Towards the Medicine of the Future in Bavaria and Germany, One Heartbeat at the Time With Confidential Computing

Florent Dufour, Leibniz Supercomputing Centre - Technical University Munich

Open Confidential Computing Conference, March 15th 2023

~$ whoami

- Florent Dufour 🇫🇷
- Computational Biologist & Data scientist
- AI, Confidential Computing, Data Privacy, and ML-Ops

1. Big data and AI Team @ Leibniz Supercomputing Centre
   - DigiMed Bayern Project
   - Teaching AI, Container Technology, and HPC
2. Ph.D. Student AI in Medicine @ TU Munich

Part I

The Bavarian Cloud for Health Research

The Bavarian Cloud for Health Research

Once Upon a Time in Bavaria...

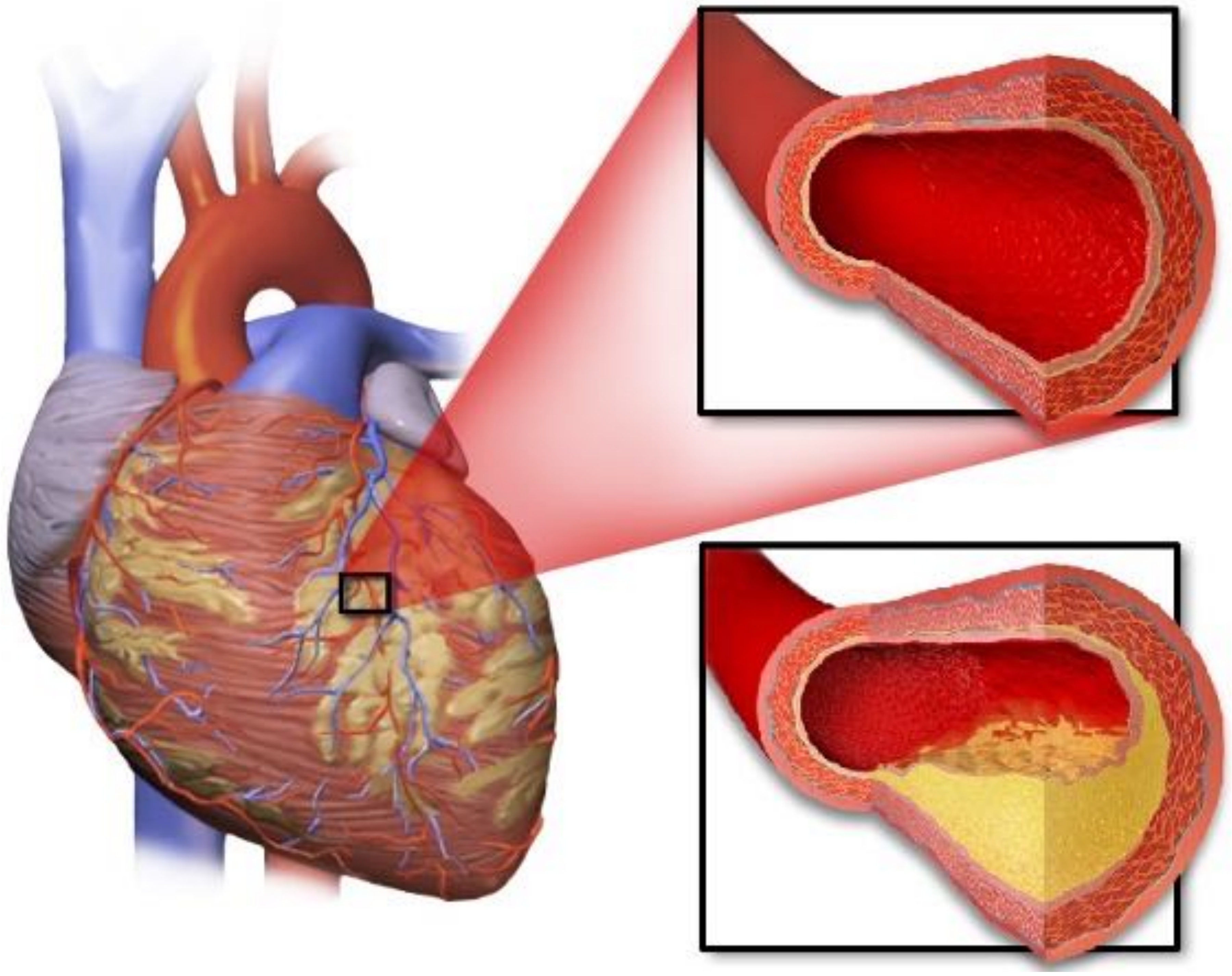
The Bavarian Cloud for Health Research

The Anti-Hero: Cardiovascular Diseases

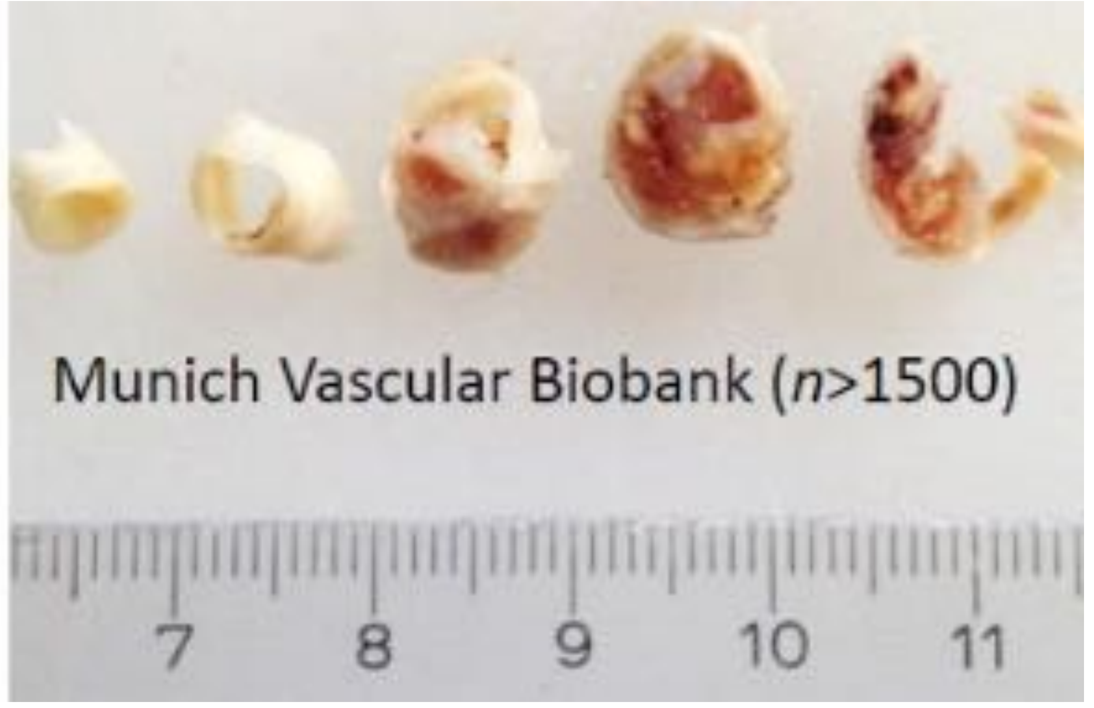
Cardiovascular diseases (CVDs) are the #1 cause of death worldwide with 18 Million deaths in 2019. That represents **32% of all deaths**. Of these, **85% were from heart attacks and strokes** [1]

In Germany, 46,207 (13.4%) and 15,026 (4.4%) people died from myocardial infarction and stroke, respectively, in 2018 [2]

In the EU, CVDs cost €210 billion in 2017 53% health system + 26% lost productivity + 21% informal care [3]



Atherosclerosis: abnormal deposition of cholesterol esters and other fats in the inner wall layer of arterial blood vessels

