Translational Data Science for Population Health

Marco Spruit | Research Overview

1 Febr 2021

Leiden University
Campus The Hague

Agenda

1. *Introductions*
   - About me, Data Science concepts
2. Research projects showcase
   - STRIPA, COVIDA, SNPcurator
3. Research agenda
   - Research Framework & Instantiation
4. Discussion
   - Alignment with NLAIC wg B&PP

Translational Data Science for Population Health

Introduction: Marco Spruit

Engineer

1993
- Information Retrieval programmer, ZyLAB Europe

1995
- Big Data system developer, Royal Netherlands Navy (Intelligence & Security Service)

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

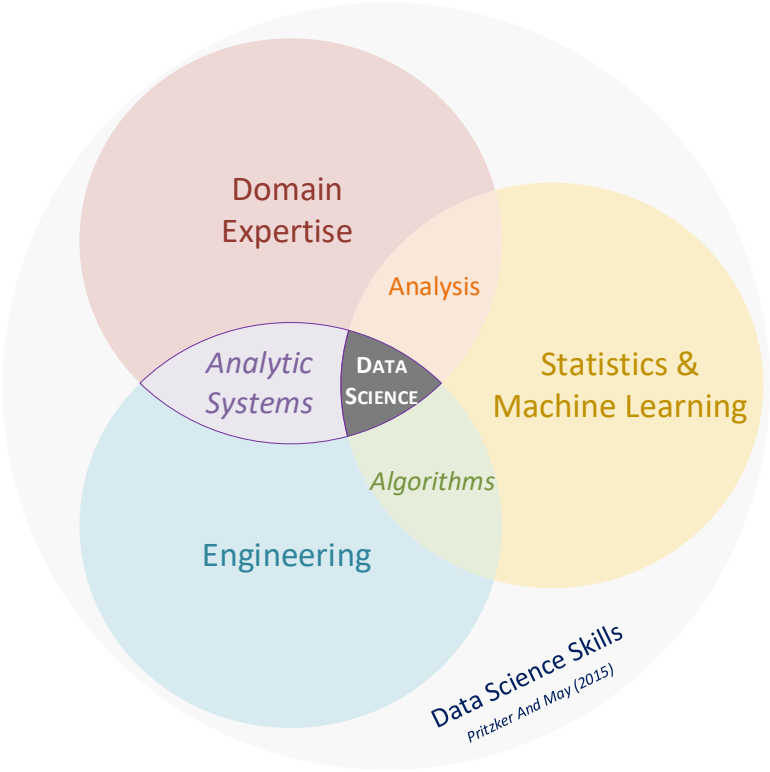
2020
- Professor Advanced Data Science in Population Health, LUMC/Leiden University
  - PH Living Lab
  - CAIRE Lab

Introduction: Data Science Continuum

- **Applied data science** is the knowledge discovery *process* in which analytical applications are designed and evaluated to improve the daily practices of domain experts (Spruit & Jagesar, 2016 ; Spruit & Lytras, 2018)
- **Self-service data science** is the knowledge discovery process in which analytic systems are designed and evaluated to empower domain professionals to perform their own data analyses on their own data sources *without coding* in a reliable, usable and transparent manner within their own daily practices (Spruit & Vries, 2020; Ooms & Spruit, 2020)

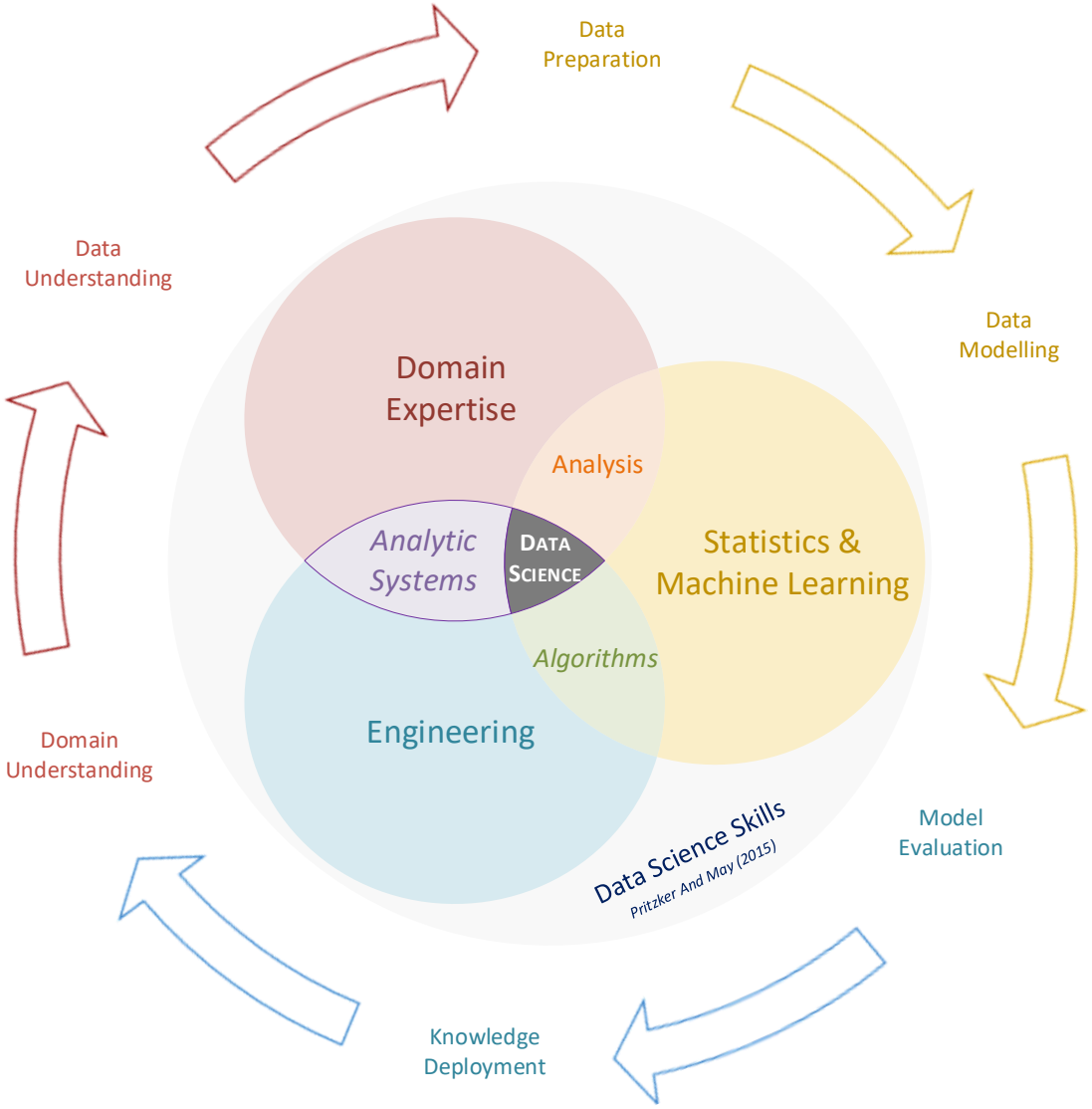
- **Translational Data Science** = *Applied* + *Self-service* data science

