

Use Case 2 - Vorhersagemodelle für Asthma/COPD und Hirntumoren

Stefan Lenz

29.03.2019

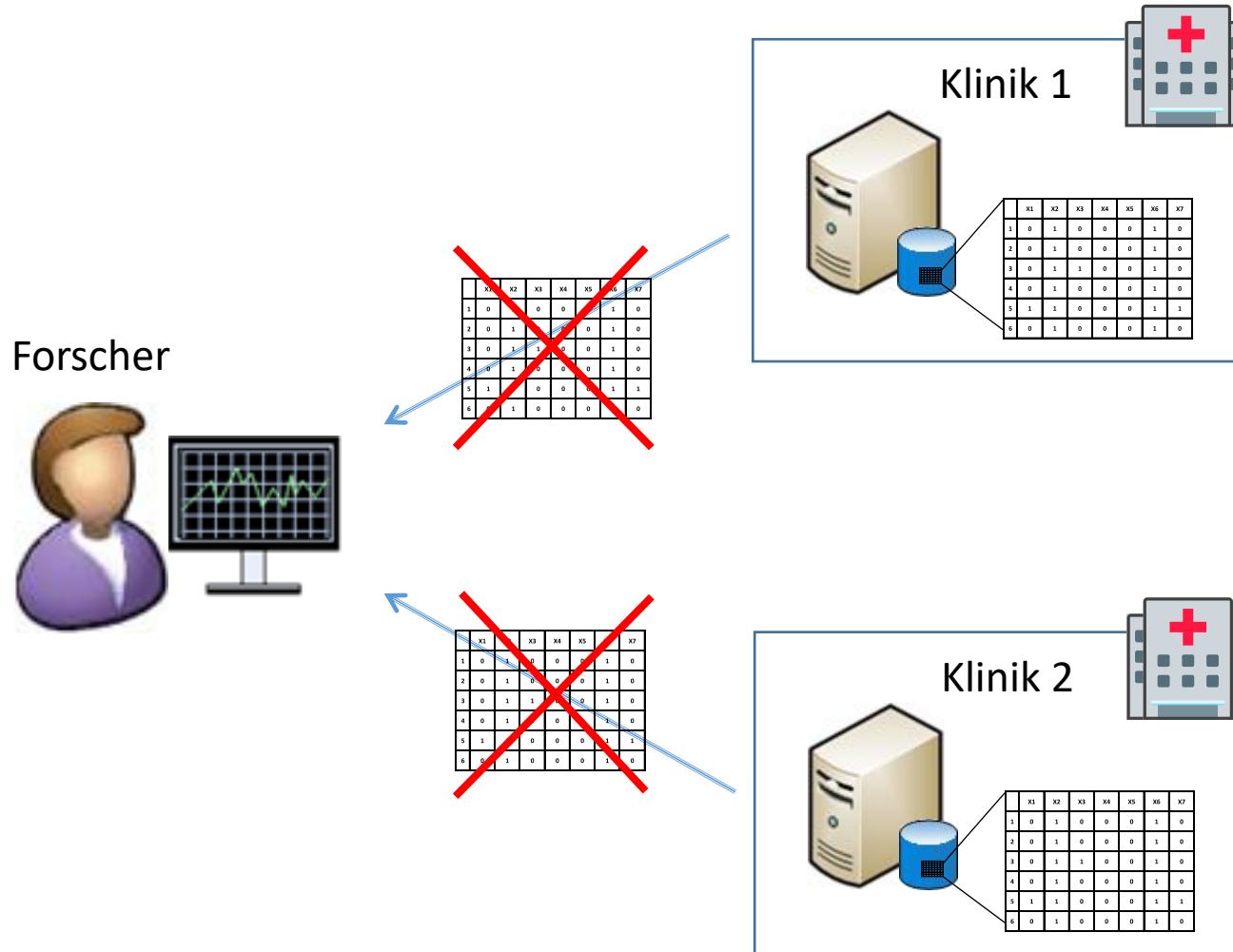
MIRACUM-Symposium

- Modelle zur Vorhersage von prognostisch relevanten Untergruppen entwickeln
- Vorhersagemodelle in die klinische Praxis bringen
- Das Ganze am Beispiel von
 - Asthma/COPD und
 - Neuroonkologie

