

InPEX 2026, Niteroi, Brazil Computing Continuum Session 1

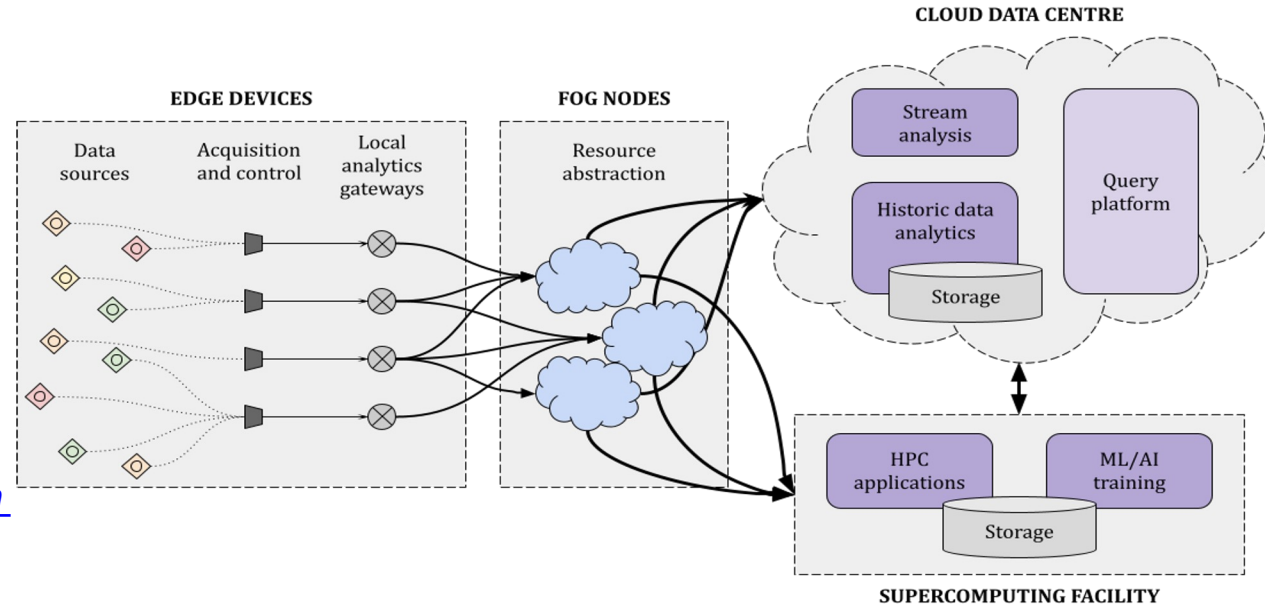
Gabriel Antoniu (Inria), Mario Acosta (BSC), Carlos Jaime Barrios Hernandez (UIS/INRIA),
Marcio Castro (UFSC), Nicola Ferrier (Northwestern University), Maria Girone (CERN),
Kento Sato (RIKEN), Hiroyuki Takizawa (Tohoku University)

InPEX 2026 - Computing Continuum Sessions

International Post-Exascale Initiative

- Session1: Fri, April 24, 9am-11am
 - Use case presentations and updates
- Session2: Fri: April 24, 2pm-4pm
 - Discussion on shared challenges and solutions

Shared notes: <https://tinyurl.com/inpex-continuum>



Use case characterization criteria

1. Access to HPC resources (for science communities)
2. Interface to a federation of resources (e.g. EuroHPC)
3. Co-design for the post-exascale systems
4. Portability of the benchmarks and codes
5. Deployment issues of the workflows and applications
6. Metrics for different deployment scenarios
7. Semantics and quality of data
8. Resources provisioning/scheduling across the continuum
9. End-to-end workflow control
10. Multitenancy
11. Data logistics
12. Energy efficiency
13. Reliability
14. Real-time data streaming at high rates