Introduction: Data Science Skills



Pritzker, P., & May, W. (2015). NIST Big Data interoperability Framework (NBDIF): Volume 1: Definitions. *NIST Special Publication, 1500*(1). https://doi.org/10.6028/NIST.SP.1500-1r2

Introduction: Data Science Process

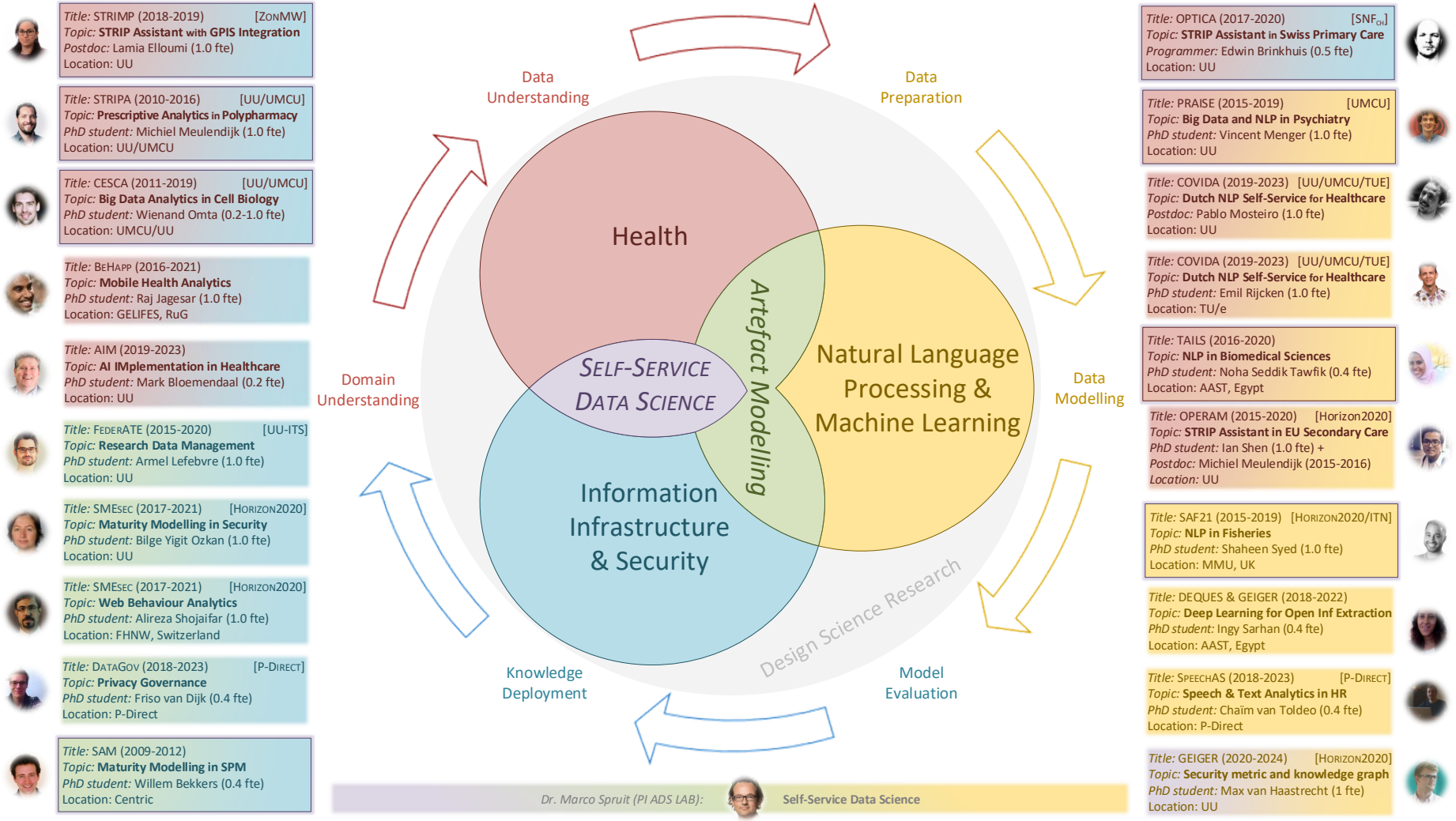


Agenda

1. Introductions
   - About me, Data Science concepts
2. *Research projects showcase*
   - STRIPA, COVIDA, SNPcurator
3. Research agenda
   - Research Framework & Instantiation
4. Discussion
   - Alignment with NLAIC wg B&PP

Translational Data Science for Population Health

Projects: Applied Data Science Lab



Title: STRIMP (2018-2019) [ZonMW]
Topic: **STRIP Assistant with GPIS Integration**
Postdoc: Lamia Elloumi (1.0 fte)
Location: UU

Title: STRIPA (2010-2016) [UU/UMCU]
Topic: **Prescriptive Analytics in Polypharmacy**
PhD student: Michiel Meulendijk (1.0 fte)
Location: UU/UMCU

Title: CESCA (2011-2019) [UU/UMCU]
Topic: **Big Data Analytics in Cell Biology**
PhD student: Wienand Omta (0.2-1.0 fte)
Location: UMCU/UU

Title: BeHapp (2016-2021)
Topic: **Mobile Health Analytics**
PhD student: Raj Jagesar (1.0 fte)
Location: GELIFES, RuG

Title: AIM (2019-2023)
Topic: **AI IMplementation in Healthcare**
PhD student: Mark Bloemendaal (0.2 fte)
Location: UU

Title: FederATE (2015-2020) [UU-ITS]
Topic: **Research Data Management**
PhD student: Armel Lefebvre (1.0 fte)
Location: UU

Title: SMEsec (2017-2021) [Horizon2020]
Topic: **Maturity Modelling in Security**
PhD student: Bilge Yigit Ozkan (1.0 fte)
Location: UU

Title: SMEsec (2017-2021) [Horizon2020]
Topic: **Web Behaviour Analytics**
PhD student: Alireza Shojaifar (1.0 fte)
Location: FHNW, Switzerland

Title: DataGov (2018-2023) [P-Direct]
Topic: **Privacy Governance**
PhD student: Friso van Dijk (0.4 fte)
Location: P-Direct

Title: SAM (2009-2012)
Topic: **Maturity Modelling in SPM**
PhD student: Willem Bekkers (0.4 fte)
Location: Centric

Title: OPTICA (2017-2020) [SNF_{CH}]
Topic: **STRIP Assistant in Swiss Primary Care**
Programmer: Edwin Brinkhuis (0.5 fte)
Location: UU

Title: PRAISE (2015-2019) [UMCU]
Topic: **Big Data and NLP in Psychiatry**
PhD student: Vincent Menger (1.0 fte)
Location: UU

Title: COVIDA (2019-2023) [UU/UMCU/TUE]
Topic: **Dutch NLP Self-Service for Healthcare**
Postdoc: Pablo Mosteiro (1.0 fte)
Location: UU

Title: COVIDA (2019-2023) [UU/UMCU/TUE]
Topic: **Dutch NLP Self-Service for Healthcare**
PhD student: Emil Rijcken (1.0 fte)
Location: TU/e

Title: TAILS (2016-2020)
Topic: **NLP in Biomedical Sciences**
PhD student: Noha Seddik Tawfik (0.4 fte)
Location: AAST, Egypt

Title: OPERAM (2015-2020) [Horizon2020]
Topic: **STRIP Assistant in EU Secondary Care**
PhD student: Ian Shen (1.0 fte) +
Postdoc: Michiel Meulendijk (2015-2016)
Location: UU

Title: SAF21 (2015-2019) [Horizon2020/ITN]
Topic: **NLP in Fisheries**
PhD student: Shaheen Syed (1.0 fte)
Location: MMU, UK

Title: DEQUES & GEIGER (2018-2022)
Topic: **Deep Learning for Open Inf Extraction**
PhD student: Ingy Sarhan (0.4 fte)
Location: AAST, Egypt

Title: SpeechAS (2018-2023) [P-Direct]
Topic: **Speech & Text Analytics in HR**
PhD student: Chaïm van Toldeo (0.4 fte)
Location: P-Direct

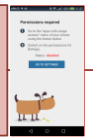
Title: GEIGER (2020-2024) [Horizon2020]
Topic: **Security metric and knowledge graph**
PhD student: Max van Haastrecht (1 fte)
Location: UU

Projects: Analytic Systems

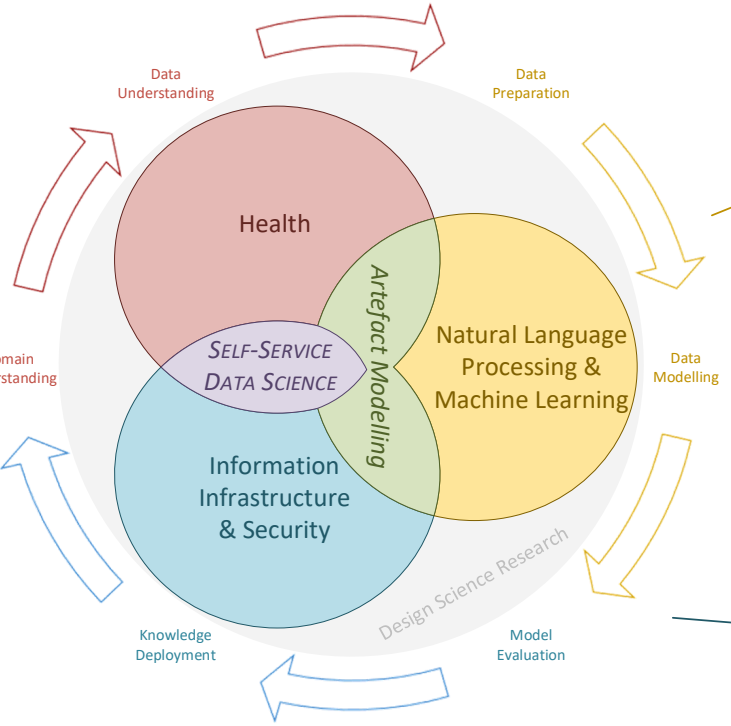
Sample #1

Title: STRIP ASSISTANT
Location: http://stripa.eu/english/
Description: Prescriptive analytics in polypharmacy

Title: BEHAPP
Location: https://behapp.org
Description: Passive mobile health analytics in psychiatry



Title: HC StratoMineR
Location: https://corelifeanalytics.com
Description: Big data analytics in cell screening



Sample #2

Title: DEDUCE
Location: http://ads.science.uu.nl/deduce/
Description: Deidentification analytics in Dutch medical text

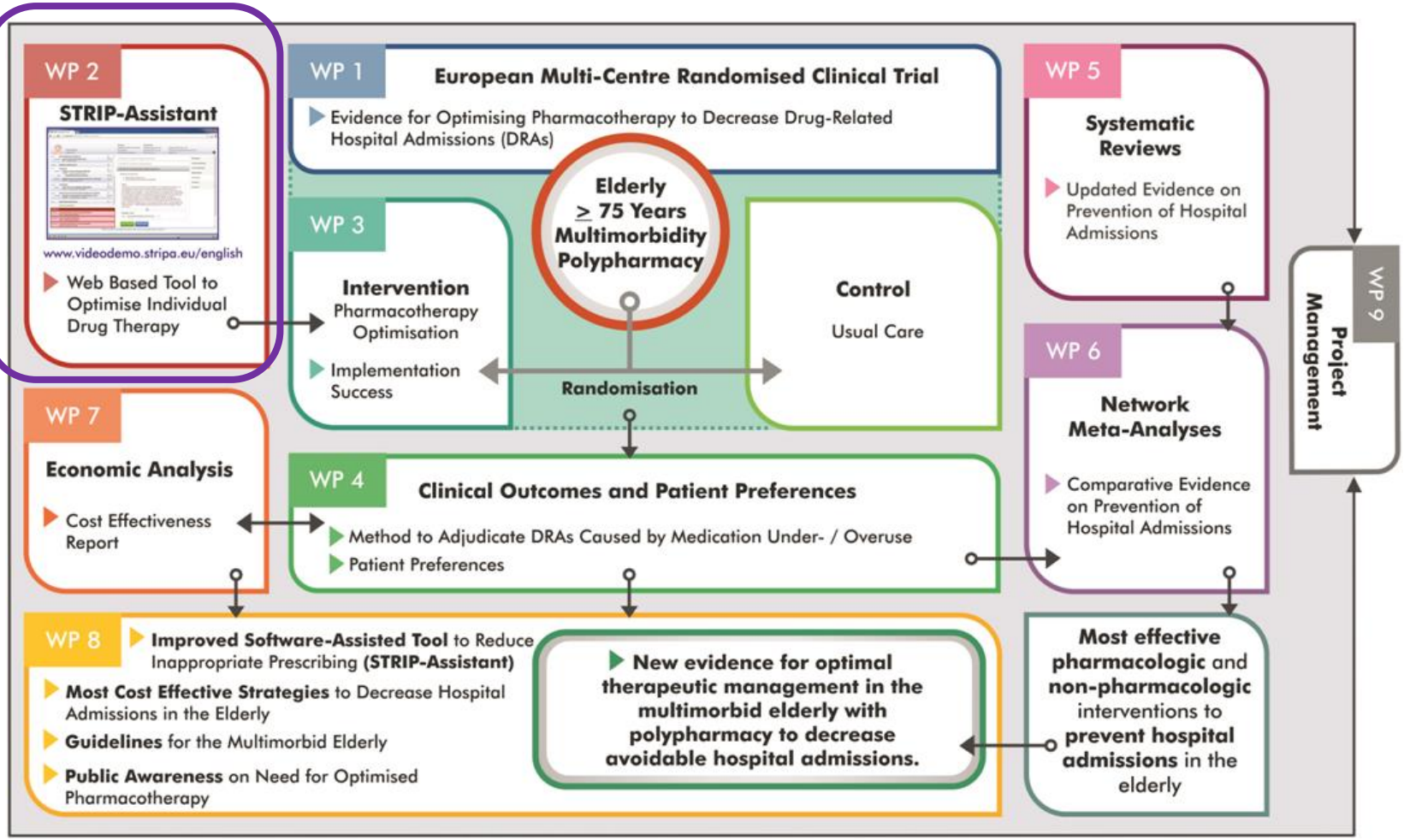
Sample #3

Title: SNPCURATOR
Location: http://snpcurator.science.uu.nl
Description: Text analytics in genome wide association studies

Title: CYSEC
Location: https://smesec.eu
Description: Behaviorial analytics on cybersecurity advice for SMEs

Sample #1: STRIP Assistant (STRIPA)

- A Clinical Decision Support System for Medication Reviews



url: https://bit.ly/stripa3-nl-preview

Sample #1: STRIP Assistant (STRIPA)

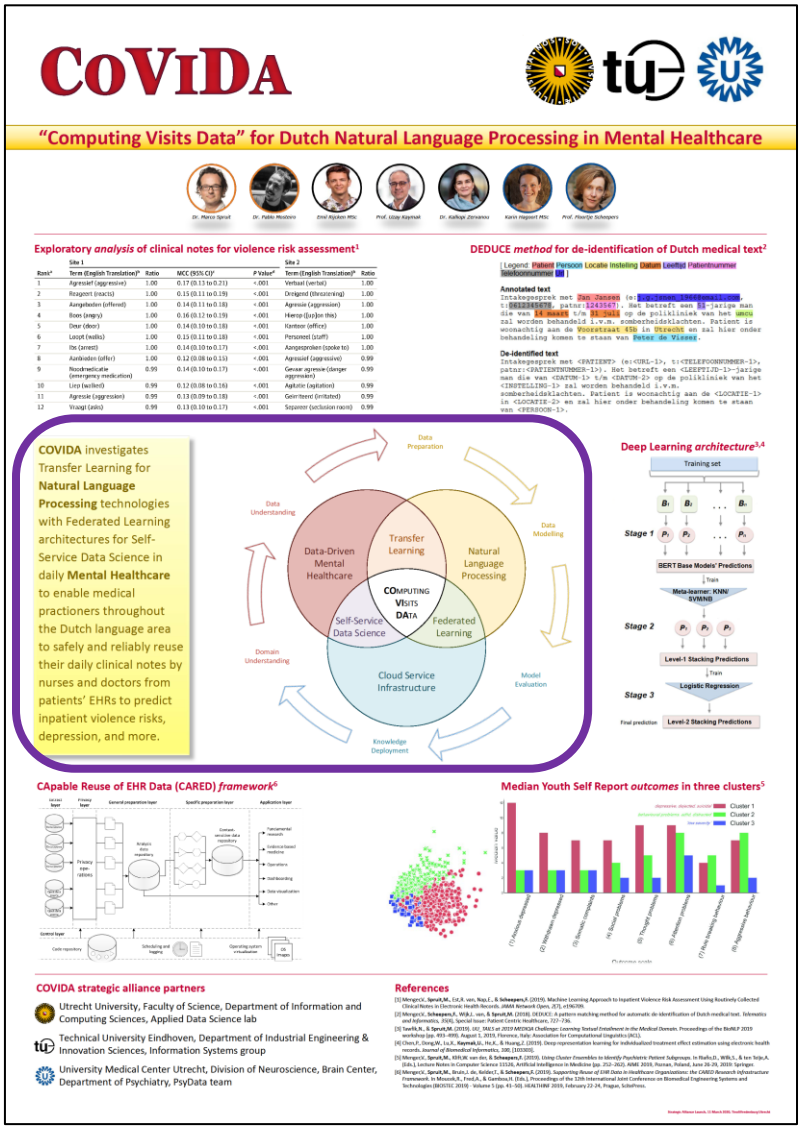
OPERAM Dr. Marco Spruit managing — V2020.05b.04a / IE / EN / SS

Add new patient

Patient 1	M	75	No	No	No	START
Patient 2	F	77	No	No	No	START
Patient 3	M	99	No	No	Yes	START
Patient 4	F	79	No	No	Yes	START
Patient 5	F	81	No	No	Yes	START
Patient 6	F	86	No	No	Yes	START
Patient 7	M	78	No	No	No	START
Patient 8	M	80	No	No	Yes	START
Patient 9	F	74	No	No	No	START
Patient 10	F	77	No	No	No	START
Patient 11	M	72	No	No	No	START
Patient 12	F	70	No	No	Yes	START
Patient 13	M	81	No	No	Yes	START
Patient 14	M	70	No	No	Yes	START
Patient 15	M	73	No	No	Yes	START
Patient 16	M	77	No	No	Yes	START
Patient 17	F	67	No	No	Yes	START
Patient 18	M	76	No	No	Yes	START
Patient 19	F	74	No	No	Yes	START
Patient 20	F	80	No	Yes	Yes	START
Patient 21	M	80	No	No	Yes	START
Patient 22	M	79	No	No	Yes	START
Patient 23	F	87	No	No	Yes	START