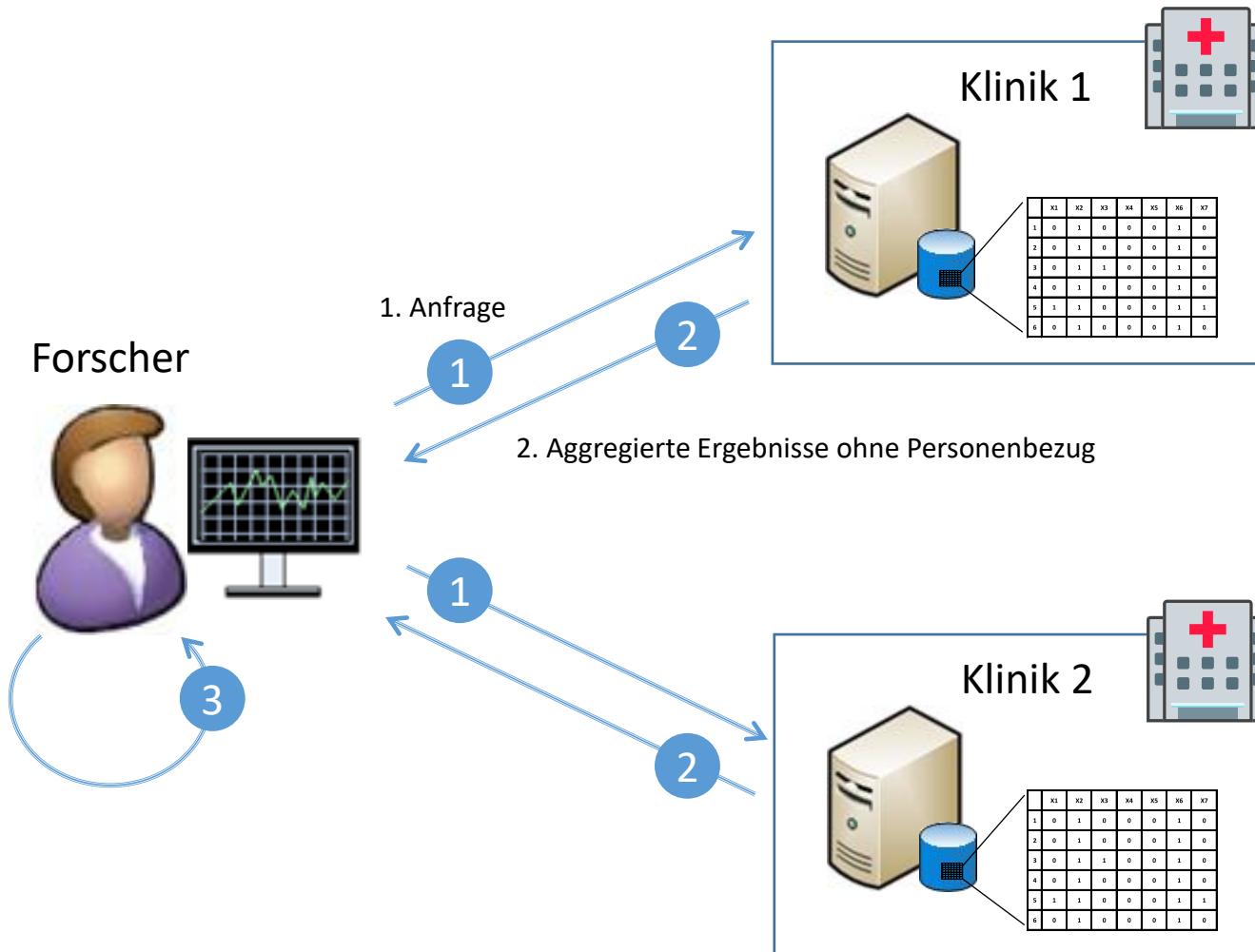
Ziele im Use Case 2

- Modelle zur Vorhersage von prognostisch relevanten Untergruppen entwickeln Muss konkreten Nutzen bringen!
- Vorhersagemodelle in die klinische Praxis bringen
- Das Ganze am Beispiel von
 - Asthma/COPD und
 - Neuroonkologieunsaubere klinische Routinedaten!hochdimensionaler Datensatz mit genetischen Daten!

