



AI-supported, healthcare-related use of cancer registry data – AI-CARE



Alexander Katalinic
www.ai-care-cancer.de

Project Partners



Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH



UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR SOZIALMEDIZIN
UND EPIDEMIOLOGIE



Krebsregister
Saarland ●●●●



ZENTRUM FÜR
KREBSREGISTERDATEN





Main questions

1. Can the complex cancer registry data be processed, improved and merged using artificial intelligence (AI) methods in such a way that they are more accessible than before for oncological quality assurance and research?
2. How can analyses of cancer registry data using AI methods usefully supplement the classic analysis spectrum of oncological health services research?



8 Major Work Packages of AI-CARE

1. Central data centre (application platform) for Cancer Registry Data
2. CR Data preparation for AI – Usage
3. Data Reports
4. Federated Learning
5. Data optimization
6. AI-based explorative analysis
7. AI-based survival analysis
8. Open-source repository

Please, consider this is work in progress !!!

Work Package 1

Central data application platform
for Cancer Registry Data



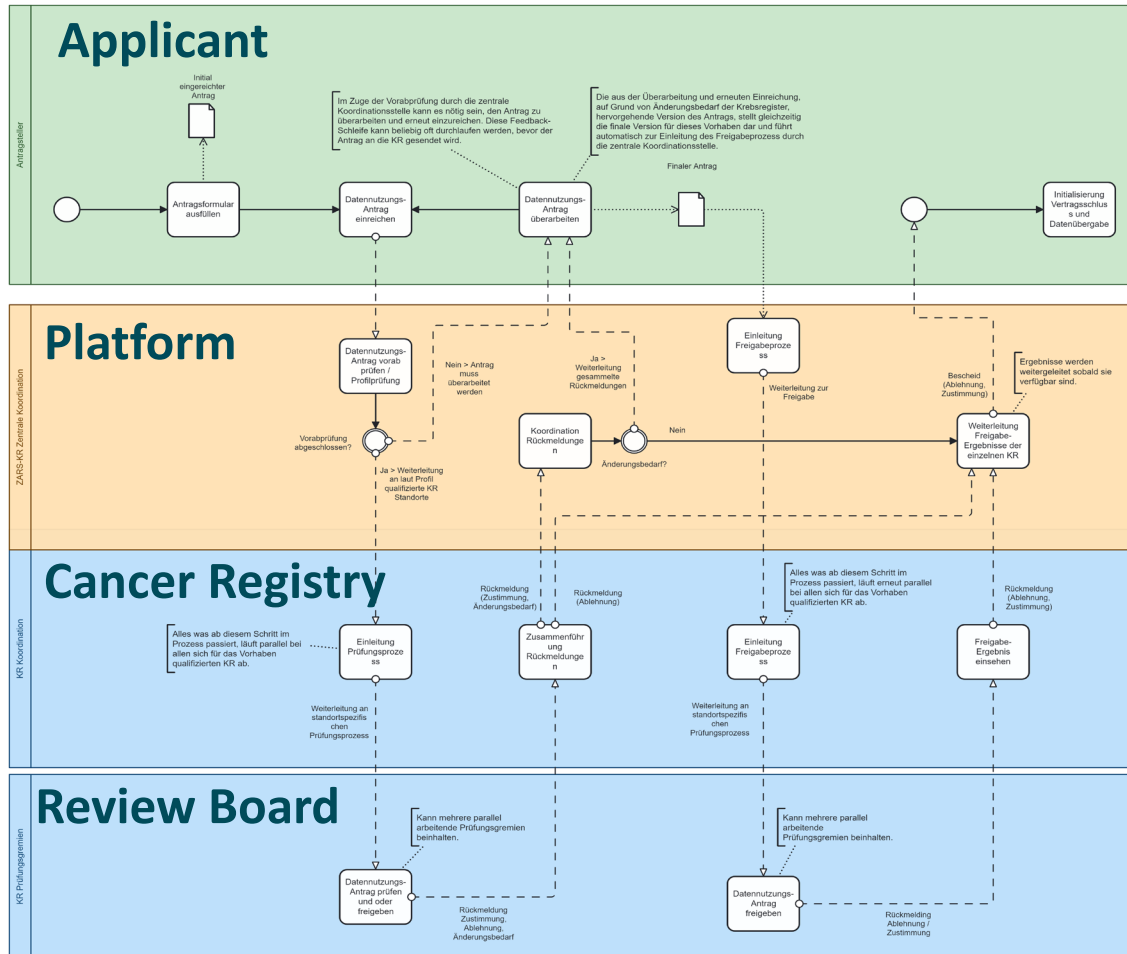
WP 1 – Establishment of a central data application platform

- Goal:
 - A single point for data usage application process for all 15 cancer registries at the same time
 - A single point for terms of usage
 - Consideration of local conditions
 - A single point of contact
- Current Situation:
 - 15 local cancer registries with 15 different legal frameworks and process to apply for data
 - 15 different data usage applications
 - 15 different contracts and terms of usage

