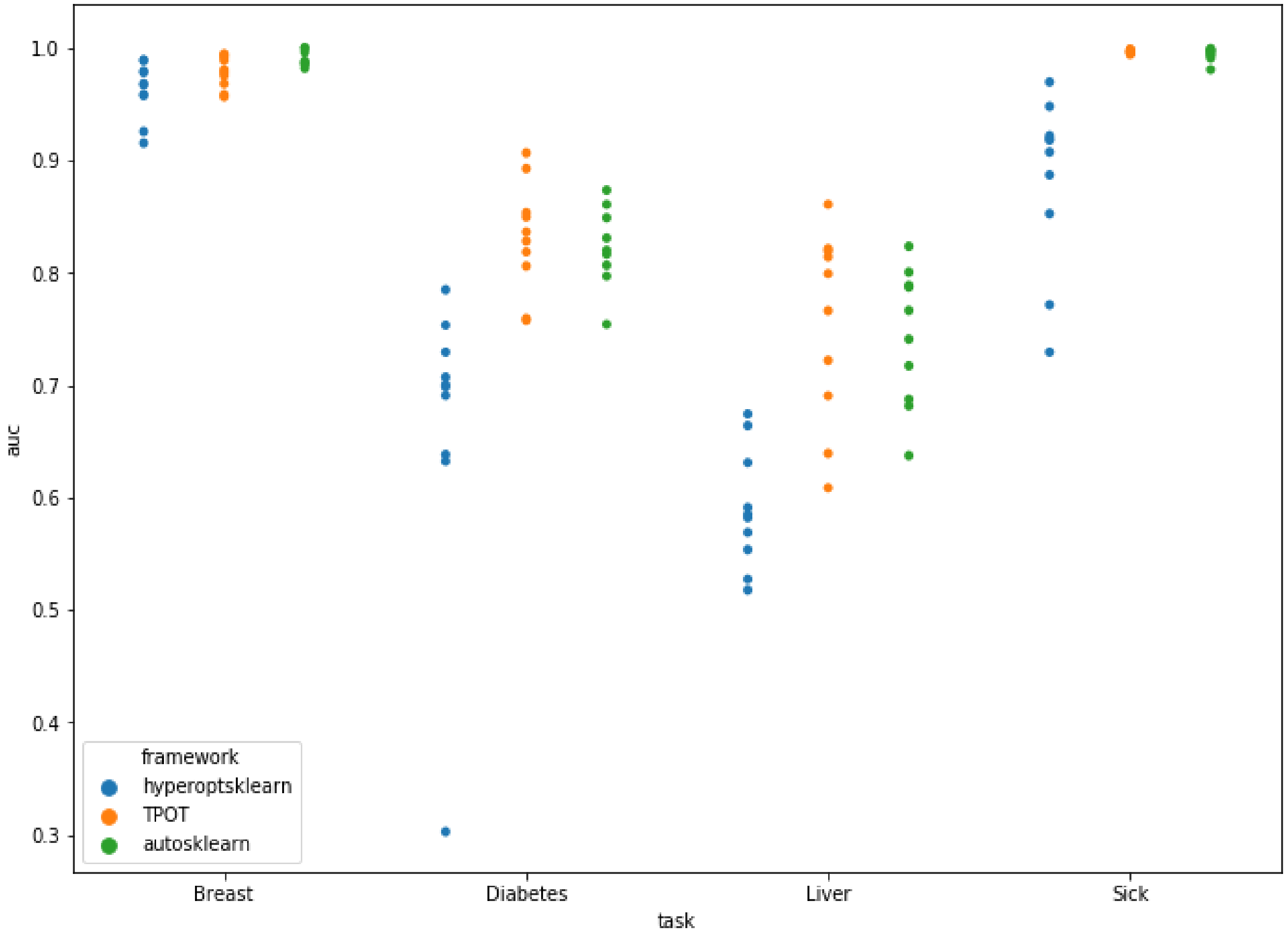


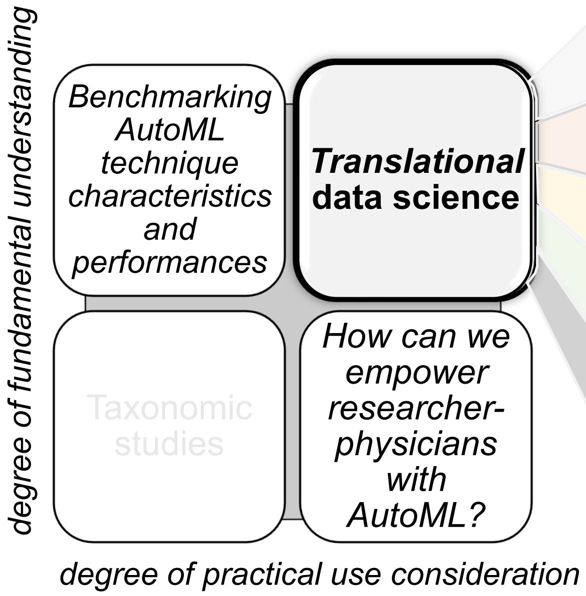
Example: Automated Machine Learning in Healthcare



- When benchmarking with 4 hour budget, again, TPOT and Auto-Sklearn perform best