InPEX 2026 - Computing Continuum – Use Cases

International Post-Exascale Initiative

Updated use cases

- ODISSEY / SKA-DDF pipeline – radio astronomy continuum workflow (update)
- HEP use cases / SPECTRUM project – HPC integration and data-intensive science roadmap (update)
- ESiWACE / HPCW – climate and weather benchmarking, toward operational readiness (update)
- Urgent computing / National Data Platform – event-triggered data and AI workflows (update)

New use cases

- High-performance Imaging and Advanced AI Analytics to Inspect Road Infrastructure (Chile, SCALAC)
- LAC sustainable HPC testbed (SCALAC)
- Brazil federated infrastructure (with SCALAC and RedClara)
- LAGO large-scale computing support for astrophysics (LAC, Europe)
- Power-grid digital twin (BSC)
- AI pipeline infrastructure toward the FugakuNEXT era (RIKEN)

Other known use cases (not updated)

- Micro-meteorology control
- Wildfire digital twin
- Disaster digital twin

Updated challenges for known use cases (1/2)

International Post-Exascale Initiative

ODISSEE / SKA Radio astronomy

- Real-time stream processing at very high rates, with limited buffer capacity.
- Scalability of data logistics across the full path: instruments, online reduction, offline analysis and provenance.
- Federated access and resource provisioning across HPC centers.
- Deployment, portability, workflow control, scalability and reliability across heterogeneous subsystems.
- Energy efficiency and sustainability over long infrastructure lifetimes.

HEP use cases SPECTRUM, SHAKE CoE

- Standardizing interfaces and policies across HPC centers
- Securing strategic, long term access to HPC centers
- Efficient workflow and data orchestration
- Software optimization for HPC
- Scientific domains represented in the HPC roadmap
- Try to align specific domain needs and all scientific domains for the HPC topic

Updated challenges for known use cases (2/2)

International Post-Exascale Initiative

ESiWACE / HPCW

- Co-design and representativeness of benchmarks for upcoming post-exascale hardware.
- Portability, Fortran optimization and reproducibility tooling (CMake, Spack, CI/CD).
- End-to-end web / cloud / HPC workflows and standard data-transfer protocols.
- Real-time metrics, validation and benchmarking for different deployment scenarios.
- Operational allocation models for weather/climate services and long climate projections.

Urgent computing / National Data Platform (USA)

- QoS under uncertainty, hard deadlines and runtime cost/benefit trade-offs.
- Dynamic discovery and aggregation of available resources across the continuum.
- Edge integration, streaming data and data-at-rest for urgent applications.
- Provenance, explainability, reliability and support for dynamic execution.
- Co-design of the continuum as an operational system, not only an infrastructure stack.

New use cases – challenge synthesis

International Post-Exascale Initiative

Road infrastructure inspection

High-resolution imaging + AI: memory-heavy vision models, scarce labels, storage/I/O bottlenecks, imaging-to-HPC pipeline.

LAC sustainable HPC testbed

Testbed to measure altitude, humidity, power stability, corrosion, energy, reliability and lifetime-extension trade-offs.

Brazil federated infrastructure

Federation without SSO yet: data cloud + edge + distributed HPC, with automated placement and scheduling.

LAGO large-scale computing for astrophysics

Distributed science support: access, federation, workflow control, data logistics, privacy, scheduling and streaming.

Power-grid digital twin

Edge sensors + cloud orchestrator + HPC simulations; COMPSs agents, provenance and real-time actuation.

FugakuNEXT AI pipeline

Secure AI4Science workflows: model lifecycle management, secure use of open models, inference deployment, monitoring, traceability, seamless integration of components and mixed-workload scheduling.