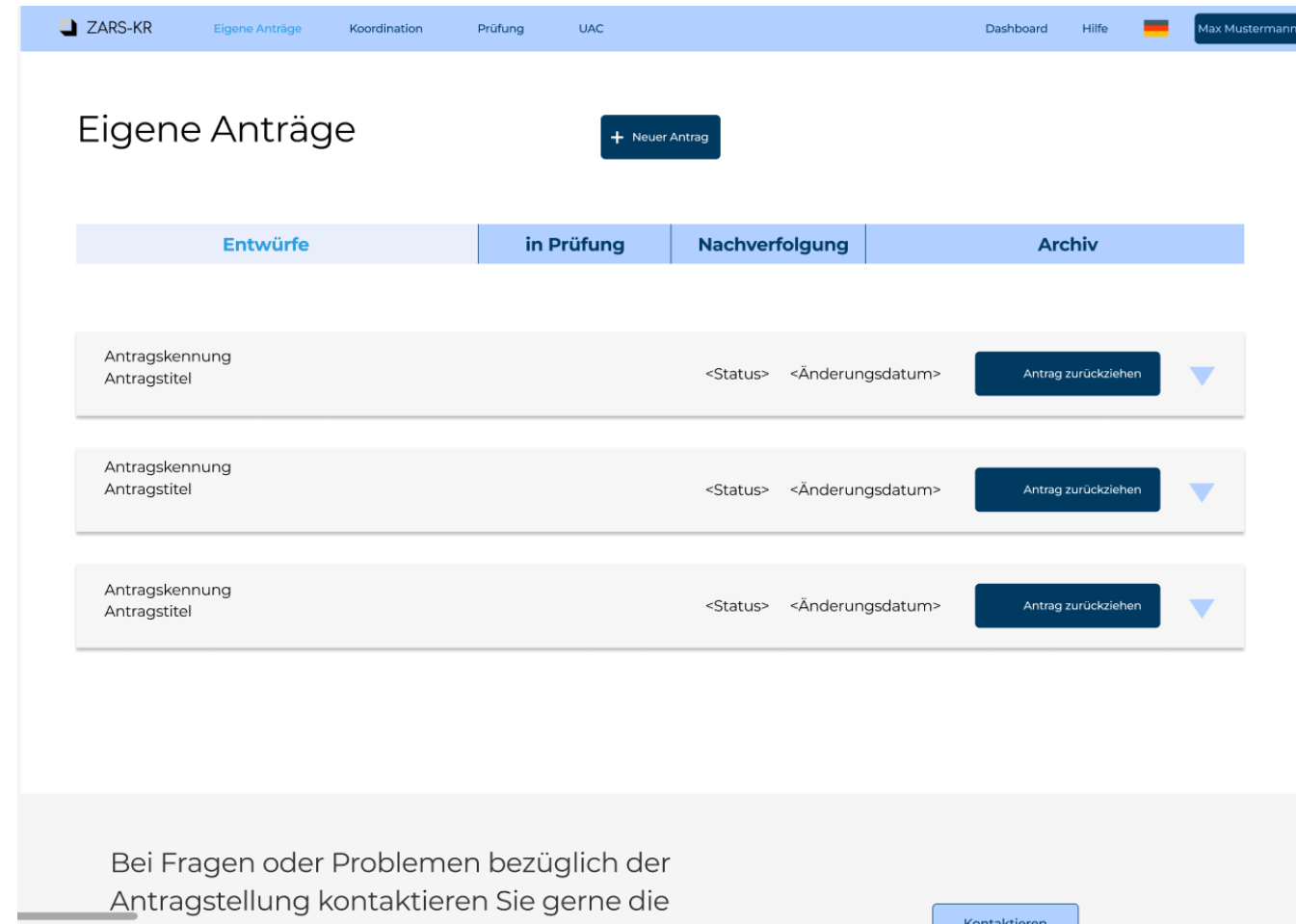
WP 1 – Establishment of an application platform

- A survey among all German cancer registries has determined the current legal framework and the data usage application process
- Development of a cross site data usage application process
- Software “ProSkive” will be used for processing data usage applications
 - is already in use for the German medical informatics initiative (MII)
 - will be further developed to meet the demands that the survey has revealed

WP 1 – Establishment of an application platform



generic process model for all locations



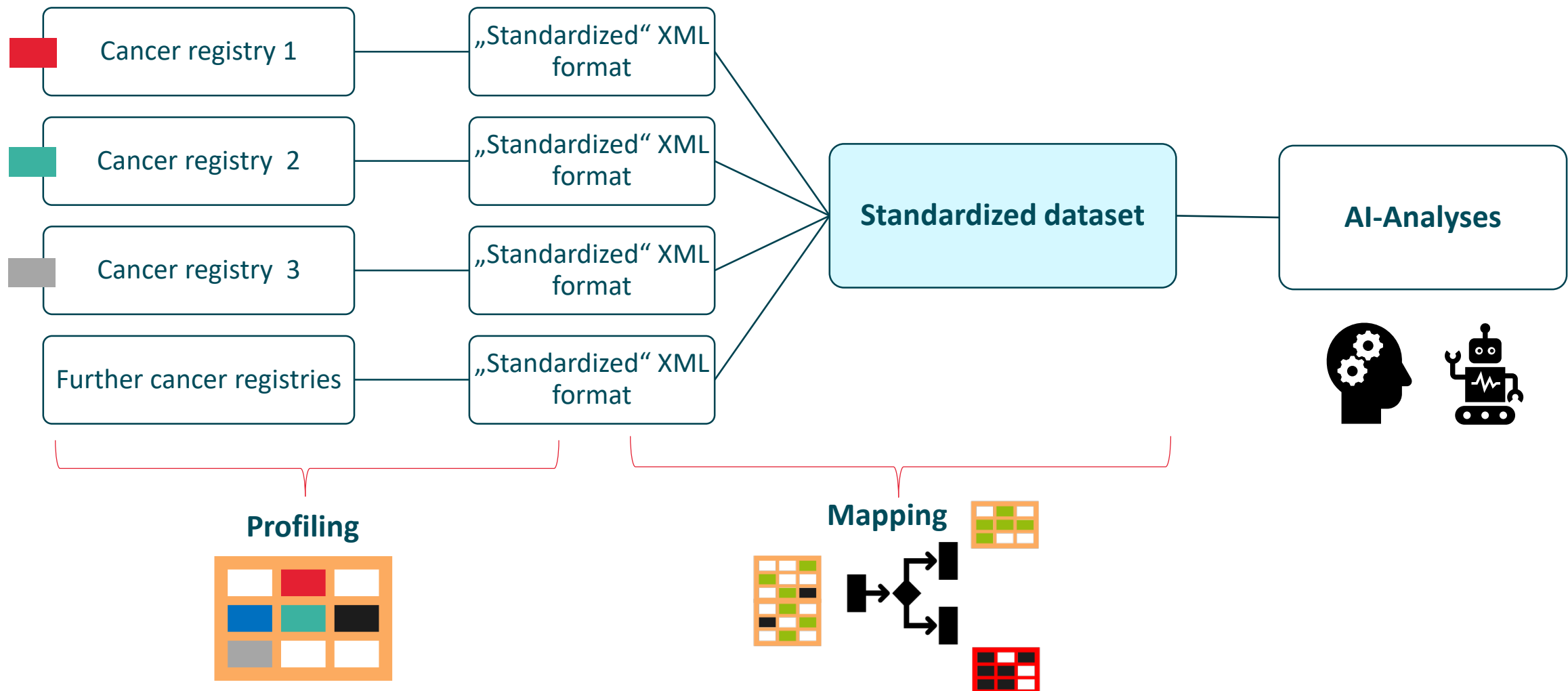
screenshot of the planned application platform

