Distributed Computing in Fusion Research

EOSC Marketplace ask me anything webinar

Andrew Lahiff, United Kingdom Atomic Energy Authority

EOSC Future with C-SCALE DICE Data Infrastructure Capacity for EOSC EGI-ACE OpenAIRE NEXUS Reliance

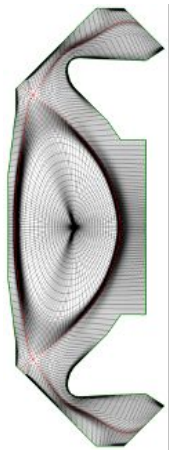
The EOSC Future, C-SCALE, DICE, EGI-ACE, OpenAIRE-Nexus and Reliance projects are funded by the European Union Horizon Programme calls INFRAEOSC-03-2020 and INFRAEOSC-07-2020.

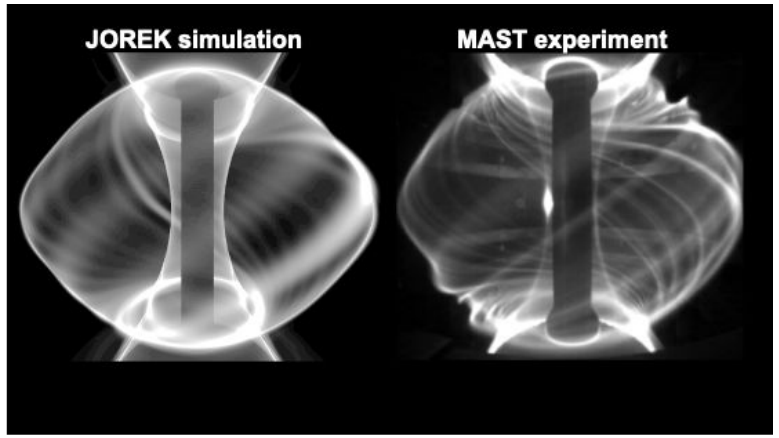
Introduction

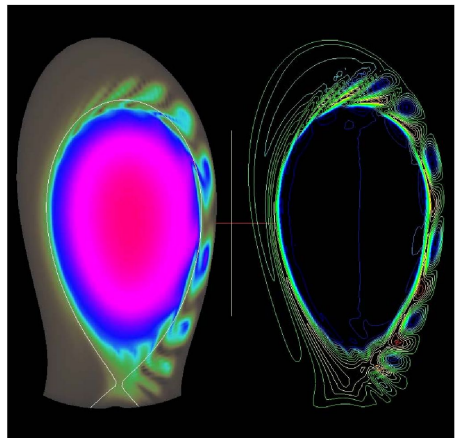
- Nuclear fusion
  - Energy source of the Sun & stars
  - Light nuclei fusing to form a heavier nucleus releases energy
- Recreating fusion on Earth
  - Typically heat deuterium & tritium to over 100 million degrees
  - Eventual aim is to use the energy produced to generate electricity for the grid
- Computing in Fusion
  - Plasma modelling, materials research, engineering, data processing, uncertainty quantification, rendering, machine learning, …
- EGI-ACE: Integrating distributed clouds, HPC and storage

Fusion use case in EGI-ACE

- Example Fusion application: JOREK
  - Simulation of MHD instabilities at the edge of Tokamak plasma
  - Dynamics highly dependent on edge plasma pressure
- Resource requirements
  - High-fidelity: large numbers of CPUs with low-latency interconnects
  - On a single node: limited fidelity, limited physics

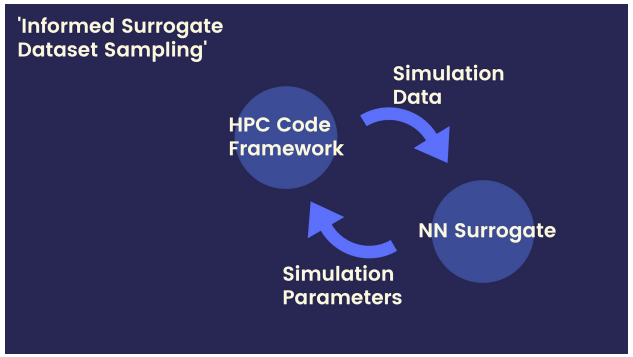






Fusion use case in EGI-ACE (cont'd)

