Project Number:	720270	Project Title:	Human Brain Project SGA1

Document Title:	High-Performance Analytics and Computing Platform specification update
Document Filename:	SGA1_D7.6.1_FINAL
Deliverable Number:	SGA1 D7.6.1 (Update of RUP D7.7.2) (D42.1, D11)
Deliverable Type:	Report
Work Package(s):	WPs 7.1, 7.2, 7.3, 7.4, 7.5, 7.6
Dissemination Level:	PU (= public)
Planned Delivery Date:	SGA1 M30 / 30 Sep 2016
Actual Delivery Date:	SGA1 M30 / 30 Sep 2016

Authors:	Thomas LIPPERT, JUELICH (P20), SP Leader Thomas SCHULTHESS, ETHZ (P18), SP Leader
Compiling Editors:	Anna LÜHRS, JUELICH (P20), T7.6.1, T7.7.1, SP Manager
Contributors:	Anna LÜHRS, JUELICH (P20), T7.6.1, T7.7.1, SP Manager Cristian MEZZANOTTE, ETHZ (P18), WP7.5 Colin MCMURTRIE, ETHZ (P18), WP7.5, Technical Coordinator Boris ORTH, JUELICH (P20), T7.6.1, T7.7.1, SP Manager
SciTechCoord Review:	EPFL (P1): Jeff MULLER, Martin TELEFONT UHEI (P47): Martina SCHMALHOLZ, Sabine SCHNEIDER
Editorial Review:	EPFL (P1): Guy WILLIS, Colin McKINNON, Lauren ORWIN

Abstract:	*This deliverable replaces the Ramp-Up Phase Deliverable D7.7.2 and provides an update of the High Performance Analytics & Computing (HPAC) Platform specification.* A key feature of the HPAC Platform is its integration with the HBP Collaboratory, a Web Portal that provides access to many services and tools of the HBP. After an introductory overview of the overall goals and layout of the Platform, the report describes typical examples of how the HPC Platform is used since its first release in March 2016. From these use cases the functional and non-functional requirements are derived. The software and hardware components used for implementing these requirements are described, and the processes to get access to the Platform resources are outlined. Architecture diagrams illustrate how the hardware and software components relate to each other, as well as how they are combined for providing services to the users. The relations to the other HBP Platforms and necessary parallel activities are described.

Keywords:	High Performance Computing (HPC), Platform, Collaboratory, use cases, functional requirements, non-functional requirements, software, hardware, physical architecture, architecture overview, research infrastructure

Table of Contents

List of Figures

List of Tables

1. Executive Summary

The Human Brain Project [1] (HBP) is a European Commission Future and Emerging Technologies Flagship. The HBP aims to put in place a cutting-edge, ICT-based scientific Research Infrastructure for brain research, cognitive neuroscience and brain-inspired computing. The Project promotes collaboration across the globe, and is committed to driving forward European industry.

The High Performance Analytics & Computing (HPAC) Platform Subproject (SP7) is one of the HBP's eleven scientific subprojects. During the project's Ramp-up Phase (RUP) it turned out that in addition to brain simulations as the initial main driving use case, data analytics is another, similarly important and challenging research area in the HBP, that requires resources provided by SP7. Consequently, SP7 was renamed from High Performance Computing Platform to High Performance *Analytics* & Computing Platform, to reflect this refocusing. The HPAC Platform has been released for the wider scientific community in project month 30 of the Ramp-up Phase, that is in March 2016, already under this new name.

Over the 10-year duration of the HBP, SP7 will build, integrate and operate the HBP's HPAC Platform. This comprises all hardware and software components of the supercomputing, data and visualisation infrastructure required to:

- Run large-scale, data intensive, interactive brain simulations up to the size of a full human brain,
- Manage the large amounts of data used and produced by the simulations, and
- Manage complex workflows comprising concurrent simulation, data analysis and visualization workloads.

The HPAC Platform comprises supercomputing capabilities at Forschungszentrum Jülich (JUELICH), Germany, the Swiss National Supercomputing Centre (CSCS) in Lugano, Switzerland, the Barcelona Supercomputing Center (BSC) in Barcelona, Spain, and the Consorzio Interuniversitario del Nord Est italiano per il Calcolo Automatico (CINECA) in Bologna, Italy. The Platform also provides access to high-fidelity visualisation systems at RWTH Aachen University (RWTH), Germany, and EPFL in Geneva, Switzerland, as well as cloud storage with a dedicated capacity of 3 Petabytes at the Karlsruhe Institute of Technology (KIT), Germany.

All of these systems are connected logically through UNICORE (Uniform Interface to Computing Resources) and physically through the high-speed network of PRACE (Partnership for Advanced Computing in Europe), with a bandwidth of 10 Gigabits per second, and other secure connections. A key feature of the HPAC Platform is its integration with the HBP Collaboratory that enables to seamlessly start and manage HPC jobs from this web portal.

This report specifies in detail the High Performance Analytics & Computing Platform Version 2.0. It is a revised version of the HPC Platform specification that was published as Ramp-up Phase Deliverable D7.7.2[2] in version 1.0. In the next update of this "living" document, the integration of the federated data infrastructure into the HPAC Platform will be described. The architecture of this infrastructure is being defined and implemented in the Federated Data Pilot Project (FeDaPP), a project between CSCS, JUELICH and CINECA. The infrastructure will be operated under the name of FENIX.

This document is intended for a technical and scientific audience. After an introductory overview of the overall goals and layout of the Platform, the report describes typical examples of how the HPC Platform is used since its first release in March 2016. From these use cases the functional and non-functional requirements are derived. The software and hardware components used for implementing these requirements are described, and the processes to get access to the Platform resources are outlined. Architecture diagrams illustrate how the hardware and software components relate to each other, as well as how they are combined for providing services to the users. The relations to the other HBP Platforms and necessary parallel activities are described.

2. The High Performance Analytics & Computing (HPAC) Platform

2.1 HPAC Platform: Overall Goals

2.1.1 The Human Brain Project

The Human Brain Project (HBP) is a European Commission Future and Emerging Technologies Flagship. It aims to put in place a cutting-edge, ICT-based scientific Research Infrastructure for brain research, cognitive neuroscience and brain-inspired computing. The Project promotes collaboration across the globe, and is committed to driving forward European industry. It was launched in October 2013, and is scheduled to run for ten years. The HBP has the following main objectives:

- Create and operate a European scientific Research Infrastructure for brain research, cognitive neuroscience, and other brain-inspired sciences
- Gather, organise and disseminate data describing the brain and its diseases
- Simulate the brain
- Build multi-scale scaffold theory and models for the brain
- Develop brain-inspired computing, data analytics and robotics
- Ensure that the HBP's work is undertaken responsibly and that it benefits society.

The HBP develops six Information & Communication Technology (ICT) Platforms, dedicated respectively to Neuroinformatics, Brain Simulation, High-Performance Analytics and Computing (HPAC), Medical Informatics, Neuromorphic Computing and Neurorobotics. This research infrastructure will make it possible to federate neuroscience data from all over the world, to integrate the data in unifying models and simulations of the brain, to validate the results against empirical data from biology and medicine, and to make them available to the world scientific community.

The resulting knowledge on the structure and connectivity of the brain will open up new perspectives for the development of “brain-like” computing systems incorporating unique characteristics of the brain such as energy-efficiency, fault-tolerance and the ability to learn. The HBP’s models and simulations will also enable researchers to carry out *in silico* experiments on the human brain that cannot be done *in vivo* for practical or ethical reasons.

2.1.2 The HPAC Platform

The HPAC Platform Subproject (SP7) is one of the HBP’s eleven scientific subprojects. Led and coordinated by Forschungszentrum Jülich, it currently involves 14 partner institutions, including four leading European supercomputing centres and several universities. Its 10-year mission is to build, integrate and operate the hardware and software components of the supercomputing and data infrastructure required to

- Run large-scale, data intensive, interactive brain simulations up to the size of a full human brain,
- Manage the large amounts of data used and produced by the simulations, and
- Manage complex workflows comprising concurrent simulation, data analysis and visualization workloads.

This infrastructure as a whole forms the *HPAC Platform*. The HPAC Platform was released to the HBP Consortium in project month 20 of the Ramp-up Phase (RUP) and to the wider scientific community in month 30. The HPAC Platform is accessible in a seamless and intuitive manner through the HBP Collaboratory.

The HPAC Platform comprises supercomputing capabilities at Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich (JUELICH), Germany, at the Swiss National Supercomputing Centre (CSCS) in Lugano, Switzerland, the Barcelona Supercomputing Center (BSC) in Barcelona, Spain, and the Consorzio Interuniversitario del Nord Est italiano per il Calcolo Automatico (CINECA) in Bologna, Italy. The Platform also provides access to high-fidelity visualisation systems at RWTH Aachen University (RWTH), Germany, and EPFL in Geneva, Switzerland, as well as cloud storage with a dedicated capacity of 3 Petabytes at the Karlsruhe Institute of Technology (KIT), Germany.

In the first version of the HPC Platform specification (RUP Deliverable D7.7.2), the Platform comprised the following hardware components:

- HBP Supercomputer JUQUEEN at JUELICH
- HBP Development System BlueBrain IV hosted by CSCS
- HBP Molecular Dynamics Supercomputer MareNostrum at BSC
- HBP Massive Data Analytics Supercomputer FERMI at CINECA
- HBP Cloud Service at KIT
- HBP High Fidelity Visualisation Systems: CAVE installation and a tiled display wall at RWTH

BlueBrain IV has been replaced in the HPAC Platform by Piz Daint at CSCS at the end of the Ramp-up Phase. The GPU cluster JURECA at JUELICH and the cluster Pico at CINECA have been integrated into the HPAC Platform at the same time. These additional systems were added since their architecture is particularly suitable for image processing and data analytics use cases. FERMI reached the end of its operational phase in July 2016 and has been replaced in the HPAC Platform by CINECA’s new system Marconi.

Due to the integration of additional systems, it has been decided to drop the original “HBP labels” that were used for the first four supercomputers in the Platform, i.e. “HBP Supercomputer” for JUQUEEN, “HBP Development System” for BlueBrain IV, “HBP Molecular Dynamics Supercomputer” for MareNostrum III and “HBP Massive Data Analytics Supercomputer” for FERMI.

The pilot systems that are currently deployed as part of the third and last phase of the HBP Pre-Commercial Procurement (PCP) will be integrated into the HPAC Platform in autumn 2016, as soon as they are fully operational.

Another display wall at EPFL has been integrated into the HPAC Platform during the Ramp-up Phase as well.

In summary, the following hardware components are part of the HPAC Platform at the end of September 2016:

- JUELICH: JUQUEEN and JURECA
- CSCS: Piz Daint
- CINECA: Marconi, Pico
- BSC: MareNostrum
- HBP Cloud Service at KIT
- HBP High Fidelity Visualisation Systems: CAVE installation and display wall at RWTH, display wall at EPFL

The HPC systems at JUELICH, CSCS, BSC and CINECA are logically connected through UNICORE (Uniform Interface to Computing Resources) and physically through the high-speed network of PRACE (Partnership for Advanced Computing in Europe), with a bandwidth of 10 Gigabits per second, and other secure connections. A Memorandum of Understanding between PRACE and the HBP was signed in January 2015 for this purpose. The HBP Cloud storage at KIT is connected through regular internet connections. A dedicated network connection between RWTH and JUELICH enables a fast data transfer also to this visualisation site.

During the Ramp-up Phase, advanced software for interactive visualisation, resource and data management has been developed in SP7, and is available to the users (see also Ramp-up Phase Deliverable D7.7.5[3]). The documentation can be found in the HPAC Platform Guidebook (https://hbp-hpc-platform.fz-juelich.de/), as defined in Ramp-up Phase Deliverable D7.7.6[4]. The HPAC Platform Guidebook is a living document that is regularly updated.

During SGA1, the research activities in SP7 on interactive visualisation (WP7.3) and dynamic resource management (WP7.4, formerly RUP-WP7.2) are continued. In addition, data-intensive supercomputing (WP7.2) has been added as another important research area that aims to link extreme scale data processing challenges to the exploitation of scalable compute resources. The development of simulation technology (WP7.1) has been integrated into SP7 in order to strengthen the direct collaboration between simulation developers and the HPC experts.

2.2 HPAC Platform: Use Cases

In the first version of this section we presented some typical examples how we expect the HPC Platform to be used during the Ramp-up Phase. With the start of SGA1, the HBP introduced Co-Design Projects (CDPs) as a strategy to refine six ICT Platforms through scientific use cases and to strengthen cross-Platform collaboration. The six CDPs during the SGA1 phase of the project aim to answer scientific questions in a multi-disciplinary effort. The new set of SP7 use cases is a subset of parts of the CDPs. This list will be complemented in the next update with specific use cases for the federated data part of the HPAC Platform that is developed in FeDaPP.

The original set of use cases was mainly focused on brain simulation, which was expected to be the main driving community for building the HPAC Platform. Since it turned out that data analysis and image processing are similarly important and demanding, the use cases described in the following cover all three research areas. Some of the use cases described below were in parts already described in RUP Deliverable D7.7.5, Annex J. The use cases are in different implementation phases and therefore some are already in a more advanced stage than others.

Roles

The HPAC Platform users can be grouped in two main categories: those that need computing resources themselves, and others who need access to datasets and results produced by the first group and that are stored at the HPC centres. The HPAC Platform offers software tools and services developed by the Platform partners for both types of users. The execution of some of these tools, for example for visualisation and data management, does not necessarily require a supercomputer, but they can be used on standard computers and notebooks. Other software, like parallel programming frameworks, can be used on both types of architectures. The HPAC Platform Guidebook (https://hbp-hpc-platform.fz-juelich.de/) describes for every software tool and library on which systems it can be used and where to find more documentation about it.

In the first version of the specification we already defined the following eight user roles:

- Computational Scientific User (CSU) – A user with scientific development skills and comfort in launching command line HPC jobs.
- Biological Scientific User (BSU) – A user with scientific expertise, but limited technical skills in programming and/or HPC.
- Scientific User (SU) – A scientific user, either a CSU or a BSU.
- Scientific Developer (SCIDEV) – A user developing software to directly realise certain scientific objectives. This user is usually working in close collaboration with scientists, both CSUs and BSUs.
- Developer (DEV) – A user developing software to realise engineering, operational and/or scientific objectives.
- Portal User (PU) –A user who accesses Platform functions through the Collaboratory or other web portals.
- Service User (ServU) – A user who accesses Platform functions through a programmatic Service Client API.
- Infrastructure Personnel (INFRA) - An infrastructure system administrator or developer, typically responsible for deploying and monitoring Platform services that are offered directly to customers.

Computational Scientific Users (CSU), Scientific Developers (SCIDEV) and Developers (DEV) mainly belong to the above described, first category of users. Biological Scientific Users

(BSU), Portal Users (PU) and Service Users (ServU) fit more into the second category but they might also have a need for computing time themselves.

2.2.1 Simulation and analytics workflow: NEST – Elephant - SpiNNaker

Use Case ID: SP7-UC-001

Collaborating SPs: SP5, SP6, SP9

2.2.1.1 Description

A NEST simulation of a neural network model and the simulation of the same model on SpiNNaker (SP9) ideally lead to the same results. This use case aims at implementing a workflow that starts a simulation from the Collaboratory on both backends, i.e. a NEST simulation on an HPC system and another simulation of the same model on SpiNNaker, and that analyses and compares the results with the analysis tool Elephant (formerly SP5, now SP9). This way it can be evaluated for the first time if the NEST-HPC and the NM implementations lead to similar or the same results when simulating the same model of a neural network. A second step will be to also compare experimental data with the simulation results. The implementation of this use case will happen in two phases (see Figure 1 and Figure 2).

2.2.1.2 Problems to be solved

The following problems require special attention and need to be solved in co-design with the users for implementing the use case successfully.

- Provenance tracking of the entire, complex workflow is needed in order to enable reproducibility.
- The results of the neuromorphic simulation and the experimental datasets need to be transferred to the HPC storage and/or to the Collab storage for the analysis step with Elephant.
- The simulation and analysis results need to be visualised, which may require downloading the data from the HPC storage to the user's computer.
- A seamless integration with the Collaboration is required to start the workflow from this portal, which in particular requires submitting and managing HPC jobs.
- The analysis tool Elephant is not parallelised yet. Therefore, using a parallel programming framework to enable (trivial) parallelisation of advanced analysis routines would allow the analysis of significantly larger datasets directly on a supercomputer.

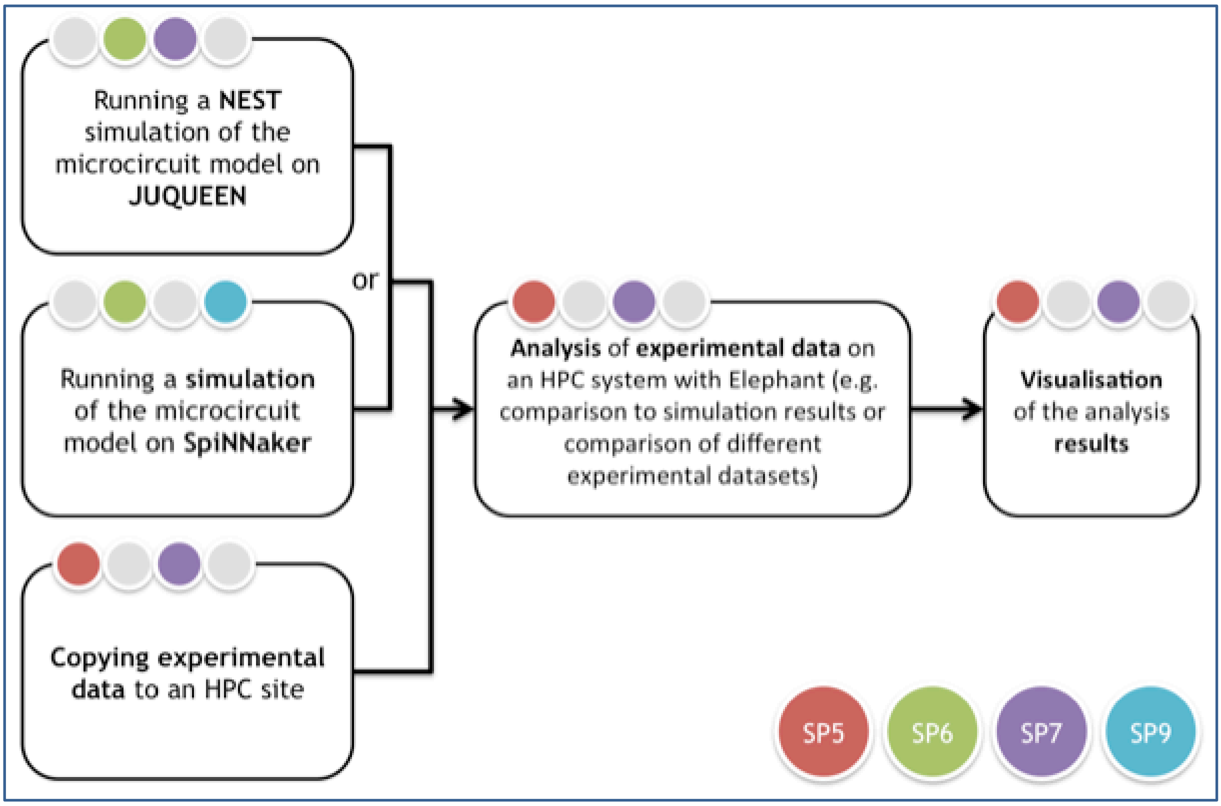

Figure 1: NEST-Elephant-SpiNNaker workflow implementation phase 1

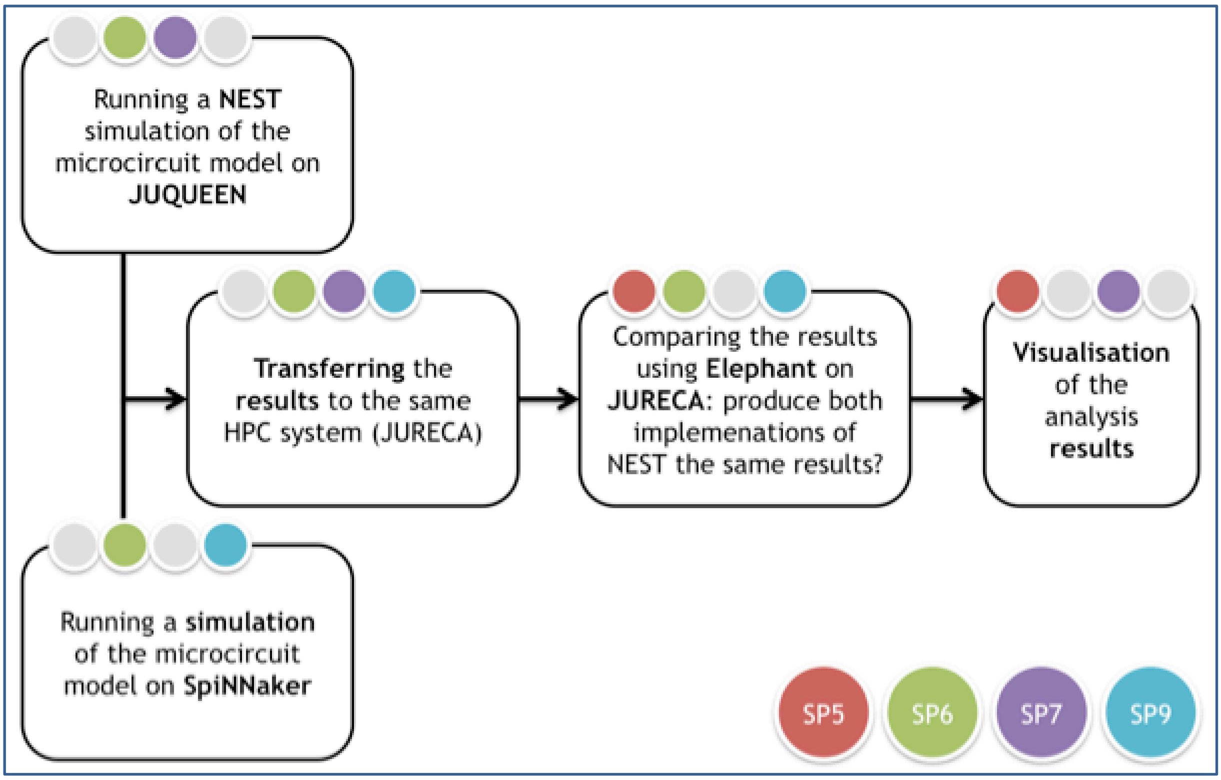

Figure 2: NEST-Elephant-SpiNNaker workflow implementation phase 2

2.2.2 Simulation: Circuit simulation with Neuron

Use Case ID: SP7-UC-002

Collaborating SP: SP6

2.2.2.1 Description

This use case consists of the following six steps:

1) Creation of a circuit on a supercomputer.
2) Release of this circuit: the data is visible to the HBP community through the Collaboratory.
3) A user from another group configures a Neuron simulation based on the circuit model in the Collaboratory.
4) The user launches the simulation from the Collaboratory.
5) The user analyses the simulation results:
    a. With a predefined analysis that is available in a Collab, and/or
    b. By writing an own analysis, and/or
        - Writing and committing the analysis
        - Defining the dependencies
        - Selecting this newly defined analysis in the Collab
        - Selecting the simulation results in the Collab
        - Selecting the target system and job requirements
    c. By performing an interactive analysis using a notebook in the Collaboratory.
        - Launching a notebook from the Collaboratory that has access to the simulation reports and that has some predefined installed software
        - Executing the Python commands from the notebook that access the simulation reports
        - Saving the notebook
6) The user launches the simulation reports and analysis in the Collaboratory to make the results available and reproducible.

2.2.2.2 Problems to be solved

The following problems require special attention and need to be solved in co-design with the users for implementing the use case successfully.

- Provenance tracking of the entire, complex workflow is needed in order to enable reproducibility.
- Visualisation of the simulation results.
- A seamless integration with the Collaboratory is required to run the workflow from this portal, which in particular requires submitting and managing HPC jobs.

2.2.3 Simulation: Interactive neurorobotics simulation

Use Case ID: SP7-UC-003

Collaborating SPs: SP10

2.2.3.1 Description

The frontend of the Neurorobotics Platform (NRP) enables the user to design virtual experiments with robots, to start the simulation from the Collaboratory and to visualise the running experiment. The neural simulation backend can either be neuromorphic systems (SP9) or NEST. This backend is currently deployed on OpenStack Virtual Machines (VM) managed by Puppet. The VMs used for Gazebo, the robot simulation, additionally require access to GPUs. The NEST simulations require a few cluster nodes with MPI that are managed by Slurm.

The currently used compute resources limit the maximum number of concurrent users that can run their virtual experiments in parallel. Therefore, the same services should be deployed on the HPAC Platform as well to reach a higher scalability. This requires access to a few nodes of at least one HPC system, some of which should contain GPUs for running Gazebo. Additionally, suitable deployment mechanisms are required, which at least in a first phase needs to be container-based.

2.2.3.2 Problems to be solved

The following problems require special attention and need to be solved in co-design with the users for implementing the use case successfully.

- Enabling container-based deployment of the simulations on an HPC system, including GPU support for Docker.

2.2.4 Data analysis: Texture analysis in high-resolution images

Use Case ID: SP7-UC-004

Collaborating SP: SP5

2.2.4.1 Description

The goal of this project is the implementation of a workflow for texture analysis in high-resolution imaging data for the brain atlas that is based on machine learning. It is part of a larger analysis workflow, i.e. the results of this project are intermediate results of the larger workflow and not meant to be provided directly to "end users". The results must be quality-checked by humans before they can be processed further in the larger workflow. It is planned to realise this manual interaction through a web-based viewer and it can also be used to improve the learning algorithm. When the result has been approved, it should be possible to signal this to the larger workflow so that it continues.

2.2.4.2 Problems to be solved

The following problems require special attention and need to be solved in co-design with the users for implementing the use case successfully.

- Provenance tracking of the entire, complex workflow is needed in order to enable reproducibility.
- Integration of a viewer for HDF5 files into the Collaboratory, that also has access to the data stored in the HPAC Platform infrastructure.
- The same storage or repository needs to be accessible from the HPC system on which the data is produced, by the viewer and by web server(s) used to transfer the data.

2.2.5 Data analytics: Characterising disease pathways using machine learning

Use Case ID: SP7-UC-005

Collaborating SP: SP8

2.2.5.1 Description

The aim of this project is to characterise complete disease pathways, from the molecular level, up to observable disorders of cognition and behaviour, and to identify unique combinations of biological and clinical signals associated with specific pathways. To pursue this goal, data mining will be used to identify biological signatures of disease across different scales (biological, anatomical, physiological and clinical variables). The data mining will use established machine learning algorithms running on HPC. It is planned to demonstrate the capabilities of using HPC resources for large scale machine learning.

Datasets:

- Research data: imaging, SNP, clinical variable (size <1TB)

Methods:

- Semi-supervised clustering algorithm
  - Matlab, R, native or docker
  - Output: Rules-based classification (size <1GB)
  - Algorithms and benchmarks: Density-based algorithm compared to the use of state of the art "black box" methods
  - Reference time on 120 HPC nodes: less than 1 hour
- Support vector machine classifier trained and cross-validation
  - Matlab, R, native or docker
  - Output: Image based weight (size <1GB)
  - Algorithms and benchmarks: The results were compared to clinical diagnostic performed by neurologists (expert knowledge)
  - Reference time on 120 HPC nodes: less than 1 hour
- Deep learning algorithm
  - Matlab, R, Java
  - Output: Neural set for feature learning (size <1GB)
  - Algorithms and benchmarks:

    Neural net/stacked auto-encoder

    Compared atlas based features vs. random based features

    Compared to clinical label
  - Reference time on 120 HPC nodes: less than 12 hours
- Monte-Carlo bi-clustering
  - Output: Co-modules between gene expression and brain morphologies (size <1GB)
  - Algorithms and benchmarks: Compared to separated clustering approach
  - Reference time on 120 HPC nodes: less than 24 hours

Implementation:

1) Identify multimodal clinical data [done]
2) Data pre-processing [done]
3) Data transfer
4) Implementing a test of different algorithms
5) Model configuration
6) Benchmark algorithms

2.2.5.2 Problems to be solved

The following problems require special attention and need to be solved in co-design with the users for implementing the use case successfully.

- Container-based deployment of the software on HPC systems.

2.2.6 Image processing: 3D-Polarised Light Imaging workflow

Use Case ID: SP7-UC-006

Collaborating SP: SP2

2.2.6.1 Description

3D-Polarized Light Imaging (3D-PLI) is a technique used to extract the orientation of nerve fibres at micrometre scale. A postmortem human brain is cut into about 2500 sections that are imaged using 3D-PLI microscopy, which results in about 1.3 µm pixel size and 80,000 times 100,000 pixels per section. The polarisation filters used in the imaging process are rotated 18 times so that 18 different images per section are available, which leads to about 750 GB storage consumption per section. The HPC system JURECA is used to run the complex workflow (see Figure 3) that extracts the nerve fibre orientations from the raw data. To speed up the workflow processing, it has already been ported to use the UNICORE workflow engine, which allows analysing a section in the order of hours instead of weeks. The goal of this use case is to integrate the existing UNICORE workflow into the infrastructure of the HBP.

2.2.6.2 Problems to be solved

The following problems require special attention and need to be solved in co-design with the users for implementing the use case successfully.

- The UNICORE workflow engine needs to be well integrated with the other parts of the HPAC Platform to enable a seamless usage inside the HBP environment.
- Provenance tracking of the entire, complex, UNICORE-driven workflow is needed in order to enable reproducibility.
- Establishing access mechanisms to the HPC storage for non-HPC users to enable sharing of results.

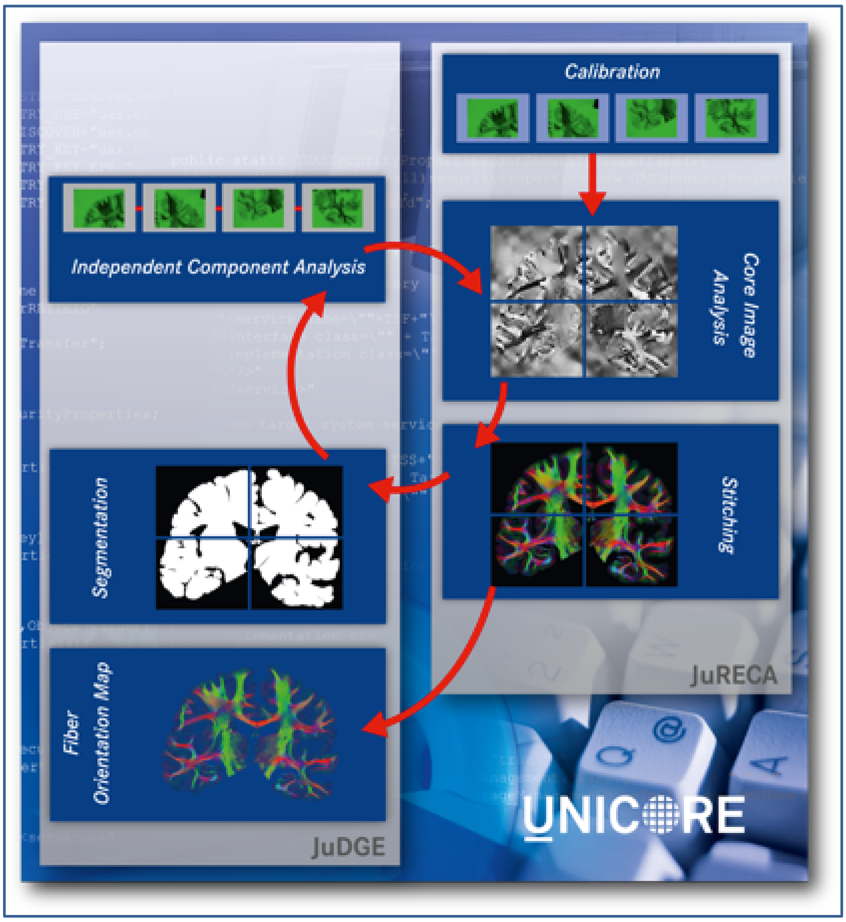

Figure 3: UNICORE-based Polarized Light Imaging workflow

2.3 HPAC Platform: Functional Requirements

Here we list the functional requirements to the HPAC Platform that follow from the use cases descibed in the previous section.

2.3.1 Computing capabilities

Requirement ID: SP7-FR-001

2.3.1.1 High Performance Computing capabilities

- The HPAC Platform should provide production-quality HPC capabilities for
  - Brain simulations
  - Molecular dynamics (subcellular) simulations
  - Massive data analytics
  - Image processing
  - Code development, testing and optimisation

2.3.1.2 General computing capabilities

- The Platform should provide servers to configure virtual machines for the users that they can configure themselves.
- The virtual machines should be configurable to have access to the storage infrastructure of the Platform.

2.3.2 Memory and storage capabilities

Requirement ID: SP7-FR-002

- The HPAC Platform provides traditional file system repositories and Cloud-based storage for retrieving, processing and managing HBP data.
- The storage for various data files should be accessible from the Collaboratory.
- The storage should be suited to accommodate
  - Low-density, i.e., large numbers of small files
  - High-density, i.e., smaller numbers of large files
- The HPAC Platform should provide two types of repositories:
  - Active data repositories for storing temporary slave replica of large data objects with master copies being kept in archival repositories. These repositories are close to computational or visualisation resources and their availability is similar to the attached computational or visualisation resources. If data gets lost, it needs to be replicated again (no data loss, but performance impact).
  - Archive data repositories for storing large data objects permanently. The stores are optimised for capacity, reliability and availability (tier-1 level availability). Data is not replicated.
- The HPAC Platform should provide software tools for leveraging future hierarchical storage architectures for data analytics workflows.

2.3.3 Visualisation capabilities

Requirement ID: SP7-FR-003

- High-resolution, 3D and fully-immersive visualisation capabilities.

- The visualisation capabilities should have access to the storage capabilities in order to enable visualising datasets stored in the HPAC Platform infrastructure.

2.3.4 Network

Requirement ID: SP7-FR-004

- The computing, storage and visualisation capabilities of the HPAC Platform should be connected with a fast, secure network.
- The users of the HPAC Platform should have access to the computing and storage capabilities using a standard internet connection.

2.3.5 Middleware

Requirement ID: SP7-FR-005

2.3.5.1 Middleware configuration

- The computing, storage and visualisation capabilities of the HPAC Platform should be logically connected by a middleware installation.
- The middleware should use the authentication and authorisation services provided by the HPAC Platform.

2.3.5.2 API

The middleware should provide a REST API that allows

- Managing HPC jobs, and
- Uploading, downloading and transferring data between the integrated sites and systems.

2.3.5.3 Portal

The middleware should provide a website or portal that provides the same functionality as the REST API, and that is integrated in or accessible from the Collaboratory.

2.3.6 Workflow management and provenance tracking

Requirement ID: SP7-FR-006

2.3.6.1 Workflow engine

The middleware should provide a workflow engine that enables the users to configure, start and monitor their workflows.

2.3.6.2 Workflow provenance tracking

The middleware should provide mechanism for provenance tracking of entire workflows.

2.3.7 Authentication and authorisation

Requirement ID: SP7-FR-007

- The HPAC Platform should provide a single sign-on mechanism to all infrastructure components so that the same account can be used for accessing all services for which the user has the required access permissions.
- Users must be uniquely identifiable and have explicit access permissions for all services of the Platform.
- Access permissions for all resources of the HPAC Platform must be granted to the user separately, based on the sites' policies and procedures.
- The HPAC Platform user account should be linked to the user's central HBP account.

- User management and access to individual Platform resources should be governed by a common, HBP-wide access policy and supplementary, site-specific policies where necessary.
- The tasks executed by the end users will in some cases interact with authenticated services operated by other parts of the HBP. This will need to address the possibility of authenticated session timeouts and the possible unavailability of authentication tokens in various tiers of the HPAC Platform job service.
- A Computer Emergency Response Team (CERT) (see Section 2.4.5) should be in place to revoke a user's authentication token in case it was compromised.

2.3.8 Resource accounting

Requirement ID: SP7-FR-008

- The HPAC Platform should allow the management of resources based on project specific quotas and data permissions. Quotas should be per group and/or per user and target the following resources:
  - Computing hours and memory usage
  - Data storage
  - Network transfers
- By default, all users have no quota for HPC resources and large data repositories.
- Access to the HPC resources and repositories is provided subject to the successful submission of a peer-reviewed proposal by an HBP PI to a European or national call available for the respective system (see also Section 2.7).

2.3.9 Access and integration

Requirement ID: SP7-FR-009

- The HPAC Platform should provide APIs to allow services developed in other parts of the HBP to
  - Interface with the middleware
  - Access the storage
  - Submit and manage HPC jobs
- All main services of the HPAC Platform should be integrated with and available from the Collaboratory as the central access point to the HBP infrastructure:
  - Job submission and management
  - Storage browser and permission management
  - Monitoring
  - Support and documentation

2.3.10 HPC job management

Requirement ID: SP7-FR-010

2.3.10.1 Submission of HPC jobs from Collaboratory apps and other services

- The HPAC Platform should allow to submit and manage HPC jobs from Collaboratory apps based on the user's HPAC Platform account and system-specific quota.
- The HPAC Platform should provide an API that enables other services to submit and manage jobs based on the user's HPAC Platform account and quota.

2.3.10.2 Command-line based access

The HPAC Platform should allow interactive, command-line based access for Developers.

2.3.10.3 Automatic dispatch of HPC jobs to the best-suited HPC resource

- A list of all available HPC resources and services and their technical characteristics should be available in the Collaboratory.
- A service should make use of execution estimation services in the HPAC Platform.
- Jobs submitted to the HPAC Platform should contain information on suitable HPC resources.
- The HPAC Platform should provide meta-scheduling capabilities such as cross-site synchronised scheduling and shortest time-to-execution scheduling.
- The HPAC Platform should provide a data staging mechanism to move data, if necessary, to the location of the best-suited HPC resource before job execution.

2.3.11 Data and storage management

Requirement ID: SP7-FR-011

2.3.11.1 Metadata service

The HPAC Platform should provide a metadata service. Functionality for marking data as hidden to other users is required.

2.3.11.2 Data transfer and replication services

The HPAC Platform should provide services for

- Uploading data to a repository.
- Downloading data from a repository.
- Initiating the reliable transfer of a dataset between the repositories at different sites.
- Seamless replication of linked data ("link upgrade") across sites when required.
- Staging of linked data, i.e. the automatic migration of data to the HPC resources where it will be needed.

2.3.11.3 Access and permission management

- The HPAC Platform should provide a regularly updated list of potential destinations for output data.
- The HPAC Platform should provide a mechanism to generate links to the datasets that can be shared with users, who do not have access to the data repository themselves, so that they get access to the dataset.
- Granular permission granting scenarios should be supported, and tools should be available for managing the permissions.
- Ability to explicitly show or hide data files to other users/groups.
- Ability to explicitly grant read/write/delete permissions to users/groups, for individual files and/or folders.
- Ability for data administrators to prevent certain data from being modified/deleted.
- Authentication and authorisation for access to data is required. It should be integrated with the central HBP authentication and authorisation services, and should support appropriate delegation schemes.

2.3.11.4 Storage and quota browser

- The HPAC Platform should provide a storage browser that allows to search for datasets and where they are stored.
- The HPAC Platform should provide a service or tool to query the user's or group's storage quota per repository or site.

2.3.12 Resource monitoring

Requirement ID: SP7-FR-012

- Resources available to the HPAC Platform should be registered in the central registry.
- Each site should monitor the usage of compute, network and storage resources and provide this information in a commonly agreed format.
- The HPAC Platform should provide a central monitoring service that makes the monitoring information available for users and system administrators (INFRA). There might be separate services for users and system administrators that provide the information at different levels of detail as appropriate.
- The HPAC Platform should allow the service to report on health checks performed on various HPC resources. Health checks should report at least the following:
  - Resource availability
  - Resource degradation/failure
  - Upcoming scheduled downtime
- The HPAC Platform should provide a central monitoring of the most important services that are required for the Platform to be operational.

2.3.13 Software management and deployment

Requirement ID: SP7-FR-013

2.3.13.1 Software deployment

- The HPAC Platform should provide mechanisms for the deployment of software on the HPC systems.
- The installations of the same software on different systems should always deliver the same results.
- Users should be able to install updates for their already existing software.

2.3.13.2 Container-based software packaging and deployment

The HPAC Platform should provide container-based software packaging and deployment mechanisms to deploy software on the HPC systems. The additional layer of abstraction introduced by using container-based deployment of software instead of traditional software deployment leads to a higher flexibility with respect to the architecture used for executing the software. If containers are used for providing services, these services are scalable with respect to the number of users that can be served since the service can easily replicated.

2.3.14 Development environment

Requirement ID: SP7-FR-014

2.3.14.1 Interactive access

The Platform should allow (command-line based) interactive access for selected HBP developers

2.3.14.2 Development environment

The Platform should provide a full software development environment, including but not limited to:

- Parallel programming environment
- Compilers
- Debuggers
- Performance analysis tools

2.3.14.3 Dynamic resource management

The HPAC Platform should provide mechanisms to run a brain simulation and an analysis or visualisation concurrently in a way that the available resources are efficiently shared between the applications.

2.3.15 Simulation technology

Requirement ID: SP7-FR-015

2.3.15.1 Scalable simulation technology

The HPAC Platform should provide scalable brain simulation codes that run and are optimised for the available HPC systems.

2.3.15.2 Coupling simulation to analysis and visualisation

- The HPAC Platform should provide mechanisms for coupling a running brain simulation to an analysis or visualisation.
- The HPAC Platform should provide mechanisms for streaming and steering of brain simulations.

2.3.16 Interactive visualisation

Requirement ID: SP7-FR-016

The HPAC Platform should provide interactive visualisation tools enabling an *in situ* visual analysis of simulation data.

2.3.17 Documentation and user training

Requirement ID: SP7-FR-017

Documentation and training should help users with different backgrounds and experience to make effective use of the HPAC Platform.

- The HPAC Platform should be documented in such a way that new users are up and able to execute simple workflows in less than two hours after getting access to compute resources.
- The documentation should be available online and from the Collaboratory.
- Introductory and advanced, HPC-related topics, e.g. for Scientific Developers, should be the subject of suitable training courses and workshops.

2.3.18 Support

Requirement ID: SP7-FR-018

2.3.18.1 User support team

- The HPAC Platform should establish a user support team with at least one member of each HPC site.

- The user support should set up a functional email address that serves as the main way to contact the support team.
- The HPAC Platform should support its users in using the provided software tools.

2.3.18.2 Ticket system

- The HPAC Platform should setup a trouble ticket system.
- Tickets in this ticket system should be opened by sending an email to the user support email address.

2.3.18.3 Support for software development

- The HPAC Platform should develop and make available a guide with best practices in software development for projects using HPC systems.
- The HPAC Platform should support its users, mainly Developers and Scientific Developers, in establishing these best practices in their projects.

2.3.18.4 Support for parallelisation and code migration

- The HPAC Platform should support its users in parallelising their applications.
- The HPAC Platform should support its users in migrating their codes to new HPC architectures.

2.3.18.5 Support for resource allocation applications

The HPAC Platform should support its future users in applying for HPC allocations:

- Choosing the most appropriate HPC architecture and system.
- Choosing a call for computing time applications for this system.
- Getting access to the HPC system for and preparation of the application.

2.3.19 Mapping of Functional Requirements and Use Cases

Table 1 below shows the relationship between functional requirements and use cases. The grey fields indicate for each functional requirement the use case(s) it is intended to address.

	SP7-FR-001	SP7-FR-002	SP7-FR-003	SP7-FR-004	SP7-FR-005	SP7-FR-006	SP7-FR-007	SP7-FR-008	SP7-FR-009	SP7-FR-010	SP7-FR-011	SP7-FR-012	SP7-FR-013	SP7-FR-014	SP7-FR-015	SP7-FR-016	SP7-FR-017	SP7-FR-018
SP7-UC-001	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SP7-UC-002	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SP7-UC-003	■	■		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SP7-UC-004	■	■		■	■	■	■	■	■	■	■	■	■	■			■	■
SP7-UC-005	■	■		■	■	■	■	■	■	■	■	■	■	■			■	■
SP7-UC-006	■	■		■	■	■	■	■	■	■	■	■	■	■			■	■

Table 1: Relationship between Functional Requirements and Use Cases

2.4 HPAC Platform: Non-Functional Requirements

In this section we list non-functional requirements to the HPC Platform as a whole.

2.4.1 Scalability and flexibility

Requirement ID: SP7-NFR-001

The HPAC Platform should be constructed such that it has, in principle, the flexibility to grow and accommodate current and future quantitative requirements. Any upgrade of the Platform will of course be constrained by technological feasibility and available budget. This in in particular needed since it is very challenging to gather the requirements from potential, future users outside the HBP. In detail, the Platform should be scalable with respect to

- Expected number of concurrent users
- Expected resource consumption per user
- Data volumes and transfer speeds

2.4.2 Interoperability between sites

Requirement ID: SP7-NFR-002

The HPAC Platform should provide the necessary tools and infrastructure to allow an efficient cooperation of the participating sites and to increase the scientists' productivity. In particular:

- Common user environments should be deployed on the various HPC sites.
- Software should be deployed in a consistent manner on the various HPC sites.
- Resolutions to problems identified at one site should be communicated appropriately to the other HPC sites for consistent system improvement.

2.4.3 Performance

Requirement ID: SP7-NFR-003

- It should be stressed that optimising HPAC Platform performance is not simply about maximising compute performance in terms of Flops, but improving the overall time-to-solution, taking into account all hardware and software components.
- All the hardware and software components of the HPAC Platform should be balanced out to offer fair resource allocation to the various workflows and to avoid bottlenecks.
- Services should be allowed to run with the required Quality of Service to ensure successful completion.

2.4.4 Data protection (backup/retention policies, disaster/recovery)

Requirement ID: SP7-NFR-004

- The HPAC Platform should offer various data protection levels (backups, RAID, data retention policies).
- Users should be able to classify their data with regard to the required protection level (e.g., short term storage vs. long term archiving).
- A disaster recovery strategy should be provided.

2.4.5 IT Security (CERT/CSIRT Team, Security Forum, Security Policies)

Requirement ID: SP7-NFR-005

The HPAC Platform should have an Information Technology Security Policy, comprising of the following elements:

- Each multi-domain infrastructure lacks information about the security level the other partners have implemented. Mutually exchanging security polices or at least providing indicative information on security policies implemented will help to generate a web of trust between the participating sites. Only with this "Web of Trust" close collaboration between these sites can be initiated. The HPAC Platform should establish such trust between sites by exchanging security policies and working together in setting up and handling the infrastructure in a secure manner.
- An HBP-AUP/ToU (Acceptable Use Policy/Terms of Use) should be elaborated, so that the users of the HPAC Platform are aware under which conditions they are allowed to use the infrastructure provided. Users not adhering to these regulations may be sanctioned by revoking their user account and/or access permissions.
- The sites being part of the HPAC Platform should implement an operational IT Security Team (CERT/CSIRT) to handle any security incidents coming up during operation of the Platform. The team should consist of at least one staff member of every site. Optimally, the CERTs of all sites work closely together and exchange any security relevant information.

2.4.6 HPAC Platform validation

Requirement ID: SP7-NFR-006

- The HPAC Platform should be in close, regular contact with its current and potential future users to validate if the functionality of the HPAC Platform fits well to their requirements and to find out if these requirements change over time.
- The HPAC Platform should take the necessary measures to ensure and maintain the usability of the Platform for the HBP and the wider neuroscience community in the best possible way.

2.4.7 Ethics and society

Requirement ID: SP7-NFR-007

- The HPAC Platform should assure that the resources are only used in compliance with applicable, national and international law. This in particular, but not exclusively concerns topics like data protection, dual use and export control.
- The HPAC Platform should assure that only data is processed using the provided infrastructure that was collected in compliance with applicable, national and international law.
- The HPAC Platform should be in regular contact with the Ethics and Society Subproject (SP12) and the Ethics Advisory Board to get their support and advice on the above mentioned topics.

2.4.8 Forward-looking activities

Requirement ID: SP7-NFR-008

The HPAC Platform should investigate new HPC architectures and technologies in general, and their relevance for neuroscience research in order to keep the infrastructure state-of-the-art.

2.5 HPAC Platform: Software and Services

In this section we describe the software components of the HPAC Platform. For each functional requirement of section 2.3, we list the software functions and services used to implement these requirements. It is also indicated at which sites of the HPAC Platform the software is or should be installed.

2.5.1 *Computing capabilities*

Related functional requirement: SP7-FR-001

Most of the software categories described in the rest of this section will be required to meet this requirement.

2.5.2 *Memory and storage capabilities*

Related functional requirement: SP7-FR-002

Most of the software categories described in the rest of this section will be required to meet this requirement.

2.5.3 *Visualisation capabilities*

Related functional requirement: SP7-FR-003

The visualisation sites of the HPAC Platform, EPFL and RWTH, maintain the required drivers and frameworks that are needed to use the visualisation capabilities.

Function	Software	Description	Site
Visualisation frameworks and libraries	Virtual Reality Toolkit for the development of virtual reality applications with a focus on immersive, scientific visualisation.	ViSTA	RWTH
	An open-source software system for 3D computer graphics, image processing and visualisation.	Vtk	RWTH, EPFL
	An open-source scientific workflow and provenance tracking system that supports data exploration and visualisation.	VisTrails	RWTH
Visualisation on tiled displays	Software environment for interactively driving and using large-scale tiled displays	DisplayCluster, Deflect	RWTH, EPFL, user’s computer

Table 2: Visualisation frameworks

2.5.4 Network

Related functional requirement: SP7-FR-004

The HPAC Platform will leverage existing network connections and thus rely on the related software and configurations used by the network providers.

2.5.5 Middleware

Related functional requirement: SP7-FR-005

UNICORE is the middleware used in the HPAC Platform.

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=60

Function	Software	Description	Site
Middleware	UNICORE	Uniform Interface to Computing Resources providing a seamless interface for preparing and submitting jobs to a variety of computing resources.	All sites
User management	UNICORE Registry	Registry server providing information about available HPC services.	At least two sites for high-availability configuration
Web portal	UNICORE Portal	The UNICORE Portal is a generic web interface to the UNICORE grid middleware, providing seamless and secure access to high-performance computing, file systems and other resources. User functions include job submission and management, storage access, data transfer and more. User authentication is integrated with the HBP OIDC server.	JSC
API	UNICORE REST API	The UNICORE REST API provides a RESTful interface to the UNICORE services.	All sites

Table 3: Middleware software

2.5.6 Workflow management and provenance tracking

Related functional requirement: SP7-FR-006

UNICORE is the middleware used in the HPAC Platform that also provides a workflow engine. A provenance tracking mechanism for UNICORE-based workflows is under development that is strongly driven by the demands from the HBP community.

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=60

Function	Software	Description	Site
Workflow engine	UNICORE workflow engine	The UNICORE Workflow engine is used to control the execution of multiple UNICORE jobs at one or multiple execution sites, dealing with dependencies between jobs and handling any required data movement.	All sites
Provenance tracking for workflows	UNICORE provenance tracking	A provenance tracking mechanism for UNICORE-based workflows is under development. It will be based on the PROV standard.	Not deployed yet.

Table 4: Software for workflow management and provenance

2.5.7 Authentication and authorisation

Related functional requirement: SP7-FR-007

Function	Description	Proposed Software	Site
Authentication and authorisation	Global User Administration System (UAS) on top of existing local UASs	Lightweight Directory Access Protocol (LDAP)	All sites
	Service container that supports SOAP Web Services and RESTful services	UNICORE/X	All sites
	UNICORE user database, shareable between Grid nodes, supports multiple logins and projects	XUUDB	All sites
	Network authentication protocol	Kerberos	CSCS
	User authentication for web services	Unity	JSC

Table 5: Software for authentication and authorisation

Traditionally, every HPC system has its own user database and uses its own authentication and authorisation mechanisms. Implementing single sign-on requires that every user is mapped by a general authentication system to the special user ID required for the specific

HPC resource. Since the HPC systems integrated into the HPAC Platform are not exclusively used by the HBP, it cannot be anticipated that the login mechanisms will be changed to fit the HBP single sign-on prerequisites. Therefore, an alternative approach has been chosen.

A User Administration System (UAS) is built on top of existing local user administration systems, which does not make any assumptions about these systems. This makes the UAS as little invasive as possible with a minimum of local adaptations required.

Every site can perform administrative tasks for their own users. Site-specific and/or system-specific policies define which tasks can be performed.

All HBP sites have read access to the UAS of other sites, so that, if users are added or attributes are changed, other sites are able to verify this and make corresponding changes. The UAS will be implemented in a secure way, so that unauthorised persons cannot tamper with the information available. It uses a uniform namespace, with a strictly defined set of allowed attributes and values. This way all information in the UAS is stored in exactly the same way at each site.

To provide this kind of UAS, the Lightweight Directory Access Protocol (LDAP) has been selected for the implementation. LDAP is widely used and accepted as a user administration system and directory service in general. LDAP services can also be adapted quite easily to interface with different information systems (back-end databases). The LDAP protocol is used for communication, but the repository itself can be almost any system of choice.

The infrastructure enabling this mechanism in the HPAC Platform consists of a master LDAP server and some slave servers. The master HBP LDAP server is hosted by one of the sites in the federation. Setting up a backup master server at another site to minimise downtimes of the authentication and authorisation services is possible and being considered. Slave LDAP servers are located at the sites providing hardware resources for the HPAC Platform. New user accounts are first created on the master LDAP server that then feeds this information into the slave servers. Also all updates of user accounts or groups are first applied at the master server and then propagated to the slave servers. For updating the local user administration databases, all sites developed tools and procedures to incorporate newly created and modified information of the LDAP servers into their local user administration and authorisation facilities.

The central HBP user accounts are administered using an OIDC server. The HBP HPC user accounts are mapped onto the user’s central HBP account to enable single sign-on also from services provided by other parts of the HBP, that use the OIDC server for verifying the user’s identity, e.g. for services running in the Collaboratory framework.

For using services based on or through UNICORE, the user’s identity is first verified by the OIDC server using the HBP username and password. The OIDC server returns an OIDC token in case of a successful authentication. This OIDC token is then used to access the UNICORE services. UNICORE passes this token to the Unity server at JSC that validates the token by contacting the OIDC server. In case of a successful validation, Unity returns the user’s distinguished name (DN) back to UNICORE. All information in Unity is transient and created on the fly when a user logs in or an OIDC token is validated.

The UNICORE user database XUUDB maps the DN to the user IDs (uid) and group IDs (gid) of the HPC systems. The slave LDAP servers at the HPC sites periodically update the user information in the XUUDB, i.e. the mapping of the DN to uid and gid. These user records are also synchronized with the master LDAP server.

The whole process happens in an automated way that is completely transparent to the user. The HBP HPC uid, that a user receives when the first HPC allocation is granted, will be reserved for this user also at the other sites, so that a user with multiple allocations has the same uid at all sites.

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=66

2.5.8 Resource accounting

Related functional requirement: SP7-FR-008

The resource accounting is handled locally at the HPC sites since the HPC systems integrated into the HPAC Platform are not exclusively used by the HBP and therefore a central accounting mechanism cannot be put in place. Quota are granted per user and/or per group. Granting significant amounts of HPC resources is subject to scientifically and technically, independently peer-reviewed applications to national or international calls for computing time proposals. The calls available for the systems in the HPAC Platform are available in the online documentation.

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=68

2.5.9 Access and integration

Related functional requirement: SP7-FR-009

The API for accessing the HPAC Platform is the UNICORE REST API described in section 2.5.5. HPC systems and storage can also be accessed using the UNICORE Portal or from a workflow defined in the UNICORE workflow engine (see section 2.5.6). Command-line access to the HPC systems is possible as well.

The integration with the Collaboratory as the central access point to the HBP infrastructure is realised as follows: As described in section 2.5.7, the HBP HPC accounts are linked with the HBP accounts, which is necessary for a seamless integration of the HPAC Platform into the Collaboratory. The UNICORE Portal is integrated into the HPAC Platform Collab (https://collab.humanbrainproject.eu/#/collab/264/nav/4306). It is also possible to use the UNICORE REST API from IPython/Jupyter notebooks in the Collaboratory. The following software is under development for enabling the additionally required functionality.

Function	Description	Proposed Software	Site
Storage browser	An HPC storage browser enables the user to see all datasets stored in one of the integrated repositories for which the user/the group has at least read permissions.	Collaboratory app is under development. A beta version is deployed.	Storage at all sites is visible.

Table 6: Software and tools for Platform access and integration

2.5.10 HPC job management

Related functional requirement: SP7-FR-010

The submission of jobs from Collaboratory apps and other services is possible via the UNICORE Portal or using the UNICORE REST API (see section 2.5.5).

Command-line based access for submitting and managing HPC jobs is available for all HPC systems integrated into the HPAC Platform.

The management of the computational resources provided by the HBP HPC machines is handled by job scheduler software (see Table 7: Software for HPC job management) to maximise the overall productivity. The scheduler is responsible for managing jobs (program execution) on the machines by allocating partitions for the user on the compute nodes, and returning job output and error files to the users. Because of site-specific, local, technical or contractual requirements, each site continues to use its own job scheduling system. However, through the adoption of a middleware (see sections 2.5.5 and 2.5.6), users have a single entry point for the submission of computational jobs, and they can use the HPC systems without entering the complexity underneath. Complementing job submission, the middleware provides data staging functionality, both for data residing on the HPAC Platform (i.e. other HPC sites or the Cloud storage), as well as storage that is external to it.

Function	Description	Proposed Software	Site
HPC Job management	Job scheduling and management (site-specific)	LoadLeveler - IBM parallel job scheduling system	Cineca
		PBS - Portable Batch System: job scheduling computer software	Cineca
		Slurm - Open-source resource manager	BSC (MinoTauro), CSCS, JSC
		LSF - Platform Load Sharing Facility: commercial job scheduler	BSC (MareNostrum)

Table 7: Software for HPC job management

2.5.11 Data and storage management

Related functional requirement: SP7-FR-011

Function	Description	Proposed Software	Site
Data transfer and replication	Data streaming library and file transfer tool	uFTP	All sites
	High-speed, reliable, and secure data transfer protocol	GridFTP	All sites
	Distributed parallel file system: provides online storage management, scalable access, and integrated information lifecycle management tools capable of managing petabytes of data and billions of files	GPFS	All sites
	Remote file caching technology that ensures seamless and high performance access to files no matter geographical location	GPFS/AFM	The adoption of the AFM technology will be evaluated during the course of the project.
	Integrated Rule-Oriented Data-management System, a community-driven, open source, data grid software solution	iRODS	The adoption of the iRODS technology will be evaluated during the course of the project.

	System for storing and retrieving huge amounts of data, distributed among a large number of heterogeneous server nodes, under a single virtual file system tree with a variety of standard access methods	dCache	The adoption of the dCache technology will be evaluated during the course of the project.
	Scalable I/O library for parallel access to task-local files	SIONlib	All sites
Cloud Storage	Amazon-S3 compatible interface	CEPH	KIT
Access permission management	A tool enabling the users to organise themselves as "Virtual Organisations" (VO), i.e. to define groups and manage the permissions of these groups.	Development of this tool is planned. Existing solutions to be tested are Perun, Unity and Keystone	All sites
Storage and quota browser	*See section 2.5.9*		

Table 8: Software for data and storage management

Transfer protocols like GridFTP, UNICORE uFTP are considered as options to enable file transfers between the GPFS file systems at the different HPC centres. This is a step towards establishing a distributed file system.

AFM is a powerful remote file caching technology that ensures high performance access to file based information no matter where it might be geographically located. This feature is useful for the HBP since large volumes of file based data are shared and require high levels of information availability no matter where it resides. The latest AFM version will be tested as access method to distribute HBP data over the different HPC file systems.

The Cloud storage located at KIT is accessible through an Amazon S3-compatible interface. It is used as a repository for secondary data: the user will be able to upload/download data to/from the cloud storage through a standard Internet connection.

Scientific user(s) can preserve and share data in the Cloud storage using a persistent URL. This feature is essential for use cases where persistence of data localisation and content is required.

The Cloud storage is based on Ceph, a highly scalable and performing distributed storage system. Ceph provides a RESTful interface, allowing a range of basic operations such as uploading, downloading and deleting objects, as well as sharing data with HTTP URLs.

Ceph uses replication to ensure data reliability and recovery. It offers a block device interface, which can be used as back-end for a Cloud computing system like OpenStack. The users can have access to on-demand computing resources by launching virtual machines at the KIT Cloud when needed.

2.5.12 Resource monitoring

Related functional requirement: SP7-FR-012

Function	Description	Proposed Software	Site

Monitoring	Monitoring tool to detect infrastructure problems by executing periodic, automated, user-level testing of software and services	Based on Nagios/INCINGA	All sites
	Service container that supports SOAP Web Services and RESTful services	UNICORE/X	All sites
	Graphical monitoring of batch system controlled clusters	LLview	All sites
	Collaboratory app testing and showing the availability of all infrastructure components known to UNICORE	Own development, available from the HPAC Platform Collab	All sites

Table 9: Software for resource monitoring

Monitoring in the HPAC Platform is based on UNICORE/X, a service container supporting SOAP Web Services and RESTful services, and ICINGA. All sites operate UNICORE/X instances for making monitoring information available to a central monitoring service. CSCS, JSC and BSC use Nagios for the system monitoring; CINECA and KIT have alternate solutions in place. CSCS and JSC in addition have ICINGA installations in place.

A Collaboratory app is available that shows the availability of all infrastructure components.

2.5.13 Software management and deployment

Related functional requirement: SP7-FR-013

Function	Description	Proposed Software	Site
Software deployment	Dynamic modification of a user's environment	Environment Module packages	All sites
	Software repository	Git, apt, yum	All sites
Software packaging	Automated deployment of software inside software containers	Docker	All sites

Table 10: Software management and deployment tools

The availability of scientific software for the users will be managed by the Environment Modules commands. On a complex HPC system, where it is necessary to make available a wide choice of software packages in multiple versions, it can be difficult to set up the user environment to allow users to always find the required executables and libraries. The Modules package provides for the dynamic modification of a user's environment via *module files*: each *module file* contains the information needed to configure the shell for an application.

Examples of software distribution mechanisms are git, apt or yum. Mechanisms for container-based software packaging and deployment, e.g. with Docker, are being investigated.

2.5.14 Development environment

Related functional requirement: SP7-FR-014

Function	Description	Proposed Software	Site
Programming	General purpose, high-level programming language	Python	All sites
	Python modules	NumPy, SciPy	All sites
	Fortran77 and Fortran90 compilers	ifort, gfortran	All sites
	C and C++ compilers	Icc, icpc, gcc, g++	All sites
	Blue Gene/Q Cross-compiler	bgq-xl	JSC
Parallel programming	Standardised and portable message-passing system	Message-Passing Interface (MPI) library	All sites
	API supporting multi-platform shared-memory parallel programming	OpenMP	All sites
	Framework for programs that are executable across heterogeneous platforms, e.g. CPUs and GPUs	OpenCL	All sites
	APIs and frameworks for using GPUs (GPGPU)	CUDA, OpenACC	BSC, CSCS, JSC
	Extension of OpenMP with new directives to support asynchronous parallelism and heterogeneity	OmpSs	All sites
	Python bindings for OmpSs	PyCOMPs	All sites
Debugging	GUI-based source code defect analysis tool to control over processes and thread execution and visibility into program state and variables	TotalView	All sites
	Profiling tool	gprof	All sites
	Tool to detect memory management and threading bugs, code profiler	Valgrind	All sites
Profiling and performance analysis	Highly-Scalable call-path profiling and trace analysis	Score-P, Scalasca	All sites
	Sophisticated trace measurement and visualisation	Extrae, Paraver	All sites

Collaborative code development	Free and open source distributed version control system	Git	All sites
	Open-source, centralised version control system	Subversion	All sites
Dynamic resource management	Dynamic Load Balancing (DLB) and resource management	DLB and resource management libraries	All sites
Libraries and tools	Basic numerical libraries	BLAS, LAPACK, FFT, GSL, GMP, SPRNG	All sites
	Parallel numerical libraries	ScaLAPACK, ARPACK, PARPACK, PETSc	All sites
	Library for Accelerated Math Applications, framework for building efficient, extensible and flexible solvers for sparse linear systems and in application domains that involve - sparse and dense - numerical linear algebra. Supports shared and distributed memory compute architectures, including accelerators, through an easy to use interface.	LAMA	All sites
	Multi-GPU, Distributed Memory 3D Fast Fourier Transform	FFT	All sites
	Hyperbolic Cross Fast Fourier Transform, a high dimensional pseudospectral methods sparse grid based library (serial version)	HCFFT	All sites
	A tool for fast automatic 3D-stitching of teravoxel-sized microscopy images	Terastitcher	Cineca

Table 11: Software and tools for developers

In order to provide Developers with a complete environment for the development of their software, a set of tools and libraries are available on the machines:

- Compilers for the most used programming languages
- Debuggers, profilers and performance analysis tools for testing, correction and improvement of code
- Tools for collaborative code development

2.5.15 Simulation technology

Related functional requirement: SP7-FR-015

Function	Description	Proposed Software	Site

Simulation	Spiking neural network simulator	NEST	All sites
	Empirically-based simulator of neurons and networks of neurons	NEURON	All sites
	Exact stochastic simulator of reaction-diffusion systems in arbitrarily complex 3D geometries	STEPS	All sites
	Molecular dynamics simulator for simulations of proteins, lipids and nucleic acids	GROMACS	All sites
	Molecular dynamics simulator	Tremolo-X	All sites
	Flexible numerical software environment for solving partial differential equations on huge clusters	UG 4	All sites
	High-performance, multi-compartment neural network simulator	NestMC (under development)	All sites
Streaming and steering	Software enabling the coupling of simulation and visualisation	Monsteer, ZeroEQ and ZeroBuf	All sites

Table 12: Software and tools related to simulation technology

The HPAC Platform supports the development and optimisation of brain simulation codes and their deployment on the HPC systems. This simulation technology will be enhanced with new features and optimisations in the course of the HBP.

The simulators need be linkable to analysis or visualisations at runtime, e.g. to Livre or RTNeuron. The HPAC Platform develops mechanisms enabling streaming and steering of simulations, i.e. Monsteer, ZeroEQ and ZeroBuf.

2.5.16 Interactive visualisation

Related functional requirement: SP7-FR-016

Function	Description	Proposed Software	Site
Interactive visualisation tools and libraries	Visualisation of structural plasticity	MSPViz	RWTH, EPFL, user's computer
	Visualisation of 2d-layered data, e.g. spike data resulting from a NEST simulation	VIOLA	RWTH, EPFL, user's computer
	Search and filtering operations using both data contents and metadata	neuroFiReS	RWTH, EPFL, user's computer

	Navigation through circuit data at different levels of abstraction	NeuroScheme	RWTH, EPFL, user's computer
	Creating neuronal meshes from a minimal skeletal description	NeuroLOTs	RWTH, EPFL, user's computer
	Visualisation of simulation results	VisNEST, RTNeuron	RWTH, EPFL, user's computer
	Parallel rendering framework to create and deploy parallel, scalable OpenGL applications	Equalizer	RWTH, EPFL, user's computer
	Large-scale Interactive Volume Rendering Engine	Livre	RWTH, EPFL, user's computer
	Handy, fast, and versatile 3D/4D/5D Image Visualisation & Analysis System for Bioimages & Surface Objects	Vaa3D	Cineca
Provenance	Creation of provenance tracks in context of interactive analysis tools and visualization applications	InDiProv	RWTH, EPFL, user's computer

The HPAC Platform develops and provides a variety of visualisation tools.

2.5.17 Documentation and user training

Related functional requirement: SP7-FR-017

The infrastructure, software and services of the HPAC Platform are documented for users and developers in the High Performance Analytics and Computing Platform Guidebook that is available as a public website (https://hbp-hpc-platform.fz-juelich.de/). This Guidebook is also integrated into the HPAC Platform Collab in the Collaboratory.

The HPAC Platform user training consists of workshops, trainings, courses and schools organised by the SP7 partners, which includes the PRACE-PATC trainings, that are offered for all users of the HPC systems. Topics covered are e.g. programming languages, programming methods and a variety of HPC-related subjects, both on introductory and advanced level. These trainings are complemented with special courses and workshops that are organised in addition for the HBP and wider neuroscience society. The training programme is available as part of the Guidebook (https://hbp-hpc-platform.fz-juelich.de/?page_id=583).

The documentation and user training of the HPAC Platform have also been described in the Ramp-up Phase Deliverable D7.7.6.

2.5.18 Support

Related functional requirement: SP7-FR-018

Function	Description	Proposed Software	Site
Ticket system	A ticket system to collect and handle all requests and bug reports related to SP7 and the HPAC Platform	Request Tracker (RT)	Installed at BSC, all sites contribute to answering the tickets

All user requests, questions and bug reports can be sent to the functional email address hpac-support@humanbrainproject.eu. If an email is sent to this email address, a ticket in the ticket system will be opened automatically. The member of the support team on duty assigns this new ticket to the respective site- or system-specific queue. The tickets will be answered by a staff member of one of the HPC centres, or by the SP7 management and coordination team, depending on the topic.

2.6 HPAC Platform: Physical Architecture

In this section we describe the hardware components of the HPAC Platform.

2.6.1 Computing capabilities

Related functional requirement: SP7-FR-001

2.6.1.1 High Performance Computing Capabilities

The high performance computing capabilities are provided by BSC, CINECA, CSCS and JSC. The first version of the HPC Platform, as it was called at that time, comprised the following HPC systems:

- HBP Supercomputer JUQUEEN at JUELICH
- HBP Development System BlueBrain IV hosted by CSCS
- HBP Molecular Dynamics Supercomputer MareNostrum at BSC
- HBP Massive Data Analytics Supercomputer FERMI at CINECA

During the Ramp-up Phase it turned out that data analytics is a second driving area in the HBP in addition to brain simulations that were initially expected to be the main, neuroscientific user community of the HPC systems. Therefore, it has been decided in the last months of the Ramp-up Phase to integrate additional HPC systems into the HPAC Platform that are also operated by the same four HPC centres. The architectures of these systems are well suited for data analytics and image processing applications.

BlueBrain IV has been replaced in the HPAC Platform by Piz Daint at CSCS at the end of the Ramp-up Phase. The GPU cluster JURECA at JUELICH and the cluster Pico at CINECA have been integrated into the HPAC Platform at the same time. FERMI reached the end of its operational phase in July 2016 and has been replaced by CINECA's new system Marconi that is now integrated into the HPAC Platform instead.

Due to the integration of additional systems it has been decided to drop the "HBP labels" that were until then used for the initial four supercomputers, i.e. "HBP Supercomputer" for JUQUEEN, "HBP Development System" for BlueBrain IV, "HBP Molecular Dynamics Supercomputer" for MareNostrum III and "HBP Massive Data Analytics Supercomputer" for FERMI.

In the following, the currently integrated six HPC systems and their architecture are briefly described.

2.6.1.1.1 JUQUEEN

JUQUEEN is located at the Jülich Supercomputing Centre (JSC) of Forschungszentrum Jülich.

Architecture	IBM Blue Gene/Q system
Peak performance	5.9 PFlops
System configuration	28 racks (7 rows à 4 racks) – 28,672 nodes (458,752 cores) Rack: 2 midplanes à 16 nodeboards (16,384 cores) Nodeboard: 32 compute nodes Node: 16 cores
Processors	IBM PowerPC® A2, 1.6 GHz, 16 cores per node
Memory	16 GB SDRAM-DDR3 per node

Main memory	448 TB
Networks	5D Torus — 40 GB/s; 2.5 μsec latency (worst case) Collective network — part of the 5D Torus; collective logic operations supported Global Barrier/Interrupt — part of 5D Torus, PCIe x8 Gen2 based I/O 1 GB Control Network — System Boot, Debug, Monitoring

Table 13: JUQUEEN specification

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=34

2.6.1.1.2 JURECA

JURECA is located at JSC, Forschungszentrum Jülich.

Architecture	T-Platforms V-Class
Peak performance	1.8 (CPU) + 0.44 (GPU) Petaflops
System configuration	1872 compute nodes and 12 visualisation nodes with 45,216 CPU cores in total

Processors	1872 compute nodes
	• Two Intel Xeon E5-2680 v3 Haswell CPUs per node
	  ○ 2 x 12 cores, 2.5 GHz
	  ○ Intel Hyperthreading Technology (Simultaneous Multithreading)
	  ○ AVX 2.0 ISA extension
	• 75 compute nodes equipped with two NVIDIA K80 GPUs (four visible devices per node)
	  ○ 2 x 4992 CUDA cores
	  ○ 2 x 24 GiB GDDR5 memory
	12 visualisation nodes
	• Two Intel Xeon E5-2680 v3 Haswell CPUs per node
	• Two NVIDIA K40 GPUs per node

Memory	Compute nodes: DDR4 memory technology (2133 MHz) • 1605 compute nodes with 128 GB memory • 128 compute nodes with 256 GB memory • 64 compute nodes with 512 GB memory Visualisation nodes: • NVIDIA K40 GPUs: 2 x 12 GB GDDR5 memory • 10 nodes with 512 GB memory • 2 nodes with 1024 GB memory
Networks	Mellanox EDR InfiniBand with non-blocking fat tree topology 100 GiB/s

Table 14: JURECA specification

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=933

2.6.1.1.3 Piz Daint

Piz Daint is located at Centro Svizzero di Calcolo Scientifico (CSCS) – Swiss National Supercomputing Centre, ETH Zürich, Lugano, Switzerland.

Architecture	Cray XC30
Peak performance	7.787 PFlops (theoretical)
System configuration	28 cabinets with 5,272 nodes and 42,176 cores in total
Processors	Compute nodes: 8-core 64-bit Intel SandyBridge CPU (Intel Xeon E5-2670) NVIDIA Tesla K20X with 6 GB GDDR5 memory
Memory & storage	Capacity/node: 32 GB (DDR3-1600) 6 GB non-ECC (GDDR5) Bandwidth/node: 51.2 GB/s DDR3 250.0 GB/s non-ECC GDDR5 Total system memory: 169 TB DDR3 32 TB non-ECC GDDR5 System storage capacity: 2.5 PB Parallel file system peak performance 117 GiB/s
Main memory	32GB host memory on each compute node
Networks	The nodes are connected by the “Aries” proprietary interconnect from Cray, with a dragonfly network topology Peak network bisection bandwidth: 33 TB/s

Table 15: Piz Daint specification

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=940

2.6.1.1.4 MareNostrum III

MareNostrum III is located at Barcelona Supercomputing Centre (BSC), Spain.

Architecture	IBM iDataPlex system
Peak performance	1.1 PFlops
System configuration	52 racks 3,056 main nodes 42 heterogeneous compute nodes
Processors	Main nodes (3,056 nodes): • 2x Intel SandyBridge-EP E5-2670/1600 20M 8-core at 2.6 GHz Xeon Phi nodes (42 nodes): • 2x Intel SandyBridge-EP E5-2670/1600 20M 8-core at 2.6 GHz • 2x Xeon Phi 5110 P
Memory	Main nodes:

	• 64 nodes with 8 x 16 GB DDR3-1600 DIMMS (8 GB/core) • 64 nodes with 16 x 4 GB DDR3-1600 DIMMS (4 GB/core) • 2880 nodes with 8 x 4 GB DDR3-1600 DIMMS (2 GB/core) Xeon Phi nodes: • 8 x 8 GB DDR3-1600 DIMMS (4 GB/core)
Main memory	2 PB
Networks	Infiniband FDR10 Gigabit Ethernet

Table 16: MareNostrum III specification

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=47

2.6.1.1.5 Marconi

Marconi is located at Cineca, Bologna, Italy.

Architecture	Intel OmniPath Cluster
Model	Lenovo NeXtScale
Peak performance	2 PFlops
System configuration	1512 nodes 36 cores/node 54,432 cores in total
Processors	2 x 18-cores Intel Xeon E5-2697 v4 (Broadwell) at 2.30 GHz, 8 cores each
Memory	128 GB RAM per node
Main memory	17 PB local storage
Networks	Intel OmniPath

Table 17: Marconi specification

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=1169

2.6.1.1.6 Pico

Pico is located at Cineca, Bologna, Italy.

Architecture	Linux Infiniband Cluster
System configuration	74 nodes with in total 1080 cores **51 compute nodes** • 20 cores each and 128 GB memory. • Processor: Intel Xeon (Ten-Core) E5-2670v2 2.50 GHz • A standard scientific computational environment, pre-installed applications in the visualization domain, as well as data analysis, post-processing and bioinformatics; access via ssh; submitting of large analysis in a PBS batch environment

	2 big memory nodes The two nodes, big1 and big2, equipped with 32 cores and 0.5 TB/ 40 cores -and1 TB of RAM for specific activities, which require a remarkable quantity of memory. Both are HP DL980 servers. Can be accessed upon request. • Big1 node: 8 Quad-Core ARK Intel(R) Xeon(R) E7520 processors with a clock of 1.87GHz and 512 GB of RAM, and a NVIDIA Quadro 6000 graphics card. • Big2 node: 4 Ten-Core Intel(R) Xeon(R) E7-2860 with a clock of 2.26GHz and 1024 GB of RAM. **Viz nodes** • 2 x (20 core, 128 GB memory, 2 x GPU NVIDIA K40) + 2 x (16 core, 512 GB memory, 1 x GPU NVIDIA K20) • Rremote visualization environment is defined on this partition, taking advantage from the large memory and the GPU acceleration. **BigInsights nodes** • 4 x 16 core nodes, 64 GB memory/node, 32 TB local disk space/node + 1 x 20 core node, 128 GB memory) • IBM solution for Hadoop applications is available **Other nodes** • 13 x 20 core nodes, 128 GB memory • Used for internal activities in the domain of Cloud computing, large Scientific Databases, hadoop for science. • Special projects can be activated on this partition, selected by special calls.

Table 18: Pico specification

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=942

2.6.1.1.7 *Pilot systems*

The pilot systems deployed in the third, final phase of a Pre-Commercial Procurement during the Ramp-up Phase are located at Jülich Supercomputing Centre, Forschungszentrum Jülich. The systems will be integrated into the HPAC Platform as soon as they are operational.

JULIA (Cray pilot)

The key technologies used for the Cray pilot system are

- KNL-based compute nodes
- 100 Gbps network technology (Omni-Path)
- NVRAM technologies
- Coherent software stack

More information about JULIA: https://hbp-hpc-platform.fz-juelich.de/?page_id=1063

JURON (IBM-NVIDIA pilot)

The key technologies used for the pilot system are

- POWER8′ + NVIDIA Tesla P100 interconnected via NVLink

- 100 Gbps network technology (InfiniBand EDR)

More information about JURON: https://hbp-hpc-platform.fz-juelich.de/?page_id=1073

2.6.1.2 General computing capabilities

An analysis of the SGA1 use cases will show which additional, general computing resources are needed in particular. KIT provides virtual machines for HBP researchers. This service is based on an OpenStack deployment. CSCS provides lightweight VMs based on Docker and Shifter.

2.6.2 Memory and storage capabilities

Related functional requirement: SP7-FR-002

2.6.2.1 HPC storage

All four HPC centres provide large storage facilities that can be accessed from the HPC systems.

In general, the HPC centres differentiate between the following kinds of repositories or file systems. The technical implementation differs between the sites:

- A home file system/directory (per user) for source code, binaries, libraries, applications, small datasets
- A scratch or work file system/directory (per user and/or group) as temporary storage for larger datasets, potentially with automatic clean-up of unused files
- An archive file system for long-term storage of results that is at some sites not directly accessible from the compute nodes of the HPC systems
- A file system/directory used as software repository for globally installed software and libraries

Details: https://hbp-hpc-platform.fz-juelich.de/?page_id=676

2.6.2.2 HBP Cloud storage

KIT provides Cloud storage with a capacity of currently 480TB, which is accessible through an Amazon S3-compatible interface.

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=51

2.6.3 Visualisation capabilities

Related functional requirement: SP7-FR-003

Three high-fidelity visualisation systems at RWTH Aachen University and at EPFL are available and can be booked by HBP scientists through a web form in the Guidebook (https://hbp-hpc-platform.fz-juelich.de/?page_id=442).

More information: https://hbp-hpc-platform.fz-juelich.de/?page_id=53

2.6.3.1 AixCAVE

RWTH operates the AixCAVE, a 5.25 x 5.25 x 3.3 m^3 fully immersive virtual reality environment using 24 HD-DLP projectors with a resolution of 1920x1200 pixels for rear projection (4 for each wall, 8 for the floor plane). The projectors are driven by a 24-node visualisation cluster, each containing two Intel Xeon Westmere CPUs (6 Cores at 2.7 GHz), 24 GB RAM and two NVIDIA Quadro 6000 GPUs. The complete volume of the AixCAVE is tracked by an infra-red-based tracking system with eight ARTTRACK1 and two tracking cameras.

2.6.3.2 Tiled Display Walls

RWTH operates a Tiled Display Wall, a display composed of six tiles (Full HD resolution and 46'' size each, total resolution of 5760 x 2160 pixels), is driven by a 7-node cluster equipped with one Intel Xeon E5 (4 Cores at 3.7 GHz), 16 GB RAM, and one NIVIDIA GTX 780Ti GPU each. The Display Wall is further equipped with two ARTTRACK2 tracking cameras for interaction.

EPFL also operates a Tiled Display Wall with a display composed of twelve tiles (Full HD resolution for each tile, total resolution of 7680 x 3240 pixels), driven by a 2-node cluster equipped with three GPUs each. The Display Wall is further equipped with a PQLab multi-touch interface supporting up to 32 touch points.

2.6.4 Network

Related functional requirement: SP7-FR-004

A Memorandum of Understanding between the PRACE-3IP project and the HBP has been signed in January 2015. It allows the HBP to use the fast, secure PRACE network for connecting the HPC systems at BSC, Cineca, CSCS and Jülich, which enables the transfer of large datasets between these sites. BSC, Cineca and Jülich have dedicated 10 Gb/s links. CSCS is connected with a 1 Gb/s IPsec tunnel. The plannings for also connecting CSCS with a faster link have already started.

Data transfers between an HPC centre and the visualisation sites RWTH and EPFL are enabled through dedicated connections between Jülich-RWTH and CSCS-EPFL. Data transfers to KIT can be made through regular Internet links, as no requirement for a high speed network is expected.

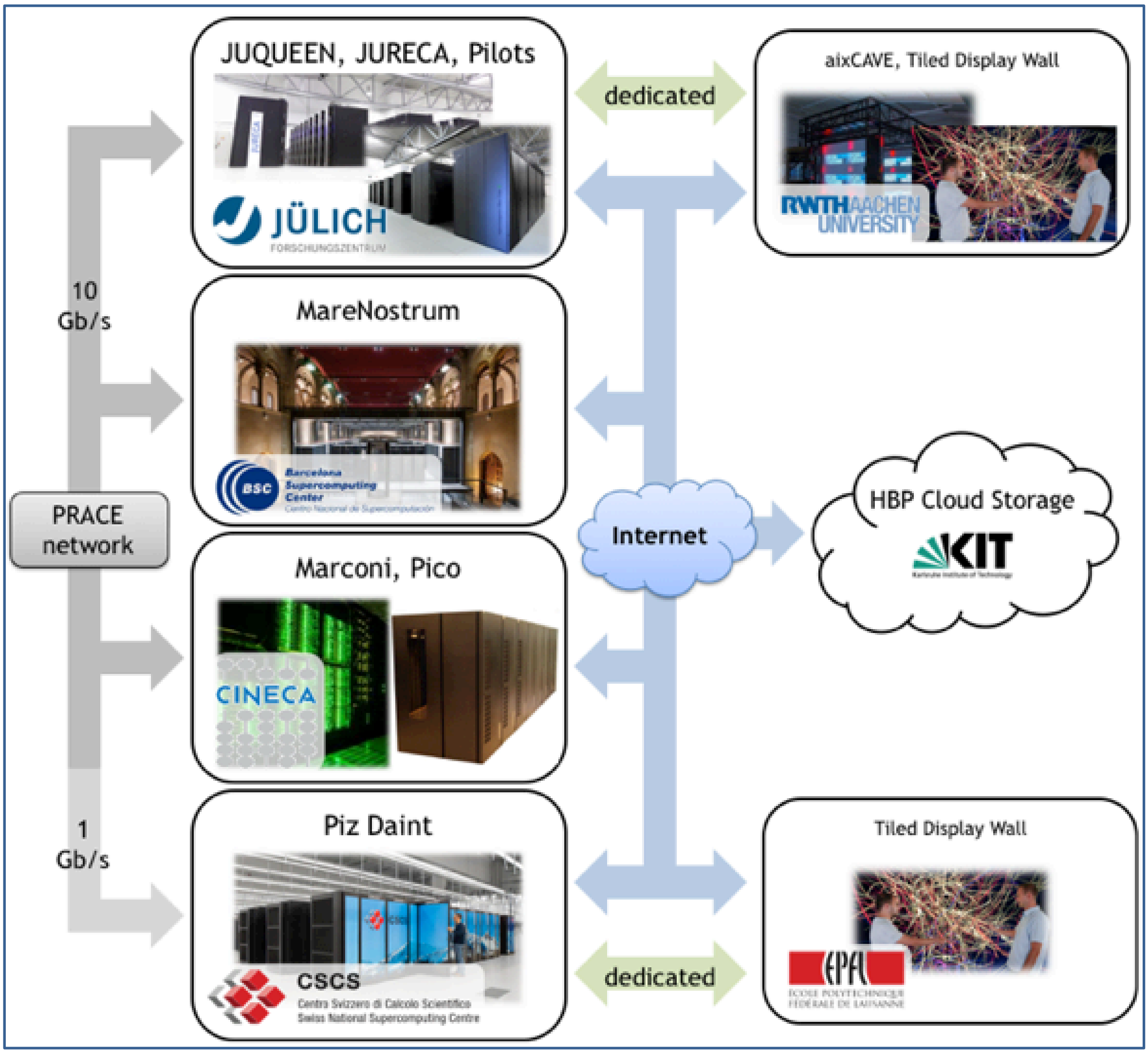

Figure 4: Network overview

Figure 4 summarises the different types of network connections between the main SP7 sites.

2.6.5 *Middleware*

Related functional requirement: SP7-FR-005

BSC, Cineca, CSCS, JSC and KIT operate servers on which the relevant UNICORE components are installed and configured, i.e. UNICORE gateway, UNICORE/X, XUUDB, UNICORE TSI and UFTPD daemons.

JSC additionally operates a central UNICORE registry server that is used as a common entry point and a central Unity server (UNICORE authentication service).

2.6.6 *Authentication and authorisation*

Related functional requirement: SP7-FR-007

Each site of the HPAC Platform provides a (virtual) machine or server to host the distributed LDAP service.

2.6.7 *Documentation and user training*

Related functional requirement: SP7-FR-017

JSC operates a web server for hosting the HPAC Platform Guidebook website.

2.6.8 Support

Related functional requirement: SP7-FR-018

BSC operates a web server for hosting the HPAC Platform ticket system.

2.7 HPAC Platform: Access to the HPC systems

2.7.1 General principles

The access to tier-0 and tier-1 systems is subject to the following general principles since the HBP currently does not own supercomputer or data centre resources.

- No large-scale in-kind contributions of resources are foreseen in SGA1.
- For SGA1, available resources are PRACE/European or national resources. In this context, applications are subject to the rules of the providers of these resources.
- Preparatory and development resources are provided under separate conditions, as explained in more detail below.

2.7.2 Large-scale Projects

2.7.2.1 European resources (PRACE[5]; called Tier-0)

The following is subject to a signature of PRACE 2.0 foreseen in September 2016.

PRACE hosting members are committed to fund and deliver large PRACE RI computing and data management resources. Under PRACE 2.0 there will be five hosting members: France, Germany, Italy, Spain and Switzerland. Contributing general partners are other countries that do not contribute hardware resources but e.g. support. All 27 EU member states are members of PRACE.

In the coming second phase of PRACE operations (PRACE 2.0), PIs from hosting members can apply starting with call 14 in October 2016. These resources are open to all sciences in a competitive manner. Applications will underlie the peer review process of PRACE.

In general, 30 % of the available resources of the contributing hosting members will be given to PRACE for PI from the hosting members.

In general, PIs from contributing general partners will have access to apply for up to 10 % of resources provided by the HMs.

2.7.2.2 National resources (called Tier-1)

In general, PIs being citizen of the country that offers resources can apply. Applications are under the national peer review rules.

Some countries accept large-scale applications PIs from foreign countries:

- Jülich Supercomputing Centre (JUELICH) will consider those applications to a certain extent. They will be subject to the transparent scientific peer review under the John von Neumann Institute for Computing. There is no restriction applied for Co-PIs and collaborators.
- The national resources provided by the User Lab at Swiss National Supercomputing Centre (CSCS) are open to all scientists, irrespective of their nationality, provided the projects comply with US Export Control. These resources are subject to a competitive allocation process with transparent peer review.
- For national calls at BSC it is recommended but not mandatory to have a Spanish collaborator in the project.
- The national calls in Italy are open to PIs affiliated to an Italian institution, while no restriction is applied for the Co-PI and collaborators. The Italian resources are provided subject to a competitive allocation process with transparent peer review. However, Cineca additionally provides moderate amounts of computing time to HBP scientists.

The following national calls are available for the systems integrated into the HPAC Platform:

- BSC: MareNostrum
- CINECA: Marconi and Pico
- CSCS: Piz Daint
- JSC: JUQUEEN and JURECA

2.7.2.3 Future HBP assigned resources

In Europe, PRACE is acknowledged as the European HPC infrastructure by the European Commission and all 27 member countries of PRACE. Funding HBP supercomputing resources, the EC expects all large-scale allocations to be provided through PRACE. Therefore, future HBP resources provided by the EC will undergo the PRACE peer review. The full amount of these resources is foreseen to be provided to the HBP.

2.7.3 Preparatory and Development Projects

The members of the HPAC Platform provide preparatory and development access. The former typically is intended to prepare larger allocation projects and range from 5,000 to 100,000 core hours, depending on the systems. Development projects serve major algorithm and code development purposes. Preparatory and development allocation requests are handled locally by these sites and typically require only technical reviews. The size of development projects at CSCS can be as large as one million core hours.

2.8 HPAC Platform: Architecture Overview

This section provides an overview of the HPAC Platform architecture describing how the components listed in sections 2.5 and 2.6 relate to each other. The following colour code is used to differentiate between different types of components:

(orange)	Service-Oriented Architecture (SOA) or functional service
(dark green)	Infrastructure or hardware
(light green)	Software
(purple)	FeDaPP component
(red)	Documentation or report

Some components that are developed as part of the Federated Data Pilot Project (FeDaPP) are already mentioned in the following. The next update of this document will be used to fully integrated the FeDaPP products and services into the HPAC Platform.

2.8.1 Federated computing services

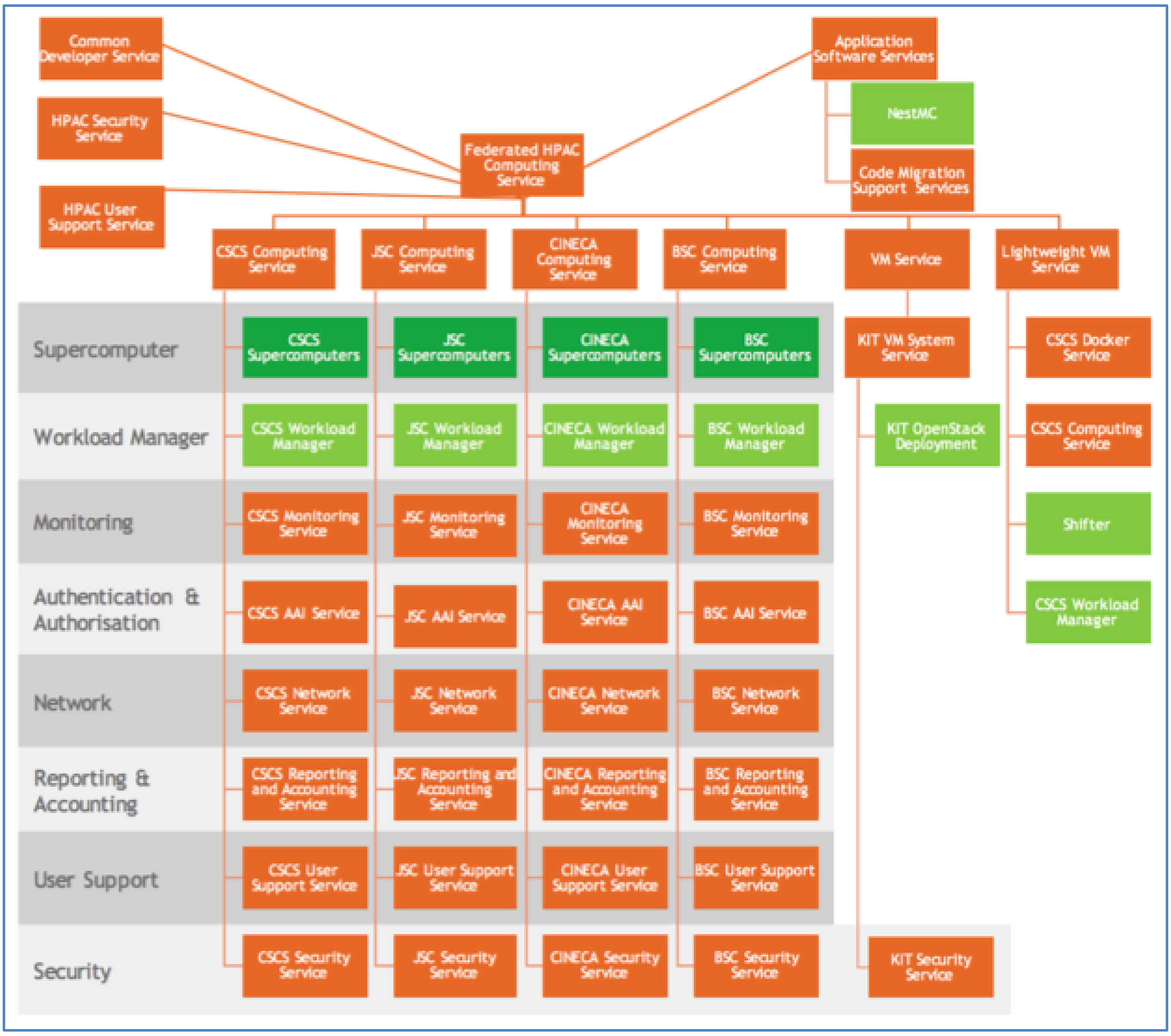

Figure 5: Federated computing services

The federated computing services (see Figure 5) comprise the computing services of the HPC centres BSC, CINECA, CSCS and JSC, virtual machines (VMs) provided by KIT (SP7-FR-001) and lightweight VMs based on Docker and Shifter that are available at CSCS (SP7-FR-013). For offering the computing services at the HPC sites, the supercomputers are integrated into the Platform (SP7-FR-009), a workload manager (SP7-FR-010) is in place for the HPC job management. In addition, monitoring (SP7-FR-012), an authentication and authorisation infrastructure (AAI; SP7-FR-007), a network for fast access to the systems (SP7-FR-004), reporting and accounting (SP7-FR-008), user support (SP7-FR018) and security services (SP7-NFR-005) are in place at each of the sites. The VM service at KIT is based on an OpenStack deployment; security mechanisms are also in place for the provided VMs.

Part of the federated computing services are also common developer services (SP7-FR-018.3), which comprises support for users to integrate bug tracking, source control, continuous integration and other software engineering best practices into their development processes. The application software services are meant to support the users of the HPAC Platform in migrating their codes to hybrid and accelerated architectures (SP7-FR-018.4). It also comprises the development of the scalable simulator NestMC (SP7-FR-015). Security services and user support are important additional elements of the federated computing services.

2.8.2 Federated data services

Figure 6: Federated data services

The federated data services (see Figure 6) are structured in a similar way as the federated computing services. CINECA, CSCS and JSC make active and archive repositories available by integrating them into a federated data infrastructure in the context of FeDaPP (SP7-FR-002). These repositories are complemented by upload and download services enabling the user to transfer own data into the provided repositories or to download results from there. In addition, fast transfer services for data transfers between the three centres are being configured (SP7-FR-011). BSC will be integrated into the federated data infrastructure developed in FeDaPP in the future. KIT additionally provides an S3 storage. All five sites providing data services have monitoring, AAI, fast network connectivity, reporting and accounting, user support and security services in place.

2.8.3 UNICORE services

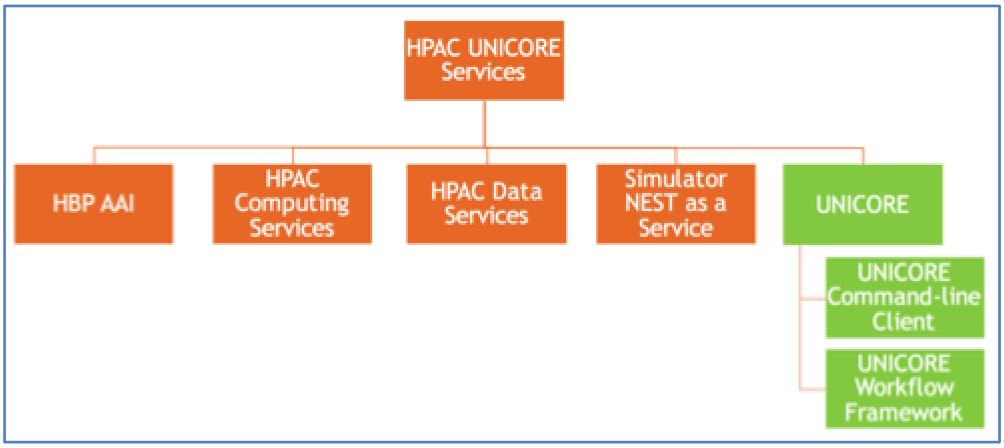

Figure 7: UNICORE services

The middleware UNICORE (see Figure 7; SP7-FR-005) is used to integrate the computing and data services into the HPAC Platform. It provides the UNICORE workflow framework for implementing workflows and a command-line client as a low-level access to the infrastructure. The UNICORE services rely on the HBP authentication and authorisation infrastructure (AAI) services for enabling a mapping between HBP user accounts and HPC accounts (SP7-FR-009). UNICORE can also be used to start a NEST simulation from a Jupyter notebook in the Collaboratory as a layer between the workload manager of the HPC system and the notebook.

2.8.4 Monitoring services

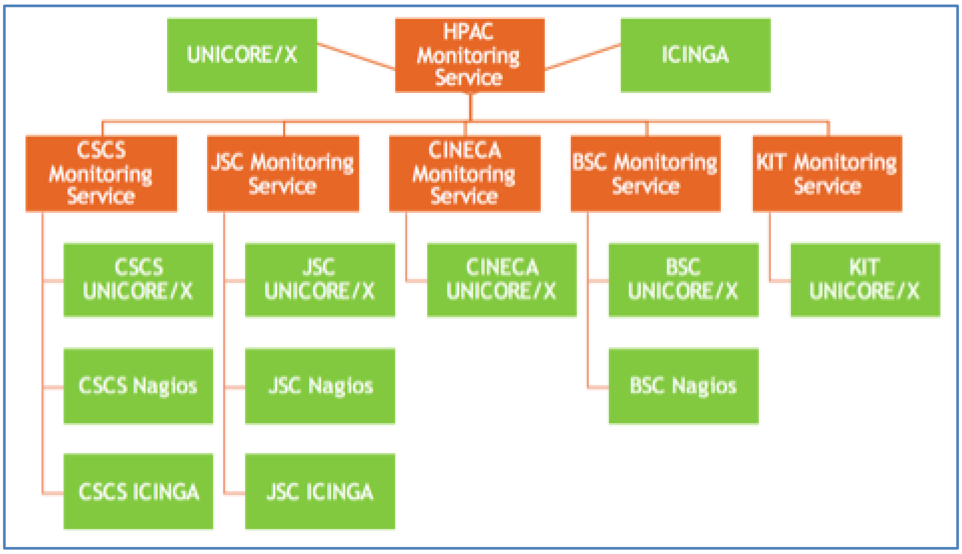

Figure 8: Monitoring services

The monitoring services (see Figure 8; SP7-FR-012) are based on UNICORE/X, a service container supporting SOAP Web Services and RESTful services, and ICINGA. BSC, CSCS, CINECA, JSC and KIT operate UNICORE/X instances for making monitoring information available to a central monitoring service. CSCS, JSC and BSC use Nagios for the system monitoring; CINECA and KIT have alternate solutions in place. CSCS and JSC in addition have ICINGA installations in place.

2.8.5 *Authentication and authorisation services*

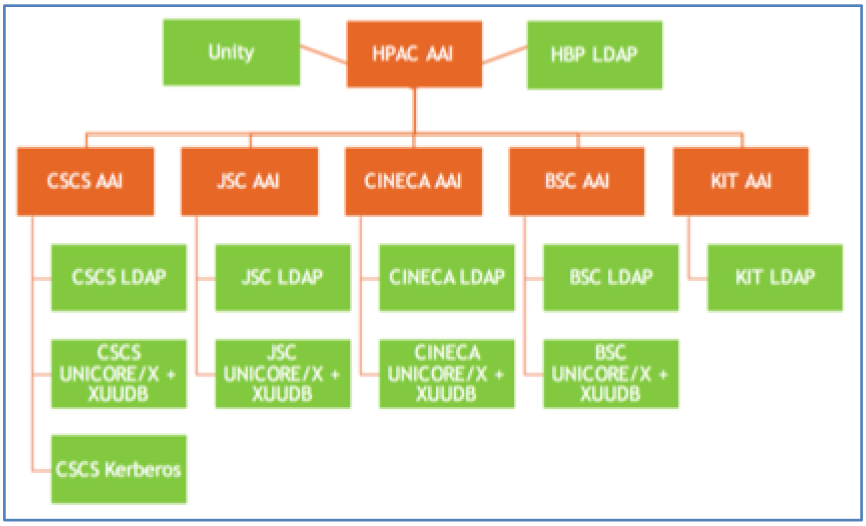

Figure 9: Authentication and authorisation services

The authentication and authorisation infrastructure (AAI; SP7-FR-007; see Figure 9) is based on LDAP, UNICORE, Unity and Kerberos. The HPC centres and KIT operate LDAP instances for maintaining the list of current users and their access permissions, which are kept up-to-date by the sites. XUUDB instances, the UNICORE user database, are used to map the user's accounts at the different sites with each other. UNICORE/X provides the REST API to query the user databases at the sites. Unity can be used by services for the user authentication.

2.8.6 Network services

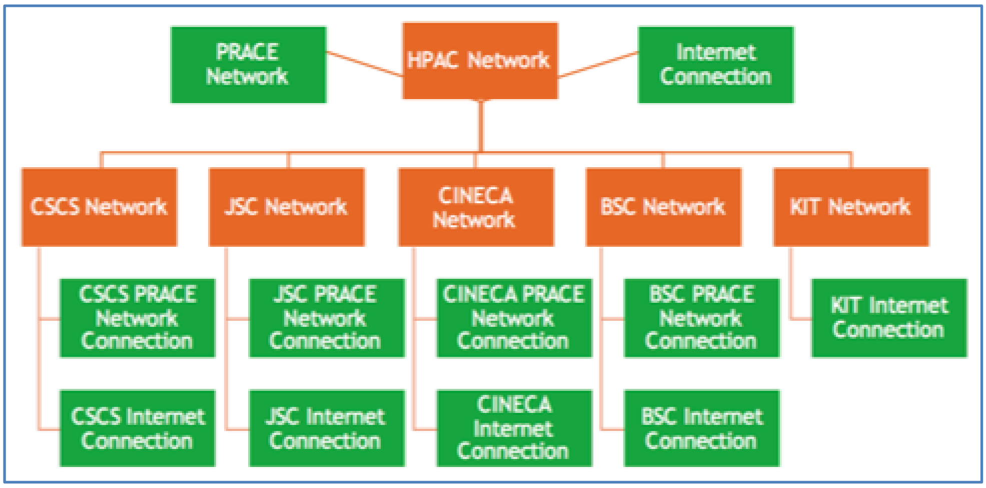

Figure 10: Network services

The HPAC Platform uses two different networks (SP7-FR-004, see Figure 10). The HPC sites are connected to the PRACE network, which is used for data transfer between these sites. The resources at these four sites can be accessed by the users using standard internet connections, which is the usual way for uploading data to or downloading data from an HPC storage. The resources at KIT, i.e. the S3 storage and virtual machines, are accessible through the standard internet connections as well.

2.8.7 Reporting and accounting services

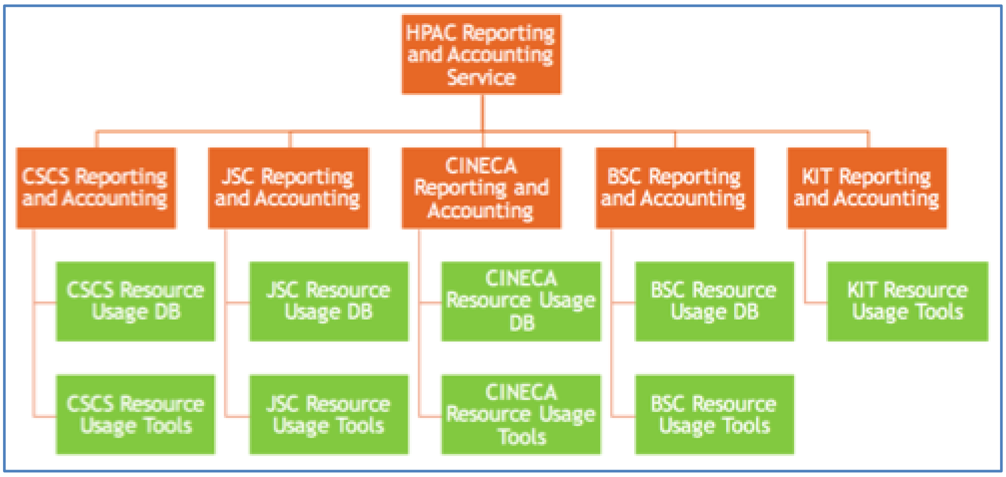

Figure 11: Reporting and accounting services

Reporting and accounting (SP7-FR-008) of the resource usage are handled locally and internally by the sites. They have tools databases in place for collecting and monitoring the resource usage.

2.8.8 *User support and documentation*

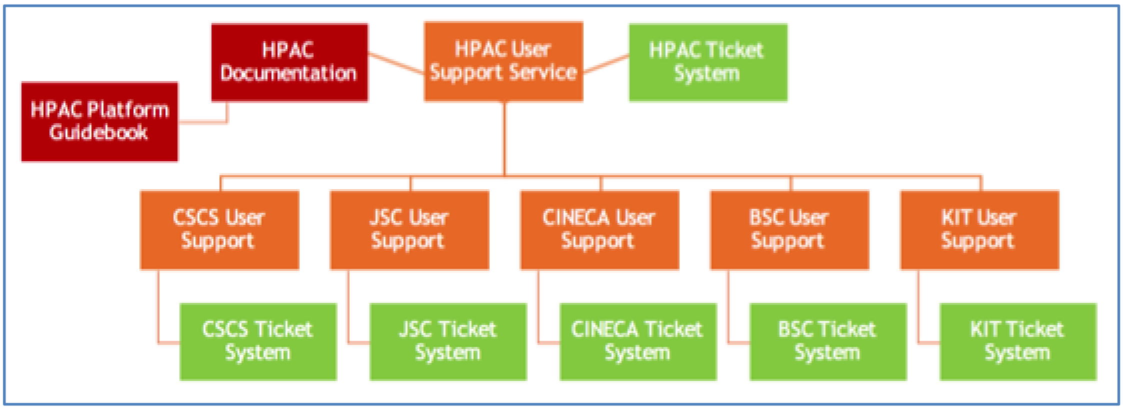

Figure 12: User support and documentation services

For supporting the users of the HPAC Platform (SP7-FR-018, see Figure 12) a support team with members from the HPC sites and KIT is established. The team can be reached through a central HPAC support email address (hpac-support@humanbrainproject.eu). When an email is sent to this address, a ticket in the HPAC ticket system is opened. The support team member on duty moves this ticket into the queue of the site that should deal with this ticket. This (semi-)automatically opens a ticket in the internal ticket system at the respective site, so that the local support team can deal with it.

The documentation (SP7-FR-017) of the HPAC Platform is available in the “HPAC Platform Guidebook”, a publicly accessible website (https://hbp-hpc-platform.fz-juelich.de/) that is also available from the HPAC Collab in the Collaboratory (https://collab.humanbrainproject.eu/#/collab/264/nav/2378).

2.8.9 Security

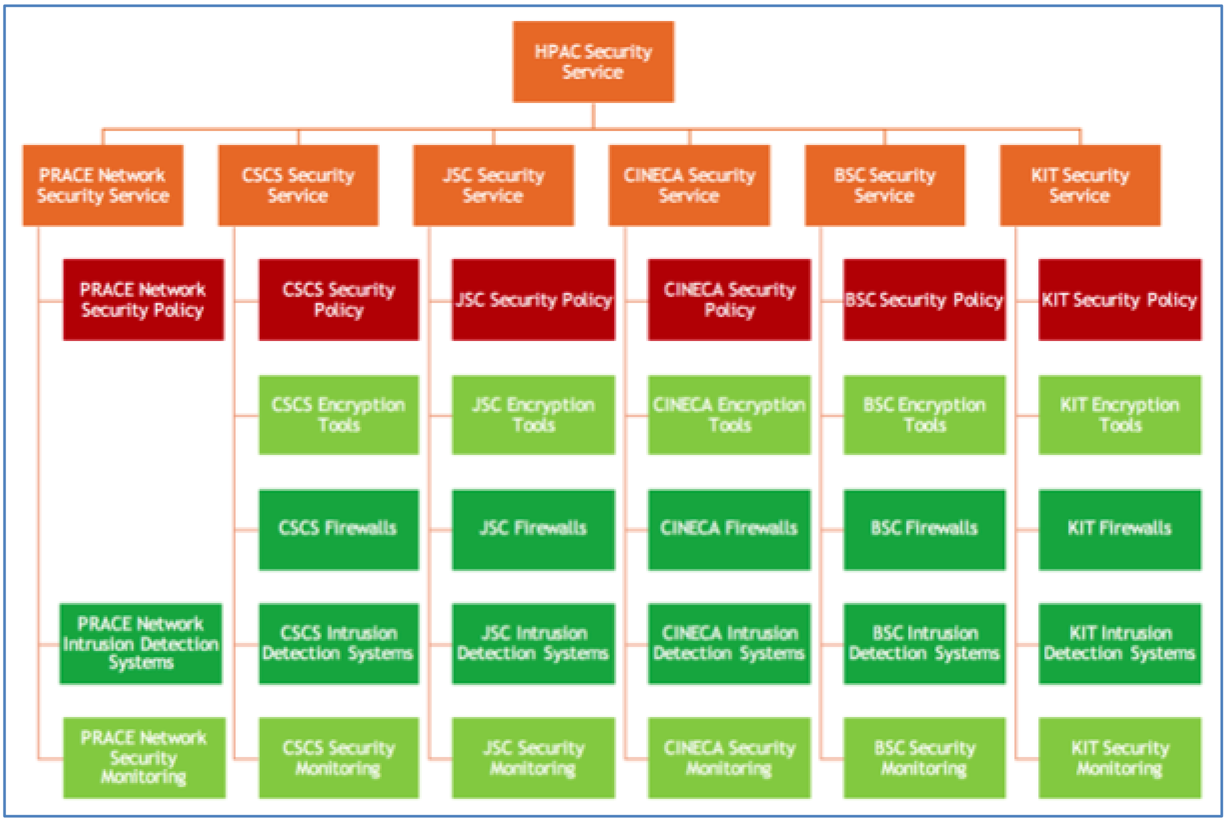

Figure 13: Security services

All sites have mechanisms in place to ensure the security (SP7-NFR-005, see Figure 13) of infrastructure and services. The sites have their own security policies defined that comply with applicable, national and European law. The same applies to the PRACE network. Encryption tools are available in order to enable secure operations like data transfers or access to the resources. Firewalls, intrusion detection systems and monitoring services protect critical infrastructure at the sites.

2.9 HPAC Platform: Relations to other Platforms

The HPAC Platform serves all other HBP Subprojects. The HBP Platforms provide services on different levels and build their services on top of each other as shown in Figure 14. The HPAC Platform is the base infrastructure for the HBP comprising storage and compute capabilities. The Neuroinformatics Platform (SP5) provides data services that run on the infrastructure and use services provided by SP7. The Brain Simulation Platform (SP6) uses the data services provided by SP5 and the HPAC Platform for their research and developments. The Neurorobotics Platform (SP10) runs simulations on Neuromorphic (SP9) and HPC systems, in the latter case using the simulation technology developed in SP6 and SP7. The neuroscience Subprojects SP1 to SP4 use the infrastructure provided by the Platforms for their research. The IT infrastructure at the hospitals integrated in the Medical Informatics Platform (SP8) exists in parallel to the HPAC Platform. This is necessary since the medical data must not leave the hospitals. SP8 also plans to use HPC resources for some of their data analytics use cases that require significant compute resources (see also Section 2.2.5).

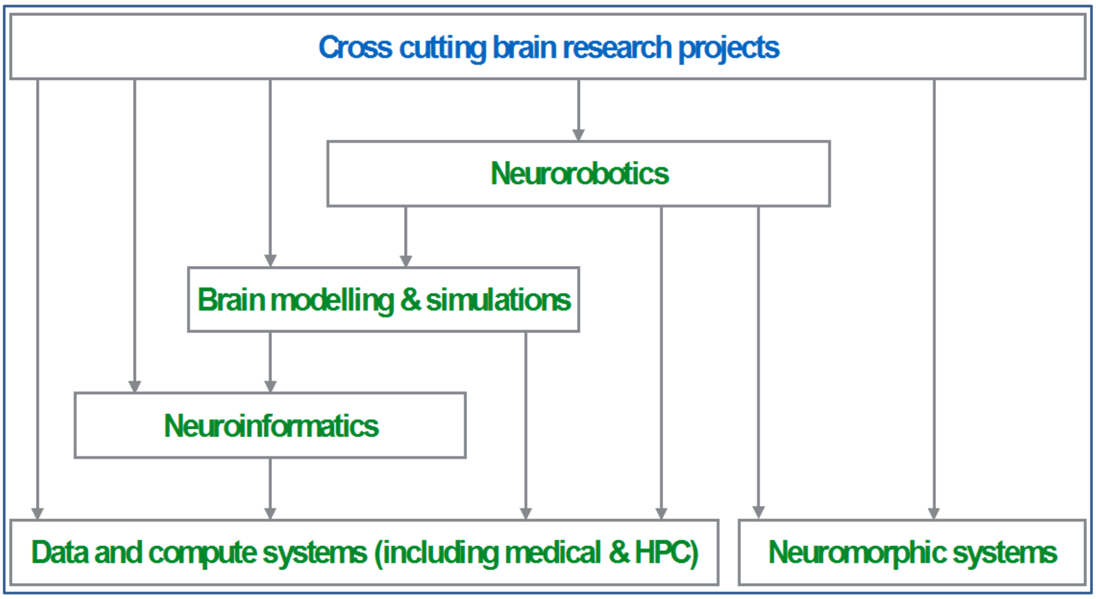

Figure 14: Relationship of the HBP Platforms

The HPAC Platform provides a number of HPC-related services to the other Subprojects, for example

- HPC user management: user authentication, authorisation, group membership
- HPC job management: submission and management of batch jobs
- HPC data management: management of high volumes of experimental and simulation data, and making the data available for subsequent computational tasks
- HPC tools: visualisation, programming tools, performance analysis tools
- HPC systems monitoring
- HPC resource usage monitoring and accounting, including quotas for compute time and storage space

All these services are made available through the provision of easy-to-use programming APIs and libraries. Some services can also be used through the middleware UNICORE, e.g. submission and management of HPC jobs. Key services are also integrated into and/or accessible from the Collaboratory. Making the same functionality accessible through different access methods is important due to the diverse community in the interdisciplinary Human Brain Project. Where some users (SCIDEV and DEV) are familiar working with APIs and libraries directly from the command-line, others prefer to work with (graphical) user interfaces since they are users but not developers of software (PU).

The HBP aims to build a Service-Oriented Architecture (SOA). From an architectural point of view, the HBP Platforms provide different kinds of services, as shown in Table 19. The infrastructure at the data centres BSC, CINECA, CSCS and JSC, the infrastructure at the hospitals and the Neuromorphic computing systems form the base infrastructure of the HBP. The compute and data services provided by SP7 and SP9 form the Infrastructure as a Service (IaaS) layer. The HPAC Platform and the Neuromorphic Computing Platform provide Platform as a Service (PaaS) on top of the IaaS layer. The Neuroinformatics and Medical Informatics Platforms provide Data as a Service using the underlying PaaS. Software as a Service in the areas of Neuroinformatics, Brain Simulations, Neurorobotics and Medical Informatics is made available through the Collaboratory as the Gateway to the HBP infrastructure.

Gateway	**Collaboratory**	
Software as a Service	Neuroinformatics, Brain Simulations, Neurorobotics, Medical Informatics	
Data as a Service	Neuroinformatics	Medical Informatics
Platform as a Service	HPAC Platform	Neuromorphic Computing Platform
Infrastructure as a Service	Compute & data infrastructure services (incl. HPC)	Neuromorphic compute services
Base infrastructure	Public data centres (BSC, Cineca, CSCS, Jülich), IT infrastructure at the hospitals, Neuromorphic computing systems	

Table 19: Structure of services in the HBP infrastructure

More information about the relation of the HBP Platforms can be found in Appendix 3 *White Paper "Transforming the Human Brain Project Platforms into a Community-Driven Infrastructure for Brain Research* of the Framework Partnership Agreement[6].

2.10 HPAC Platform: Necessary Parallel Activities

2.10.1 Access Committee

As a general rule, any substantial resource allocation on the HPAC Platform must always be based on a peer-reviewed project proposal. This requires suitable application, reviewing, ranking and allocation procedures. Under the precondition that the HBP will have certain amounts of resources at the hosting sites at its disposal in the future, the project-internal distribution of these resources should be organised by an appropriate Access Committee with HBP-internal and external members. The membership of the Access Committee should be defined such as to ensure an appropriate balance between the countries and disciplines present in the HBP and the broader user community.

2.10.2 Coordination with other European Research Initiatives and Infrastructures

2.10.2.1 PRACE

All four supercomputing centres of the HPAC Platform are hosting members of the Partnership for Advanced Computing in Europe (PRACE). Therefore, and to be in alignment with the EC's strategy for HPC, the provision of computing resources by these centres and access to these resources should be coordinated with PRACE. Existing and proven infrastructures and procedures should be used and leveraged wherever possible. Starting with the 14th Call for Proposals for Project Access, HPC resources at all four HPC centres of the HPAC Platform will be available through PRACE for scientists in Europe. This is important since most of the national calls for proposals (GCS/NIC for Jülich, ISCRA for Cineca, RES for BSC) have strict eligibility criteria that either only allows applications from scientists working in the same country or that at least strongly recommend such an affiliation. PRACE calls enable all HBP PIs irrespective of their affiliation to apply for HPC resources at any of the centres.

2.10.2.2 EUDAT

HPAC Platform members BSC, CINECA and JSC are also members of EUDAT[7], the European Data Infrastructure. Through these partners, the HBP is in close contact with EUDAT to work out which EUDAT services might be useful for the HBP, and how they could be exploited. In a first meeting of HBP and EUDAT representatives in January 2014, the EUDAT services B2SHARE for data sharing and B2FIND for metadata harvesting where identified as first candidates for further evaluation.

2.10.2.3 EGI

The European Grid Infrastructure[8] (EGI) is a federation of over 350 resource centres. It is a publicly funded e-infrastructure with the aim to give scientists access to more than 650,000 logical CPUs, 500 PB of disk and tape storage and 21 federated cloud providers to drive research and innovation in Europe. The infrastructure provides both high throughput computing and cloud compute/storage capabilities. The resources are provided by over 325 resource centres across 56 countries in Europe, the Asia-Pacific region, Canada and Latin America.

The HPAC Platform and the European Grid Infrastructure (EGI) are investigating the possibilities for a future collaboration in the areas of general computing. First meetings took place in August and September 2016.

2.10.3 Interaction with external HPC technology providers

The HBP Pre-Commercial Procurement (PCP), conducted by SP7 during the Ramp-Up Phase, aimed to obtain architectural solutions that should permit an interactive use of large-scale supercomputers for complex workflows comprising concurrent simulation, analysis and visualisation workloads. The goal of the HBP PCP was to have suppliers of HPC technology competitively research, develop and integrate novel technologies in the areas of dense memory integration, scalable visualization and dynamic resource management in order to enable such “interactive supercomputing”. It resulted in two cutting-edge demonstrators developed by Cray and a consortium of IBM and NVIDIA, which are deployed as pilots at JUELICH and used not only for testing and benchmarking, but also productively for neuroscience applications. The pilot systems will become part of the current HPAC Platform, and the R&D results of the HBP PCP will help shaping future versions of it.

The HBP PCP was carried out in parallel to the setup of the initial HPC Platform, as a forward-looking activity involving external stakeholders, in this case developers of relevant HPC technologies. In the same spirit, SP7 must continue to work with such stakeholders under SGA1, to enable them to provide innovative and competitive solutions required for implementing and advancing the HPAC Platform.

This goal requires collecting knowledge created across the Project, in particular in those parts concerned with applications (like SP6, the Brain Simulation Platform), and to organize them in a suitable way.

To facilitate co-design, it will also be necessary to complement knowledge created in the Project with information on developments in HPC technologies and architectures from outside the Project. It is therefore planned to continuously assess the relevant roadmaps.

A key aspect of knowledge management is sharing this knowledge with relevant external stakeholders, which will mostly be commercial operators. This involves dissemination, sharing, exchange and communication of data and methodologies, training in how to use and understand this information, and interpreting the knowledge.

3. Glossary

API – Application Programming Interface

AUP – Acceptable Use Policy

CAVE - a high resolution, 3D fully immersive virtual reality environment where projectors are directed to three, four, five or six of the walls of a room-sized cube

CERT – Community Emergency Response Team

Collaboratory – see *HBP Collaboratory*

Compute resource - a computer or collection of computers where a Job can be executed

CSIRT - Computer Security Incident Response Team

GUI – Graphical User Interface

HBP Collaboratory – the web portal through which the web accessible components of the HBP Platforms are made available

Host - A single operating system instance, running on virtualised or real hardware

ICT – Information & Communication Technology

Job - an instance of an execution of a Task on a compute resource

Parameter - a low data-density discrete data element, primarily meant to denote a value that one might enter into a single form element. It might also be used to refer to a richer configuration document containing a group of settings.

Platform - software components: libraries, services, APIs and their documentation that are to be used to build portals or cockpits.

RAID - a data storage technology that combines multiple disk drive components into a logical unit for the purposes of data redundancy and performance improvement.

REST - acronym for REpresentational State Transfer, for a definition see http://en.wikipedia.org/wiki/Representational_state_transfer

Service - a software function performed by a third party for a user or another service.

Site - A collection of hosts collected together in a single location. The grouping is potentially arbitrary.

SOA – Service-Oriented Architecture

Task - A logical software unit that can be executed on a Compute resource. It may or may not be visible as a Service. A Task identifies its dependencies and its default parameters.

ToU – Terms of Use

Workflow - A tree of decision structures and Tasks. It may or may not be visible as a Service.

4. References

[1] https://www.humanbrainproject.eu

[2] https://www.humanbrainproject.eu/documents/10180/1055011/SP07_D7.7.2_FINAL.pdf

[3] https://www.humanbrainproject.eu/documents/10180/1055011/SP7+D7.7.5+FINAL.pdf

[4] https://www.humanbrainproject.eu/documents/10180/1055011/SP7+D7.7.6+FINAL.pdf

[5] http://www.prace-ri.eu/

[6] https://www.humanbrainproject.eu/en_GB/framework-partnership-agreement

[7] https://www.eudat.eu/

[8] https://www.egi.eu/