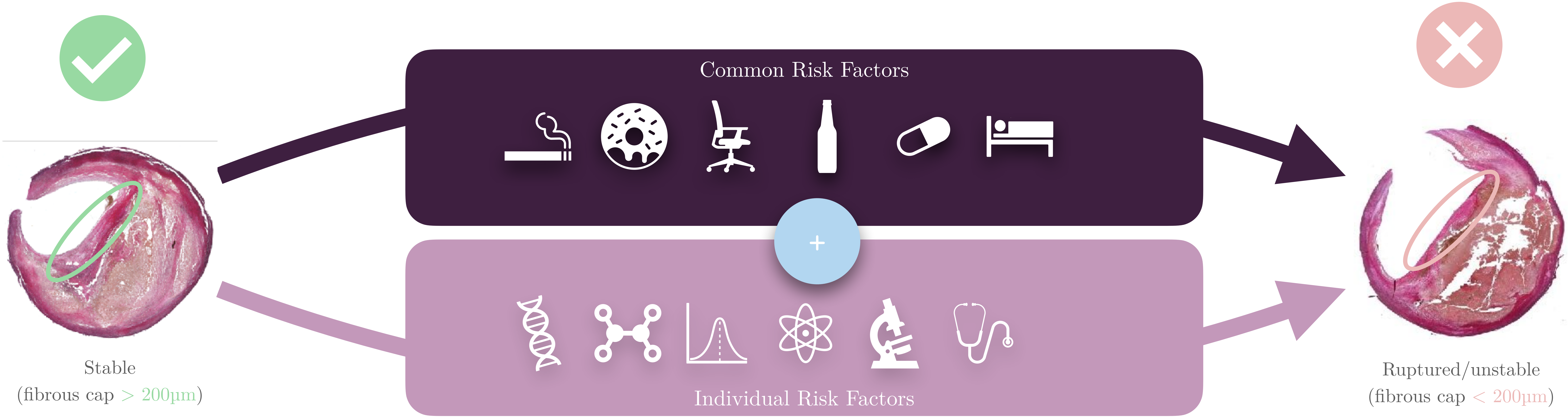


[1] adapted from WHO for 2016 **Mortensen et al., 2019

[2] Statisiches Bundesamt

[3] https://ehnheart.org/cvd-statistics/cvd-statistics-2017.html

The Anti-Hero: Cardiovascular Diseases



The Bavarian Cloud for Health Research

The Genesis: DigiMed Bayern Project

Workpackage	Medicine	Biology	Data	IT	Legal	Healthcare	Society
1. Atheroscl./Heart		+	+	+	+	○	○
2. Stroke	•	+	+	○	○	○	○
3. Fam. Hyp. chol.	•	+	+	○	○	+	+
4. Epidemiology	•	+	+	+	○	○	○
5. Multi-Omics	+	•	+	+	○	○	○
6. IT Infrastructure	○	○	+	•	+	○	○
7. Ethics & Legal	○	○	○	○	•	+	•
8. Project Mngmt & Communic.	+	○	○	+	+	•	•

• Focus + Active ○ Involved

14 institutions

100+ researchers

€25 Million

https://digimed-bayern.de

Gefördert durch

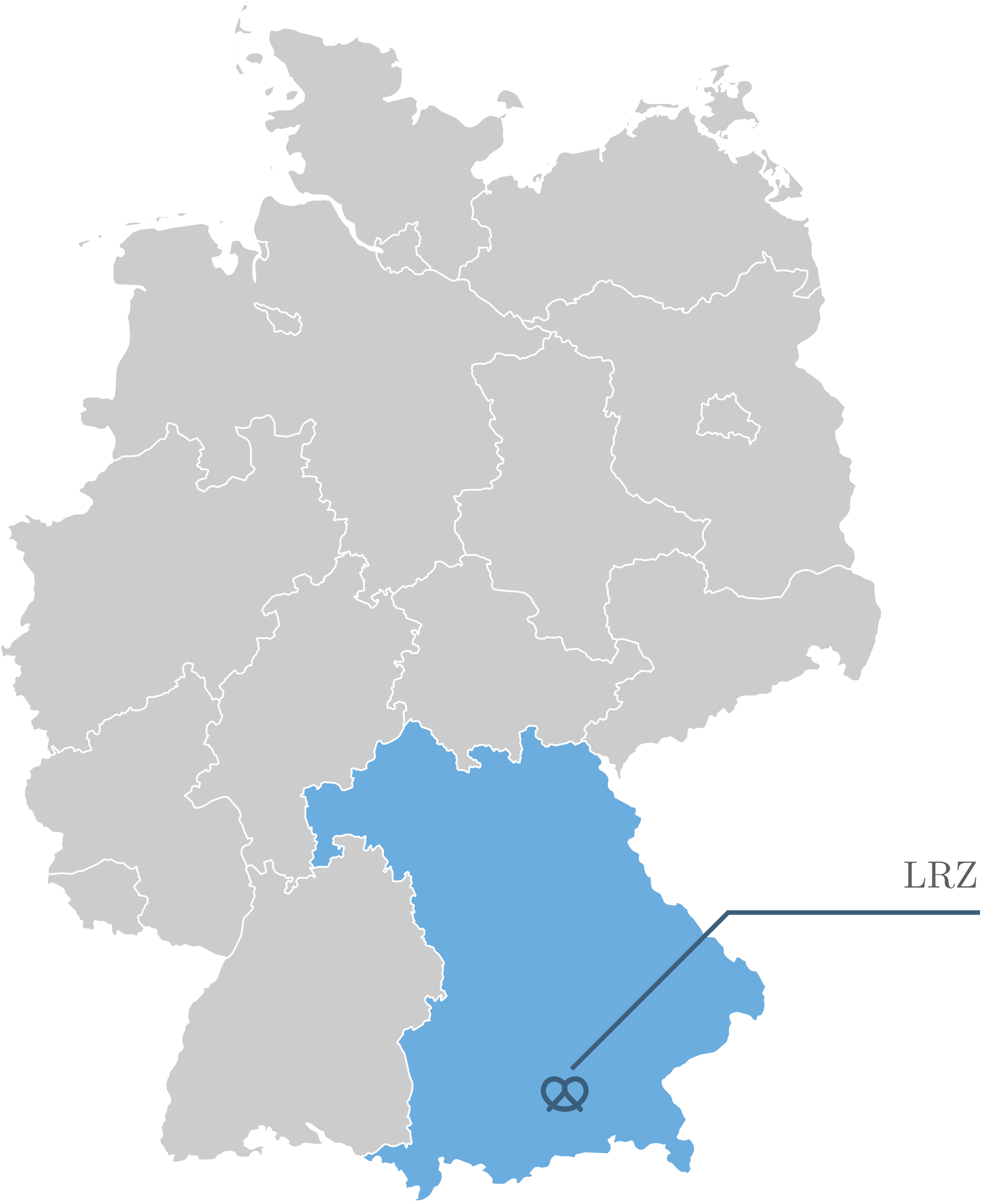
Wissenschaftliche Leitung

Geschäftsführende Leitung

Open Confidential Conference 2023

The Bavarian Cloud for Health Research

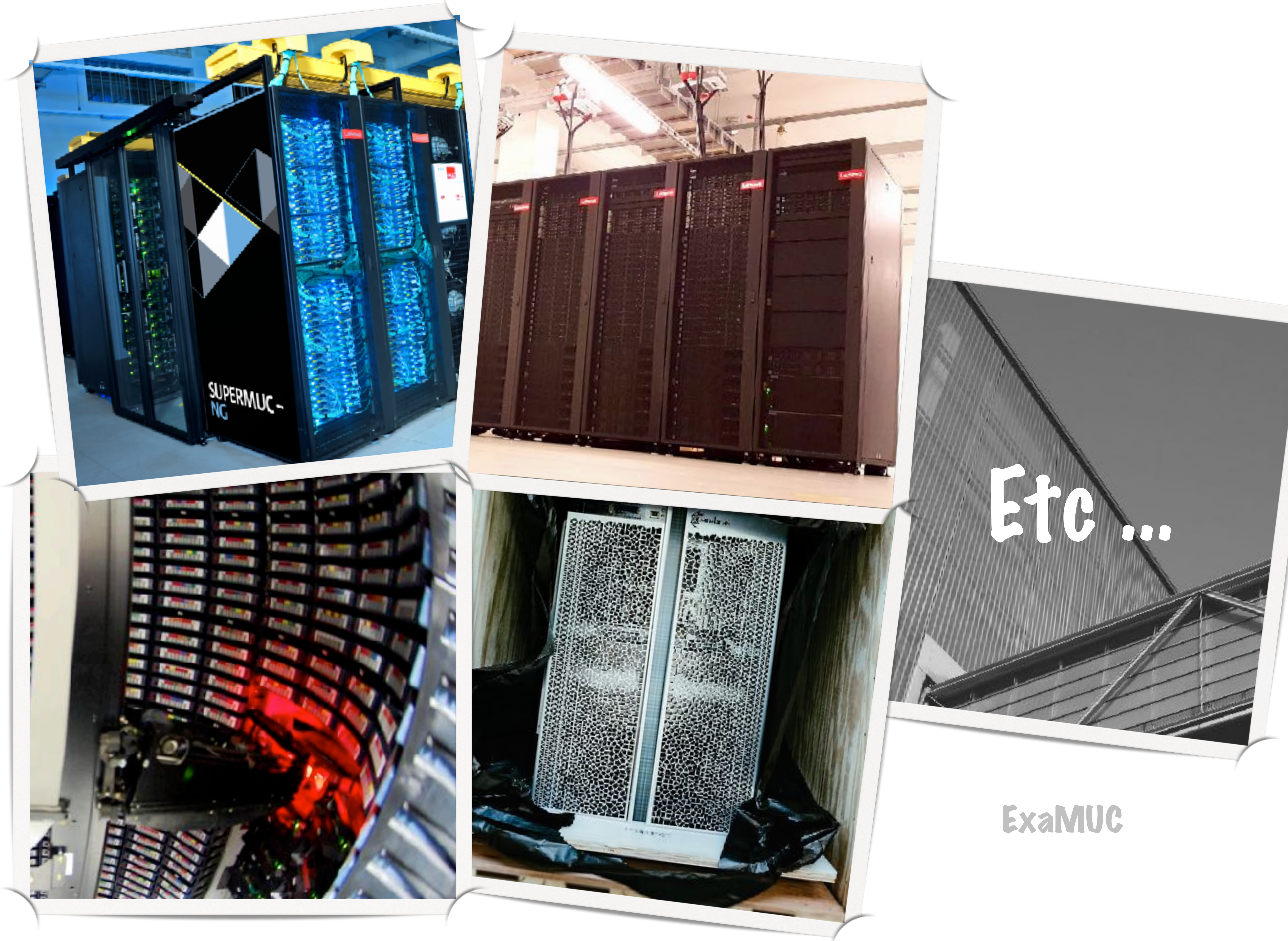
The Leibniz Supercomputing Centre (LRZ)



