

# CASTIEL 2

Castiel-2 and the NCCs project, possibilities for SMEs and research

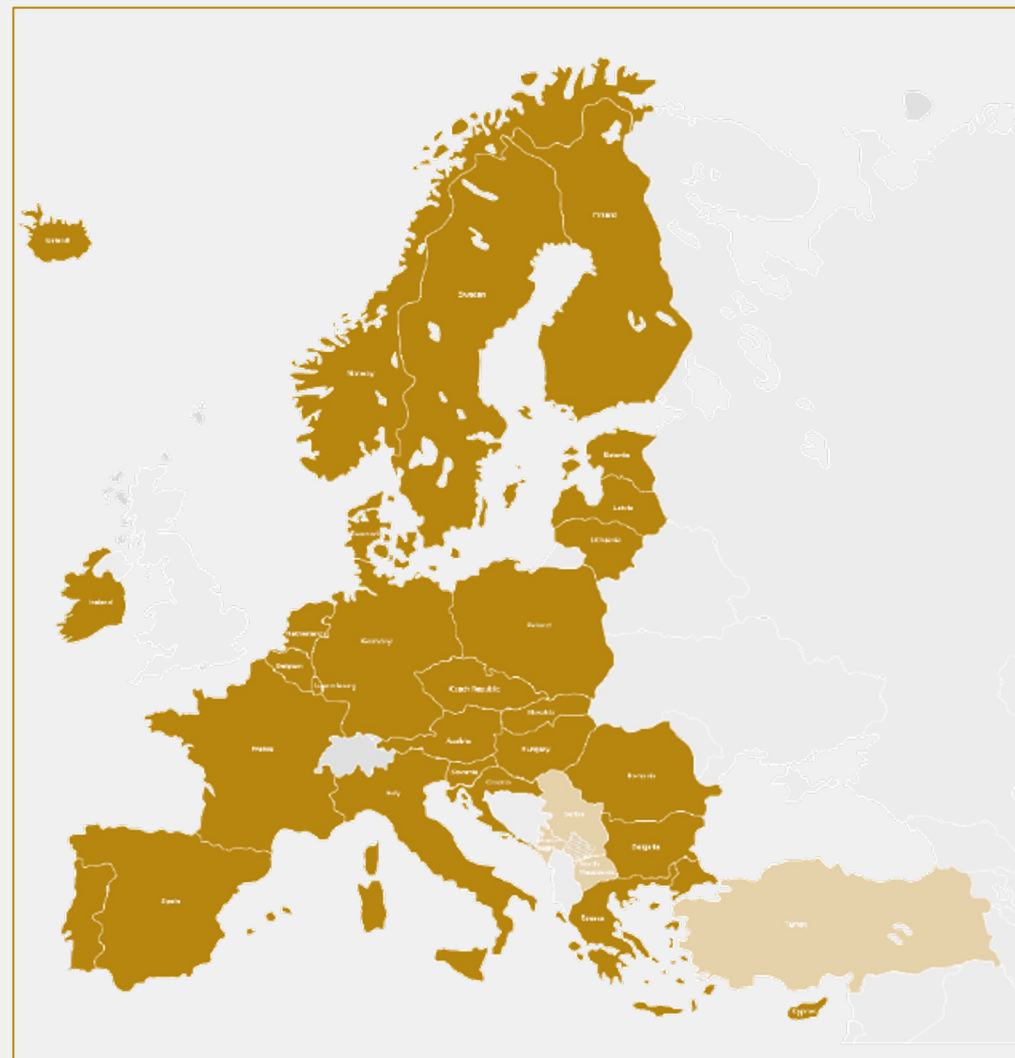
**November 22nd, 2023**

*Marie-Françoise Gerard, Teratec*



# Castiel-2 and the NCCs project, possibilities for SMEs and research

*HPC/HPDA/AI and Industry  
collaboration in Europe*



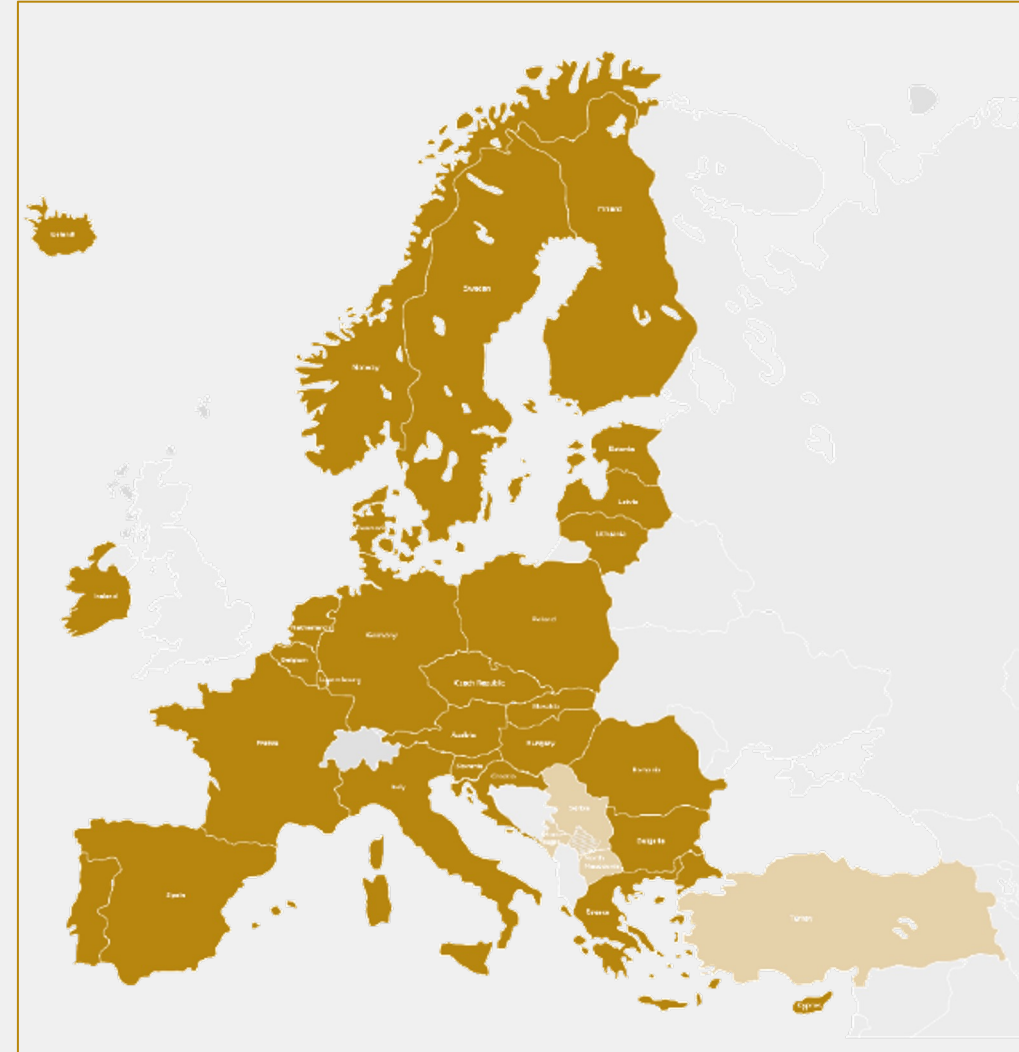
# What are EuroCC (2) and CASTIEL(2) ?