Open 'https://stripa.nl/OPERAMIE/dashboard?patient=8' in een nieuw tabblad

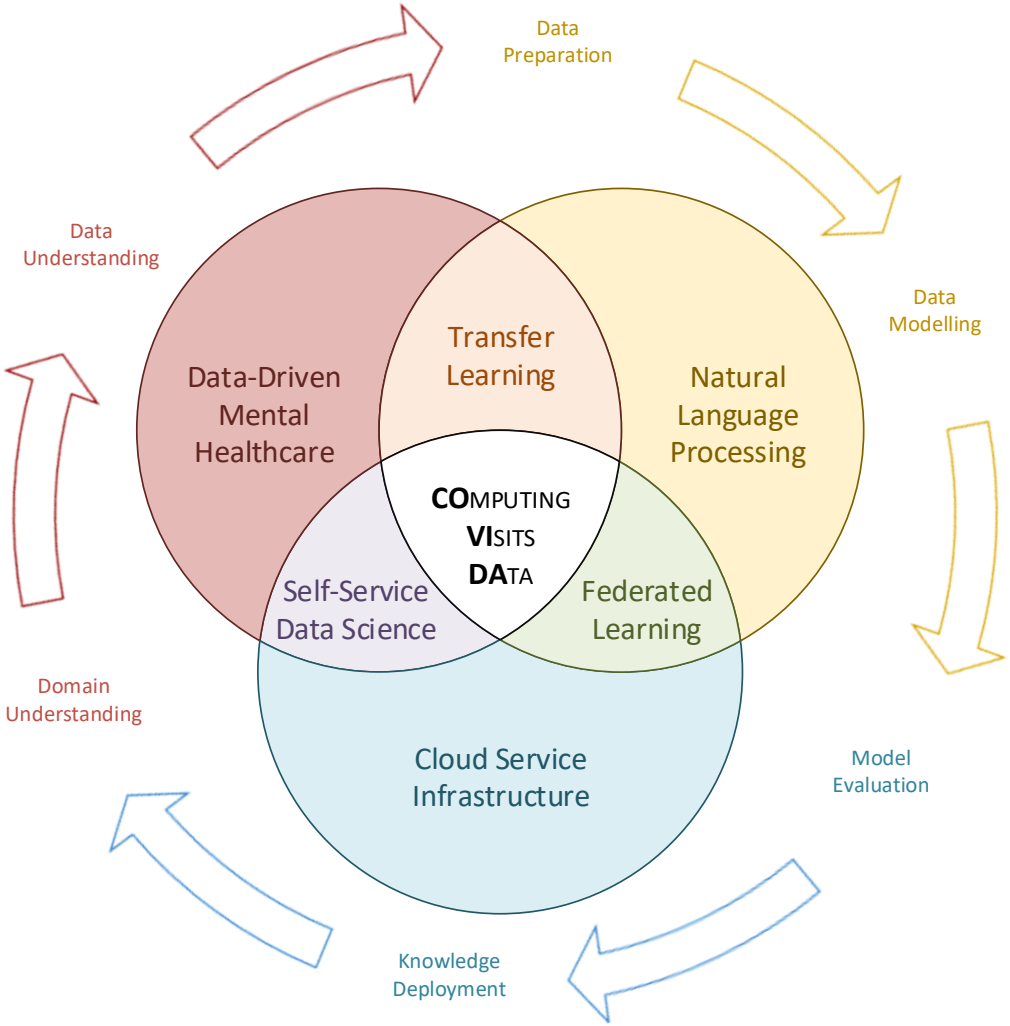
Sample #2: Computing Visits Data (COVIDA)



url: https://bit.ly/covida-poster

Sample #2: Computing Visits Data (COVIDA)

COVIDA investigates Transfer Learning for **Natural Language Processing** technologies with Federated Learning architectures for Self-Service Data Science in daily **Mental Healthcare** to enable medical practioners throughout the Dutch language area to safely and reliably reuse their daily clinical notes by nurses and doctors from patients' EHRs to predict inpatient violence risks, depression, and more.



url: https://bit.ly/covida-poster

Sample #2: Computing Visits Data (COVIDA)

- The DEDUCE method for de-identification of Dutch medical text

[Legend: Patient Persoon Locatie Instelling Datum Leeftijd Patientnummer Telefoonnummer Url]

Annotated text

Intakegesprek met Jan Jansen (e:j.g.jsnen_1966@email.com, t:0612345678, patnr:1243567). Het betreft een 51-jarige man die van 14 maart t/m 31 juli op de polikliniek van het umcu zal worden behandeld i.v.m. somberheidsklachten. Patient is woonachtig aan de Voorstraat 45b in Utrecht en zal hier onder behandeling komen te staan van Peter de Visser.