Work Package 2

Data optimization



Standardization of heterogeneous cancer registry data for AI-applications

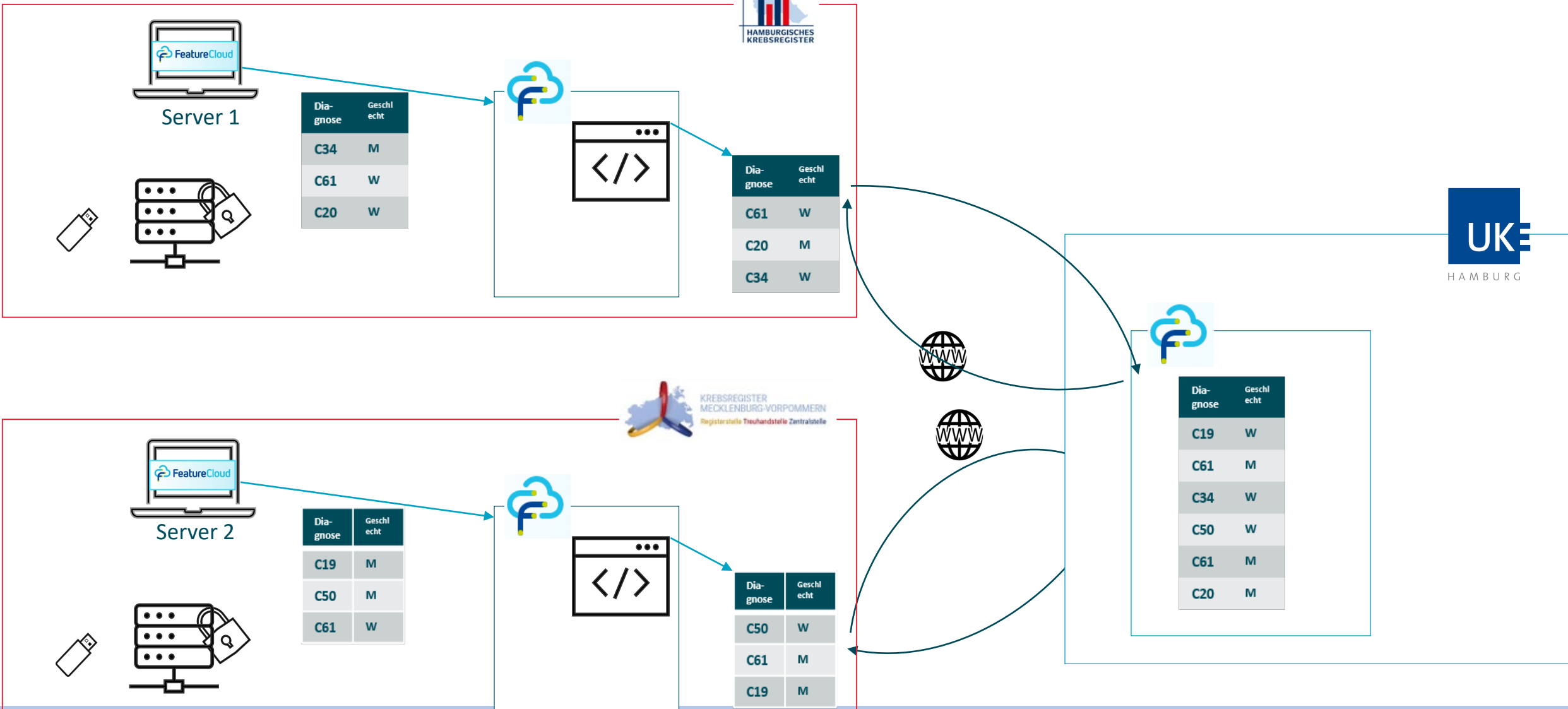


Work Package 4

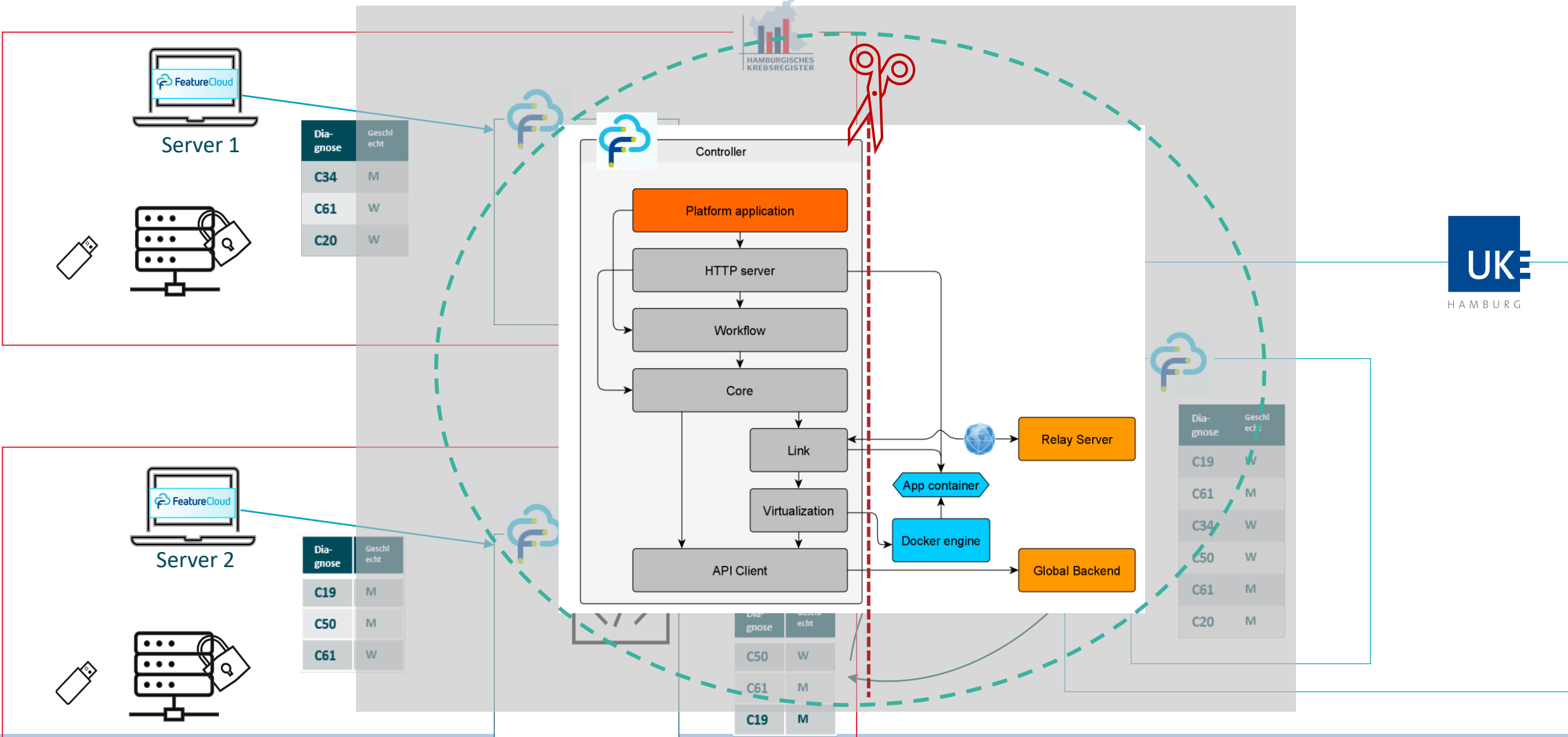
Federated Learning



WP2: Enabling federated learning in cancer registries