The Use Case vs. Challenges Overview Matrix (don't struggle to read the text ;))

International Post-Exascale Initiative

	SKA - DDF pipeline	HEP use cases	ESIWACE	Urgent computing	Micro-meteorology control	Wildfire Digital Twin	Disaster Digital Twin Project	High-performance Imaging and Advanced AI Analytics to Inspect Road Infrastructure	Additional use case 1 from LAC (HPC Testbed Sustainable Infrastructure)	Additional use case 2 from LAC (Federated Infrastructure from Brazil)	LAQG Project - Large Scale Computing Support	Power grid digital twin	Relevant existing R&D efforts related to the challenge	Important strategic actions	Relevant technical milestones
Access to HPC resources (as a community)	X (exploitation of final data products offline)	X	X	X (Urgent computing needs to harness all available resources achieve its objectives)	X	https://doi.org/10.1145/3744641					X	X (run simulations that generate, store and share data - power grid)	EU-CASTEL2 initiative (not successful)	EuroHPC projects comes with no resources. Communities can benefit EuroHPC resources are provided along projects and CoEs granted	JP-HPC manages the accesses to HPCs and the large storages, and each university may have the account system for accessing their resources. Network?
Interface to federation of resources (e.g. multiple EuroHPC centers)	X (exploitation of final data products offline)	X	X (seeks opportunity to participate with domain cases in the development of the federated infrastructure)	X (End to end workflows require the federation of resources across the digital continuum, including HPC centers)			X		X	X	X		JP (RKEN): Working on Data Acquisition System for AI for Science by Global: JPU Osaka; Working on data aggregation infra named ONION (make data searchable) and high-speed data transfer service on campus EU: Destination Earth studies on federated challenges JP (COE) - meta-scheduling	JP (RKEN): Euro system and community building around Global Work in close collaboration with EPF during 2025. This corresponds to the development of the EPF Testbed (development phase). Assessment to be ready end of Q1 2026. Domain communities do need to participate and provide feedback (co-design)	JP (RKEN): Integration of Global with other data sharing service because different organization uses different data sharing services. JPU Osaka: meta-data management of aggregated data
Co-design for the post-exascale systems	X (especially for offline component, but downscaled solutions could be reused for online component in order to maximize efficiency)	X	X (seeks opportunity to participate in the co-design of the benchmark to assess upcoming HW for the Post-Exascale)	X (co-design of the digital continuum, both architecturally and operationally)					X	X	X		EU: DARE and Center of Excellence as ESWACE3 JP (RKEN): Fugaku/UXT	Collaborate with the community CoEs to ensure that we have an useful ongoing while the new hardware is developed JP (RKEN): Selection of system software (quantum)	
Portability of the benchmarks and codes	X (exploitation of final data products offline)	X (porting to GPUs or new accelerators)	X (best efficiency possible in different CPU machines and porting to GPUs or new accelerators)	X (essential as workflows will have to run on available resources, which may not be known until runtime)		Portability of the benchmarks and codes			X	X	X		EU: ODSESE project CoEs as ESWACE3		
Deployment issues of the workflows and applications	X (across the continuum: each sub-systems handle a portion of the overall pipeline)	X (having CVMFS as a baseline solution available at all centers would be largely beneficial (either full installation or possible at user level))	X (need automation/on-demand workflow submission)	X		Deployment issues of the workflows and applications			X	Scientific workflows of reservoir simulations and bioinformatics in HPC and clouds.	X		EU: ODSESE project, workflow managers (Autosub) and initiatives as Spack on development	community (vertical) CoEs to ensure the adaptation for domain communities	
Metrics for different deployment scenarios			X (need to collect metrics in real time)	X (Evaluation of cost/benefit at runtime is essential)		https://doi.org/10.1145/3731545.3743643			X		X				
Semantics and quality of data				X		very important for decision support and fire intel			X		X				
Resources provisioning	X (mostly for exploitation of final data products offline)	X (an optimal utilization would require longer provisioning - e.g. 72 hours)		X		Resources provisioning			X		X				
End-to-end workflow control, workflow management	X (for the online part)	X (need custom interfaces to submit pilot-based workflows)	X (web - Cloud- HPC)	X	Custom workflow control	End-to-end workflow control, workflow management	X (the urgent workflow consists of 200+ tasks, but only a small number of tasks need to be executed on HPC systems)		X	Workflow management systems for big-of-tasks and multi-workflow applications	X	X (sensors and instruments in the edge, cloud used to host orchestrate and set of tools, HPC for simulation)	EU: ODSESE, Blow4HPC, CoEs as ESWACE3 JP (RKEN): No project	Apply for transversal CoEs for common look and a joint funded collaboration with community (vertical) CoEs to ensure the adaptation for domain communities JP (RKEN):	Interoperable workflow control/ management tool JP (RKEN): Standardization of workflow system because there are many workflow systems. These workflow community systems which requires a common support
Multi-tenancies	X (exploitation of final data products offline)			X		Multi-tenancies			X		X				
Data logistics	X (across the continuum, from real-time data stream to offline data analysis: several data topologies and formats as well as provenance tracking strategies)	X (need transient storage; limited connectivity; need standard data transfer protocols)	X (need standard data transfer protocols)	X (Data, both real-time streaming and at rest, is critical to urgent applications)	Real-time data from sensors are used	Data logistics			X		X	X (data flows from edge to cloud and HPC)			
Reproducibility	X (for the online part)			X (Provenance, explainability is critical)		Reproducibility			X		X	X (partially supported by COMPS provenance)	JP (U Osaka): Working on Data resource management system, EU (brno): E2C lab project, https://doi.org/10.1145/3731545.3743643	JP (U Osaka) reproducibility assurance of compilation on supercomputers as a computing center production of domain workflow managers as part of funded projects	JP (U Osaka): working on container job trace
Hardware Heterogeneity	X (performance / efficiency issue for the online part, portability issue for the offline part)	X (large interest in High Level Frameworks for heterogeneous computing (KOKOS, Apaka, SYCL...))	X	X	Main simulation tasks currently run	Hardware Heterogeneity	X (use of multiple data centers in emergency)		X		X	X (the infrastructure is very heterogeneous by nature)			
Cybersecurity and Privacy	X (mostly for exploitation of final data products offline)	X (need federated authentication/identity management compatible with the WLOG trust model - which is adhering the AARC Blueprint)		X		Cybersecurity and Privacy			X		X				

Shared challenge priorities

International Post-Exascale Initiative

Strong themes across updated known use cases

- Federated access, policy alignment and long-term resource commitments
- End-to-end workflow control across HPC, cloud, edge and instruments
- Data logistics, transient storage, standard transfers and provenance
- Portability across GPUs/accelerators and heterogeneous platforms
- Scheduling / resource allocation that matches scientific production timelines
- Reliability, reproducibility, sustainability and energy efficiency

Emerging themes from new use cases

- Secure AI workflow orchestration and model lifecycle management
- Sensor/instrument-to-HPC streaming and near-sensor versus HPC trade-offs
- Operational federation for LAC/Brazil, including identity, data cloud and edge
- Sustainable infrastructure testbeds and environmental operation metrics
- Agentic workflows, digital twins and real-time actuation/decision support
- Zero-trust / privacy and adaptive multiscale storage for hybrid environments

Key next action: a validated mapping from each challenge to concrete blockers, owners, projects and milestones

Next steps

International Post-Exascale Initiative

Make the matrix usable as a collaboration and roadmapping tool

1. Fill the weak columns

Fill missing data

2. Validate with use-case owners

Replace generic "X" marks by concrete blockers, required capabilities and initial approaches.

3. Prioritize shared workstreams

Start with most cited challenges: workflow control, scheduling/resource allocation, federation/access, data logistics/streaming, heterogeneity/portability and reliability.

4. Define paper outline, write paper

Synchronize to validate the paper outline, asynchronous collaborative writing process

5. Convert priorities into actions

Synergy to SPE-EU, ETP4HPC, SPECTRUM/SRIDA, ODISSEE, ESIWACE4, NDP, SCALAC and national infrastructure/funding initiatives.

Immediate operational action: schedule a first online follow-up within 6 weeks
In preparation, freeze an owner-validated version of the matrix before prioritization.