- Research and Innovation Action (RIA)
- **Phase 1:** 01/09/2020  
24 months duration

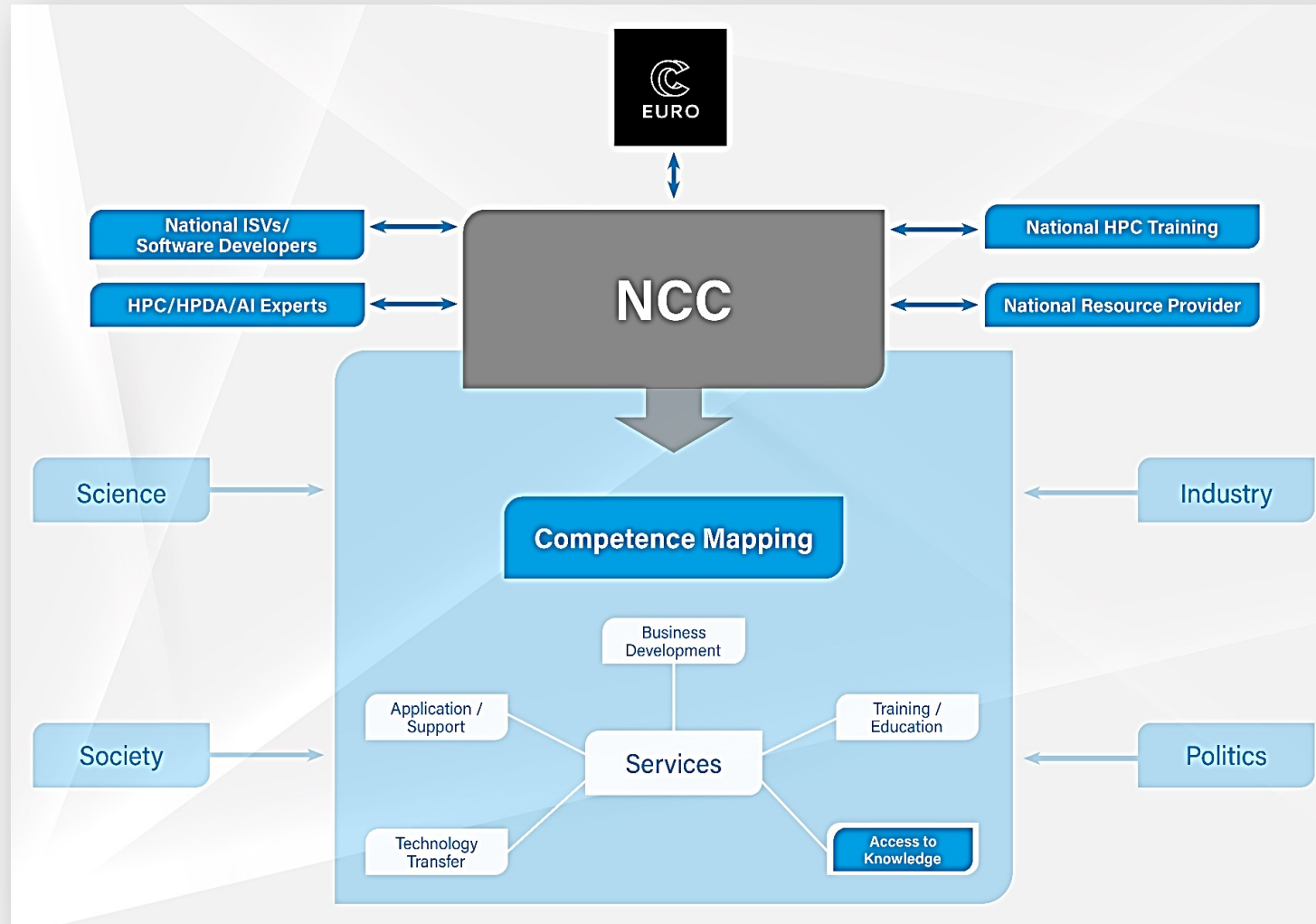
Now phase 2!  
2023 – 3 years

 EuroCC

 @EuroCC\_project



# National Competence Centres in the framework of EuroHPC





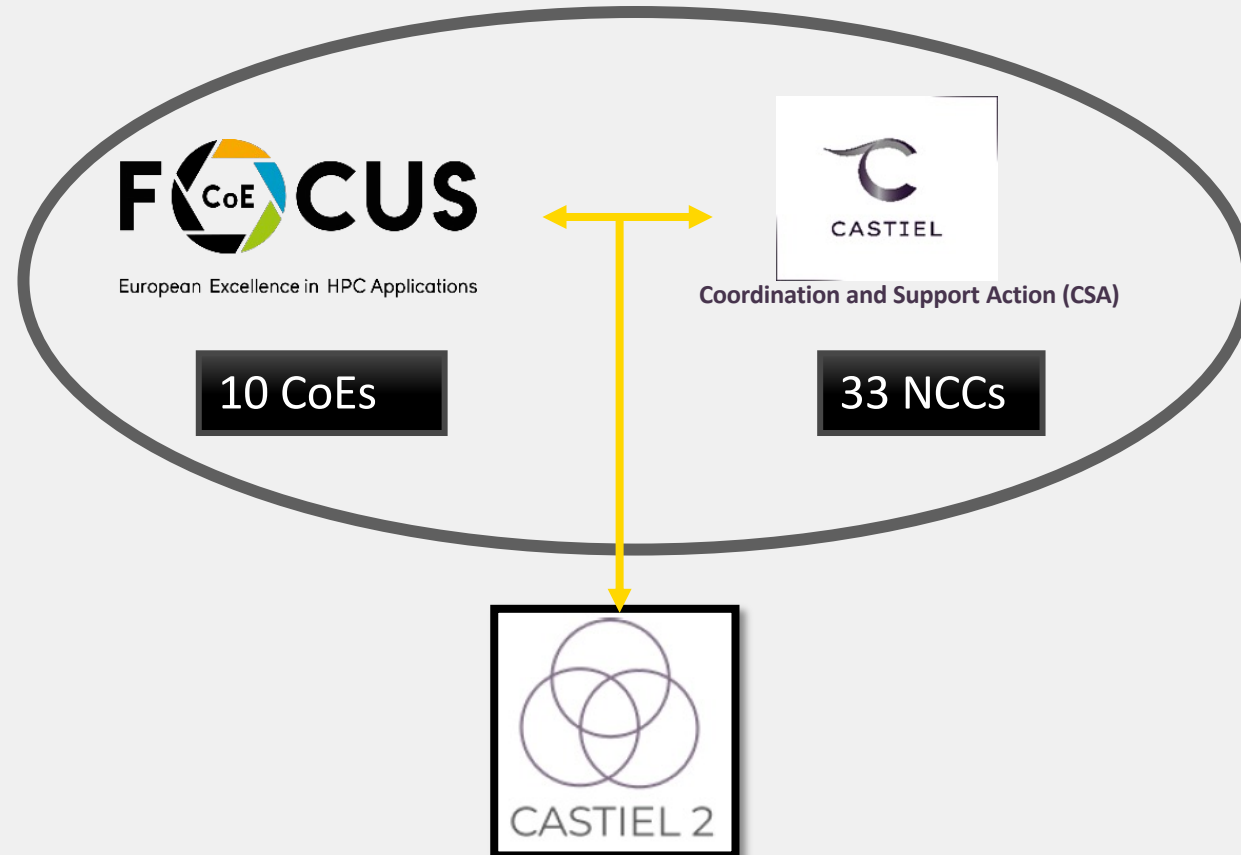
# Setup: Same structure in the 33 countries

