# HPC-Cloud-Big Data Convergent Architectures and Research Data Management: The LEXIS Approach

Stephan Hachinger,[a,*] Jan Martinovič,[b] Olivier Terzo,[c] Marc Levrier,[d] Alberto Scionti,[c] Donato Magarielli,[e] Thierry Goubier,[f] Antonio Parodi,[g] Piyush Harsh,[h] Florin-Ionut Apopei,[i] Johannes Munke,[a] Rubén J. García-Hernández,[a] Martin Golasowski,[b] Mohamad Hayek,[a] Frédéric Donnat,[k] Laurent Ganne,[j] Cédric Koch-Hofer,[j] Giacomo Vitali,[c] Paolo Viviani,[c] Danijel Schorlemmer,[l] Emanuele Danovaro,[m] Andrea Parodi,[n] Seán Murphy[o] and Aaron Dees[p]

[a]*Leibniz Supercomputing Centre (LRZ) of the BAdW,*
*Boltzmannstr. 1, Garching b.M., Germany*

[b]*IT4Innovations National Supercomputing Center, VŠB – Technical University of Ostrava,*
*Studentská 6231/1B, Ostrava, Czech Republic*

[c]*Advanced Computing and Applications, LINKS Foundation,*
*Via Pier Carlo Boggio, 61, Torino, Italy*

[d]*Atos,*
*Campus Teratec, 2 rue de la Piquetterie, Bruyères-le-Châtel, France*

[e]*Avio Aero**,*
*Via I. Maggio, 99, Rivalta di Torino, Italy*

[f]*CEA LIST,*
*Bâtiment 565, Gif-sur-Yvette, France*

[g]*CIMA Foundation,*
*Via A. Magliotto, 2, Savona, Italy*

[h]*Cyclops Labs GmbH,*
*Neptunstrasse 63, Zürich, Switzerland*

[i]*TESEO S.P.A.,*
*Corso Alexander Fleming, 25, Druento, Italy*

[j]*Atos,*
*1 Rue de Provence, Échirolles, France*

[k]*Outpost24,*
*2405 Route des Dolines, Valbonne, France*

[l]*Helmholtz-Zentrum Potsdam – Deutsches GeoForschungsZentrum (GFZ),*
*Wissenschaftpark "Albert Einstein", Telegrafenberg, Potsdam, Germany*

[m]*European Centre for Medium-Range Weather Forecasts (ECMWF),*
*Shinfield Park, Reading RG2 9AX, United Kingdom*

[n]*School of Engineering, Zürcher Hochschule für Angewandte Wissenschaften (ZHAW),*
*Technikumstrasse 9, Zürich, Switzerland*

*Speaker

**A business of GE aviation.

*ᵖ Irish Centre for High-End Computing (ICHEC),*
*Tower Building, Technology & Enterprise Campus, Grand Canal Quay, Dublin 2, Ireland*

*E-mail:* hachinger@lrz.de

The LEXIS project (Large-scale EXecution for Industry & Society, H2020 GA825532) provides a platform for optimised execution of Cloud-HPC workflows, reducing computation time and increasing energy efficiency. The system will rely on advanced, distributed orchestration solutions (Atos YSTIA Suite, with Alien4Cloud and Yorc, based on TOSCA), the High-End Application Execution Middleware HEAppE, and new hardware capabilities for maximising 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 complex turbomachinery and gearbox systems in Aeronautics, ii) Tsunami simulations and earthquake loss assessments which are time-constrained to enable immediate warnings and to support well-informed decisions, 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 visualisation functionality.

As part of its back-end, LEXIS builds an elaborate system for the handling of input, intermediate and result data. At its core, a Distributed Data Infrastructure (DDI) ensures the availability of LEXIS data at all participating HPC sites, which will be federated with a common LEXIS Authentication and Authorisation Infrastructure (with unified security model, user database and policies). The DDI leverages best of breed data-management solutions from EUDAT, such as B2SAFE (based on iRODS) and B2HANDLE. REST APIs on top of it will ensure a smooth interaction with LEXIS workflows and the orchestration layer. Last, but not least, the DDI will provide functionalities for Research Data Management following the FAIR principles ("Findable, Interoperable, Accessible, Reusable"), e.g. DOI acquisition, which helps to publish and disseminate open data products.

# 1. Introduction

When researchers from science and companies explore large-scale computing, data science and Big Data solutions, they will quickly discover two separate ecosystems. On the one hand, an IT service-based ecosystem, mostly "Cloud native" and backed by tech giants such as Google and Amazon. On the other hand, an ecosystem of raw High-Performance-Computing (HPC) power on the world's top supercomputers, unique machines which are notoriously difficult to work on. This split in philosophies has generated absurd situations such as scientists not understanding the notion of reliable, cloud-based IT services, or industrial users stuck with isolated virtual machines and databases and unable to parallelise their workloads. All in all, mixed access to Cloud and HPC systems has remained a troublesome topic for users and involves two separate, sometimes steep, learning curves.

In this setting, the EU project "Large-Scale Execution for Industry & Society" (LEXIS [1], H2020 Grant Agreement no. 825532) supports scientific and industrial users with demanding, data-driven applications with a platform addressing basic demands: orchestration of computational/data-driven workflows and research data management. The workflows with automated execution and data handling are run on a rich, geographically distributed backend of Cloud, HPC and data systems at internationally recognised computing and data centres: IT4Innovations National Supercomputing Center (IT4I, Ostrava, CZ), Leibniz Supercomputing Centre (LRZ, Garching b.M., D), European Centre for Medium-Range Weather Forecasts (ECMWF, Reading, UK) and Irish Centre for High-End Computing (ICHEC, Dublin, IRL). Thus, systems with an aggregate computing power of more than 30 PFlop/s are made available via the LEXIS portal in a "HPC-as-a-Service" (and, clearly, Infrastructure-as-a-Service-/IaaS-Cloud) approach.

To make best use of its computing backend, LEXIS uses a novel, open-source orchestration software from ATOS (Yorc – YSTIA Orchestrator, [2]). While Yorc workflows are based on the standard TOSCA language (cf. e.g. [3]), programming is not necessary to specify workflows, as a convenient web UI with template components is provided – Alien4Cloud (Application LIfecycle ENablement for Cloud, [4]). The actual authentication and job submission bridge for HPC systems in this context is provided by the "HEAppE" middleware from IT4I (High-End Application Execution Middleware, [5, 6]). With this setup, the LEXIS Orchestration System can not only execute jobs on the OpenStack IaaS-Cloud infrastructures of IT4I and LRZ, but also utilise arbitrary HPC-system accounts to execute jobs on existing HPC computing time grants on behalf of the user. Together with a LEXIS Authentication and Authorisation Infrastructure (AAI), based on the Keycloak Identity and Access Management solution [7], the Yorc-HEAppE setup allows for geographically distributed computing with a single sign on (SSO), in the tradition of international grid-computing infrastructures (e.g. [8, 9]).

The LEXIS "Distributed Computing Infrastructure" (DCI) clearly needs a counterpart on the data side – i.e. a "Distributed Data Infrastructure" (DDI), via which data at the multiple centres can be accessed, read and written in a uniform manner. In a European/EOSC context, the B2SAFE framework of the EUropean DATa initiative (EUDAT [10, 11]) aims to implement this, also fulfilling basic prerequisites for FAIR ("Findable, Interoperable, Accessible, Reusable", cf. [12]) Research Data Management. Based on the Integrated Rule-Oriented Data System (iRODS [13]), B2SAFE provides functionalities e.g. for cross-site data replication. With iRODS and B2SAFE installed at

the LEXIS sites, and the sites forming an iRODS federation, data are referenced via one uniform "directory tree" across all data centres. Yet, iRODS allows the LEXIS Orchestration System to identify the location(s) where data are actually stored – which is, e.g., important for workflows following a "compute-to-data" paradigm, or for a dynamic, automated choice of computing resources during the run, taking into account the load of different systems/clusters. For accelerating data transfers and data compression, decompression, encryption and decryption, "Data Nodes" with large NVDIMM and NVMe storage were installed at each site, which can be used directly or as software-controlled Burst Buffers.

