Quarterly meeting
April 22nd, 2021

EPI

The meeting starts at

13:05

Agenda

13:05-13:15	Introduction to EPI
13:15-13:50	Break-out sessions: - Policies — by Corrette Ploem & André Krom - Algorithms — by Giovanni Cinà - Architecture — by Douwe Lycklama - Organization — by Floortje Scheepers & Sander Klous - Use cases — by Rianne Fijten
13:50-14:15	Discussion of insights gained
14:15-14:30	*Coffee-break*
14:30-15:40	Progress updates PhD students: - Rosanne Turner – Real-time evidence collection in data stream - Corinne Allaart – Vertical Distributed Learning for CVA rehabilitation - Saba Amiri – Privacy Preserving Distributed Machine Learning - Milen Kebebe – Automating normative control for Healthcare research eFlint specification for regulatory documents in DIPG-research - Jamila Alsayed Kassem – A dynamic infrastructure to support health applications
15:40-15:55	EPI Architecture update: implementation & integrations
15:55-16:00	Any other business
16:00	End

Who are we? - research institutions, healthcare providers and the private sector working together

Research institutions

vrije Universiteit

CWI

Universiteit van Amsterdam

Healthcare providers

St Antonius een santeon ziekenhuis

prinses máxima centrum kinderoncologie

UMC Utrecht

parnassia groep Specialist in geestelijke gezondheid

Private sector

SURF

KPMG

PHILIPS

Ministerie van Economische Zaken

NWO

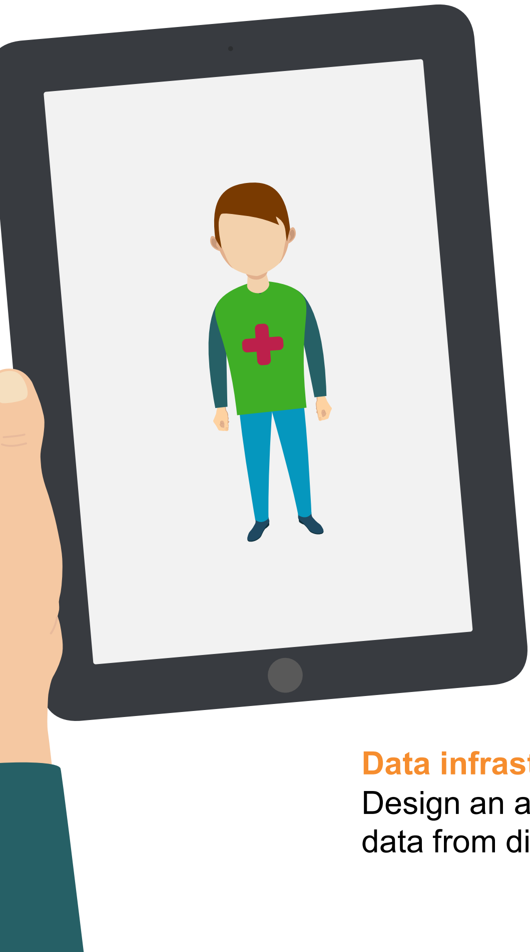
Commit 2 Data connecting business and science

santeon

white box

ZonMw

Our research objective: designing, distributing and saving adaptive data in a secure infrastructure



The outcome of the EPI project is a digital health twin for self-join management

- All data will be collected of a patient
- Inform health decisions and avoiding unnecessary treatment Empower self/joint management of disease
- Able to perform with data gathered from different sources
- Deal with the variability, ownership, data protection and privacy issues

Distribution of Data & Algorithm
Making accurate predictions while preserving privacy constraints of remote data sources

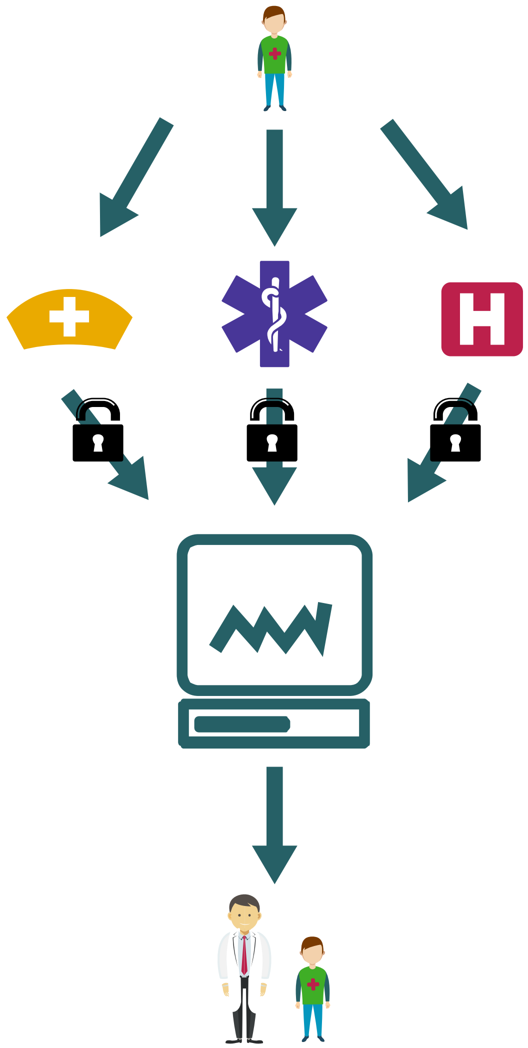
Regulatory constraints and data governance
Automating the process of data sharing with different legal constraints

Data infrastructure
Design an architecture for the data from different sources

Adaptive health diagnosis
The models should be able to keep learning from new data and treatments

Analyzing interventions
Develop models that can predict the effectiveness interventions

Use case 1: A complete overview of data of CVA patients is missing



ISSUE

- Patient data is spread over institutions
- A complete overview of the patient data is missing, which makes prediction of outcomes (survival, functional status, quality of life) difficult

AIM

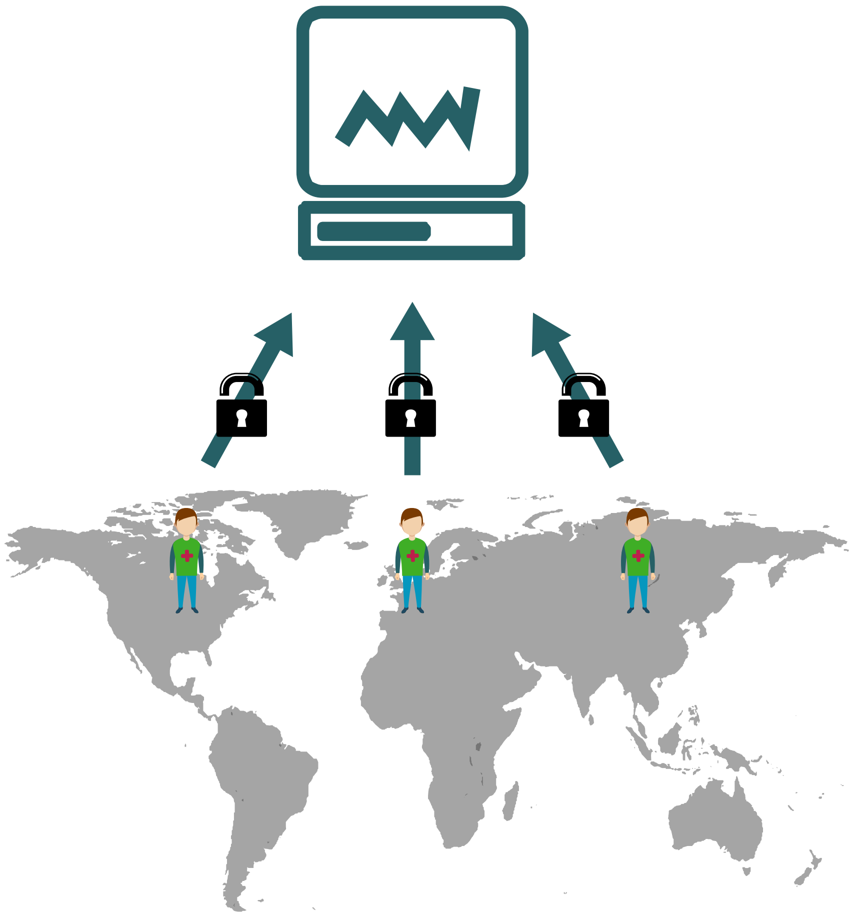
- Understand how prediction can support shared decision-making and self-management and result in better health outcomes
- Set up a prediction model for individual patients to inform them on their outcomes
- The tool will be used in shared decision making

METHOD

- Using incomplete data to predict outcomes

ST ANTONIUS
een santeon ziekenhuis

Use case 2: Data consensus differs between institutions and countries of DIPG patients

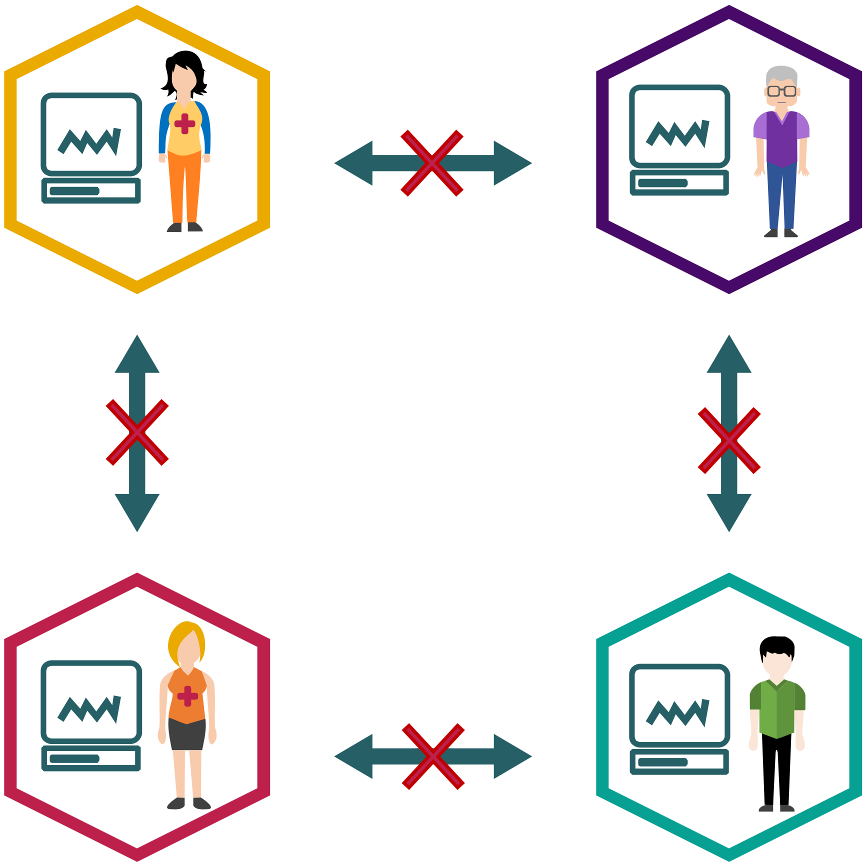


ISSUE
- Number of Diffuse Intrinsic Pontine Glioma (DIPG) patients is small
- Patients are spread across different countries with different data consensus
- Which makes it difficult to get results on treatment effects

AIM
- Integrate different data platforms into one communicating data infrastructure
- Find patterns in data that predict (length of) survival and QoL of DIPG patients
- Create self-learning algorithms that will improve predictive capacities of the DIPG survival prediction model

METHOD
- Use advanced prediction methods and overcome legal issues such as working with consent

Use case 3: Data of psychiatry patients has to stay in its hub complicating integral research



ISSUE

- There is a huge variability between patients in trials.
- Adjust patients to the right medication takes too long
- In daily practice, the variation between patients is even bigger because of strong selection bias in most clinical trials, because data can not be shared

AIM

- Support in decision making about medication and adjustment to medication, which accelerates the process
- Confidence in medication treatment and adherence is increased
- Duration of treatment and suffering are decreased

METHOD

- Developing an algorithm to predict best treatment, while not sharing data, only algorithms and expertise

UMC Utrecht

Break-out sessions in progress

Reconvene at 13:50

Discussion of insights gained

Coffee-break ☕

The meeting resumes at

14:30

Progress updates PhD students

RQ: Real-time evidence collection in data streams

Updates april 2021

- Rosanne J. Turner
- Supervisors and collaborators within EPI:
  - Prof. Peter Grünwald (CWI)
  - Prof. Floor Scheepers (UMCU)
  - Karin Hagoort (UMCU)
  - Dr. Aki Harma (Philips)
  - Roel van Est (Parnassia)

CWI Centrum Wiskunde & Informatica

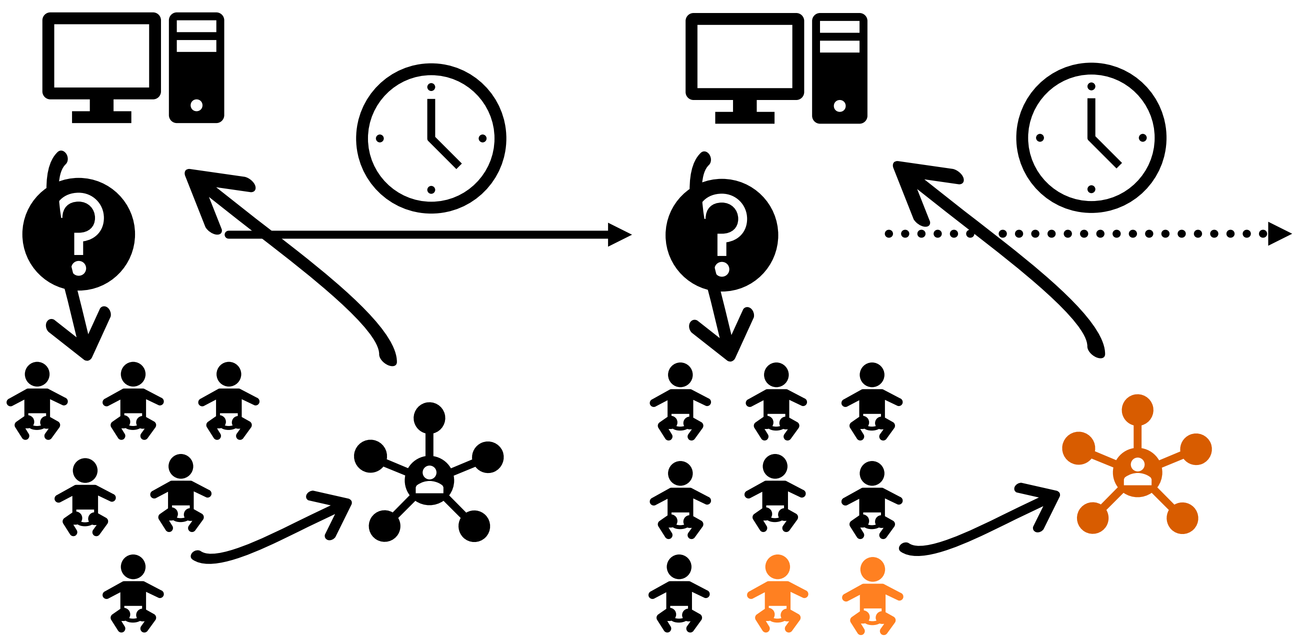
UMC Utrecht

parnassia groep

PHILIPS

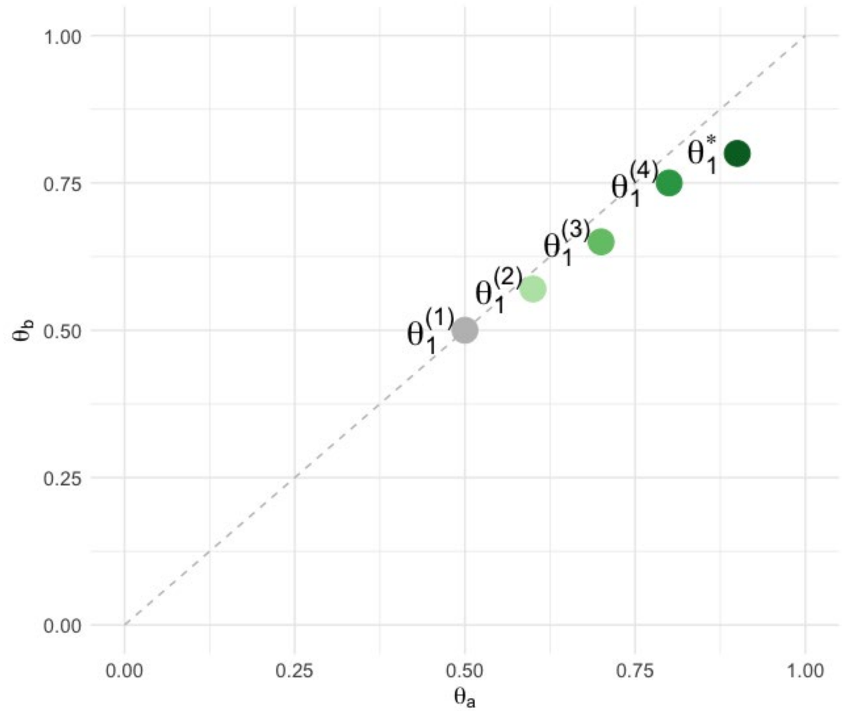
Real-time analysis: safe statistics

- Collect and adjust inference about evidence for treatment strategies **in real time**