Task	EuroCC 2
X.1	NCC Management
X.2	Training and Skills Development
X.3	Services to and Interaction with Industry
X.4	Services to and Interaction with Academia and Public Administration
X.5	Service Portfolio and Competence Management, Additional Services
X.6	Collaboration
X.7	Awareness Creation and Communication

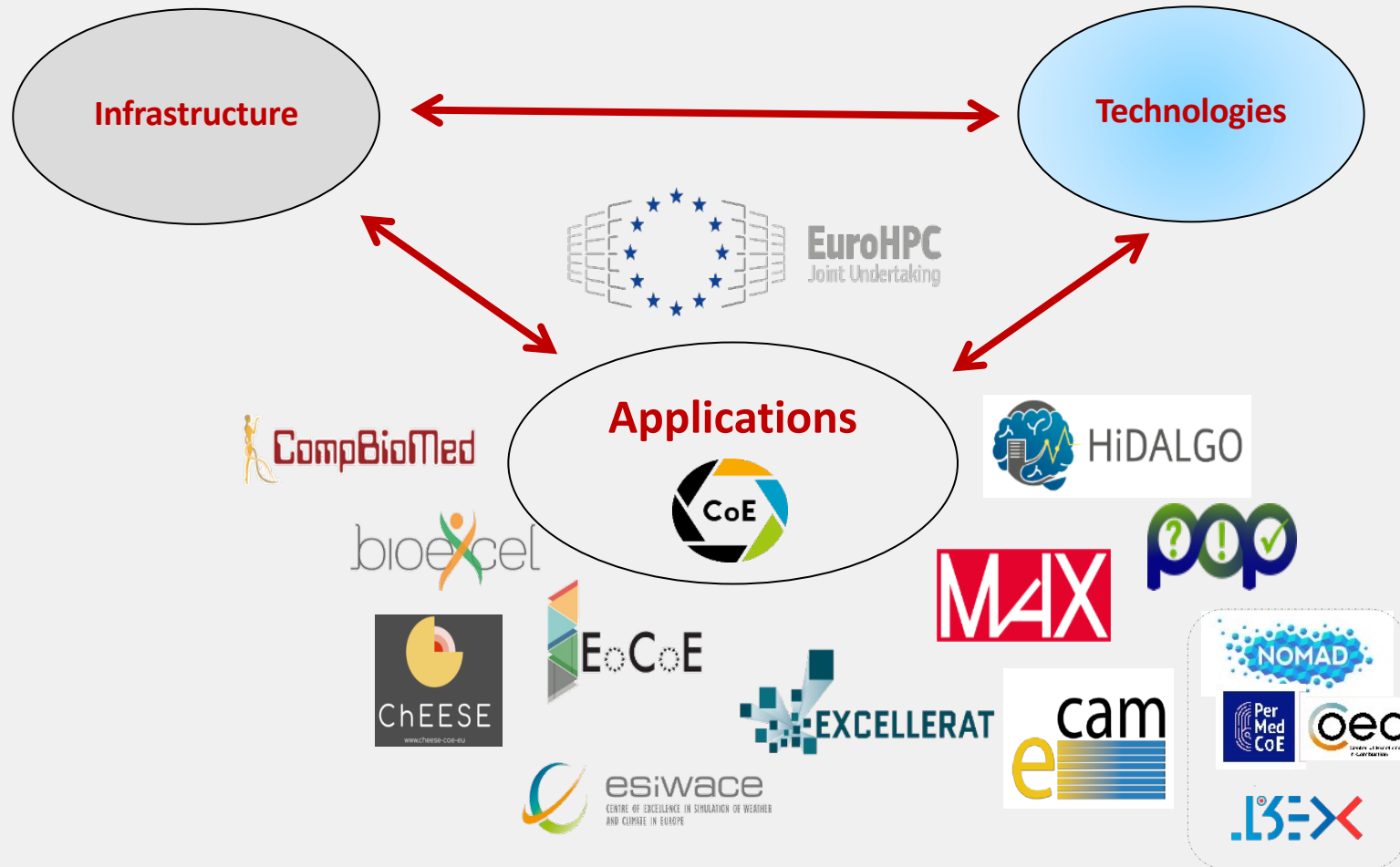
CASTIEL-2 coordinates interactions among NCCs & CoEs to put in common knowledge & best practices

# CASTIEL-2 = Phase II of CASTIEL-1 & Focus CoE

Started: 01/01/2023, 3 years duration  
+ new members: 10 CoEs



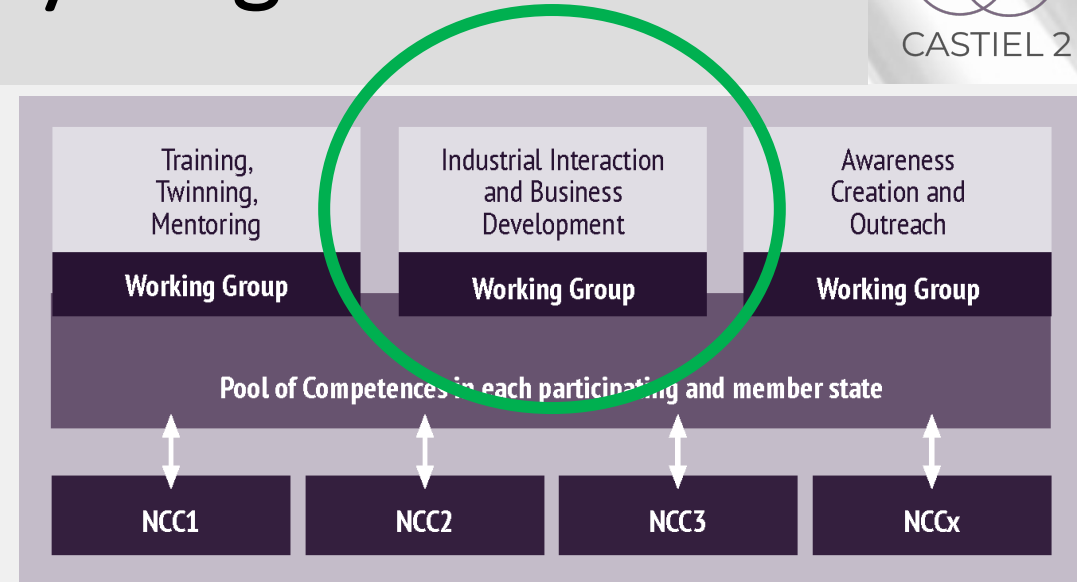
# Centres of Excellences (CoEs) : focus on a specific sector/field





