Neue Methoden für die nationale Infrastruktur der MII am Beispiel der FLAME-Plattform

Prof. Dr. Fabian Prasser
Prof. Dr. Oliver Kohlbacher
Prof. Dr. Daniel Rückert

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Modul 3 - Methodenplattformen

„Die Methodenplattformen werden zentrale technisch-methodische Bedarfe der Medizininformatik-Initiative konsortienübergreifend adressieren und Lösungen entwickeln die für die standortübergreifende Datennutzung grundlegend sind.“

https://www.gesundheitsforschung-bmbf.de/de/medizininformatik-ausbau-und-erweiterung-16099.php

Übersicht & Ziele

PrivateAIM - Methodenplattform im Rahmen der aktuellen Förderphase der MII

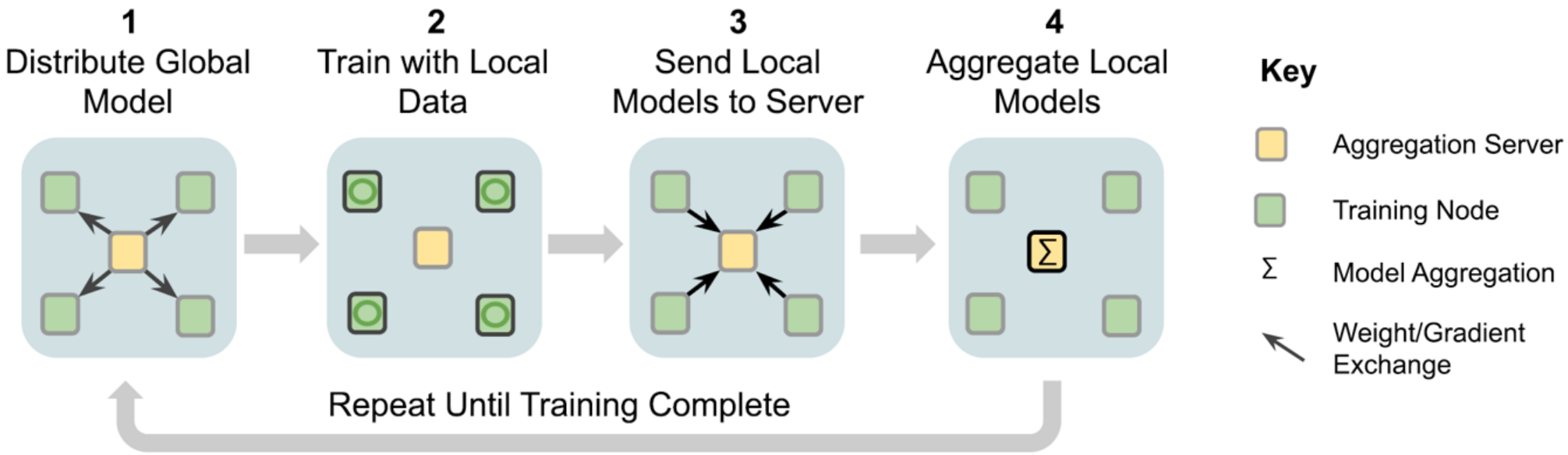
Ziel des Projekts:

> *"Das Ziel von PrivateAIM ist die Entwicklung einer föderierten Plattform für maschinelles Lernen (ML) und Datenanalytik, genannt FLAME, für die Medizininformatik-Initiative zu entwickeln, bei der die Analysen zu den Daten kommen und nicht die Daten zu den Analysen„*

"Code to Data"-Paradigma - Die Daten bleiben dort, wo sie sind, um die Privatheit optimal zu schützen und große Datenmengen (Bilder, Omics) verarbeiten zu können.

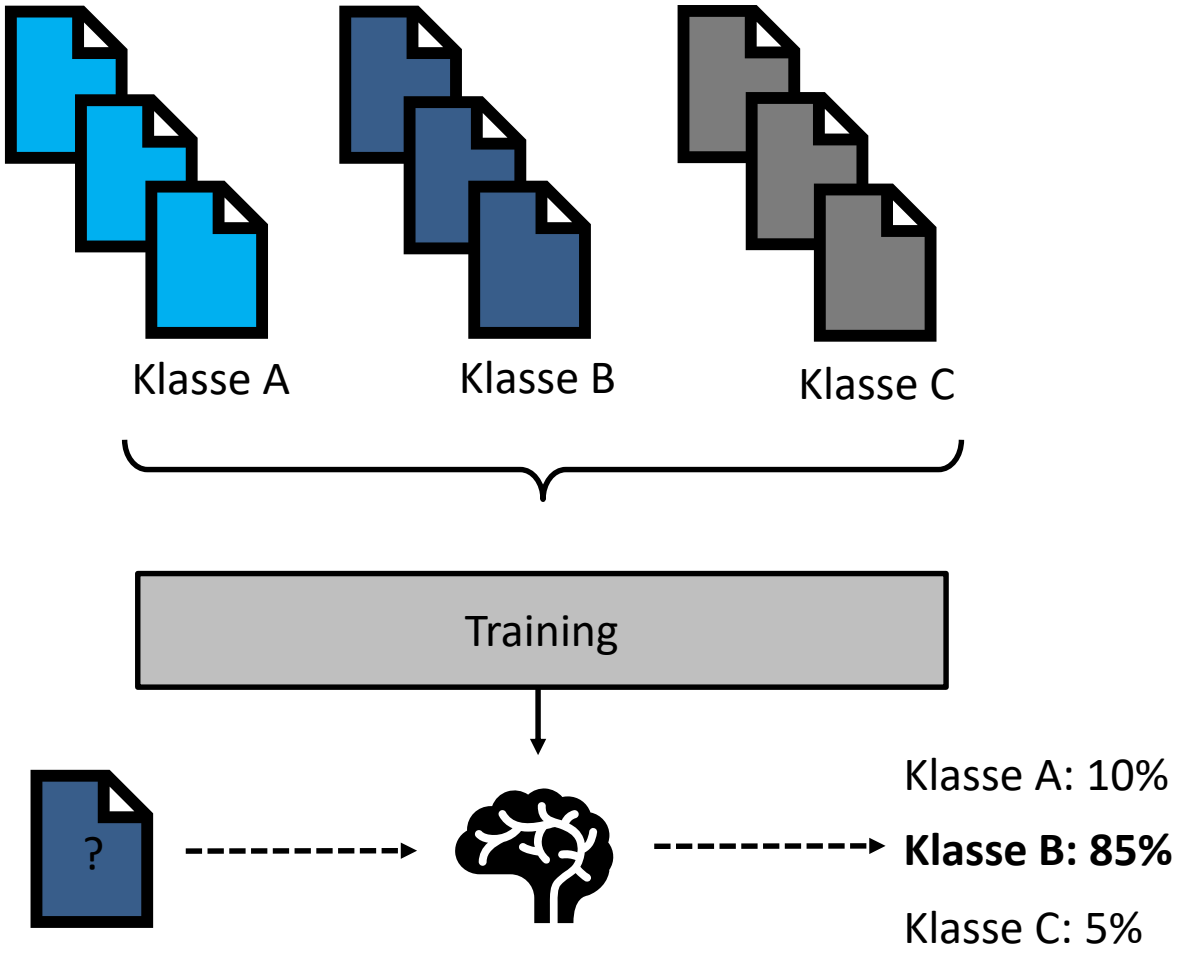
Was bedeutet „Code to Data“?

Beispiel: Federated Learning



Quelle: Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, Bakas S, Galtier MN, Landman BA, Maier-Hein K, Ourselin S, Sheller M, Summers RM, Trask A, Xu D, Baust M, Cardoso MJ. The future of digital health with federated learning. NPJ Digit Med. 2020 Sep 14;3:119

Warum ist das schwierig? (1)



„Membership Inference“

- Eingabedaten, die von einem Modell mit hoher Zuverlässigkeit klassifiziert werden, sind wahrscheinlich ähnlich wie die Trainingsdaten

„Attribute Disclosure“

- Bei bekannter Klasse können mögliche Eingabewerte rekonstruiert werden (Modellinversion)

Datenlecks und Rekonstruktion

- Speicherung von Trainingsdaten durch Sprachmodelle
- Rekonstruktion von Bildern

Literatur: Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 3-18). IEEE.