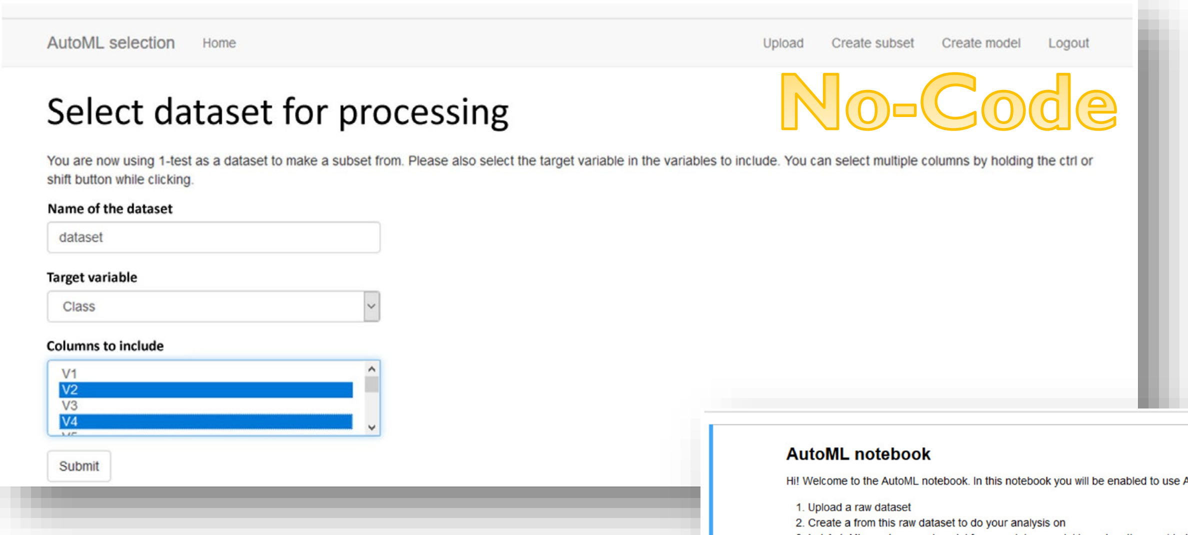
<4-hour budget>

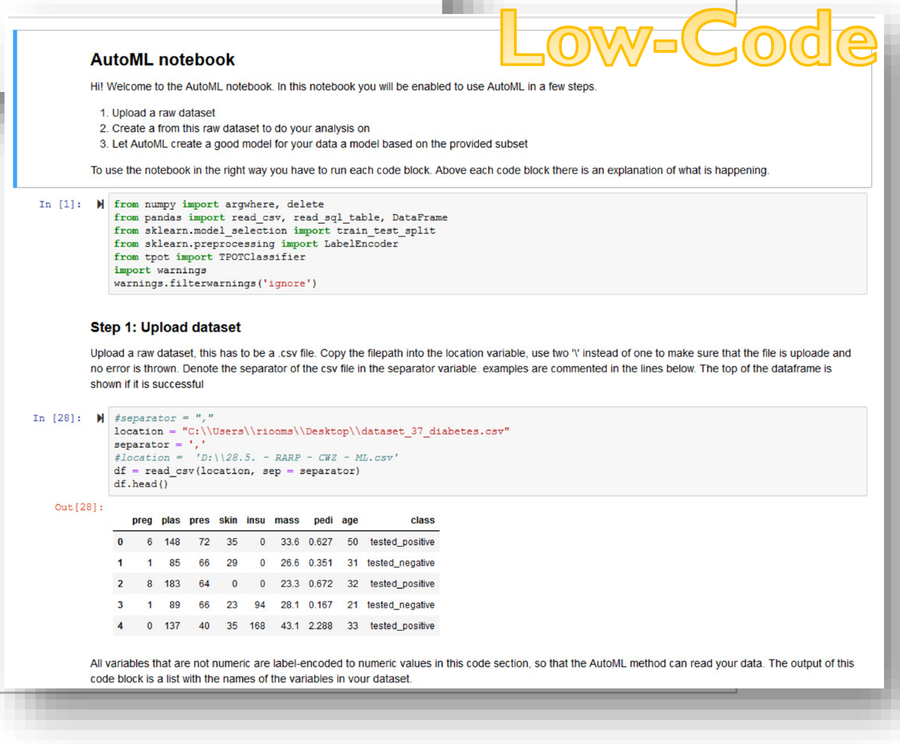




Example: Automated Machine Learning in Healthcare

- *Evaluation:* Webapp vs Notebook deployment

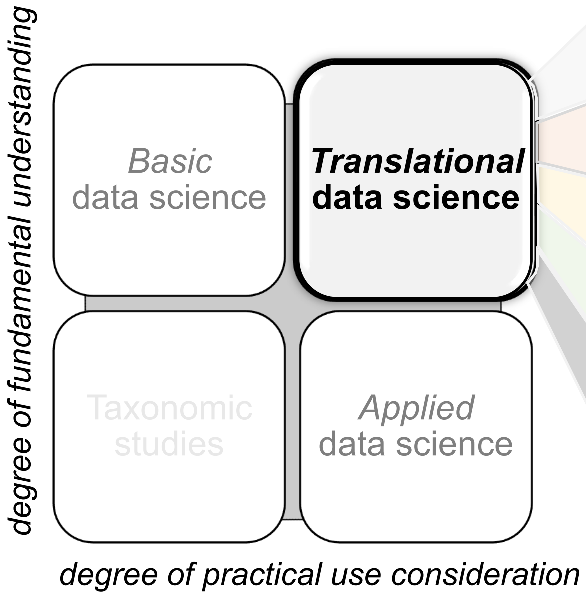




- *Artefact A: GUI /Flask app*
- *Artefact B: Interactive code/ Jupyter notebook*

- *Findings:* "best of both"…
- ***A** preferred for 'basic' ops*
- ***B** preferred for modeling*

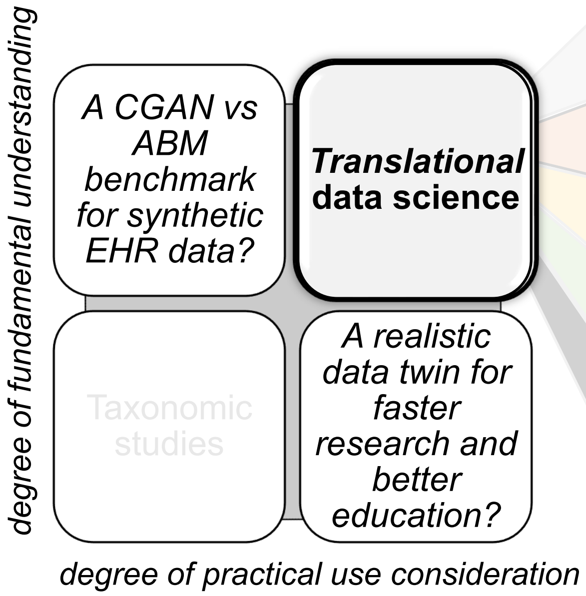
PS: PyCaret for Self-Service Data Science?



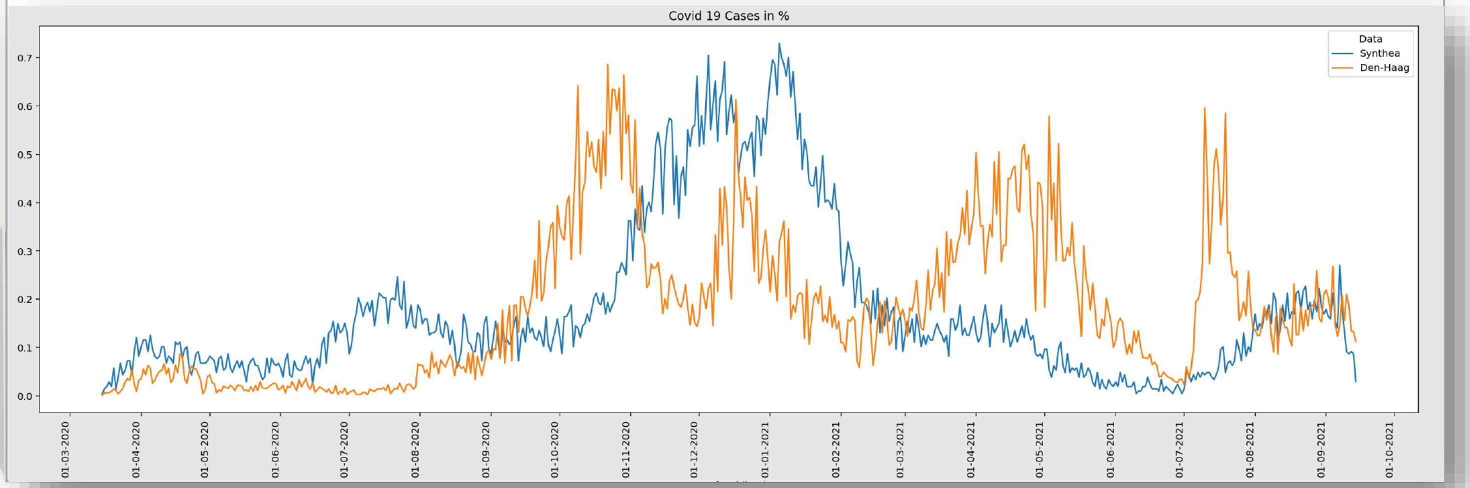

```
best = compare_models(sort='R2')
```

