WP2 AI- & HPC-Cross Methods at Exascale – Monthly Meeting

Prof. Dr. – Ing. Morris Riedel et al.
School of Engineering & Natural Sciences, University of Iceland

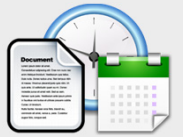
2022-07-29, RAISE WP2 Monthly Meeting July 2022, Online

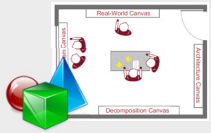
@ProfDrMorrisRiedel @Morris Riedel @MorrisRiedel @MorrisRiedel
https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg morris@hi.is

WP2 Meeting July – Welcome & Agenda

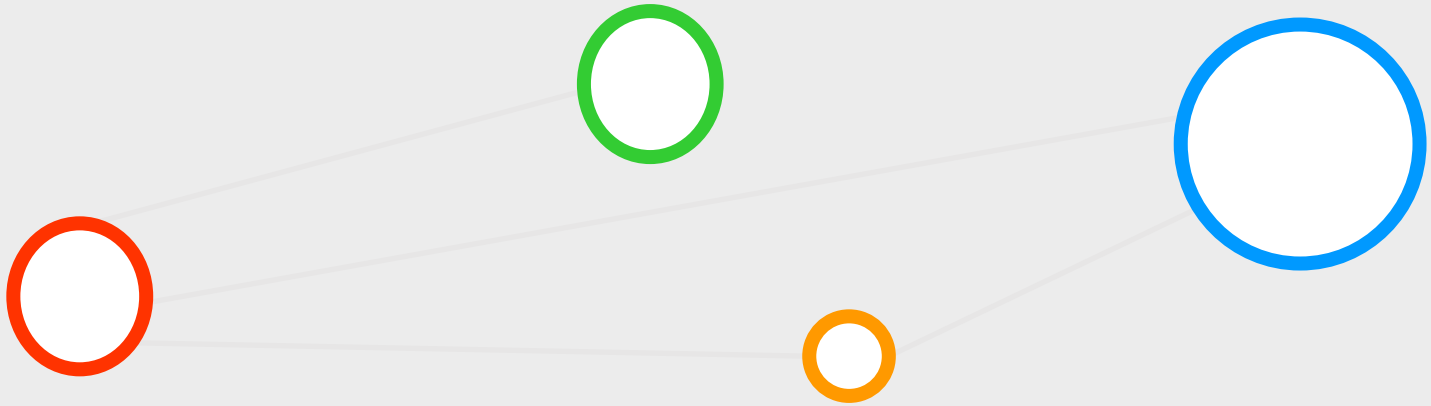
1. Approval of minutes from Monthly Meeting June 2022
   - (All), ~5 Min
2. Review WP2 Status on Interaction Rooms
   - (Morris Riedel, Matthias Book, Helmut Neukirchen), ~10 Min
3. September Review Preparation (M21)
   - (Morris, Andreas, et al.), ~20 Min
4. Realization of AI Framework
   - (Eray, Morris et al.), ~10 Min
5. Status WP2 Training Plans
   - (Morris et al.), ~5 Min
6. Compelling Scoreboard Review & Next Steps
   - (All), ~10 Min

Thanks for joining in vacation period!

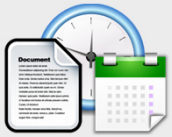




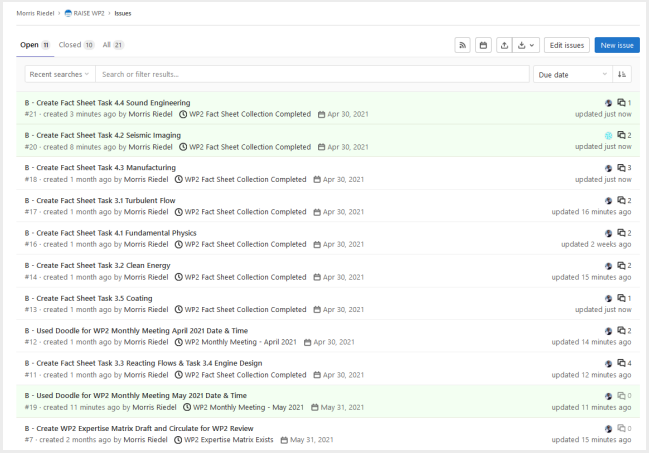
Agenda Item (1) – Minutes Approval – June 2022

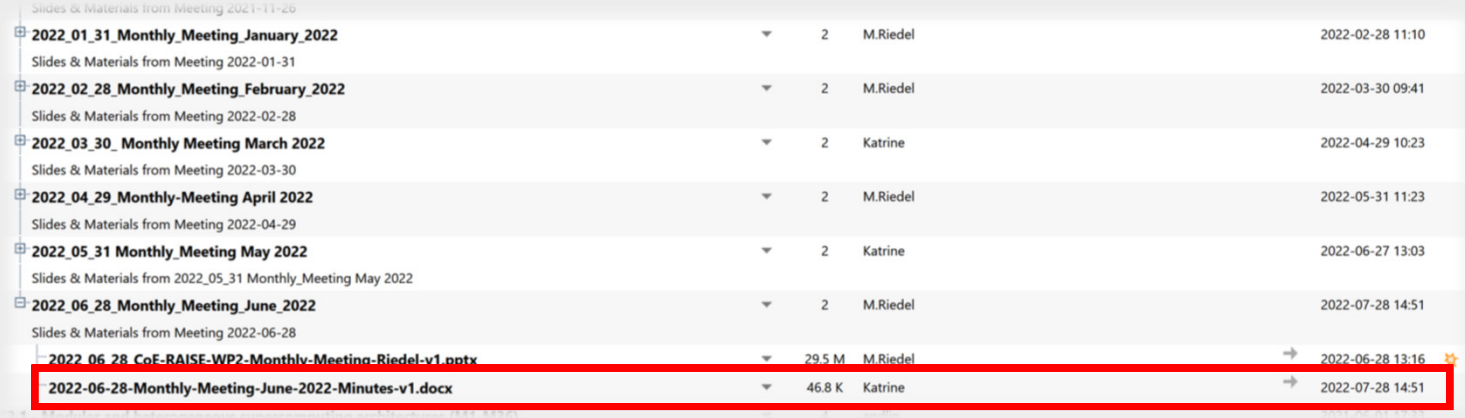


Minutes Approval – Monthly Meeting June 2022

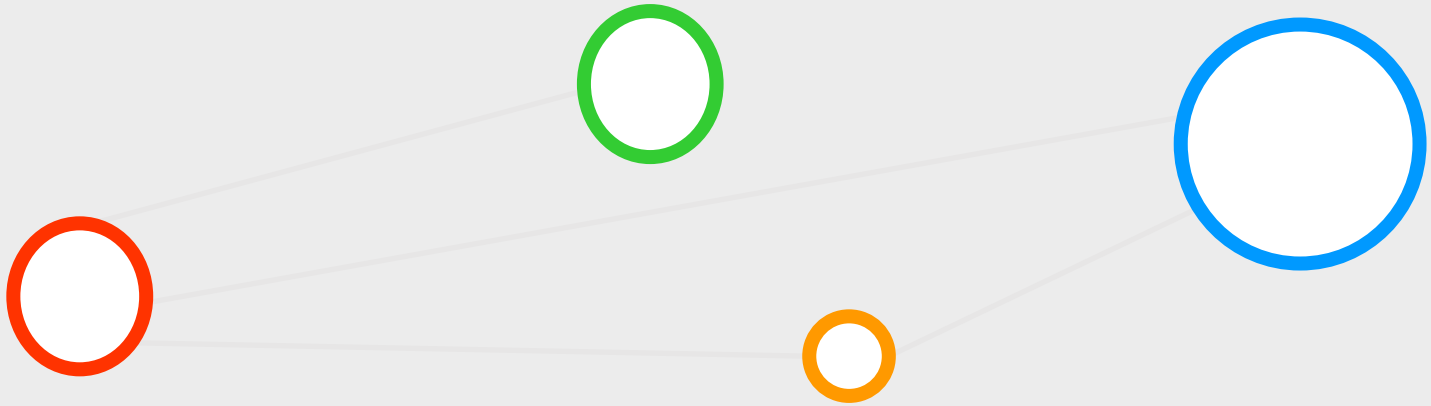


- Minutes available in BSCW
  - https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/3340884
  - TBD(all): Any objections or additions/changes?

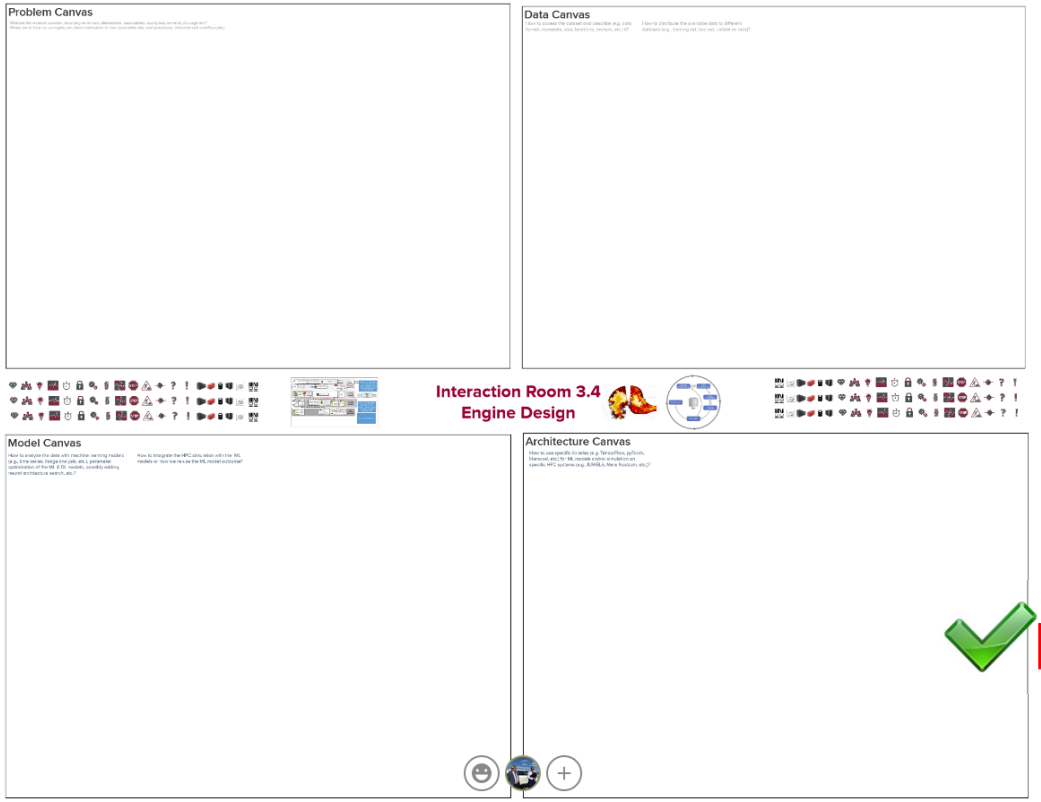


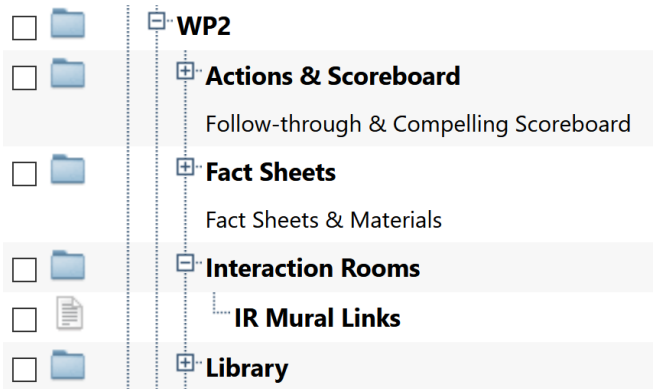


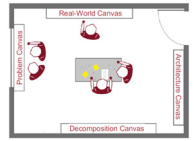
Agenda Item (2) – Review WP2 Status on Interaction Rooms