The Leibniz Supercomputing is located at the North of Munich, Bavaria

SuperMUC-NG

LRZ Compute Cloud



ExaMUC

Data Science Storage & Archive

Future Computing, Artificial Intelligence, and Quantum Computing

The Bavarian Cloud for Health Research

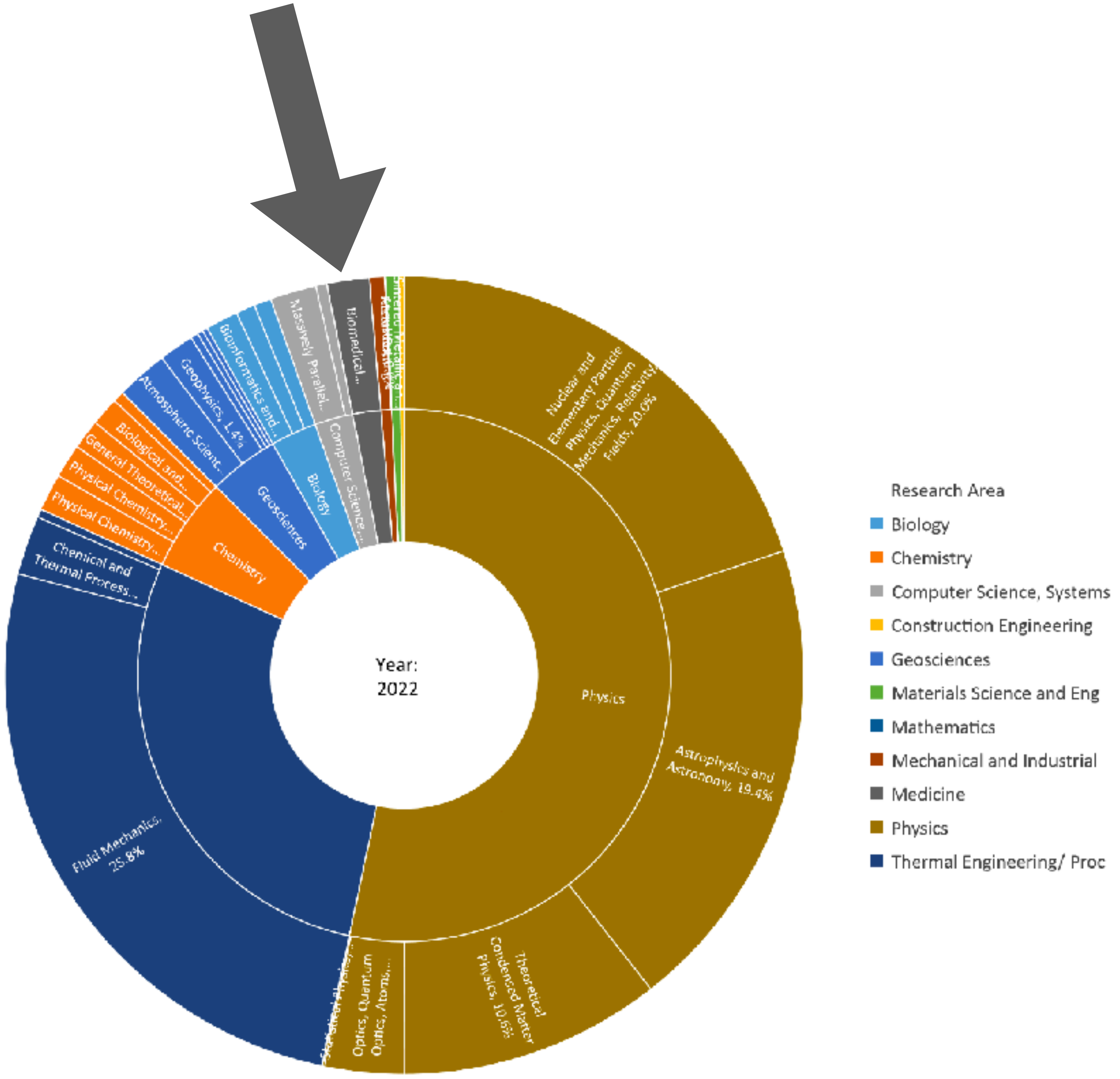
The Leibniz Supercomputing Centre (LRZ)

HPC and AI Resources

- > 7k nodes / ~350k Cores / ~800 TB RAM (SuperMUC-NG + LX)
- > 2000 M core-hour / year
- 70 PB Storage + 260 PB Archives
- ~50 GPUs
- Other accelerators: WSE 2

Scientific Cloud

- OpenStack & CEPH
- 200 Nodes
- 32 × 2 GPUs Nodes
- ~2PB raw storage
- 100G Fabric
- 40000 vCPU capacity with overcommitment
- 2000 users and 1500 active VMs



