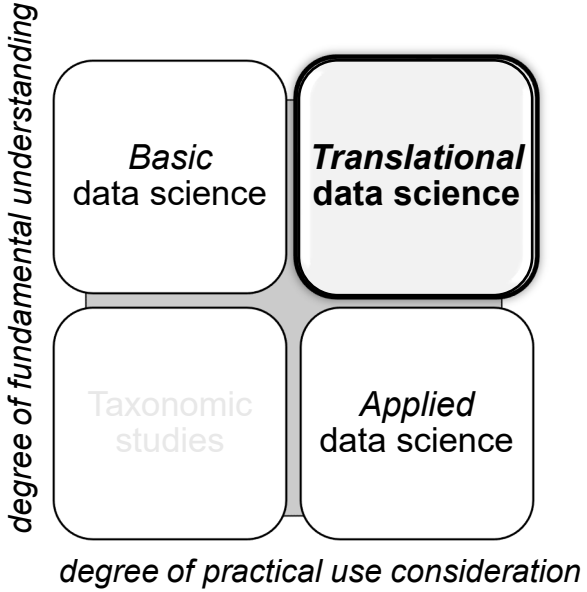
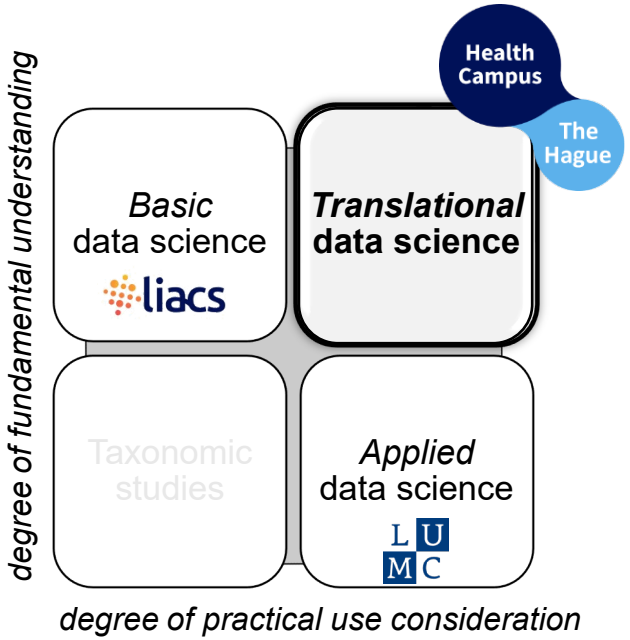
Health Campus Den Haag

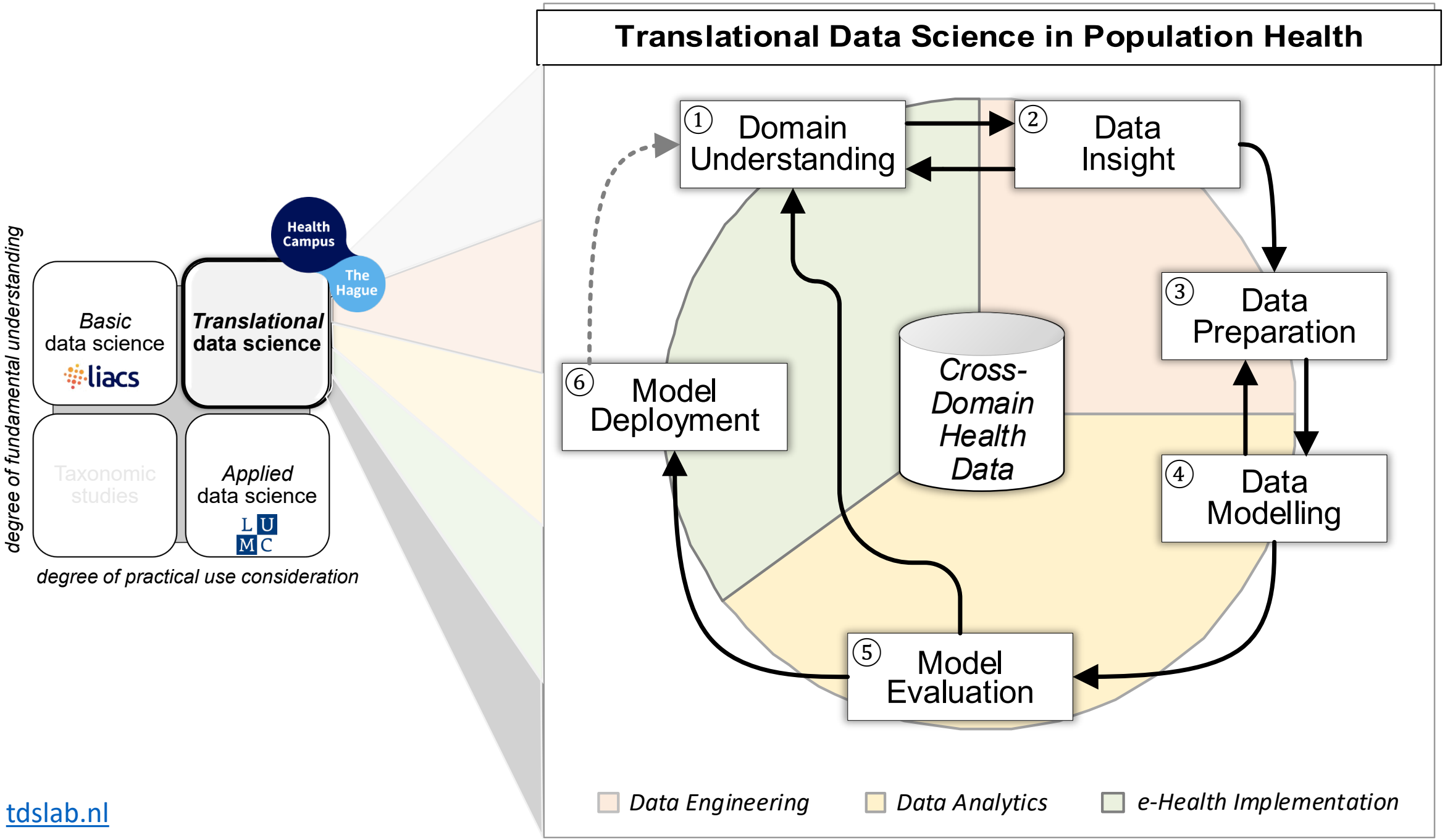
Translational Data Science in Population Health