Interaction Rooms via MURAL Boards & Milestone Inputs







IR Mural Links

IR3.1 Turbulent Flow: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377866397/8613c384d54f66fb5e78599ff307a4ce8a9090c0?sender=u15e3008bb41d6628a5bb5701

IR3.2 Clean Energy: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377887905/cb44cca3eedd3bb9964fbfa36af16b1bfcce085f?sender=u15e3008bb41d6628a5bb5701

IR3.3 Reactive Flows: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377959022/0c363886f24833ecb19b025d87324b57fd50e2db?sender=u15e3008bb41d6628a5bb5701

IR3.4 Engine Design: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377976343/8d7aba6be09af3b2ffd305d2f709c53661ac889d?sender=u15e3008bb41d6628a5bb5701

IR3.5 Coating: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377991014/7a5d7e1eaf230178342d1e1d4a84d656d9055d52?sender=u15e3008bb41d6628a5bb5701

IR4.1 Fundamental Physics: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378007555/6f0d5285fcaec5eafa515bd6676c84d8b4879d39?sender=u15e3008bb41d6628a5bb5701

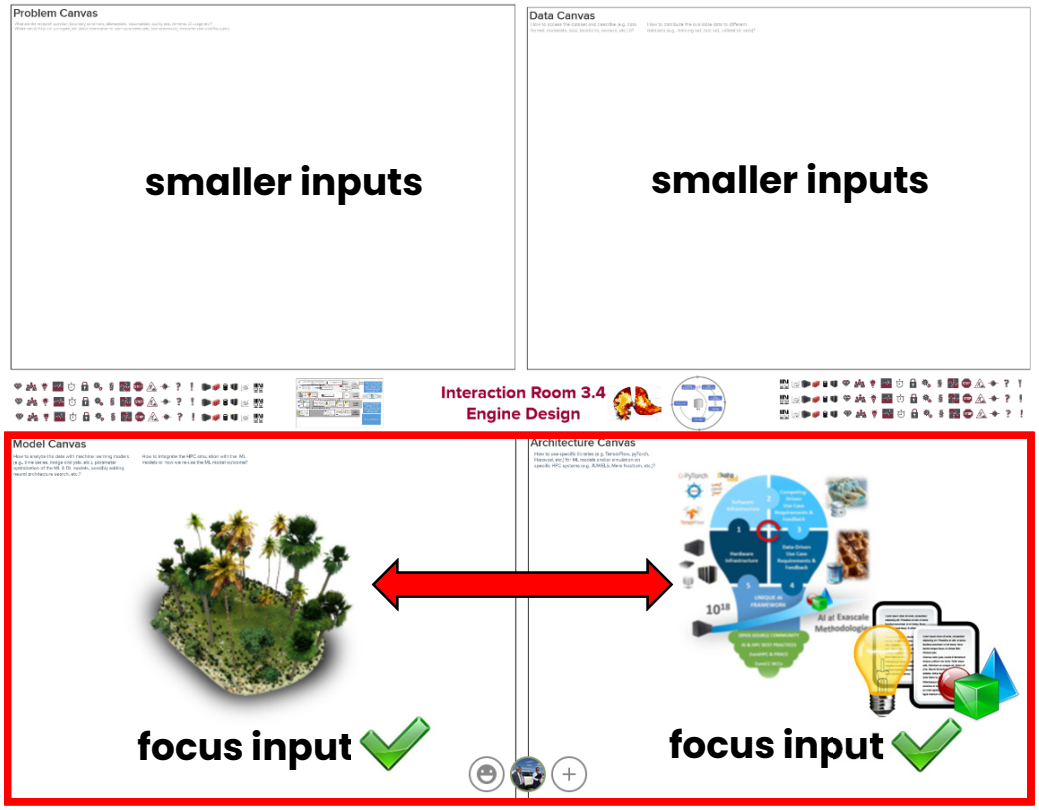
IR4.2 Seismic Imaging: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378023838/a0b9503abb837ae3e28af4bb8d9adbec33874998?sender=u15e3008bb41d6628a5bb5701

IR4.3 Manufacturing: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378038069/93df6fa7a41093f4eaae7be9d72979de2ba42b9d?sender=u15e3008bb41d6628a5bb5701

IR4.4 Sound Engineering: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378050431/b5fa122190024040 59f90a4bbb0101fa379a8503?sender=u15e3008bb41d6628a5bb5701

- 2nd Iteration with all teams completed – Strategy for Task 3.4 has been discussed!
- https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/3591551

MURAL Board contents for Deliverables & Milestones



- WP2
  - Actions & Scoreboard
    - Follow-through & Compelling Scoreboard
  - Fact Sheets
    - Fact Sheets & Materials
  - Interaction Rooms
    - IR Mural Links
  - Library

IR Mural Links

IR3.1 Turbulent Flow: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377866397/8613c384d54f66fb5e78599ff307a4ce8a9090c0?sender=u15e3008bb41d6628a5bb5701

IR3.2 Clean Energy: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377887905/cb44cca3eedd3bb9964fbfa36af16b1bfcce085f?sender=u15e3008bb41d6628a5bb5701

IR3.3 Reactive Flows: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377959022/0c363886f24833ecb19b025d87324b57fd50e2db?sender=u15e3008bb41d6628a5bb5701

IR3.4 Engine Design: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377976343/8d7aba6be09af3b2ffd305d2f709c53661ac889d?sender=u15e3008bb41d6628a5bb5701

IR3.5 Coating: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621377991014/7a5d7e1eaf230178342d1e1d4a84d656d9055d52?sender=u15e3008bb41d6628a5bb5701

IR4.1 Fundamental Physics: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378007555/6f0d5285fcaec5eafa515bd6676c84d8b4879d39?sender=u15e3008bb41d6628a5bb5701

IR4.2 Seismic Imaging: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378023838/a0b9503abb837ae3e28af4bb8d9adbec33874998?sender=u15e3008bb41d6628a5bb5701

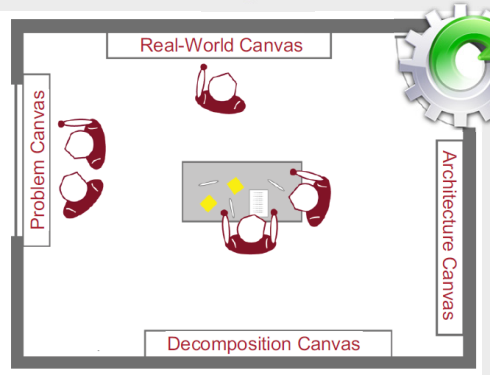
IR4.3 Manufacturing: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378038069/93df6fa7a41093f4eaae7be9d72979dc2ba42b9d?sender=u15e3008bb41d6628a5bb5701

IR4.4 Sound Engineering: https://app.mural.co/t/matthiasbook8855/m/matthiasbook8855/1621378050431/b5fa12219002404059f90a4bbb0101fa379a8503?sender=u15e3008bb41d6628a5bb5701

Interaction Room Status & Discussions – WP3/WP4 Overview

- WP3 (second round IRs)
  - T3.1: Turbulent Flow → done
  - T3.2: Clean Energy → done
  - T3.3: Reactive Flows → done
  - T3.4: Engine design → done
  - T3.5: Coating → done
- WP4 (second round IRs)
  - T4.1: Fundamental physics → done
  - T4.2: Seismic imaging → done
  - T4.3: Manufacturing → done
  - T4.4: Sound engineering → done
- 2nd iteration of Interaction Rooms → done

Continously Updating



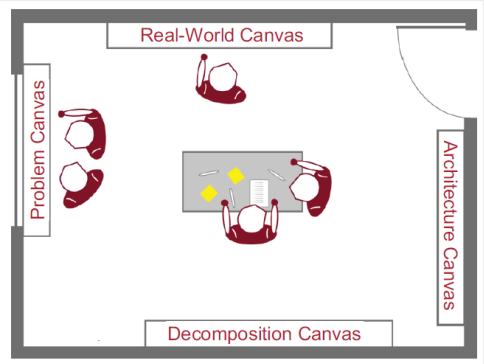
Use Case	AE	PIML	ANNs	CNN		NO	SMs			GNN	IN	LSTM	GRU
Details	*CAE*		*RBF-ANN*	*U-Net*	*RESNET*	*FNO*	*AR*	*ARMA*	*ARIMA*		*JEDI-net*		
AI for turbulent boundary layers	X	X											
AI for wind farm layout optimization			X				X	X	X				
AI for data-driven models in reacting flows				X						X			
Smart 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

- Next round Interaction Rooms after Review
  - Carve out more details on AI/HPC methods
  - Contribute to the Unique AI Framework
  - Update our HPC/AI Methods Matrix

Cross HPC/AI Methods Table – IR Results (D2.10)

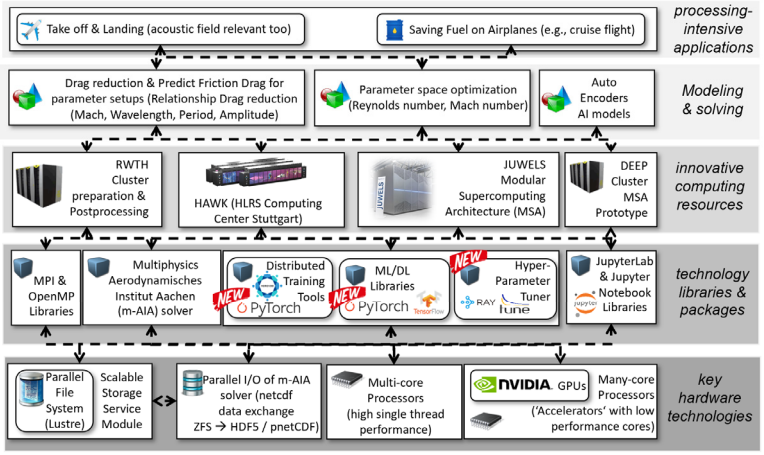
Input to Deliverable D2.10

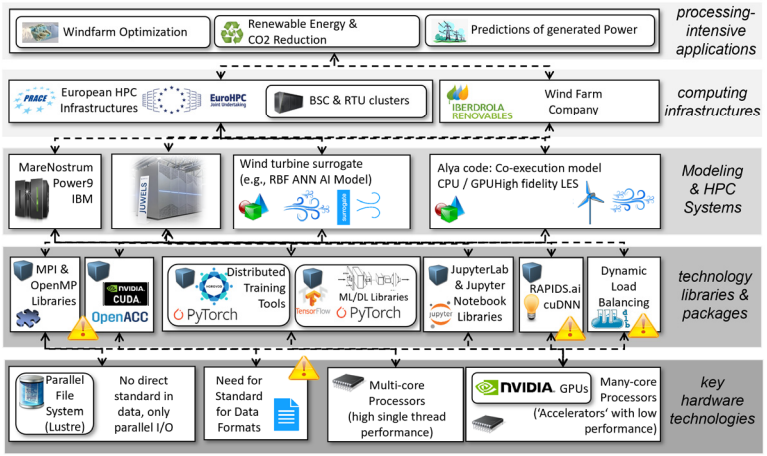
- Interaction Rooms
  - Update of Matrix
  - Components relatively constant & common
  - Methods change & new methods added (e.g., Transformers, RFs)