WP2: Enabling federated learning in cancer registries



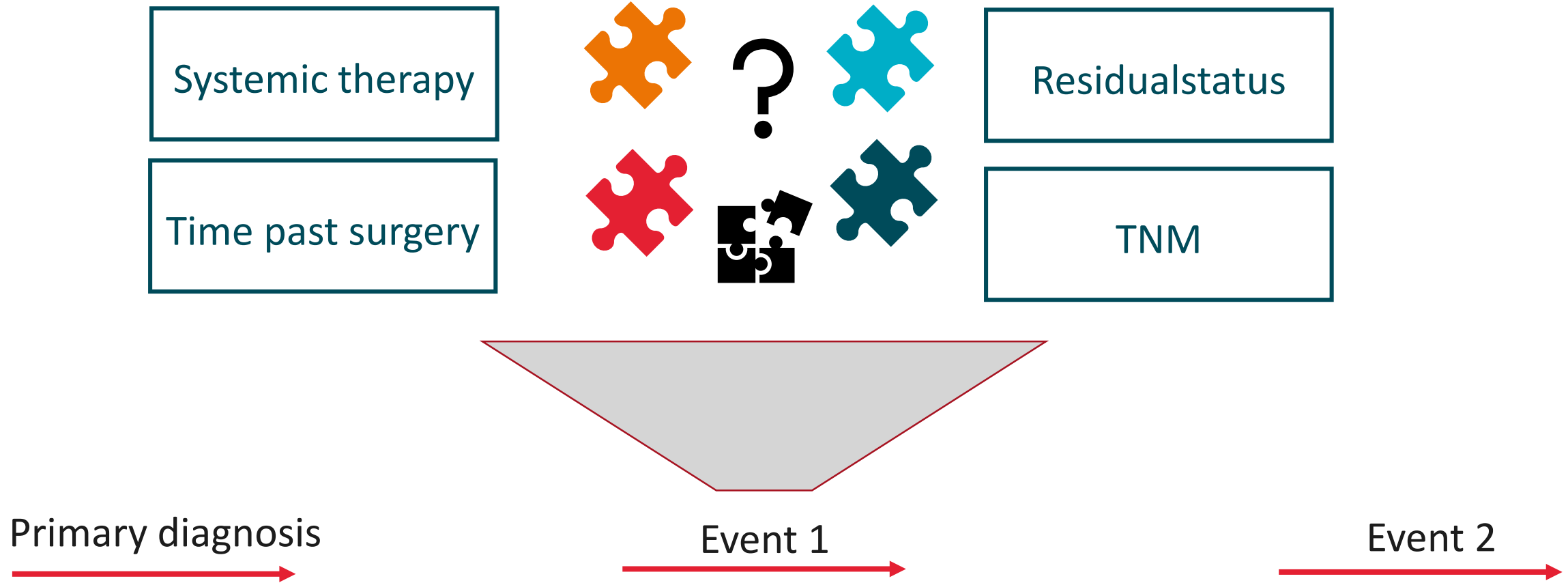
Work Package 5

Data optimization

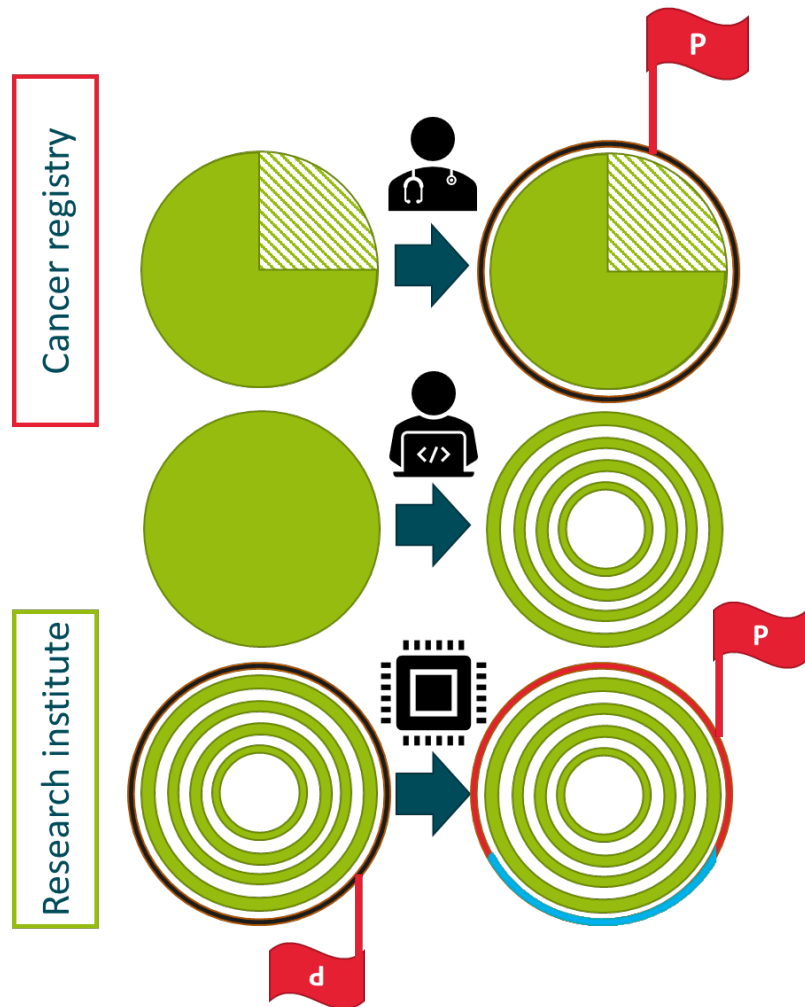


AI-supported classification of events into recurrence and progress

Recurrence or progression?



AI-supported classification of events into recurrence and progress



Manual labelling of events by domain experts 

Information reduction of cases (privacy protection)

Training of ML-Algorithms to classify events into recurrence and progress

```
HistGradientBoosting Results, 5-fold-crossvalidation
fit_time : (0.04, 0, 2)
score_time : (0.01, 0, 2)
test_accuracy : (0.92, 0, 2)
test_precision_1 : (0.87, 0, 2)
test_recall_1 : (0.81, 0, 2)
test_f1_score_1 : (0.83, 0, 2)
test_precision_0 : (0.93, 0, 2)
test_recall_0 : (0.96, 0, 2)
test_f1_score_0 : (0.94, 0, 2)
```

280s

Work Package 6

Exploring approaches for automatic stratification of patients



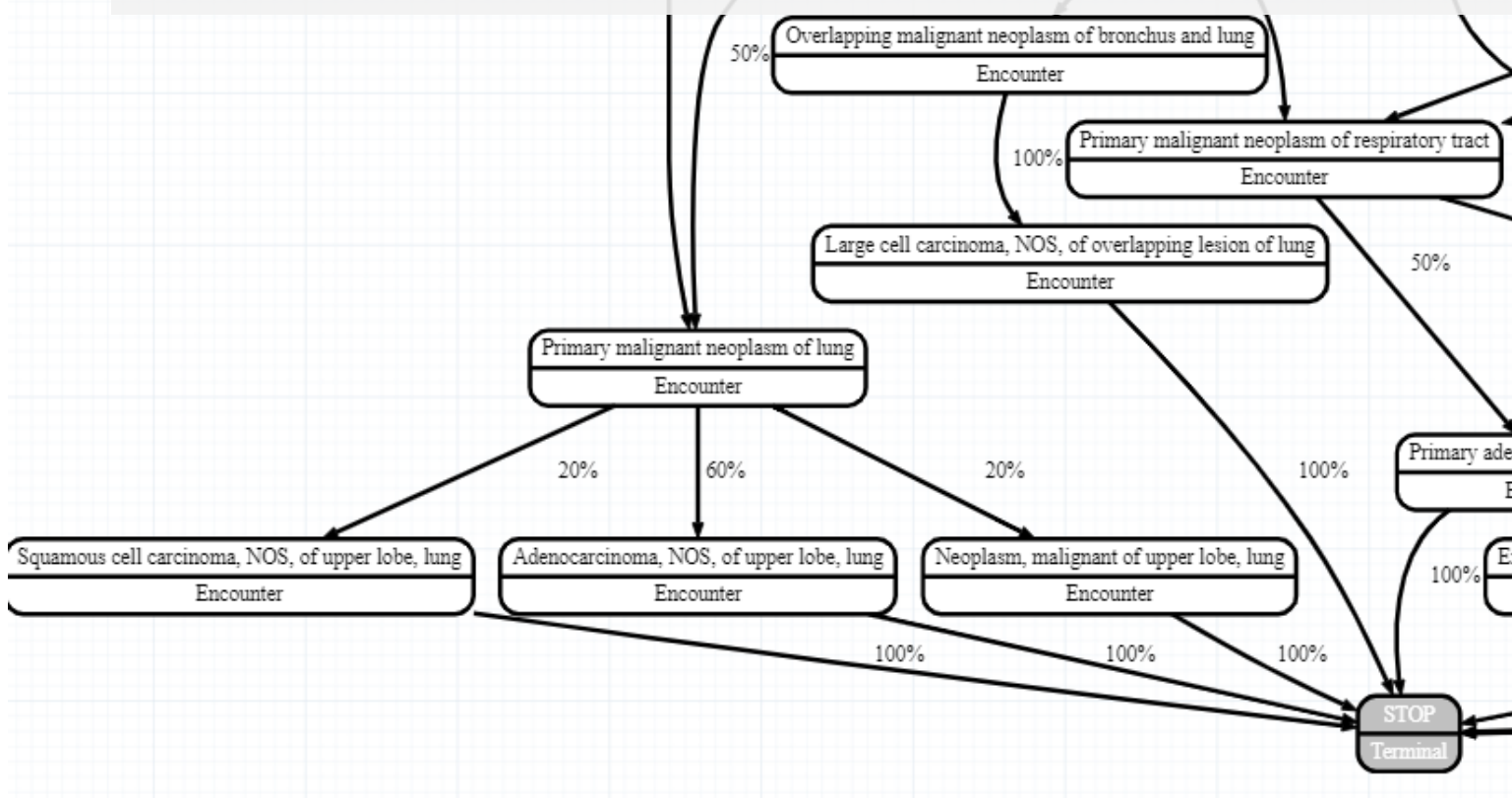
WP6: Exploring approaches for automatic stratification of patients AI-CARE

Option 1: Generative models with explicit (symbolic) knowledge representation

Representation of therapies in a probabilistic sense

Complexity limited, but fully interpretable and verifiable

Area of application: Synthetic data that demonstrably contains no identifiable data



$$p(y|x)$$

$$p(x)$$

Using generative models with explicit knowledge
 → Symbolic representation

WP6: Exploring approaches for automatic stratification of patients

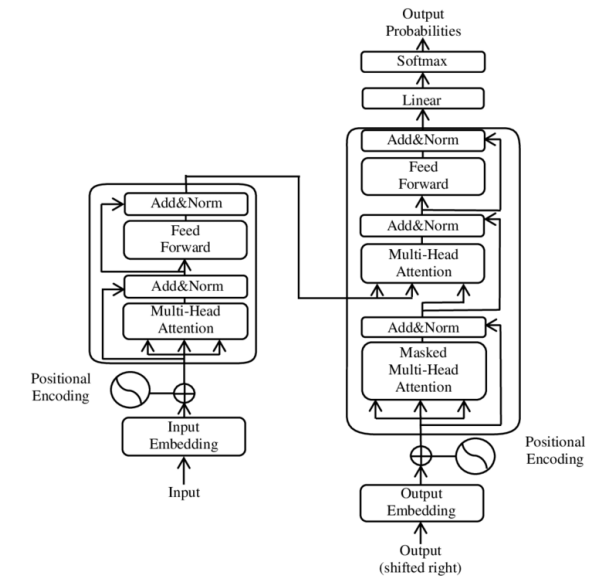
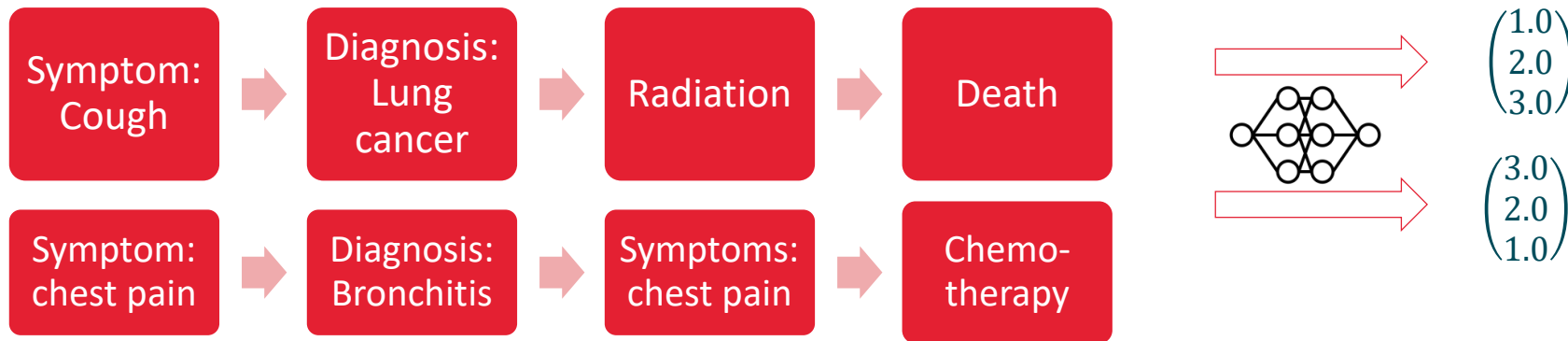
Option 2: "Deep learning" (not interpretable or verifiable)