https://doku.lrz.de
https://lrz.de/hpcbooks

The Bavarian Cloud for Health Research

User Requirements and Decisions

DigiMed Bayern
lrz

Citrix

2FA

We don't know

Not exposed to the internet

Bare Metal

No Encryption

VMs

Expose to the internet

MacBook Pro

Slurm

MySQL

Neo4J

Excel

We don't know

Require collaboration

We don't know

NGS pipelines

Machine Learning

FTP

Jupyter Notebooks

Mongo DB

Windows VNC

Time Series

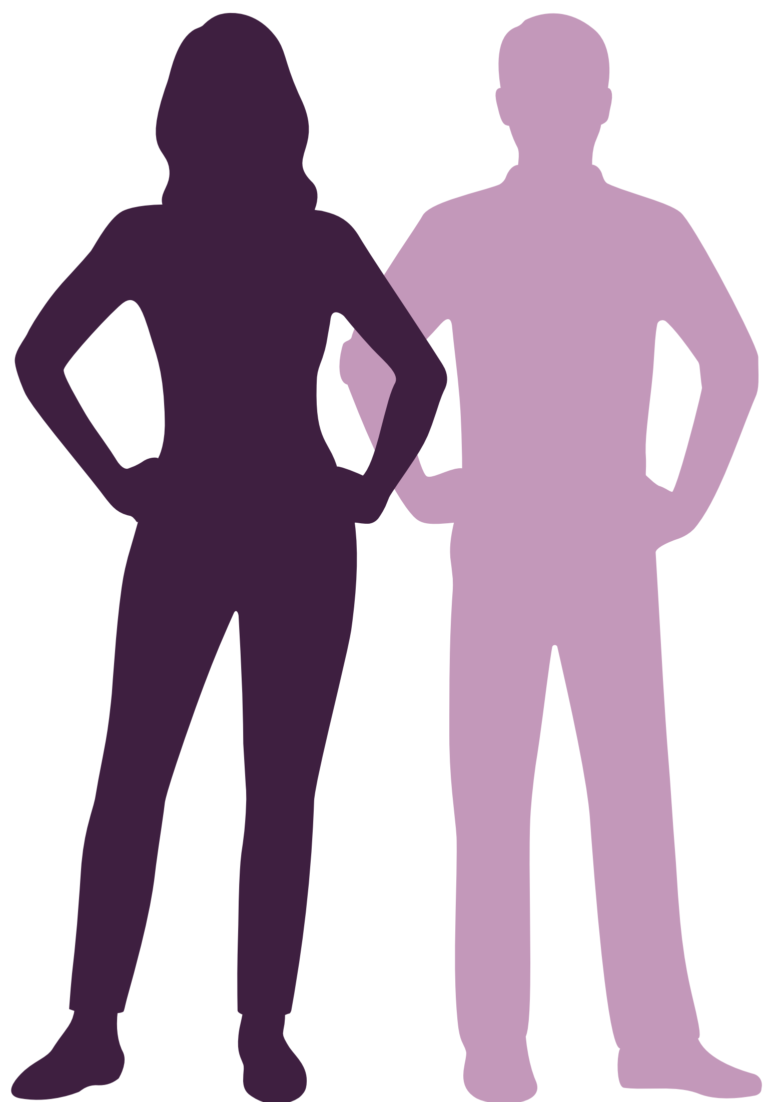
Long running jobs

Data Archive Only

GPUs

Scattered ASCII Files

800 cores



16 Cores

We don't know

Image processing

6.5 PB

7'602 KB

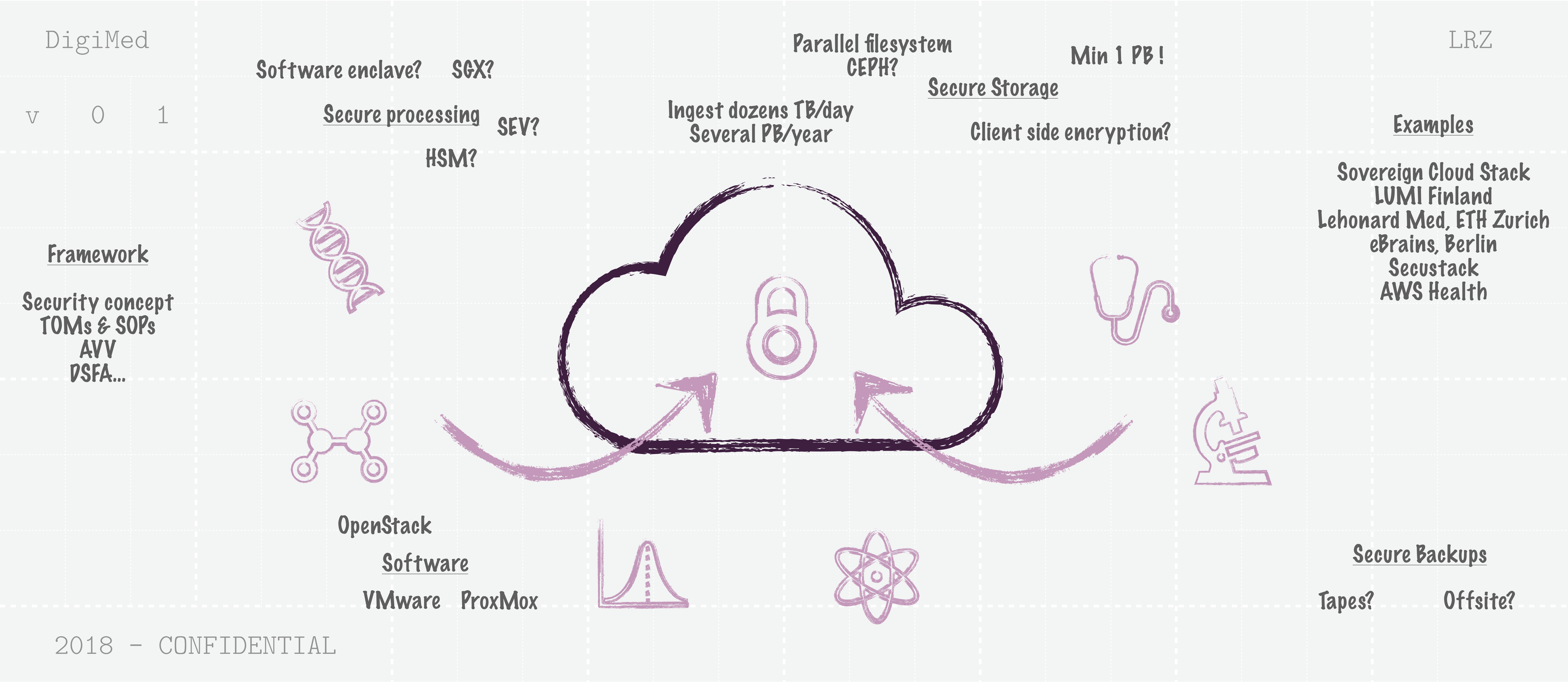
S3 buckets

The Bavarian Cloud for Health Research

User Requirements and Decisions

The Bavarian Cloud for Health Research

Brainstorming



The Bavarian Cloud for Health Research

4 Design Principles for a Community IaaS

DigiMed

LRZ

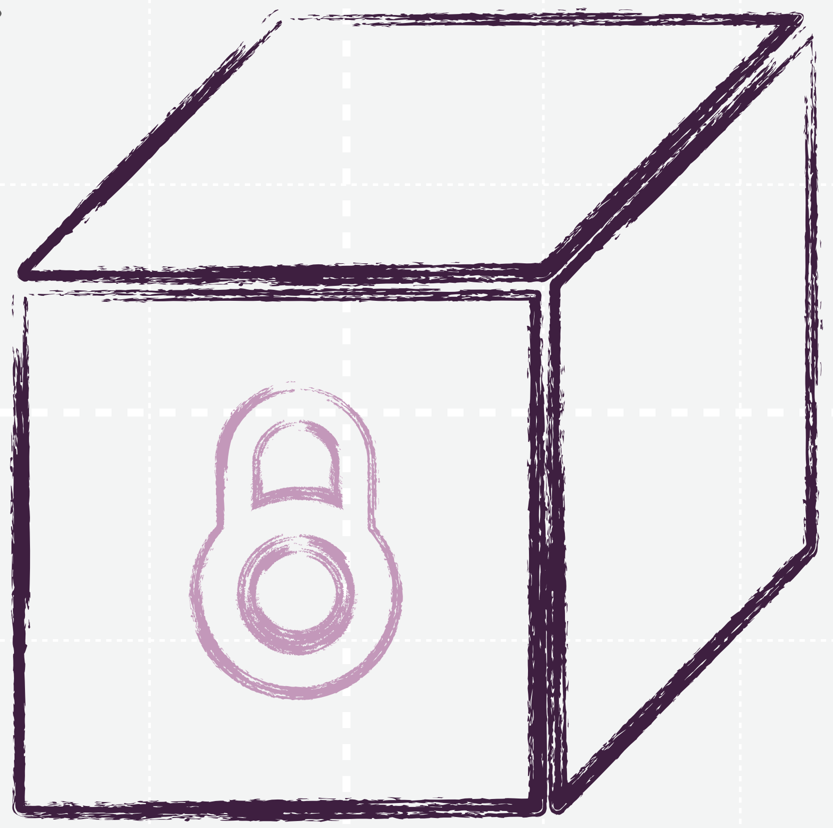
v 0 1

Secure

AMD-SEV/SNP
End to end data encryption
Multi tenancy
Pseudonimized data
...

Low op. overhead

100% Self service
OpenStack and Quobyte
...



Scalable

100G Fabric
Solid management layer
Open source Technologies
...

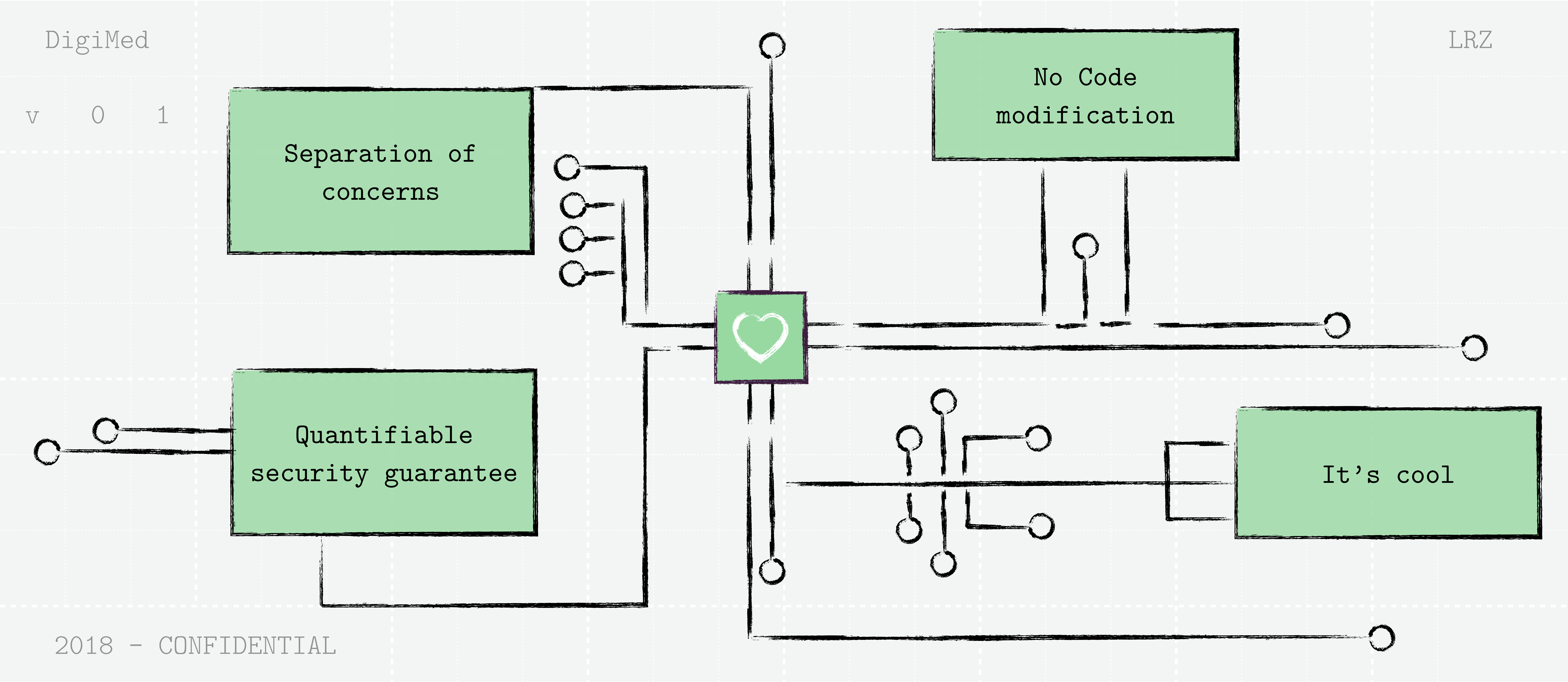
Cost efficient

Commodity Hardware
Scale out Capability
Reduced operation
...

2018 - CONFIDENTIAL

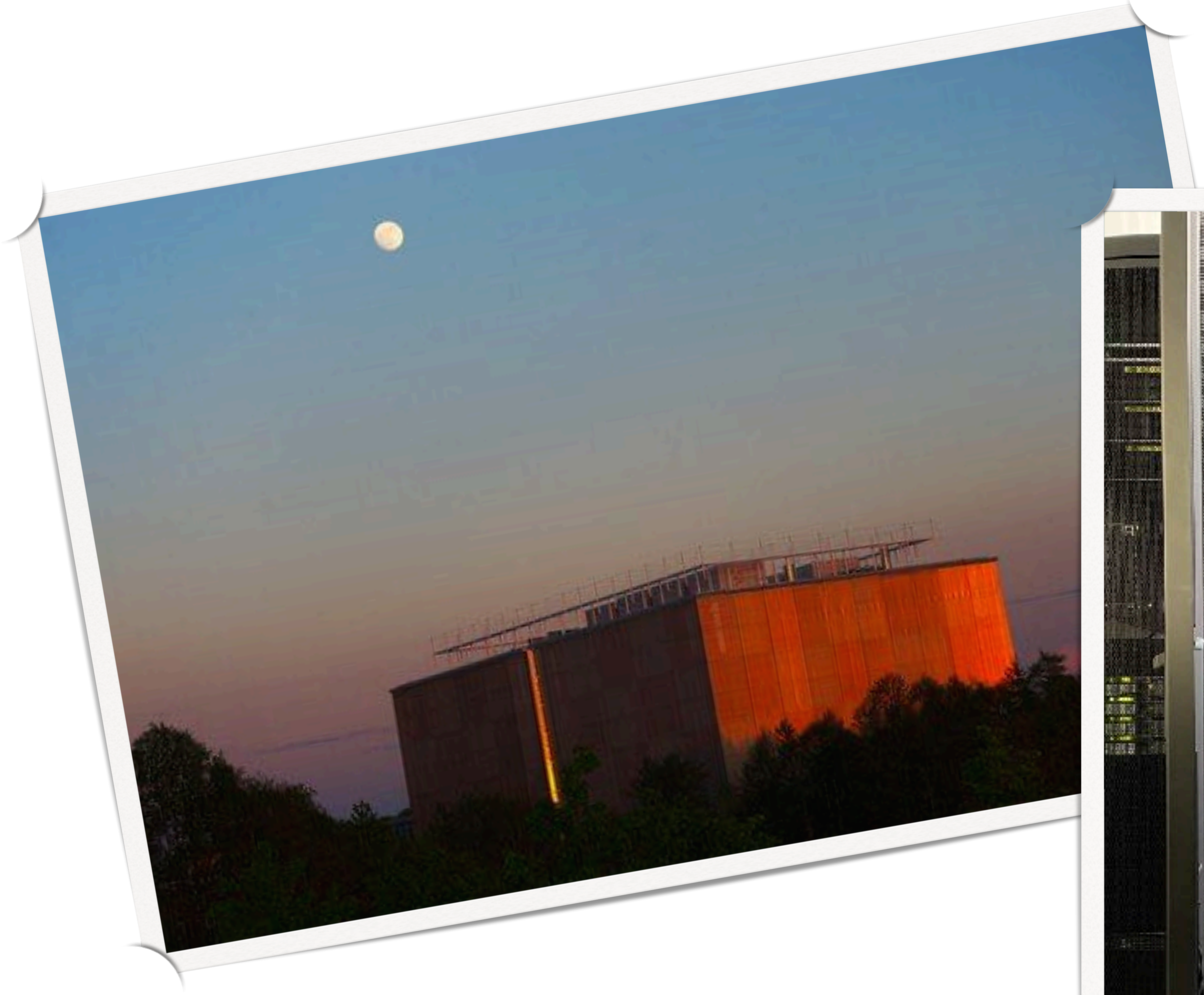
The Bavarian Cloud for Health Research

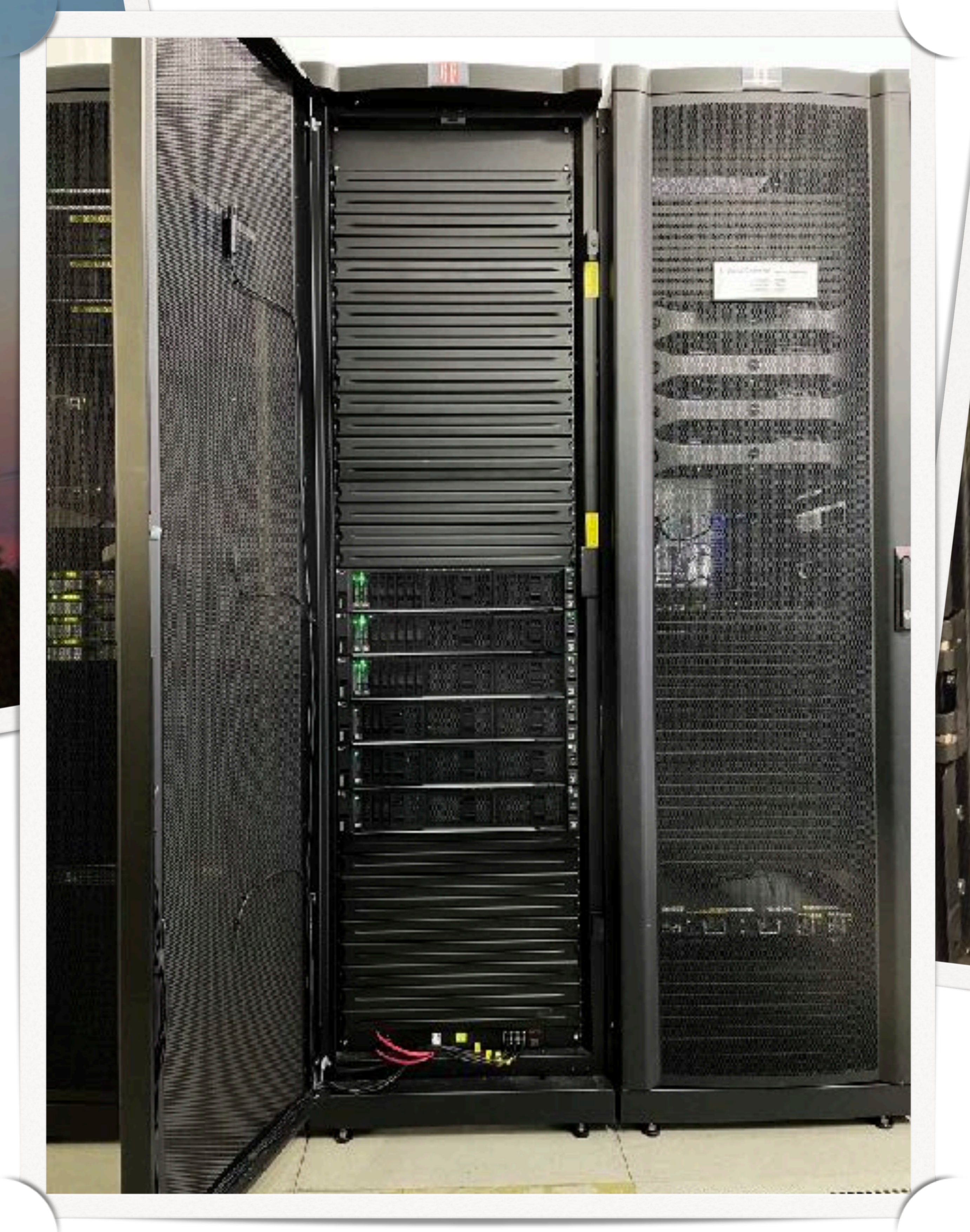
Confidential Computing With AMD-SEV Is the Magic



The Bavarian Cloud for Health Research

A Baby Cloud Was Born





The Bavarian Cloud for Health Research

Quobyte Filesystem

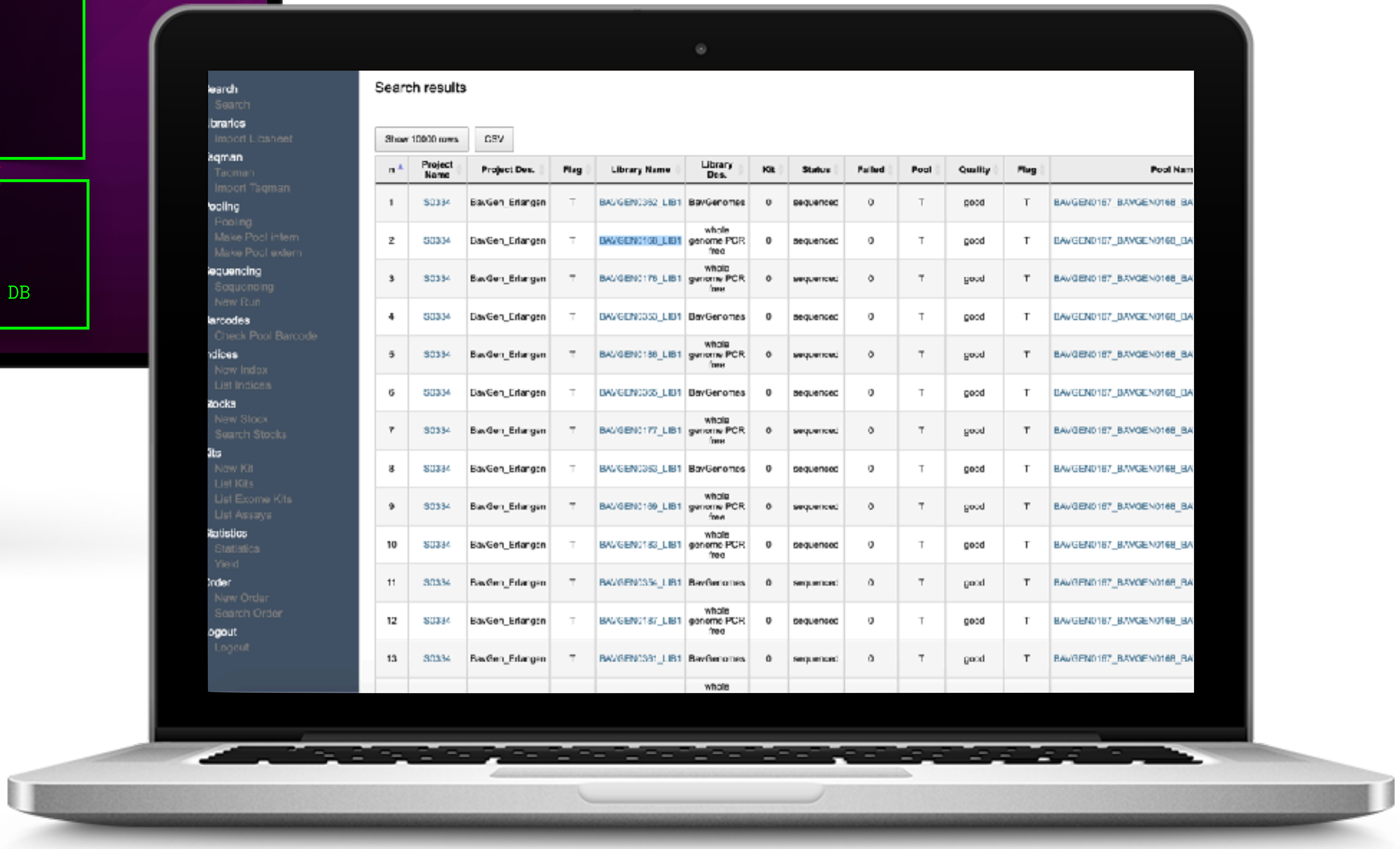
Quobyte is a **distributed filesystem** made for horizontal scale-out providing large-scale data storage and processing. It is suited for scientific research, and data analytics, and cloud computing.

Architecture	Access	Security	Features
1 PB	Quobyte Client	Multi-tenancy	Redundancy
4 Nodes	NFS/SMG/S3	TLS/AES encryption	Stripping
70 Drives	Share volumes	Immutable files	Snapshots
20% - 80% SSD - HDD	OpenStack drivers	Certs/access keys	(Geo replication)
	Hadoop / HDFS	FS event log	

Storage is powered by the Quobyte Filesystem

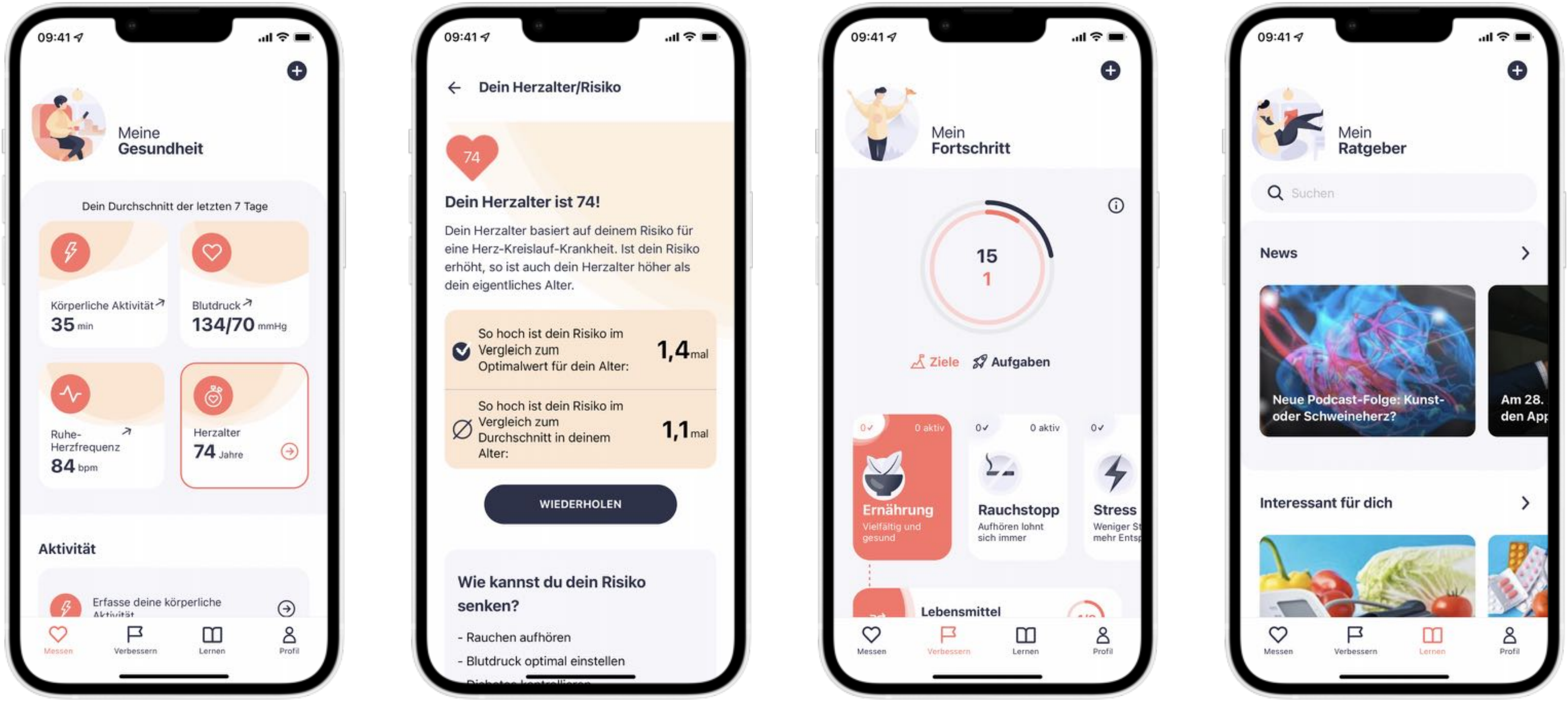
The Bavarian Cloud for Health Research

Example of Workflows: Ent-To-End NGS Pipelines



The Bavarian Cloud for Health Research

Example of Workflows: the HerzFit App



Open Confidential Conference 2023

The Bavarian Cloud for Health Research

Lessons Learned: Moral of the Story

1 Big data = Big Problems

- Surprisingly, not (always) technically
- What is criticality, value, risk ratio of data?
- Chicken and Egg problem before data upload: No framework...
- Solution: Take baby steps (*e.g.*, start with public datasets)

2 Money isn't always the bottleneck

- Difficult to recruit in academia for IT, but we have the money
- Users want to pay for an academic cloud: no competition
- Users need a legal framework, you can't *really* buy it like you would with hardware