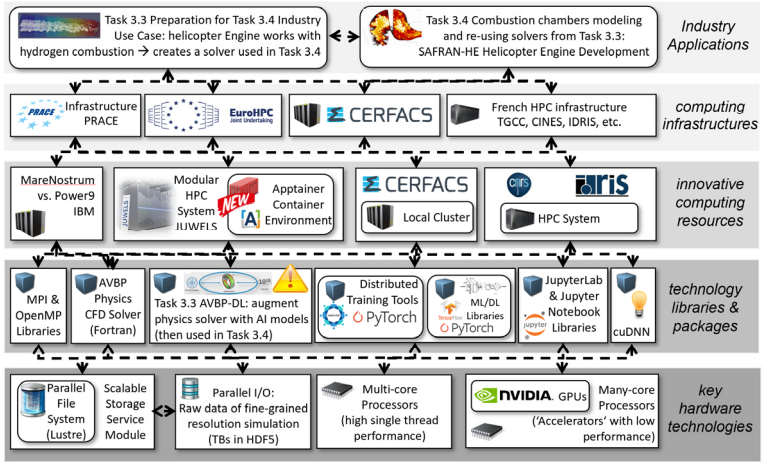


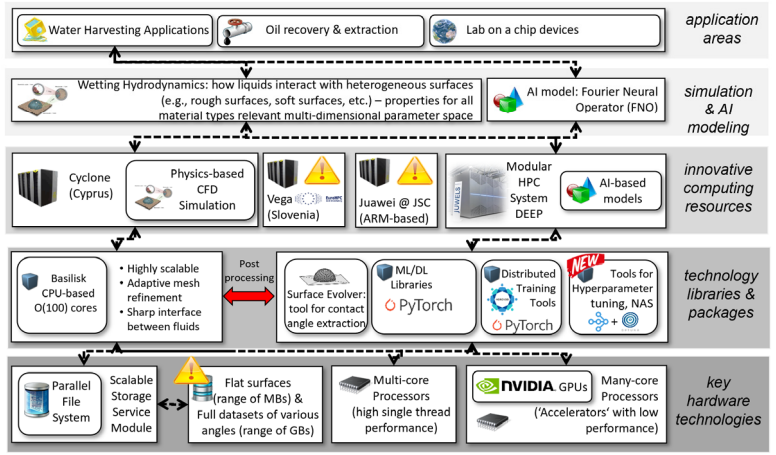
Use Case	AE	PINN	ANNs		CNN		NO	GNN		RNN		GAN	TF				SVM	RF
Details	*CAE*		*ANN*	*RBF-ANN*	*U-Net*	*RES NET*	*FNO*	*MLPF*	*GAT*	*LSTM*	*GRU*	*WGAN*	*MViT*	*ViViT*	*Swin*			
AI for turbulent boundary layers	X	X	X									X						
AI for wind farm layout optimization				X												X		
AI for data-driven models in reacting flows					X				X									
Smart models for next generation aircraft engine design					X				X									
AI for wetting hydrodynamics	X	X					X			X								
Event reconstruction and classification at the CERN HL-LHC use case								X										
Seismic imaging with remote sensing for energy applications	X	X				X	X			X	X					X	X	X
Detect-free metal additive manufacturing	X		X									X	X	X	X			
Sound Engineering	X		X															

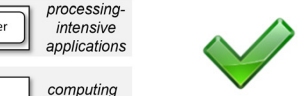
Co-Design of SW Framework – IR Results (see D2.10)





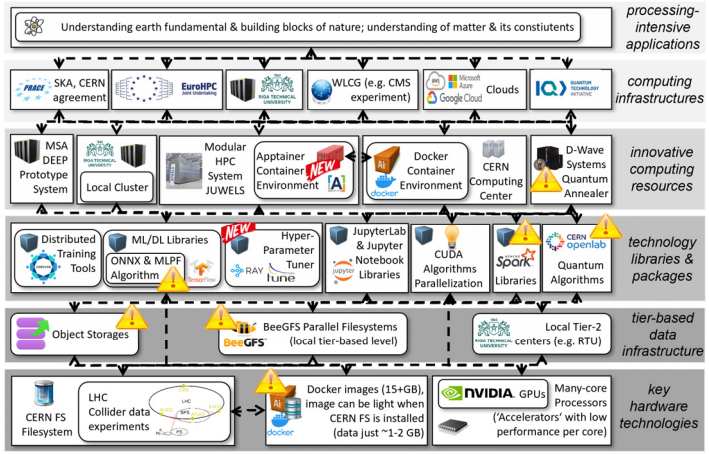


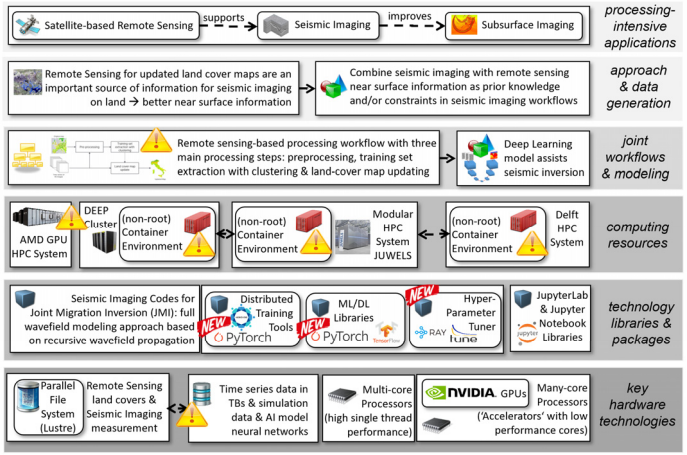


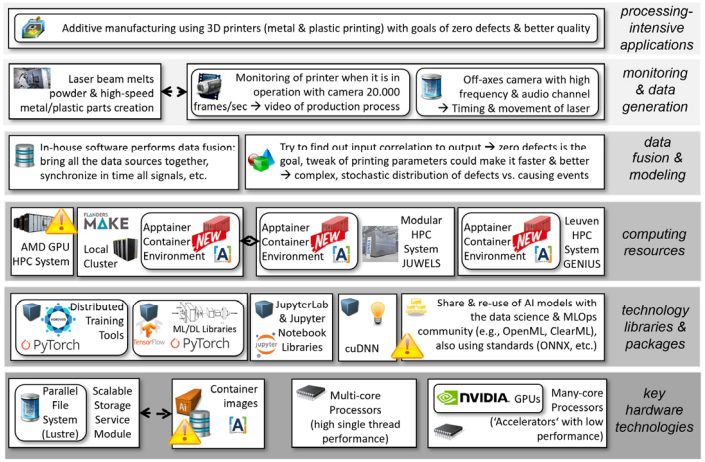


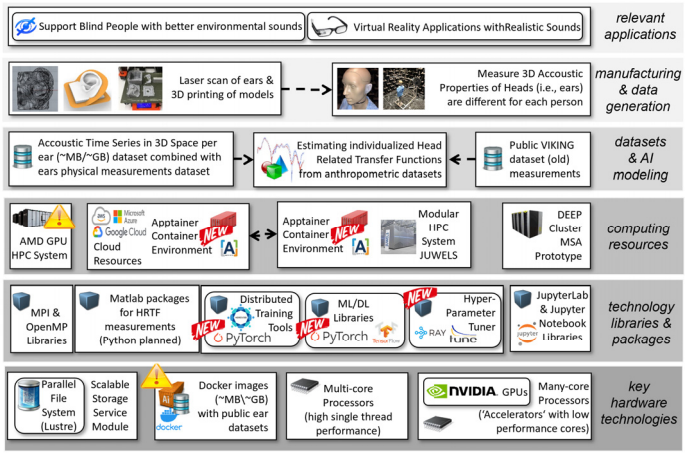
Fact Sheets available for each use case (Task 3.3 / Task 3.4 in one)

Co-Design of SW Framework – IR Results (see D2.10)



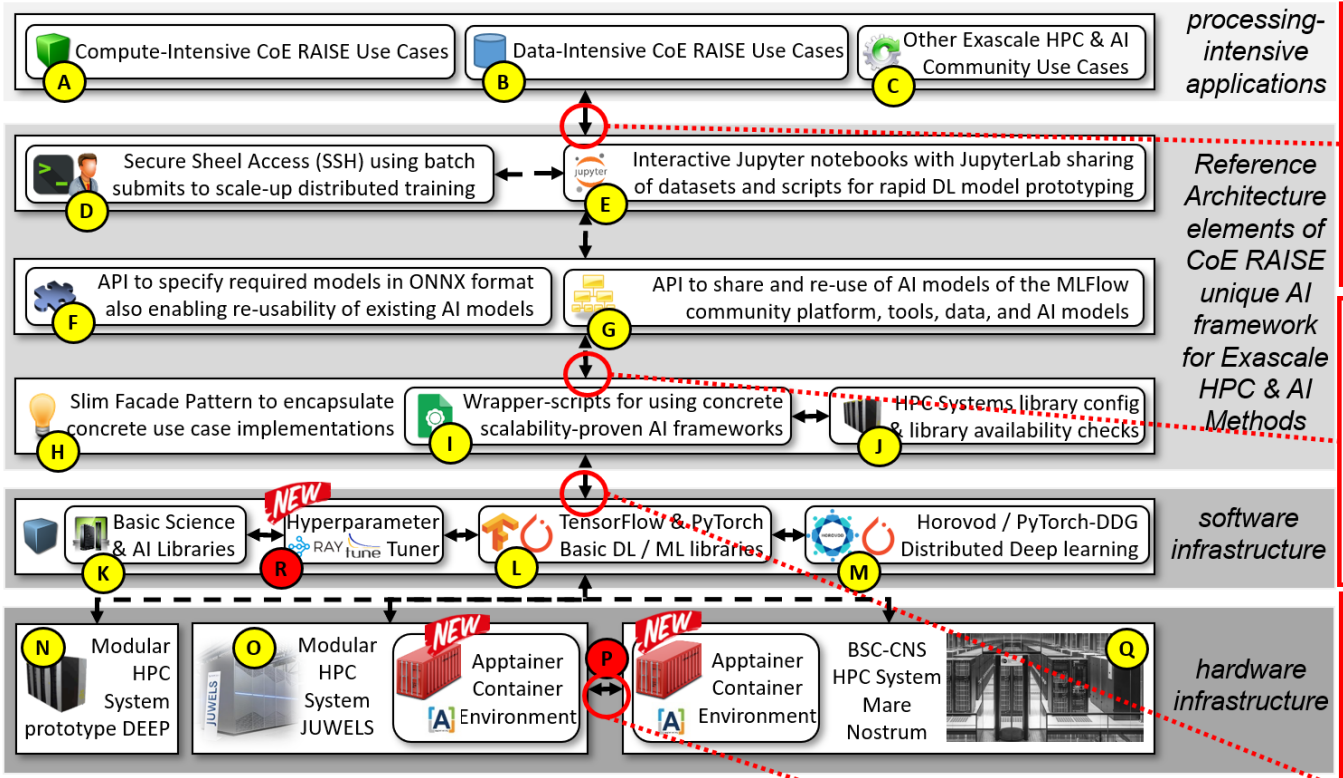






Fact Sheets available for each use case

Realization of SW Framework – IR Results (see D2.10)



✓ 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 NEW

❖ Parts of the framework layout plan is to provide lightweight & abstract Python APIs building on ONNX enabling exchange with MLFlow, OpenML, ClearML, etc.

