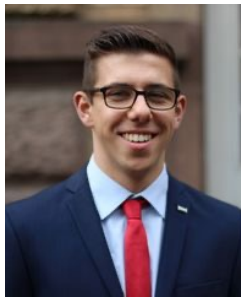

Join us for the
2022 NephroCAGE Symposium
Aug 16, 2022 @ 6.15 am PDT / 9.15 am EDT / 3.15pm CEST
Konstantin D. Pandl
Research Associate
Karlsruhe Institute of Technology
NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
2nd International NephroCAGE Symposium, Aug 16, 2022
Hasso Plattner Institut
KIT Karlsruher Institut für Technologie
THE UNIVERSITY OF BRITISH COLUMBIA
Université de Montréal
POLYTECHNIQUE MONTRÉAL
GenomeQuébec
Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag
PIRCHE
CHARITÉ
Centre universitaire de santé McGill
McGill University Health Centre
CHUM
GenomeCanada
Genome BritishColumbia

KIT project team

M. Sc. Florian Leiser



- Research associate at the KIT since 2021
- M.Sc. in information systems in 2021
- Research interests: health information systems, informed ML, federated learning

M. Sc. Konstantin Pandl

- Research associate at KIT since 2019
- M. Sc. in electrical engineering and information technology in 2018
- Research interests: machine learning, digital health, distributed systems

Dr. Scott Thiebes

- Research associate with Prof. Sunyaev since 2014
- Research interests: digital transformation, information privacy and critical digital health infrastructures, trustworthy AI, persuasive digital health technologies

Prof. Dr. Ali Sunyaev

- Professor at KIT and director of the AIFB since 2018
- Previous professorships at the University of Cologne and Kassel
- Research interests: trustworthy artificial intelligence, innovative health IT solutions

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Goals of the work package

Design a **federated learning infrastructure**

1

Develop this infrastructure and **deploy** it at each site

2

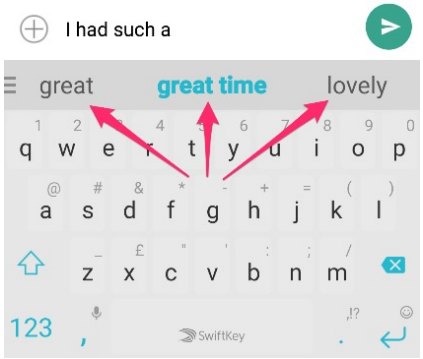
Evaluate the infrastructure's utility

3

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

History of federated learning

- Introduced by Google in 2017
- **Initial use case:** high-quality, machine learning (ML)-based word suggestions for the Android keyboard



- **Problem:** ML process typically runs on a large data set in the cloud, but keyboard inputs are too sensitive to share them with a cloud server
- **Solution:** federated learning

Image: https://askubuntu.com/questions/1125541/word-prediction-like-android-keyboard

NEPHROCAGE
German-Canadian consortium on AI for improved kidney transplantation outcome
2nd International NephroCAGE Symposium, Aug 16, 2022

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Functional principle of federated learning

- **Goal:** ML across private data silos
- **Key idea:** train ML models on local data and only exchange ML models
- Process consists of repeated rounds; each round comprises 4 steps
- Central model update through averaging the local model parameters

