

The LEXIS approach to Searching and Managing FAIR Research Data in Converged HPC-Cloud-Big Data Architectures

Stephan Hachinger¹, Jan Martinovic², Olivier Terzo³, Cédric Koch-Hofer⁴, Martin Golasowski², Mohamad Hayek¹, Marc Levrier⁴, Alberto Scionti³, Donato Magarielli⁵, Thierry Goubier⁶, Antonio Parodi⁷, Sean Murphy⁸, Florin-Ionut Apopei⁹, Carmine D'Amico³, Simone Ciccia³, Massimo Sardo⁷, Danijel Schorlemmer¹⁰

¹Leibniz Supercomputing Centre (LRZ), Bavarian Academy of Sciences and Humanities, Garching, Germany

²IT4Innovations, VŠB - Technical University of Ostrava, Ostrava-Poruba, Czech Republic

³Advanced Computing and Applications, LINKS Foundation, Torino, Italy

⁴ATOS, Paris, France

⁵Avio Aero, Torino, Italy

⁶CEA, LIST, Paris, France

⁷CIMA Research Foundation, Savona, Italy

⁸Cyclops Labs GmbH, Winterthur, Switzerland

⁹TESEO, Torino, Italy

¹⁰GFZ German Research Centre for Geosciences, Potsdam, Germany

The enormous amounts of data generated in modern industry, business and science pose a significant challenge to those extracting actionable intelligence from data using various filtering and analysis techniques. In this "Big Data" setting, the LEXIS project (Large-scale EXecution for Industry & Society) provides a platform for optimised execution of Cloud-HPC workflows, reducing computation time and energy efficiency. The system will rely on advanced, distributed orchestration solutions (Bull Ystia Orchestrator, based on TOSCA and Alien4Cloud technologies), the High-End Application Execution Middleware HEAppE, and new hardware capabilities for maximizing efficiency in data processing, analysis and transfer (e.g. Burst Buffers with GPU- and FPGA-based data reprocessing).

LEXIS handles computation tasks and data from three Pilots, based on representative and demanding HPC/Cloud-Computing use cases in Industry (SMEs) and Science: i) Simulations of of complex turbo-machinery and gearbox systems in Aeronautics, ii) Earthquake and Tsunami simulations which are accelerated to enable accurate real-time analysis, and iii) Weather and Climate simulations where massive amounts of in situ data are assimilated to improve forecasts. A user-friendly LEXIS web portal, as a unique entry point, will provide access to data as well as workflow-handling and remote visualization functionality.

The "LEXIS Data System" constitutes the data back-end for the project. At its core, a Distributed Data Infrastructure (DDI) ensures the availability of LEXIS data at all participating HPC sites, and provides functionality for FAIR ("Findable, Interoperable, Accessible, Reusable") Research Data Management. The DDI leverages best of breed data-management solutions of EUDAT, including B2SAFE and B2HANDLE. Via DOI acquisition, open data products can be published and disseminated. Exposing metadata via standardized REST interfaces, the DDI will be a best-practice example for demonstrating the connection of Research Data Infrastructures to specialized and general-purpose search facilities (internal data search, B2FIND, GeRDI, web search engines). Making research data findable in such ways proves essential for data sharing and re-use within research communities.