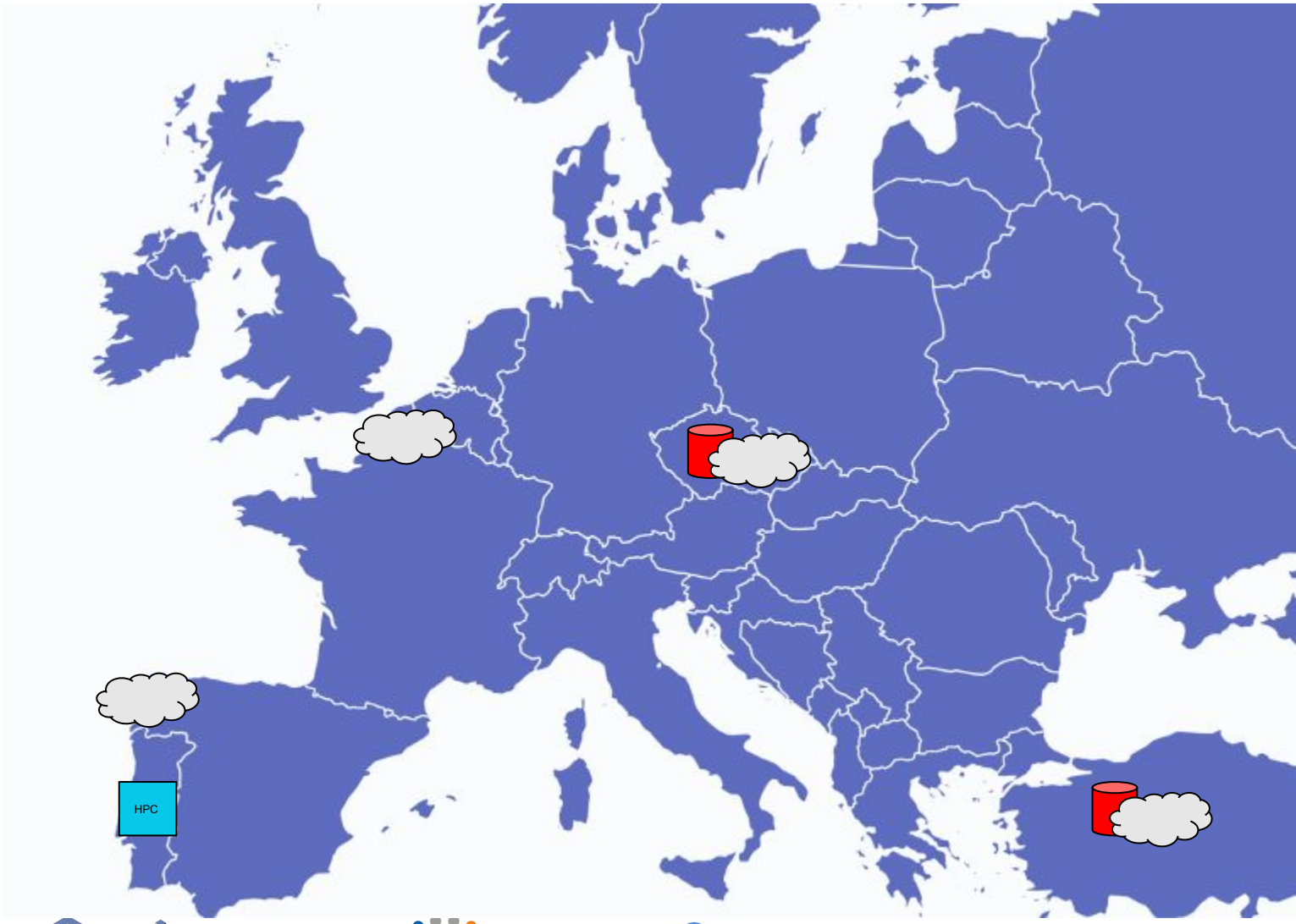
- Building NN surrogate models in an efficient & effective manner
  - Traditional approach involves arbitrary scans across a range of a parameters
  - May be unaware of more complex behaviour & nuances
- An alternative approach to overcome these difficulties
  - Physics-informed neural network
  - Cyclic workflow
    - Gaps in knowledge identified
    - Trigger the simulation code to generate additional data points