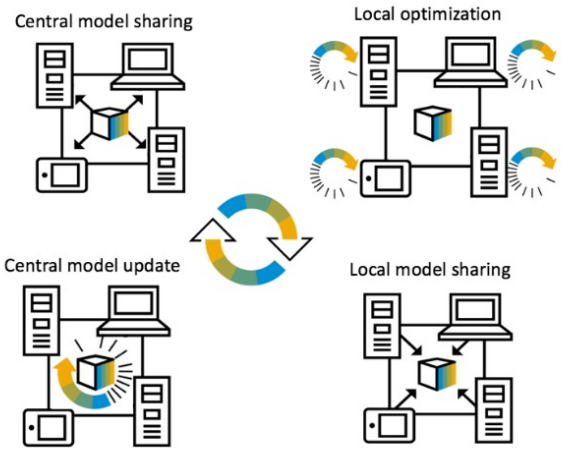
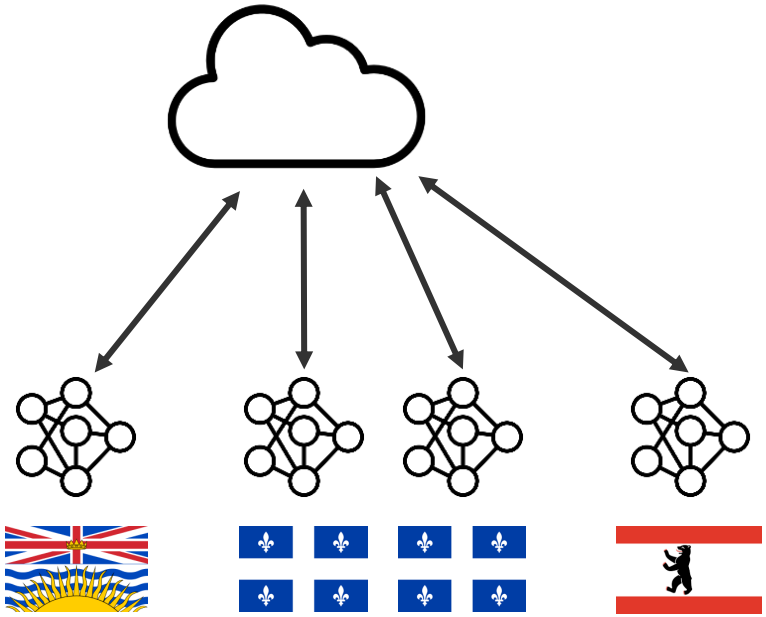
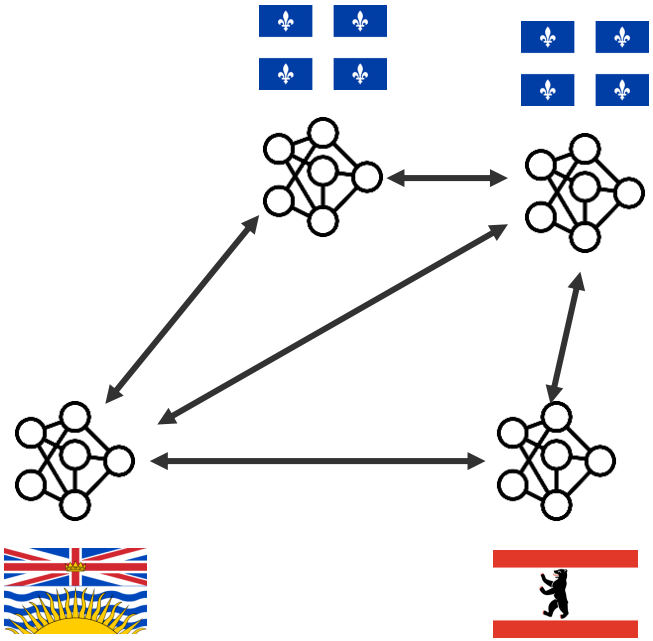


Image: https://medium.com/sap-machine-learning-research/client-sided-differential-privacy-preserving-federated-learning-1fab5242d31b

Centralized and Decentralized Learning Infrastructure





2nd Int'l NephroCAGE Symposium, Aug 16, 2022

Konstantin D. Pandl, Research Associate, KIT

Centralized: each clinic server communicates with a single cloud server to exchange data

Decentralized: each clinic communicates with other clinics to exchange data

Icons from flaticon.com

Supported by: Federal Ministry for Economic Affairs and Climate Action, on the basis of a decision by the German Bundestag

Collaborative learning principles

- **Federated learning**: local optimization, then sharing updates, then aggregation, then distribution to every node
- **Braintorrent**: Local optimization, then sharing of local updates with global node, then aggregation, no redistribution [1]
- **Institutional incremental learning**: One node optimizes, then shares with one other node which re-optimizes the model [2, 3]

- Due to the challenging aggregation of random forests, we started with developing and deploying **institutional incremental learning**