Das PrivateAIM-Konsortium

15 Teilnehmer aus allen vier MII-Konsortien (und darüber hinaus)

Koordinatoren

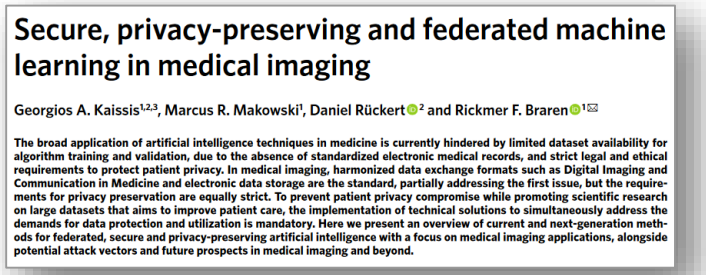
- Oliver Kohlbacher (U Tübingen)
- Fabian Prasser (Charité)
- Daniel Rückert (TU München)

Drei assoziierte Nachwuchsgruppen

- Datenschutzbewusstes Training von ML-Modellen auf medizinischen Daten (Tübingen)
- Vertrauenswürdiges Maschinelles Lernen (Essen)
- Integration von Multimedia-Objekten und PACS-Umgebungen (Kiel)

Charité - Universitätsmedizin Berlin (Charité)	Prof. Dr. Fabian Prasser
Helmholtz Center for Information Security (CISPA)	Prof. Dr. Mario Fritz
Deutsches Krebsforschungszentrum (DKFZ)	Dr. Ralf Omar Floca
University of Tübingen (EKUT)	Prof. Dr. Nico Pfeifer
Ludwig-Maximilians-Universität München (LMU)	Prof. Dr. Ulrich Mansmann
TMF e.V. (TMF)	Dr. Sebastian C. Semler
Technische Universität München (TUM)	Prof. Dr. Daniel Rückert
Friedrich-Alexander-Universität Erlangen-Nürnberg (UKER)	Prof. Dr. Thomas Ganslandt
University of Freiburg (UKFR)	Prof. Dr. Harald Binder
University Hospital Heidelberg (UKHD)	Prof. Dr. Christoph Dieterich
University of Cologne (UKK)	Prof. Dr. Oya Beyan
Leipzig University Medical Center (UKL)	Prof. Dr. Toralf Kirsten
University Hospital Tübingen (UKT)	Prof. Dr. Oliver Kohlbacher
Ulm University (UKU)	Prof. Dr. Hans Kestler
Medical Faculty Mannheim, Heidelberg University (UMM)	Prof. Dr. Martin Lablans

Vorarbeiten

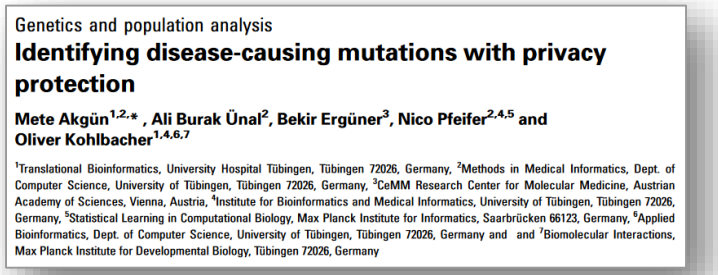


Secure, privacy-preserving and federated machine learning in medical imaging

Georgios A. Kaissis[1,2,3], Marcus R. Makowski[1], Daniel Rückert[2] and Rickmer F. Braren[1]

The broad application of artificial intelligence techniques in medicine is currently hindered by limited dataset availability for algorithm training and validation, due to the absence of standardized electronic medical records, and strict legal and ethical requirements to protect patient privacy. In medical imaging, harmonized data exchange formats such as Digital Imaging and Communication in Medicine and electronic data storage are the standard, partially addressing the first issue, but the requirements for privacy preservation are equally strict. To prevent patient privacy compromise while promoting scientific research on large datasets that aims to improve patient care, the implementation of technical solutions to simultaneously address the demands for data protection and utilization is mandatory. Here we present an overview of current and next-generation methods for federated, secure and privacy-preserving artificial intelligence with a focus on medical imaging applications, alongside potential attack vectors and future prospects in medical imaging and beyond.

