



# Relevance of OpenML in CoE RAISE

Prof. Dr. – Ing. Morris Riedel

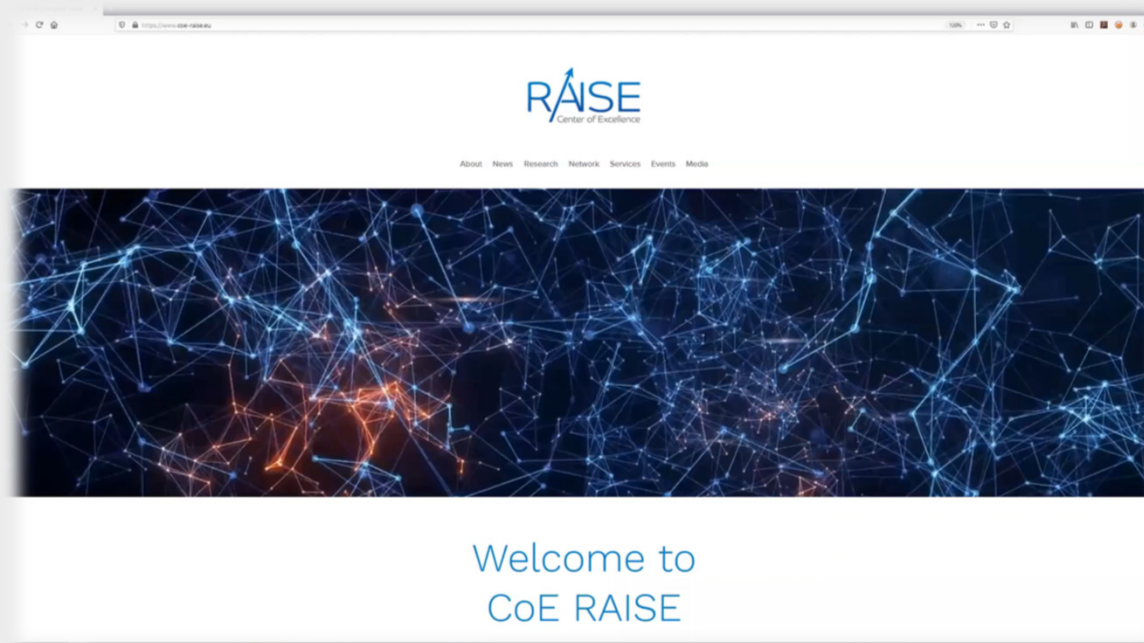
School of Engineering & Natural Sciences, University of Iceland, Iceland  
Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

*2022-05-31, RAISE CoE Training Using OpenML for sharing datasets, algorithms, and experiments, Online*

 @ProfDrMorrisRiedel  @Morris Riedel  @MorrisRiedel  @MorrisRiedel  
 <https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg>  morris@hi.is



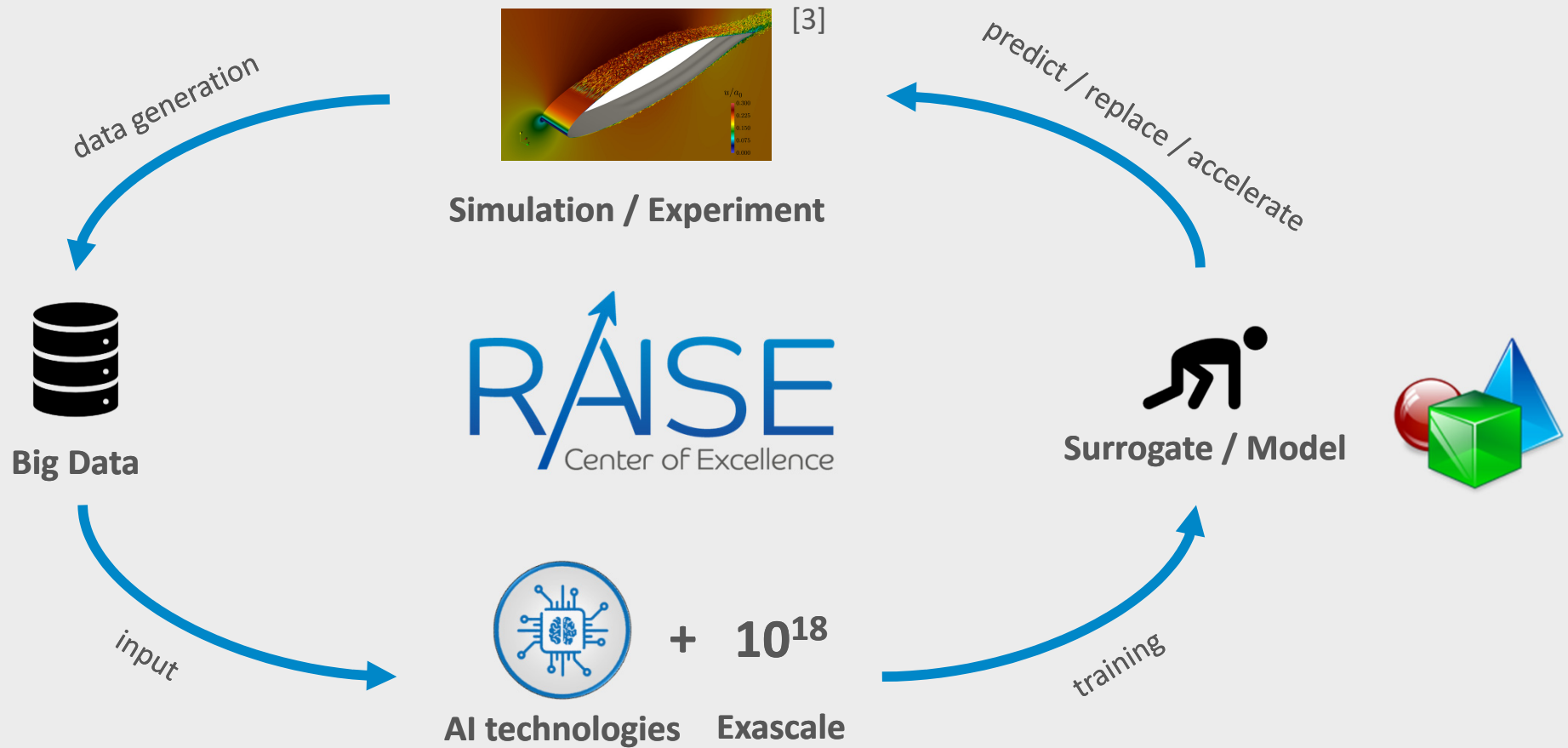
# CoE RAISE Web Page & More Information



<https://www.coe-raise.eu>

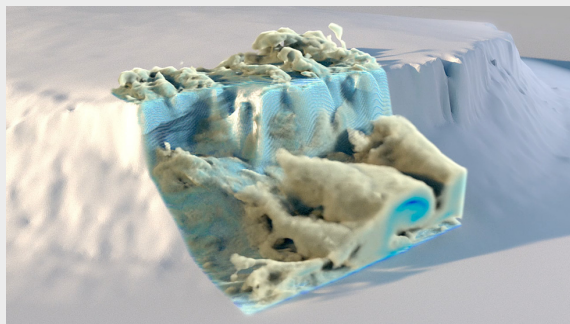
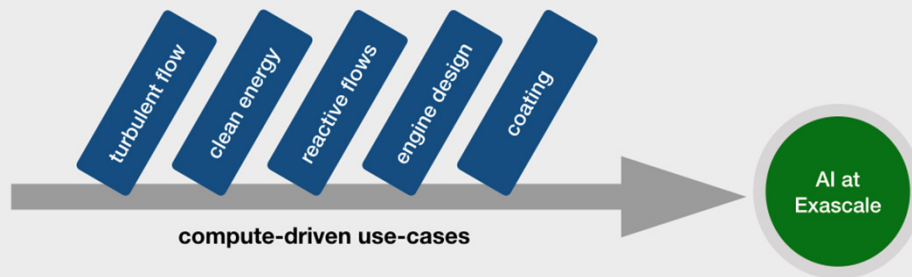


# CoE RAISE – Motivation & Approach

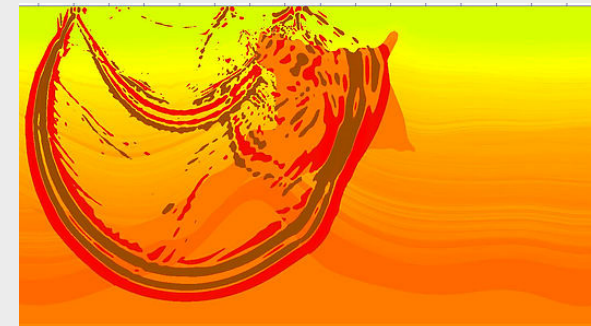
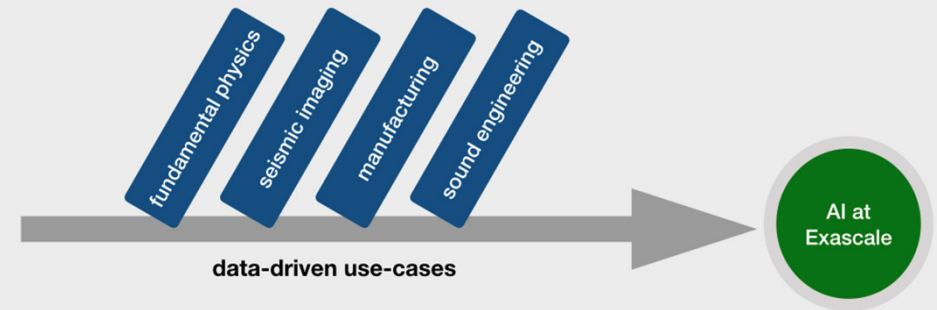


# Use Cases in CoE RAISE

➤ Two kinds of use cases:



Example from use case "AI for wind farm layout": Turbulence generated by a cliff on Bolund Island, Denmark.



Example from use case "Seismic imaging with remote sensing - oil and gas exploration and well maintenance": Snapshot from a wavefield.

# Compute- and Data-driven Use Cases – Data & Modeling

AI for turbulent boundary layers

AI for wind farm layout optimization

AI for data-driven models in reacting flows

Smart models for next-generation aircraft engine design

AI for wetting hydrodynamics

Event reconstruction and classification at the CERN HL-LHC

Seismic imaging with remote sensing - oil and gas exploration and maintenance

Defect-free metal additive manufacturing

Sound Engineering



# Example: Impact of Quantum Computing in Optimization

AI for turbulent boundary layers

AI for wind farm layout optimization

AI for data-driven models in reacting flows

QU

BIT

Macro scale ~ 1mm

Micro scale ~ 1µm

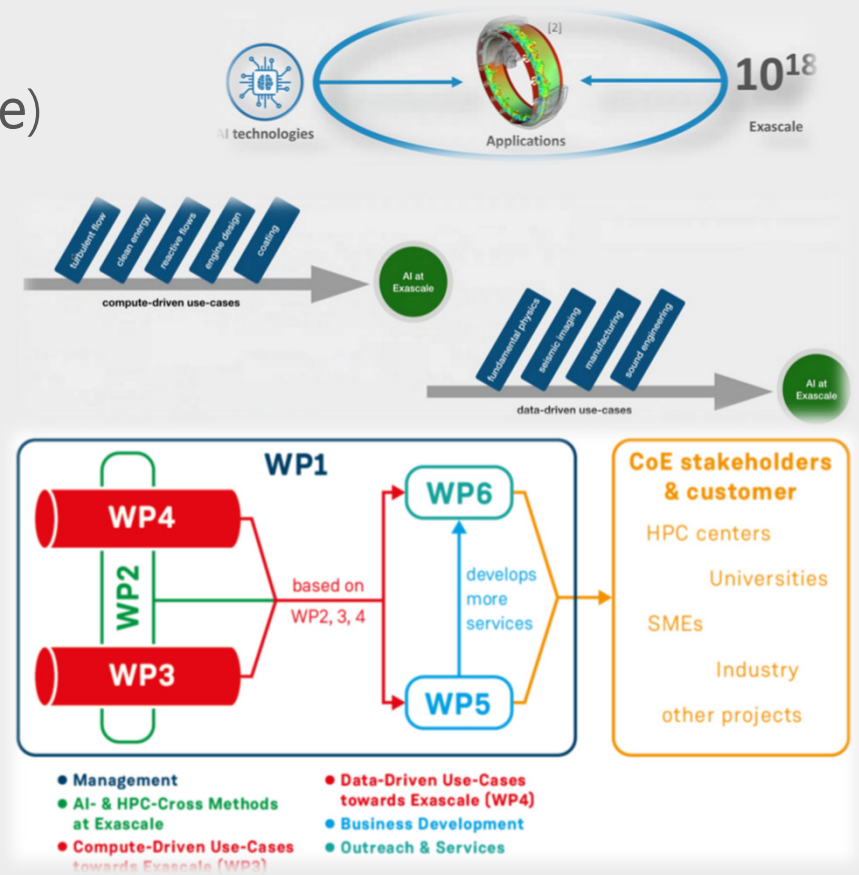
Seismic imaging with remote sensing - oil and gas exploration maintenance

Defect-free metal additive manufacturing

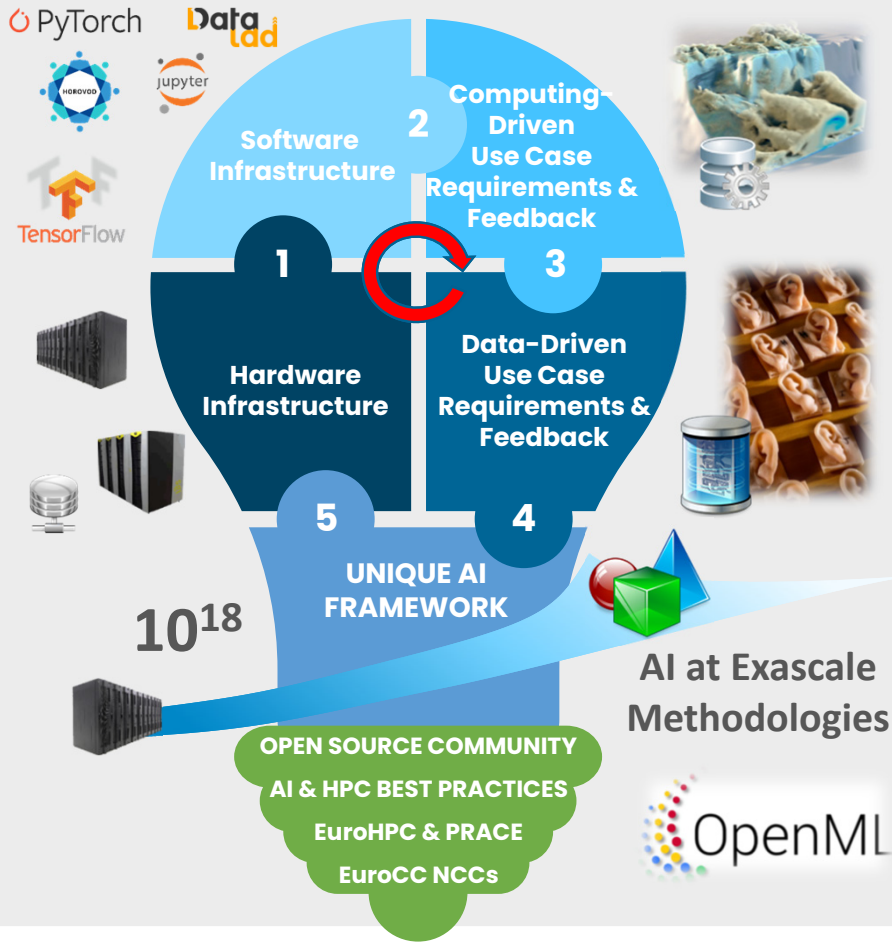
Sound Engineering

# WP2 – AI- & HPC-Cross Methods at Exascale in a nutshell

- WP3 (Compute-Driven Use-Cases towards Exascale)
- WP4 (Data-Driven Use-Cases towards Exascale)
- Developments in these WPs will be supported by the cross-linking activities of WP2
  - E.g. scaling machine & deep learning codes with frameworks like Horovod/Deepspeed
  - E.g. introduction to new AI methods such as Long-Short Term Memory (Time series)
  - E.g. data augmentation approaches
  - E.g. benchmarking HPC machines and offer also pre-trained AI algorithms (i.e., transfer learning)
  - E.g. offer neural architecture search methods for hyperparameter – tuning in semi-automatic way



# Towards AI & HPC at Exascale with CoE RAISE Results



## Hardware Infrastructure

Prepare & Document available production systems at partners' HPC centers  
 Examples: JUWELS (JUELICH), LUMI (UoICELAND), DEEP Modular Prototypes, JUNIQ (JUELICH), etc.



## Software Infrastructure

Prepare & Document available open source tools & libraries for HPC & AI useful for implementing use cases  
 Examples: DeepSpeed and/or Horovod for interconnecting N GPUs for a scalable deep learning jobs

## Computing-driven Use Cases Requirements & Feedback

Use cases with emphasize on computing bring in co-design information about AI framework & hardware

Examples: Use feedback that TensorFlow does not work nicely, so WP2 works with use cases on

## Data-driven Use Cases Requirements & Feedback

Use cases with emphasize on data bring in co-design information about AI framework & hardware  
 Examples: Deployment blueprint by using AI training on cluster module & inference/testing on booster

## → UNIQUE AI FRAMEWORK

Living design document & software framework blueprint for using HPC & AI offering also pretrained AI models



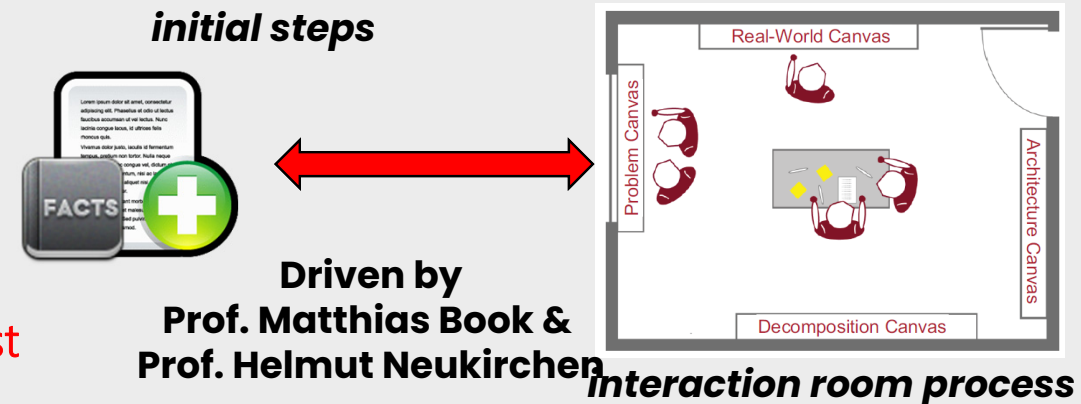
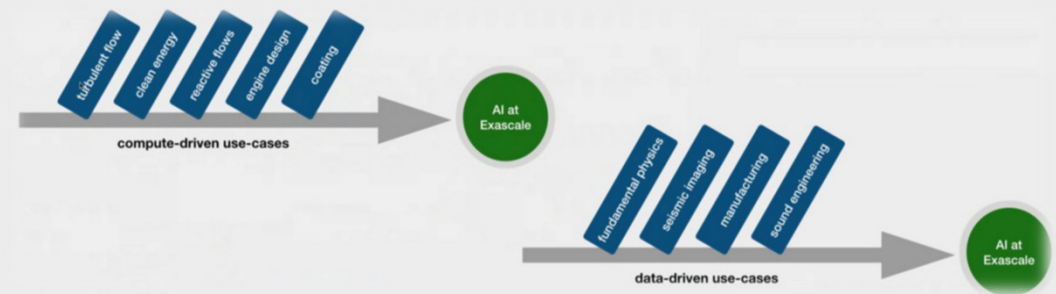
# Selected Techniques to Identify Cross-Methods for HPC & AI

## ➤ Fact Sheets

- Foster initial understanding
- Living document & each Fact Sheet per WP3/WP4 Use Case
- *(Experience from many other EU projects)*

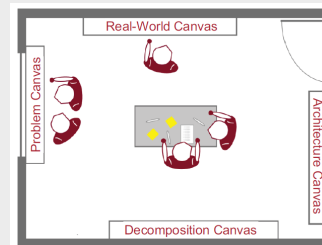
## ➤ Selected Contents

- Short Application Introduction
- Clarify Primary Contacts
- Codes/Libraries/Executables
- HPC System Usage Details
- Specific Platforms & 'where is what data'?
- **Machine/Deep Learning Approaches of Interest**



## ➤ CoR RAISE Interaction Room Process as Next Step

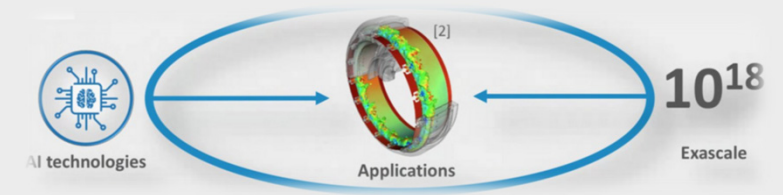
- Supports the proper software engineering design of the unique AI framework blueprint
- Expecting to work with WP3 & WP4 experts in an open minded way
- Process will be guided by Prof. Dr. Matthias Book (University of Iceland)
- Supported by Software Engineering & testing expert Prof. Dr. Helmut Neukirchen (University of Iceland)



### HPC Systems Engineering in the Interaction Room

Matthias Book

with Morris Riedel, Jülich Supercomputing Centre / UoI and Helmut Neukirchen, University of Iceland



➤ CoE RAISE @ YouTube: <https://www.youtube.com/channel/UCAdIZ-v6cWwGdapwYxdN7dg>

➤ **Methology as one CoE RAISE outcome**

Book, M., Riedel, M., Neukirchen, H., Goetz, M.: **Facilitating Collaboration in High-Performance Computing Projects with an Interaction Room**, in conference proceedings of the 4th ACM SIGPLAN International Workshop on Software Engineering for Parallel Systems (SEPS 2017), October 22-27, 2017, Vancouver, Canada

# RAISE Initial Identified AI/HPC Methods: Graph Example



BEFORE

Table 6: Use-case vs. AI-methods matrix.

Use-Case vs. AI-Methods	DA	NAS	AE	TL	PF	PIDL	LSTM
Turbulent boundary layers	X	X	X	X	X	X	
Wind farm layout optimization	X			X	X	X	
AI for data-driven models in reacting flows				X	X	X	
Smart models for next-generation aircraft engine design	X	X		X	X	X	
Wetting hydrodynamics		X	X			X	X
Event reconstruction and classification at the CERN HL-LHC		X		X			X
Seismic imaging with remote sensing - oil and gas exploration and well maintenance	X	X		X			
Defect-free metal additive manufacturing		X				X	X
Sound engineering	X	X		X			

H2020-INFRAEDI-2018-2020



CoE RAISE

Center of Excellence "Research on AI- and Simulation-Based Engineering at Exascale"

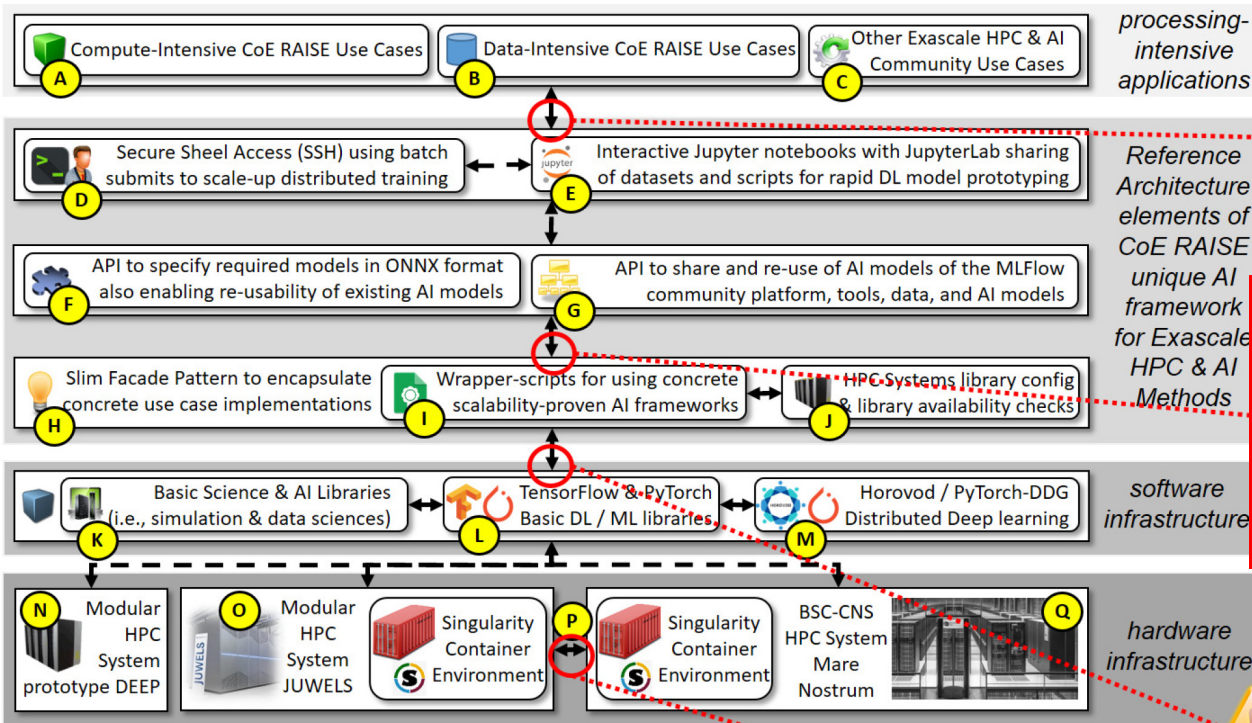
Grant Agreement Number: 951733

MS2  
AI/HPC Methods  
Draft



Use Case	AE	PIML	ANNs	CNN	NO	SMs	GNN	IN	LSTM	GRU
<b>Details</b>	CAE		RBF-ANN	U-Net RESNET	FNO	AR ARMA ARIMA		JEDI-net		
AI for turbulent boundary layers	X	X								
AI for wind farm layout optimization			X			X X X				
AI for data-driven models in reacting flows				X			X			
Smart models for next generation aircraft engine design				X			X			
AI for wetting hydrodynamics					X					
Event reconstruction and classification at the CERN HL-LHC use case							X	X		
Seismic imaging with remote sensing for energy applications	X			X						
Detect-free metal additive manufacturing	X			X						
Sound Engineering									X	X

# Role of Hyperparameter Tuning in RAISE Unique AI Framework



✓ RQ1, RQ2, RQ4, RQ5

- ❖ Parts of the framework layout plan is to provide Kernels for Jupyter notebooks with correct version setups of modules for specific HPC Systems

✓ RQ3, RQ6

- ❖ Parts of the framework layout plan is to provide a lightweight and abstract Python API building on ONNX enabling also exchanges via MLFlow/ClearML

✓ RQ1, RQ2, RQ8, RQ9

- ❖ Parts of the framework layout plan is to provide a lightweight Python API that abstracts from low level versioning of AI packages (with proven scalability) and is harmonized with different available HPC system module versions

✓ RQ6, RQ7, RQ8, RQ9

- ❖ Part of the framework layout plan is to provide containers in Singularity with prepackaged datasets & software stacks needed for AI agnostic to hardware & good I/O performance

**Legend:**  
 Tangible outputs of RAISE WP2 as part of the unique AI framework layout



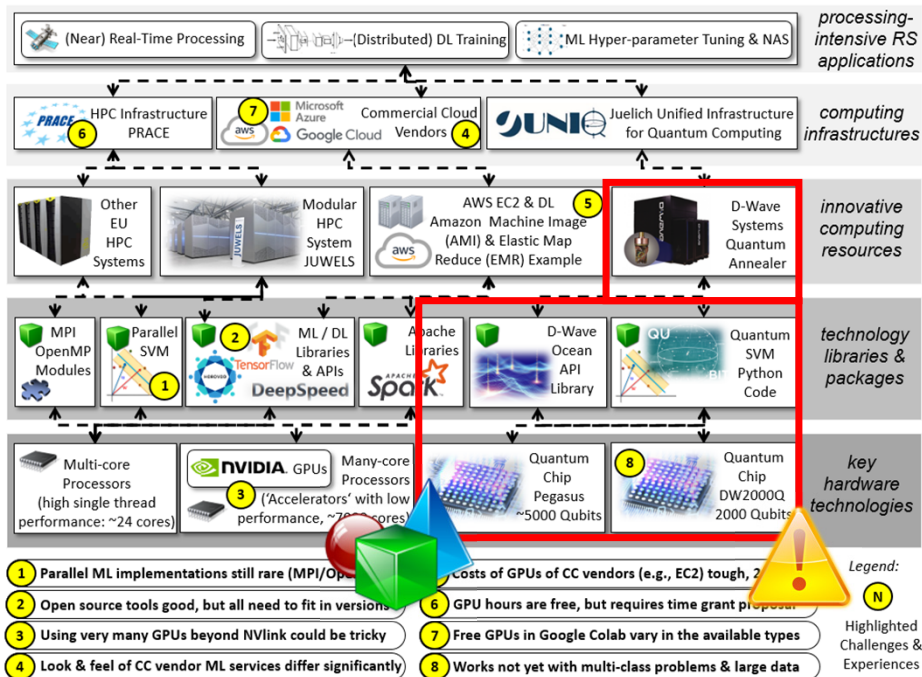
OpenML



**Continuously Updating**



# Innovative Quantum Computing Hardware as 'Accelerators'

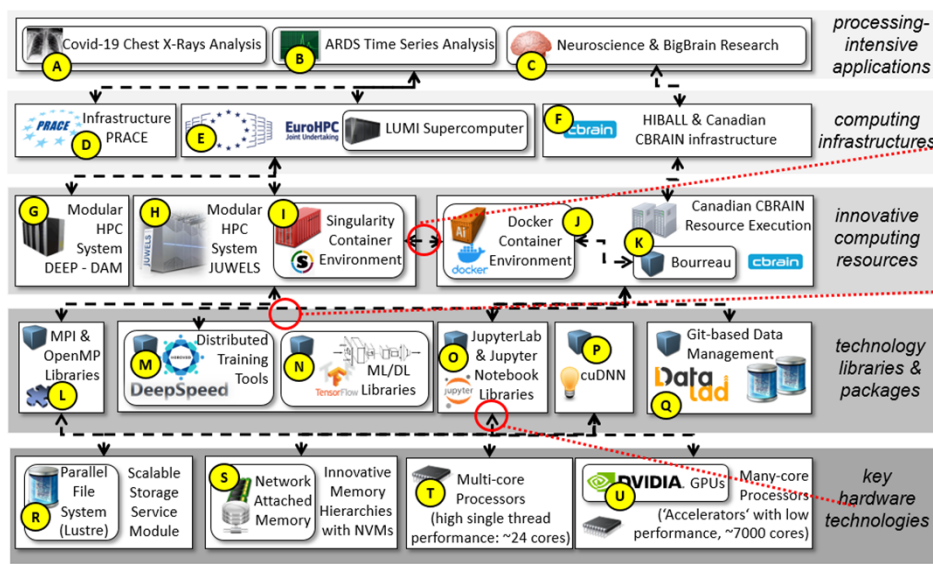


- 1 Parallel ML implementations still rare (MPI/OpenMP)
  - 2 Open source tools good, but all need to fit in versions
  - 3 Using very many GPUs beyond NVlink could be tricky
  - 4 Look & feel of CC vendor ML services differ significantly
  - 5 Costs of GPUs of CC vendors (e.g., EC2) tough, 2000 Qubits
  - 6 GPU hours are free, but requires time grant proposal
  - 7 Free GPUs in Google Colab vary in the available types
  - 8 Works not yet with multi-class problems & large data
- Legend: ● Highlighted Challenges & Experiences

Riedel, M., Cavallaro, G., Benediktsson, J.A.: Practice and Experience in using Parallel and Scalable Machine learning in Remote Sensing from HPC over Cloud to Quantum Computing, in conference proceedings of the IEEE IGARSS Conference, Brussels, Belgium, 2021, Physical and Online event, to appear <https://igarss2021.com/>



Riedel, M., Sedona, R., Barakat, C., Einarsson, P., Hassanian, R., Cavallaro, G., Book, M., Neukirchen, H., Lintermann, A.: Practice and Experience in using Parallel and Scalable Machine learning with Heterogenous Modular Supercomputing Architectures, in conference proceedings of the IEEE IDPS Conference, Heterogenous Computing Workshop (HCW), Portland, USA, 2021, Online, to appear <https://www.ipdps.org/>



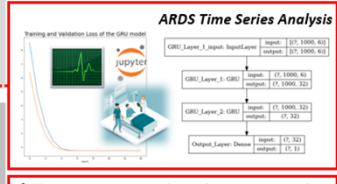
```

Some preparation
$ mkdir winterschool_cache winterschool_tmp
$ chmod -w winterschool_cache
$ export SINGULARITY_CACHEDIR=$(mktemp -d -p "$(pwd)/winterschool_cache")
$ export SINGULARITY_TMPDIR=$(mktemp -d -p "$(pwd)/winterschool_tmp")

Pull the docker image:
$ cd winterschool
$ singularity pull hvs.sif docker://glatard/hvs

Step into the container:
$ singularity shell --hvs.sif
(the prompt changes to '$Singularity')

download a dataset:
$ git config --global user.name "Your name"
$ git config --global user.email "yourname@gmail.com"
Singularity dataset install https://github.com/COMP-PCMD/comp-dataset.git
    
```



```

# /bin/bash
# load required modules
module purge
module use $SCRATCH/stages
module load Stages/2020
module load GCCCompiler/3.8.0
module load Python/3.8.5
module load Tensorflow/2.3.1-Python-3.8.5
module load OpenCV/3.8-Python-3.8.5
# activate python virtual environment
source ~/project/braining2021/ing1/conda/jupyter/conda/ing1/braining2021/activate
# ensure python packages installed in the virtual environment are always preferred
export PYTHONPATH=$(python -c 'import sys; print(sys.path[0])')
export PYTHONPATH=$(python -c 'import sys; print(sys.path[0])')
    
```



drive. enable. innovate.



The CoE RAISE project receives funding from the European Union's Horizon 2020 – Research and Innovation Framework Programme H2020-INFRAEDI-2019-1 under grant agreement no. 951733