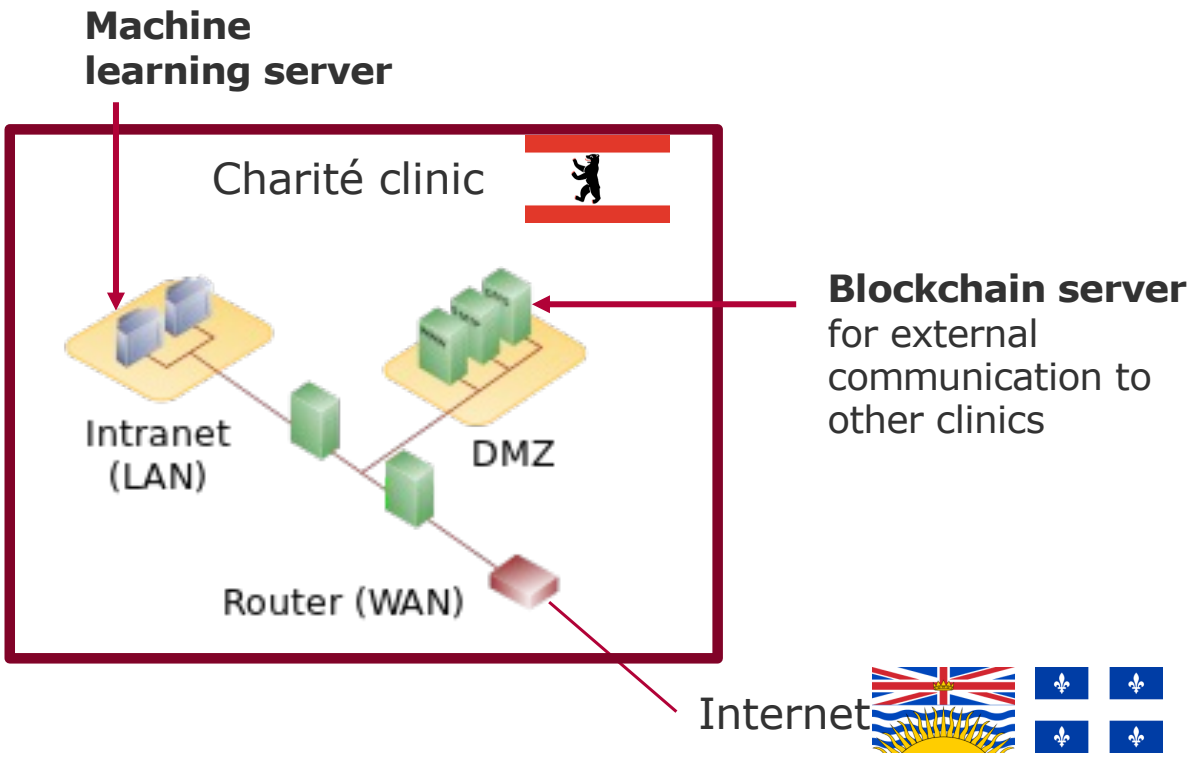
1. Roy, A.G., et al., *Braintorrent: A peer-to-peer environment for decentralized federated learning.* arXiv preprint arXiv:1905.06731, 2019.
2. Sheller, M.J., et al. *Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation*. in *International MICCAI Brainlesion Workshop*. 2018. Springer.
3. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., ... & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, *10*(1), 1-12.

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Blockchain

- Originated with the emergence of Bitcoin, today a variety of different blockchain networks exist
- In the scope of our project: **private peer-to-peer network** of **institutions**
- Characteristics:
  - **Replicated** ledger (i.e., each institution stores a copy of the blockchain locally)
  - **Immutable** (i.e., data can be written but cannot be removed)
  - Each institution has **equal rights**
  - **Transparent** and, thus, **auditable**
- Goals in the project:
  - **Communication** between the institutions **through the blockchain**
  - **Store ML models** (or a representation such as hashes) on the blockchain ledger

Proposed two-server setup in case of a demilitarized zone



Blockchain server for external communication to other clinics

2nd Int'l NephroCAGE Symposium, Aug 16, 2022

Konstantin D. Pandl, Research Associate, KIT

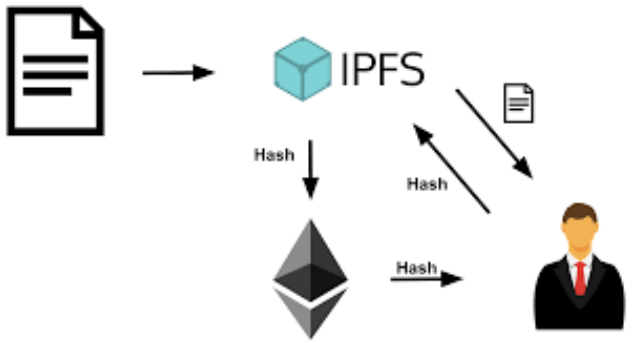
Storing data on the blockchain

- Converting ML models to the hexadecimal format, and converting it back, using Python
- Storing models and retrieving models from the blockchain
  - Successfully store models of up to 32 KB in one transaction
  - For larger transactions
    - Split model into chucks of 32 KB
    - Send one transaction containing header information of the transaction hashes and order of the following transactions which contain the actual ML models
    - Allows sending models of up to 32 MB (1000 transaction hashes can be included in the header)

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

IPFS as a peer to peer file sharing system

- For further scalability, and to keep the blockchain lightweight, we are currently exploring the InterPlanetary File System (IPFS)
  - Private peer to peer network to retrieve files by hashes
  - Other research uses it in combination with blockchains in general, and blockchain-based federated learning
  - Runs on the blockchain server

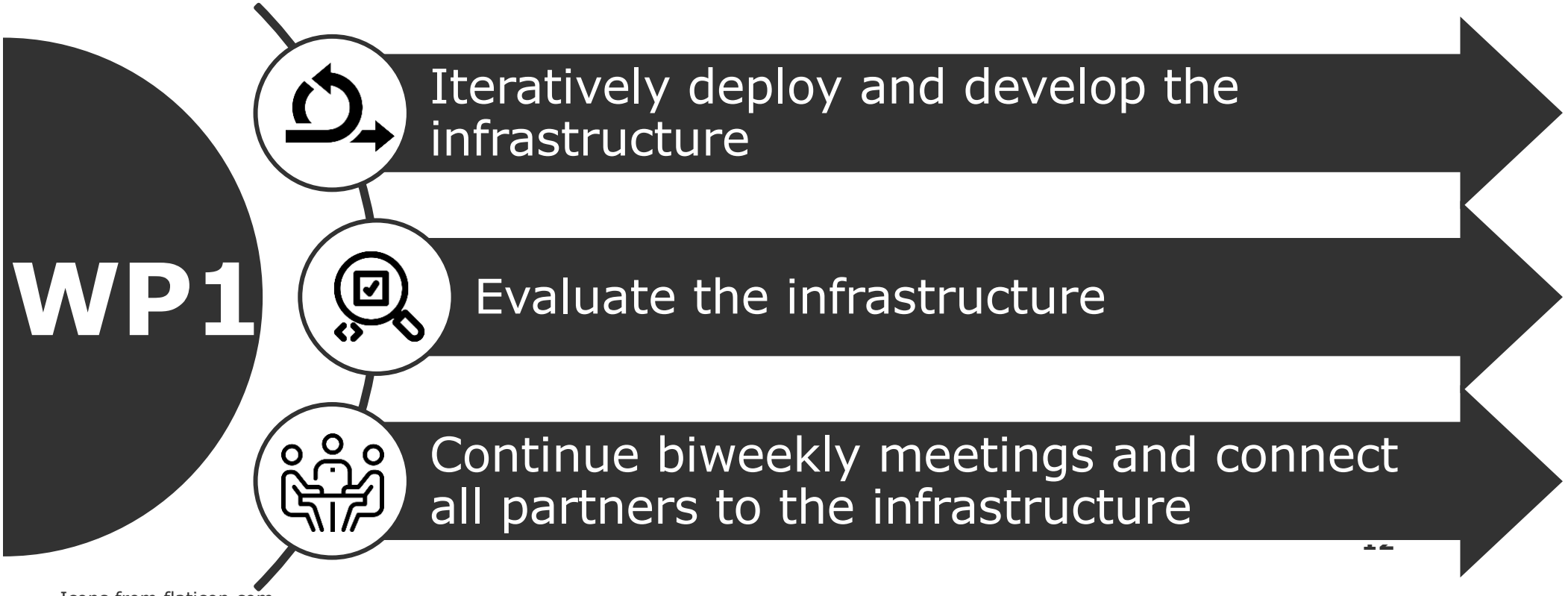


Images from https://medium.com/pinata/ethereum-and-ipfs-e816e12a3c59

Konstantin D. Pandl, Research Associate, KIT

Supported by: Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag

Next steps



Icons from flaticon.com

Supported by: Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag