RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

ExaMA Work Package 7 -- Feedback Session

Benchmarking

A. Calloo & C. Prud'homme

January 21, 2026 -- NumPEx General Assembly

ExaMA -- Exa-scale Methodologies and Algorithms

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030
PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

cea cnrs Inria

Key Takeaways

Key Takeaways from the GA

- **Modus operandi 2026**: WP7 engineers provide the *benchmark-facing skeleton* (CI/CB, packaging/containers, reporting), while research teams own *app specification, scientific content, and verification* (is my problem correctly solved?).
- **Deliverable-driven scope**: expand to additional frameworks over the coming months (FreeFEM++, HPDDM/PETSc, Samurai, TRUST, …) with multiple apps in the benchmarking pipeline; start small/medium-scale now, keep the path to exascale at project end.
- **Enablers and risks**: establish a shared dataset repository for Exa-MA (UNISTRA can provide; exploring NumPEx-level alternative), support heterogeneous packaging (containers first where needed), and strengthen GPU + observability (profiling/trace) to move beyond “run” toward “diagnose”.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030

PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

cea cnrs Inria

Discussions Summary

Main Discussion Points

- **Framework and app selection**: ensure partner representation (notably CEA) and avoid an "explosion" of apps; keep a curated set of representative scenarios connected to the actual research developments.
- **Execution and reproducibility**: converge on a pragmatic packaging strategy while keeping automated CI/CB execution, frozen configs, baselines, and regression detection.
- **From benchmarking to performance engineering**: extend the pipeline with analysis capabilities (profiling, MPI/IO traces) so that benchmark results can drive bottleneck identification and follow-up actions.

See ? feelpp.benchmarking for automated benchmarking (? GitHub ? PyPI)

Status Updates (Mon 19/01) — Inputs to WP7 Planning

- **HPC@Maths (Samurai / xtensor)**: GPU roadmap (Kokkos + CUDA), load balancing for adaptive meshes, packaging (Spack/Guix), first benchmark alignment with Exa-DI.
- **HPDDM / PETSc (WP3 interface)**: migrate dense linear algebra to PETSc objects to get portable GPU backends; next steps include PCHPDDM performance evaluation, dense Kokkos matrix support, and app-hpddm-dfn benchmarking.
- **Feel++ benchmarking**: reframe-based automation with reporting (JSON → AsciiDoc → HTML), dashboards, and Apptainer + Girder integration; current focus on benchmarking Feel++ mini-apps/demonstrators and hardening large-scale runs. See GitHub and PyPI.
- **Composyx**: presentation of the Composyx solver library with many examples and documentation; ready to integrate into the benchmarking pipeline.

Interfaces with Other WPs (inputs needed / feedback captured)

WP1-6 — Provide candidate mini-app specs + reference outputs for benchmarks; confirm metrics and scaling targets.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030

PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

cea cnrs Inria

Action Items

Action Items Identified

Action	Owner	Deadline
Confirm WP7 modus operandi (roles, contribution rules, rhythm, repo workflow)	Ansar & Christophe	02/2026
Select and freeze the initial benchmark set (3–4 frameworks, representative apps, configs)	WP7 core	05/2026
Define baseline results + regression criteria for the initial benchmarks	Javier + framework owners + app owners	03/2026
Data publication: agree on Girder layout (public/ private) + Zenodo archival policy	Christophe + Javier + Erik	03/2026
GPU readiness plan: Samurai, HPDDM, Feel++	A. Hoffmann, S. Dubois, Erik, Javier	06/2026
Integrate profiling/trace tooling (MPI/IO) into reports	Javier + Arthur	06/2026

New Collaborations

- **Samurai GPU enablement**: collaboration track with CExA (Kokkos) and NVIDIA (CUDA) to converge toward a stable GPU-capable release.
- **Cross-WP benchmarking pipeline**: leverage Feel++ benchmarking components as a NumPEx-level asset (ReFrame + reporting + dashboards) and connect to WP2 profiling initiatives to add observability to WP7 reports.
- **HPDDM/PETSc ecosystem**: align GPU-enabled dense linear algebra and app-hpddm-dfn benchmarking with WP3 discussions; contribute test datasets/ results via the shared data platform.
- **European opportunity**: monitor the (small) EU call on benchmarking software infrastructure mentioned during discussions; use it to reinforce sustainability beyond Exa-MA.

Adjustments

Adjustments to WP7 Plans

- **Clarify responsibilities**: WP7 does not "write the science code"; it provides the benchmarking substrate (automation + reporting) and supports teams in integrating, profiling, benchmarking their apps.
- **Support multiple packaging modes**: container-first when needed (e.g., large legacy frameworks), while keeping Spack/Guix recipes where feasible for developer workflows.
- **Raise the bar on outputs**: evolve from "benchmark runs" to "benchmark + diagnosis" by planning for profiling/tracing artifacts in reports.

Thank you!