De-identified text

Intakegesprek met <PATIENT> (e:<URL-1>, t:<TELEFOONNUMMER-1>, patnr:<PATIENTNUMMER-1>). Het betreft een <LEEFTIJD-1>-jarige man die van <DATUM-1> t/m <DATUM-2> op de polikliniek van het <INSTELLING-1> zal worden behandeld i.v.m. somberheidsklachten. Patient is woonachtig aan de <LOCATIE-1> in <LOCATIE-2> en zal hier onder behandeling komen te staan van <PERSOON-1>.

url: http://ads.science.uu.nl/deduce/

Sample #3: SNPcurator

- PubMed Literature mining of enriched SNP-disease associations

SNP Curator

Results for **ibd**:
- A total of **1535** articles were fetched by PubMed.
- The extracted list of abstracts was shortened to **280** via selecting those comprised of SNP mentions.
- **150** PubMed article(s) had statistical results reported withing the abstract text with a total of **524** SNP pairs.

Go back to home page Export Data to CSV File

SNP	PMID	Title	Date	Pvalue	ORvalue	Ethnicity	Patient group Size	Control group Size	Frequency	Text Evidence
rs61750370	29788244	Nonsynonymous Polymorphism in Guanine Monophosphate Synthetase Is a Risk Factor for Unfavorable Thiopurine Metabolite Ratios in Patients With Inflammatory Bowel Disease.		0.031		Caucasian	264		2	-
The SNP rs61750370 was significantly associated with 6-MMP:6-TGN ratios ≥100 odds ratio, 5.64; 95% confidence interval, 1.01-25.12; P < 0.031 in a subset of 264 Caucasian IBD patients.The GMPS SNP rs61750370 may be a reliable risk factor for extreme 6MMP preferential metabolism.										
rs61750370	29788244	Nonsynonymous Polymorphism in Guanine Monophosphate Synthetase Is a Risk Factor for Unfavorable Thiopurine Metabolite Ratios in Patients With Inflammatory Bowel Disease.		0.031		Caucasian	264		2	+
rs16969968	29688464	Smoking Interacts With CHRNA5, a Nicotinic Acetylcholine Receptor Subunit Gene, to Influence the Risk of IBD-Related Surgery.		0.05					4	+

url: https://snpcurator.science.uu.nl/

Agenda

1. Introductions
   - About me, Data Science concepts
2. Research projects showcase
   - STRIPA, COVIDA, SNPcurator
3. *Research agenda*
   - Research Framework & Instantiation
4. Discussion
   - Alignment with NLAIC wg B&PP

Translational Data Science for Population Health

Problem: Three Challenges to Appropriate Care

- Regarding *Diagnostics*:

  New, clinically relevant and personalised patient profiles are urgently needed as a starting point for diagnostics and treatment, that do justice to the impact of a patient's context and the dynamics of health. [*i.e.* NLP/ML]

- Regarding *Treatment*:

  New, personalised treatment decision support systems (DSS) are needed to improve shared decision making, treatment effectivity and outcome of especially complex patients. [*i.e.* Infrastructure]

- Regarding *Monitoring*:

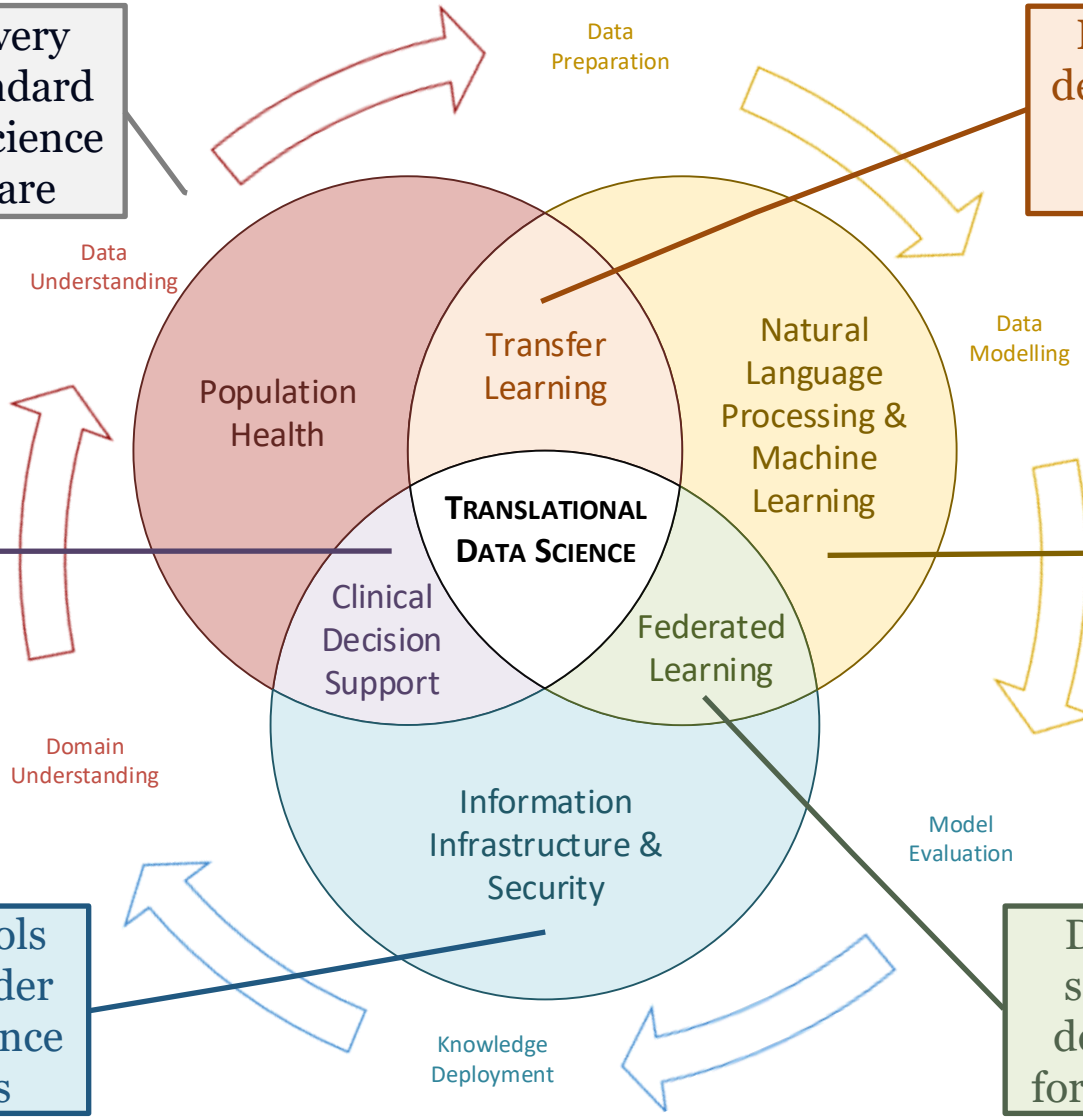
  New, patient-centred network approaches are needed to stimulate effective appropriate care. [*i.e.* Health process]

"Personalisation"

Agenda: Translational Data Science for PH

Knowledge discovery process model standard for applying data science in Dutch healthcare

Language model for deep understanding of Dutch healthcare communications



Self-service clinical decision support systems for Dutch Healthcare

Multimodal machine learning algorithms for Dutch healthcare

Governance controls for multi-stakeholder technology acceptance in daily practices

Distributed patient-sensitive knowledge deployment platform for continuous learning

Example Research Question:

What? How can we improve population health
How? with *e*-mental health and medication services
Who? for socioeconomically vulnerable older adults
Why? to promote healthy aging
Where? in the Population Health Living Lab?

Universiteit Leiden
The Netherlands

"LiLa"

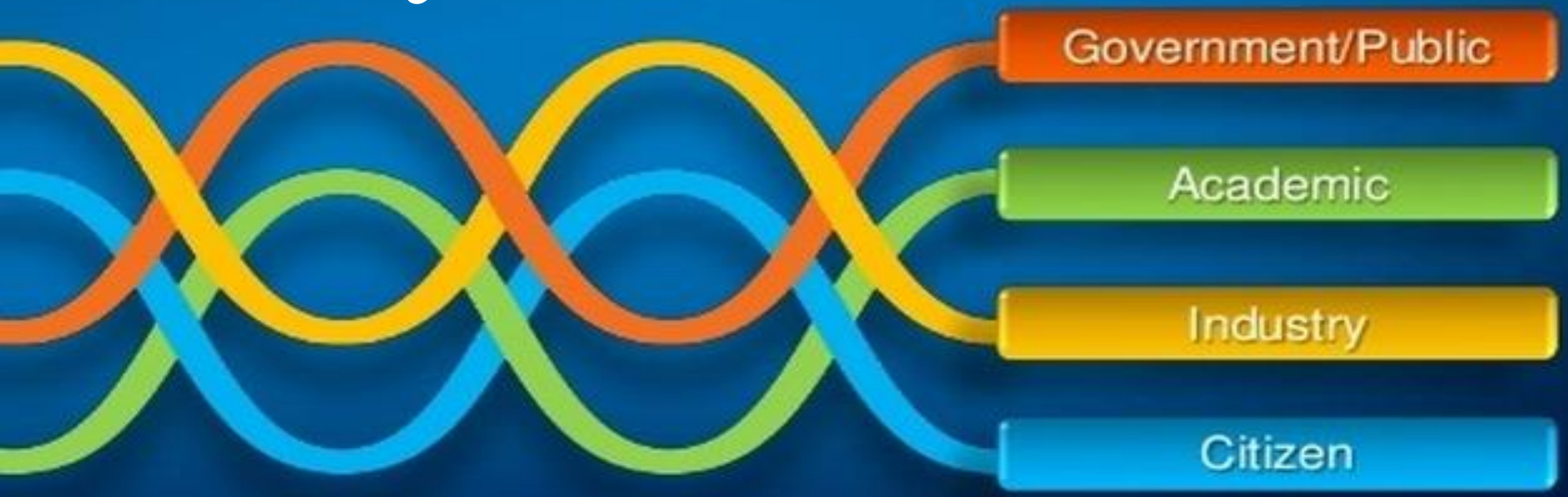
NWO

Agenda

1. Introductions
   - About me, Data Science concepts
2. Research projects showcase
   - STRIPA, COVIDA, SNPcurator
3. Research agenda
   - Research Framework & Instantiation
4. *Discussion?*
   - Alignment with NLAIC wg B&PP

Translational Data Science for Population Health

Thank you!



m.r.spruit@lumc.nl

Leiden University Medical Center

Leiden University Campus The Hague