## ❖ Coordination & Support Action

- Phase 1 Started 01/09/2020
- Phase 2 started 01/01/2023



## ❖ Missions of CASTIEL

- To contribute to the success of the activities of the National Competence Centres (NCC) of EuroCC **by identifying and closing gaps** due to the diverse levels of maturity between the different nations.
- Maximize existing **European HPC knowledge and expertise** across Europe
- Elaboration on **best practices** to strengthen the NCCs
- Support by **enabling access to knowledge, expertise, industry use cases and success stories**

**Maximize the visibility**  
of the European HPC knowledge and expertise  
across Europe

**1. Support outreach & awareness actions from the NCCs,  
and centralize their inputs**



## Voxo AB SHAPE Access to Improve Swedish Text-to-Speech Algorithms

ENCCS (EuroCC National Competence Centre Sweden). ENCCS provides high-performance computing training and support for industry, academia and public administration for free. [www.enccs.se](http://www.enccs.se)



**Industrial Organisations Involved**  
Voxo AB (<https://www.voxo.se>) is a Stockholm-based startup that specializes in extracting, analysing, and visualising voice data. Their services are used in multiple industries to provide insights and enable data-driven business development.

**VOXO**

**Technical Challenge**  
Tools such as Apple's Siri, Amazon's Alexa, and Google Home have brought text-to-speech capabilities to the masses. These conversational assistants respond to natural-language requests and reply in kind. They use machine-learning models trained on large amounts of recorded speech samples matched with the corresponding text. When the assistant wants to say something, the model is able to build new utterances that sound natural.

These big tech companies also provide APIs to access such capabilities, and those support many languages. To use them, the user has to send the text to their server and receive the generated speech back. This is not relevant when the text pertains to someone's personal data. In the European Union (EU), GDPR requires that such data be handled correctly, and in particular not transmitted outside the EU. Using a third-party API of a trans-national company cannot provide the required transparency.



Johan Widenhult, CEO Voxo AB, at the ENCCS Industry Days 2023



# Information Industry access to the JU supercomputers



EuroHPC  
Joint Undertaking



## Access to the supercomputers of the JU Industrial usage

Information collected by CASTIEL-2 for the benefit of the NCCs and CoEs members

July 2023

### Table of content

Summary of the available information about the EuroHPC-JU Supercomputers (focus on **Industry access**) ..... 2

Price list for commercial use of HPC resources (at date June 2023) ..... 7

- LEONARDO..... 7
- LUMI Computing Services..... 7
- MARENOSTRUM 5 ..... 7
- MELUXINA ..... 7
- HPC Vega resources ..... 8
- KAROLINA ..... 8
- DEUCALION..... 9
- DISCOVERER ..... 9

Technical specifications of the JU supercomputers..... 10

References ..... 10

### For each Supercomputer:

- Country
- Contact people for the industry access
- Possible usages
- Details on the possible usages
- Opportunities for “Test before use”  
& Proof of Concept Availabilities for the industry users





## Energy

3

16

## Company

Fluke® is the technology leader for time series management in the energy industry. Its PowerTUN® Technology enables and accelerates innovation.

### Challenges & Solution

The energy sector generates more data than any before – more than 8 TB in some sectors. In order to add more advanced big data analytics capabilities, the company looked for technology on a supercomputer. Parag Data selected Hadoop in the support of the Little Big Data (LBD) cluster, an HPC system at T2 Wisc. The successful integration of time series management (TSM) software on a highly parallel cluster enables Hadoop users to further develop the tools for analysis of very large data sets directly through the TSM system.

"With the guidance of H. Kuvshinov, we decided to export our time series data to parquet files. ... With all the data residing in the distributed network file system reading subsets of data – roughly 20 GB – into a Spark distributed data frame became both doable and reasonably fast."

Simon Harrison, *Future Austria: influencing Austria's*

17

## Company

Founded in 1896, 152 years ago, Tycohal is today one of the world's leading engineering companies for energy, water and infrastructure projects.

**Challenges & Solution:**  
In 2011, the federal agency of nuclear control of Belgium reported defects in the vessels of the nuclear reactor Doel and Tihange 2. Among a lot of analyses, numerical simulation has been chosen to assess the risk level.

These simulations were highly computationally demanding: single computations required a high quantity of memory (up to 1GB) for the first computation – feasibility proved up to 1GB). Due to the high number of configurations to compute (one hundred), combined with the memory needs, the use of PC infrastructure was a requirement.

hardware resources while satisfying high-quality standards in the results. Current development methodologies to face the challenges (number of configurations, memory and execution time limitations, cost economical...) involve mainly technical managers of system & structural integrity issues (structural

Feldtspäter, Institut für Virologie, Universität Bonn, 53115 Bonn, Germany

18

100

specific units of power engineering, thermal engineering and ecology, generally in the processes of fuel energy conversion and electricity production and distribution.

The main objective was to determine whether Computational Fluid Dynamics simulations could be used for the first and efficient description of the selective catalytic reduction technology (SCR) catalyst process and, therefore, as a tool to monitor the design of a computational application for the design of this technology.

The use of numerical modeling and simulation will make the process more efficient, and faster and, most importantly, reduce the experimental uncertainties of this process. It also will enable subsequent optimization of the designed solution and, finally, in high-performance computing, it will be possible to complete these simulations in a relatively short time.

By spreading up the design process of the AEM technology, time and cost savings are achieved and, in addition, the optimized AEM technology design leads to more effective NOX emissions reduction and extended technology lifetime, which has a positive impact on the environment."

Contact person at the NCT: Tonnes Kornum, [tonnes.kornum@nct.no](mailto:tonnes.kornum@nct.no)

19

Environment/climate/  
weather

2

20

[illegible]

Sweden's Meteorology & Hydrology Institute digitizes archival tabular data using LUMI

**Company**  
The Swedish Meteorological and Hydrological Institute (SMHI) is an expert authority with a global perspective and a vital task of predicting changes in weather, water and climate.

### Challenges & Solution

While processing, transfer of critical data of observations spanning decades in paper format. The architecture of the project is to capture and train a sufficiently accurate machine learning model which can handle different format of tabular data, convert handwritten text and produce machine-readable. While able to use a combination of image processing and machine learning to achieve this, the digitization pipeline is implemented in Python, using well-known open-source scientific libraries such as color-image and bioimage.

"It MPC allocation enables us to rapidly test and develop the product. [...] GPU allow faster tuning hyperparameters of the model. On CPU the neural network training takes 11 hours. On GPU the whole

Author: Michael, Scientist's employment at NMH

21

[illegible]

The multiphysics experiments of the Weather Research and Forecasting Model (WRF) on precipitation patterns of Turkey

**Company**

**Challenges & Solutions**

WRF might may provide accurate representation of the atmosphere at the surface conditions. However, such WRF model require proper parameterizations (e.g., cloud, planetary boundary layer, other physical options) to be optimized. On the other hand, simulations of WRF may require high computational resources to accomplish such optimization.

The project's scientific team was completed over the Turkey phase of the study, with 30 contributions of studies and papers to three volumes of scientific papers, including the Turkey volume. The next two volumes will be devoted to the other countries, and the third will contain the synthesis and conclusions. The project is expected to be completed by the end of 2000.