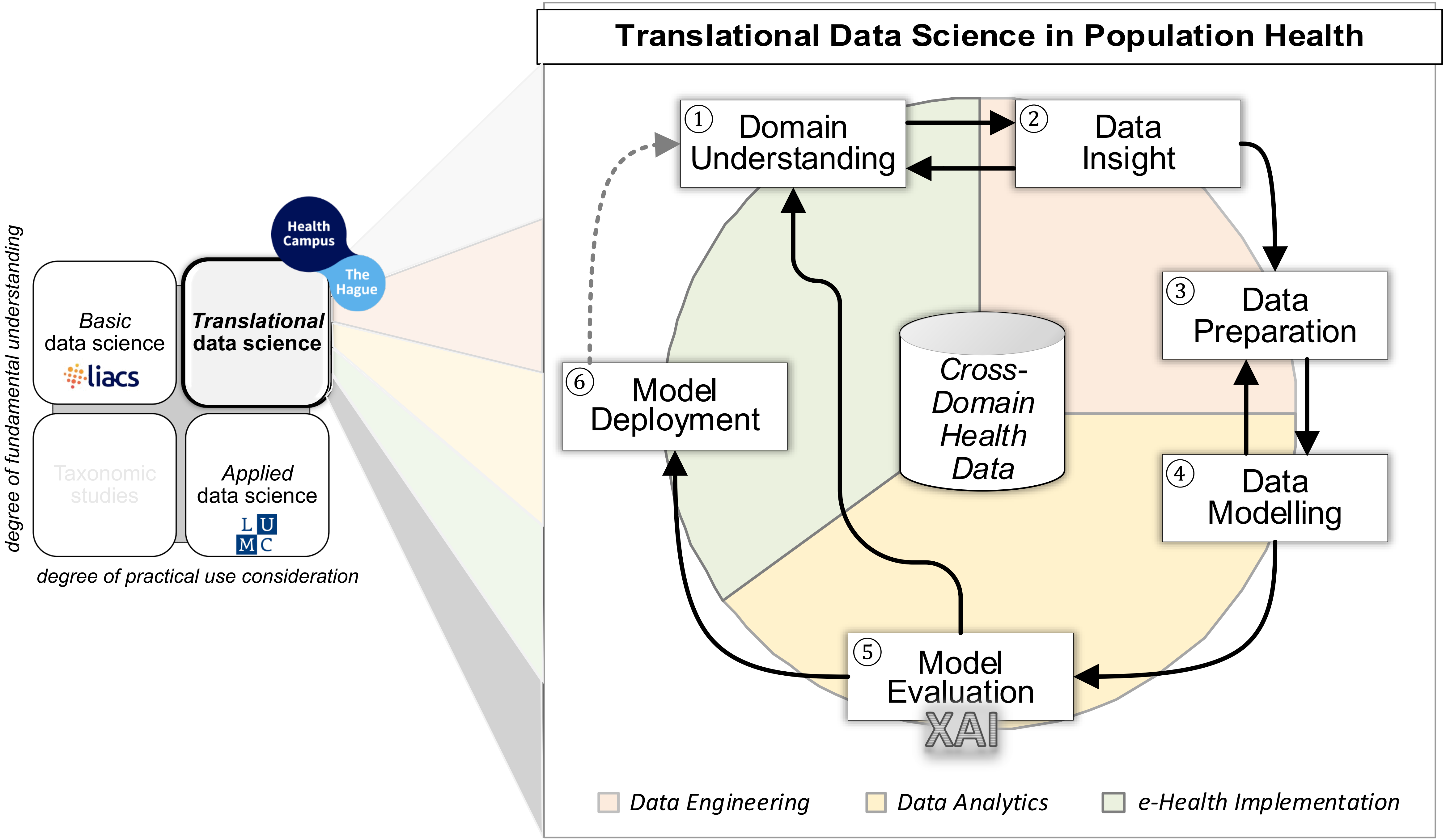
	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
rf	Random Forest Regressor	2342.1429	22959433.7357	4762.0337	0.8351	0.4097	0.2091	0.2360
gbr	Gradient Boosting Regressor	2275.4641	22815428.3119	4750.1089	0.8350	0.3858	0.1874	0.0740
ada	AdaBoost Regressor	3257.2171	23279230.3521	4807.0257	0.8339	0.4770	0.4264	0.0330
lightgbm	Light Gradient Boosting Machine	2491.6919	24030584.9610	4865.2118	0.8272	0.4153	0.2118	0.4180
et	Extra Trees Regressor	2364.4206	25237906.1980	4999.4284	0.8167	0.4283	0.2116	0.2060
catboost	CatBoost Regressor	2530.8745	25732627.8903	5042.4862	0.8134	0.4088	0.2020	1.0060
xgboost	Extreme Gradient Boosting	2931.6919	31946244.2000	5615.7612	0.7678	0.4551	0.2602	0.3410
dt	Decision Tree Regressor	3031.4152	42283353.7664	6468.0098	0.6936	0.5132	0.3181	0.0240
omp	Orthogonal Matching Pursuit	5645.3004	59119654.4986	7679.3606	0.5758	0.6831	0.6880	0.0190
ridge	Ridge Regression	4066.3599	61583179.6000	7714.4257	0.5583	0.4400	0.2707	0.0150
br	Bayesian Ridge	4072.9367	61948316.9816	7735.3075	0.5556	0.4399	0.2705	0.0210
lr	Linear Regression	4081.2541	62419186.8000	7762.5140	0.5521	0.4399	0.2702	0.6810
lar	Least Angle Regression	4081.2284	62418429.6714	7762.4664	0.5521	0.4399	0.2702	0.0210
knn	K Neighbors Regressor	4590.2544	70154126.3724	8271.6145	0.5162	0.5252	0.3131	0.0260
huber	Huber Regressor	4211.3096	80449305.4129	8799.2293	0.4214	0.4535	0.2076	0.0240
par	Passive Aggressive Regressor	5841.9890	94762106.9683	9607.9039	0.2863	0.6406	0.4609	0.0260
en	Elastic Net	8222.6689	160918301.6000	12608.0390	-0.1206	0.9079	0.9707	0.0200
lasso	Lasso Regression	8249.2145	161224220.8000	12619.7366	-0.1227	0.9107	0.9777	0.0150
llar	Lasso Least Angle Regression	8249.2145	161224210.7582	12619.7361	-0.1227	0.9107	0.9777	0.0190

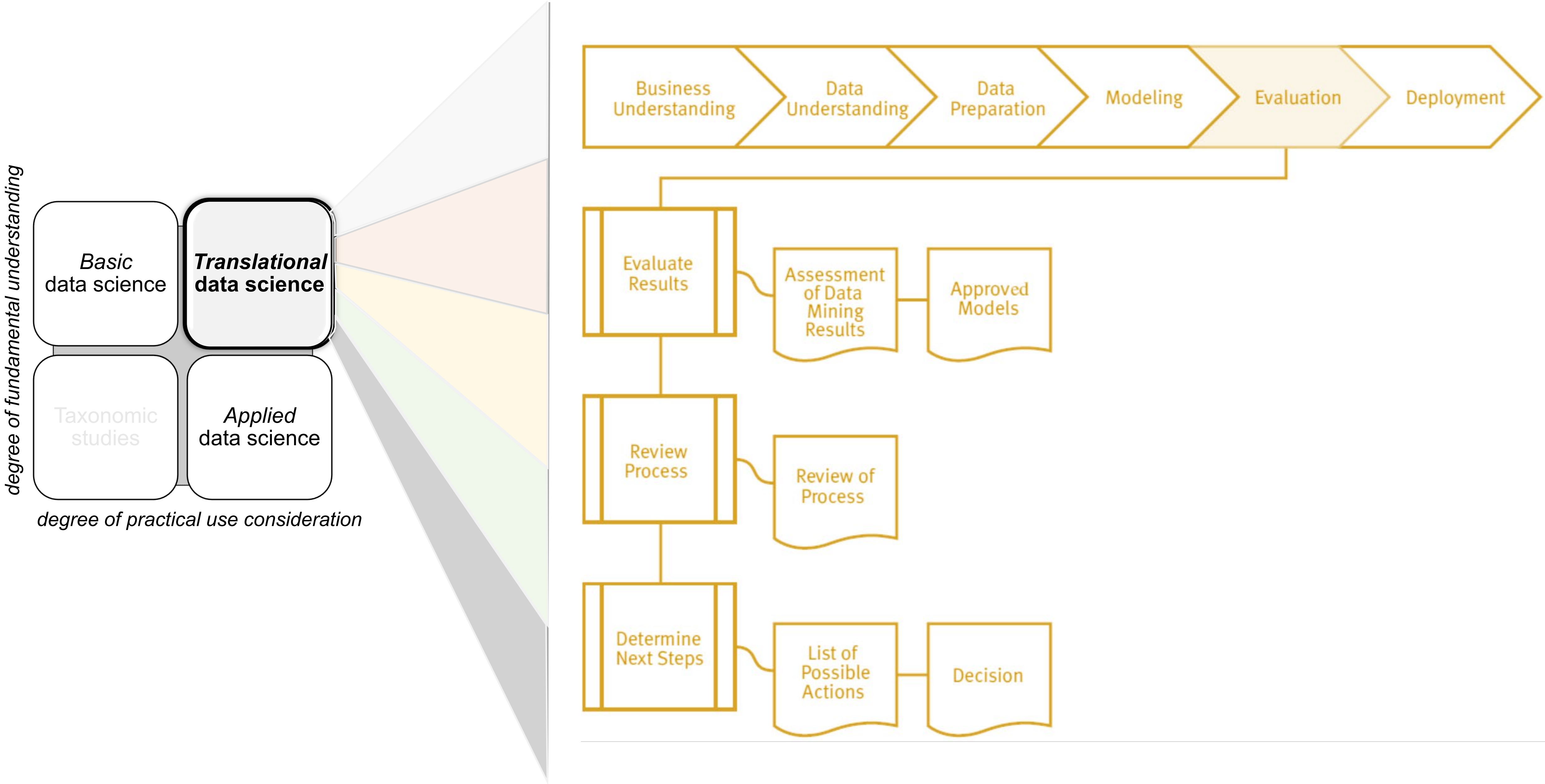
Example 2: Synthesising Virtual Patients & Population

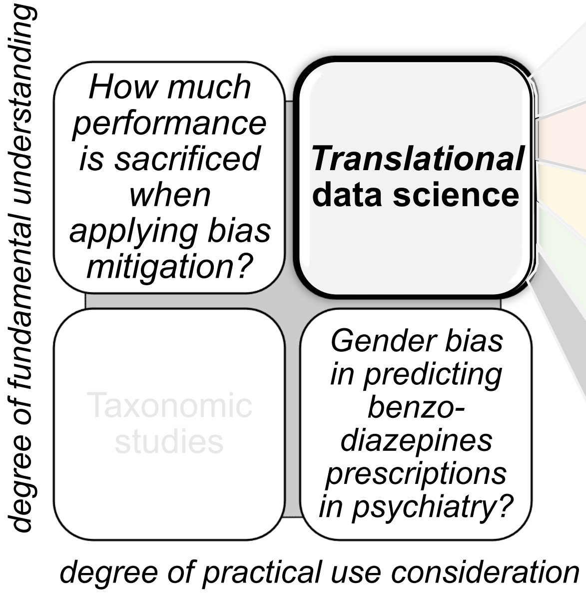


ABM approach *(Ammar Faiq)*	**CGAN approach** *(Jim Achterberg)*
• Synthea-based dataset An ELAN ‘digital twin’ is already being used in the PHM Fundamentals master course to let students analyse COVID outbreaks in The Hague region (see below) • Joint research with Statistics Netherlands (CBS) & Syntho • Workshop ‘Guidance Ethics’: many stakeholders, 50+ effects	• Thesis Evaluation Framework for synthetic EHR data (supporting heterogeneous types, time series, unpredictable quality) - tSNE extension - two-sample GoF test - evaluation metric for privacy risk through AiAs • Horizon Europe, NWO OSF



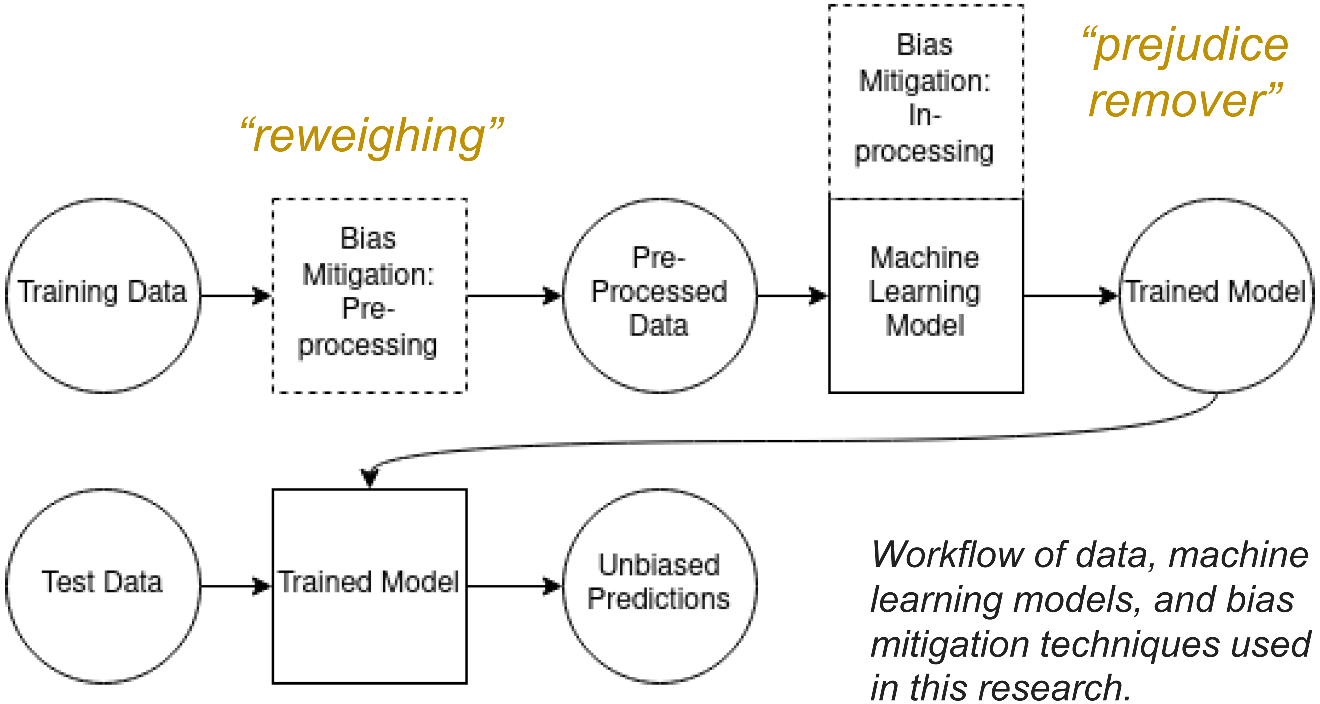

Translational Data Science in Population Health
degree of fundamental understanding
Basic
data science
liacs
Translational
data science
Health
Campus
The
Hague
Taxonomic
studies
Applied
data science
L U
M C
degree of practical use consideration
① Domain Understanding
② Data Insight
③ Data Preparation
④ Data Modelling
⑤ Model Evaluation
XAI
⑥ Model Deployment
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation


degree of fundamental understanding
Basic
data science
Translational
data science
Taxonomic
studies
Applied
data science
degree of practical use consideration
Business
Understanding
Data
Understanding
Data
Preparation
Modeling
Evaluation
Deployment
Evaluate
Results
Assessment
of Data
Mining
Results
Approved
Models
Review
Process
Review of
Process
Determine
Next Steps
List of
Possible
Actions
Decision



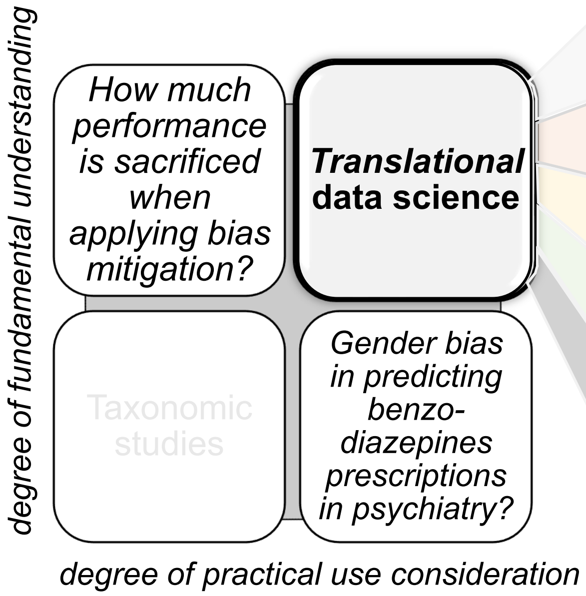
Example: Discovering Bias in Mental Health

- Effects and mitigation of gender fairness bias on a model trained to predict the future administration of benzodiazepines to psychiatric patients *(AI Fairness 360)*



Workflow of data, machine learning models, and bias mitigation techniques used in this research.

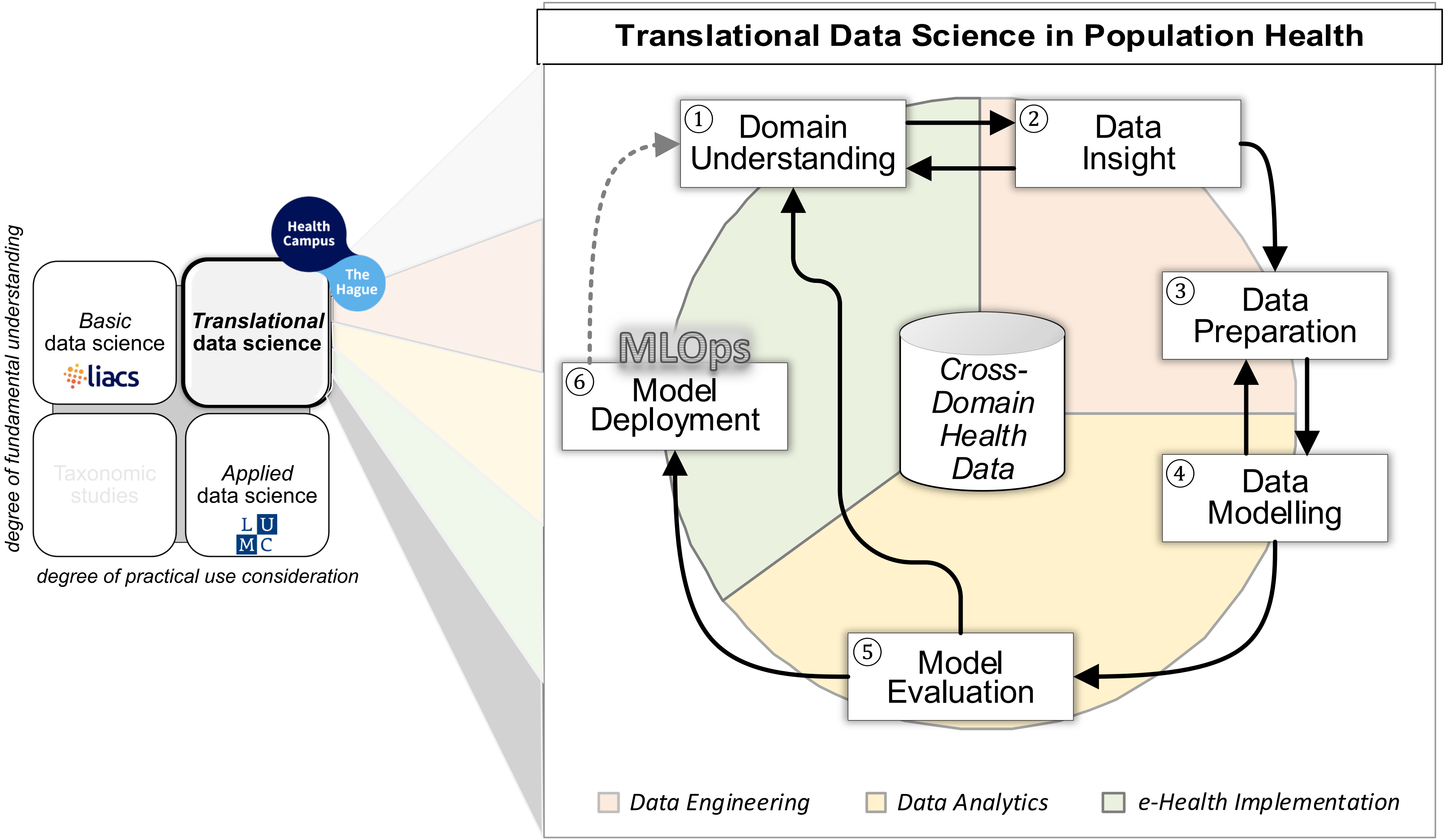
Mosteiro,P., Kuiper,J., Masthoff,J., Scheepers,F., & Spruit,M. (2022). Bias Discovery in Machine Learning Models for Mental Health. *Information, 13*(5), Advances in Explainable Artificial Intelligence, 237. [online]

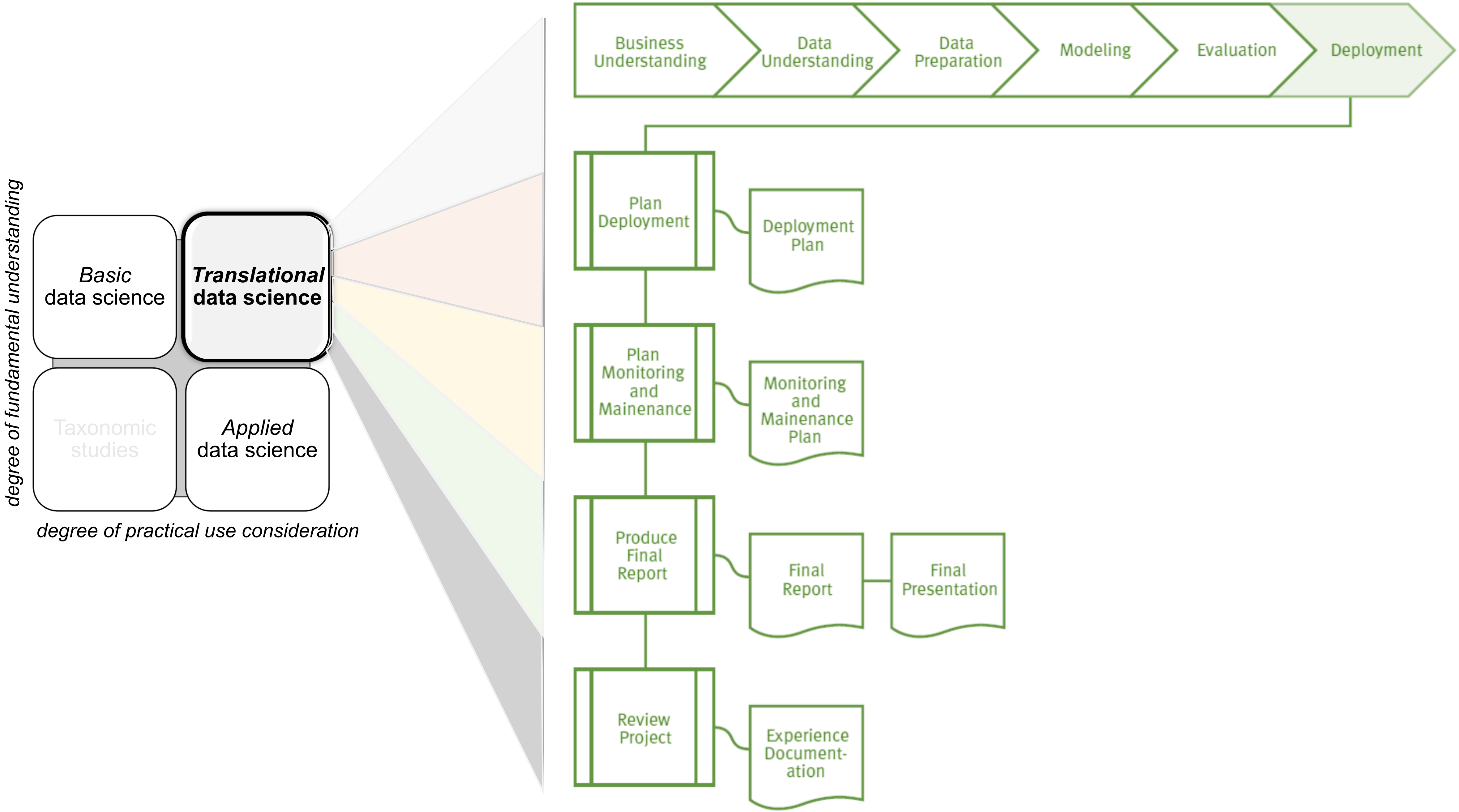


Example: Discovering Bias in Mental Health

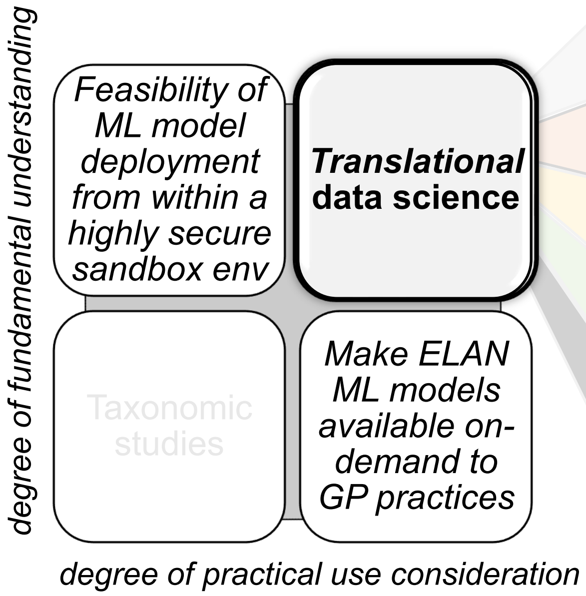
Model		Performance (comp. to baseline)	
Classifier	***Mitigation strategy***	***Δ Balanced accuracy***	***Δ F1***
Logistic regression	Predudice remover	**−0.040 ± 0.013**	−0.041 ± 0.025
Logistic regression	Re-weighing	−0.003 ± 0.013	−0.005 ± 0.013
Random forest	Re-weighing	0.003 ± 0.002	**0.005 ± 0.001**

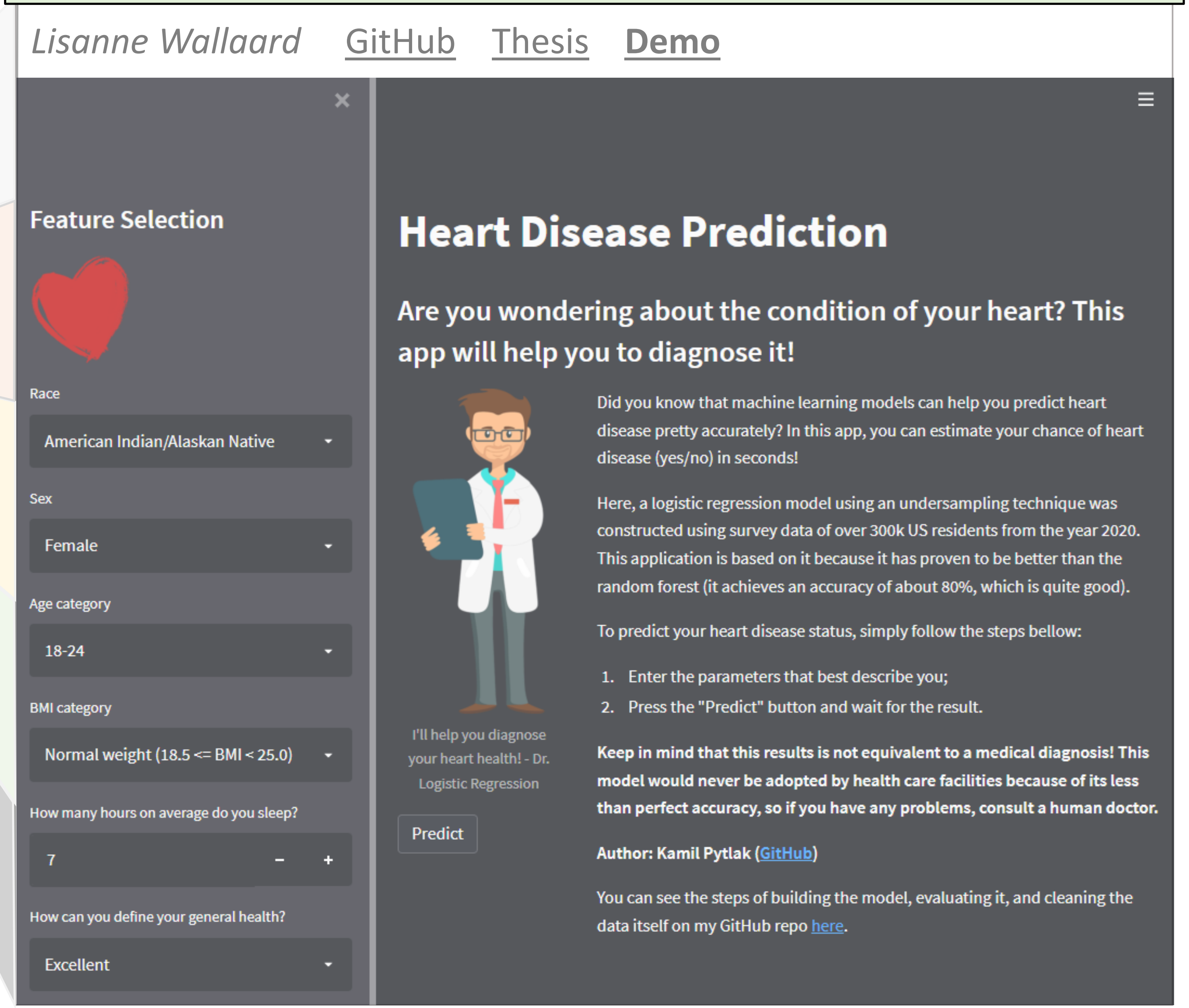
Model		Fairness metric (compared to baseline)			
Classifier	***Mitigation strategy***	***Δ Disparate impact***	***Δ Average odds difference***	***Δ Statistical parity difference***	***Δ Equal opportunity difference***
Logistic regression	Predudice remover	**0.092 ± 0.036**	0.038 ± 0.021	**0.050 ± 0.019**	0.018 ± 0.042
Logistic regression	Re-weighing	**0.075 ± 0.021**	**0.043 ± 0.017**	**0.043 ± 0.014**	0.042 ± 0.034
Random forest	Re-weighing	**0.034 ± 0.013**	**0.014 ± 0.006**	**0.013 ± 0.006**	0.014 ± 0.011


Translational Data Science in Population Health
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of fundamental understanding
degree of practical use consideration
① Domain Understanding
② Data Insight
③ Data Preparation
④ Data Modelling
⑤ Model Evaluation
⑥ Model Deployment
MLOps
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation


degree of fundamental understanding
Basic data science
Translational data science
Taxonomic studies
Applied data science
degree of practical use consideration
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Plan Deployment
Deployment Plan
Plan Monitoring and Mainenance
Monitoring and Mainenance Plan
Produce Final Report
Final Report
Final Presentation
Review Project
Experience Document-ation

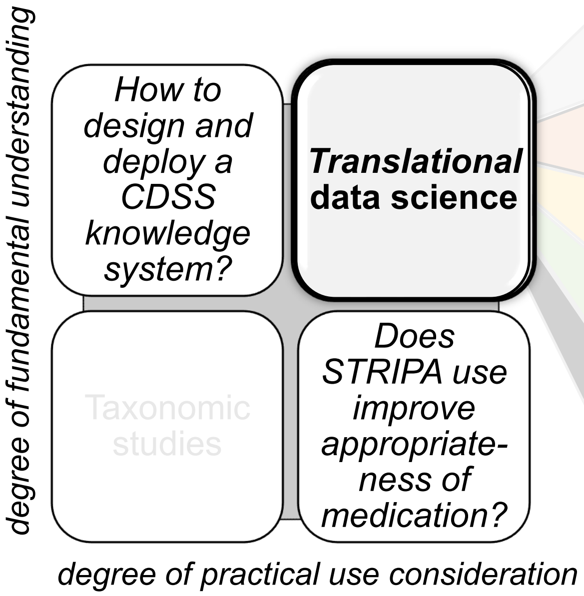
Example: Deploying Prediction Models for CDSS





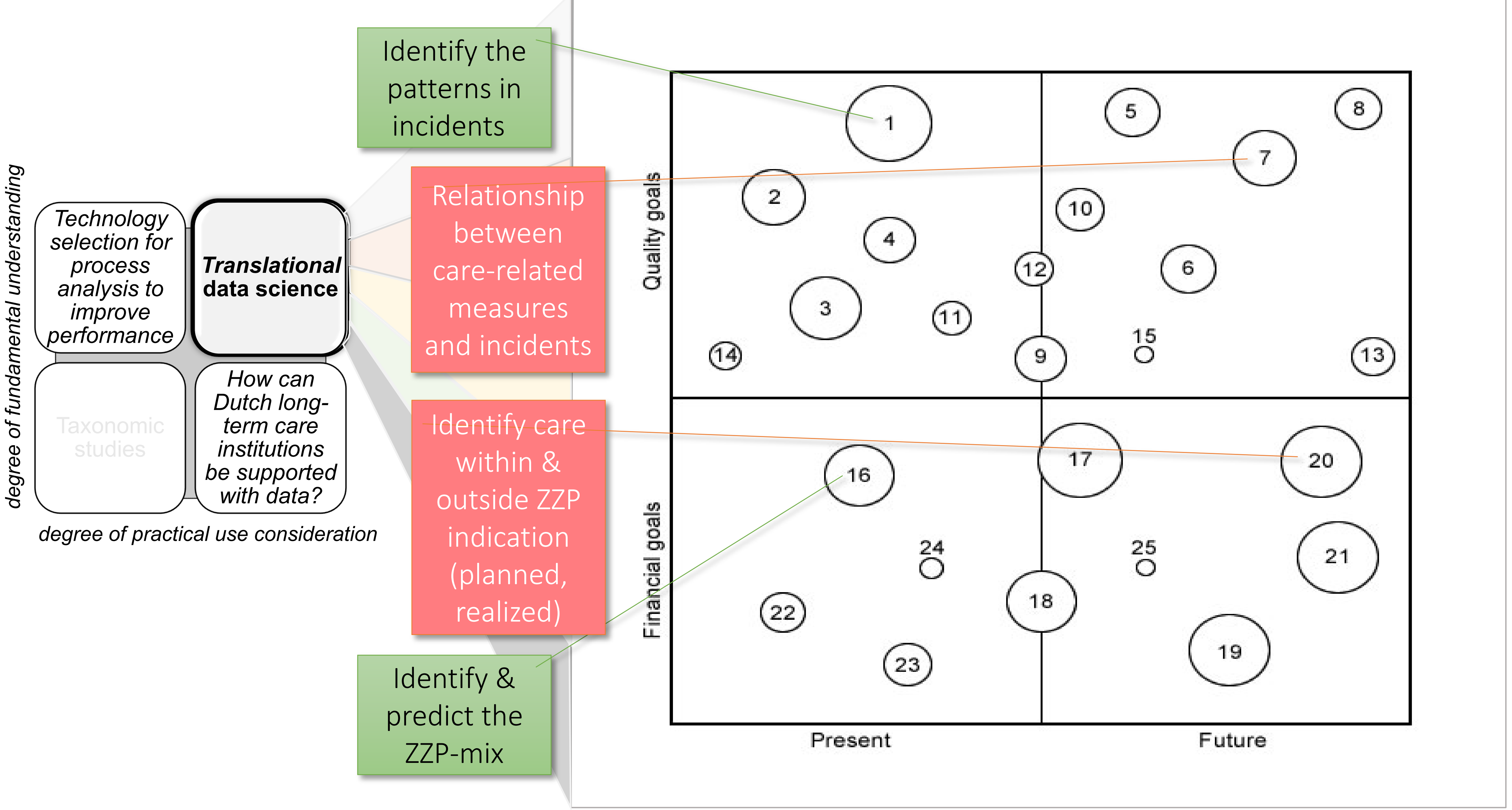
Example: STRIP Assistant for Medication Reviews

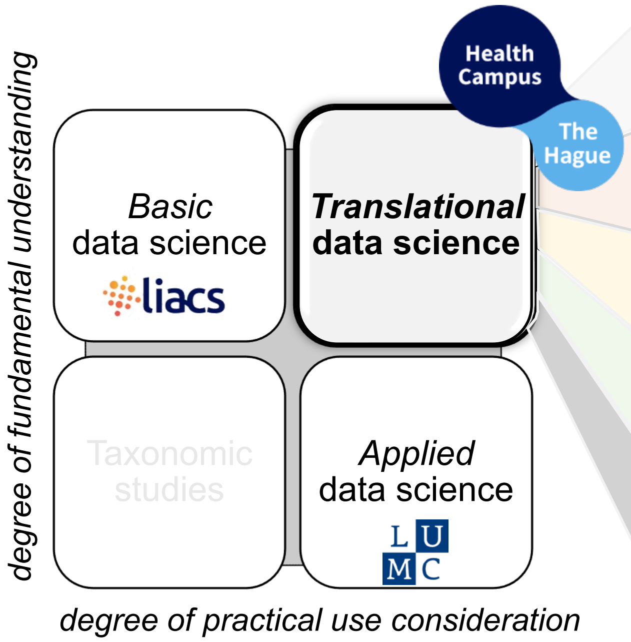
- A 10-years running research programme with big funding (OPERAM, OPTICA, STRIMP)
- Basically lost the fight with UU biz developers: *no product*

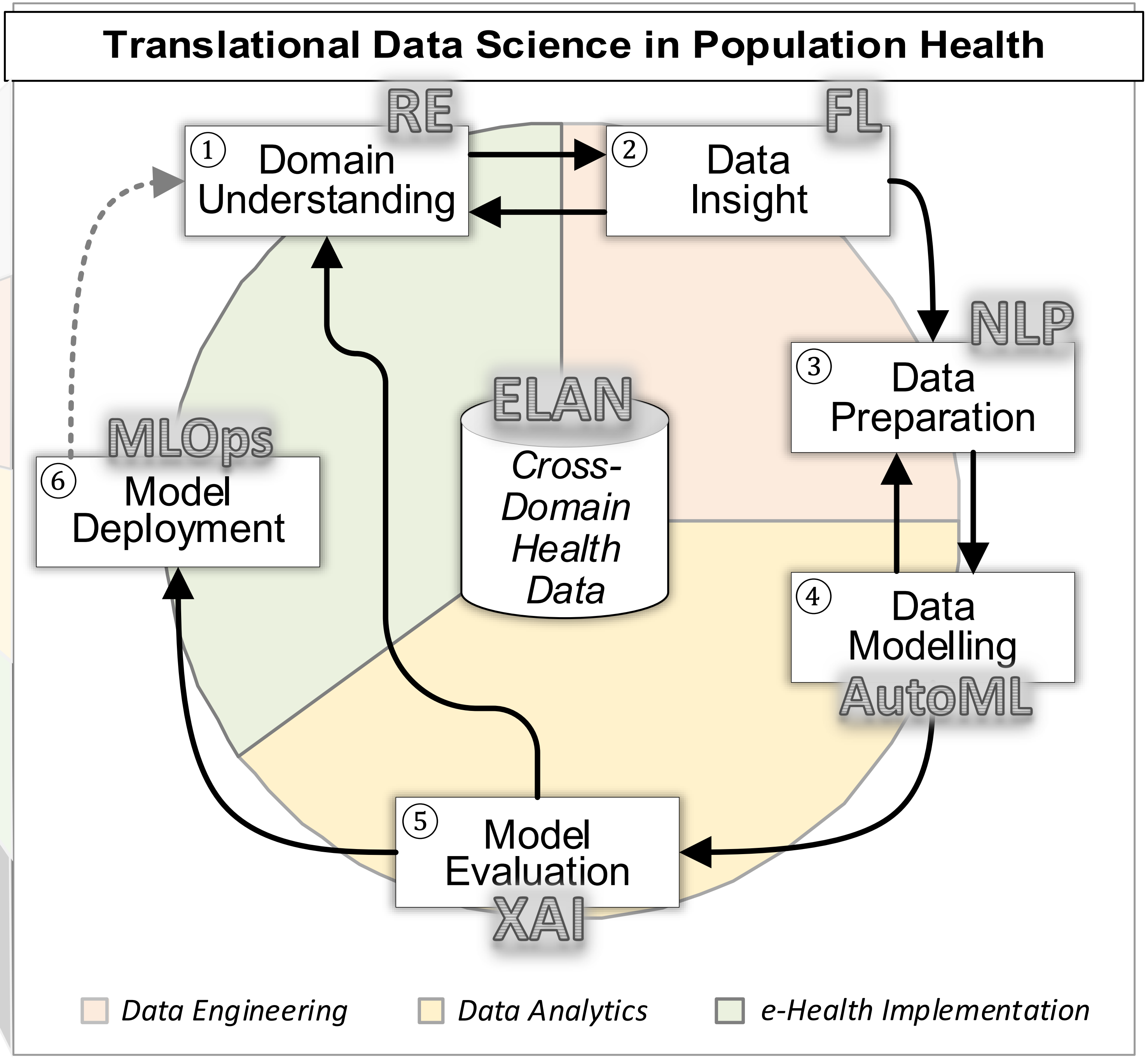


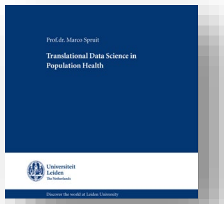
Jungo,K., ..., Spruit,M., …, Rodondi,N., Streit,S. (2023). Optimising prescribing in older adults with multimorbidity and polypharmacy in primary care (OPTICA): cluster randomised clinical trial. BMJ, 381. [online]

Example: Understanding Long-term Care









Spruit, Marco. (2022). *Translational Data Science in Population Health* (p. 20). Inaugural lecture. Leiden University. https://doi.org/10.5281/zenodo.7665858

Health Campus Den Haag

Thanks

Translational Data Science LAB : members *d.d. 29-10-2023*

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

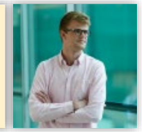
Friso van Dijk
PhD student (UU)
Privacy Governance

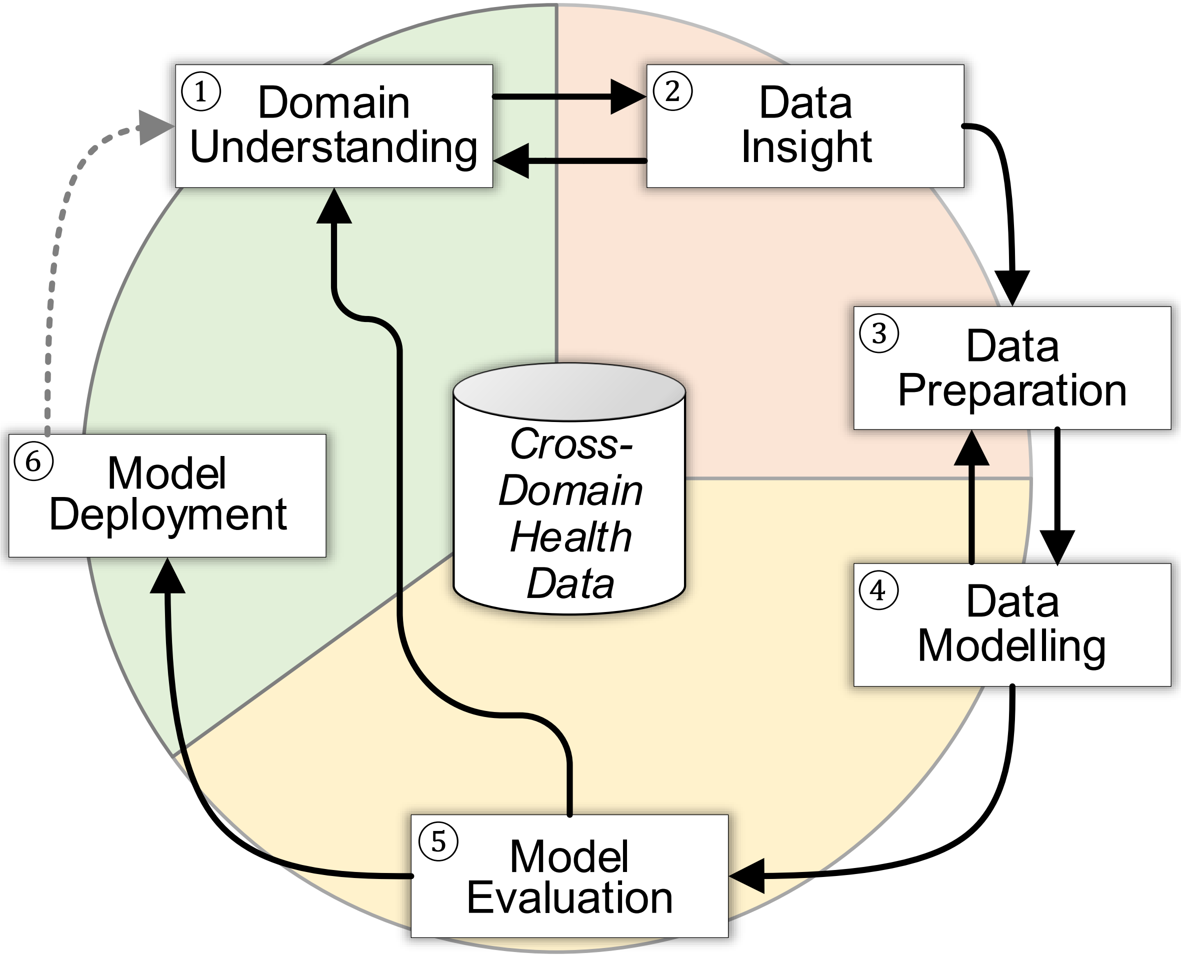
Els Roorda
PhD student LUMC
Population Health Information Systems

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

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

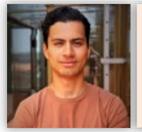
Max van Haastrecht
PhD student (LIACS)
Cybersecurity Risk Modelling & Validation



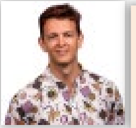


Ian Shen
PhD student (LIACS)
Healthcare and Open Science Engineering

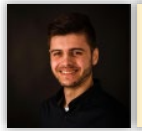
Samar Samir
PhD student (LIACS)
Federated NLP in Mental health



Bram van Dijk
PhD student (LIACS)
Mindreading with NLP



Emil Rijcken
PhD student (TUE)
Topic Modelling in NLP



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

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

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

Data Engineering *Data Analytics* *e-Health Implementation*

http://tdslab.nl

Health Campus The Hague

***Translational* data science**

tdslab.nl

Translational Data Science LAB : former members *d.d. 9-10-2023*

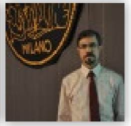
Dr. Willem Bekkers
former PhD student (UU)
Situational Process Improvement in Software Product Management

Dr. Michiel Meulendijk
former PhD & postdoc (UU)
Optimizing Medication Reviews through Decision Support

Dr. Armel Lefebvre
former PhD student (UU)
Research Data Management for Open Science

Dr. Bilge Yigit Ozkan
former PhD student (UU)
Cybersecurity Maturity Assessment & Standardisation

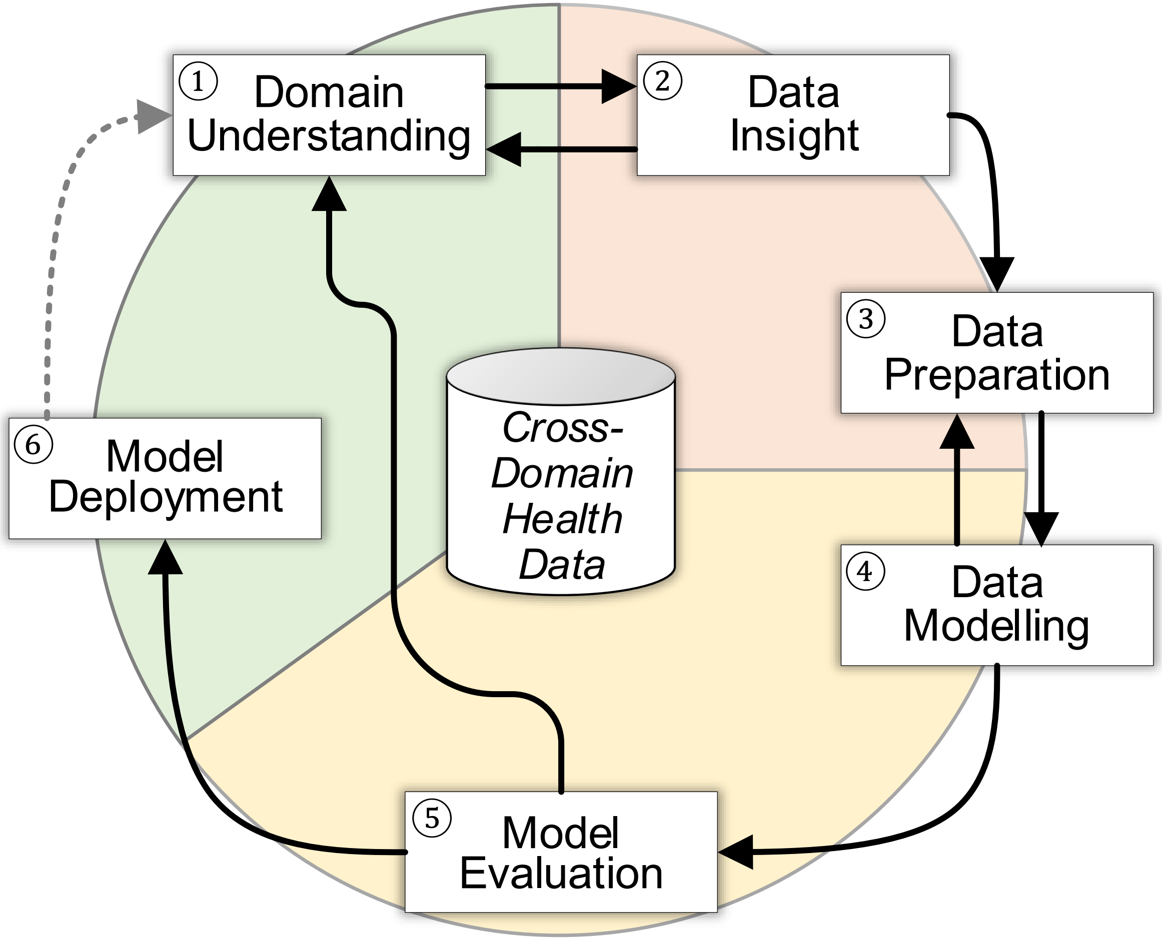
Dr. Alireza Shojaifar
former PhD student (UU)
Volitional Cybersecurity



Jan van Dijk
former PhD student (UU)
Heterogeneous Data Space Engineering

Dr. Lamia Elloumi
former postdoc (UU)
Medication Review Decision Support RCT Implementation

Edwin Brinkhuis
Software Engineer (Trumpett)
Medication Review Decision Support RCT Implementation

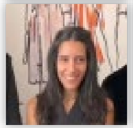


Data Engineering *Data Analytics* *e-Health Implementation*

Dr. Kalliopi Zervanou
former assistant professor (LUMC/LIACS)
Health Data Science & NLP

Dr. Shaheen Syed
former PhD student (UU)
Topic Discovery from Textual Data

Dr. Noha Seddik Tawfik
former PhD student (UU)
Text Mining for Precision Medicine



Dr. Ingy Sarhan
former PhD student (UU)
Open Information Extraction for Knowledge Representation

Chaïm van Toledo
former PhD student (UU)
Dutch NLP for HR

Dr. Vincent Menger
former PhD student (UU)
Knowledge Discovery in Clinical Psychiatry

Dr. Wienand Omta
former PhD student (UU)
Knowledge Discovery in High Content Screening

Dr. Pablo Mosteiro Romero
former postdoc (UU)
Dutch NLP for Mental Healthcare

https://tdslab.nl

Health Campus The Hague

Translational **data science**

tdslab.nl