Reduction of different treatment trajectories to a low-dimensional vector

Possibility to calculate similarities between patients

Use of so-called "transformer architectures" (Technology behind ChatGPT)

Area of application: Pre-processing before classification



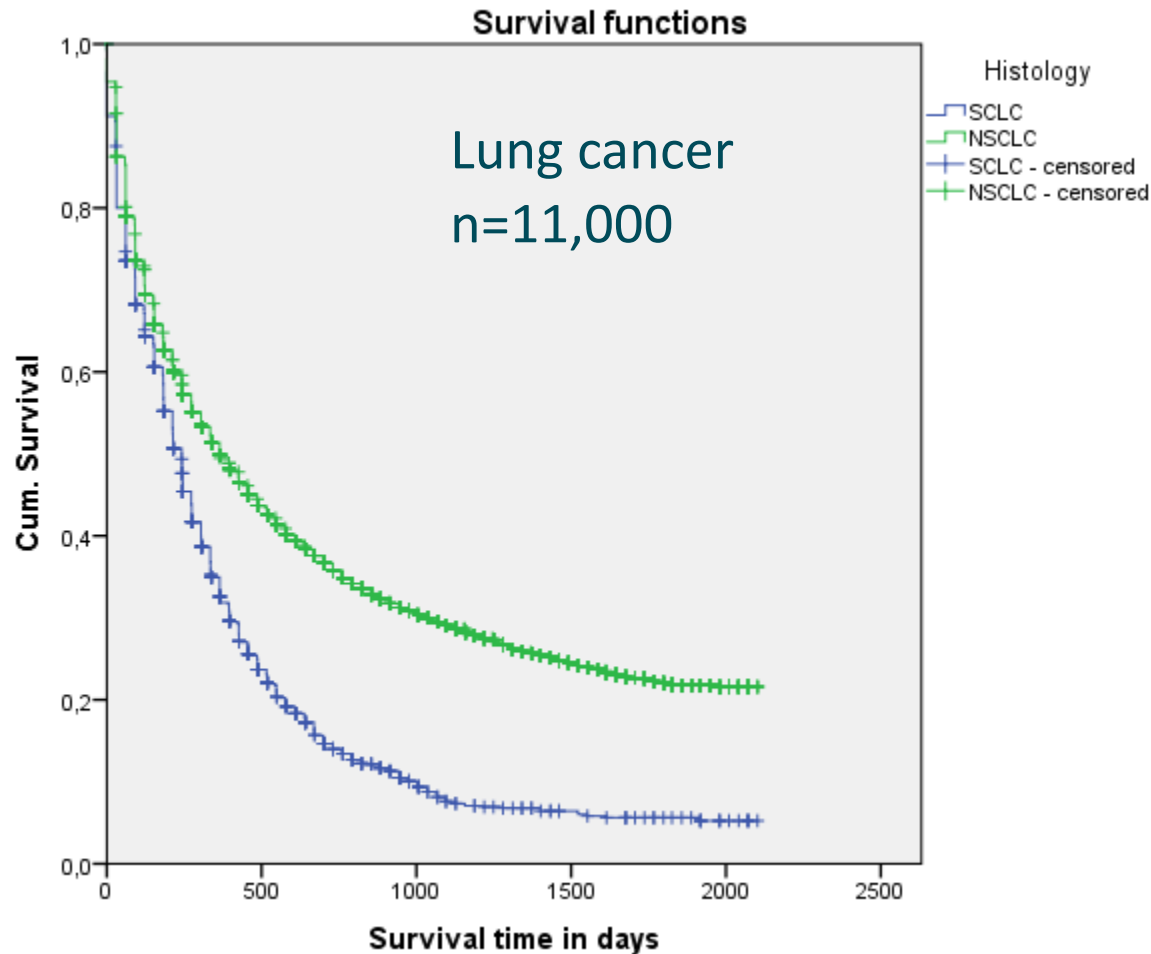
Using deep learning with implicit knowledge
 → Sub-symbolic representation

Work Package 7

Survival analyses – Comparing conventional epidemiological methods to AI-based approaches



Kaplan-Meier Method and Cox Regression



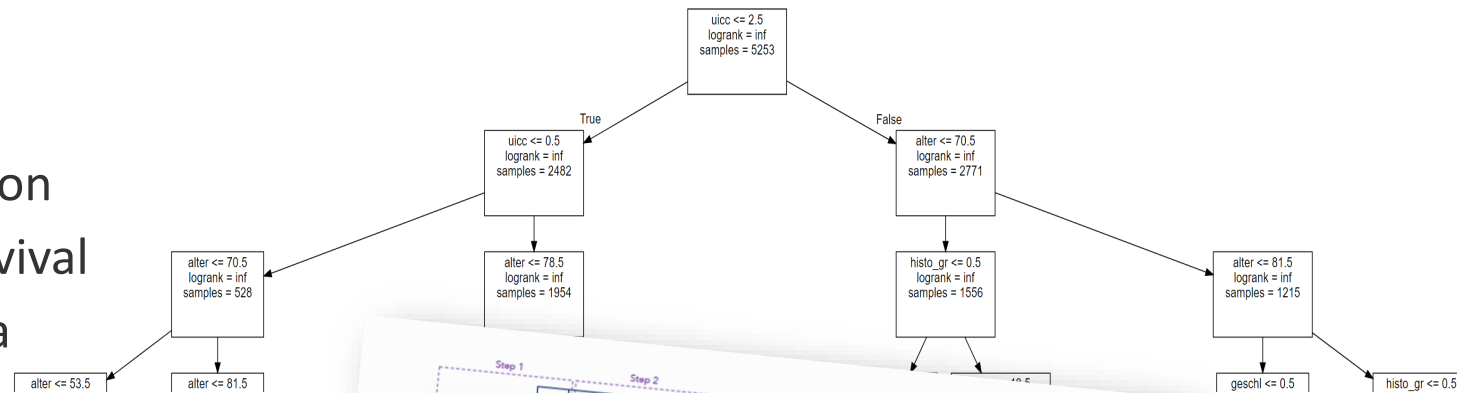
- Unique feature of epidemiological survival analysis are censored data
- Censored patients contribute survival time to the analyses, but their outcome is unknown
- Challenge for AI-approaches, because binary outcomes are preferred