✓ RQ1, RQ2

❖ 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

❖ Part of the framework layout plan is to provide containers in Apptainer with prepackaged datasets and required software stacks needed for AI models

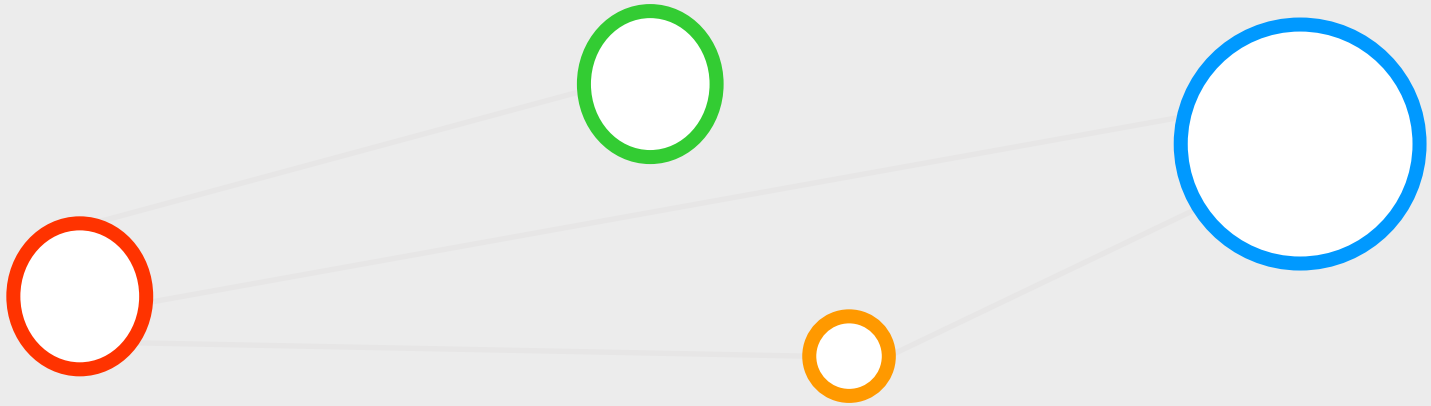
Next version might include coupler (Atos)?

Continously Updating!

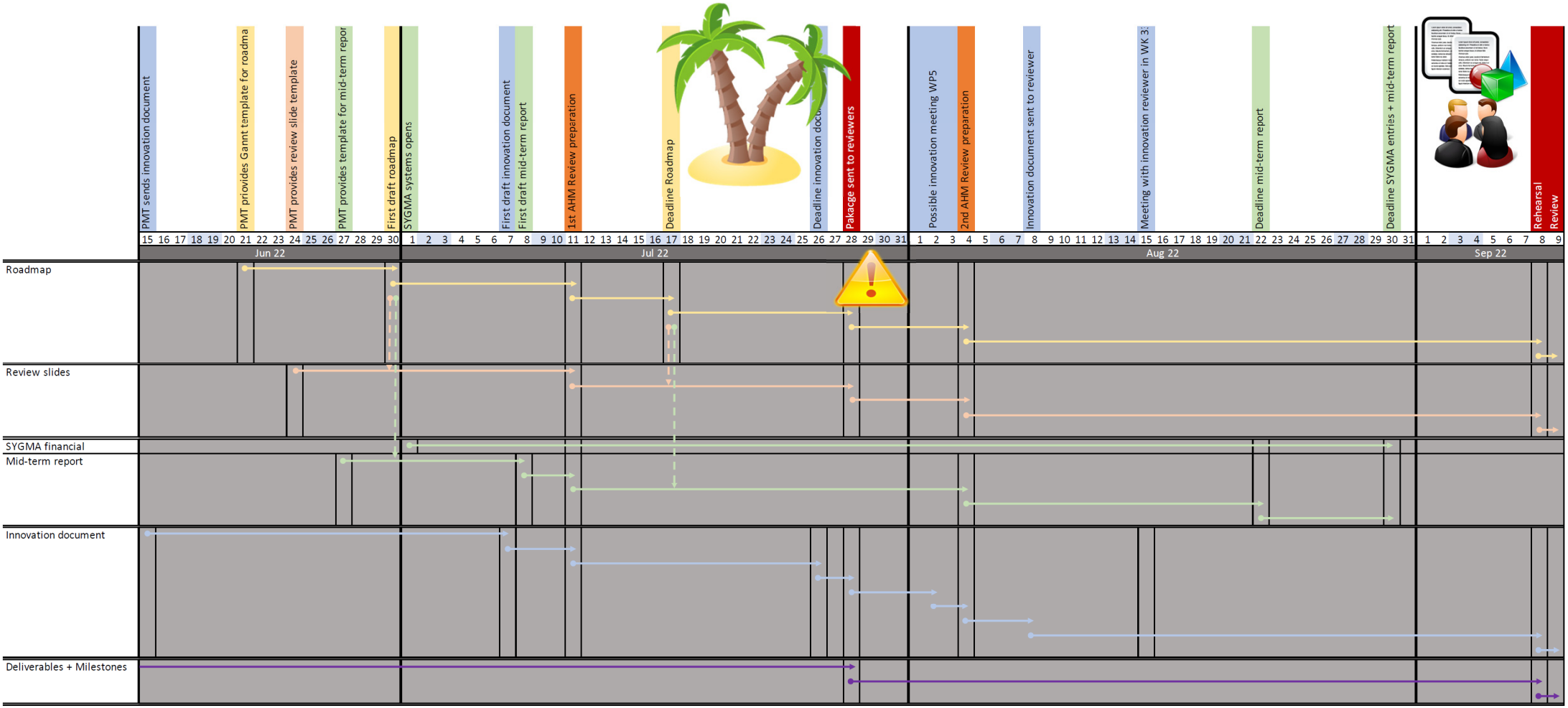
Legend:

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

Agenda Item (3) – September Review Preparation (M21)



September Review Preparation: 08.09. + 09.09. (update!)



September Review Preparation: Roadmap Document

How can we prepare?

- A roadmap document is currently in preparation
  - It aims at
    - self-reflecting our activities
    - inspiring us to think about further requirements to fulfill our promises
    - inspiring us to think about our next steps
  - Basic structure:
    - Status of the various domains before the start of CoE RAISE
    - Changes achieved so far (impact)
    - Future plans
- Contents of the WP presentations
  - Overview of the Tasks and details on what has happened in the WP (requires input from the task leaders
  - Slides will be finalized on the day of the rehearsal (there will only be a single presenter)

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

Draft

Version: 0.1
Author(s): A. Lintermann (FZJ)
Contributor(s):
Date: 17.05.2022

- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.1

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

Draft

Version: 0.1
Author(s): A. Lintermann (FZJ)
Contributor(s):
Date: 17.05.2022

4.2.1 Task 2.1 - Modular and heterogeneous supercomputing architectures

[TODO Task leader BSC]: General introduction to the task

The fundamental objective of RAISE is To develop innovative AI methods on heterogeneous HPC architectures capable of scaling towards Exascale. To this end, task 2.1 provides all the necessary support for the porting of the codes and their optimizations on heterogeneous architectures, while ensuring constant availability of Tier-3 to Tier-0 supercomputers.

For this, the HPC centers of RAISE provide their computational resources for development and testing of RAISE's software. This includes the homogeneous and heterogeneous HPC systems found at the Tier-2 and Tier-3 centers of the consortium (UOI, RWTH, and RTU) as well as the cutting-edge HPC systems of the Tier-0 and Tier-1 providers (FZJ and BSC). That is, FZJ and BSC give access to their MSA/heterogeneous supercomputers JURECA, JUWELS, and MareNostrum 4 (MN4) general cluster and its three prototypes. Together with the use-case providers, the representative applications, where necessary, are jointly prepared for such heterogeneous systems together with the HPC centers of the consortium.

The AI experts further exploit the HPC architectures and file systems for relevant algorithms, i.e., they work on the enhancement of the scalability of existing ML/DL methods on dedicated or shared HPC components using, e.g., GPGPUs or other accelerator platforms, and test I/O performance.

The developers are supported by the HPC experts by means of performance engineering activities, best practice guidelines, and system-specific tutorials and manuals. The training courses of the HPC centers are offered to the use-case providers in line with Task 6.1.

Status at the beginning of the project (SoA)

[TODO Task leader]: Provide the state-of-the-art, or the general status of the domain, if applicable

In 2018, for the first time, "most of the flops added to the TOP500 list came from GPUs instead of CPUs[9]. Since then, the relative power delivered by GPUs with respect to CPUs has continued to increase. Although the integration of HPC and AI is still at an early stage, mixed-workflow will rapidly increase to fully exploit such systems. This was the hypothesis at the proposal stage, which is now being verified. Next figure shows the new TOP500 list, where we can observe that four of the five first supercomputers include accelerators.

???

- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.2

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

Draft

Version: 0.1

Author(s): A. Lintermann (FZJ)

Contributor(s):

Date: 17.05.2022

4.2.2 Task 2.2 - Hardware prototypes

Within CoE RAISE, the developers have access to various high-performance computing hardware prototypes. These hardware prototypes are provided by the project partners Barcelona Supercomputing Center (BSC), Spain, and Forschungszentrum Jülich (FZJ), Germany. These hardware prototypes here are of experimental nature, and they represent ideal playgrounds to test available codes on hardware that is not available in production High-Performance Computing (HPC) systems. These systems hence contribute to the evolvement of simulation and data processing codes, and they have the potential to influence the co-design of next-generation HPC systems by providing feedback to the system maintainers and operators.

Status at the beginning of the project (SoA)

The status of the simulation codes relevant for Work Package 3 (WP3) and Work Package 4 (WP4) prior to Task T2.2 is given below.

- The multi-physics simulation framework m-AIA (Multiphysics-Aerodynamic Institute Aachen) developed by (RWTH) is part of Task T3.1 for the simulation of turbulent boundary flows over an actuated boundary layer. This code has been tested in major Tier-0/1 HPCs, where it has shown promising scaling performance. The m-AIA code runs on Central Processing Units (CPUs), and an accelerated version that uses General-Purpose Graphics Processing Units (GPGPUs) is planned in Task T2.1. Both distributed (MPI) and shared memory (OpenMP) parallelization approaches (in hybrid) are available in the implementation of this code.
- The multi-physics simulation code Alya developed by BSC is part of Task T3.2 in WP3 for simulating the flow over individual wind turbines as well as over a whole wind park. The Alya code has shown exceptional scaling performance on Tier-0/1 HPCs. The code utilizes a hybrid MPI/OpenMP approach and can be co-executed using both CPUs and GPGPUs in heterogeneous architectures.
- The open-source software code Basilisk, which solves partial differential equations based on adaptive Cartesian computational domains, is used by CYI in Task T3.5 for the simulation of droplet behavior on various surfaces. The code performance and scalability have been analyzed in various HPC centers. This code uses MPI parallelization.

Below is the status of Machine Learning (ML) codes and tools relevant to Task T2.2.

- ML frameworks that are theoretically capable of scaling to exascale systems, such as the distributed data parallel frameworks, e.g., PyTorch-DDP, Horovod, Heat, and DeepSpeed are openly available to the community. These ML frameworks are only tested and documented on a limited range of hard-, and middlewares, commonly x86-based CPUs or NVIDIA GPGPUs with their specific driver packages.

Eray

- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.3

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

Draft

Version: 0.1

Author(s): A. Lintermann (FZJ)

Contributor(s):

Date: 17.05.2022

4.2.3 Task 2.3 - Benchmarking on disruptive technologies

With JUPSI, the first quantum annealer (QA) by D-Wave with more than 5000 qubits in Europe has been put into operation at the Forschungszentrum Juelich at the beginning of this year. It is part of the Jülich UNified Infrastructure for Quantum computing (JUNIQ), which provides German and European researchers with access and support on various quantum systems. A QA is well suited to solve certain optimization problems much faster than a regular computer, where in the RAISE project, several of these well-suited problems are explored. Additionally the performance of the QA as an accelerator to the classical supercomputers located in Jülich as part of the modular supercomputing approach is investigated.

Status at the beginning of the project (SoA)

In general, as quantum computing systems have only become available in the last few years, the field is still quite new. One focus of current research is the development of suitable benchmarks for quantum systems [13]. Results show that the JUPSI machine outperforms earlier systems [14]. While the JUPSI machine features a large number of qubits and can solve larger problems, other factors still limit the size of the actual problems that can be computed. Based on the current state of the QA, the combination of a QA as an accelerator with a classical supercomputer in a hybrid scheme is an interesting research direction, which has also not been explored in much detail yet.

Task objectives

In this task, sub-problems from the applications in the RAISE project that are suitable for solving on the JUPSI QA are identified. Initially, machine learning methods to solve these sub-problems on a classical supercomputer are developed and then adapted to run on the QA. Benchmarking the performance of these methods on both systems generates insights into the advantages and disadvantages of such quantum systems. At the same time a technical workflow for communication and data exchange between the JUPSI and JURECA-DC-GPU machines is developed, a key factor for enabling hybrid computation which is currently not available.

Strategy to reach the objectives including already accomplished work

The task is divided into three sub-tasks:

- T2.3.1: Quantum accelerated hyperparameter tuning: in a hybrid scheme, the JUPSI machine is coupled with the JURECA-DC-GPU partition of the supercomputer. Machine learning models with different values of hyperparameters are trained partially on the supercomputer for a few epochs. Then the learning curves of these models are transferred to the QA, where a Quantum Support Vector Regression (QSVR) is

Rakesh

- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.4

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

Draft

Version: 0.1
Author(s): A. Lintermann (FZJ)
Contributor(s):
Date: 17.05.2022

4.2.4 Task 2.4 - Software design of a unique AI framework

[TODO Task leader UOI]: General introduction to the task ✅
The overall idea of this task is to simplify the usage of AI tools and libraries in complex HPC environments by providing a guiding unique AI framework that addresses the requirements of particularly HPC researchers. The CoE RAISE use cases of WP3 and WP4 serve as co-design applications of such a framework, while other tasks in WP2 analyze the scalability of the framework components towards Exascale performance. Consequently, WP2 works closely together in this task with the WP3 and WP4 researchers to ensure that the developments in the use cases align with the generalized software layout of the unique AI framework. The task shares its framework components such as HPC job scripts, Jupyter Notebooks, or Python AI scripts with the larger AI and HPC community to encourage broader adoption of the framework beyond the CoE RAISE use case applications.

Status at the beginning of the project (SoA)

[TODO Task leader]: Provide the state-of-the-art, or the general status of the domain, if applicable ✅
Researchers of computing and data-driven applications that leverage HPC systems face the challenge of choosing from a wide variety of AI tools and libraries (e.g., TensorFlow, Horovod, pyTorch, Tarantella, RayTune, etc.) that need to work together in complex HPC environments jointly. The particular challenges include heterogeneous HPC hardware that needs to work together with many different versions of AI tools and libraries that usually require a detailed understanding of a wide variety of HPC modules to choose from in the right combination. Researchers at the JSC usually spend roughly 2-3 working days each month setting up HPC job scripts, Jupyter Notebooks and AI python codes with that right combination. In addition, the complexity of choosing the right combination represents a barrier for new HPC users that do not know the underlying technical details of HPC environments. Hence, no commonly agreed-on framework guides HPC researchers on what represents the right combination for each of the different HPC systems (e.g., JSC JUWELS, BSC MareNostrum) or EuroHPC JU systems (e.g., LUMI, VEGA, MELUXINA, etc.). The state-of-the-art in HPC already observes that the heterogeneity of HPC systems is constantly growing (e.g., library differences between accelerators from different emerging vendors) and that the complexity of HPC systems towards Exascale with emerging HPC systems like SUMMIT or JUPITER is increasing. That raises the demand for a unique AI framework design that guides HPC researchers and new users of HPC to simplify the usage of AI tools and libraries and ensure their scalability towards Exascale performance.

Morris, Matthias, Helmut

- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.4 - Flowcharts

No one has a flowchart except WP2: discuss if this is really needed then

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

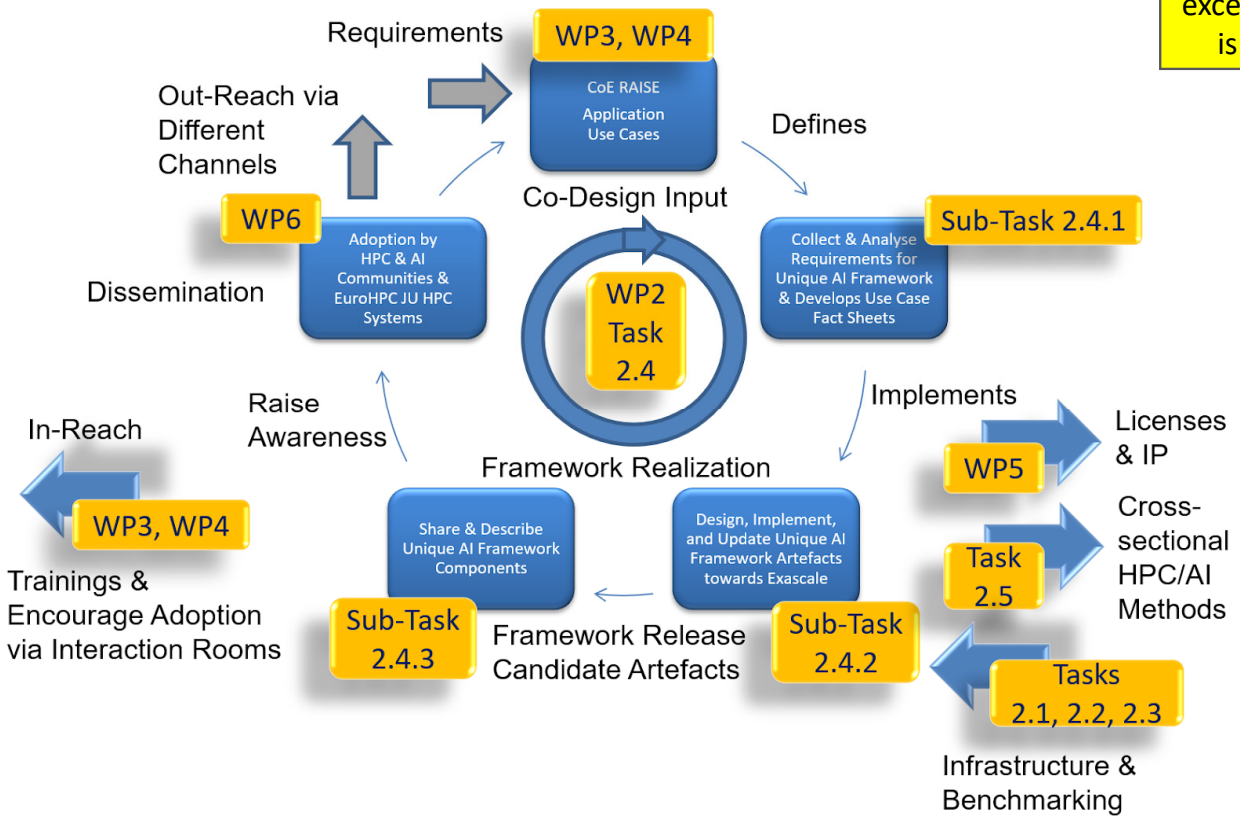
Draft

Version: 0.1

Author(s): A. Lintermann (FZJ)

Contributor(s):

Date: 17.05.2022



- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.5

RAISE Center of Excellence

RAISE Center of Excellence

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

Draft

Version: 0.1

Author(s): A. Lintermann (FZJ)

Contributor(s):

Date: 17.05.2022

Morris, Matthias, Helmut

4.2.5 Task 2.5 - Cross-Sectional AI Methods

[TODO Task leader UOI]: General introduction to the task ✅

The overall idea of this task is to complement the domain-specific application use case approaches in WP3 and WP4 with selected complementary machine and deep learning methods, including statistical methods where relevant. While Task 2.4 focuses on the software layout, Task 2.5 develops innovative AI methods based on that software layout on heterogeneous and modular HPC architectures, thus capable of scaling towards Exascale. Consequently, WP2 works closely together in this task with the WP3 and WP4 researchers to ensure that the AI developments in the use cases adopt the generalized software layout of the unique AI framework of Task 2.4. In addition, the task shares its cross-sectional AI methods via Jupyter Notebooks or Python AI scripts with the larger AI and HPC community to encourage broader adoption of the CoE RAISE AI framework beyond the CoE RAISE use case applications.

Status at the beginning of the project (SoA)

[TODO Task leader]: Provide the state-of-the-art, or the general status of the domain, if applicable ✅

Researchers of computing and data-driven applications that leverage HPC systems face the challenge of choosing from a wide variety of general AI (i.e., machine and deep learning methods) that must be adapted and prepared to work with domain-specific approaches and datasets. While many general tutorials, guidelines, and examples are available for selected simple problems (e.g., character recognition using the famous MNIST dataset), sophisticated AI materials with concrete scripts and precise approaches in various domains are still relatively rare. That is particularly the case for those domains driven by simulation sciences (e.g., computational fluid dynamics, aerodynamics, etc.) that implement numerical methods based on known physical laws. Coupling and combining those numerical methods with cutting-edge AI models is still a relatively new topic where many approaches are still state-of-the-art research topics. One of the challenges in this research area is having access to the knowledge of complex techniques in the scientific or engineering domain and the know-how to apply innovative machine and deep learning in context. Hence, novel cross-sectional AI methods that have proven to scale on HPC systems towards Exascale are required to advance the use of AI technologies in the overall use cases of WP3 and WP4.

- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap T2.5 - Flowcharts

No one has a flowchart except WP2: discuss if this is really needed then

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

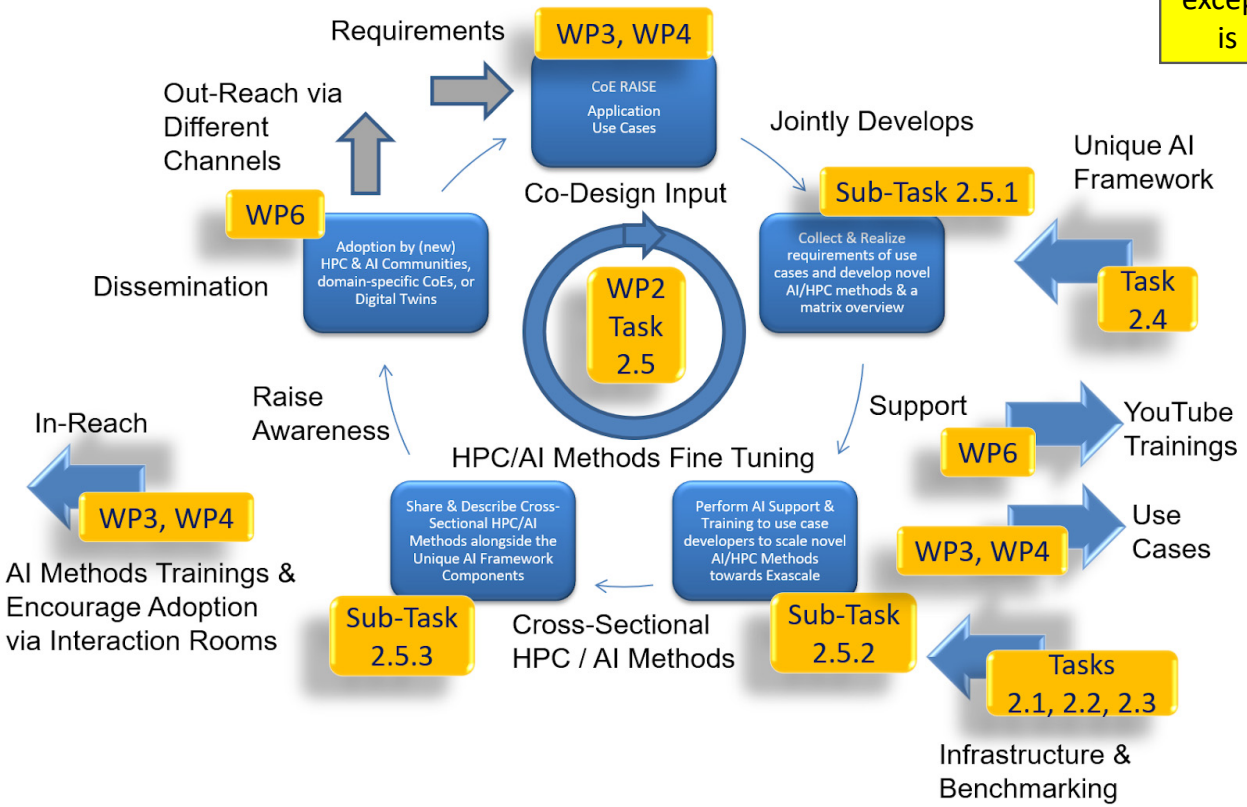
Draft

Version: 0.1

Author(s): A. Lintermann (FZJ)

Contributor(s):

Date: 17.05.2022



- TBD (Task Leaders): https://docs.google.com/document/d/1eJ5sV6Ubr_prHt80PJAmnB4fWEK6Kyvh/edit

September Review Preparation: Roadmap GANTT Example

H2020-INFRAEDI-2018-2020

CoE RAISE

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

Grant Agreement Number: 951733

Roadmap

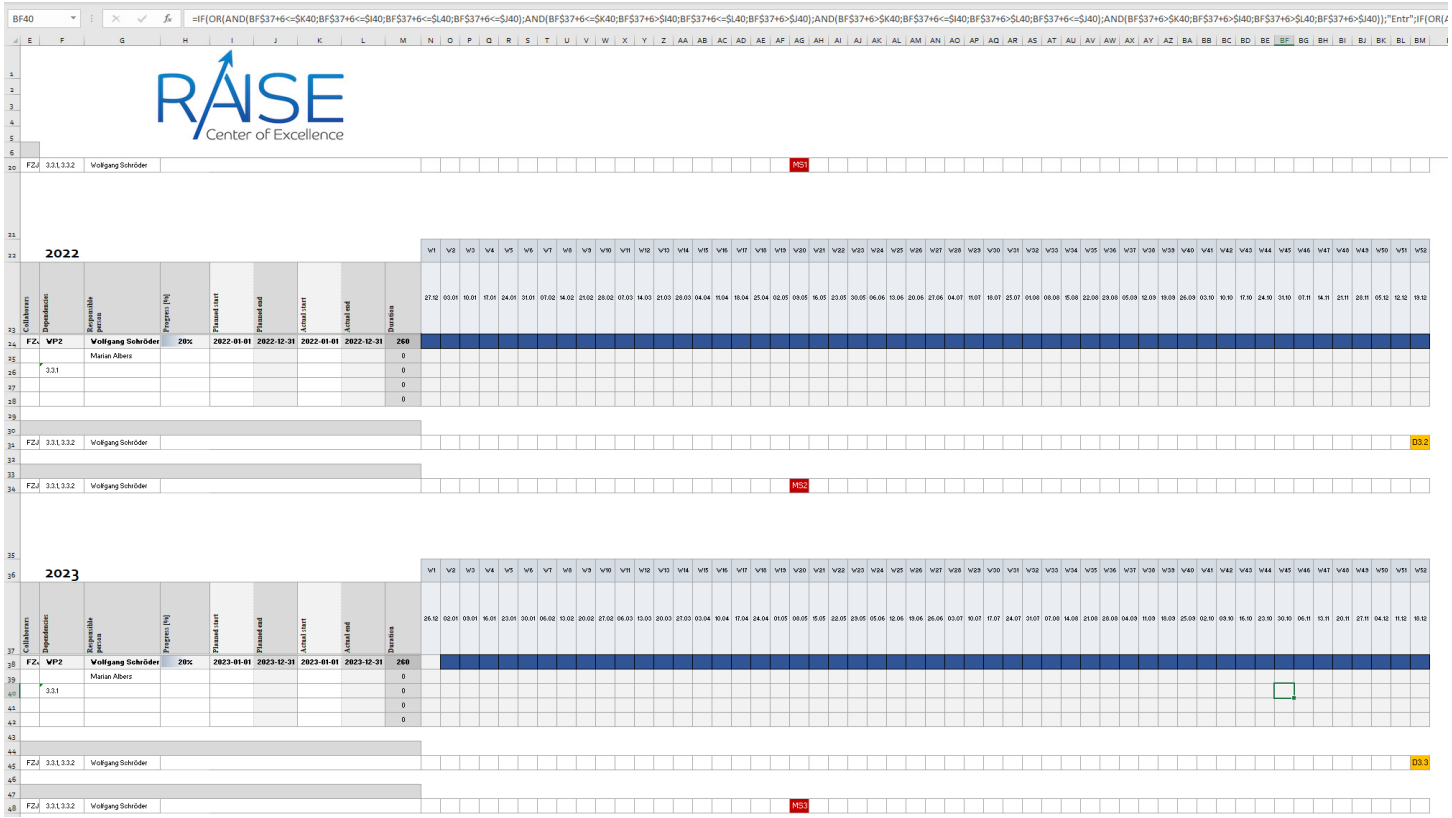
Draft

Version: 0.1

Author(s): A. Lintermann (FZJ)

Contributor(s):

Date: 17.05.2022



- TBD (Task Leaders): https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/d3937966/CoE_RAISE_Gantt_Roadmap_TX.Y.xlsx

September Review Preparation: Innovation Radar Status

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR COMMUNICATIONS NETWORKS, CONTENT AND TECHNOLOGY

Digital Excellence and Science Infrastructure
High Performance Computing and Quantum Technology

Luxembourg 09-06-2022
CNECT.C.2.001/**AE**

Dr.-Ing. Dipl.-Inform. Andreas Lintermann
Juelich Supercomputing Centre
Institute for Advanced Simulation
Forschungszentrum Juelich GmbH
52425 Juelich, Germany

Subject: Horizon 2020 Framework Programme
Project: 951733 — RAISE
Innovation Radar Information Letter

Dear Mr. Lintermann,

I am writing to inform you that, in the context of the ongoing project review, your project has been selected to also participate in the EU Innovation Radar.

The Innovation Radar is an initiative to collect structured data about the innovation profile of EU funded projects and their outputs. The goal of this is to identify high-potential innovators and innovations; their specific 'go to market' needs; and, provide targeted recommendations. The aim is ultimately to encourage and support innovators in getting their innovations 'out of the lab' and into (or at least closer to) the market. You can find more information on this initiative in this link. The Innovation Radar is a complementary tool to the Horizon Results Platform.

The project review panel will therefore also look at the innovation dimension of your project. It would be very valuable for the work of the experts if the project consortium provides information on the innovations being developed by your project. We would be very grateful if you could fill out in the attached questionnaire the following:

– a meaningful **title** for each project innovation (up to 200 characters long)

– a meaningful **description** for each project innovation (up to 1000 characters long).

Please send it back through the Portal Communication Centre (please do NOT use the option reply Expert's CoI since this is used for other purposes) one week before the review meeting, using the templates that you have received together with this letter.

I would be grateful if you could inform the other members of your consortium (if any) of this letter.

For any questions, please contact us via your Funding & Tenders Portal account.

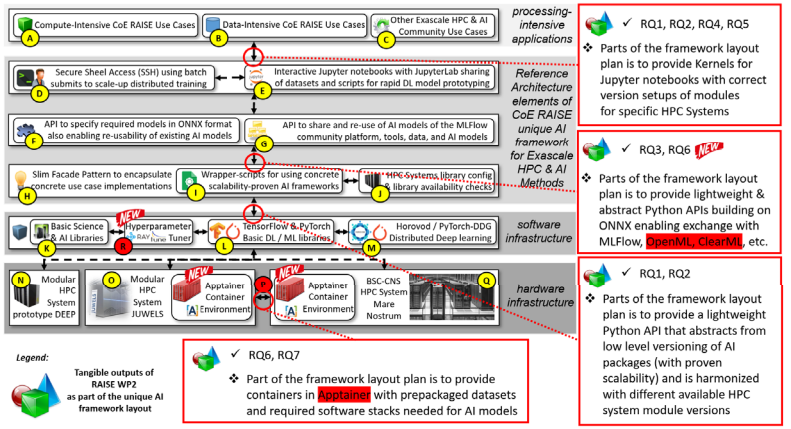
Commission européenne/Europese Commissie, 1049 Bruxelles/Brussel, BELGIQUE/BELGIË - Tel. +32 22991111
Athanasia-Charalampia.EVANGELINOU@ec.europa.eu

INNOVATION RADAR QUESTIONNAIRE UOI BLUEPRINT

1. AI Framework Components for HPC (41/500 Characters)
2. RAISE aims to overcome the challenges of working with heterogeneous HPC hardware that needs to work together with many different versions of AI tools that require a detailed understanding of choosing the suitable HPC modules. Hence, UOI and FZJ researchers generally spend roughly 2-3 working days each month setting up their environments. Therefore, RAISE provides AI framework components for HPC systems suitable for Exascale to help researchers in using AI technologies at a scale significantly. (498/500 Characters)
3. This innovation is under development
4. Significantly improved process
5. Innovative but could be difficult to convert customers
6. Only deployed as new to the organization/company (new internal processes implemented, etc.)
7. (empty due to #6)
8. Multiple owners
9. UOI, FZJ, RTU
10. Expanding to more markets (e.g., Cloud providers?)
11. Current customers
12. Emerging: There is a growing demand and few offerings are available
13. (skip due to #12)
14. Yes
15. Several major players with strong competencies, infrastructure and offerings (e.g., clouds)
16. Between 1 and 3 years
17. No
18. As:
    a. Secure, clean and efficient energy
    b. Climate action, environment, resource efficiency and raw materials
19. As:
    a. SDG 7 – Affordable and Clean Energy
    b. SDG 9 – Industry, Innovation, and Infrastructure
    c. SDG 12 – Responsible Consumption and Production (i.e., using responsible HPC cycles)
    d. SDG 13 – Climate Action
20. Mitigation potential

General Discussion:

a. Meeting my expectations
b. Yes
    i. EuroHPC JU HPC System sites and their end-users
c. Potential procurer of innovation (Public sector)
d. As:

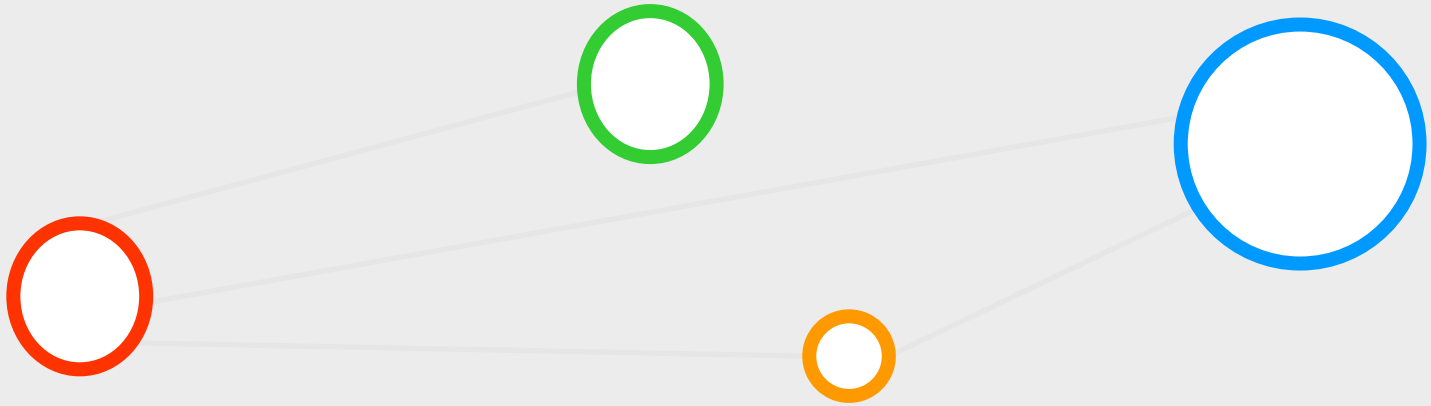


i. Are there IPR issues within the consortium that could compromise the ability of the organisation(s) to exploit new products/solutions/services, internally or in the market place? Yes!
ii. Which are the external bottlenecks that compromise the ability of project partners to exploit new products, solutions or services, internally or in the market place? Several components are created by a business partner (e.g., coupler from ATOS in their own commercial AI4SIM library)
iii. How would you rate the level of commitment of relevant organisation(s) to exploit the innovation? High

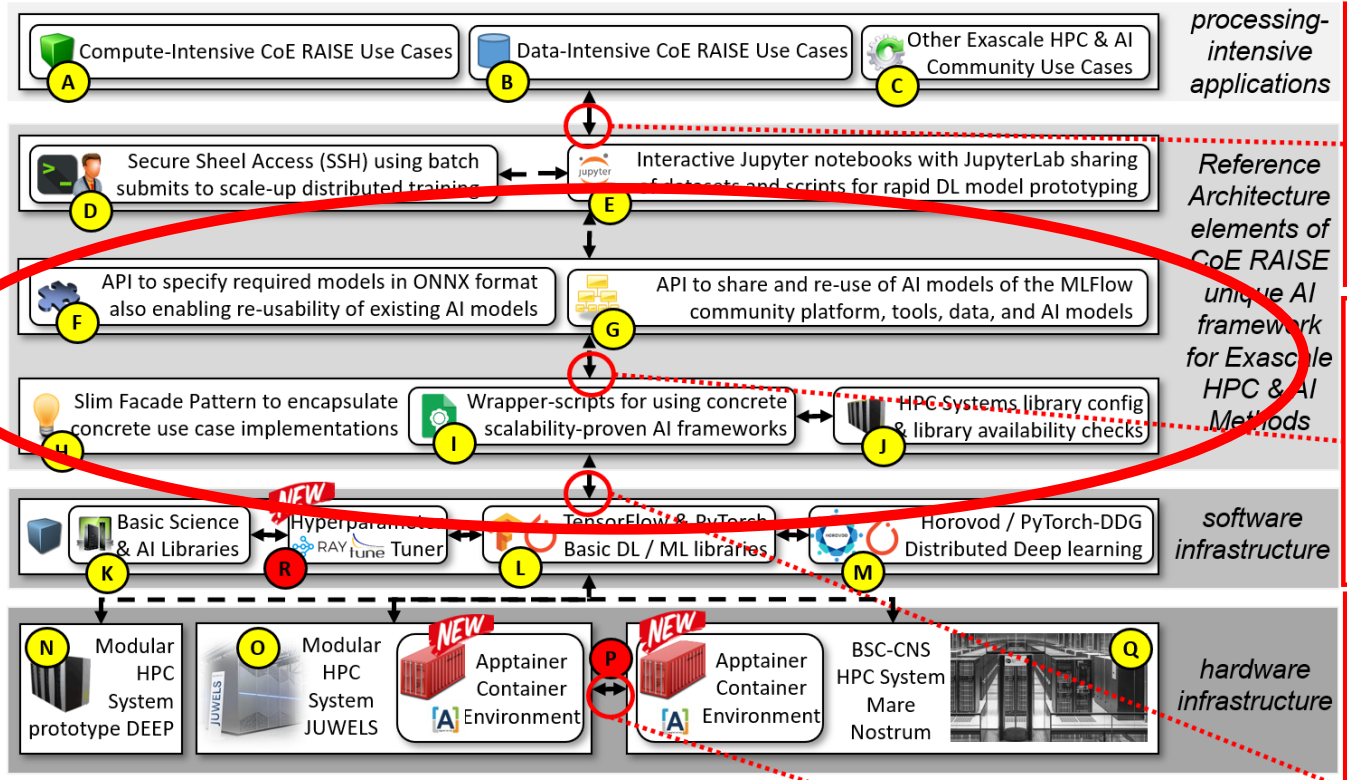
e. FZJ (and Exascale machine JUPITER coming up in the future)
f. As:
    i. Increase research budget
    ii. Easier access to available budget (also for non-EU countries)
g. Yes
h. (no one)

- TBD (Task Leaders): https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/3935782

Agenda Item (4) – Realization of AI Framework



Realization of SW Framework – HPC Systems Support



✓ 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 NEW

❖ Parts of the framework layout plan is to provide lightweight & abstract Python APIs building on ONNX enabling exchange with MLFlow, OpenML, ClearML, etc.

✓ RQ1, RQ2

❖ 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

❖ Part of the framework layout plan is to provide containers in Apptainer with prepackaged datasets and required software stacks needed for AI models

Next version might include coupler (Atos)?

Continously Updating!

Legend:

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

Realization of SW Framework – Collect Artefacts (Johannes)

- nvidia lib problem is not anymore a problem from Stages/2022

Rocco:

2022-03-04

- JUWELS Booster usage with Horovod: jobscript_juwels_tf2module_hvd_booster.sh

Reza:

2022-04-20

- GRU usage on model for turbulence flow using Jupyter notebook (on 1 node): velocity_time_series_gru.ipynb

Eric:

2022-04-22 (IR Session)

- Original Repository: https://github.com/jpata/particleflow
- usually own branches of developers and then merge to main: https://github.com/erwulff/particleflow
- MLPF has already its only repository
- script with raytune in that repository too
- model, etc.

- WP2 Wiki: https://gitlab.jsc.fz-juelich.de/riedel1/raise-wp2/-/wikis/Software-Framework-Co-Design

Realization of SW Framework – ONNX (Kristofer)

ONNX Research

Kristófer Darri Finnsson

June 2022

ONNX

ONNX is an open format built to represent machine learning models. ONNX defines a common set of operators - the building blocks of machine learning and deep learning models - and a common file format to enable AI developers to use models with a variety of frameworks, tools, runtimes, and compilers.

ONNX file format:

- Model
  - Version info
  - Metadata
  - Acyclic computation dataflow graph
- Graph
  - Inputs and outputs
  - List of Computation nodes
  - Graph name
- Computation node
  - Zero or more inputs of defined types
  - one or more outputs of defined types
  - Operator
  - Operator parameters

Tensorflow backend for ONNX

TensorFlow Backend for ONNX makes it possible to use ONNX models as input for TensorFlow. The ONNX model is first converted to a TensorFlow model and then delegated for execution on TensorFlow to produce the output. It already has the ability to use ONNX models by using onnx backend from tensorflow, see here:
github.com/onnx/tutorials/blob/master/tutorials/OnnxTensorflowImport.ipynb

ONNX to Pytorch

They also seem to have something to get torch models from onnx files here. But seems to not be as much as tensorflow: https://github.com/ENOT-AutoDL/onnx2torch

Triton Inference Server Backend

A Triton backend is the implementation that executes a model. A backend can be a wrapper around a deep-learning framework, like PyTorch, TensorFlow, TensorRT or ONNX Runtime. Or a backend can be custom C/C++ logic performing any operation (for example, image pre-processing).
https://github.com/triton-inference-server/backend

This one seems to be a backend for most of the frameworks we have been thinking about using.

Implementing a ONNX backend

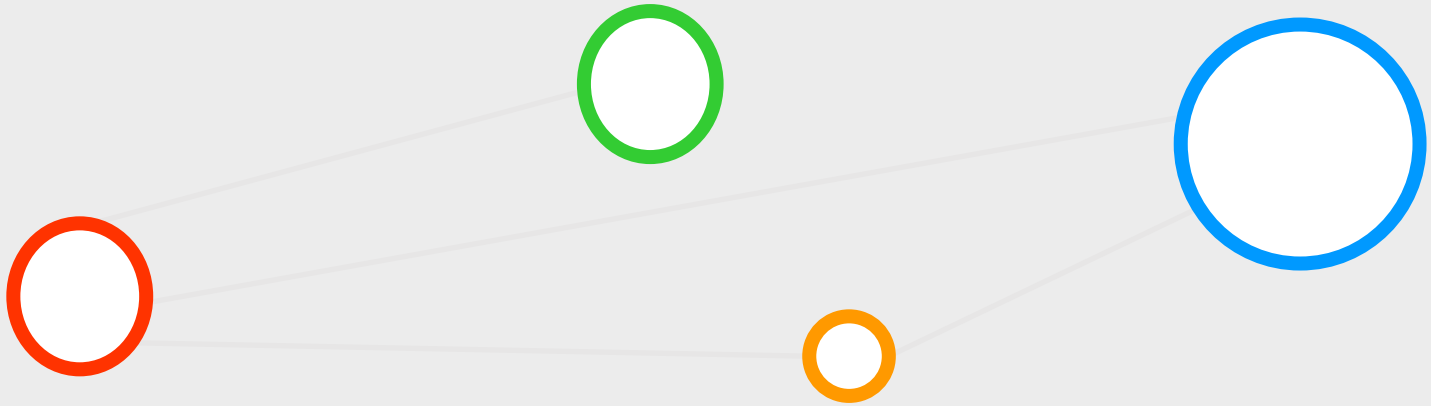
An ONNX backend is a library that can run ONNX models. Since many deep learning frameworks already exist, you likely won't need to create everything from scratch. Rather, you'll likely create a converter that converts ONNX models to the corresponding framework specific representation and then delegate the

- Research PDF available for WP2 and RAISE members (contact Katrin)

Realization of SW Framework – HPC Systems (Eray)

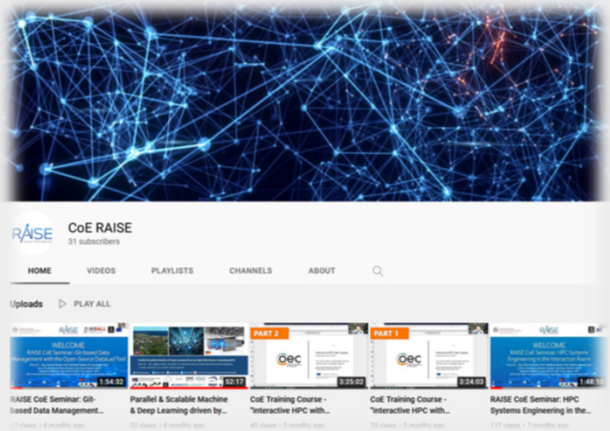
- Eray (adding maybe automated testing regularly for updates, training in August, etc.): important is outreach and uptake in the different use cases and other new community adopters and EuroHPC JU systems

Agenda Item (5) – Status WP2 Training Plans



WP2 Monthly Trainings – Review & Plan

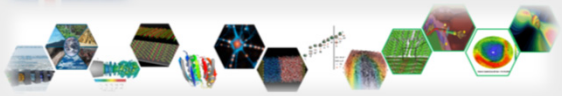
- Monthly WP2 Trainings
  - Co-organized with Icelandic National Competence Center (NCC) funded by the EuroCC project: http://ihpc.is
  - Performed since Quarter 2 of the project (April 2021)
  - Selected dates via agreement of availability of speakers
  - Used as major AI/HPC methods information/training for WP3/WP4
  - Contributed to outreach via YouTube Channel recordings: https://www.youtube.com/channel/UCAdIZ-v6cWwGdapwYxdN7dg
  - TBD(Katrín): Schedule the YouTube Training series with speakers



Plan for next months

- **Carry on with monthly WP2 trainings in the same style, but schedule on 3-4 month horizons**
- **Repeat certain trainings with advanced content and updates of activities**
- **Work better together with WP6 on releasing seminars on YouTube channel more regularly**
- **Collect slides of speakers and make them available on BSCW and/or on the RAISE Web Page**

IHPC National Competence Center (NCC) for HPC & AI in Iceland



WP2 Monthly Trainings – Review & Plan

WP2 Monthly Trainings – Review & Plan

December:
No Seminar

January:
RTU CUDA Seminar
(THANKS!)

February:
none

2022-03-31 14:00 GMT
Graph Neural Network
Morris: Intro GNNs (30 min)
Leo from Atos (45 min)
Eric from Cern (20 min)

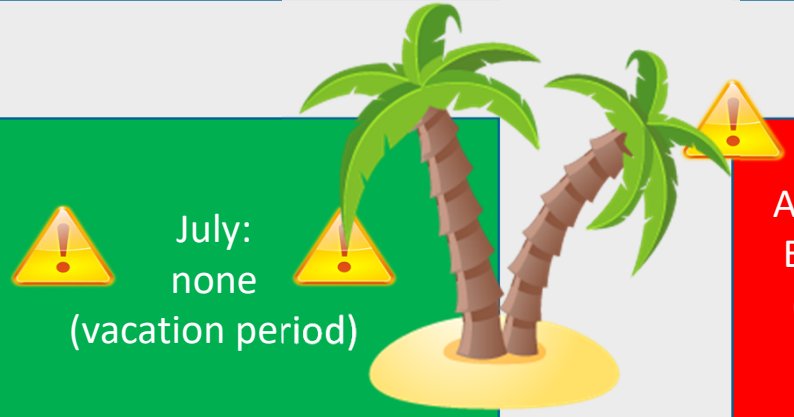
TBD (all): Please suggest further training & teaching seminars for YouTube channel on our WP2 mailing list to plan better ahead

WP2 Monthly Trainings – Review & Plan

April:
Quantum Annealing
Maybe Gabriele Examles from SVMs, Amer SVR

May:
Using OpenML for sharing datasets, algorithms, and experiments

9th of June:
Morris: GPUs in general
Arnis & Cuda @ RTU



July:
none
(vacation period)

August: Morris Framework + Eray Tooling with modules? (might be good before the review) + Marcel?

ATOS: affects of change in persons?

September:
EOSC – NI4OS-Europe or TREX Project (in scheduling) Request Project Partners? (continous integration ATOS)???
→ Katrin: check and schedule

TBD (all): Please suggest further training & teaching seminars for YouTube channel on our WP2 mailing list to plan better ahead

WP2 Monthly Trainings – Review & Plan

October:
Project Partners?

November:
Project Partners?

December:
Project Partners?

Wait for the feedback from Review:
Otherwise: idea would be to have also trainings on the adoption of sw framework parts again & AI/HPC Methods

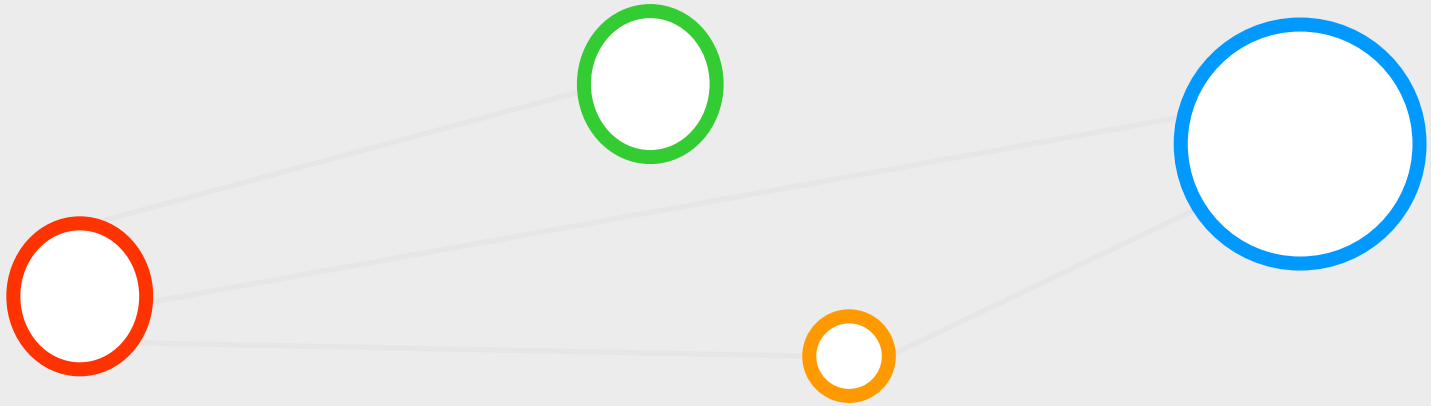
January:
Project Partners?

February:
Project Partners?

March
Project Partners?

TBD (all): Please suggest further training & teaching seminars for YouTube channel on our WP2 mailing list to plan better ahead

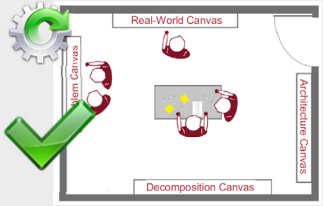
Agenda Item (6) – Compelling Scoreboard Review



Compelling Scoreboard Review – Use Case Progress

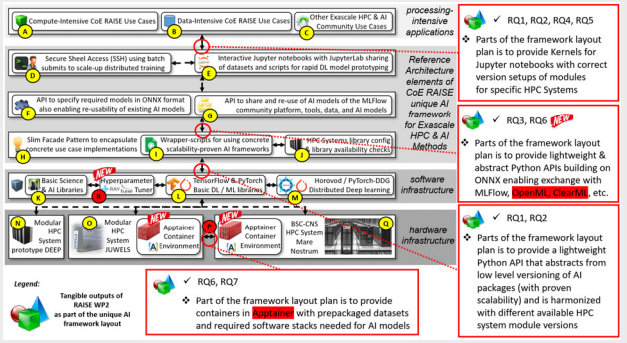
RAISE Center of Excellence

- Performing joint Interaction Room sessions: to identify novel AI methods and build unique AI framework parts
- Realization of SW framework design started → initial collection in WP2 Wiki page RAISE (Jupyter notebooks, etc.)
- Identified lots of problems → SW framework concept required
- Collection of software artefacts started for the realization of the framework design bottom-up by use case elements

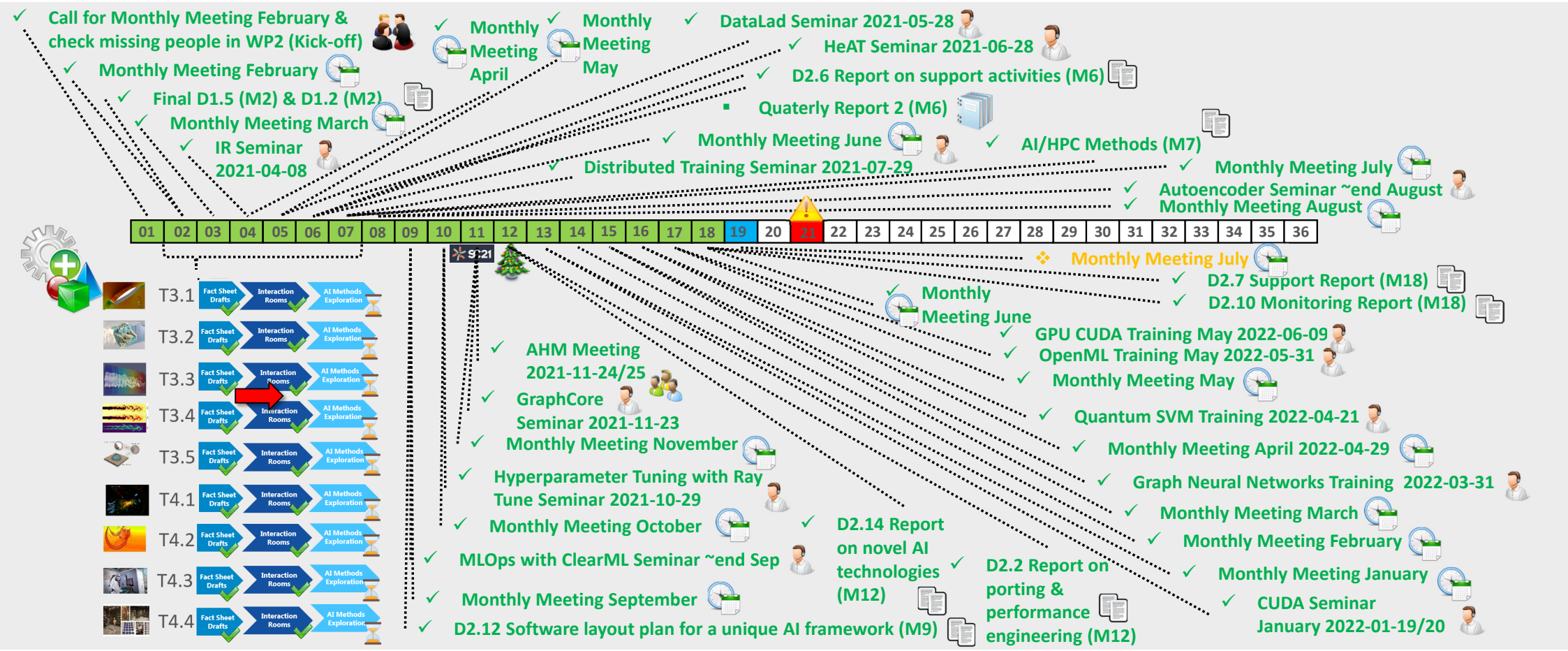


Lessons Learned from IR Rooms with use cases: Agreement & common components between use cases realistic → Framework Co-Design ok

Use Case	AE	PINN	ANNs		CNN		NO	GNN		RNN		GAN	TF				SVM	RF
Details	*CAE*		*ANN*	*RBF-ANN*	*U-Net*	*RES NET*	*FNO*	*MLPF*	*GAT*	*LSTM*	*GRU*	*WGAN*	*MVIT*	*ViViT*	*Swin*			
AI for turbulent boundary layers	X	X	X									X						
AI for wind farm layout optimization				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	X					X			X								
Event reconstruction and classification at the CERN HL-LHC use case								X										
Seismic imaging with remote sensing for energy applications	X	X				X	X			X	X					X	X	X
Detect-free metal additive manufacturing	X		X									X	X	X	X			
Sound Engineering	X		X															



Compelling Scoreboard Review & Next Steps



Agenda Item (6) – Next Steps & Follow-Through

- 08.09 - 09.09.2022 (Review Preparations start)
  - One day in-person rehearsal
  - One day full event in Juelich
  - WP leaders, not really task leaders
  - Every Organization takes part
  - Instructions & schedule given by PMT
- Clarify Flowcharts with PMT
  - Affects some Tasks in WP2
  - Affects almost all other WPs (no flowcharts yet)
  - Either more template-oriented (e.g. Rose and my statement) or skip (discussions with WP4)
- Other items?
  - Task 2.1
    - LUMI (get access via UICE), Puhuri access
    - Mare Nostrum (machine end of the year)

drive. enable. innovate.

RAISE Center of Excellence

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

Follow us: