

Kathrin Sobe – ZMI – PARS-Workshop 2026

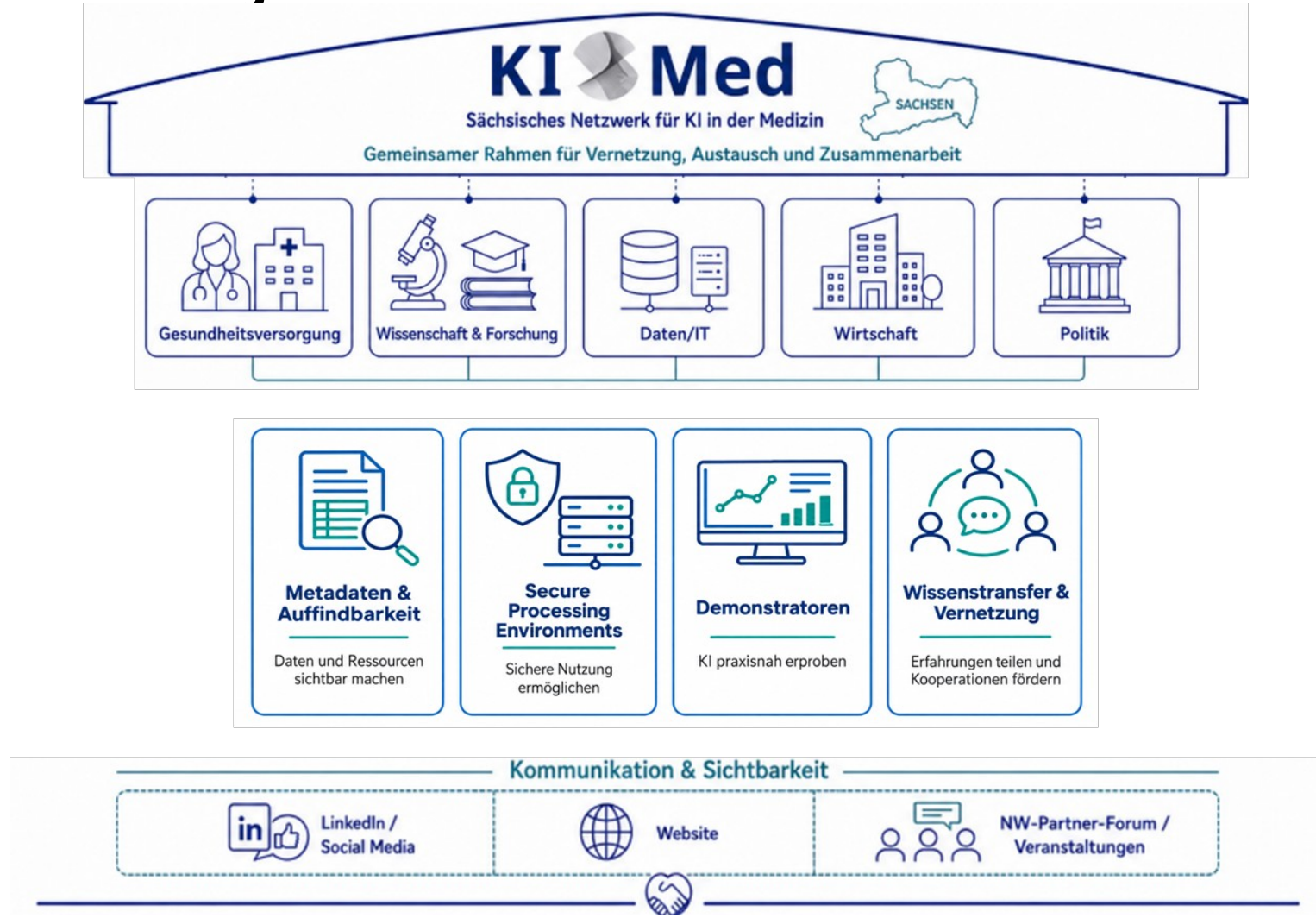
Enabling Secure HPC processing for medical research in Saxony

Experiences on utilizing the ZIH secure workflow infrastructure for processing medical data – an exemplary application

Overview

- Motivation: Project KIMed
- Secure processing environments: SPEs/TREs
- ZIH Secure Workflow
- Exemplary use case
- Other secure processing environments
- Conclusion and Outlook