Building NN-based surrogate models

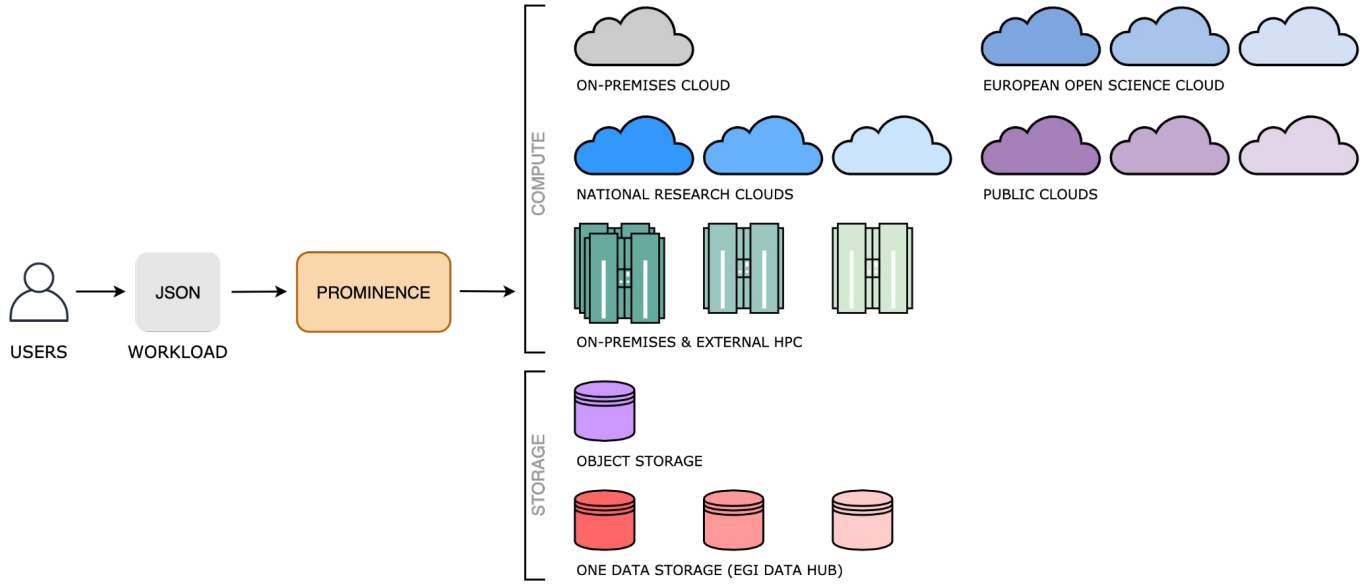
Fusion resources in EGI-ACE

- Clouds
  - TUBITAK
  - University of Lille
  - CESGA
  - CESNET
- HPC
  - LIP
- Storage (OneData)
  - TUBITAK
  - CESNET



PROMINENCE: Unifying access to resources

- Platform which abstracts away multiple clouds & HPC clusters
  - Appears like a single batch-system to users
  - Users don't have to worry about provisioning clusters or infrastructure
  - Jobs are automatically directed to appropriate resources
  - Supports opportunistic usage of idle resources: improves efficiency



EOSC pilot
The European Open Science Cloud for Research Pilot Project

EOSC-hub

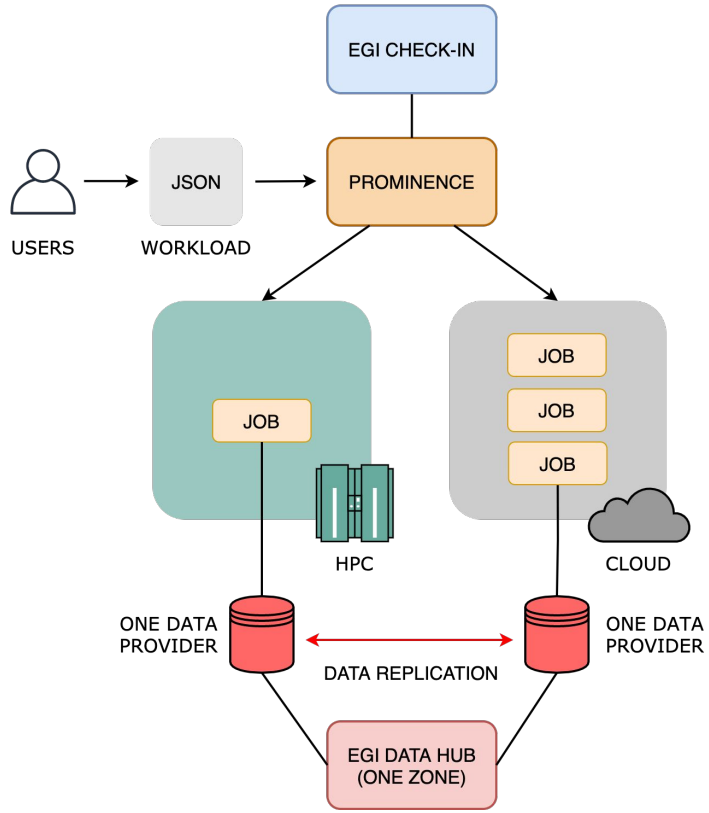
EOSC Future with C-SCALE DICE Data Infrastructure Capacity for EOSC EGI-ACE OpenAIRE NEXUS Reliance

PROMINENCE: In use

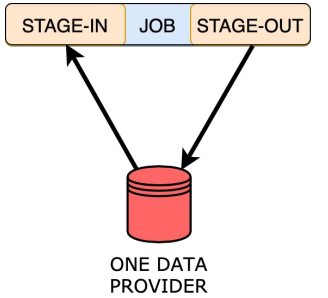
- Example running a multi-node MPI job

```
{
  "name": "lammps-example01",
  "resources": {
    "cpus": 32,
    "disk": 10,
    "memory": 32,
    "nodes": 8
  },
  "artifacts": [
    {
      "mountpoint": "lammps-example01-01:/data",
      "url": "lammps-example01-01.tgz"
    }
  ],
  "tasks": [
    {
      "cmd": "lmp_intel_cpu_intelmpi -pk intel 0 -sf intel -in sample.in",
      "image": "alahiff/lammps-intel-avx512-2018:latest",
      "runtime": "udocker",
      "workdir": "/data",
      "type": "intelmpi"
    }
  ]
}
```

Example JSON description of an Intel MPI job



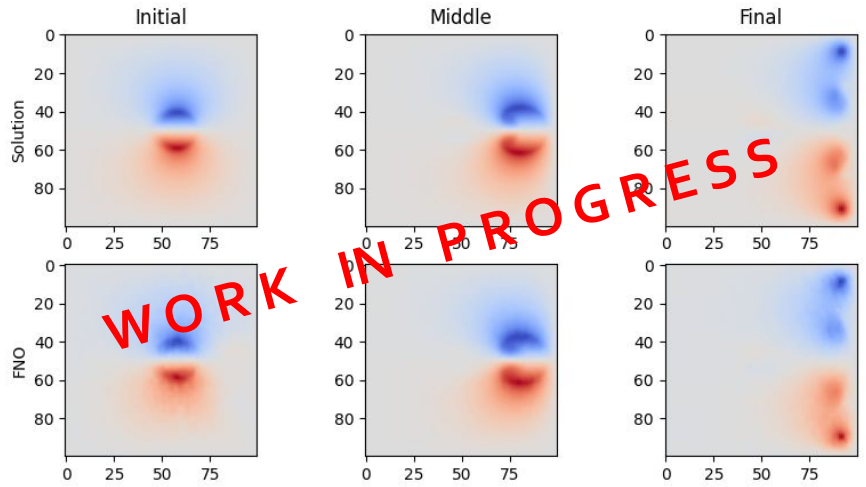
Data access on clouds & HPC:

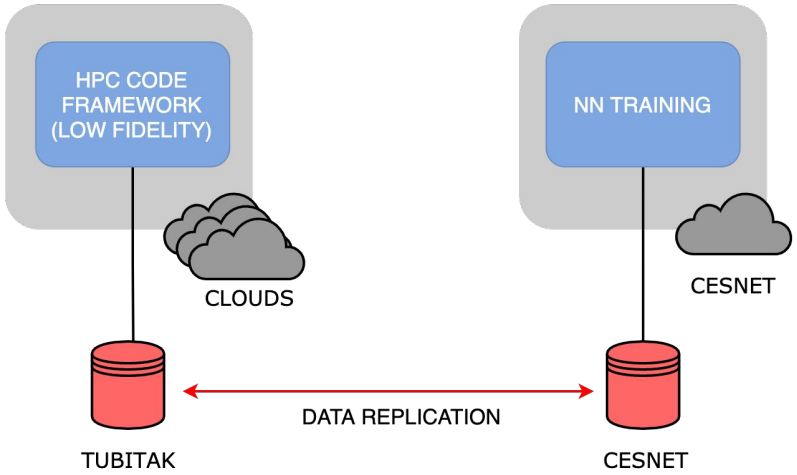


On clouds also POSIX-like access (typically not possible on HPC)

Example results

- Using a neural network (Fourier Neural Operator) to model the density, vorticity & potential from MHD simulations

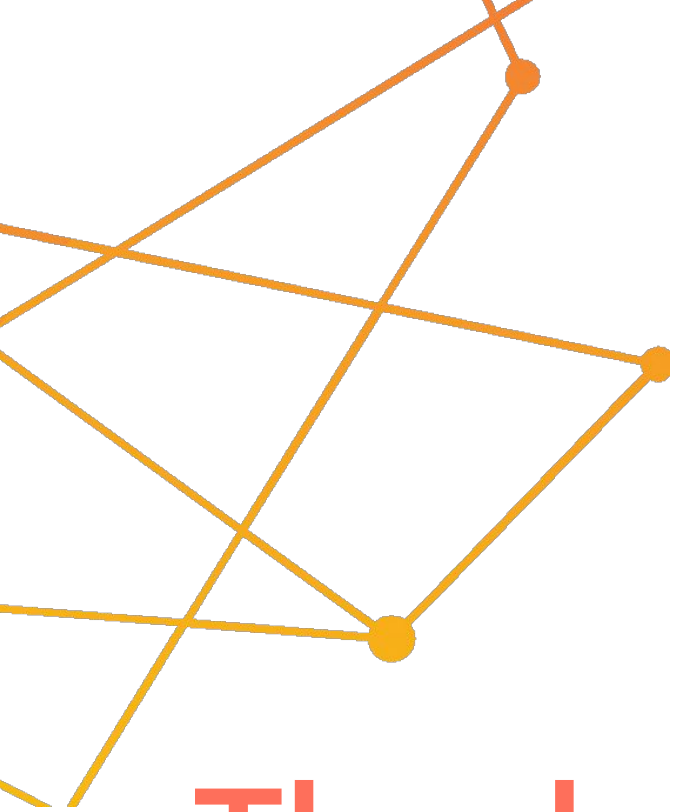




- Next steps
  - Initial results (above) used clouds only
  - Next will use an HPC cluster via PROMINENCE

Challenges

- Creating containers which perform optimally on multiple HPC clusters
- POSIX-like access to OneData not available on HPC clusters
  - Users aren't typically able to use FUSE
- HPC clusters with restricted outgoing internet access



Thank you

EOSC Future Together with C-SCALE DICE Data Infrastructure Capacity for EOSC EGI-ACE OpenAIRE NEXUS Reliance