



UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING,
MECHANICAL ENGINEERING AND COMPUTER SCIENCE



WP₂ AI- & HPC-Cross Methods at Exascale

Prof. Dr. – Ing. Morris Riedel

School of Engineering & Natural Sciences, University of Iceland

2021-01-22, RAISE Kick-Off Meeting, Online



@ProfDrMorrisRiedel



@Morris Riedel



@MorrisRiedel



@MorrisRiedel



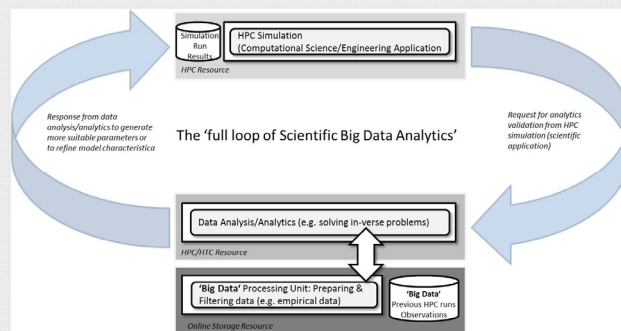
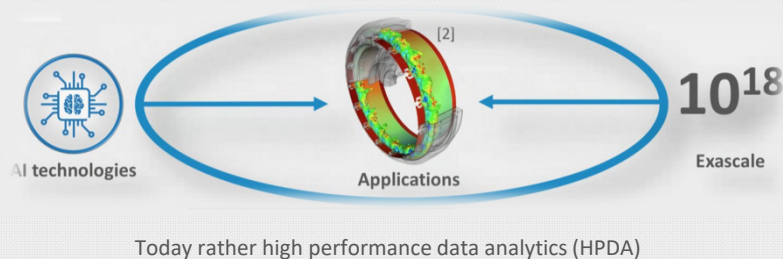
<https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg>

morris@hi.is



WP2 Vision – Intertwined HPC Simulations & AI – ‘full loop’ ?

➤ What means AI & HPC Cross Methods at Exascale?



Lippert, T., Mallmann, D., Riedel, M.: **Scientific Big Data Analytics by HPC**, in Symposium proceedings of NIC Symposium 2016 – publication Series of the John von Neumann Institute for Computing (NIC), NIC Series 48 (417), ISBN 978-3-95806-109-5, February 11-12, 2016, Juelich, Germany

Energy Meteorological In-Situ Big Data Analytics

Morris Riedel^{1,2}, Jonas Berndt^{1,3}, Charlotte Hoppe^{1,3}, Hendrik Elbern^{1,3}

contact: m.riedel@fz-juelich.de
j.berndt@fz-juelich.de

¹Institute of Energy and Climate Research (IEK-8), Forschungszentrum Jülich GmbH, Jülich, Germany
²University of Iceland, Reykjavik, Iceland
³Rheinisch Institute for Environmental Research at the University of Cologne, Cologne, Germany

Background & Objective

The stochastic nature of weather imposes wind and solar power as an uncertain source of electrical energy. Stable power grid management and energy trade on stock markets call for improvement of probabilistic wind and solar power forecasts. The major potential lies in the improvement of the underlying weather forecast.

We make use of various perturbation techniques in the frame of a regional meteorological ensemble with 1024 members to capture extreme error events and to improve skill scores of short and shortest range forecasts of wind speed at hub height (~100m) and irradiance.

A data mining application shall identify the relationship between observation compliance and perturbation techniques. This information serves as a basis to improve the further generation of ensemble members within a particle filter algorithm.

Meteorological Ensemble

- An ultra large ensemble version of the Weather Research and Forecast model (WRF) as part of ESIRAS (Ensemble for Stochastic Integration of Atmospheric Simulation), which provides a comprehensive probability density evolution of the model state.
- Computational efficient implementation on the JUQUEEN, which realizes communication between the ensemble members by introducing a second stage of MPI parallelism
- Initial values and lateral boundary values from the global ECMWF and GFS ensembles
- A broad variety of state-of-the-art techniques of uncertainty representation within the model (SKEDS – Stochastic Kinetic-Energy Backscatter Scheme, SPPT – Stochastic Perturbed Parameterization Tendency, perturbation of surface values, etc.)

Stamp plots of a 12-member selection with either GFS or ECMWF boundary and initial conditions and SKEDS model perturbation.

Coupled Forecast-Analysis System

Energy Meteorological In-Situ Analytics Application Loop

Computational Resources
The amount of computational resources is proportional to the ensemble member number. As an elementary example, 1024 members require 65536 cores on the IBM BlueGene Q system (JUQUEEN) to accomplish a 24 hour simulation in approximately 1.5 hours (the NWP comprises the majority, with I/O tasks and on-line data mining being of secondary relevance).

Amount of data to be managed

In-Situ Analytics Application Loop

The coupled forecast-analysis system combines the meteorological forecast, particle filtering and data miner in one application loop. Due to high computational demands, special focus is

Data Mining Methodology

- Classification methodology trains a model of the data given training set $T = (x_1, y_1), \dots, (x_n, y_n)$
- Supervised classification problem: Experts provide labels y_i data of WRF ensembles x_i quality
- Multi-class design enabling scientists to label with an increasing range of quality classes
- The trained model is then used with unseen WRF data to assign it to a quality class
- Depending on the quality class predicted by the model WRF, ensembles are canceled/continued
- Chosen algorithm to create a model are Support Vector Machines (SVM) with kernel methods

In this simplified 2D example of a two class problem (red & blue WRF ensemble members), SVM will achieve the optimal decision boundary between both classes. While many lines will separate both classes in this example, SVM will automatically learn via the training set the blue line as shown in the illustration. The interesting property of this blue line is that it offers the best generalization out of sample. In other words, once the training data has been used to train the model, the model will work quite well with unseen WRF ensemble members.

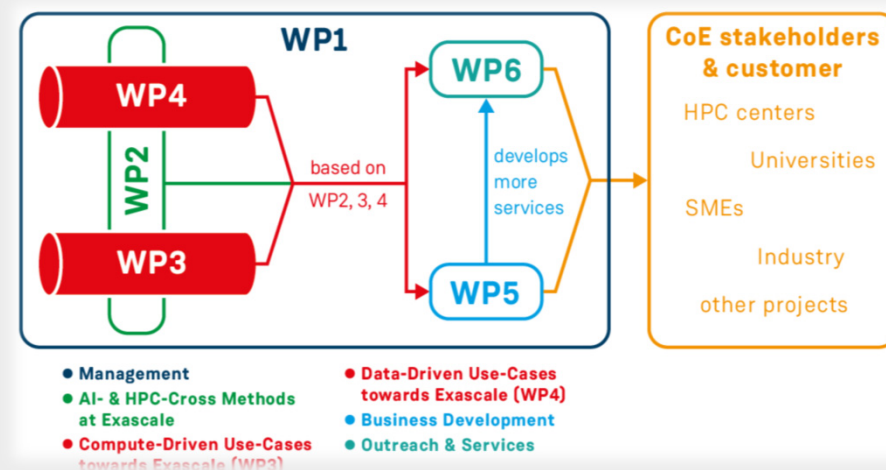
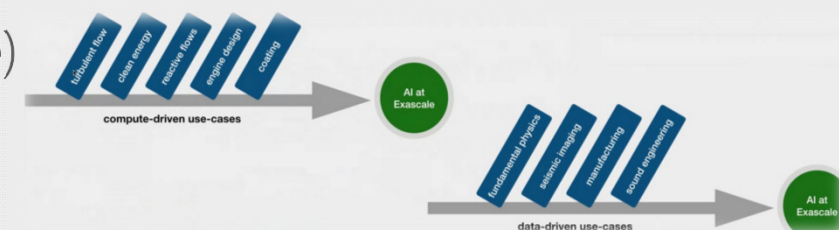
- Train a model with support vectors (cf. orange data in figure) is computationally complex
- SVM needs to find the best decision boundary (aka points most far away from existing points)
- It is a constraint optimization problem solved inherently with sequential minimal optimization
- The optimization problem aims to maximize the margin (above orange background color)

WP2 Partners & Funding Levels

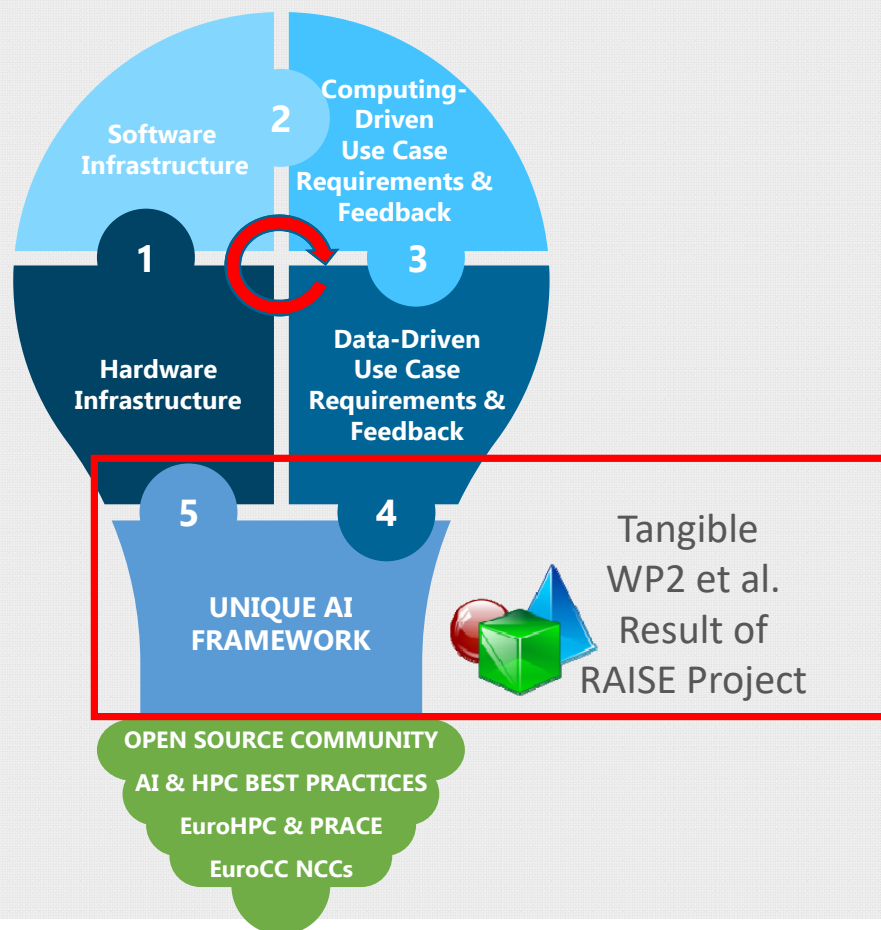
Work package title	AI- and HPC-Cross Methods at Exascale											
Participant No.	1	2	3	4	5	6	7	8	9	10	11	TOT
Short name of participant	FZJ	UOI	CYI	RWTH	BSC	CERN	CERFACS	BULL	RTU	FM	SAFRAN	
PM per partner	43	24	0	10	8	8	0	8	22	12	0	135
Start month	1			End month								36

WP2 & Connections to other WPs – ‘WP2 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



WP2 Objectives – 'AI & HPC at Exascale backbone'



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 (→ WP3)

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 pyTorch

Data-driven Use Cases Requirements & Feedback (→ WP4)

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

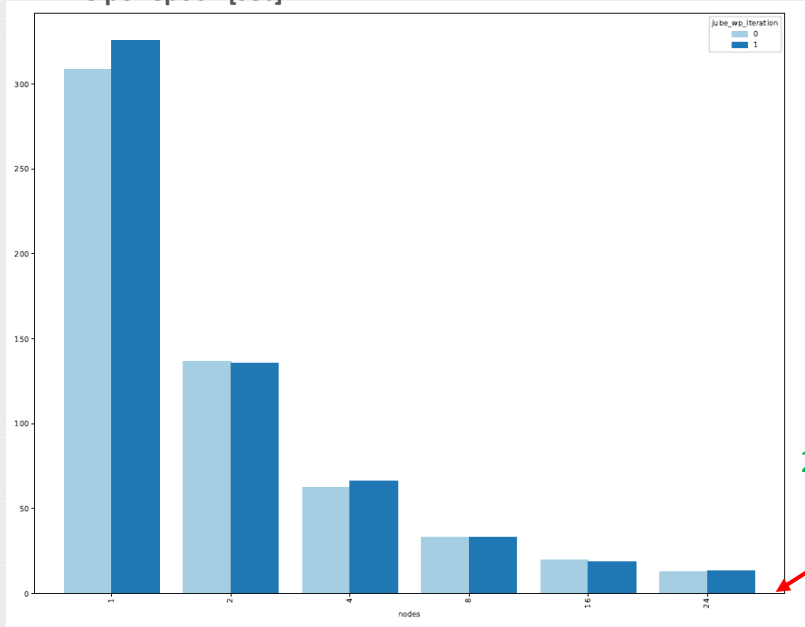
Examples: Deployment blueprint by using AI training on cluster module & inference/testing on booster

UNIQUE AI FRAMEWORK

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

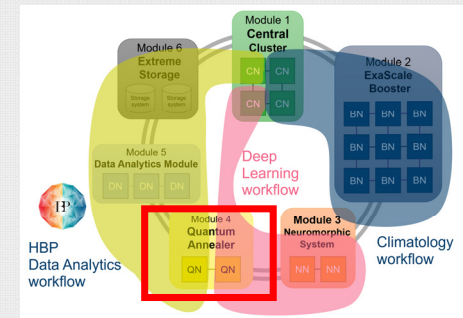
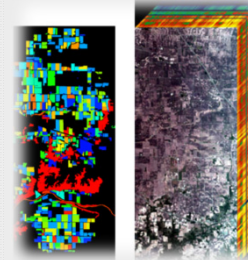
WP2 Objectives – Can AI do Exascale & use Disruptive Technologies?

Time per epoch [sec]



Example from 2019:
Using partition of the JUWELS
system has 56 compute nodes,
each with 4 NVIDIA V100 GPUs
(equipped with 16 GB of memory)

24 nodes x 4 GPUs = 96 GPUs



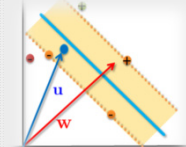
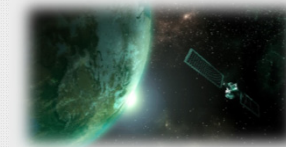
```
In [ ]: from quantum_SVM import *

# Hyperparameters
B=[2,3,5,10]
K=[2,3]
xi=[0.1,0.5]
gamma=[-1,0.125,0.25,0.5,1,2,4,8]
n_experiments=len(B)*len(K)*len(xi)*len(gamma)

hyperparameters=np.zeros([n_experiments,4], dtype=float)

path_data_keys='input_datasets/calibration/*id_dataset*/*'
data_key = 'id_dataset*calibrain*'
path_outs='outputs/calibration/*id_dataset*/*'

trainacc=np.zeros([fold], dtype=float)
trainauc=np.zeros([fold], dtype=float)
trainaup=np.zeros([fold], dtype=float)
```



Sedona, R., Cavallaro, G., Jitsev, J., Strube, A., Riedel, M., Book, M.: [SCALING UP A MULTISPECTRAL RESNET-50 TO 128 GPUS](#), in conference proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), September 26 – October 2nd, 2020, Virtual Conference, Hawaii, USA, to appear

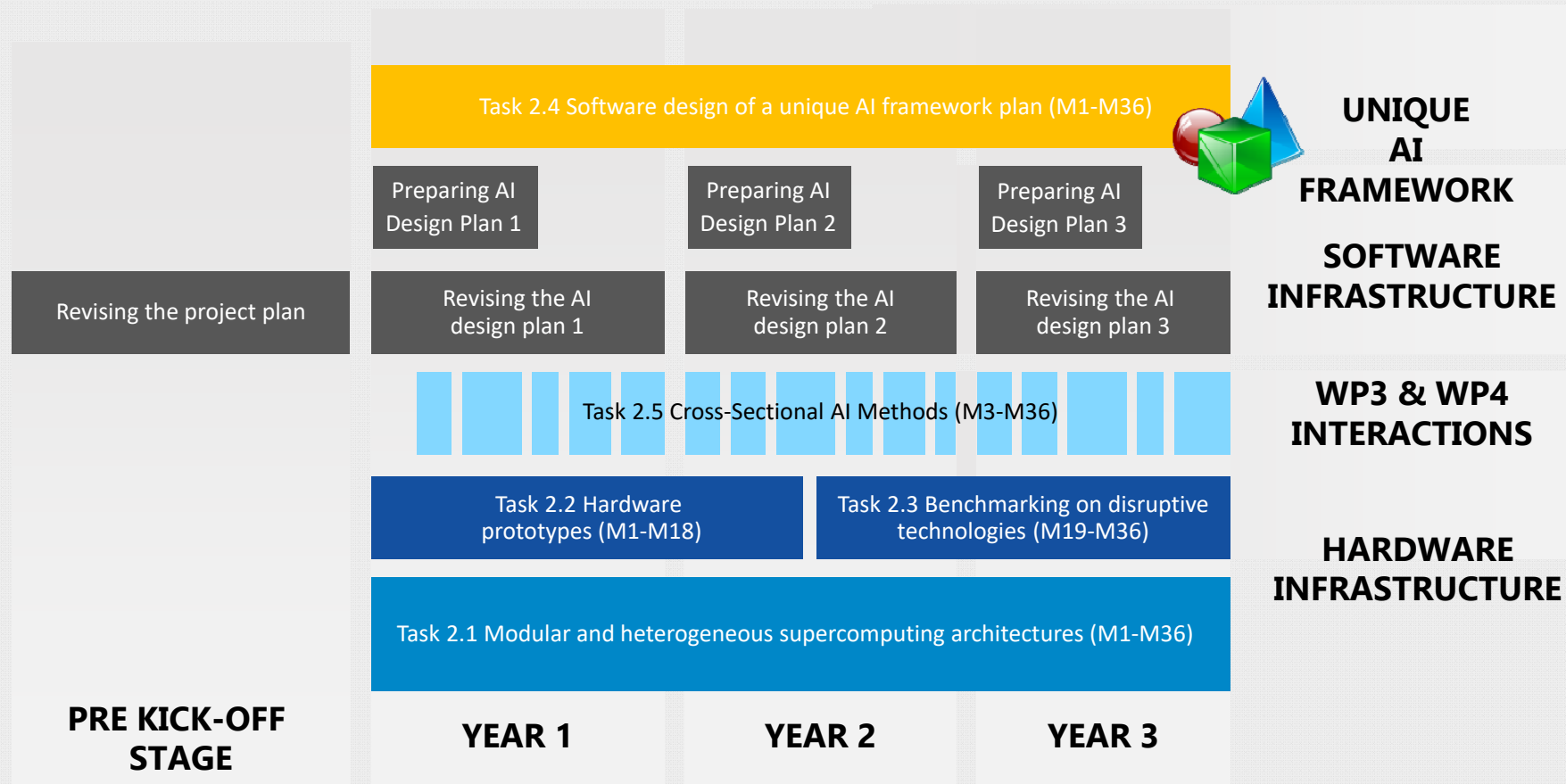
Sedona, R., Cavallaro, G., Jitsev, J., Strube, A., Riedel, M., Benediktsson, J.A.: [Remote Sensing Big Data Classification with High Performance Distributed Deep Learning](#), Journal of Remote Sensing, Multidisciplinary Digital Publishing Institute (MDPI), Special Issue on Analysis of Big Data in Remote Sensing, 2019

Cavallaro, G., Willsch, D., Willsch, M., Michielsen, K., Riedel, M.: [APPROACHING REMOTE SENSING IMAGE CLASSIFICATION WITH ENSEMBLES OF SUPPORT VECTOR MACHINES ON THE D-WAVE QUANTUM ANNEALER](#), in conference proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), September 26 – October 2nd, 2020, Virtual Conference, Hawaii, USA, to appear

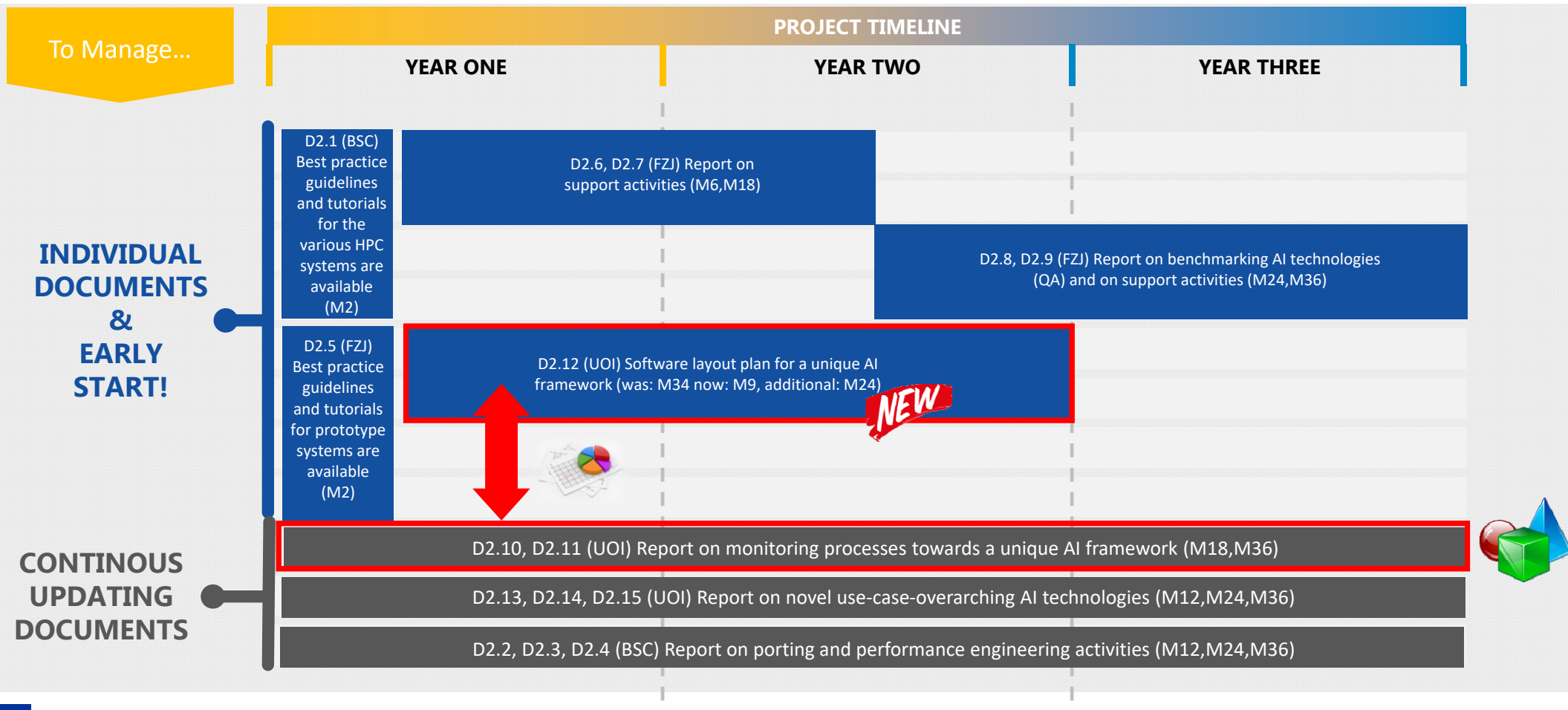
WP2 Tasks



TIME →



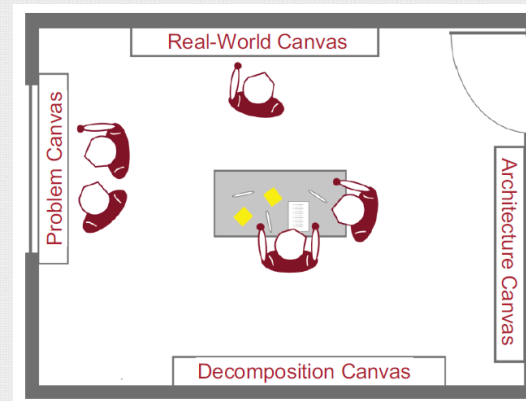
WP2 Work on Deliverables & Milestones



WP2 Expectations & Requirements

➤ Start: Interaction Room Process

- 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** (Software Engineering, 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



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



drive. enable. innovate.



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