Privacy-Preserving Machine Learning



Genetics and population analysis

Identifying disease-causing mutations with privacy protection

Mete Akgün[1,2,*], Ali Burak Ünal[2], Bekir Ergüner[3], Nico Pfeifer[2,4,5] and Oliver Kohlbacher[1,4,6,7]

[1]Translational Bioinformatics, University Hospital Tübingen, Tübingen 72026, Germany, [2]Methods in Medical Informatics, Dept. of Computer Science, University of Tübingen, Tübingen 72026, Germany, [3]CeMM Research Center for Molecular Medicine, Austrian Academy of Sciences, Vienna, Austria, [4]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen 72026, Germany, [5]Statistical Learning in Computational Biology, Max Planck Institute for Informatics, Saarbrücken 66123, Germany, [6]Applied Bioinformatics, Dept. of Computer Science, University of Tübingen, Tübingen 72026, Germany and [7]Biomolecular Interactions, Max Planck Institute for Developmental Biology, Tübingen 72026, Germany

Secure Multi-Party Computation

TECHNICAL NOTE

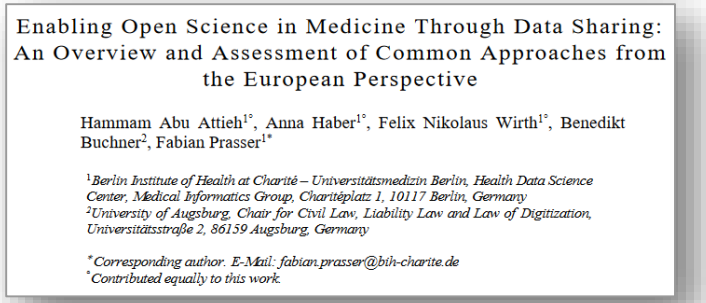
A scalable software solution for anonymizing high-dimensional biomedical data

Thierry Meurers[1,*], Raffael Bild[2], Kieu-Mi Do[3] and Fabian Prasser[1]

[1]Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Medical Informatics, Charitéplatz 1, 10117 Berlin, Germany; [2]School of Medicine, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany and [3]Faculty of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

*Correspondence address. Thierry Meurers, Berlin Institute of Health at Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. E-mail: thierry.meurers@charite.de https://orcid.org/0000-0001-8168-7067

Data Anonymization



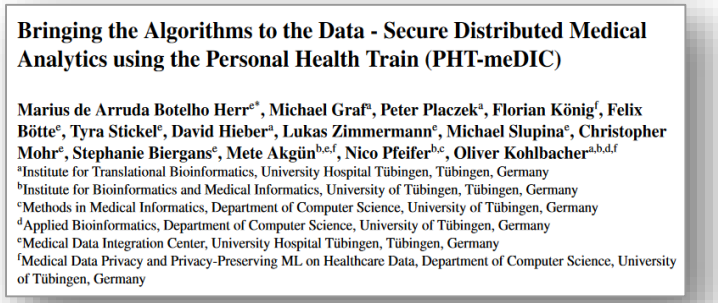
Enabling Open Science in Medicine Through Data Sharing: An Overview and Assessment of Common Approaches from the European Perspective

Hammam Abu Attieh[1°], Anna Haber[1°], Felix Nikolaus Wirth[1°], Benedikt Buchner[2], Fabian Prasser[1*]

[1]*Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Health Data Science Center, Medical Informatics Group, Charitéplatz 1, 10117 Berlin, Germany*
[2]*University of Augsburg, Chair for Civil Law, Liability Law and Law of Digitization, Universitätsstraße 2, 86159 Augsburg, Germany*

*Corresponding author. E-Mail: fabian.prasser@bih-charite.de
°Contributed equally to this work.

Technico-Legal Analyses



Bringing the Algorithms to the Data - Secure Distributed Medical Analytics using the Personal Health Train (PHT-meDIC)

Marius de Arruda Botelho Herr[e*], Michael Graf[a], Peter Placzek[a], Florian König[f], Felix Bötte[e], Tyra Stickel[e], David Hieber[a], Lukas Zimmermann[e], Michael Slupina[e], Christopher Mohr[e], Stephanie Biergans[e], Mete Akgün[b,e,f], Nico Pfeifer[b,c], Oliver Kohlbacher[a,b,d,f]

[a]Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany
[b]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany
[c]Methods in Medical Informatics, Department of Computer Science, University of Tübingen, Germany
[d]Applied Bioinformatics, Department of Computer Science, University of Tübingen, Germany
[e]Medical Data Integration Center, University Hospital Tübingen, Tübingen, Germany
[f]Medical Data Privacy and Privacy-Preserving ML on Healthcare Data, Department of Computer Science, University of Tübingen, Germany

PHT Implementation

Wirth et al. BMC Med Inform Decis Mak (2021) 21:242
https://doi.org/10.1186/s12911-021-01602-x

BMC Medical Informatics and Decision Making

RESEARCH Open Access

Privacy-preserving data sharing infrastructures for medical research: systematization and comparison

Felix Nikolaus Wirth*, Thierry Meurers, Marco Johns and Fabian Prasser

Data Sharing Architectures

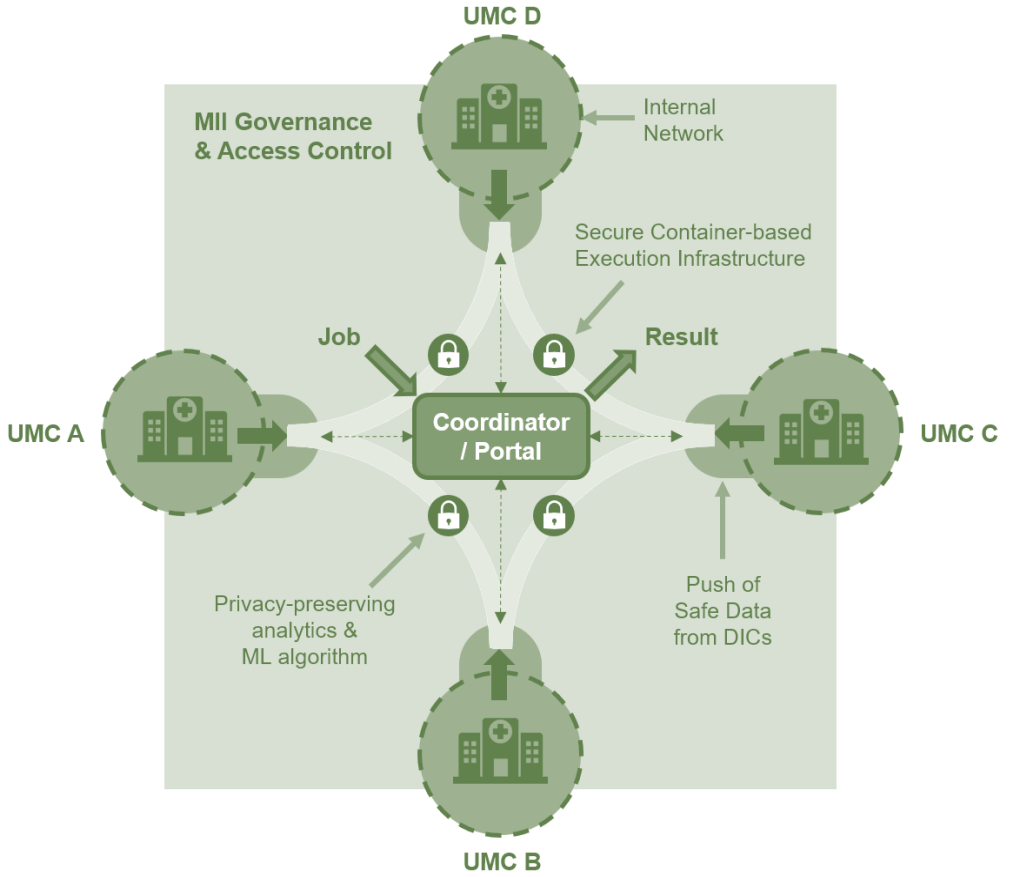
Eckpfeiler

Entwicklung

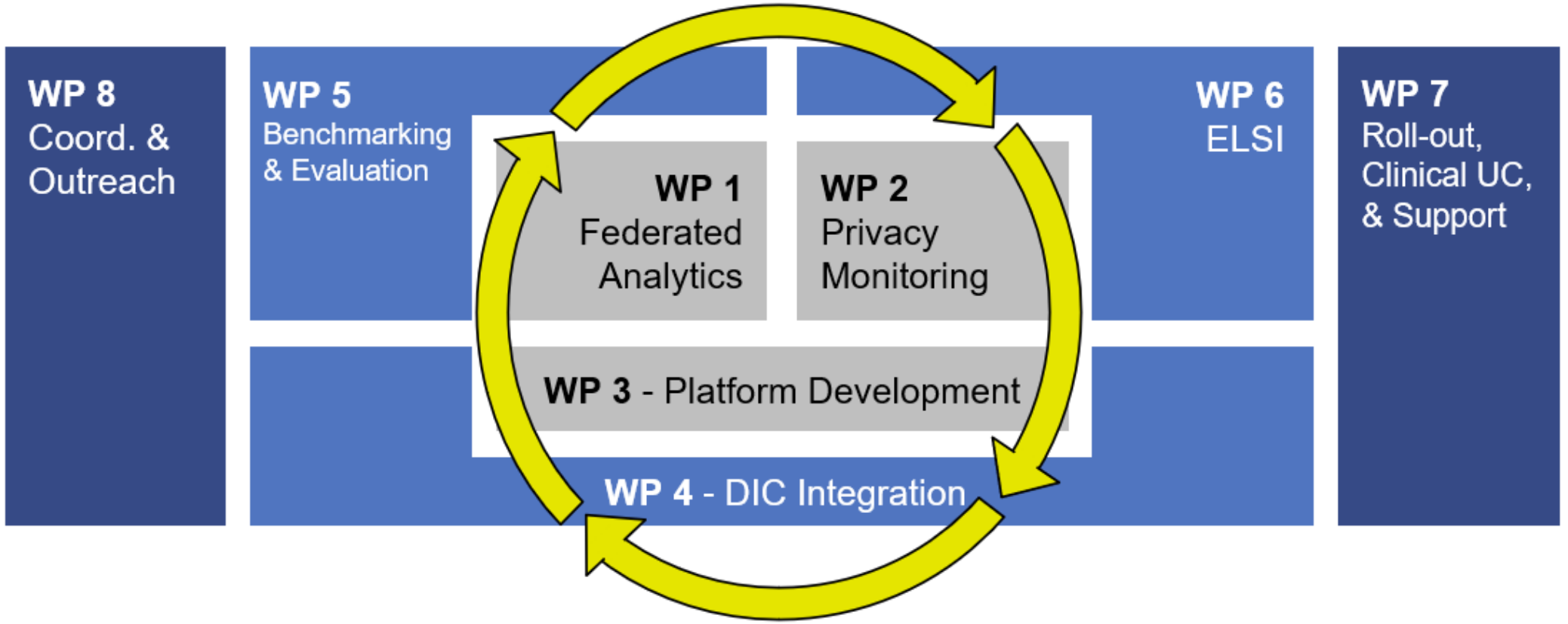
- Innovativer Methoden für föderiertes Lernen
- Robuster Datenschutzgarantien für föderierte Ansätze
- Praktikable Plattform für verteilte Analysen

Einsatz dieser Lösungen in einer konsistenten Plattform an allen MII-Standorten

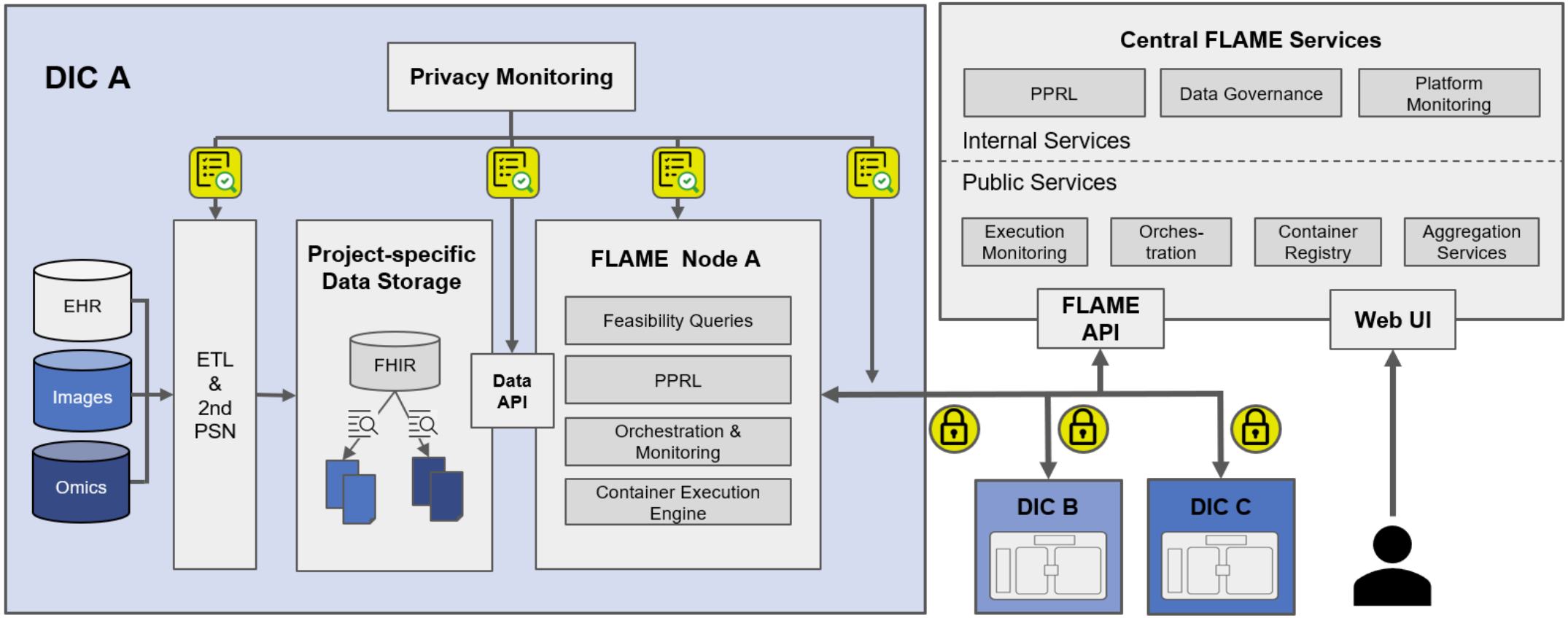
Unterstützung anderer (klinischer) Anwendungsfälle innerhalb der MII



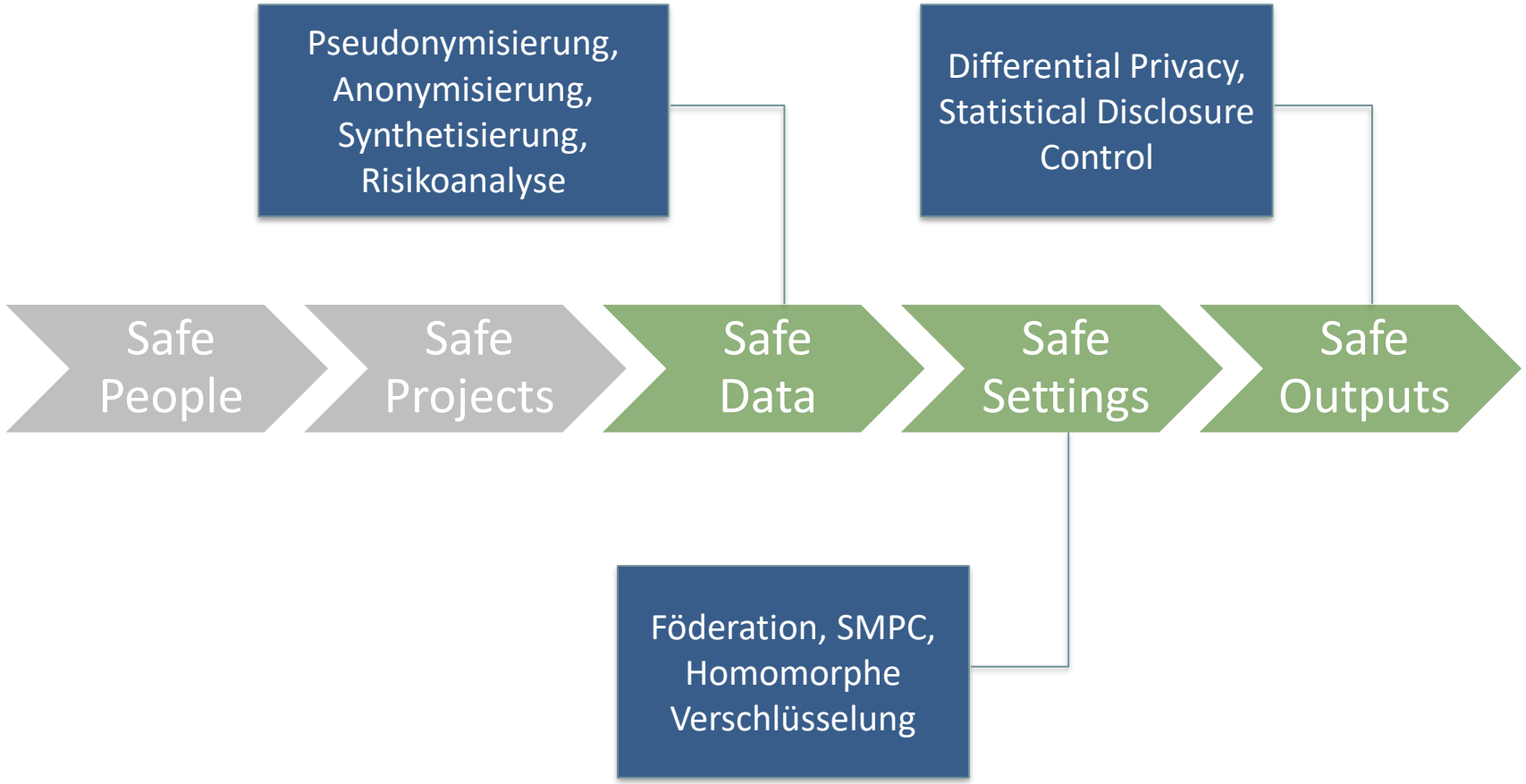
Arbeitspakete



Die FLAME Plattform



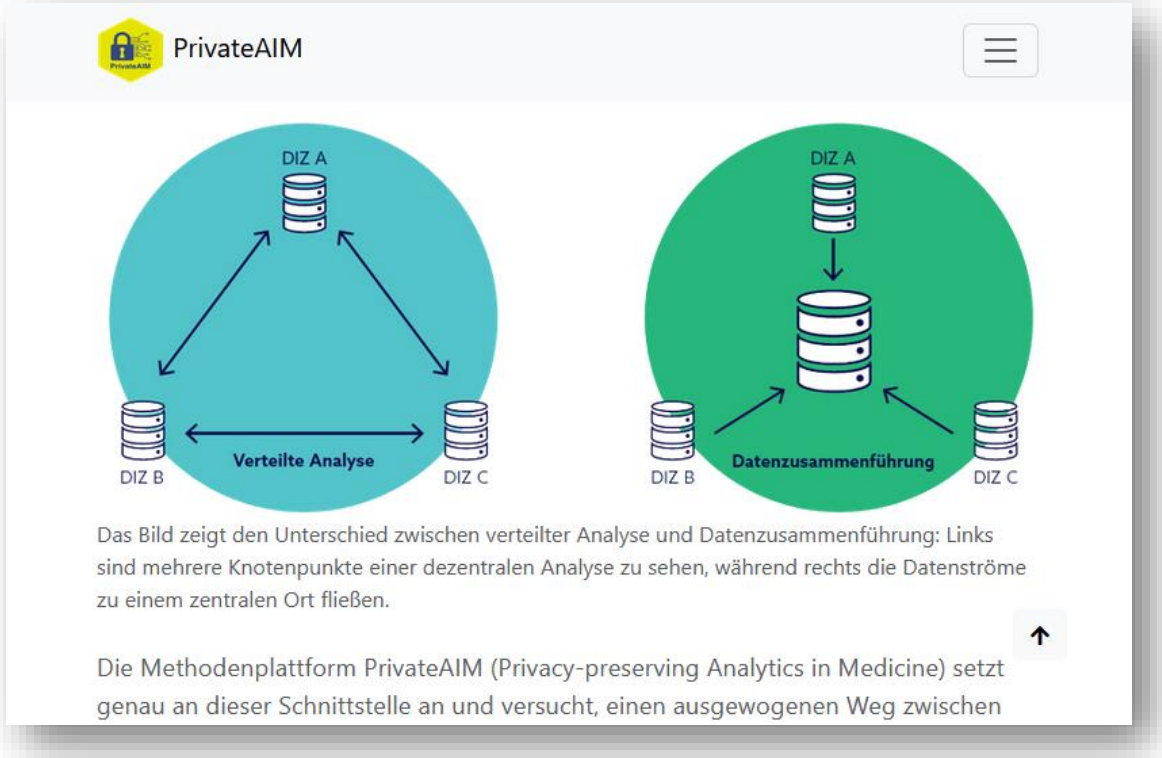
Welche Schutzmaßnahmen bietet die FLAME-Plattform?



Aktueller Status nach einem Jahr

- Erste Version der FLAME-Plattform kurz vor Fertigstellung („Minimum Viable Product“)
- Deployment an Pilotstandorten in PrivateAIM (Erlangen, Berlin, Leipzig, Tübingen)
- Anschließend Evaluation
- Dann Weiterentwicklung und Roll-Out

Danke für Ihre Aufmerksamkeit!



https://privateaim.de