*TÜTAK's enhanced forecasting in Türkiye benefits climate-dependent domains by improving resource prediction, hydrometeorological forecasting, and climate change mitigation. This capacity aids in better water*

sectors in risk and operational handling." Unless it is

22



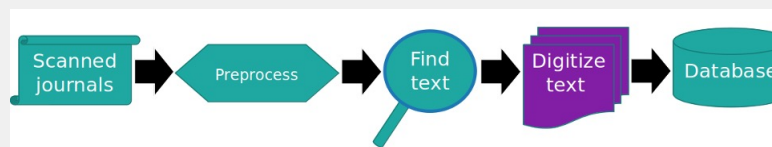
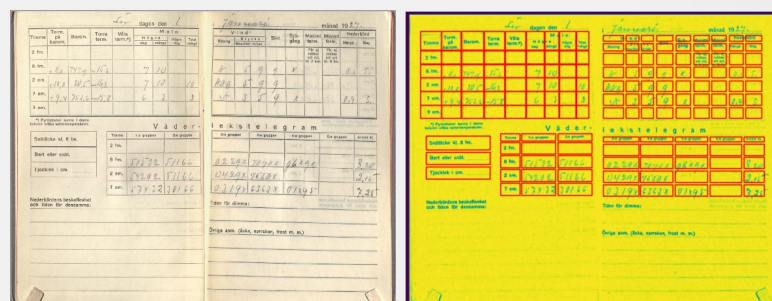
## Sweden's Meteorology & Hydrology Institute digitizes archival tabular data using LUMI

### Company

The Swedish Meteorological and Hydrological Institute (SMHI) is an expert authority with a global perspective and a vital task of predicting changes in weather, water and climate.

### Challenges & Solution

SMHI, possesses troves of archival data of observations spanning decades in paper format. The ambition of the project is to optimize and train a sufficiently accurate machine learning model which can handle different forms of tabular data, convert handwritten-text and produce machine-readable. SMHI aims to use a combination of image processing and machine learning to achieve this. The digitization pipeline is implemented in Python, using well-known open-source scientific libraries such as scikit-image and TensorFlow.



### Benefits

This project aids and accelerates the digitization work from the paper archives into data, which is done manually as of now. As a result of the project, SMHI aims at digitizing numerous historical weather observations that will help a better understanding of climate, especially of the occurrence of extreme weather events.

*"A HPC allocation enables us to rapidly test and develop the product. (...) GPUs allow faster tuning hyperparameters of this model. On CPUs the neural network training takes 11 hours. On GPUs the whole training takes only 1 hour."*

**Ashwin Mohanan, Scientific programmer at SMHI**

### Full story:



**Maximize the visibility**  
of the European HPC knowledge and expertise  
across Europe

## 2. Collaborative answer for start-ups & SMEs

OPEN CALL  
EuroCC Supercomputing Accelerator  
\*\*\*  
Free service package for startups and SMEs to  
optimise computing tasks and grow with advanced technology

APPLY HERE

APPLY ON F6S

- Need more compute power to train AI models, analyse big data or run simulations?
- Want to stay ahead of the competition by using advanced computing technologies?

Apply for the EuroCC Supercomputing Accelerator and get an opportunity to grow your business while saving up to €76,000!

## Why supercomputing?

With supercomputing, or high-performance computing (HPC), you can drastically accelerate computing tasks, reduce time-to-market of your product and optimise data-intensive computations such as training machine learning models.

More and more companies worldwide invest in expensive HPC infrastructure, but you don't have to: the European Union is publicly funding HPC access for startups and small and medium enterprises (SMEs), making it possible to use powerful national supercomputing infrastructure for commercial applications.



## What do you get?

Supercomputing Accelerator offers you transfer of valuable HPC know-how tailored to your specific domain of science or industry. By applying for the EuroCC service package, you get:

- ✓ Tech feasibility check
- ✓ HPC training
- ✓ Business plan advice
- ✓ Financing advice
- ✓ Programming support
- ✓ Project support
- ✓ Access to supercomputers

## Common advertising & marketing services to a specific group: Startups & SMEs

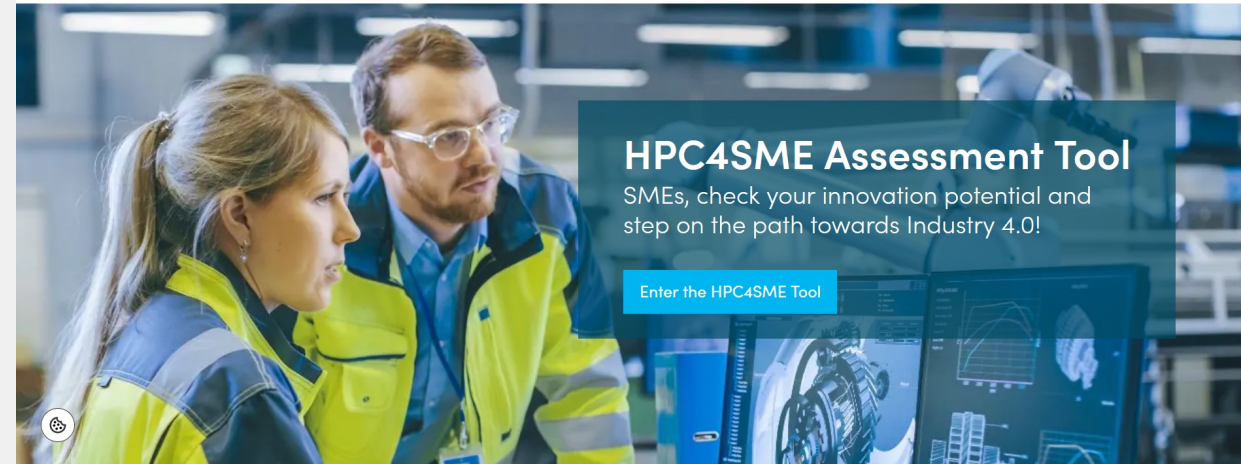
- Through the F6S platform
- To train and support the early users, the startups and some SMEs in part of a program



**Maximize the visibility**  
of the European HPC knowledge and expertise  
across Europe

### 3. Develop tools to facilitate interactions with SMEs

Automatically generated **reports and personalized recommendations** generated by the HPC4SME Automated Assessment Tool (AAT) represent a potential **information resource** for the HPC4SME AAT's **end users, enabling them to chart their course with data-driven insights**



GENERAL QUESTIONS

READINESS

CLOUD

HPC

OVERVIEW (60%)

GENERAL QUESTIONS

A1 General questions

READINESS

B1 Potential

B2 Organizational factors

B3 Internal capacity

CLOUD

C1 Confidentiality

C2 Speed

HPC

D1 Computing Demands

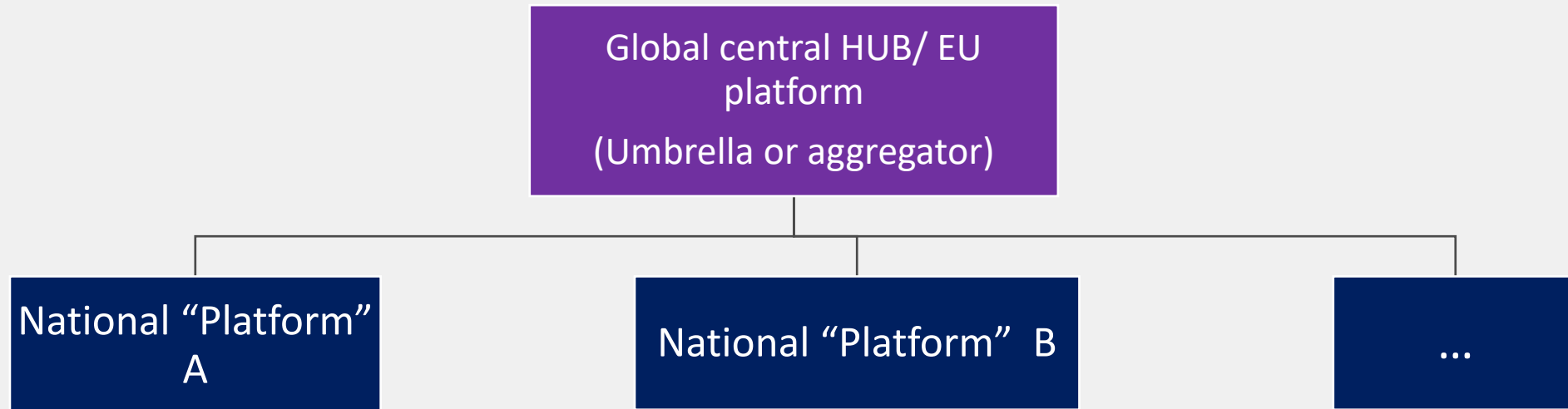
D2 Volume of simulation



## Maximize the visibility

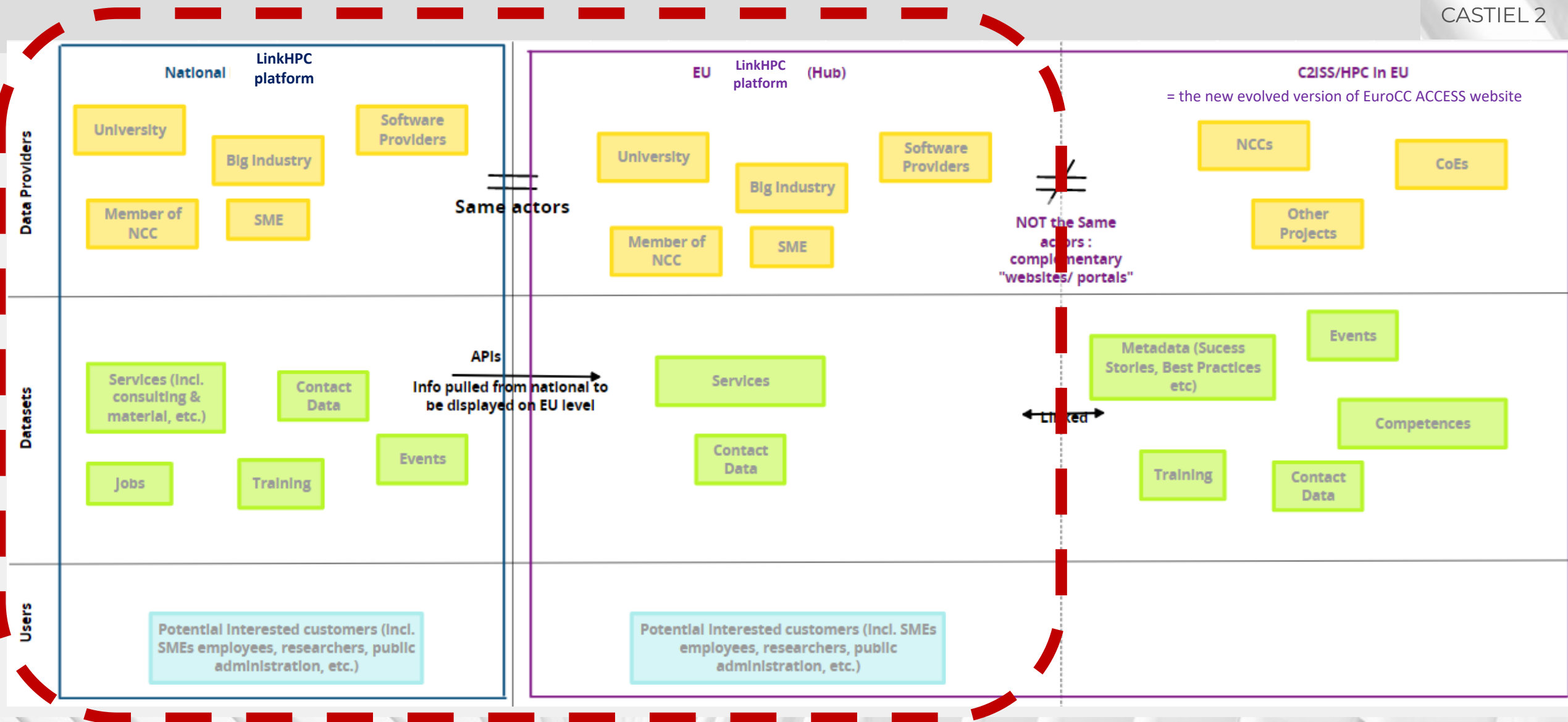
of the European HPC knowledge and expertise  
across Europe

4. Support visibility of the services offerings &  
interaction in the ecosystems  
-> Designing a LinkHPC platform

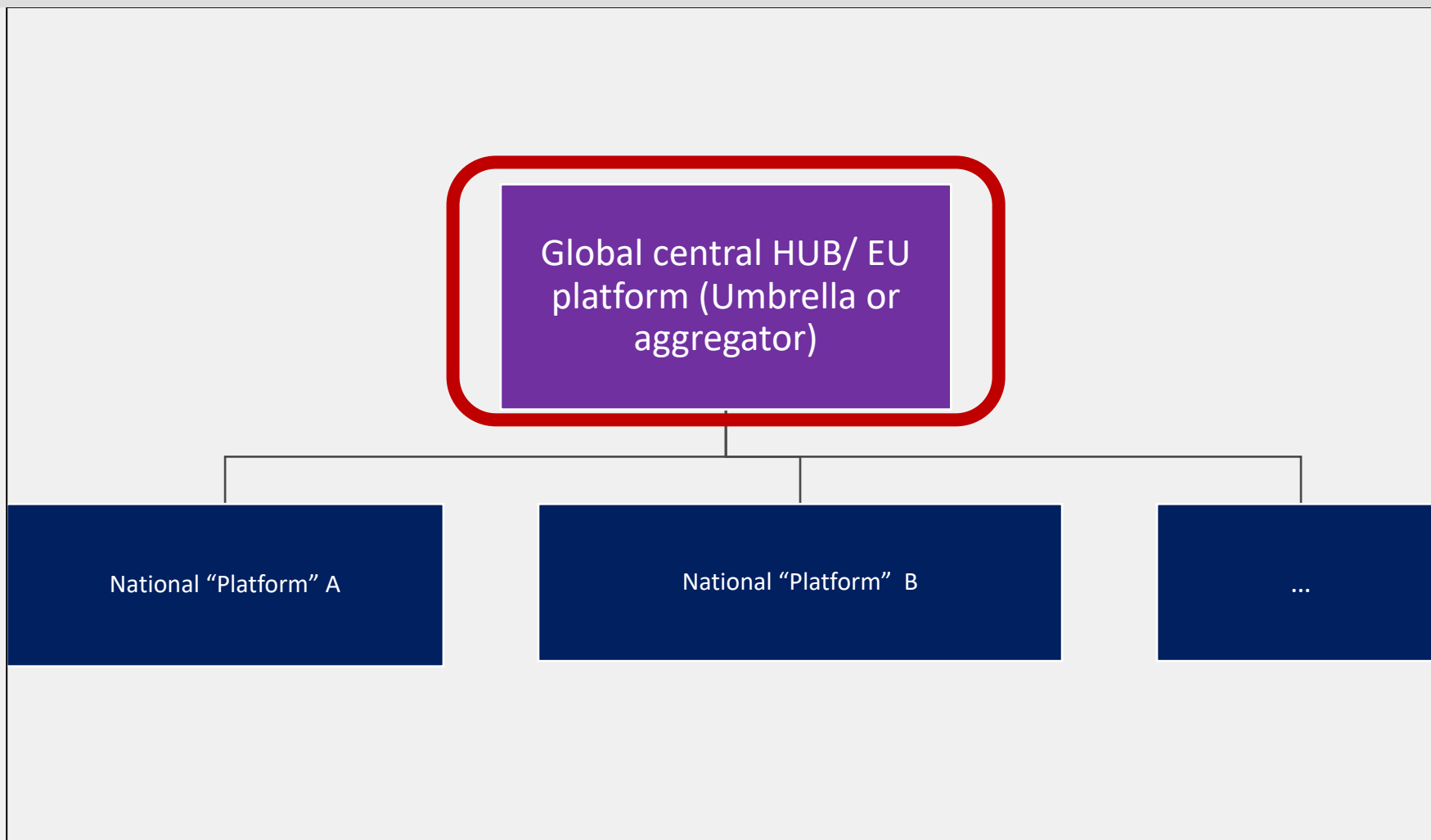


- Common same template for all countries
- Single platform for the country, local branding
- Matchmaking tool for the national ecosystem (Offers, Providers & Users)
- Able to connect with the umbrella platform
- Data: Owned by national center

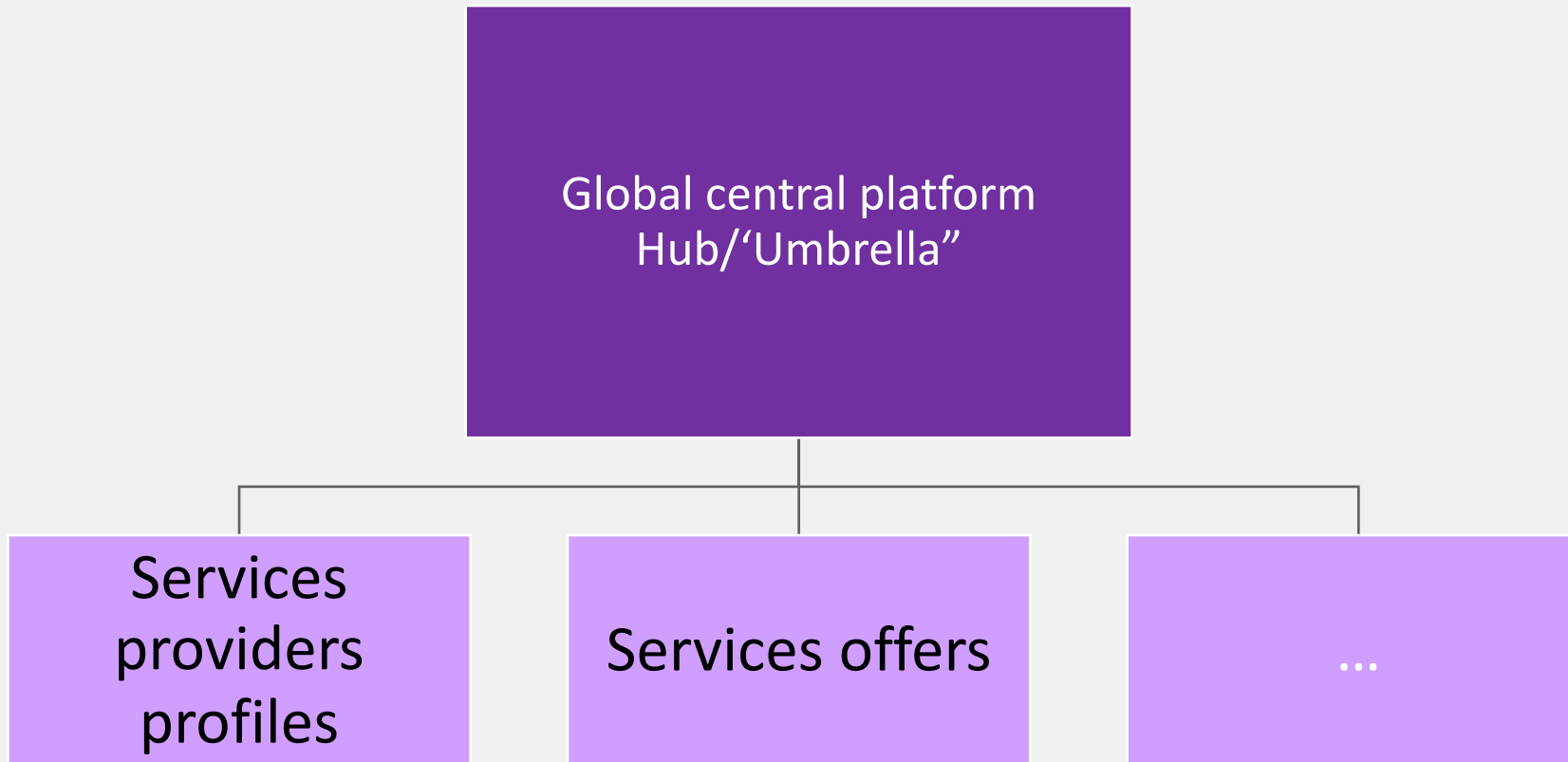
# Global concept of the LinkHPC platform



# Let's focus first on the EU national level



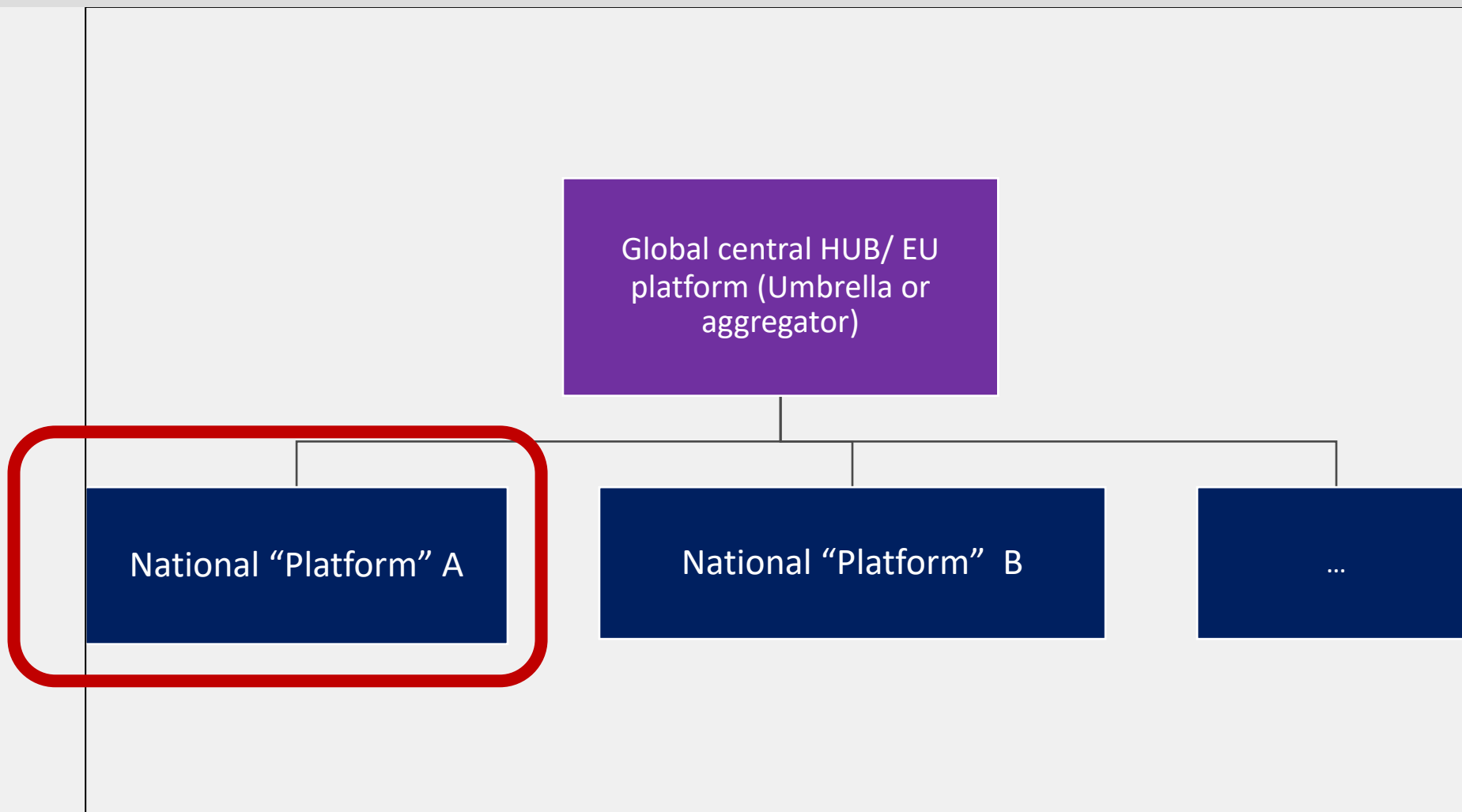
# Central unique level (=Hub, automatically filled in by one common aggregator)

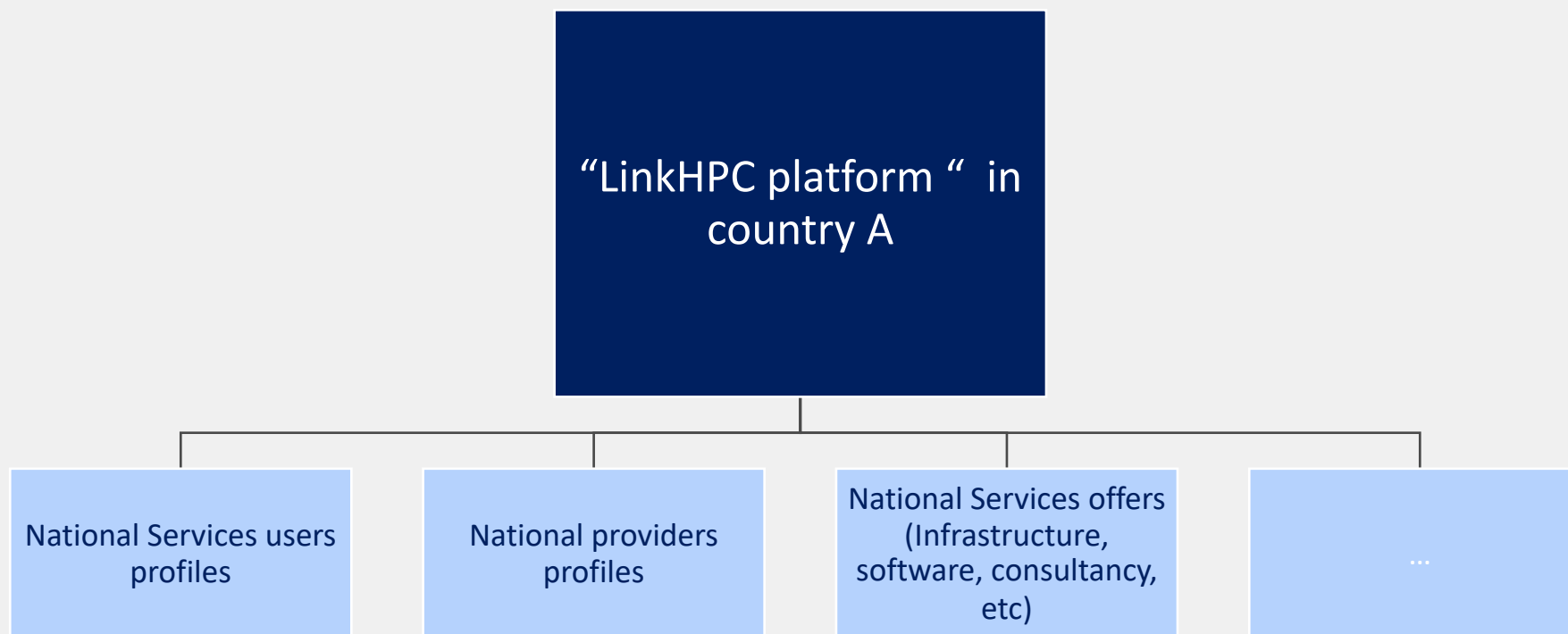


- To display in a central place the agreed local/national information
- Common overview of EU network/actors: information from all national level platforms



# Changing the focus, now on the national level





- To display at the local/national level: **all information about the market/ecosystem**
- Have a **structured visibility** and **promotion** of your ecosystem offers, users, providers, etc.

- Currently testing & working on a common template for the structure of the potential future **national LinkHPC platforms** :  
=> collection of particular specifications or feedbacks by end of 2023
- Test version = an English version of the platform developed in France and implemented in Belgium

**SMEs assessment tool**  
for the SMEs to assess their readiness

## Accelerator

- train and support
- the early users



when ready and support done through the accelerator, they will be recommended to go to

## LinkHPC platform

- stage for the SMEs & startups: nationally & EU level
- to promote their services, their offers or to look for further specific HPC+ competences or existing services
- *In prep*

# Thanks



This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101102047. The JU receives support from the Digital Europe Programme and Germany, Italy, Spain, France, Belgium, Austria.

*Learn more about EuroCC2 here!*