…Data Techniques and Methodology for Violence as a Public Health Problem

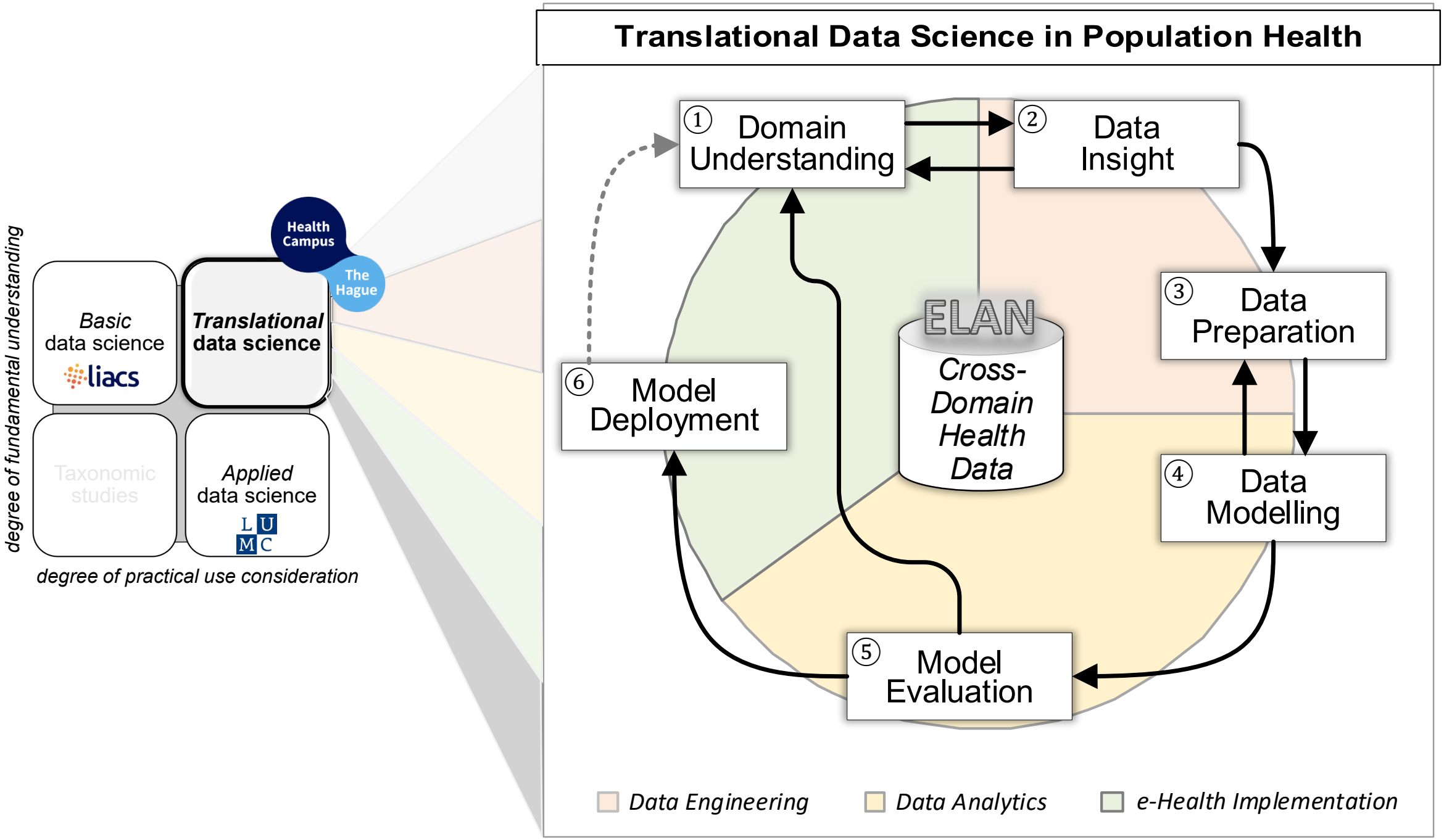
KIEM Pressure Cooker Workshop, 11 March 2024, **Marco Spruit**

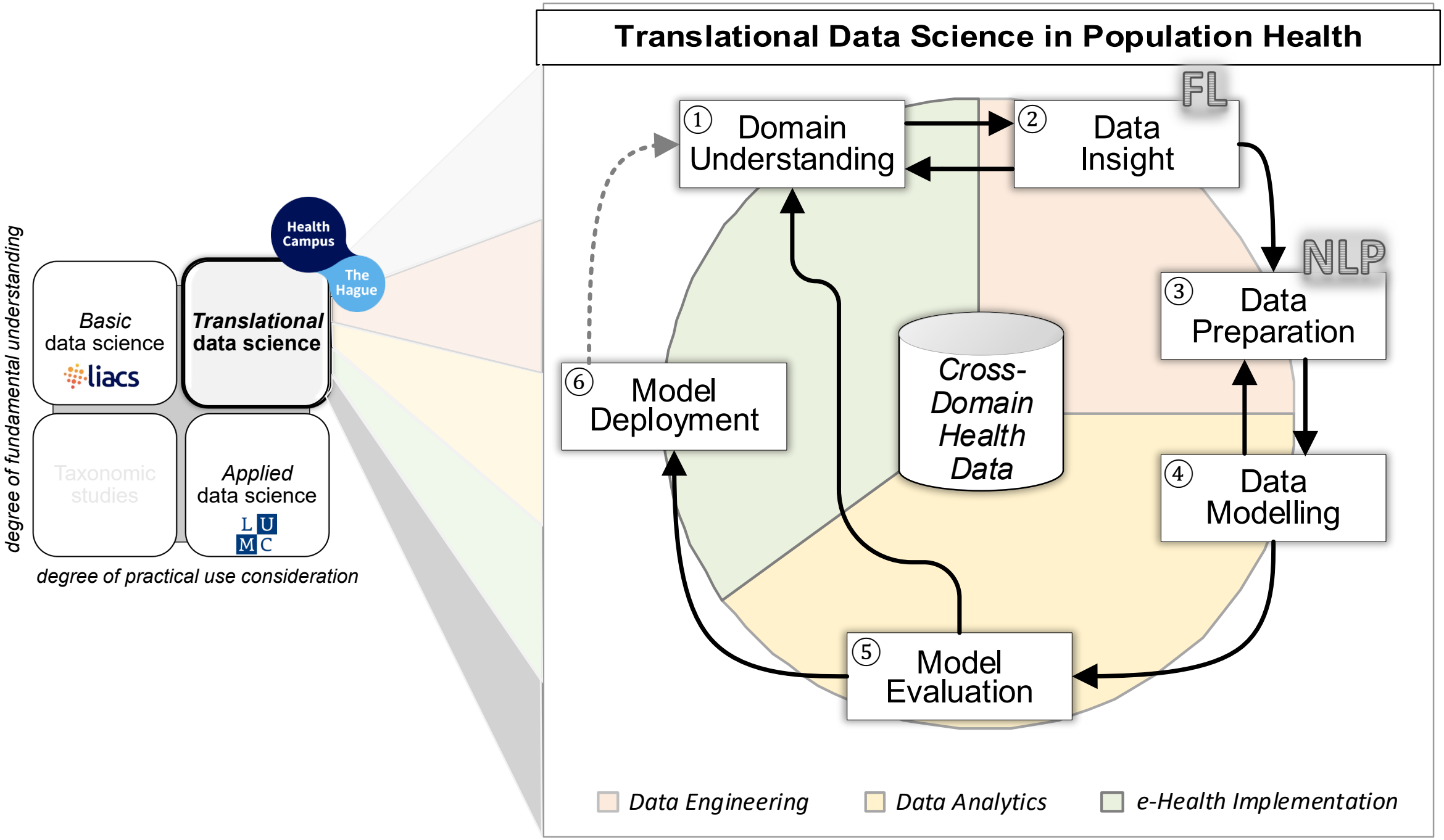

Basic
data science
Translational
data science
Taxonomic
studies
Applied
data science
degree of fundamental understanding
degree of practical use consideration

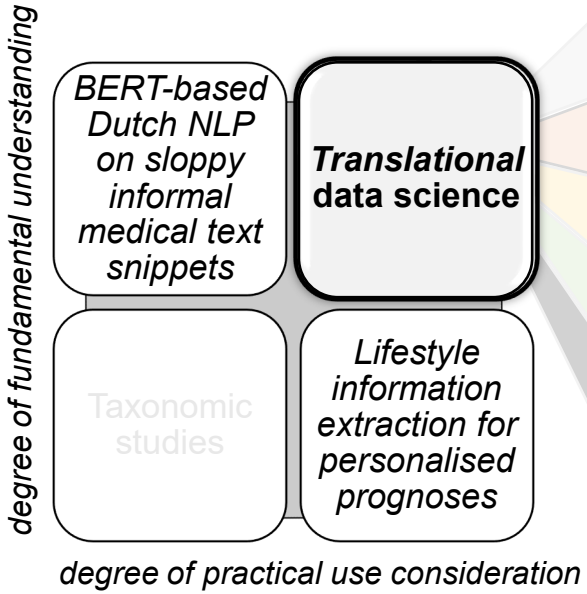

Health Campus
The Hague
Basic
data science
liacs
Translational
data science
Taxonomic
studies
Applied
data science
L U
M C
degree of fundamental understanding
degree of practical use consideration


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
Cross-Domain Health Data
Data Engineering
Data Analytics
e-Health Implementation

tdslab.nl


Translational Data Science in Population Health
degree of fundamental understanding
degree of practical use consideration
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
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
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
FL
NLP
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

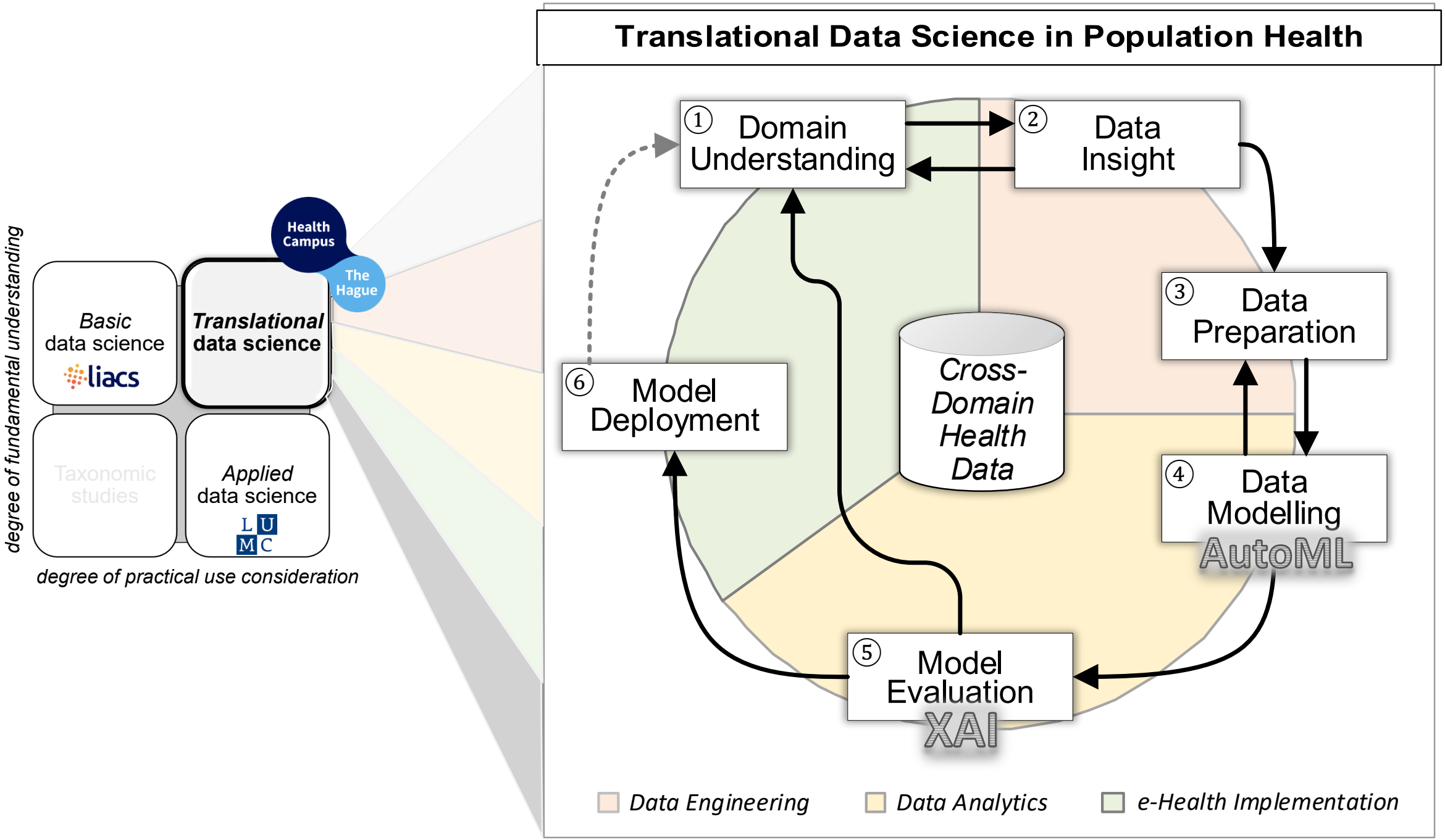


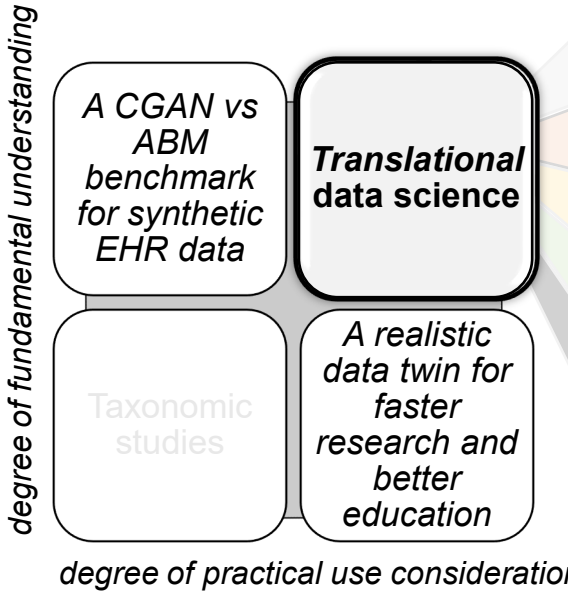
Example: EHR Lifestyle Characteristics Extraction

- Muizelaar, Haas, Dortmont, Putten & Spruit. Extracting Patient Lifestyle Characteristics from Dutch Clinical Text with BERT Models (under review)

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

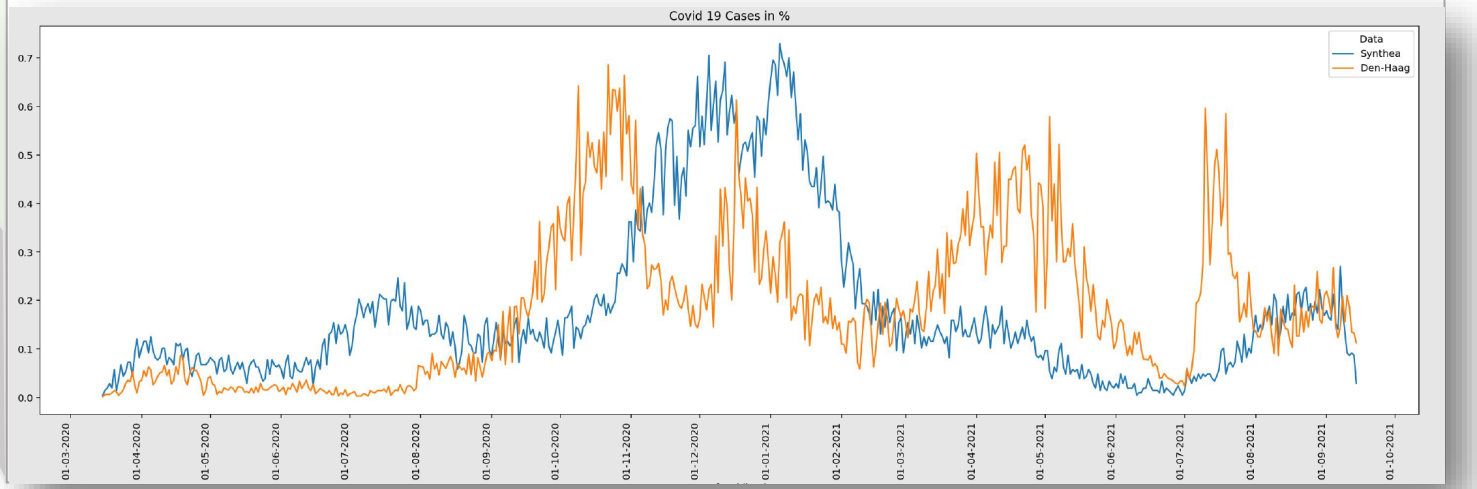
Model	Smoking	Alcohol	Drugs
String Matching	0.84	0.74	0.68
Machine Learning (SGD)	0.85	0.71	0.60
HAGALBERT	0.66	0.54	0.43
RobBERT-HAGA	0.87	0.71	0.63
belabBERT-HAGA	0.48	0.64	0.57
MedRoBERTa.nl-HAGA	0.93	0.79	0.77
BioBERT (translated)	0.91	0.72	0.52
ClinicalBERT (translated)	0.92	0.80	0.61

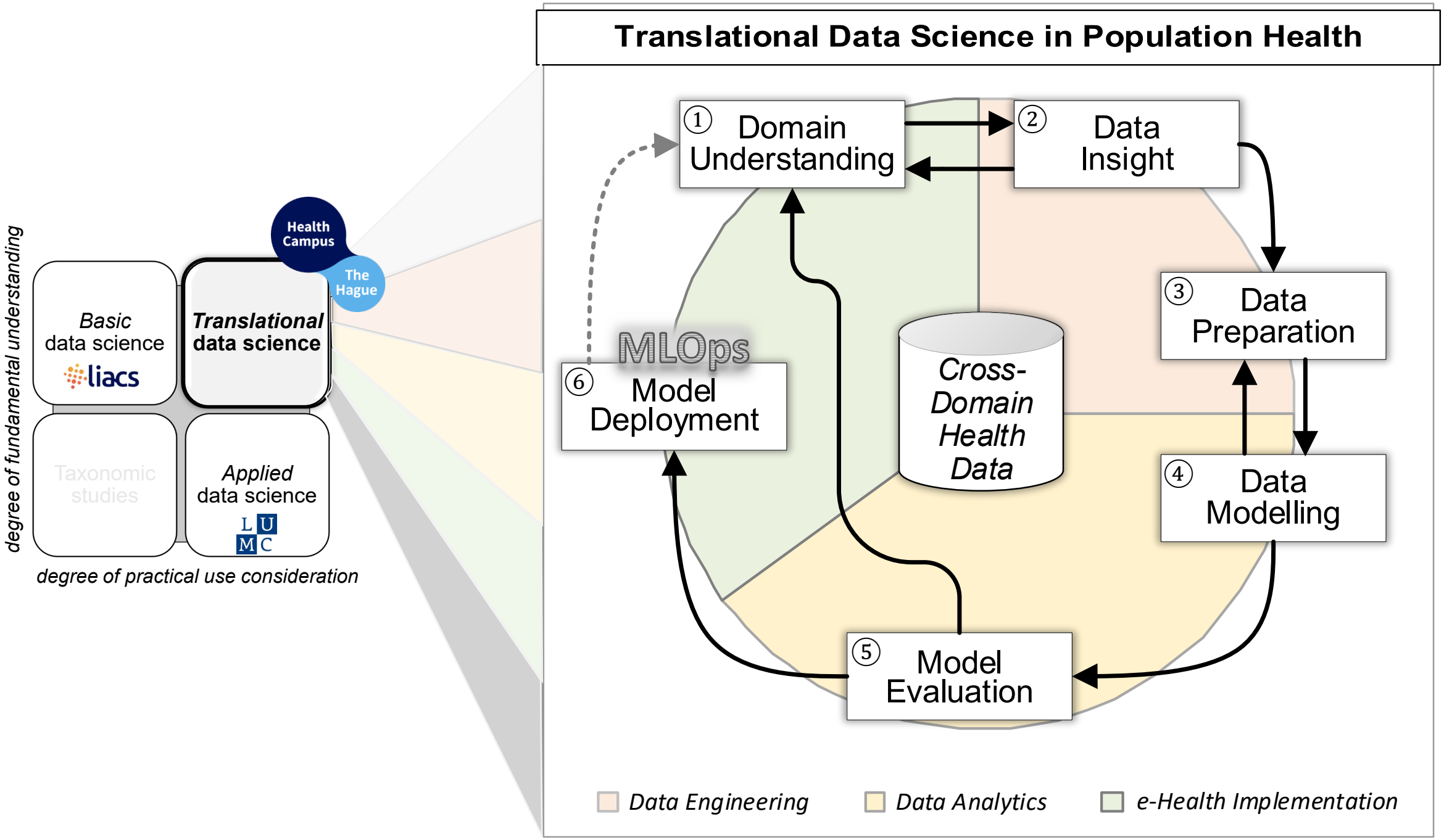

Translational Data Science in Population Health
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of fundamental understanding
degree of practical use consideration
① Domain Understanding
② Data Insight
③ Data Preparation
④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



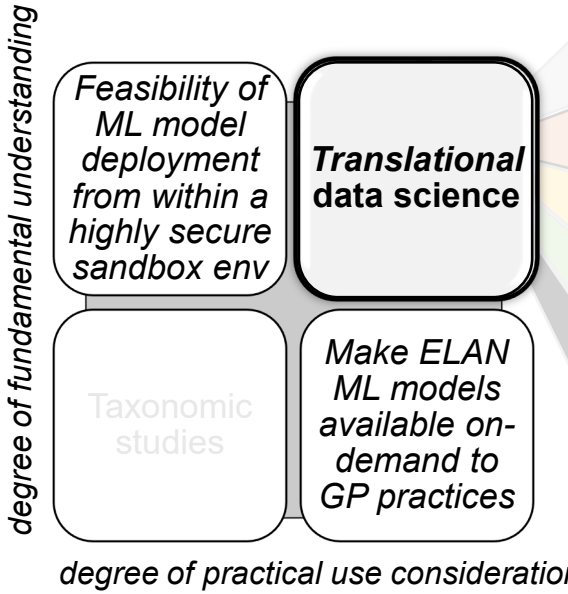
Example: Virtual Patients and Population dataset

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

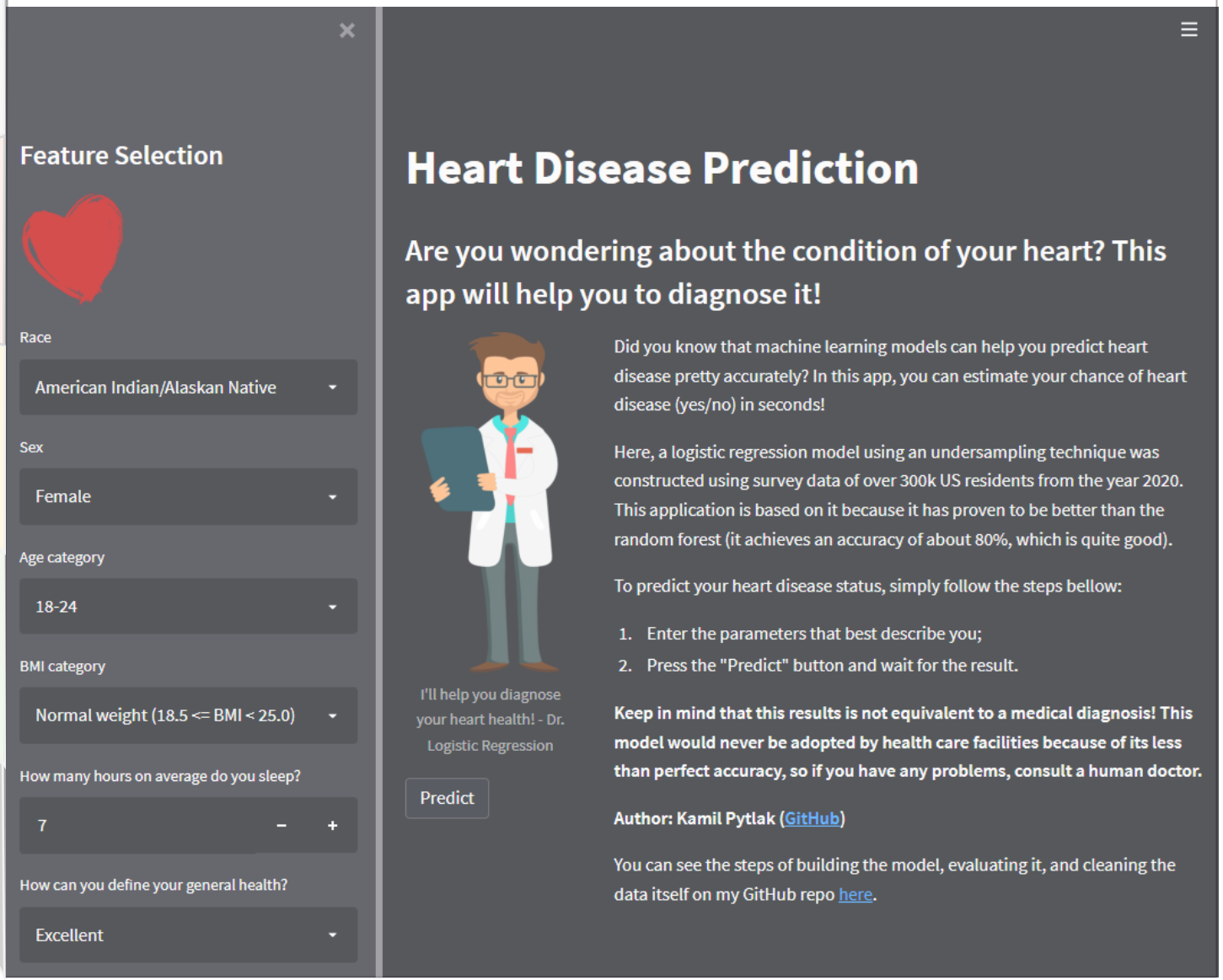


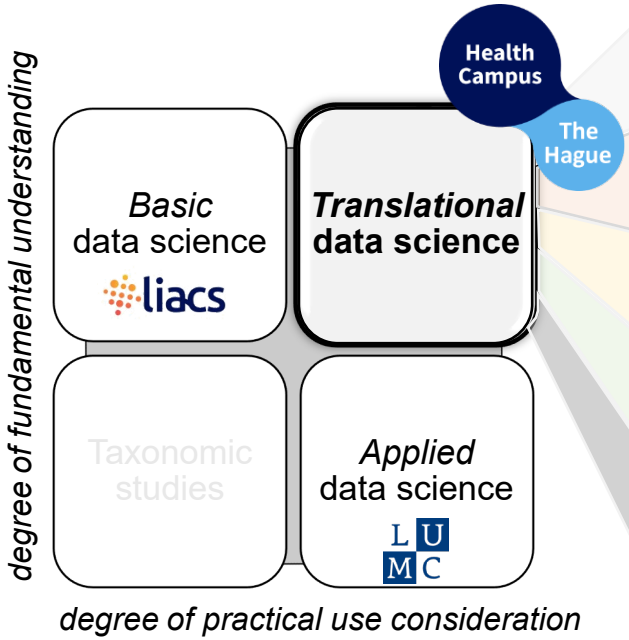

Translational Data Science in Population Health
Health Campus
The Hague
Basic data science
liacs
Translational data science
Taxonomic studies
Applied data science
LUMC
degree of fundamental understanding
degree of practical use consideration
① Domain Understanding
② Data Insight
③ Data Preparation
④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

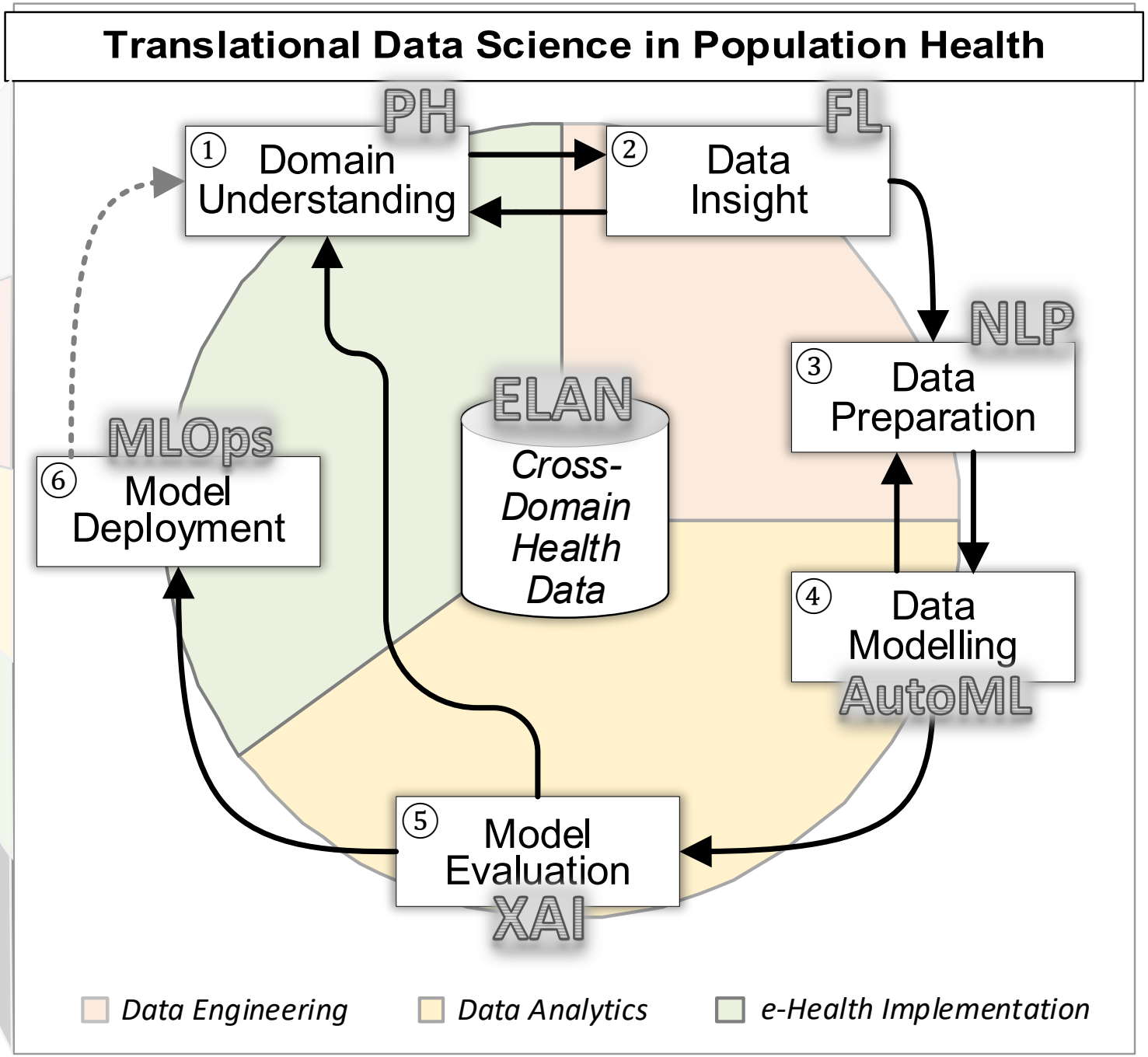
Example: Prediction Model Deployment for CDSS

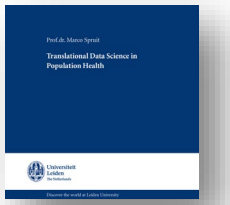


Lisanne Wallaard GitHub Thesis **Demo**



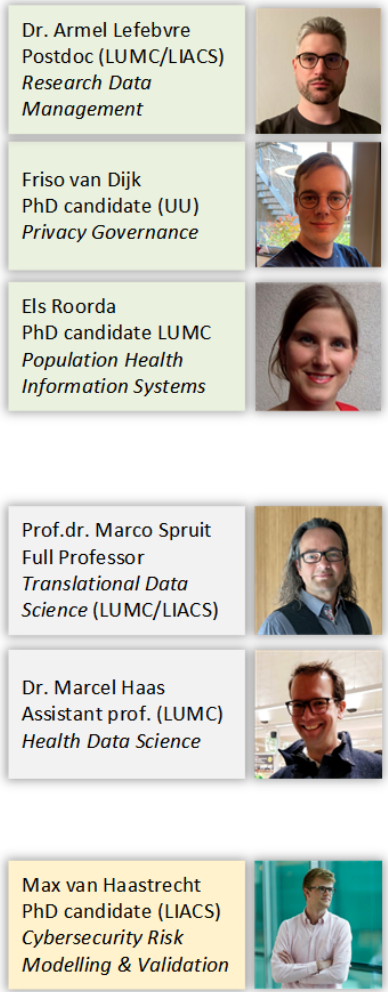






Spruit, Marco. (2022). *Translational Data Science in Population Health* (p. 20). Inaugural lecture. Leiden University. https://doi.org/10.5281/zenodo.7665858

Translational Data Science LAB : members *d.d. 14-12-2023*



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

Friso van Dijk
PhD candidate (UU)
Privacy Governance

Els Roorda
PhD candidate 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 candidate (LIACS)
Cybersecurity Risk Modelling & Validation

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

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

Bram van Dijk
PhD candidate (LIACS)
Mindreading with NLP

Emil Rijcken
PhD candidate (TUE)
Topic Modelling in NLP

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

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

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

http://tdslab.nl