Random Decision Forest and Deep Neural Networks

- There are machine learning techniques to address the problem of censored data in survival analysis:

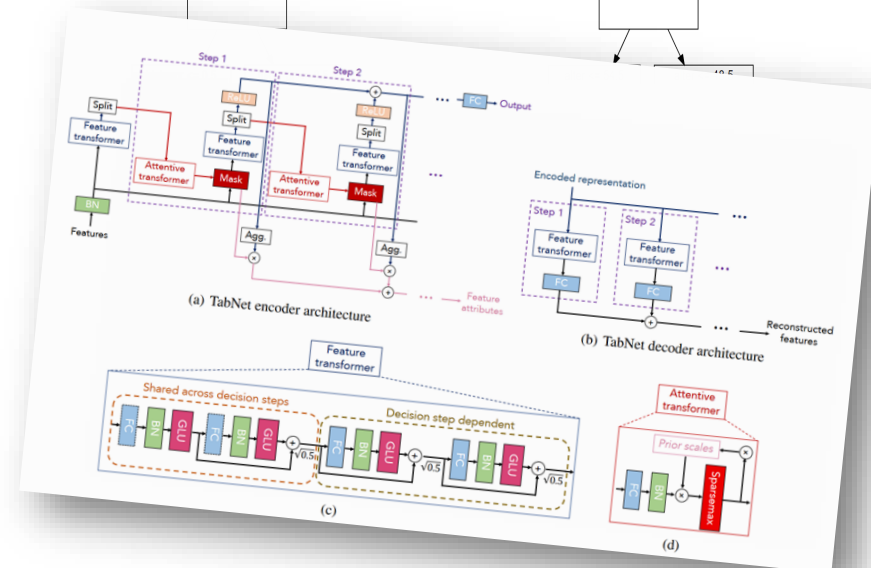
- Random Survival Forest (RSF):

RSFs are adapted Random Decision Forests for the application in survival data set with right-censored data



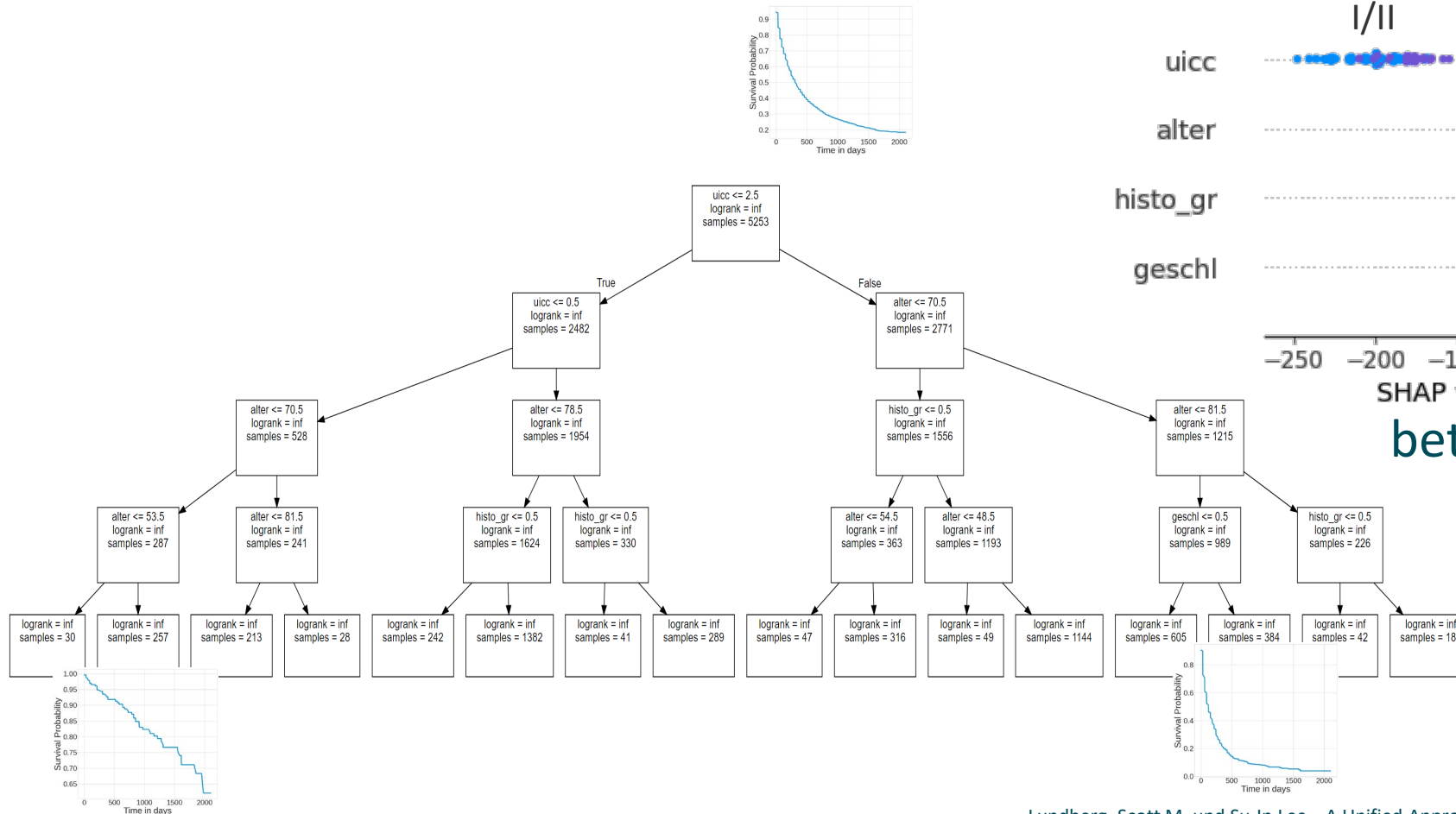
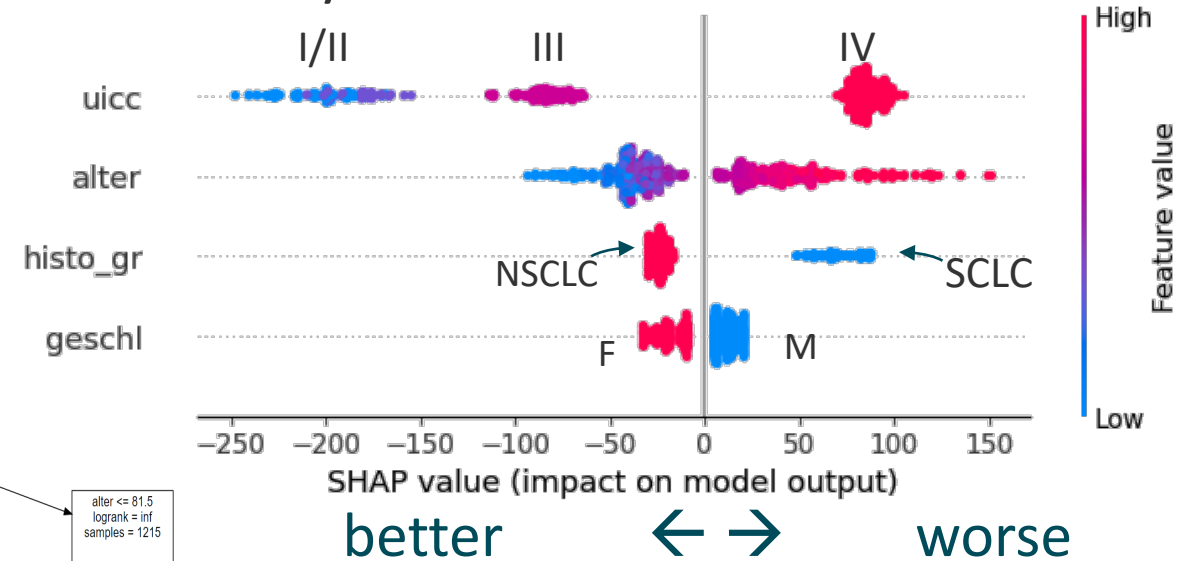
- Deep Neural Networks (DNN):

TabNet is a neural network for tabular data which can be trained with a censoring-allowing loss function



Random Survival Forest – First results

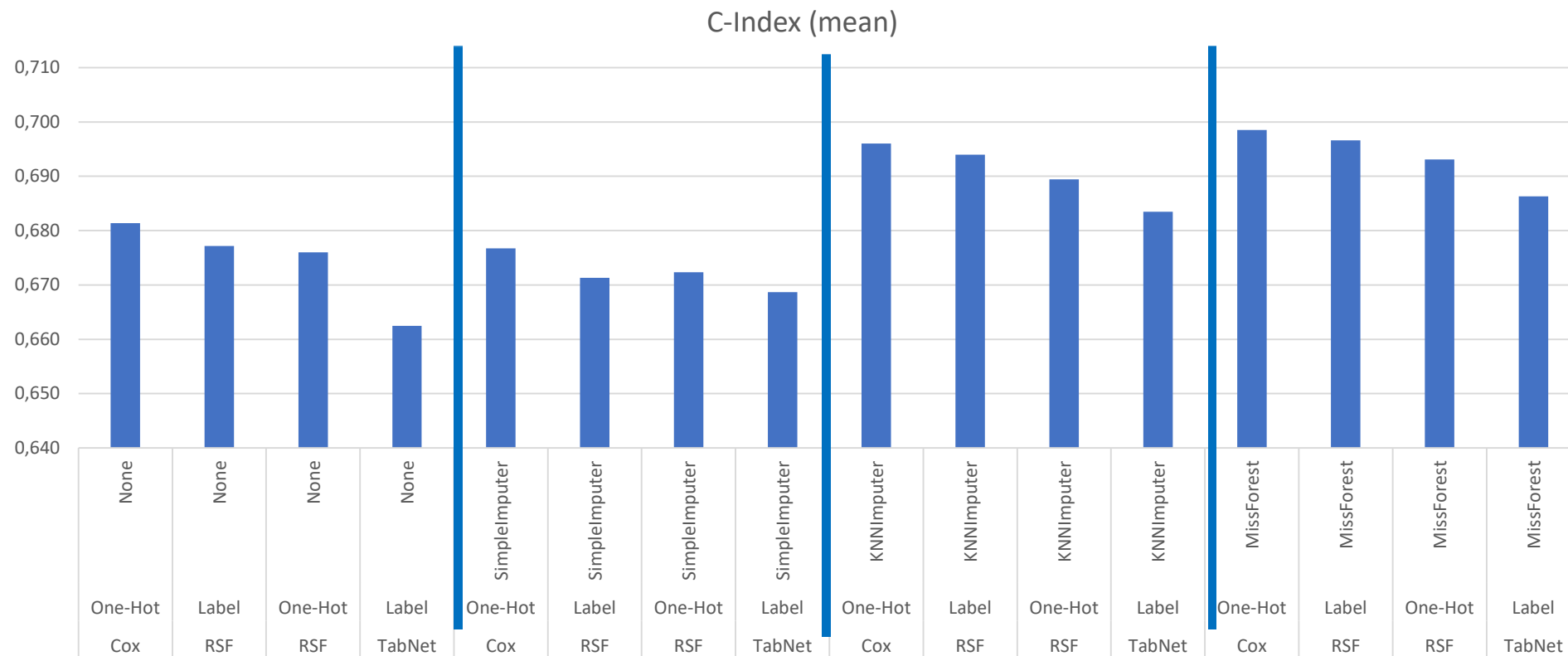
SHAP-Analysis:



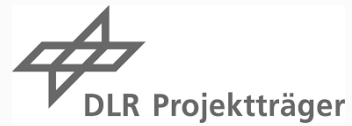
Lundberg, Scott M, und Su-In Lee. „A Unified Approach to Interpreting Model Predictions“. In *Advances in Neural Information Processing Systems*, Bd. 30. Curran Associates, Inc., 2017.
<https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>.

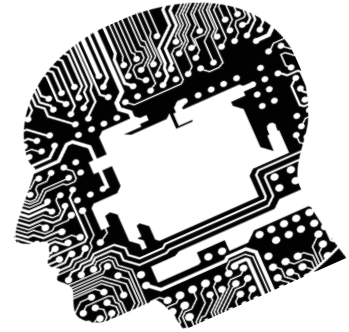
Benchmarking Cox Regression with AI-based approaches

- Preliminary results based on lung cancer data from the Schleswig-Holstein Cancer Registry (2016 – 2021; follow-up 03/2022; 10 364 cases; 14% missing in UICC)



SUMMARY





Summary

- Project is well on schedule!
- Central application platform will significantly simplify application processes for researchers
 - Integration into upcoming German Health Data Law possible
- The use of artificial intelligence in the analysis of cancer registry data is possible in principle. Whether and where AI has a relevant advantage over conventional methods will be one of the key project results!