Motivation: Project KIMed



Challenges in using health data for research

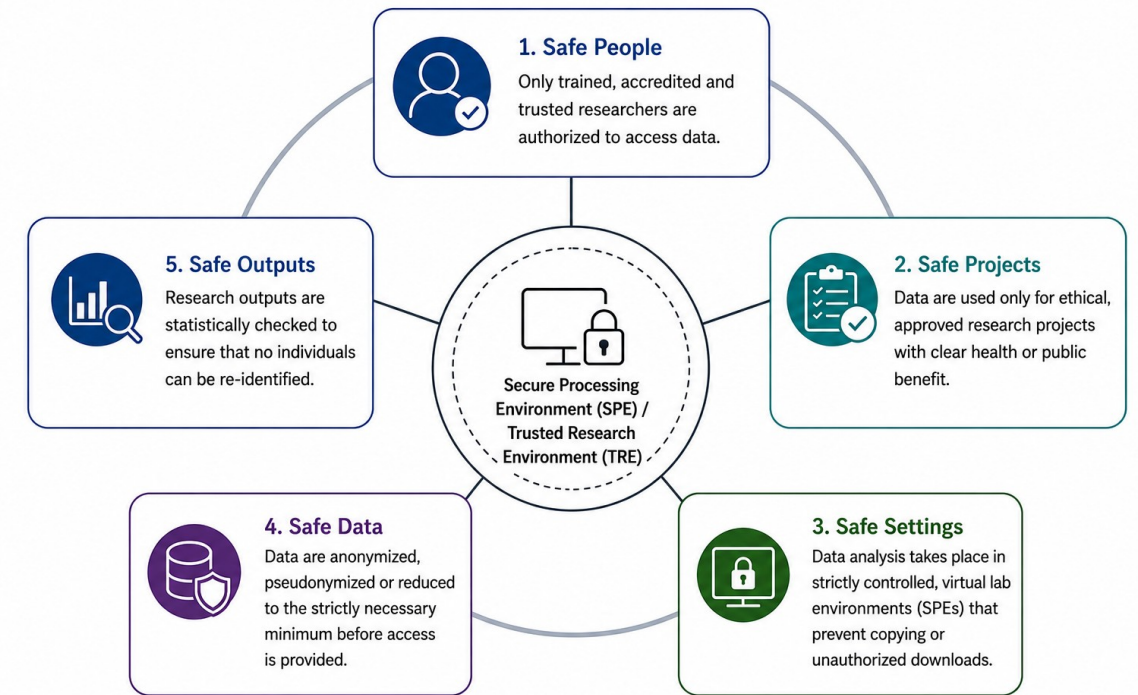
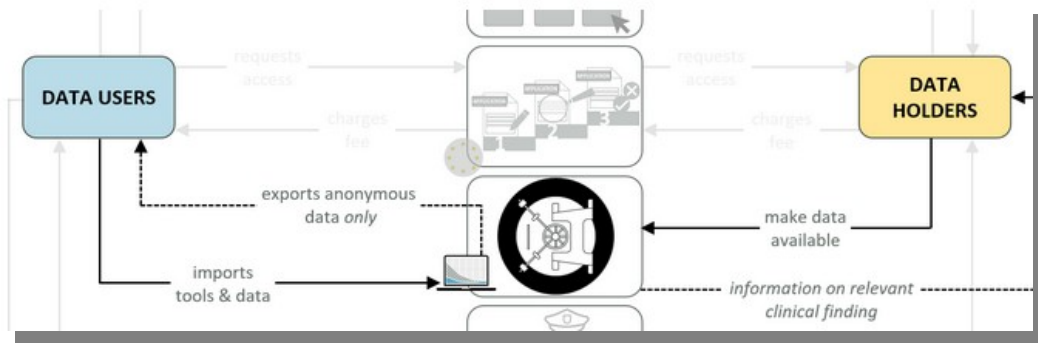


Made with Napkin

Secure processing environments (SPE)

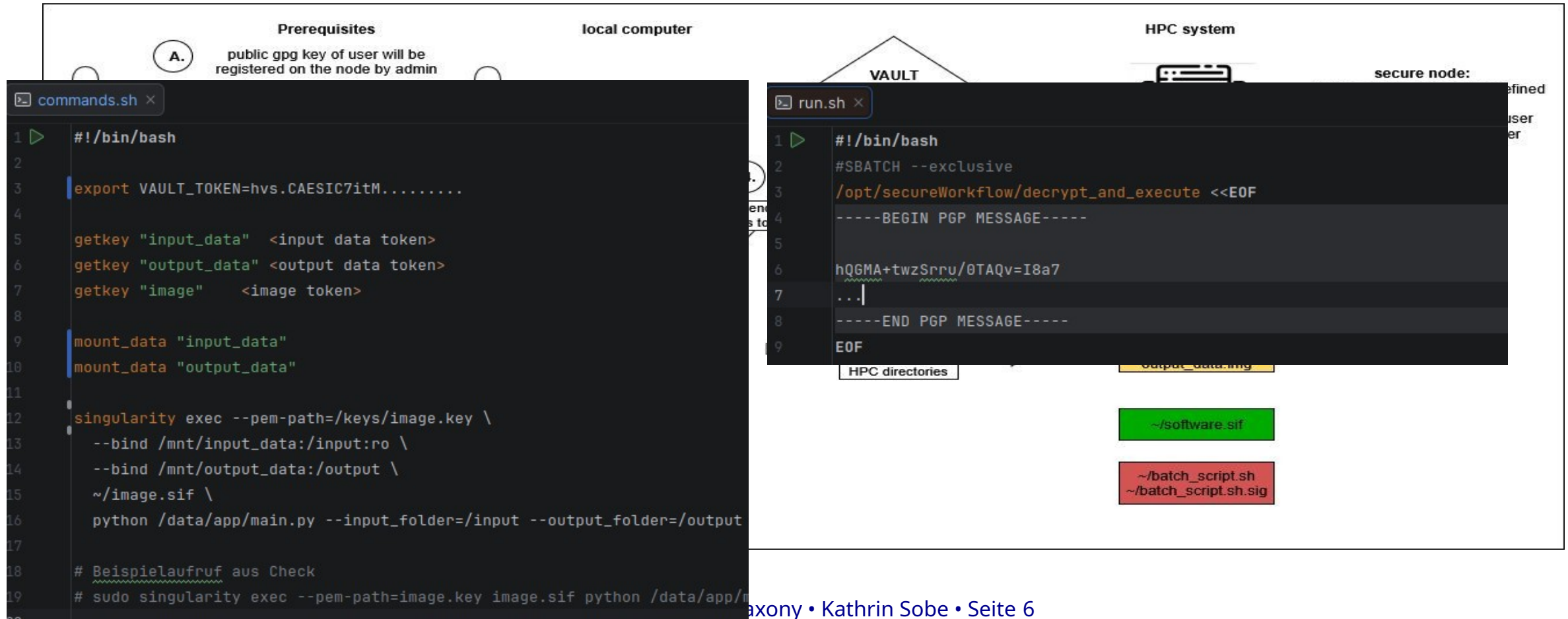
Trusted research environments (TRE), ...

- Sensitive data not handed out to researcher
- Processing in a secure environment
- Governance framework: **5 Safes**



Secure Workflow @ ZIH Dresden

- Based on Nolte 22 „A Secure Workflow for Shared HPC Systems“



Secure Workflow: user guide

 **Mention that `sudo` is usually not required for apptainer builds**
Alexander Grund authored 1 month ago

8033b090  History

user_guide.md 39.74 KiB

Code Preview Blame   

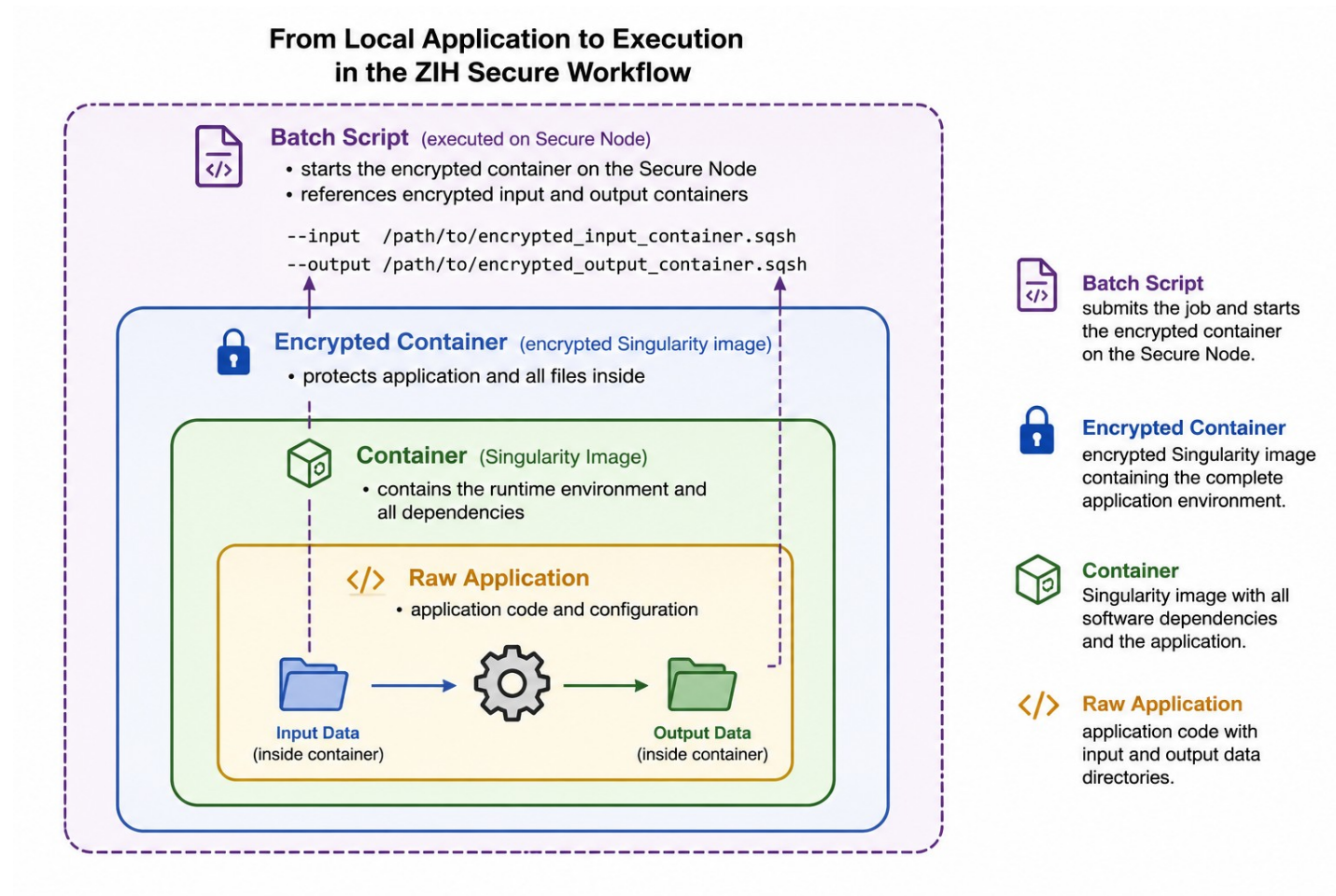
HPC Secure Workflow - user guide

This guide describes a secure, container-based workflow for processing sensitive data on an HPC (High Performance Computing) system. Basic knowledge of Linux and HPC usage is helpful in understanding this guide.

Table of Contents

- [Overview](#)
- [Requirements](#)
- [Workflow overview](#)
- [Used terms](#)
- [Prerequisites \(One-Time Setup\)](#)
 - [A. Import and generate the GPG keys](#)
 - [B. Vault user token](#)
 - [C. Passwordless access to the HPC system](#)
- [Workflow Steps \(Per Job\)](#)
 1. [Create data container](#)
 2. [Create Apptainer image](#)
 3. [The script](#)
 4. [Vault set up and keys](#)
 5. [HPC set up and data transfer](#)
 6. [Run the script](#)
 7. [Copy the results to your local computer and access them](#)
- [Checklist of files and keys](#)
- [Appendix](#)
 - [A. Additional useful Slurm parameters](#)
 - [B. Use Vault installed in an Apptainer image](#)
 - [C. Sudo privileges](#)
 - [D. Apptainer images](#)

Application preparation on local workstation



Application preparation on local workstation

```
<commands.sh>
```

```
#!/bin/bash
```

```
getkey "input_data" <input data token>
getkey "output_data" <output data token>
getkey "image" <image token>
```

```
mount_data "input_data"
mount_data "output_data"
```

```
singularity exec --pem-path=/keys/image.key \
  --contain \
  --writable-tmpfs \
  --bind /mnt/input_data:/input:ro \
  --bind /mnt/output_data:/output \
  ~/image.sif \
  python /data/app/main.py --input_folder=/input --output_folder=/output --num_epochs=25 --batch_size=128
```

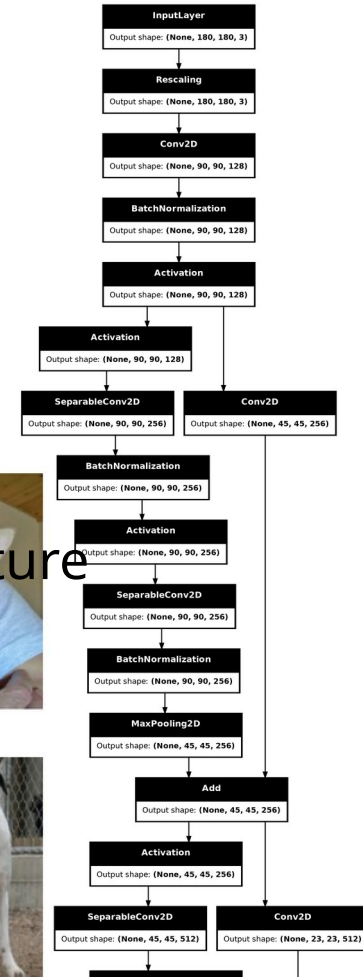
```
sudo singularity exec --pem-path=image.key image.sif python /data/app/main.py --input_folder=/data/dataset --output_folder=$PWD/result_models --num_epochs=1 --batch_size=32
```

```
singularity exec singularity_container.sif python /data/app/main.py --input_folder=/data/dataset --output_folder=$PWD/result_models --num_epochs=1 --batch_size=32
```

```
/local_home/.venv/bin/python main.py --input_folder=ImagesSmall --output_folder=results --batch_size=32 --num_epochs=1
```

Details image classification model

- Model training: tensorflow + keras, example code:
https://keras.io/examples/vision/image_classification_from_scratch/
- Kaggle data set: 23398 files (18719 training, 4679 validation)
- Xception (Extreme Inception): deep convolutional neural network CNN architecture
- Parameters: Batch size (64), num epochs (25)
- Training results:
 - Epoch 1/25: acc: 0.6736 - loss: 0.5857 - val_acc: 0.4960 - val_loss: 0.6946
 - Epoch 25/25: acc: 0.9676 - loss: 0.0800 - val_acc: 0.9457 - val_loss: 0.1273



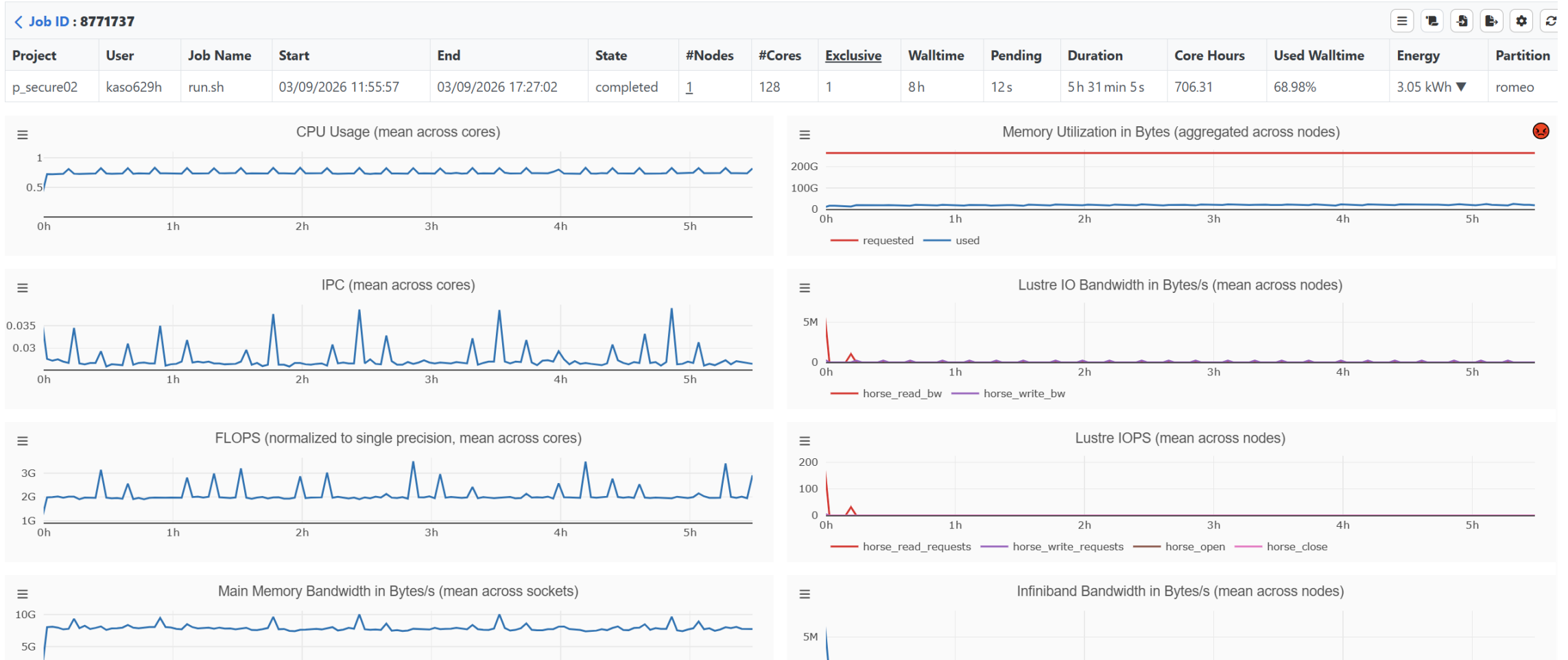
HPC run on ZIH Romeo cluster (1 isolated node)

Romeo

The cluster `Romeo` is a general purpose cluster by NEC based on AMD Rome CPUs.

- 188 nodes, each with
 - 2 x AMD EPYC CPU 7702 (64 cores) @ 2.0 GHz, Multithreading available
 - 512 GiB RAM (8 x 32 GiB DDR4-3200 MT/s per socket)
 - 181.4 GiB local storage on SSD at `/tmp`
- Login nodes: `login[1-2].romeo.hpc.tu-dresden.de`
- Hostnames: `i[7001-7186].romeo.hpc.tu-dresden.de`
- Operating system: Rocky Linux 9.6
- Further information on the usage is documented on the site [CPU Cluster Romeo](#)

Monitoring run on Romeo node (128 cores)



Related environments

SPE	Benutzer	Daten	HPC	Analyse
 Datenhotel UKE Hamburg Eppendorf	internal	internal	no	interactive
 UKE new SPE (NFDI4Health)	external	external	scalability?	n/a
 Charite	internal	internal	no	interactive
 Secure Workflow: GWDG, ZIH, URZ TRE	internal	bring your data	yes	command line
 GöTRE	n/a	n/a	yes	n/a
 Chorus Software	n/a	internal	possible	interactive

Productive access service: HDRUK portal

Health Data Research Gateway > Search > Community > News > About > Help > Data Custodian Area

HS HDRUK

Search the Gateway

Search, for example, by Dataset, SNOMED code, LOINC term, Read code, ICD10 code, MESH term, keywords
[How to search the Gateway](#)

Q Enter your search term

Datasets & BioSamples Data Uses / Research Projects Analysis Scripts & Software Publications Data Custodians Collections / Networks

Filters

There are more than 100 results. We recommend refining your search.

Showing 1-25 of 1288 results.

Sort by most relevant

Download results Save search Switch to list view

Source	Action search result	Metadata title	Data Custodian	Population Size	Date Range	Access Service	Data Standard	Cohort Discovery	BioSample availability	Structural Metadata
Health Data Research Gateway 1288	Actions	SCOT translational sample collection	Tissue Directory	Not reported	2008-	Varies based on project	OTHER	-	✓	-
Partner portals 290	Actions	ICU RTIs	Wessex Secure Data Environme...	Not reported	2018-2026	TRE/SDE	NHS DATA DICTIONARY	-	-	-
☐ Datasets with BioSamples	Actions	LIVING: Transforming supportive care for people living with lung cancer when yo...	Royal Marsden NHS Foundatio...	43	2025-2026	Varies based on project	LOCAL	-	-	-
☐ Datasets in Cohort Discovery	Actions	Sussex Integrated Dataset - Primary Care Encounter	Kent, Medway and Sussex Secure...	2.070.000	2000-2026	TRE/SDE	OMOP, LOCAL	-	-	✓
Data Custodian	Actions	BiB4All SystemOne HV Primary Care Medication	Born in Bradford	Not reported	1990-2024		LOCAL	-	-	✓
Dataset type	Actions	Real Time Data Network (RTDN)	DATA-CAN	10.000	2020-	TRE/SDE	OTHER	-	-	✓
Geographical coverage	Actions	Sussex Integrated Dataset - Primary Care Patient Demographics	Sussex ICB	Not reported	2025-	null	NHS DATA DICTIONARY, LOCAL	-	-	✓
Data standard	Actions	Civil Registration of Births and Deaths Dataset - North East and North Cumbria	North East and North Cumbri...	718.836	2013-2024		NHS DATA DICTIONARY	-	-	-
Data Custodian Network	Actions	Pharmacy	West Midlands Secure Data...	12.000.000	2020-2024	TRE/SDE	LOCAL	-	✓	Need support?
Collection name	Actions	National Audit of Care at the End of Life-				Varies based on				

healthdatagateway.org/dataset/1903

Conclusions + Outlook

Secure workflow is productive

SPEs under high development

KIMed focus SPE for saxony

Closing the regulatory gap for medical researchers

Support AI developments in medicine

<https://www.kimed-netzwerk.de/de/>

KI Med
HERZLICH WILLKOMMEN IM
Netzwerk für Künstliche Intelligenz in der Medizin
Jetzt Partner werden

KI in Medizin sicher, vernetzt
KIMed ist ein interdisziplinäres Forum für Künstliche Intelligenz (KI) in der Medizin. Internationale Akteure aus Forschung und Praxis kommen zusammen. Sie arbeiten mit medizinischen Experten zusammen. Werden Sie Teil unseres Netzwerks!
Mehr zu KIMed

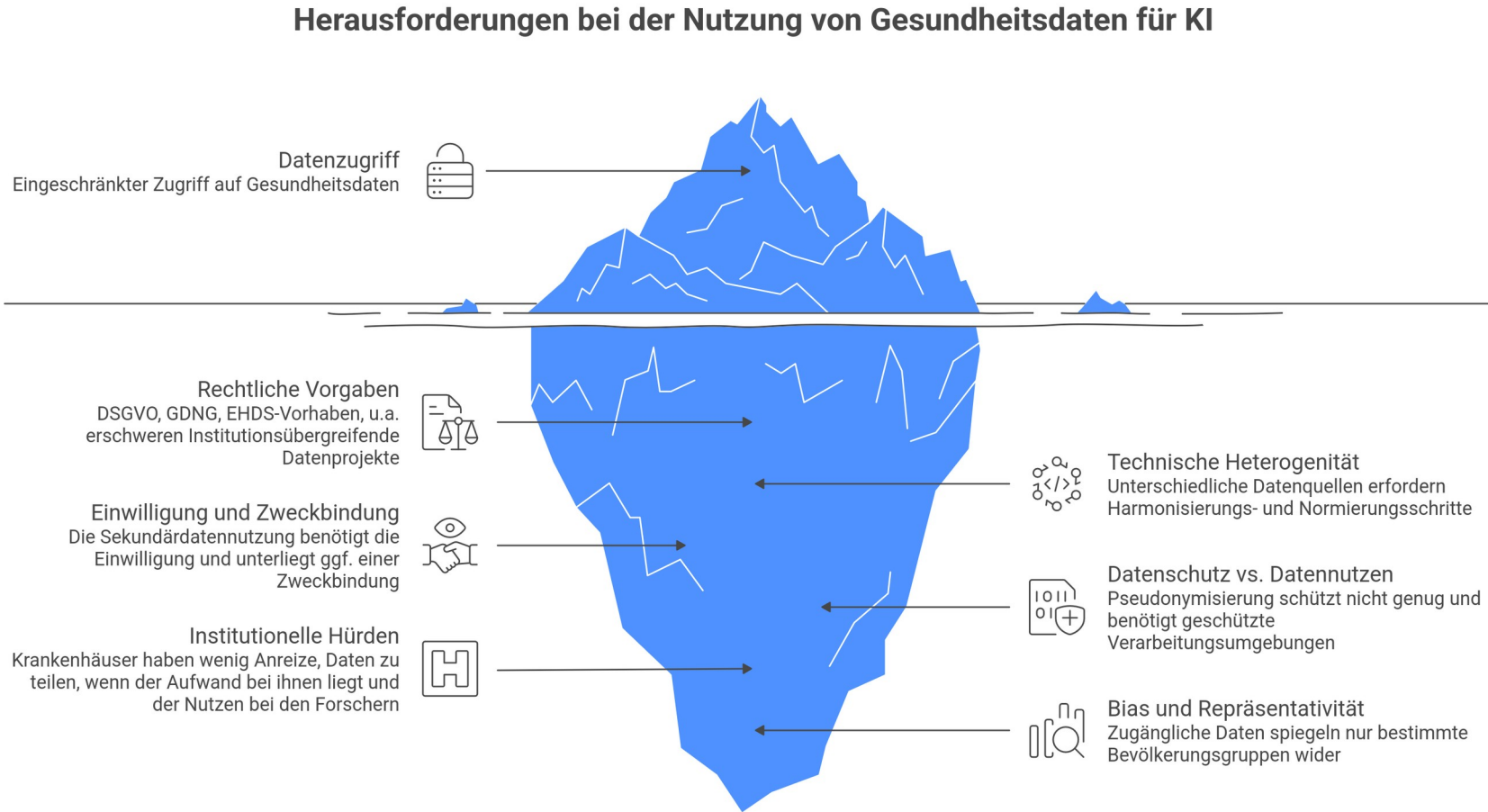
KI sichtbar machen.
Vernetze Dich mit
KI Med
Netzwerk für Künstliche Intelligenz in der Medizin

3. KIMed Netzwerk Partner Forum
KI trifft Medizin – sichere Lösungen für die Versorgung von morgen
Save the Date
27.10.2026
10:30 – 15:30 Uhr
Paulinum,
Neues Augusteum,
Leipzig

KI Med
Netzwerk für Künstliche Intelligenz in der Medizin

UNIVERSITÄT LEIPZIG
HOCHSCHULE MITTWEIDA
Technische Universität Dresden
Kofinanziert von der Europäischen Union
Land Sachsen
Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushalts.

AI developments require medical data



Made with Napkin

Secure Workflow @ ZIH Dresden

1 Encrypt Locally

2 Upload Securely



```
commands.sh x
1 #!/bin/bash
2
3 export VAULT_TOKEN=hvs.CAESIC7itM.....
4
5 getkey "input_data" <input data token>
6 getkey "output_data" <output data token>
7 getkey "image" <image token>
8
9 mount_data "input_data"
10 mount_data "output_data"
11
12 singularity exec --pem-path=/keys/image.key \
13   --bind /mnt/input_data:/input:ro \
14   --bind /mnt/output_data:/output \
15   ~/image.sif \
16   python /data/app/main.py --input_folder=/input --output_folder=/output
17
18 # Beispielaufwurf aus Check
19 # sudo singularity exec --pem-path=image.key image.sif python /data/app/main.py
```

```
run.sh x
1 #!/bin/bash
2 #SBATCH --exclusive
3 /opt/secureWorkflow/decrypt_and_execute <<EOF
4 -----BEGIN PGP MESSAGE-----
5
6 hQGMA+twzSrru/0TAQv=I8a7
7 ...|
8 -----END PGP MESSAGE-----
9 EOF
```

4 Execute on Secure Node

Details image classification model

- 25 Epochen
- Xception (Extreme Inception) is a deep convolutional neural network architecture

```
class CatDogClassifier: 1 usage  👤 Sobe, Kathrin +1
    def train_model(self): 1 usage  👤 Sobe, Kathrin +1
        ]
        model.compile(
            optimizer=keras.optimizers.Adam(3e-4),
            loss=keras.losses.BinaryCrossentropy(from_logits=True),
            metrics=[keras.metrics.BinaryAccuracy(name="acc")],
        )
        logging.info("done with model compile")

        model.fit(
            train_ds,
            epochs=self.num_epochs,
            callbacks=callbacks,
            validation_data=val_ds,
        )
        logging.info("done with model fit, callbacks are saving the model after epochs...")
```