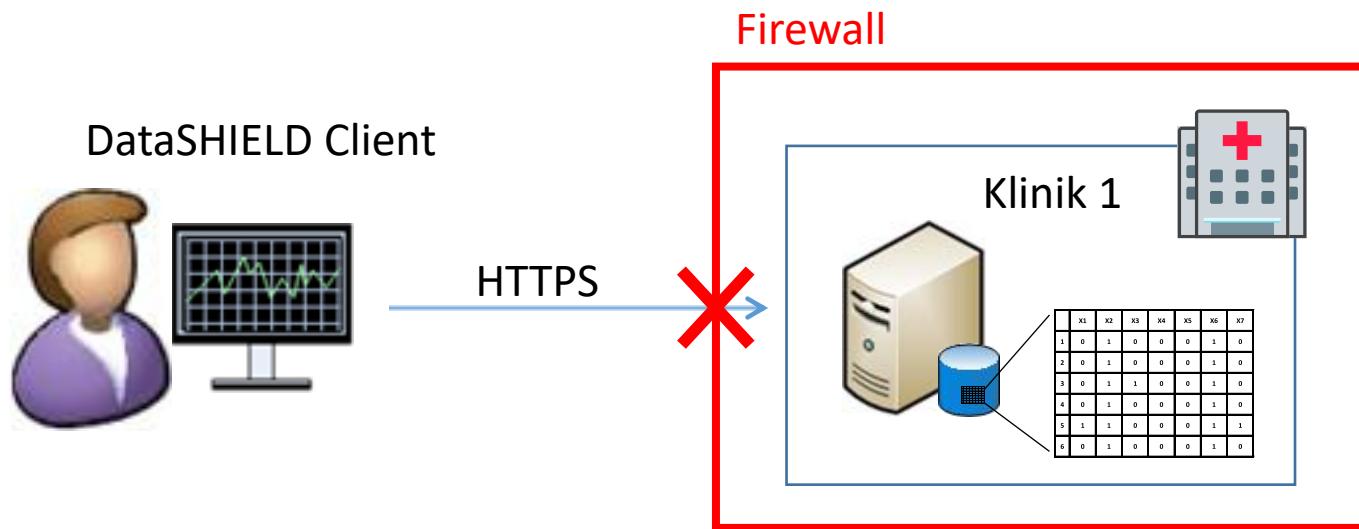
Verteilte datenschutzkonforme Analysen: Problemstellung



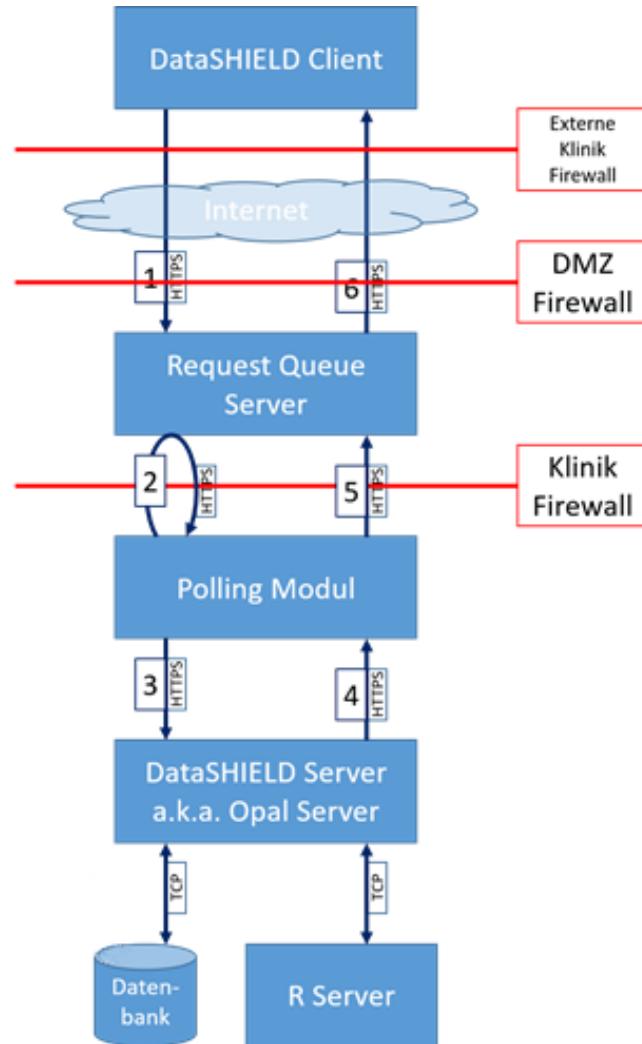
DataSHIELD: Prinzip



Firewall-Problematik



Polling-Konzept als zusätzliche Sicherheitsschicht für DataSHIELD



Implementierung der DataSHIELD Infrastruktur in MIRACUM



- Polling-Erweiterung als Software verfügbar
- Datenschutzkonzept und Informationsblatts für die DIZ-Leiter erstellt
- Erfolgreicher Test der Infrastruktur mit Freiburg und Erlangen



Julian Gründner,
Erlangen



Sebastian Schindler,
Magdeburg



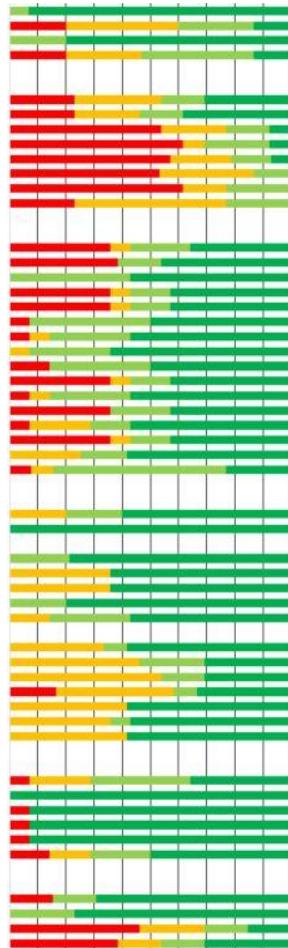
- Erfassung und Beschreibung relevanter und verfügbarer Variablen
- Abbildung auf LOINC/OMOP
- Erhebung mittels Fragebögen an die Kliniker der verschiedenen Standorte
- Separate Variablenlisten für Teile Asthma/COPD und Neuroonkologie
- Ergebnis im Confluence



Saskia Kiefer,
Freiburg

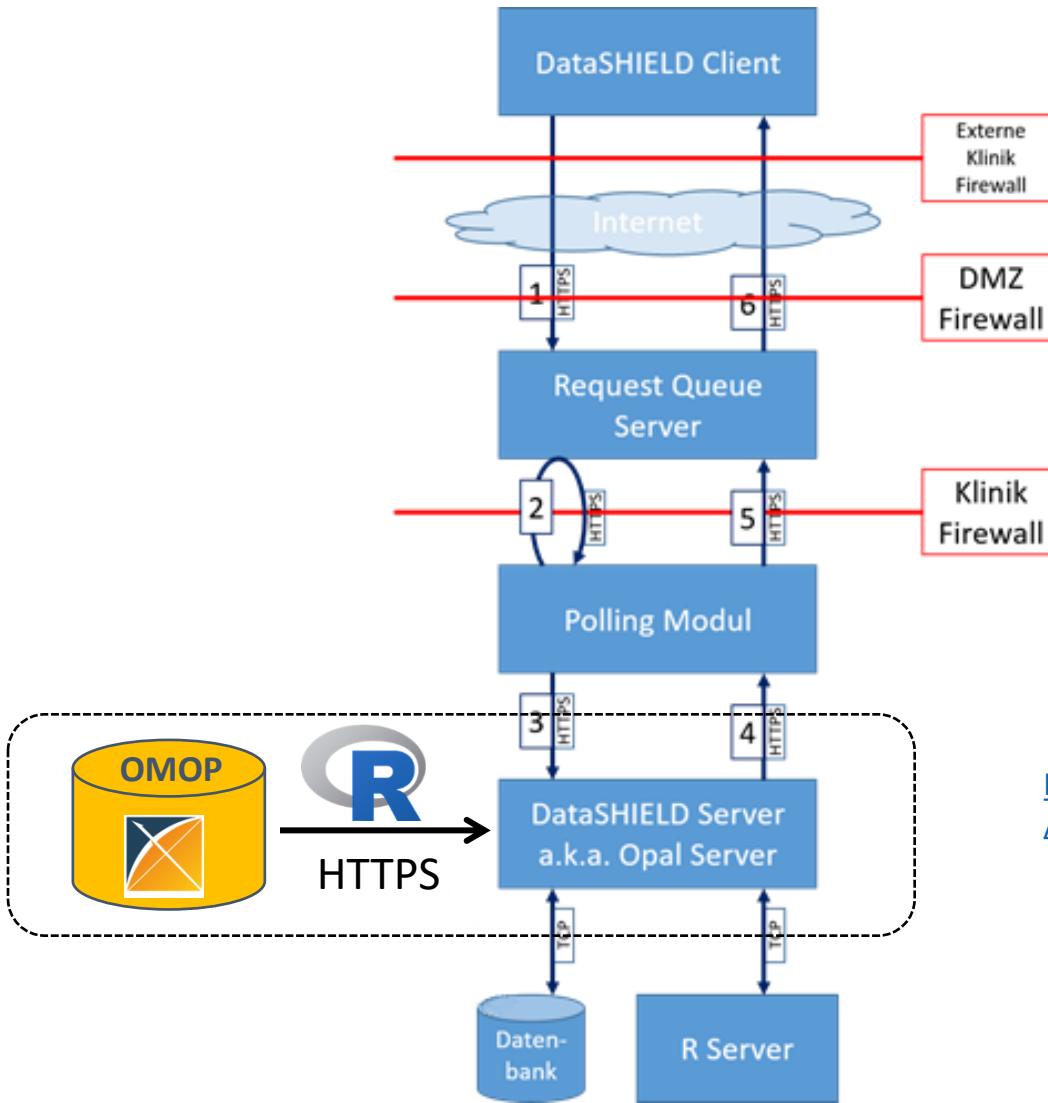


Caroline Broichhagen,
Freiburg



<https://confluence.imi.med.fau.de/pages/viewpage.action?pageId=94470934>
<https://confluence.imi.med.fau.de/pages/viewpage.action?pageId=101909917>

Datenimport für DataSHIELD: Von OMOP nach Opal



Dr. Petar Horki,
Freiburg

<https://confluence.imi.med.fau.de/pages/viewpage.action?pageId=101909933>

- Herbst 2019: Daten stationärer Asthma/COPD-Patienten bereit. Bis dahin:
 - Fragestellungen detaillieren
 - Ethikanträge
 - Installation der DataSHIELD-Infrastruktur an allen Standorten
- Einbettung von DataSHIELD in die MIRACUM-Infrastruktur, Generalisierung für zukünftige Analysen

- Prinzip: (potentiell) relevante Variablen in einer ansonsten unüberschaubaren Menge finden
- Beispiel: Welche genetischen Varianten haben Einfluss auf ein bestimmtes Krankheitsrisiko?

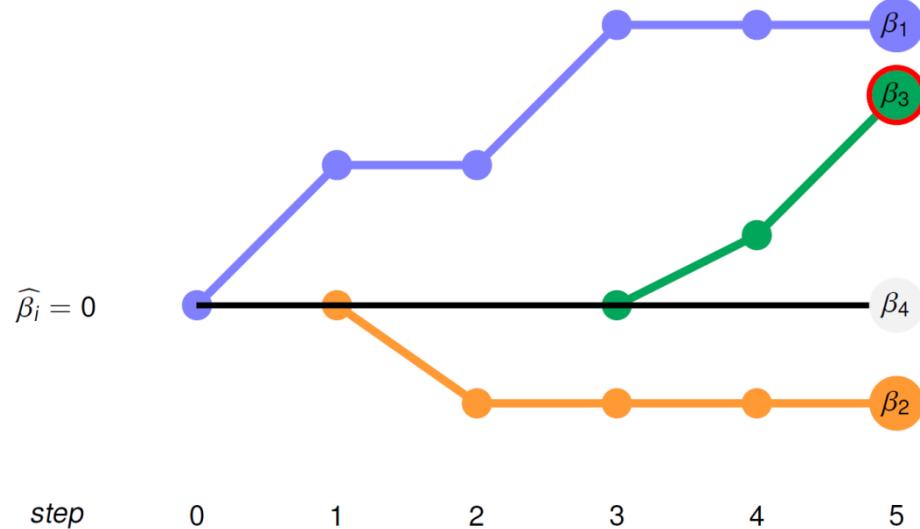
Pat-ID	X1	X2	X3	X4	...
1	1	1	0	1	1
2	0	0	0	0	0
3	1	1	0	1	0
...



Boosting zur Variablenelektion

- Boosting: schrittweise Erhöhung der Koeffizienten β_i in Regressionsmodellen, nur relevant ob $\beta_i \neq 0$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots$$



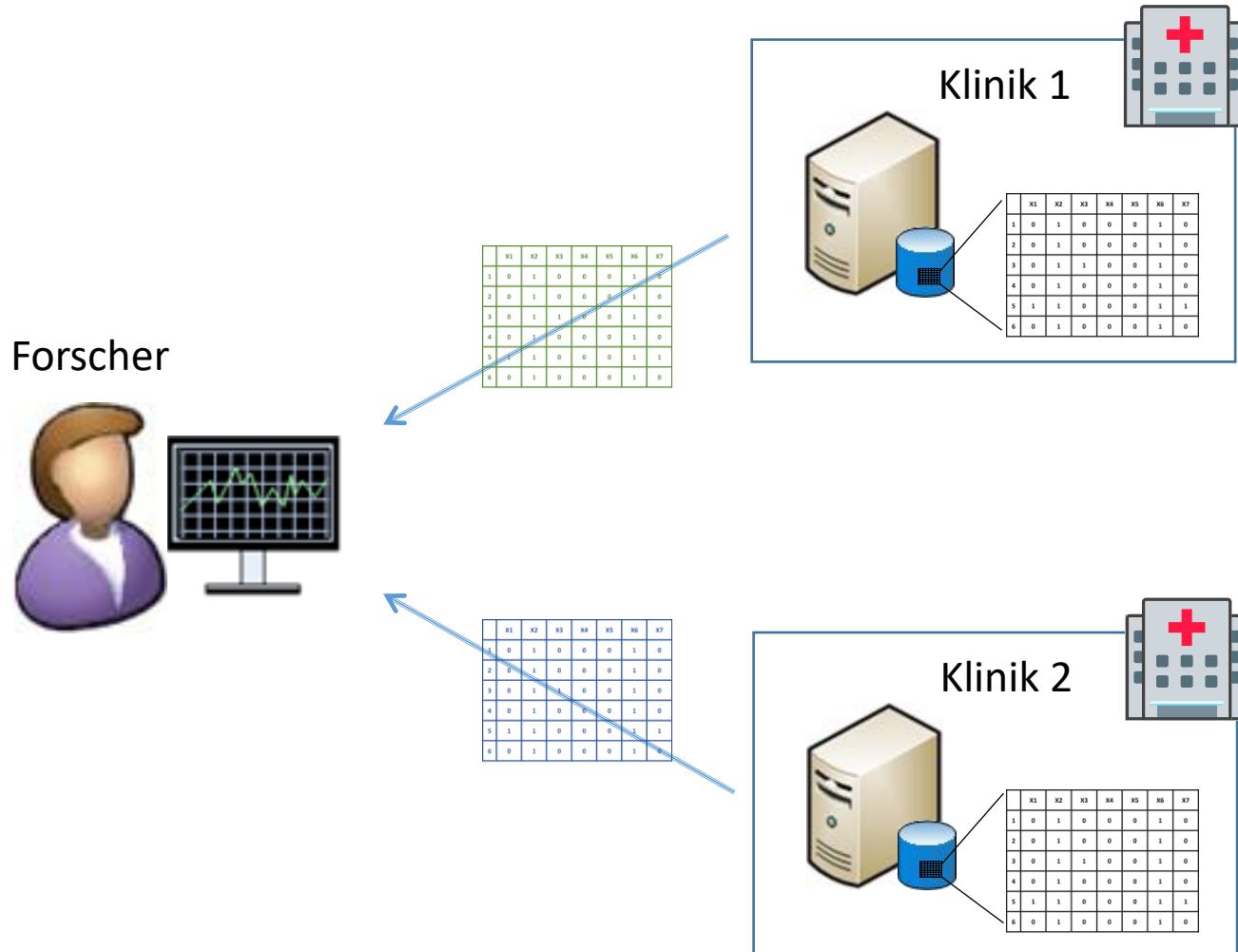
Distributed Boosting

- Boosting-Algorithmus für verteilte Daten
- Umgesetzt als Julia-Paket, aufbauend auf DataSHIELD
github.com/danielazoeller/ds_DistributedBoosting.jl
- Erstes Manuskript:
<https://arxiv.org/abs/1803.00422>
- Derzeit: Verteilte Analyse von DNA-Methylierungsdaten mit Gießen (Prof. Till Acker) und Freiburg und die Erstellung einer Publikation dazu

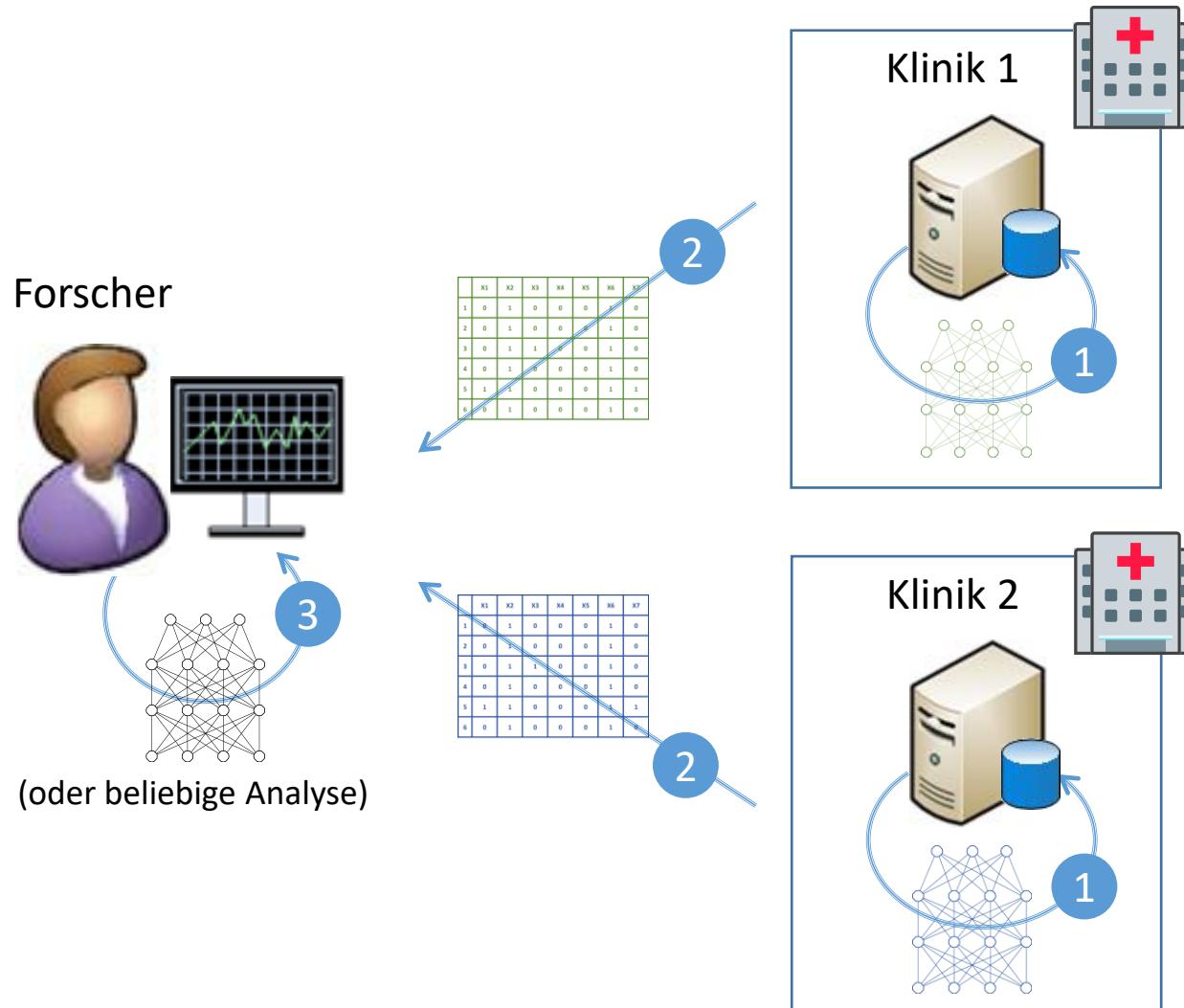


Dr. Daniela
Zöller, Freiburg

Generierung synthetischer Daten



Verteiltes Deep Learning mit generativen Modellen



2018 DataSHIELD Workshop, 06.-08.11



...



Danke für die Aufmerksamkeit