3 A cloud doesn't always fly on its own

- Running NGS in the cloud is possible (e.g., lifebit cloudOS)
- (HPC) users remain to convince
- Require pipelines / workflow refactoring
- Require bioinformaticians to become IT people

4 There's never too much paperwork

- > 130 documents to hold the consortium together
- 80% coordination vs. 20% actual hacking
- We'll share as much as we can with the community
- Some wheels need to be re-invented

5 IT is just another form of yoga

- Practice of "letting go"
- Chip shortage / war / covid: Not everything is in your control
- You don't control the users either, you can only educate
- Gap between research and operations: hard to co-design

6 Suffering as grace

- Take everything as a teaching
- Embrace the change
- It can always be worse
- Remain humble

...And They Lived Happily Forever

Part II

Privacy Preserving AI With Confidential Computing

Privacy Preserving AI With Confidential Computing

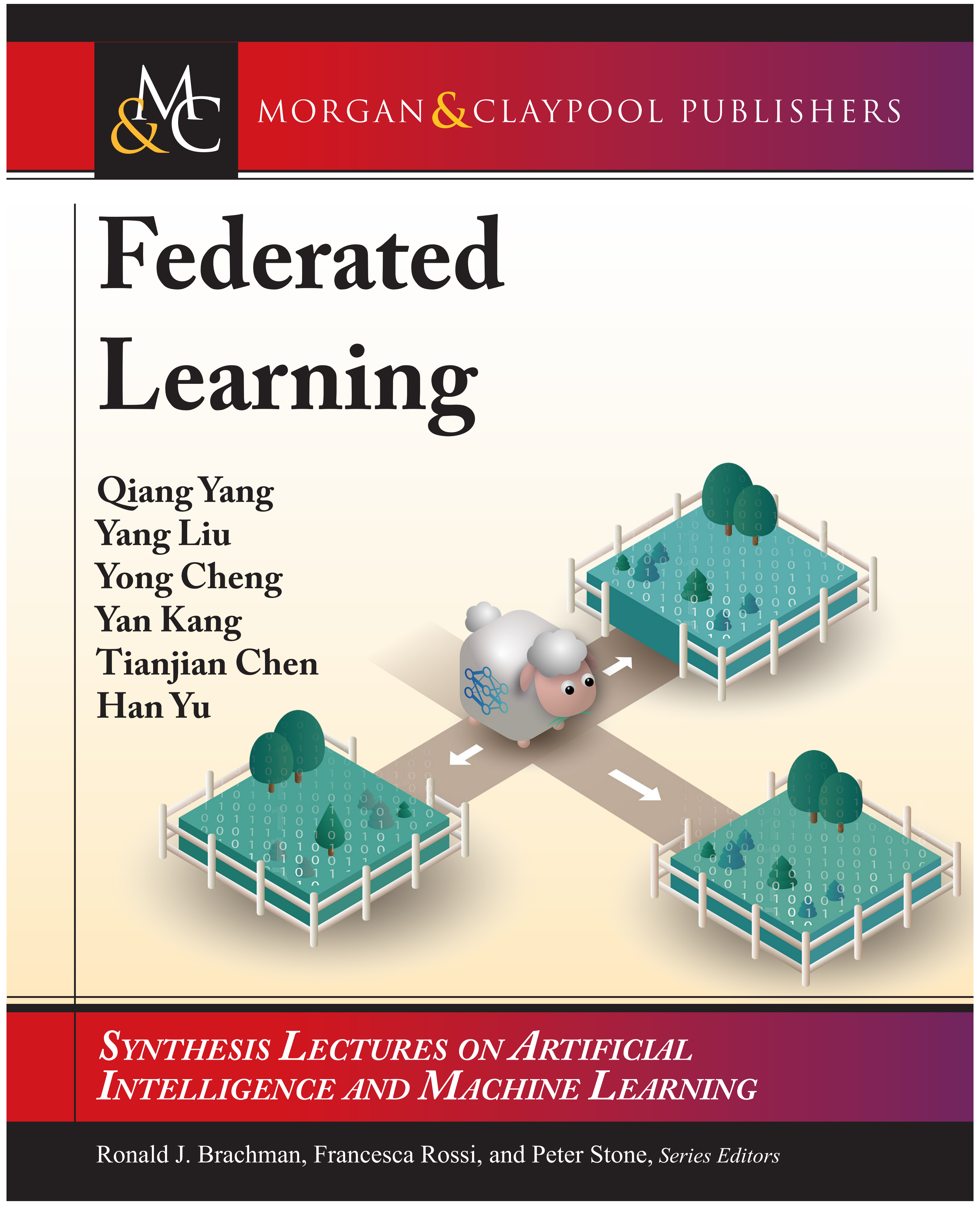
Privacy Preserving AI With Confidential Computing

Federated Learning Allow To Learn on Sensitive Datasets

“ *How is it possible to allow multiple data owners to collaboratively train and use a shared prediction model while keeping all the local training data private?*

“ *Federated machine learning (or federated learning, in short) emerges as a functional solution that can help build high-performance models shared among multiple parties while still complying with requirements for user privacy and data confidentiality.”*

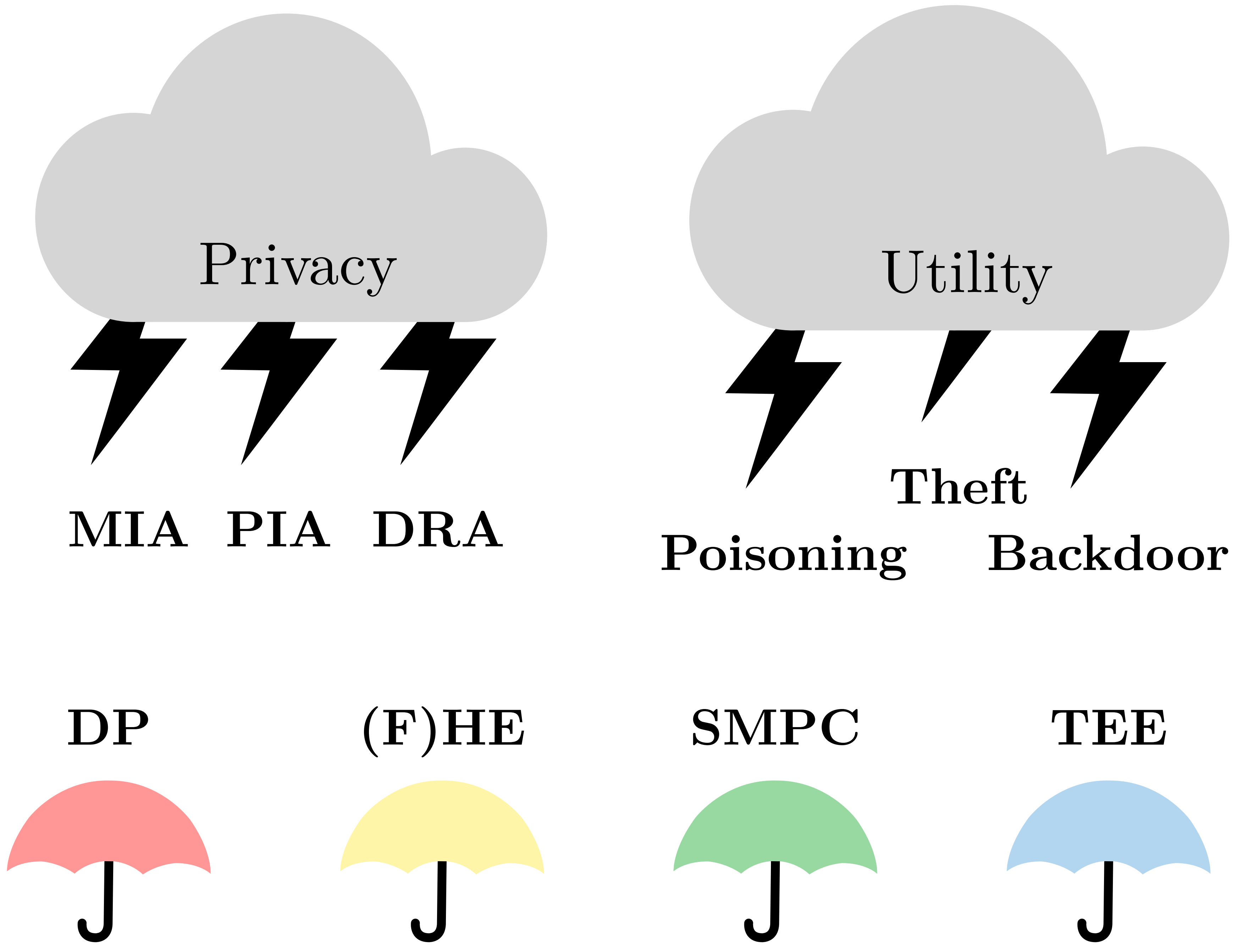
“ *The data is spread across various sites owned by different individuals or organizations, and there is no simple solution to consolidate it. Big data is a crucial element for AI and society, yet we are currently in an era of small, disconnected, and fragmented data silos.*

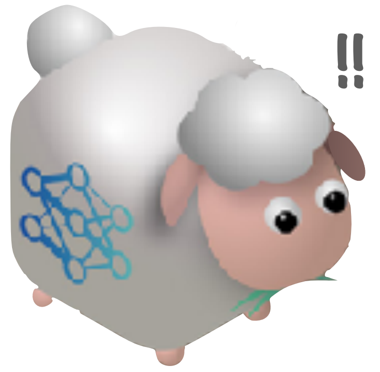


Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, “Federated Learning,” Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 13, no. 3, pp. 1–207, Dec. 2019, doi: 10.2200/S00960ED2V01Y201910AIM043.

Privacy Preserving AI With Confidential Computing

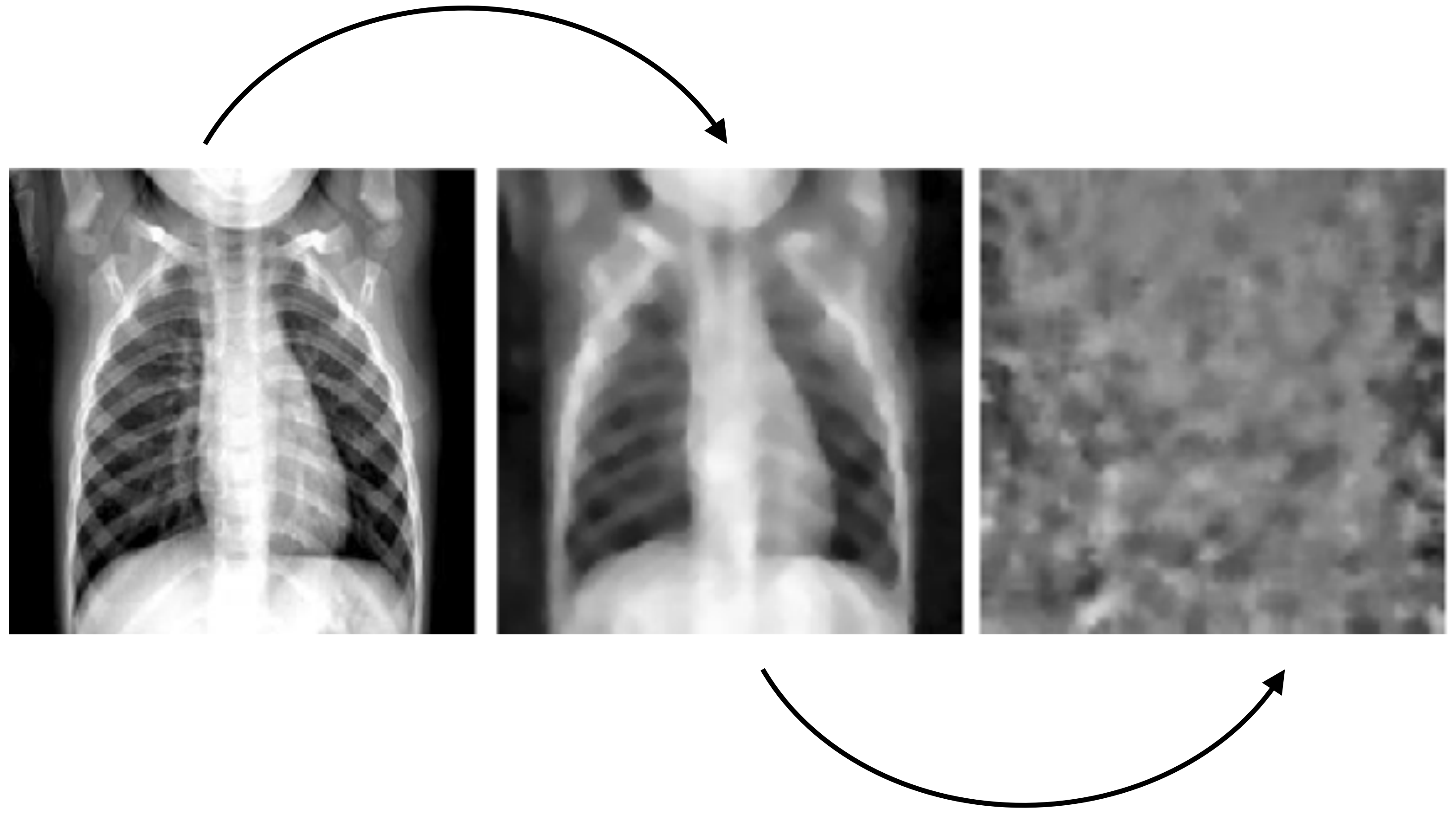
But Federated Learning Doesn't Protect Data Privacy





$$\arg\min_{x'\in[0,1]^n}\left\{1-\frac{\langle \nabla_\theta \mathscr{L}(x,y), \nabla_\theta\mathscr{L}(x',y)\rangle}{\|\nabla_\theta\mathscr{L}(x,y)\|_2\cdot\|\nabla_\theta\mathscr{L}(x',y)\|_2}\right\}$$

Where x' is the reconstruction target, x is the ground truth, y is the label, $\nabla_\theta\mathscr{L}$ is the gradient with respect to the weights, $\langle\cdot\rangle$ is the inner product in $\mathbb{R}^n$ and $\|\cdot\|_2$ is the L_2-norm. α is a hyperparameter scaling the total variation penalty over the image, $\mathrm{TV}(x)$



$$\mathbb{P}\left(\mathscr{M}(q(D))\in S\right)\le e^{\epsilon}\times\mathbb{P}(\mathscr{M}(q(D'))\in S)+\delta$$

(ϵ,δ)-DP: A mechanism $\mathscr{M}$ is (ϵ,δ)-DP iff, for all $D \equiv D'$ and all subsets S of the co-domain of $\mathscr{M}$, when a query function q is executed, the above holds

Poisoning example: An image with a 16 × 16 backdoor patch.

Predict

Eiffel Tower

G. Kaissis *et al.*, "End-to-end privacy preserving deep learning on multi-institutional medical imaging," *Nat Mach Intell*, vol. 3, no. 6, pp. 473–484, Jun. 2021
Zhu, Ligeng, Zhijian Liu, and Song Han. "Deep leakage from gradients." Advances in neural information processing systems 32 (2019).
D. Usynin *et al.*, "Adversarial interference and its mitigations in privacy-preserving collaborative machine learning," *Nat Mach Intell*, vol. 3, no. 9, Art. no. 9, Sep. 2022
N. Carlini and A. Terzis, "Poisoning and Backdooring Contrastive Learning," 2022.
L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," in *Advances in Neural Information Processing Systems*, 2019, vol. 32. Accessed: Mar. 14, 2023
C. Dwork, "Differential Privacy," in Automata, Languages and Programming, Berlin, Heidelberg, 2006, pp. 1–12. doi: 10.1007/11787006_1.
N. Carlini and A. Terzis, "Poisoning and Backdooring Contrastive Learning," 2022.

Privacy Preserving AI With Confidential Computing

Confidential Computing for Private and Secure AI

Data/Model lifespan

Complexity of AI workloads	+	TEE everywhere (client & server)	=	TEEs to the win
Heterogeneous Architectures		Low Performance overhead		
Evolution to the edge		Ever increasing resource protection		
Many attack vectors		Doesn’t reduce model utility		
		Protect model in time and space		

Privacy Preserving AI With Confidential Computing

Current Ph.D. Directions

1 **TEEs against model poisoning**

- Secret Provisioning and Attestation
- With Zero knowledge
- Attest against model poisoning

2 **TEEs with Differential Privacy**

- Use TEEs and DP in concerts
- Reduce noise need to protect the model
- Attest the privacy guarantee

3 **TEEs with GPUs**

- New generation of GPU that support TEEs
- Develop a new accelerated Federated Learning framework

4 **TEEs for Explainable AI**

- TEEs to provide reproducible and accountable decision.
- Required healthcare

Acknowledgements

BioM

Prof. Dr. Horst Domdey
Dr. Jens Wiehler
Anja Kroke
Dr. Ruoyu Sun
All the DigiMed colleagues

LRZ

Prof. Dr. Dieter Kranzlmüller
Dr. Nicolay Hammer
Dr. Peter Zinterhof
Dr. Naweiluo Zhou
Vinzent Bode

TUM

Prof. Dr. Daniel Rückert
Dr. Georgios Kaissis
All the Ph.D. Students

🖥️ du4.link/oc3
💌 florent@lrz.de
🔑 A57E 6345 F0DE 3B3E DB2D 406B 58DD B727 9BF1 44F6

We kindly acknowledge the support of the Bavarian State Ministry of Health and Care who funded this work with DigiMed Bayern (Grant No: DMB-1805-0001) within its Masterplan “Bayern Digital II”