Learn from data for adjusting evidence collection process

- With safe statistics can collect **"statistically sound" evidence for A/B testing (e.g. treatment recommendations!)** in real-time
- Learning approach: take safe test for next data block that is optimal relative to estimate $\hat{\theta}_1$ based on data seen so far
- Technical detail: take $\hat{\theta}_1$ as posterior mean for θ_1 with Beta prior on parameter space

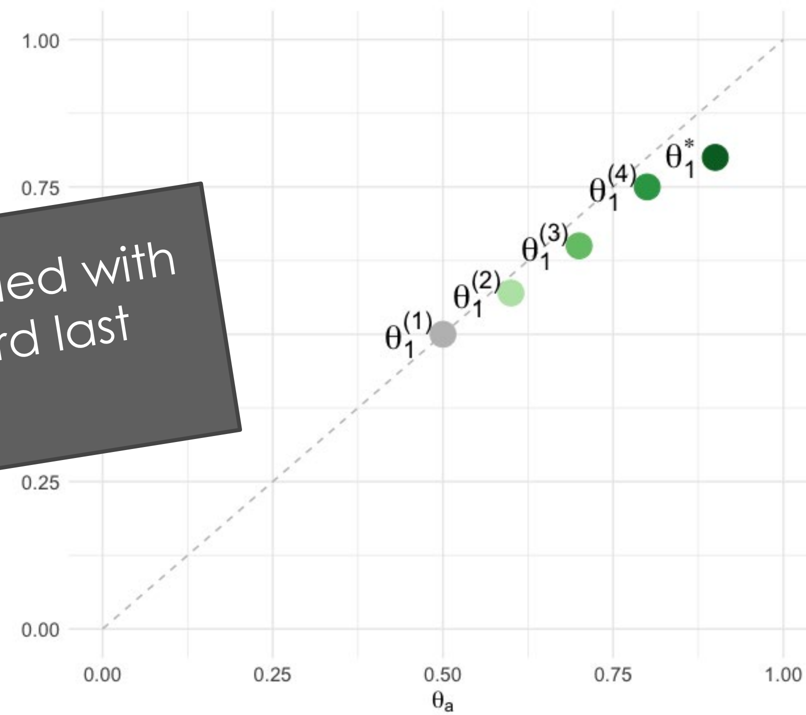


Learn from data for adjusting evidence collection process

- With safe statistics can collect **“statistically sound” evidence for A/B testing (e.g. treatment recommendations!)** in real-time
- Learning a

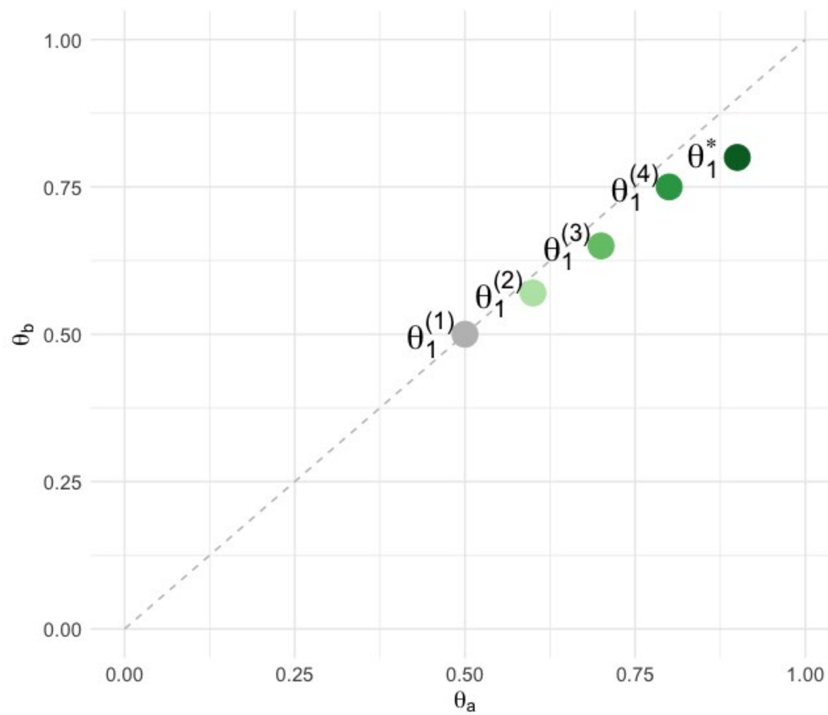
Work partly from master thesis; awarded with the VVSOR Jan van Hemelrijk award last March

- nical detail: take $\hat{\theta}_1$ as posterior mean for θ_1 with Beta prior on parameter space



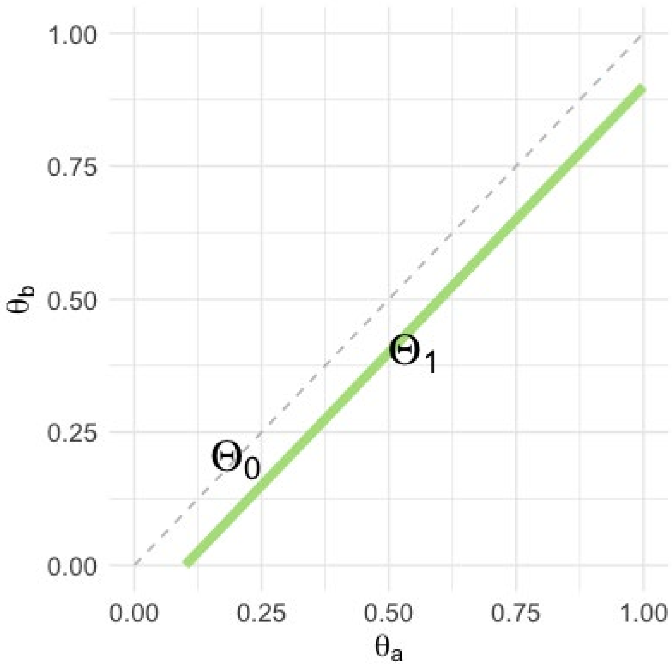
Learn from data for adjusting evidence collection process

- With safe statistics can collect **"statistically sound" evidence for A/B testing (e.g. treatment recommendations!)** in real-time
- Learning approach: take safe test for next data block that is optimal relative to estimate $\hat{\theta}_1$ based on data seen so far
- Technical detail: take $\hat{\theta}_1$ as posterior mean for θ_1 with Beta prior on parameter space

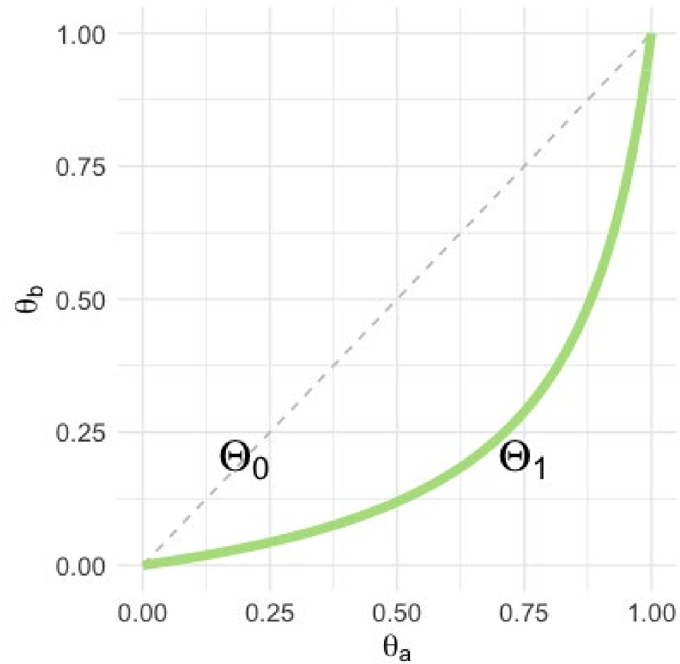


Can also add expert/ prior knowledge: restrict options!

Absolute difference



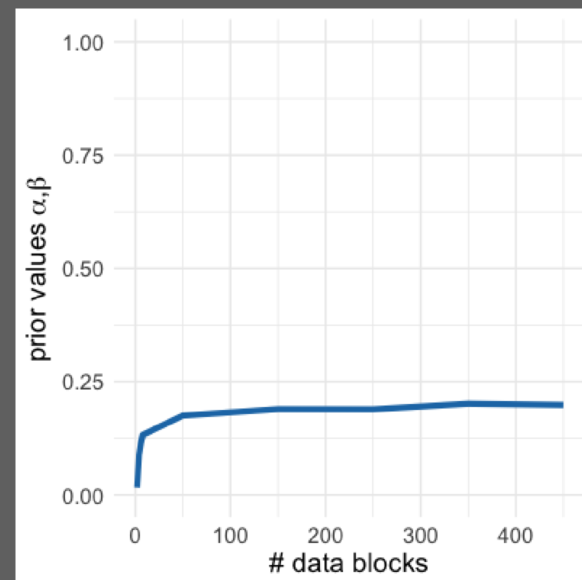
Odds ratio/ relative risk

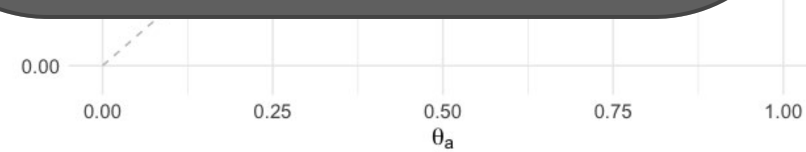


Learn from data fo
evidence collect

- With safe statistics can collect **"statistically sound" evidenc for A/B testing (e.g. treatment recommendations!)** in real-time
- Learning approach: take safe test for next data block that is optimal relative to estimate $\hat{\theta}_1$ based on data seen so far
- Technical detail: take $\hat{\theta}_1$ as posterior mean for θ_1 with Beta prior on parameter space

Progress since last meeting: Beta priors that are "optimal" w.r.t. REGRET and data collected





Real-life example: SWEPIS perinatal death study

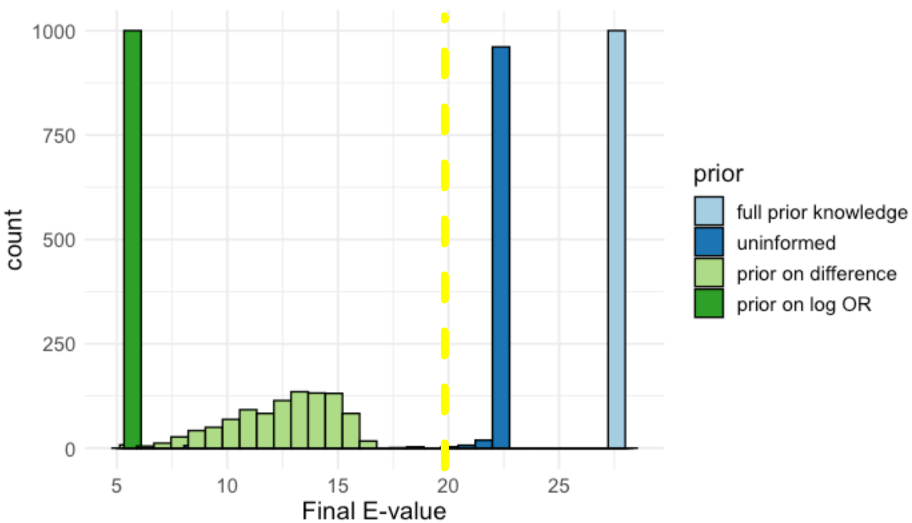
- Comparing perinatal death in labour induction at 41 or 42 weeks
- "All significance tests were two sided at the 0.05 level."

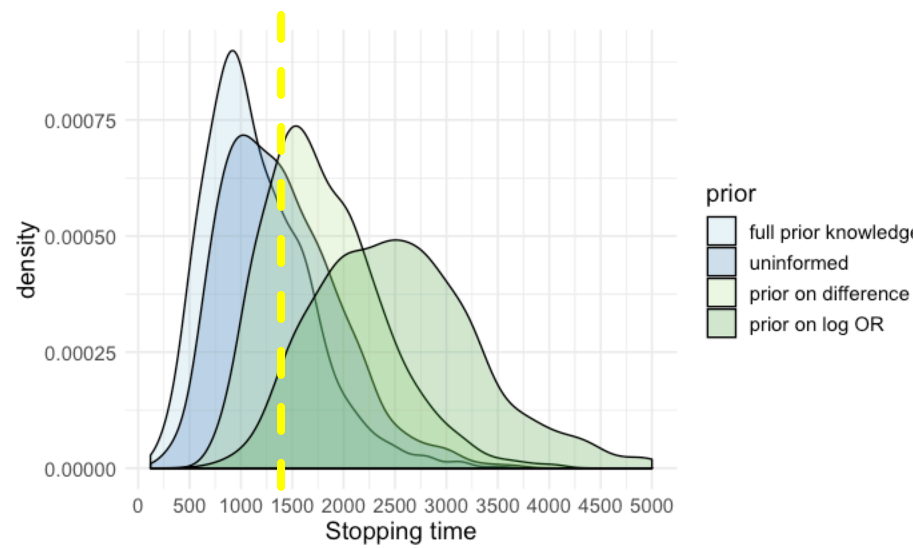
After observing 1380 births in each group:
"On 2 October 2018 the Data and Safety Monitoring Board strongly recommended the SWEPIS steering committee to **stop the study owing to a statistically significant higher perinatal mortality in the expectant management group**. Although perinatal mortality was a secondary outcome, it was not considered ethical to continue the study. **No perinatal deaths occurred in the early induction group but six occurred in the expectant management group (**five stillbirths and one early neonatal death; **P=0.03**)."

Safe testing applied to collect evidence in the SWEPIS scenario

Optionally: use knowledge from previous studies in prior of safe test

- Mean perinatal death rate at 41 weeks: 0.0001
- Difference risk between 42 and 41 weeks: 0.00318

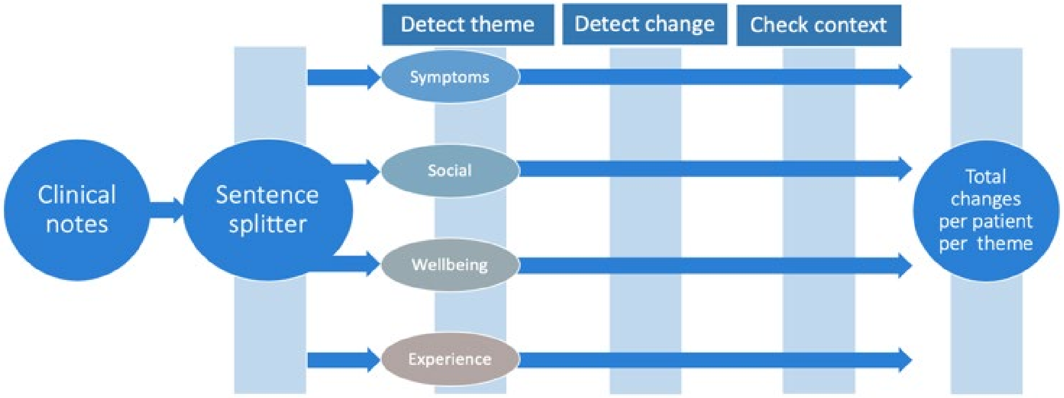




Application to use case: evaluating the usefullness of a recommender system for treatment of depression

Extracting information on the outcome of treatment trajectories from electronic health records in psychiatry

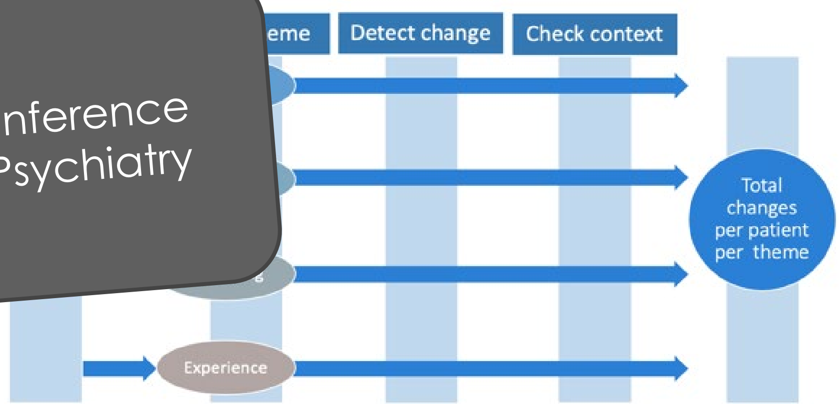
R.J. Turner[1,2], F. Coenen[2], K. Hagoort[2], F.E. Scheepers[2], P.D. Grünwald[1], and A. Härmä[3]
[1]CWI, Amsterdam, NL
[2]UMC Utrecht, Utrecht, NL
[3]Philips Research, Eindhoven, NL



Application to use case: evaluating the usefullness of a recommender system for treatment of depression

Extracting information on the outcome of treatment trajectories from electronic health records in psychiatry

R.J. Turner[1,2], F. Coenen[2], K. Hagoort[2], F.E. Scheepers[2], P.D. Grünwald[1], and A. Härmä[3]
[1]CWI, Amsterdam, NL
[2]UMC Utrecht, Utrecht, NL
[3]Philips Research, Eindhoven, NL



NLP part submitted to OCUPAI'21 conference
Full draft will be submitted to JAMA Psychiatry

Application to use case: evaluating the usefullness of a recommender system for treatment of depression

- Which types of recommendations assist clinicians the best?
- Plan: offer, in (micro-)randomized format, different forms of recommendations to clinicians based on the four outcome measures extracted from free text
  - ECT
  - Antidepressants
- Continuously analyze results with safe tests

ARTICLE Open Access

How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection

Maia Jacobs[1], Melanie F. Pradier[1], Thomas H. McCoy Jr.[2,3], Roy H. Perlis[2,3], Finale Doshi-Velez[1] and Krzysztof Z. Gajos[1]

EPI RQ4 Research Update:
Privacy Preserving Distributed Machine Learning

Saba Amiri

s.amiri@uva.nl

Supervisor: Adam Belloum

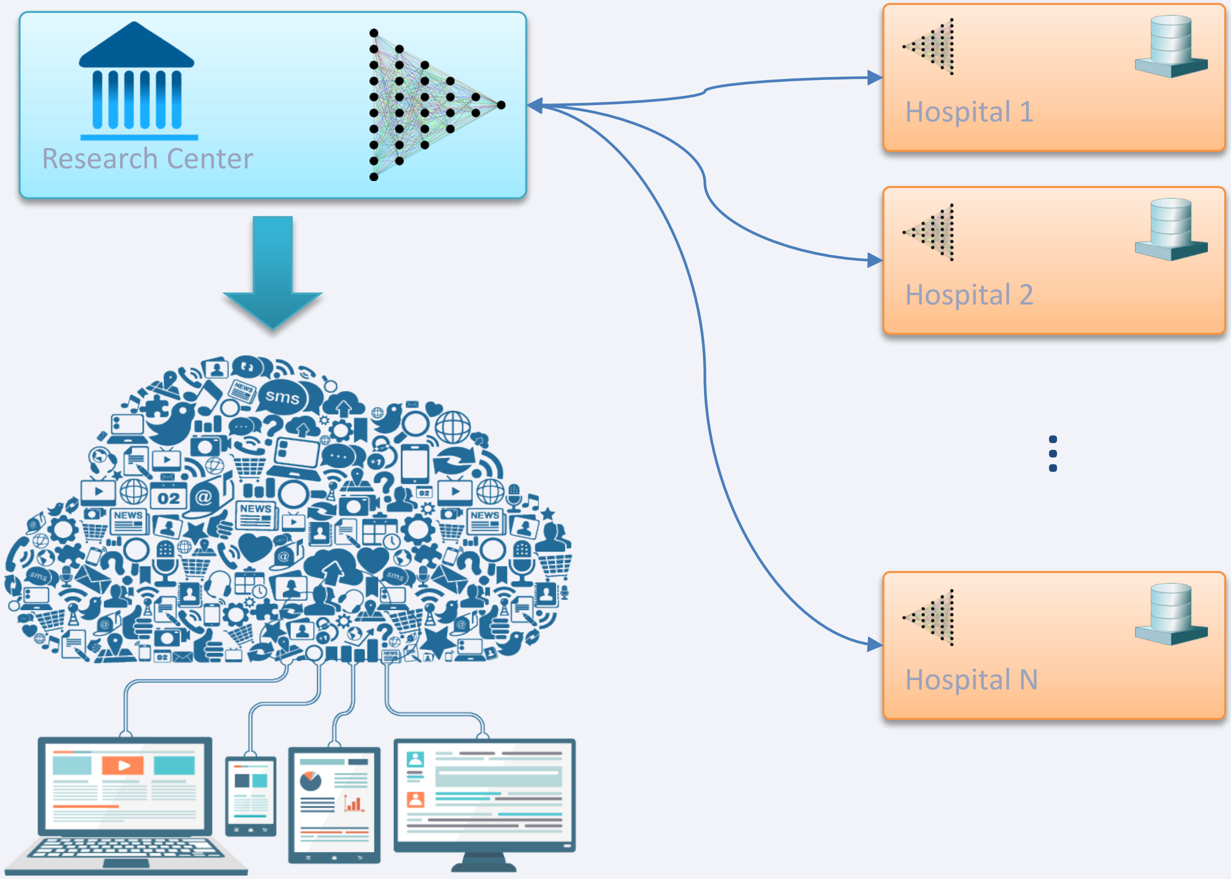
Promoters: Sander Klous, Leon Gommans

Multiscale Networked Systems Group

21 April 2021

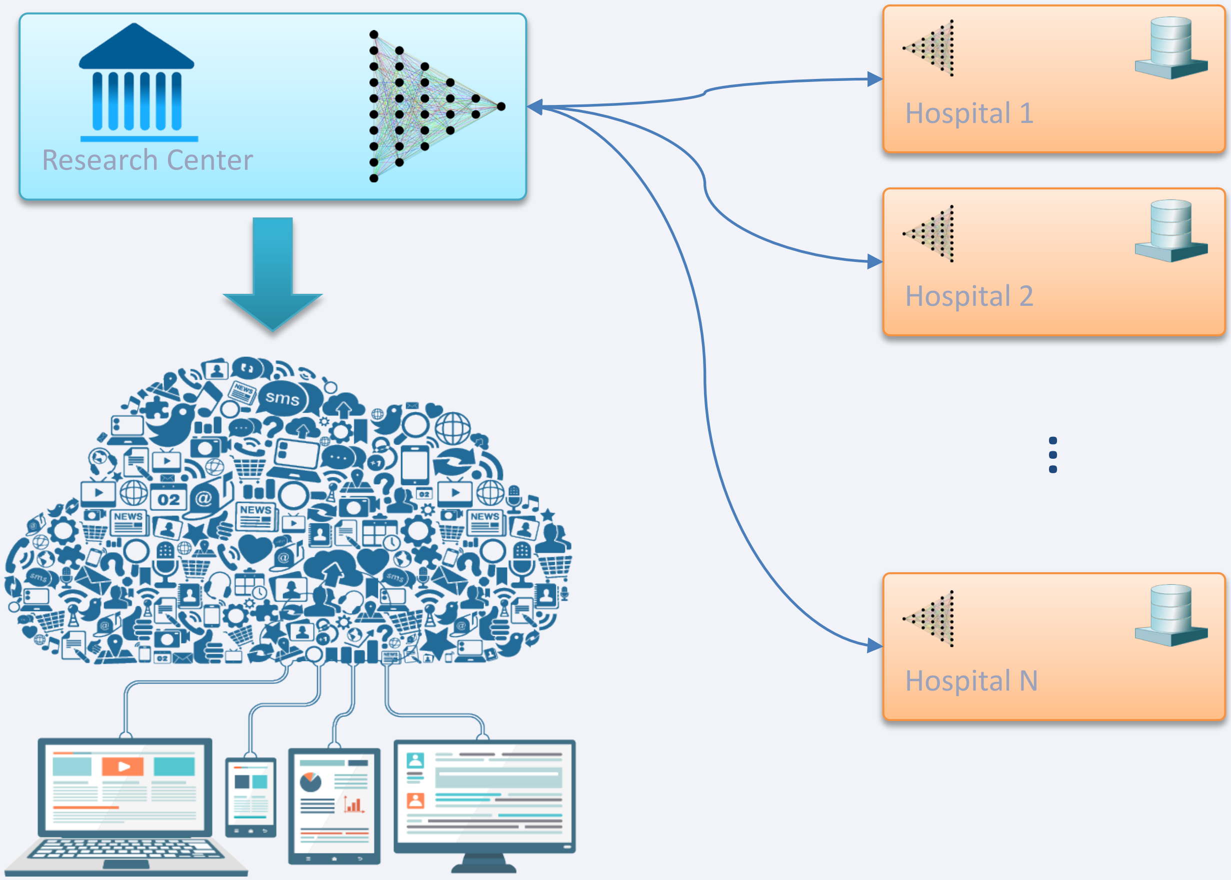
Research Domain

- Digital Health Twin



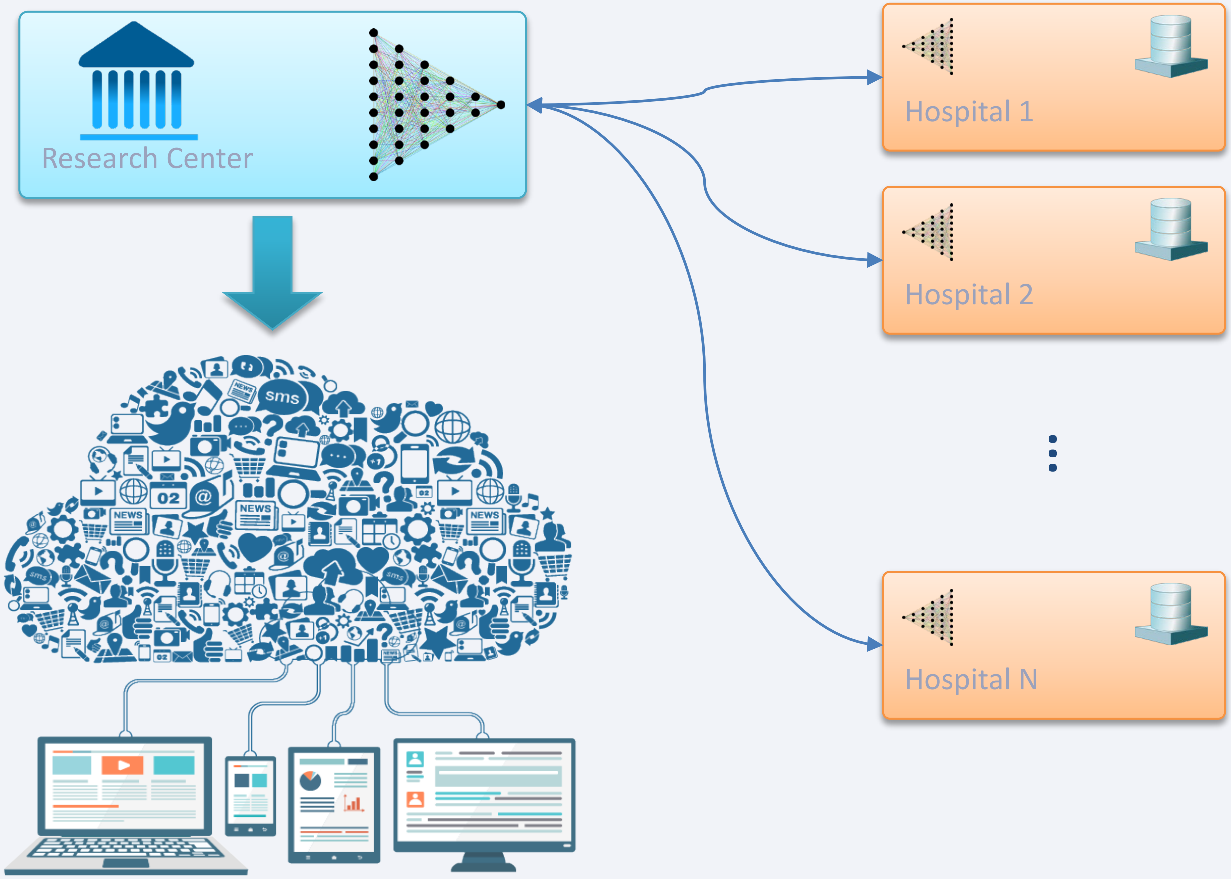
Research Domain

- Digital Health Twin
- Distributed Learning



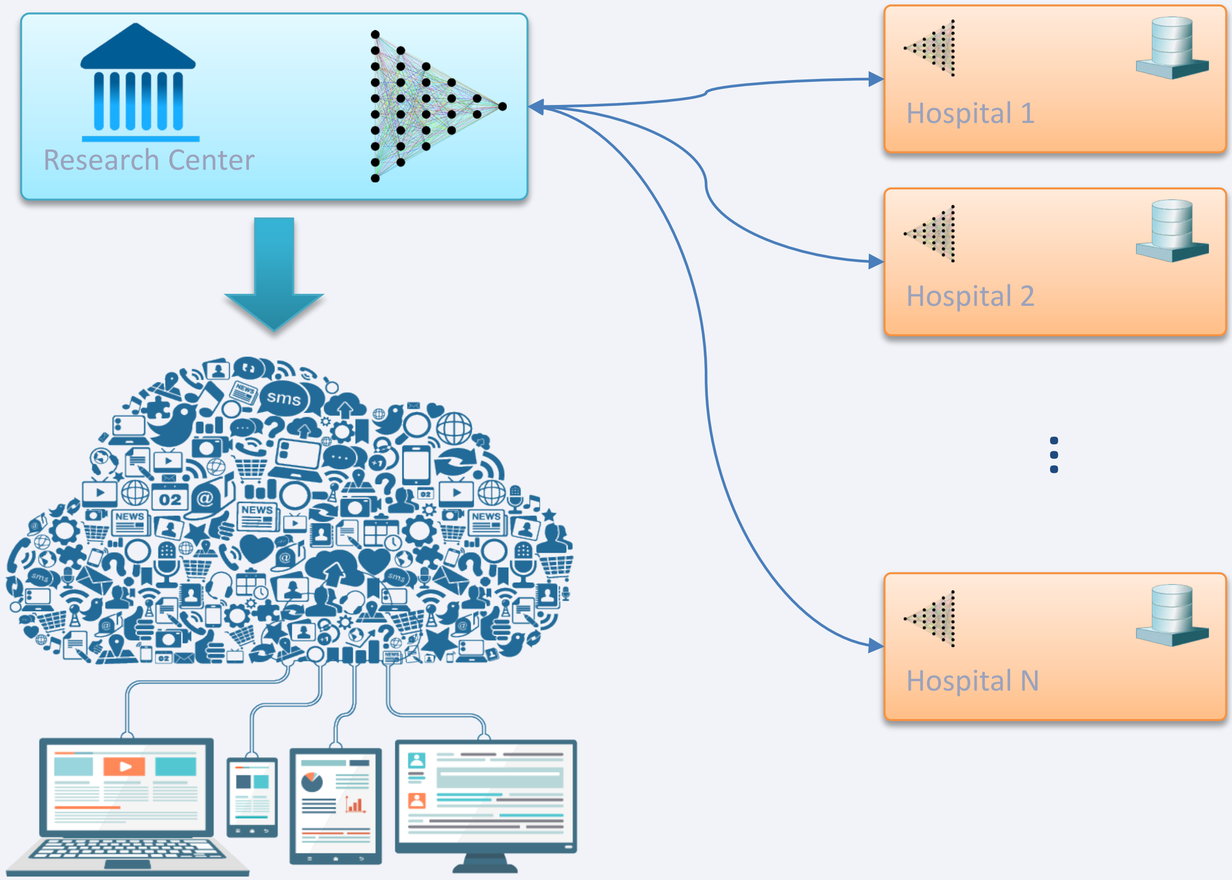
Research Domain

- Digital Health Twin
- Distributed Learning
- Privacy Preservation



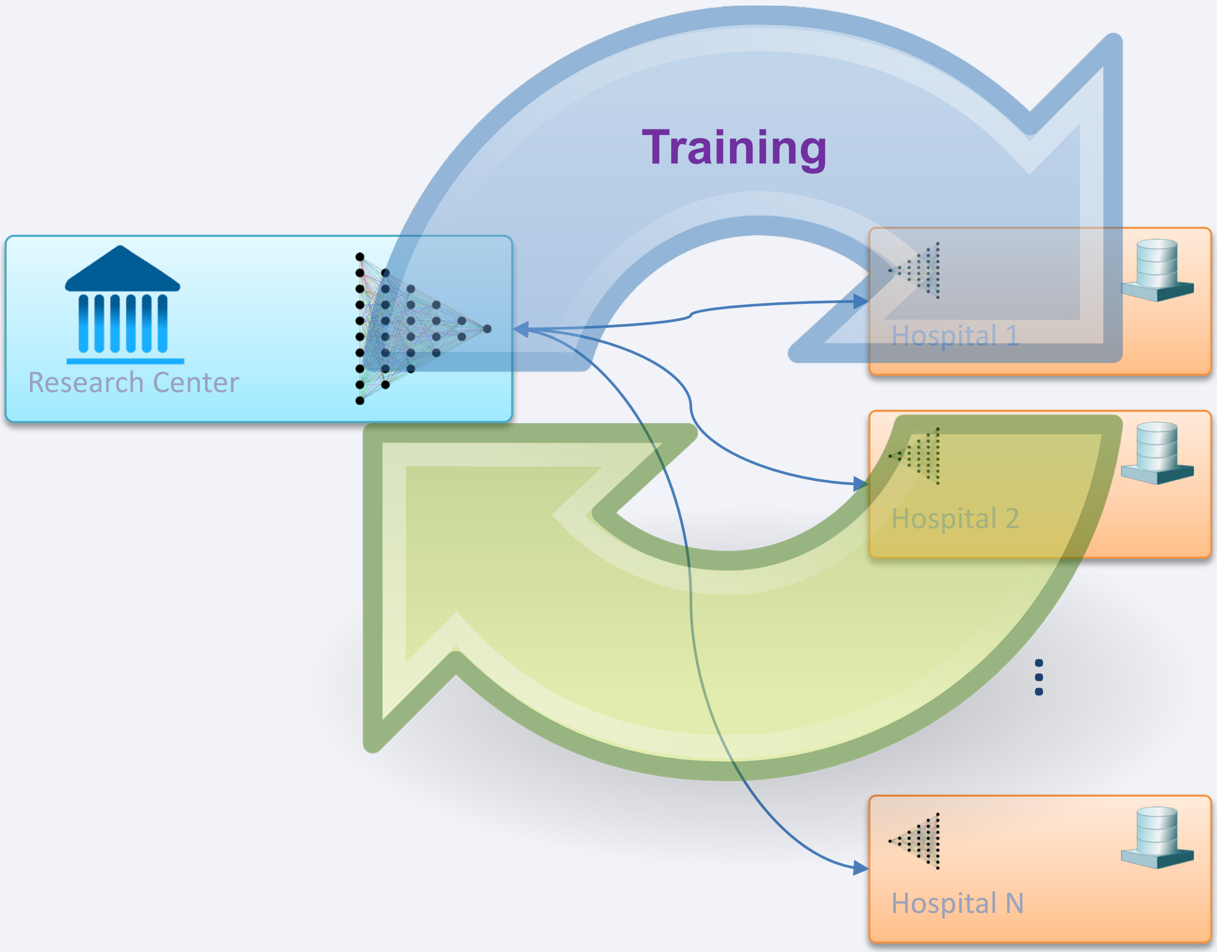
Definition of Privacy

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole



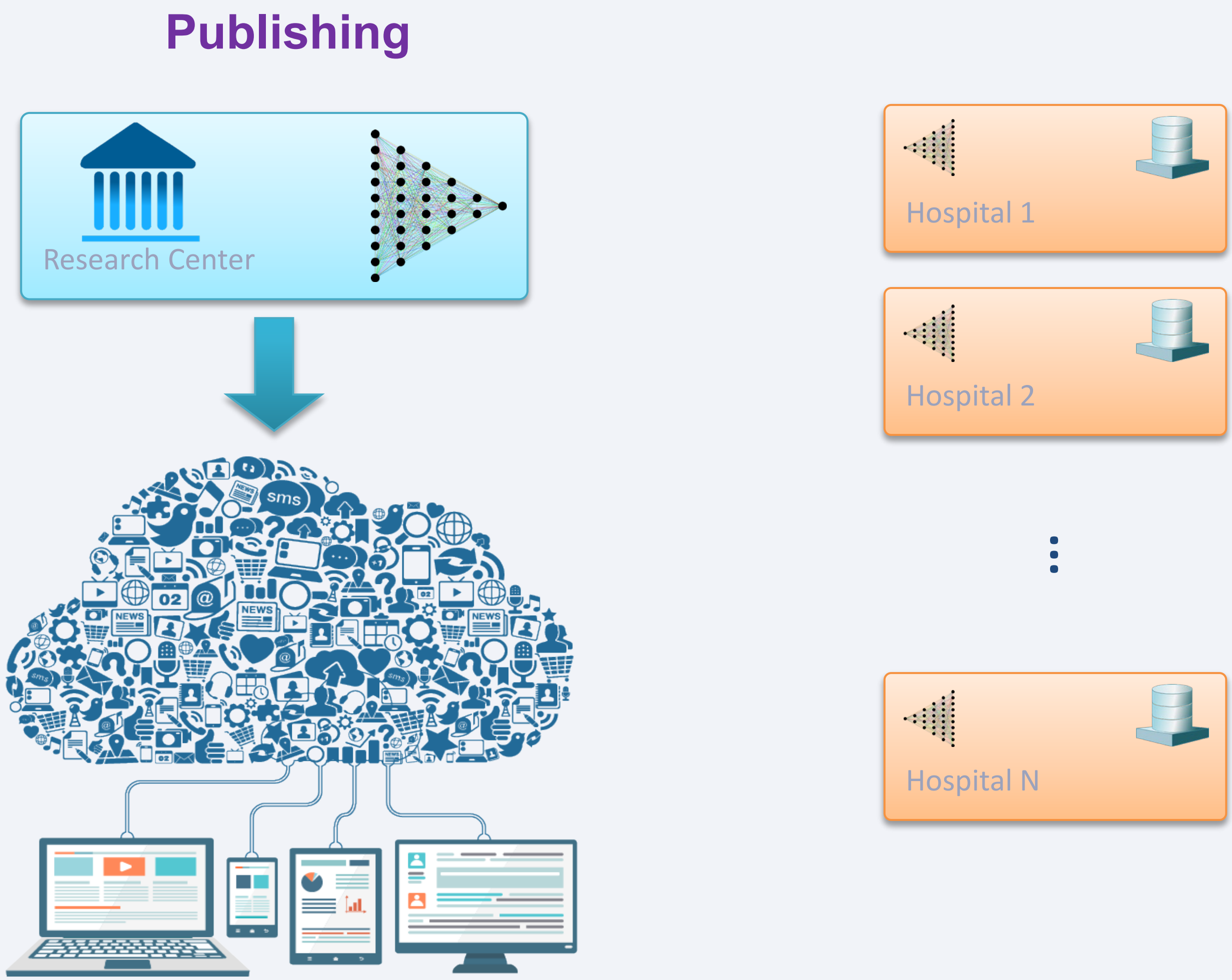
Typical Federated Learning Scenario

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole



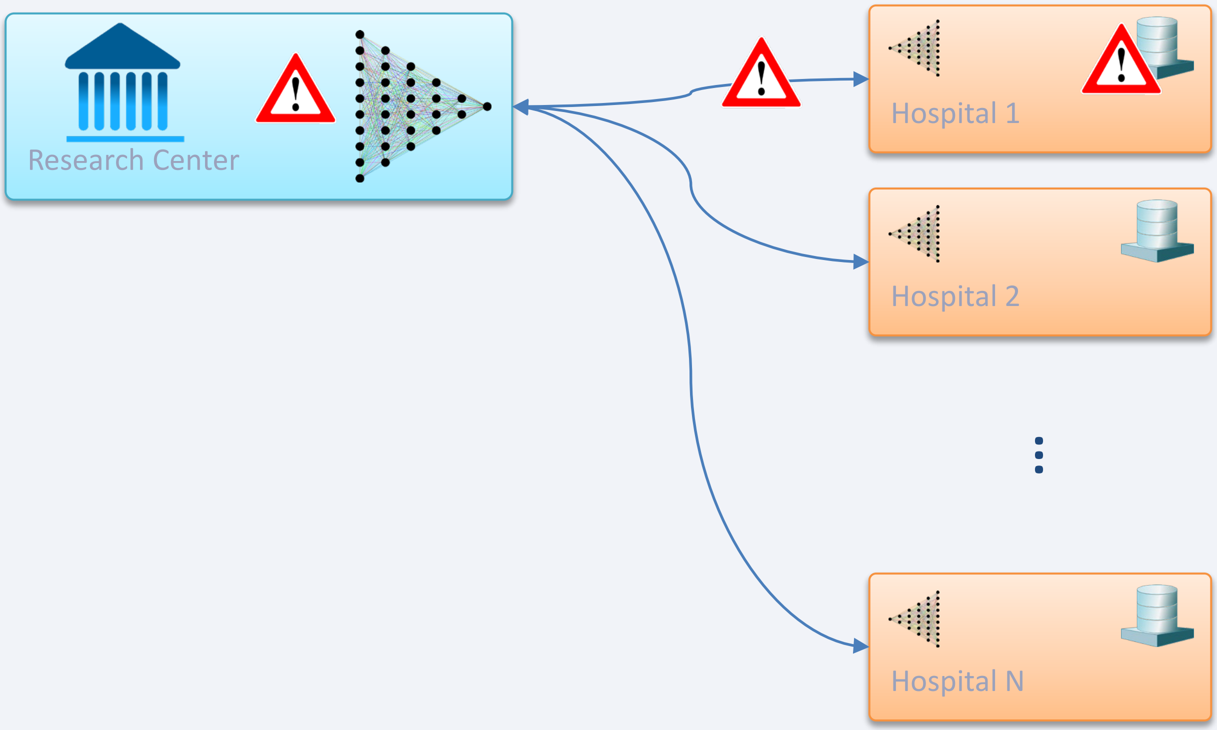
Typical Federated Learning Scenario - Publishing

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole



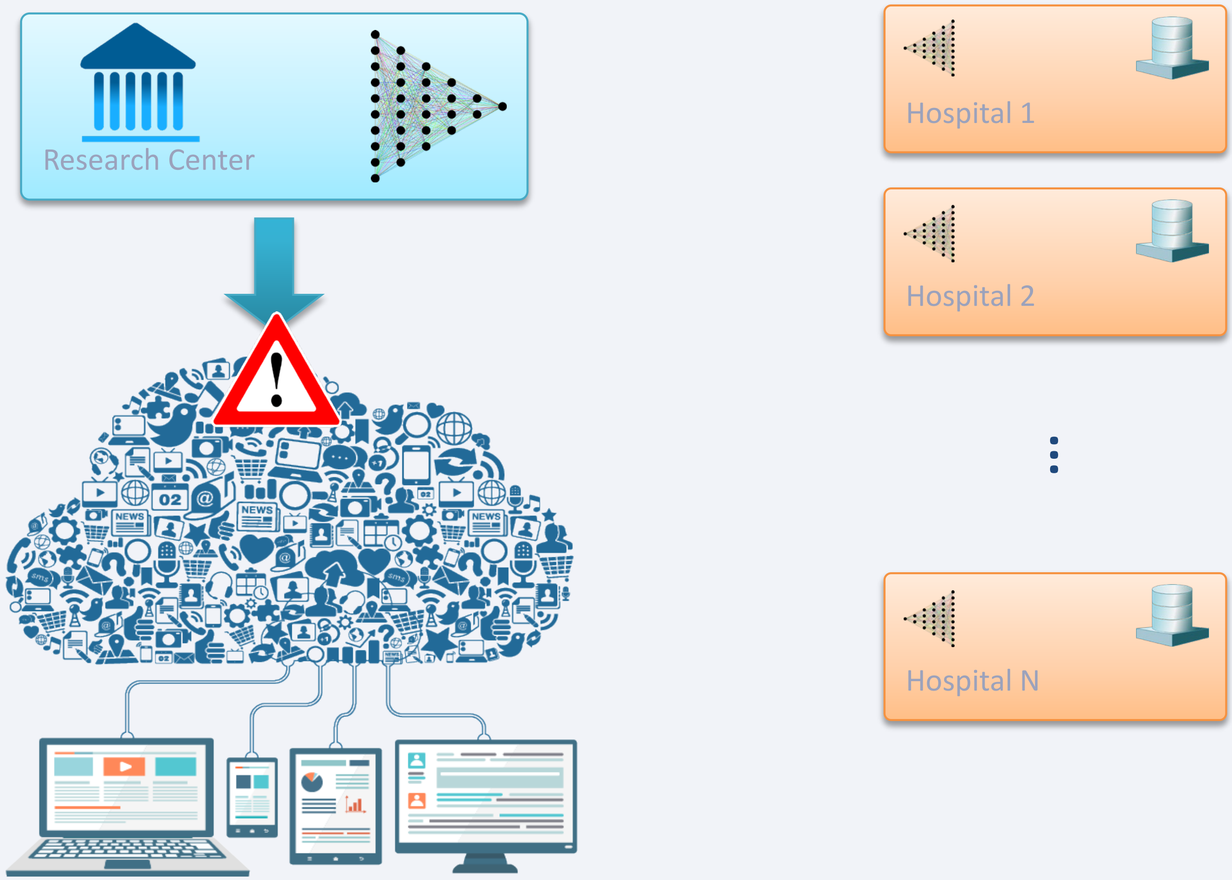
Typical Federated Learning Scenario – Training Risks

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - ☐ Data
    - ☐ Communication
    - ☐ Infrastructure



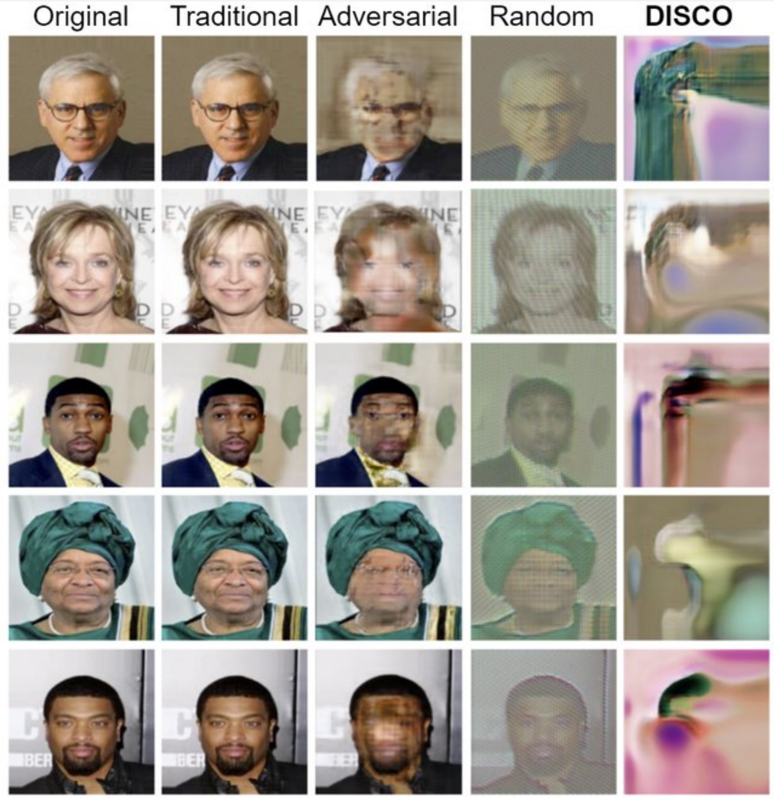
Typical Federated Learning Scenario – Publication Risks

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - Data
    - Communication
    - Infrastructure
    - **Machine learning model output**



The Need for Privacy Preserving Machine Learning

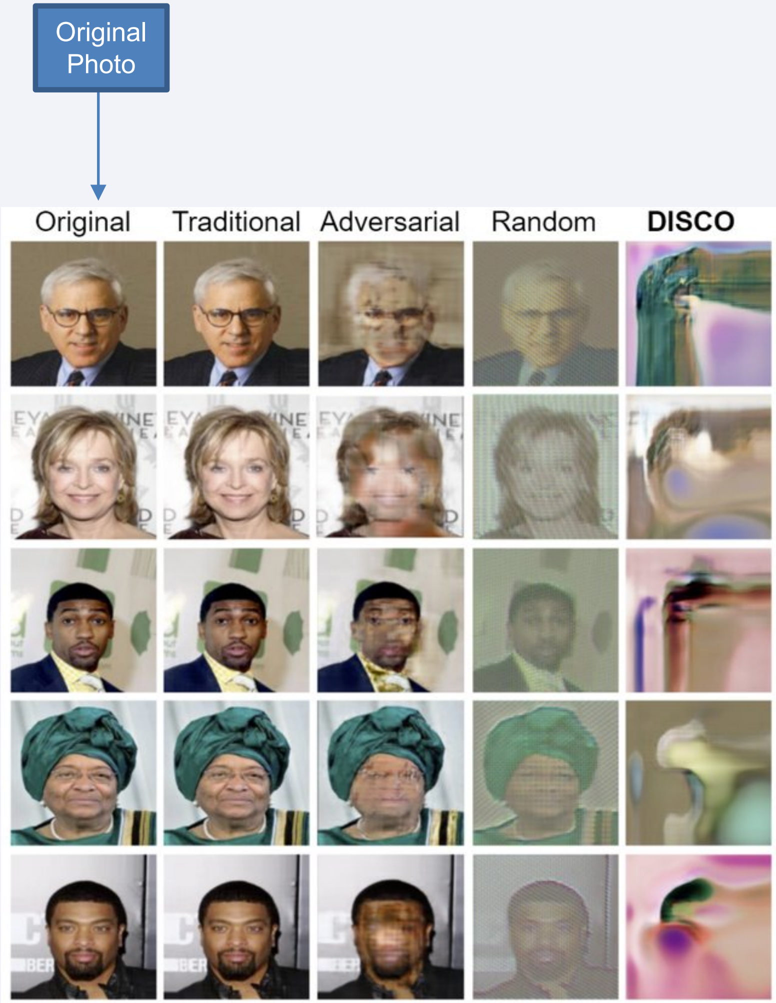
- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - ☐ Data
    - ☐ Communication
    - ☐ Infrastructure
    - ☐ **Machine learning model output**



Singh, Abhishek, et al. "DISCO: Dynamic and Invariant Sensitive Channel Obfuscation for deep neural networks." *arXiv preprint arXiv:2012.11025* (2020)

The Need for Privacy Preserving Machine Learning

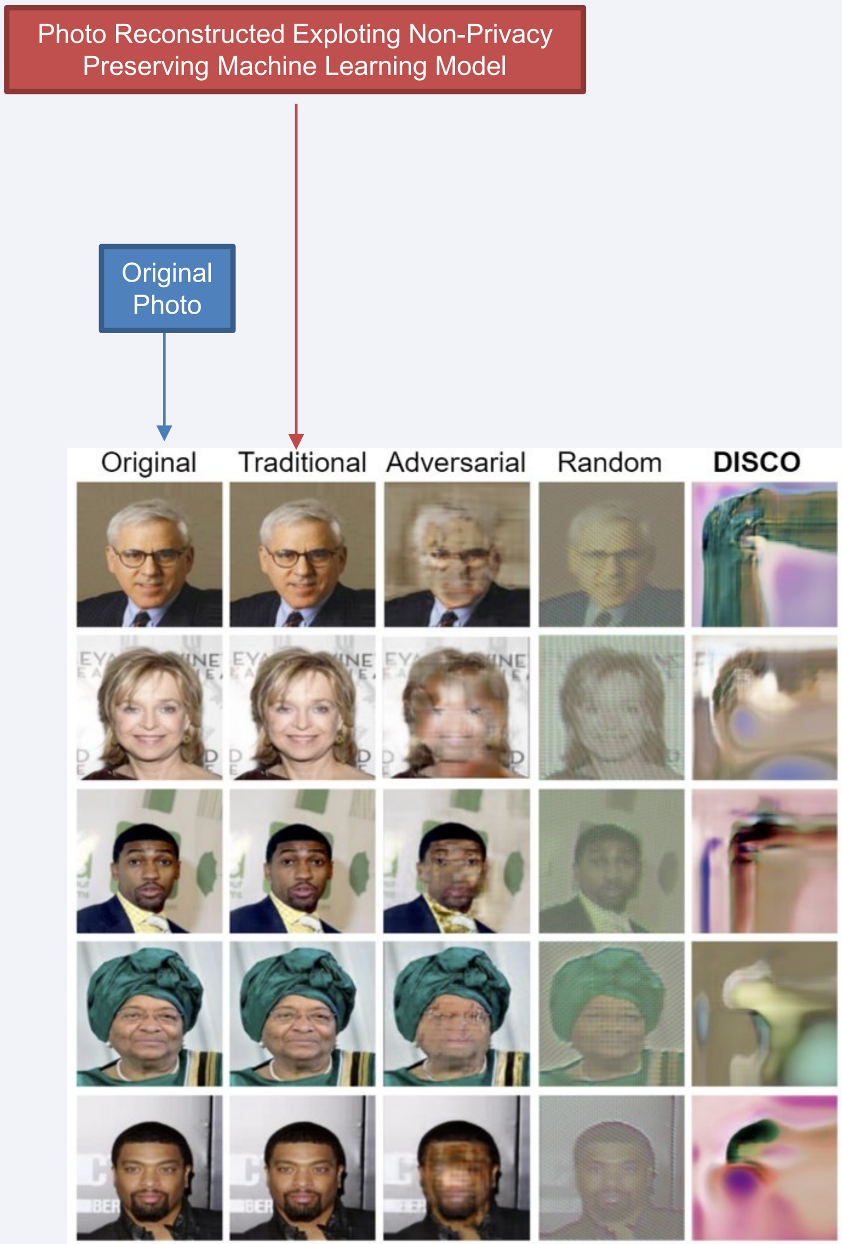
- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - ☐ Data
    - ☐ Communication
    - ☐ Infrastructure
    - ☐ **Machine learning model output**



Singh, Abhishek, et al. "DISCO: Dynamic and Invariant Sensitive Channel Obfuscation for deep neural networks." *arXiv preprint arXiv:2012.11025* (2020)

The Need for Privacy Preserving Machine Learning

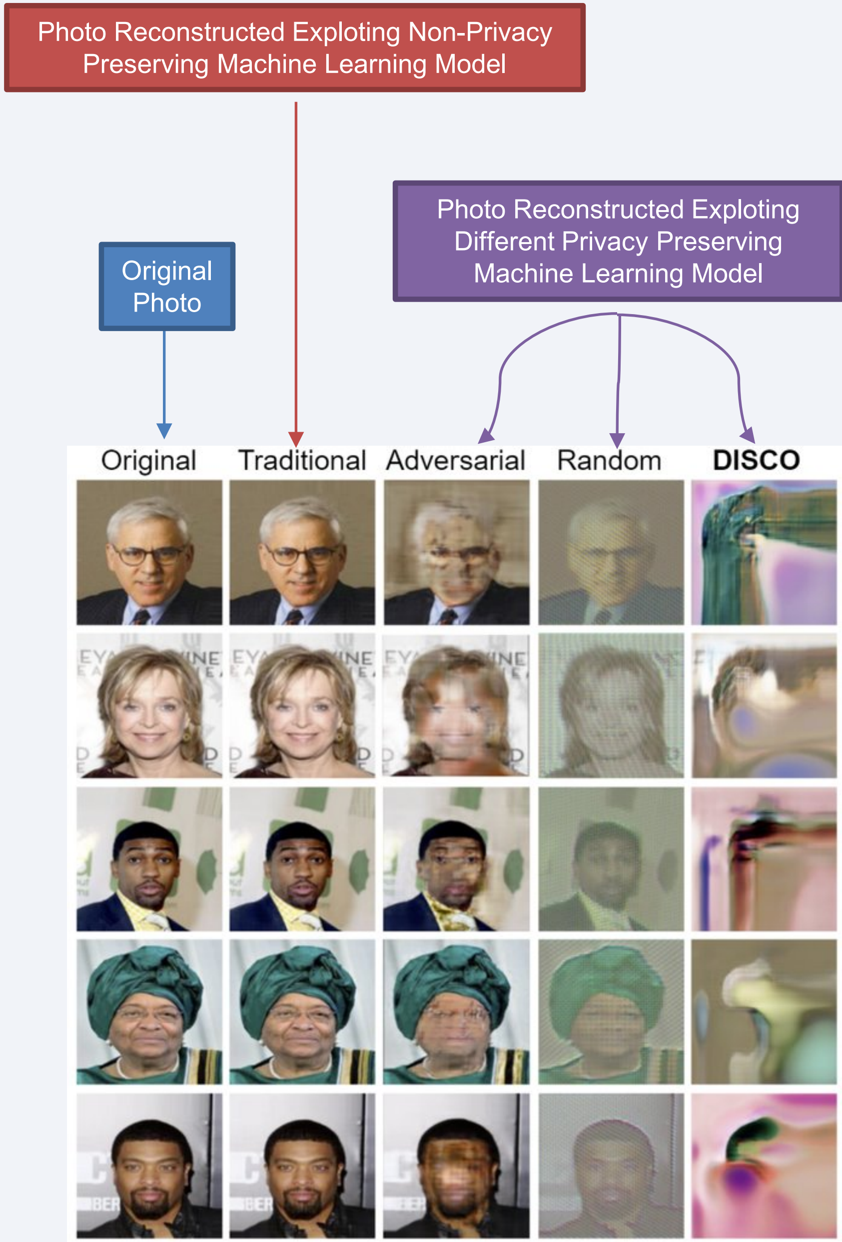
- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - Data
    - Communication
    - Infrastructure
    - **Machine learning model output**



Singh, Abhishek, et al. "DISCO: Dynamic and Invariant Sensitive Channel Obfuscation for deep neural networks." *arXiv preprint arXiv:2012.11025* (2020)

The Need for Privacy Preserving Machine Learning

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - Data
    - Communication
    - Infrastructure
    - **Machine learning model output**



Singh, Abhishek, et al. "DISCO: Dynamic and Invariant Sensitive Channel Obfuscation for deep neural networks." *arXiv preprint arXiv:2012.11025* (2020)

The Need for Privacy Preserving Machine Learning

- Digital Health Twin
- Distributed Learning
- Privacy Preservation
  - Definition: Providing record level protection to every member of the training set while gaining useful insights about the populations as a whole
  - What is not private?
    - Data
    - Communication
    - Infrastructure
    - **Machine learning model output**
  - Privacy preservation mechanism
    - Differential Privacy

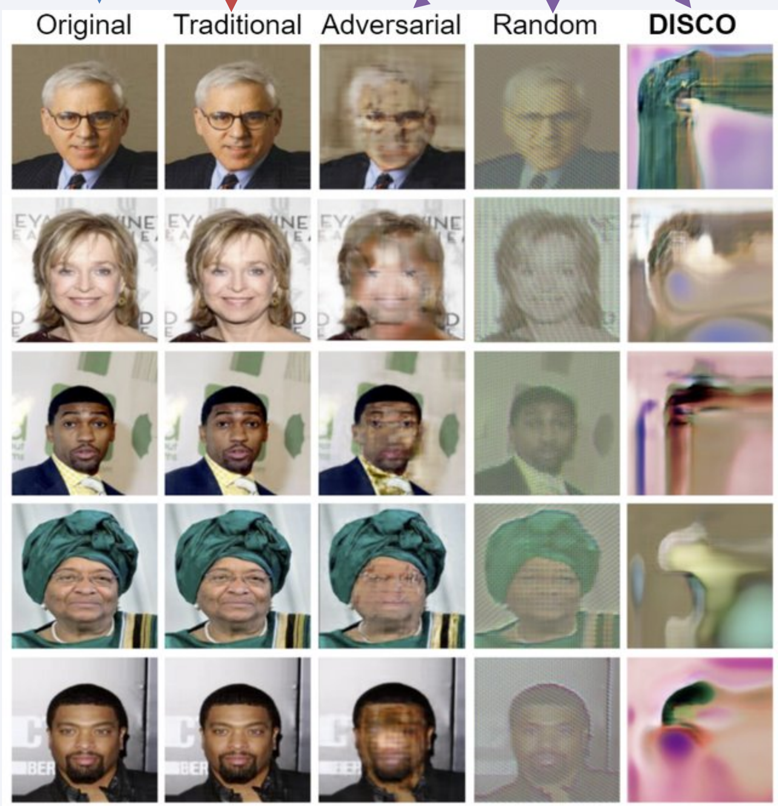
Satisfies our privacy definition

Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2014): 211-407.

Photo Reconstructed Exploting Non-Privacy Preserving Machine Learning Model

Original Photo

Photo Reconstructed Exploting Different Privacy Preserving Machine Learning Model



Singh, Abhishek, et al. "DISCO: Dynamic and Invariant Sensitive Channel Obfuscation for deep neural networks." *arXiv preprint arXiv:2012.11025* (2020)

Past and Present Activities

- Supervision of 3 B.Sc. Theses (concluded)[1]
- Supervision of 3 M.Sc. Literature reviews (concluded) [1]
- Short paper on *local differentially private federated learning through compression* (PPAI@AAAI-21) [2]
- Review paper on *differentially private synthetic data generation* submitted (pending editorial decision) [1]
- Review paper on *privacy attacks against machine learning* systems (receiving internal feedback) [1]
- Supervision of 4 M.Sc. AI, Computer science and data science theses (underway)
- Research on *local and global differentially private federated learning through compression* (experiments underway, paper being prepared)
- Research on DP distributed synthetic data generation (underway)

[1] Reports and paper available upon request; Code will be published by August 2021 depending on permission from consortium

[2] https://ppai21.github.io/files/29-paper.pdf

Lessons learned

ML system outputs are vulnerable

Current methods adequate?

Data anonymization is not enough

What is the solution?

Privacy preserving machine learning

Which privacy preserving mechanism?

Differential Privacy

What are the costs?

Added complexity
- Privacy design
- Utility/privacy trade-off

Any independent factors?

Non-i.i.d data
- Class imbalance
- Imbalanced distribution among hospitals

And if all these problems are fixed?

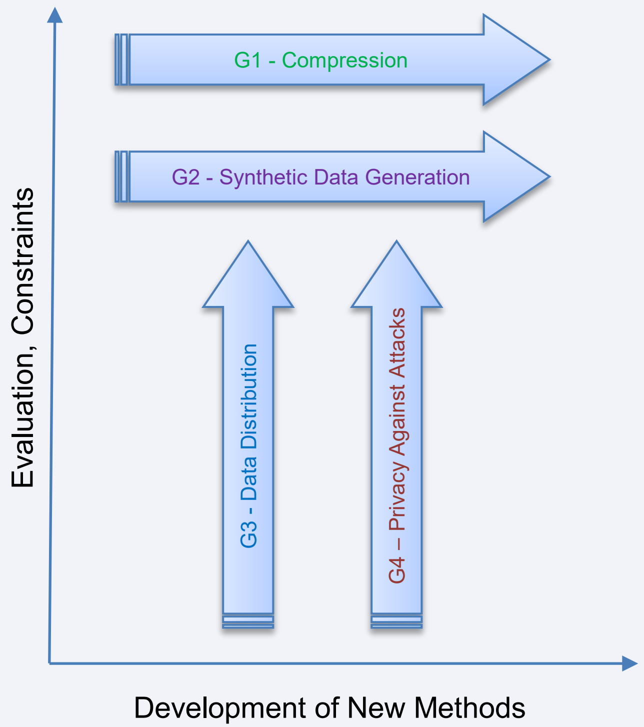
Definitions of privacy are abstract

How to use it in real world?

Deployment and governance mechanism needed

Research Goals

- G1 - Achieve Differential Privacy Through Compression?
- G2 - Generate differentially-private synthetic tabular data in a distributed setting
- G3 - Analyze the effect of non-i.i.d data on the performance of differentially private machine learning models
- G4 - Measure the privacy level of DP machine learning methods from the perspective of privacy attacks



Future Works

- [1] July 2021
  - Research on local/global compressive differentially private federated learning
  - Research on comparison of JAX framework against Pytorch for privacy preserving federated learning [1]
  - Research on privacy preserving federated learning on Vantage6 framework [2]
  - Output: paper; Code + experiments
- [2] September 2021
  - Research on distributed DP synthetic data generation using VAEs
  - Research on distributed DP synthetic data generation using GANs
  - Output: paper; Code + experiments
- [3] October 2021
  - Research on effect of non-i.i.d data in privacy preserving and non-privacy preserving federated learning
  - Output: paper; Code + experiments
- [4] April 2021 Q4, 2022 Q1
  - Research on extension of [2]
- [5] 2022 Q2
  - Utilization of the results of [3] in [2], [4]
- [6] 2022 Q2, Q3
  - Research on privacy analysis by measuring resiliency against privacy attacks

[1]JAX developed by Google (See https://opensource.google/projects/jax), Pytorch developed by Facebook (See https://pytorch.org/)
[2] Developed by Maastricht University Medical Centre+, Netherlands Comprehensive Cancer Organization (See https://vantage6.ai/vantage6)

Thank you!

My direct collaborators in chronological order

- Serge van Haag
- Boris Egelie
- Tidi Stamatiou
- Carlijn Nijhuis
- Mike Schouw
- Jetske Beks
- Willemijn Beks
- Yu Wang
- Wendy Cheung
- Simon Tokloth

RQ5: Automating normative control for Healthcare research
eFlint specification for regulatory documents in DIPG-research

Milen G. Kebede

Informatics Institute, University of Amsterdam
m.g.kebede@uva.nl

April 22, 2021

Supervisors : Prof. Tom van Engers, Dr. Thomas van Binsbergen, Dr. Dannis van Vuurden

Use-case DIPG Registry

- Diffuse Intrinsic Pontine Gliomas(DIPG) registry: rare disease repository that allows researchers to access patient data that can lead to discovering new treatment and prognosis factors.

1. **Definitions**

1) In this Regulatory Document the following terms have the meanings ascribed to them below:
   a) Access: Access to certain Datasets in accordance with Section 5.
   b) Coded: processed through pseudonymisation pursuant to the GDPR i) by or on behalf of the Member making available Data to the DIPG Registry and ii) by or on behalf of the DIPG Network making available a Dataset to a Researcher.
   c) Data: The information collected from Donors that is transferred to and stored in the DIPG Registry in Coded form.
   d) Dataset: the Data from the DIPG Registry made available for the purpose of a Project.
   e) Donor: Any individual who's Data are transferred to the DIPG Registry in compliance with the terms and conditions of this Regulation.
   f) Dutch Childhood Oncology Group (DCOG): Dutch foundation and National Paediatric Haematology-Oncology Society, in which all Dutch paediatricians and other professionals specialized in research and treatment of children with cancer (paediatric oncology) are organized. DCOG will have tasks related to the DIPG Registry as outlined under 3.
   g) Executive Committee: The board that in accordance with the DIPG Network Bylaws has the authority to make decisions with respect to:
      - Access for a Project; and

7) Access to Data shall at least be conditional to the following:
   a) The Researcher shall be responsible for obtaining the permits and approvals necessary in its own country and the policies of its institution.
   b) The Researcher shall use the Data for the approved Project only. In case of deviations or changes in the project the Executive Committee shall have the right to terminate Access without any liability at its sole discretion
   c) The Researchers shall bear sole responsibility for the handling and use of the Data in accordance with applicable law and legislation
   d) The Researcher shall not duplicate the Data or have them duplicated
   e) The Researcher shall not disclose or provide access to the Data to any third party without the prior written consent of Executive Committee.
   f) The Researcher shall report the progress of the Project and Findings in a frequency as outlined in the Letter of Approval
   g) Findings must be shared with the research community at large and therefore be scientifically published in accordance with the Terms and Conditions.

GDPR Concepts Art.4

GDPR concepts(Art.4)

(7) 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;

(8) 'processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

(11) 'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

```
Fact subject
Fact data
Fact purpose
Fact subject-of Identified by subject * data

Fact controller
Fact processor
Fact processes Identified by
  processor * data * controller * purpose

Fact consent Identified by
  subject * controller * purpose
Fact accurate-for-purpose Identified by
  data * purpose
```

DIPG Reg. terms

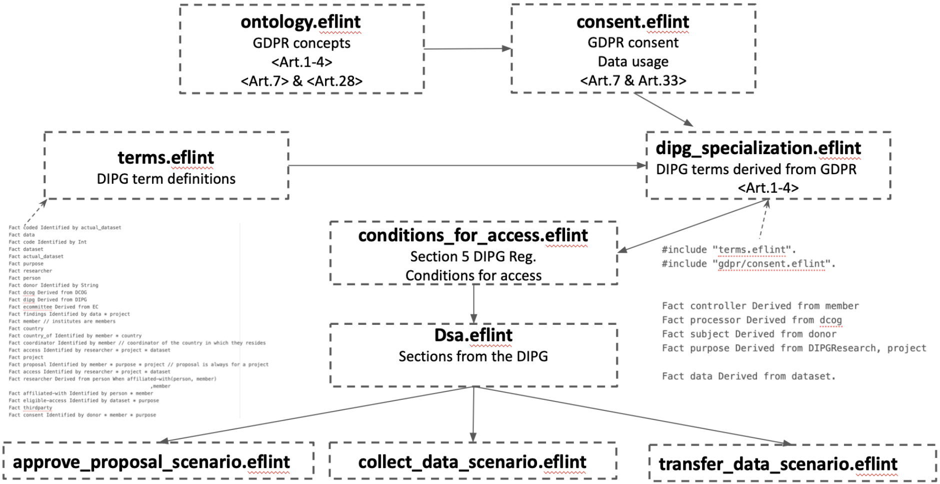
5) The Researcher shall be considered a controller under the GDPR in relation to the Dataset. The Letter of Approval shall include provisions imposing on the Researcher the obligations of the Researcher as controller under the GDPR.

DIPG Registry are performed exclusively on behalf of the DIPG Network Members. DCOG is considered a processor under the GDPR.

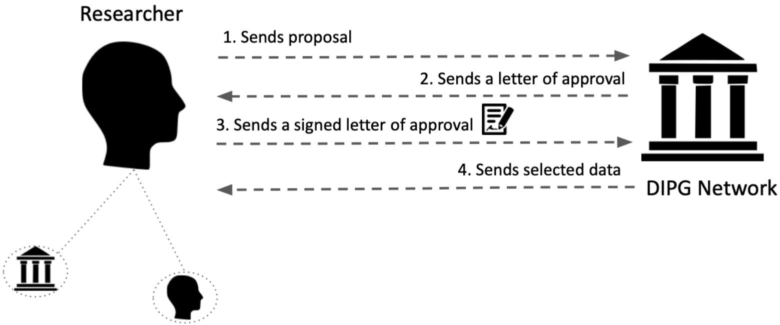
1) Members shall transfer into the DIPG Registry free of charge, Coded Data of patients that consented and are eligible to participate in the DIPG Registry. The Data in the DIPG Registry shall be used for further research purposes.

```
#include "terms.eflint".
#include "gdpr/consent.eflint".


Fact controller Derived from member
Fact processor Derived from dcog
Fact subject Derived from donor
Fact purpose Derived from DIPGResearch, project

Fact data Derived from dataset.
```

DIPG Regulatory articles



Accessing the registry



SIOP Europe DIPG Network

Policy specification with *eFLINT*

Send project proposal

```
Act propose-project
  Actor researcher
  Recipient ecommittee
  Related to project
  Creates proposal(researcher,project,
     project) When member(researcher)
         ,proposal(member, project,
     project)    When
     affiliated-with(researcher,member)
  Holds when researcher && ecommittee
```

Approve Project proposal

```
Act approve-project
  Actor ecommittee
  Recipient member
  Related to project
  Creates approved(project, member)
  Holds when ecommittee && member &&
     proposal(member, project, project)
Fact approved Identified by project *
     member
```

```
Act send-letter-of-approval
  Actor dipg
  Recipient member
  Related to project
  Creates letter-of-approval-sent(project, member)
  Holds when dipg && member && approved(project, member)
Fact letter-of-approval-sent Identified by project * member
Fact letter-of-approval-signed Identified by project *
     member
Act sign-letter-of-approval
  Actor researcher
  Recipient ecommittee
  Related to project
  Creates letter-of-approval-signed(project, researcher)
     When member(researcher)
         ,letter-of-approval-signed(project, member) When
     affiliated-with(researcher,member)
         ,duty-to-select-data(ecommittee, researcher,
     project) When member(researcher)
         ,duty-to-select-data(ecommittee, member, project)
     When affiliated-with(researcher,member)
         ,duty-to-send-data(ecommittee, researcher,
     project) When member(researcher)
         ,duty-to-send-data(ecommittee, member, project)
     When affiliated-with(researcher,member)
  Holds when
    researcher && ecommittee &&
     letter-of-approval-sent(project, researcher) &&
     member(researcher)
   ,researcher && ecommittee &&
     letter-of-approval-sent(project, member) &&
     affiliated-with(researcher, member)
```

Project approval scenario

```
+dataset(D1).
+member(STa).
+person(Eve).
+project(P1).
+affiliated-with(Eve,STa).
propose-project(Eve,EC,P1).
approve-project(EC,STa,P1).
send-letter-of-approval(DIPG,STa,P1).
sign-letter-of-approval(STa,EC,P1).
?duty-to-select-data(EC, STa, P1).
?duty-to-send-data(EC, STa, P1).
select-data(EC, STa, P1, D1).
?!duty-to-select-data(EC, STa, P1).
?duty-to-send-data(EC, STa, P1).
send-data(EC, STa, D1).
?!duty-to-send-data(EC, STa, P1).
```

Design Decisions

- The formalization that we are making matchs the formalization of the DIPG document
- Representing a member who can either be a researcher or an institution

```
        Fact researcher Derived from person When affiliated-with(person, member)
                                              ,member
```

- Signing letter of approval creates two duties

```
Duty duty-to-select-data
    Holder ecommittee
    Claimant member
    Related to project

Duty duty-to-send-data
  Holder ecommittee
  Claimant member
  Related to project
```

Future work

- Generic data sharing ontology
- Enforcing higher level policies
- Establish a connection with lower level policies

RQ5: Automating normative control for Healthcare research

eFlint specification for regulatory documents in DIPG-research

Milen G. Kebede

Informatics Institute, University of Amsterdam
m.g.kebede@uva.nl

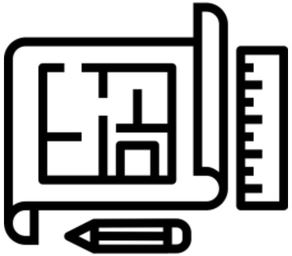
April 22, 2021

EPI Framework: A dynamic infrastructure to support health applications

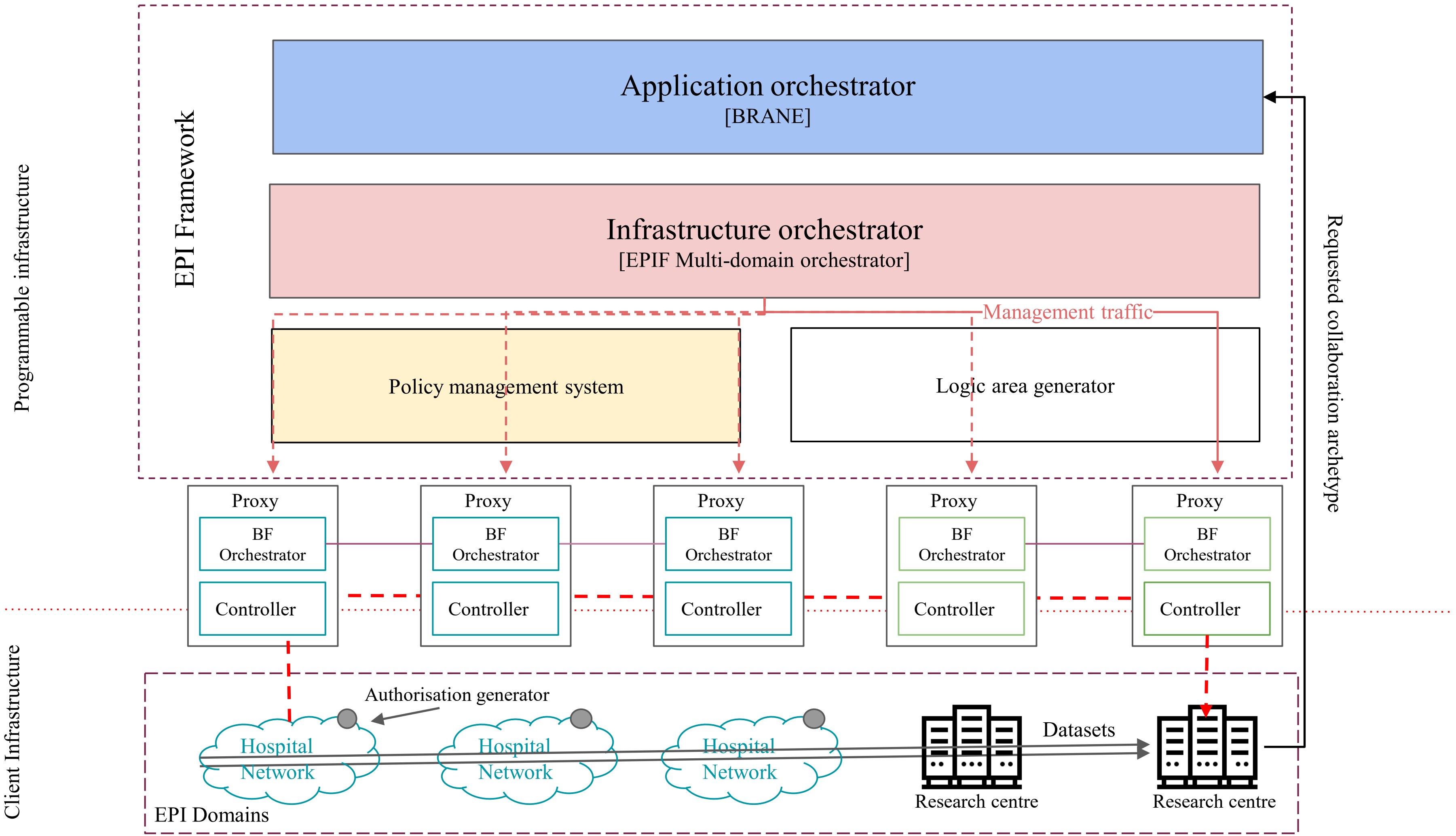
PhD: Jamila Alsayed Kassem
Daily supervisor: Dr. Paola Grosso
Co-supervisor: Dr. Axel Berg
Promotors: Prof. Dr. Cees de Laat
Prof. Dr. Anwar Osseyran

Since our last meeting

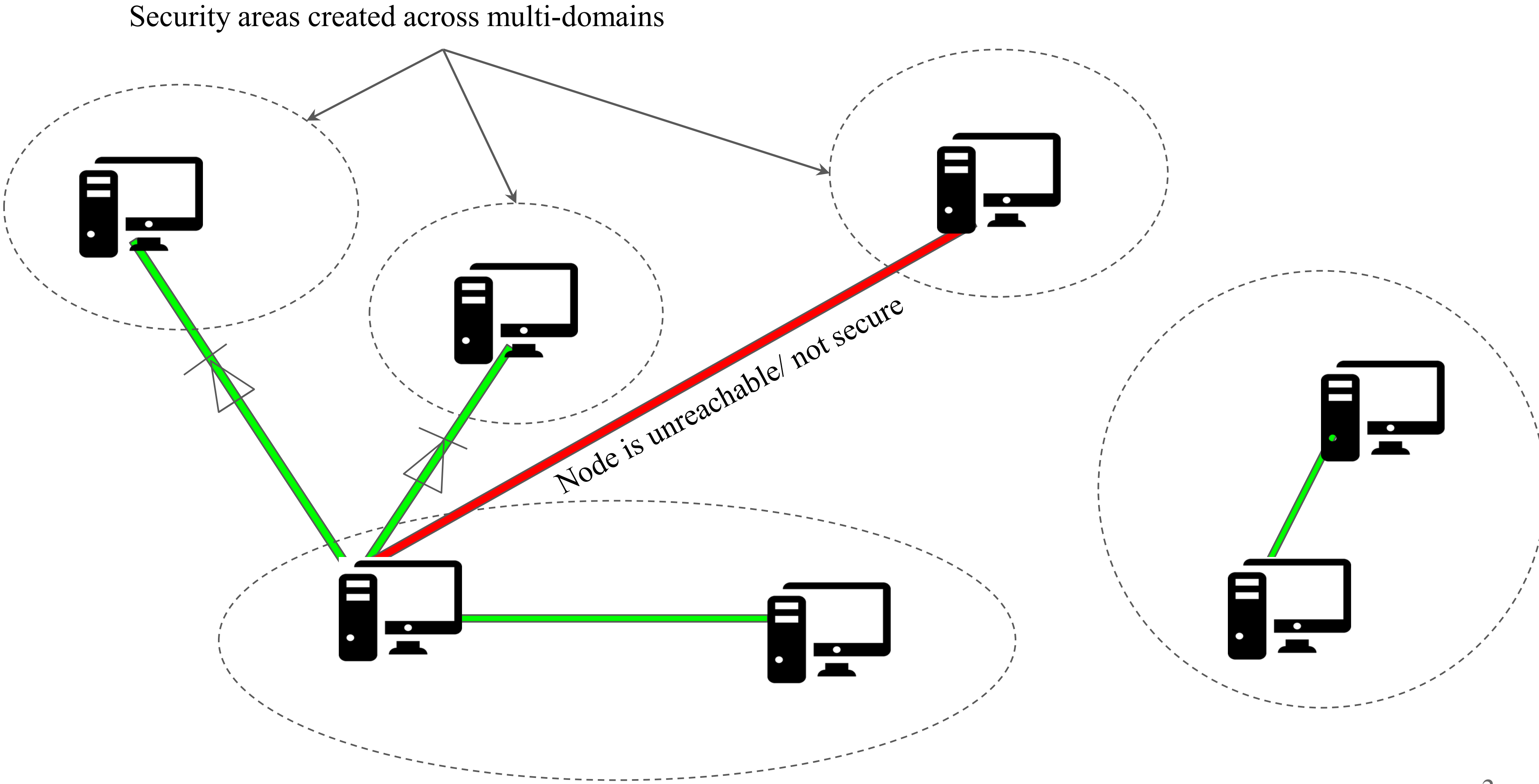
- Collaboration with BRANE to implement the framework
- Implementing a EPIF functionality
  - Redirection tools
  - Evaluating different parameters
- Paper submitted to eScience2021
- Experiment plan
  - More on redirection tools
  - Chaining BF
  - Security of bridges
- 3 Students supervision

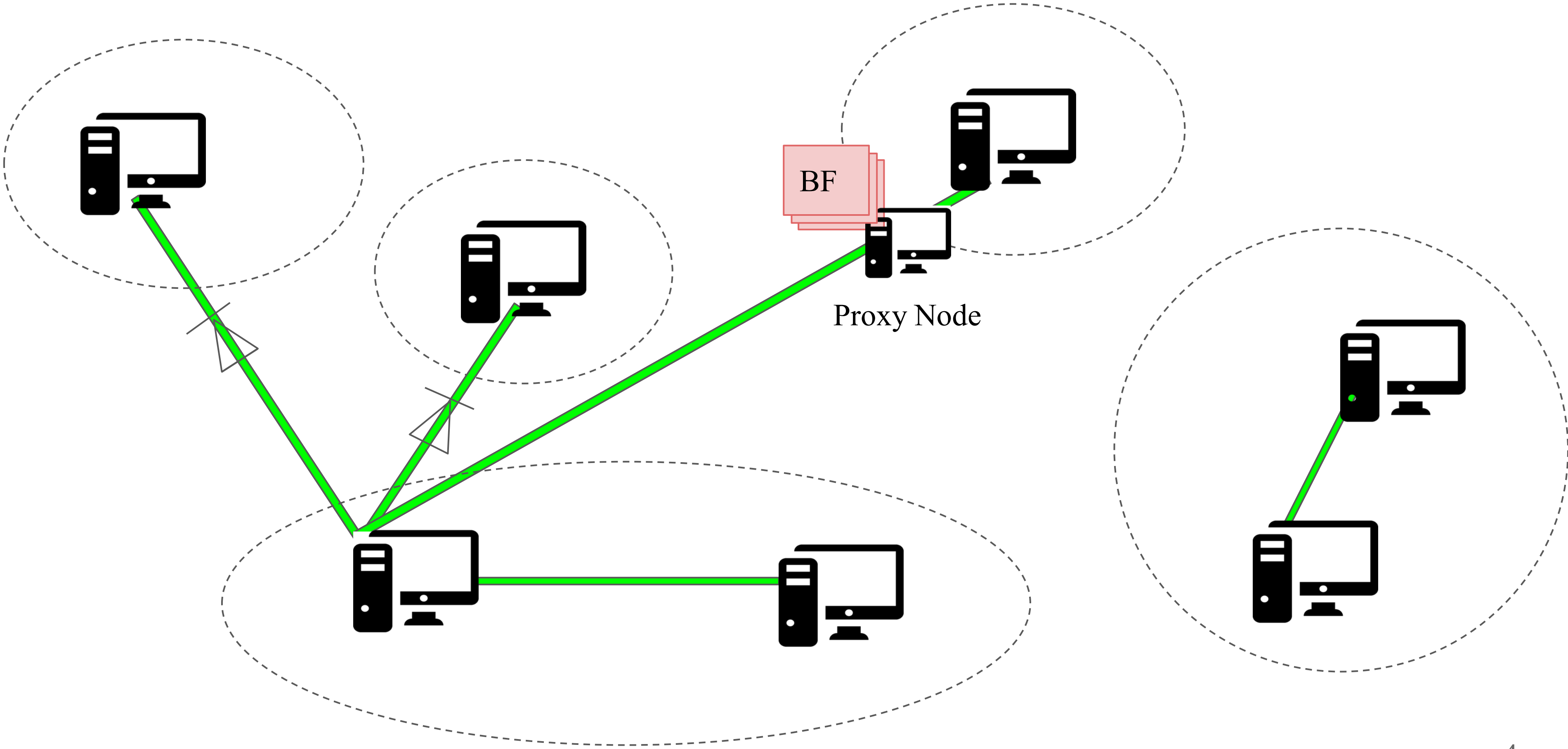


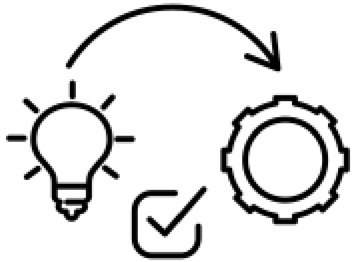
EPIF: The Architecture



Programmable infrastructure
EPI Framework
Application orchestrator
[BRANE]
Infrastructure orchestrator
[EPIF Multi-domain orchestrator]
Management traffic
Policy management system
Logic area generator
Requested collaboration archetype
Proxy
BF Orchestrator
Controller
Proxy
BF Orchestrator
Controller
Proxy
BF Orchestrator
Controller
Proxy
BF Orchestrator
Controller
Proxy
BF Orchestrator
Controller
Client Infrastructure
Authorisation generator
Hospital Network
Hospital Network
Hospital Network
Datasets
Research centre
Research centre
EPI Domains

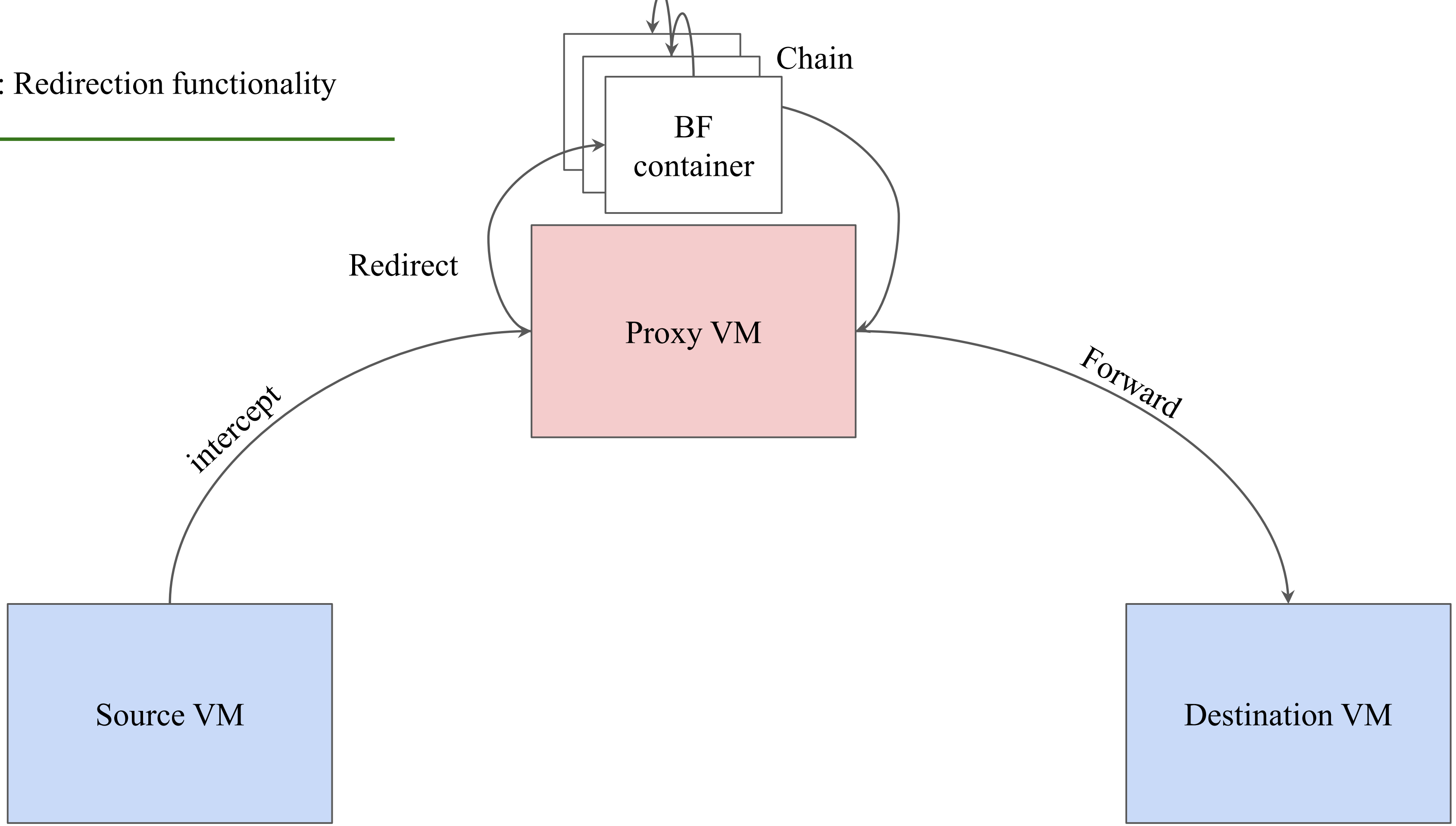

Security areas created across multi-domains
Node is unreachable/ not secure


BF
Proxy Node

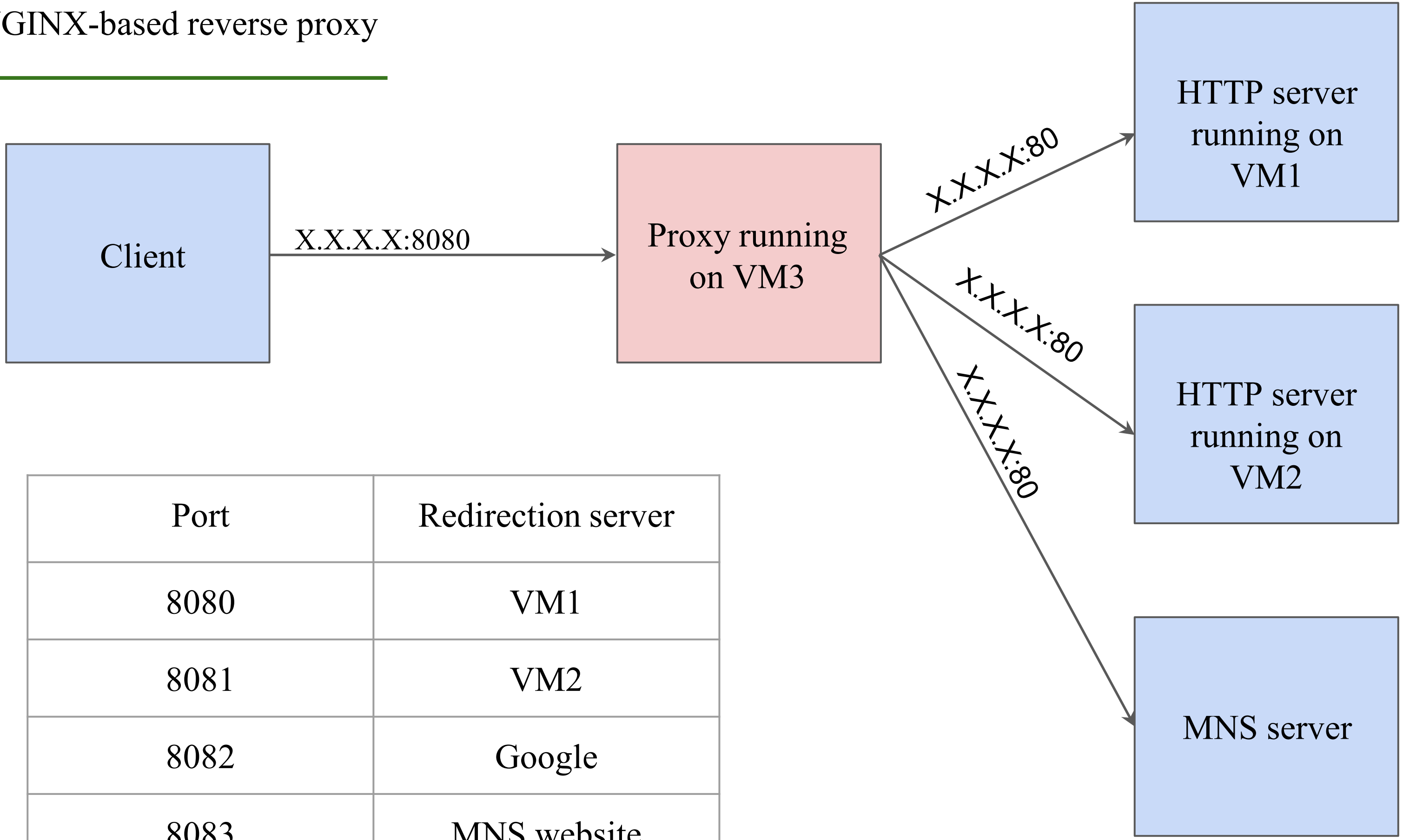


Proxy Implementations

EPIF: Redirection functionality

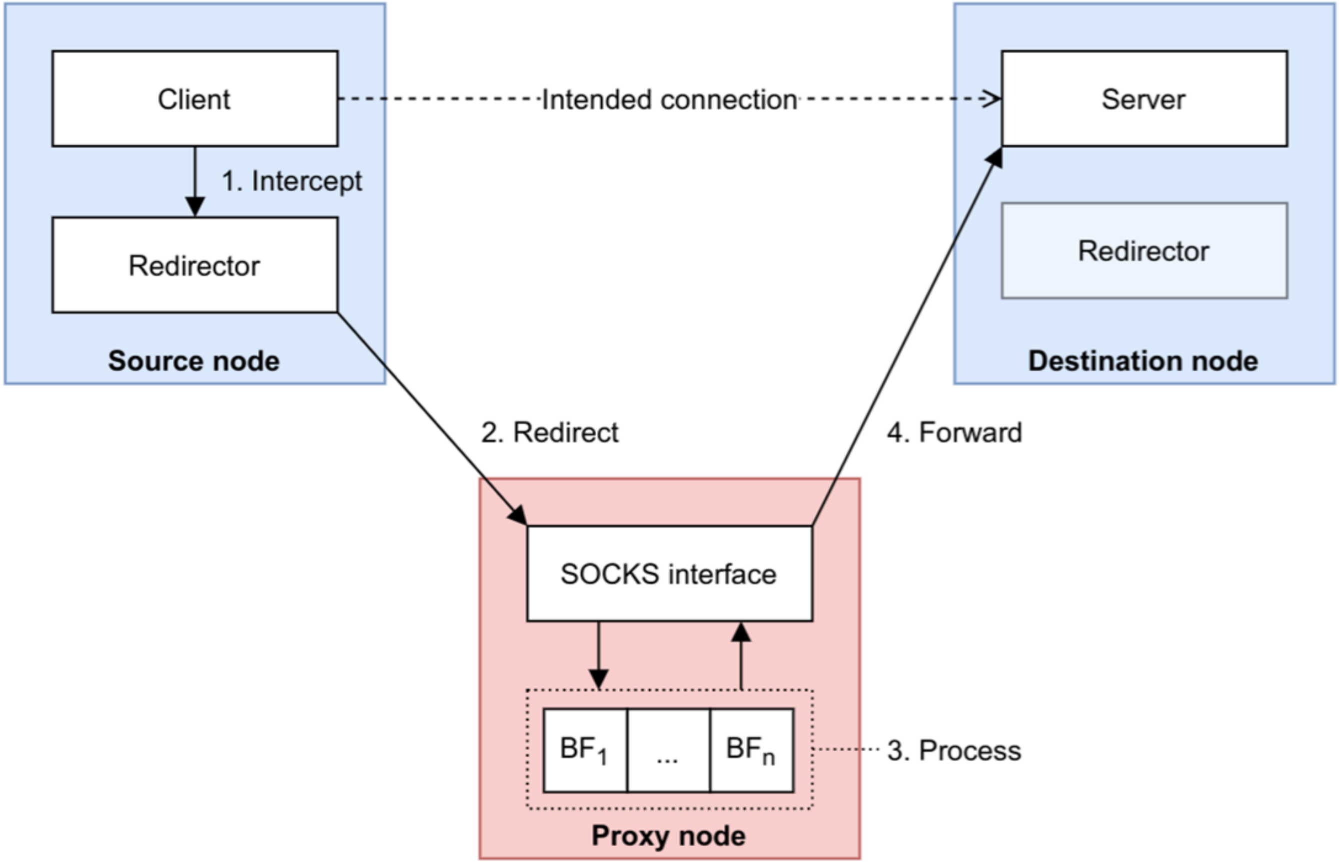


1. NGINX-based reverse proxy



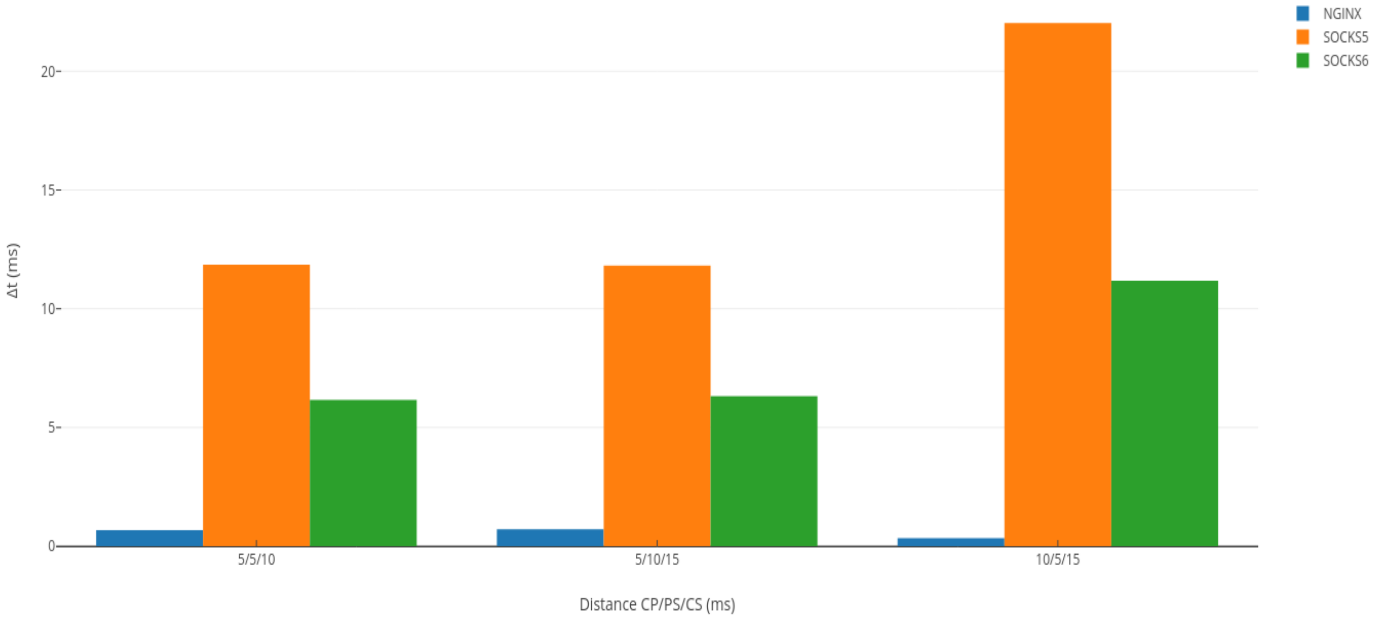
Port	Redirection server
8080	VM1
8081	VM2
8082	Google
8083	MNS website

2. SOCKS-based proxy

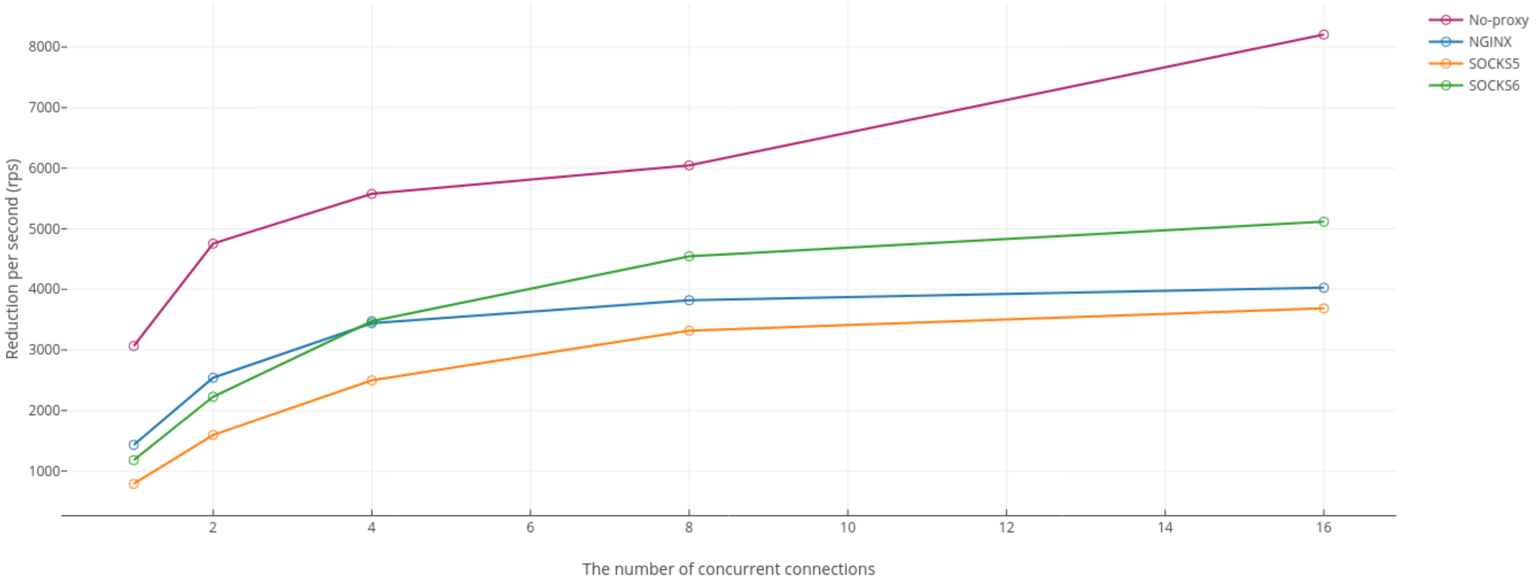


Results

Proxy overhead (ms)



Transaction processing rate (rps)



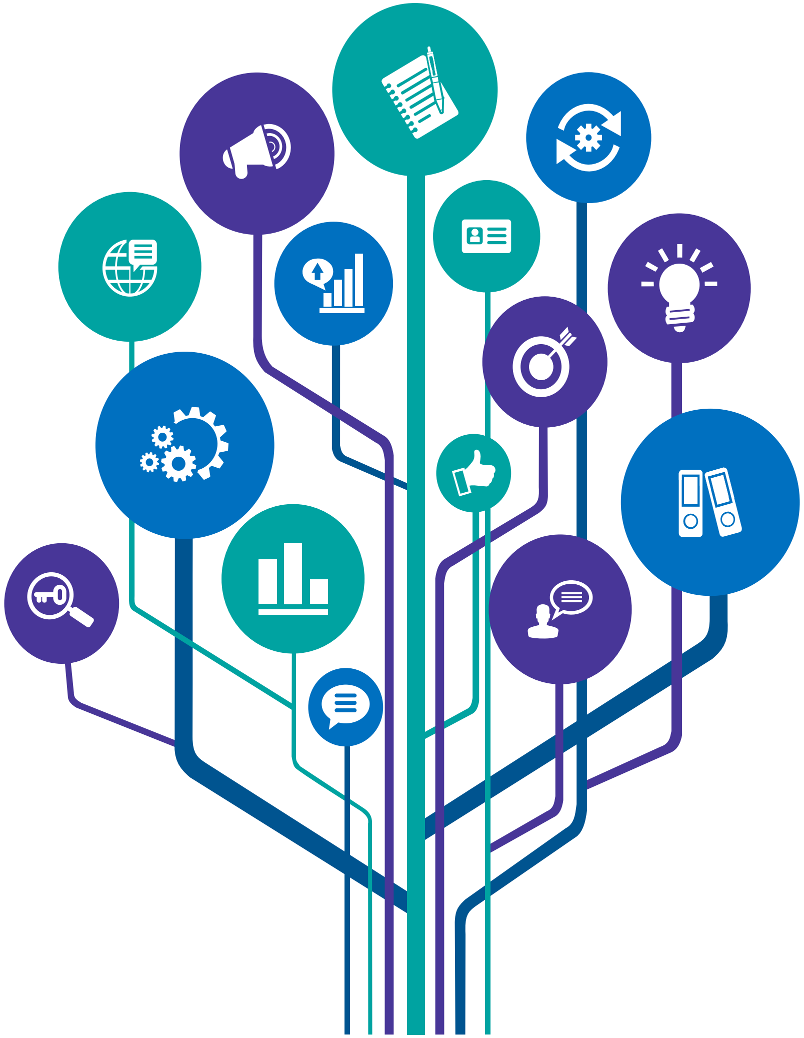
Comparison

Parameters	NGINX	SOCKS5	SOCKS6
Δt	✓		
Throughput			✓
Port scalability	✓		
Reconfiguration		✓	✓
Dynamicity		✓	✓
Security		✓	✓

Future work

- Considering more proxy implementations
- Implementing the BF chaining and uniform interfaces for BF
- Implementing Complex NF’s chaining
- Evaluating in real test-beds with SURF
- Integration with WHITEBOX
- Utilising framework and applying use cases
  - Redirection tools
  - Chaining BF
  - Security of bridges
- Integration with policy

EPI architecture: implementation & integrations



EPI implementation & integrations

Onno Valkering
o.a.b.valkering@uva.nl

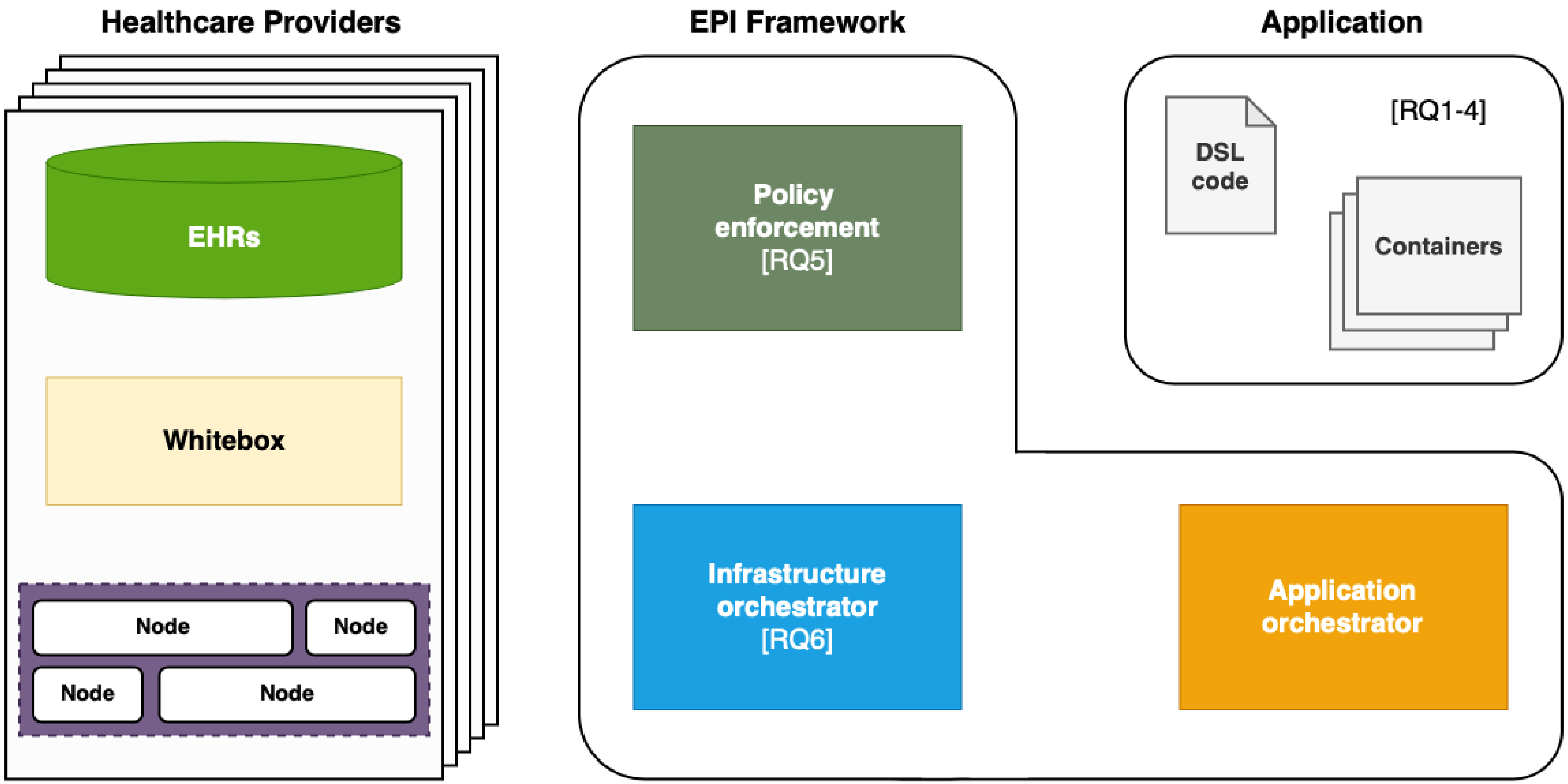
https://mns-research.nl

Thursday 22 April 2021

Table of Contents

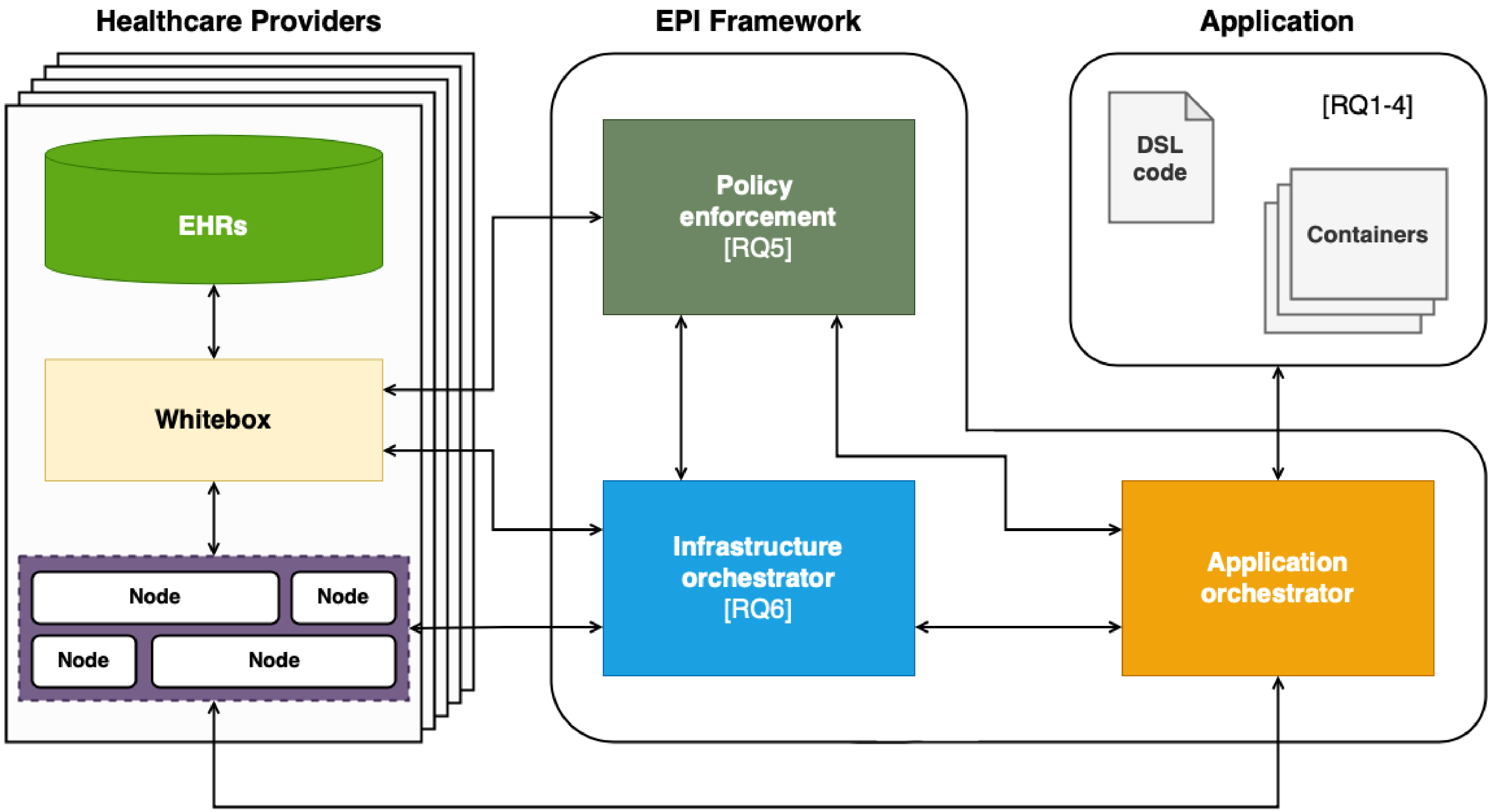
EPI architecture (simplified, high-level overview)

Todo: overlay boxes



EPI architecture (simplified, high-level overview)

Todo: overlay boxes



Implementation

- I worked on an implementation of the architecture that incorporates RQ4, RQ5, RQ6 (the three presentations before me)…
- TODO:
  - add diagram / visuals of the implementation
  - screenshots in extra slides
- A group of MSc students (±20) will implement a ML project using a similar feature set, as part of Web Services and Cloud Based systems course…

Integrations

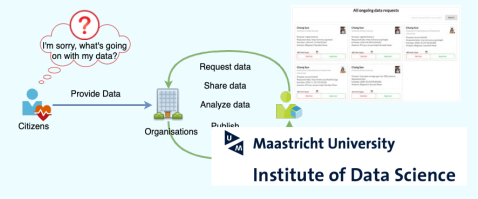
Data ingress

* ongoing

whitebox

https://whiteboxsystems.nl

ciTIzen-centric DatA pLatform (TIDAL)



https://sunchang0124.github.io/dist/homepage.html

* potential

EPI Framework

PERSONAL HEALTH TRAIN

https://pht.health-ri.nl

* ongoing

SURF Research Cloud

https://www.surf.nl/en/surf-research-cloud-collaboration-portal-for-research

netherlands eScience center

XENON

https://www.research-software.nl/software/xenon

* done

* includes

BRANE

https://brane-framework.org

* ongoing

VANTAGE

https://vantage6.ai

Infrastructure

Application

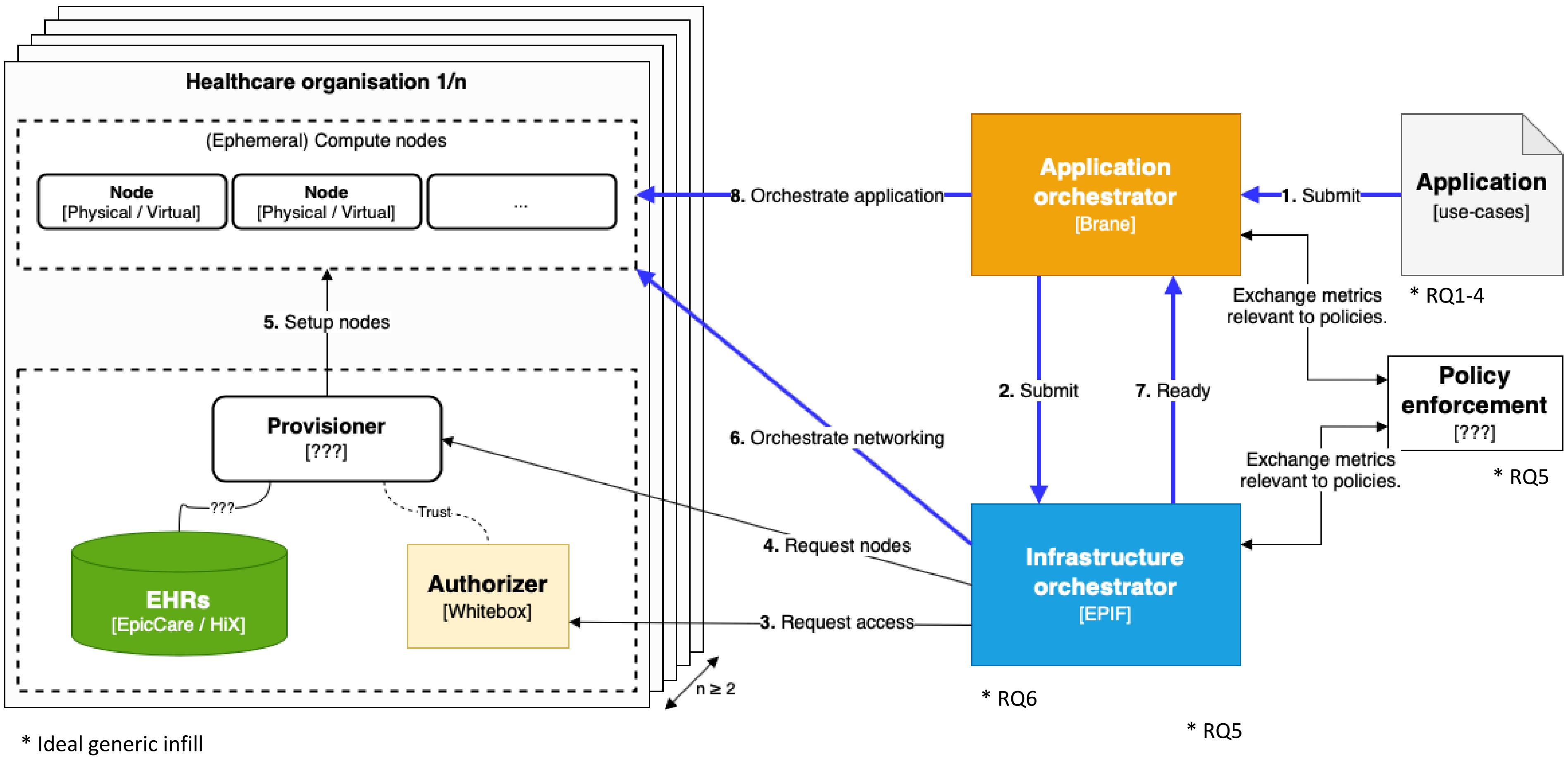
Next steps

- <u>Foremost</u>: Whitebox integration
  - Q1
  - Q2
  - Q3

- Implementation:
  - Improvements related, mainly, to fully support RQ6 proxy/bridges
  - UIs, potential integrations

Extra Slides

EPI architecture (as shown during previous meeting)



* Ideal generic infill

* Outcome of brainstorm sessions.

Any other business

Thank you!

For more information please contact:

Marloes Bons
bons.marloes@kpmg.nl
+31 6 23593759

EPI Linked-in group:
Enabling Personalized Interventions (EPI)