In order to properly implement and convey the user-friendly and service-based concept of LEXIS, a portal has been programmed, which not only allows for the control of workflows and data, but also implements business functionality that industrial and SME users would expect. This includes fine-granular accounting and billing as well as access-rights management, as envisaged in the LEXIS security and data handling concepts.

The integrative approach of LEXIS supports "Big Data, Data Science and Simulations for everyone", by lowering usage thresholds and offering a platform where not only Cloud and HPC system usage converges, but also the computing patterns of industrial, SME and academic users are advanced such that convergence and knowledge transfer are facilitated. This is also reflected in the mix of three LEXIS Pilots (groups from science and industry providing first use cases), with which the platform is co-developed. The Pilots deal with: i) data-intensive and computationally demanding modelling of turbo-machinery and gearboxes in Aeronautics, ii) Earthquake and Tsunami data processing and simulations, which are accelerated to enable accurate real-time analysis, and iii) Weather and Climate models where massive amounts of in situ data are assimilated to improve forecasts (also predicting, e.g., flash flood hazard or agricultural yields).

In this contribution, we give some more details on the status of LEXIS by the beginning of 2021, covering the LEXIS Portal (Section 2), Orchestration System (Section 3) and DDI (Section 4). We conclude with a short discussion of the LEXIS Open Call (for test users), which we have issued (Section 5) in order have our platform systematically tested and validated beyond the original Pilots, and a summary/outlook (Section 6).

## 2. LEXIS Portal

The LEXIS Portal is the unified entry point for accessing the distributed computing and data infrastructure of the platform. Input data for workflows, workflow specifications and uploads of necessary processing software (preferably as containers) are exclusively provided to the platform via the portal. Also, the portal handles downloads of results from the workflows. In cases that up- and downloads need to be driven by machines or are too big for usual HTTP-based transfers, the portal gives the user links and instructions to provide/retrieve the data, e.g., via GridFTP [14].

Furthermore, the portal serves as entry point into the system with regards to authentication. Once a user successfully logs in and thus authenticates against the LEXIS AAI, he receives an access token, which can be passed on to lower-level LEXIS services such as the DDI. Each lower-level component can thus obtain a minimalist, specific access token from the AAI, as well as information on the users' access rights. Users are members of so-called LEXIS (computational) projects, i.e. collaborations with common access to collaboration data and computing-time grants. For so-called

**Figure 1:** Screenshot of an early test version of the LEXIS Portal, with the workflow-management interface activated.

"project managers" and other administrative roles, the AAI will grant elevated rights, for example the right to add a user to a project.

The portal is based on custom programming in Go ([15], backend) and React ([16], frontend) for our special purposes. It features interfaces for data set management, user/organisation (self-)management, workflow management and finally project management. The latter includes the possibility to view consumed resources and data relevant to billing (where clearly a price model is yet to be established, as the construction of the LEXIS platform is finished up). The necessary data for fine-grained accounting and billing are collected with the Cyclops system [17], whose development is driven by project partner Cyclops Labs.

An impression of the portal (early test version) is given in Figure 1, showcasing the workflow management interface.

## 3. Platform overview and Orchestration System

The LEXIS infrastructure consists of HPC and OpenStack-based Cloud-Computing systems at LRZ, IT4I and ICHEC, as well as data systems at ECMWF. Through the European GÉANT network [18], the interconnect between the centres consents transfer rates in the range of 10 Gbit – 100 Gbit at normal days, clearly supporting distributed, data-driven workflows (as long as transfers of lots of very small pieces of data are avoided). All centres except ECWMF are connected to the DDI federation for data exchange, while all centres except ICHEC run servers of the LEXIS Weather and Climate Data API (WCDA), where curated weather and climate data (for the respective Pilot
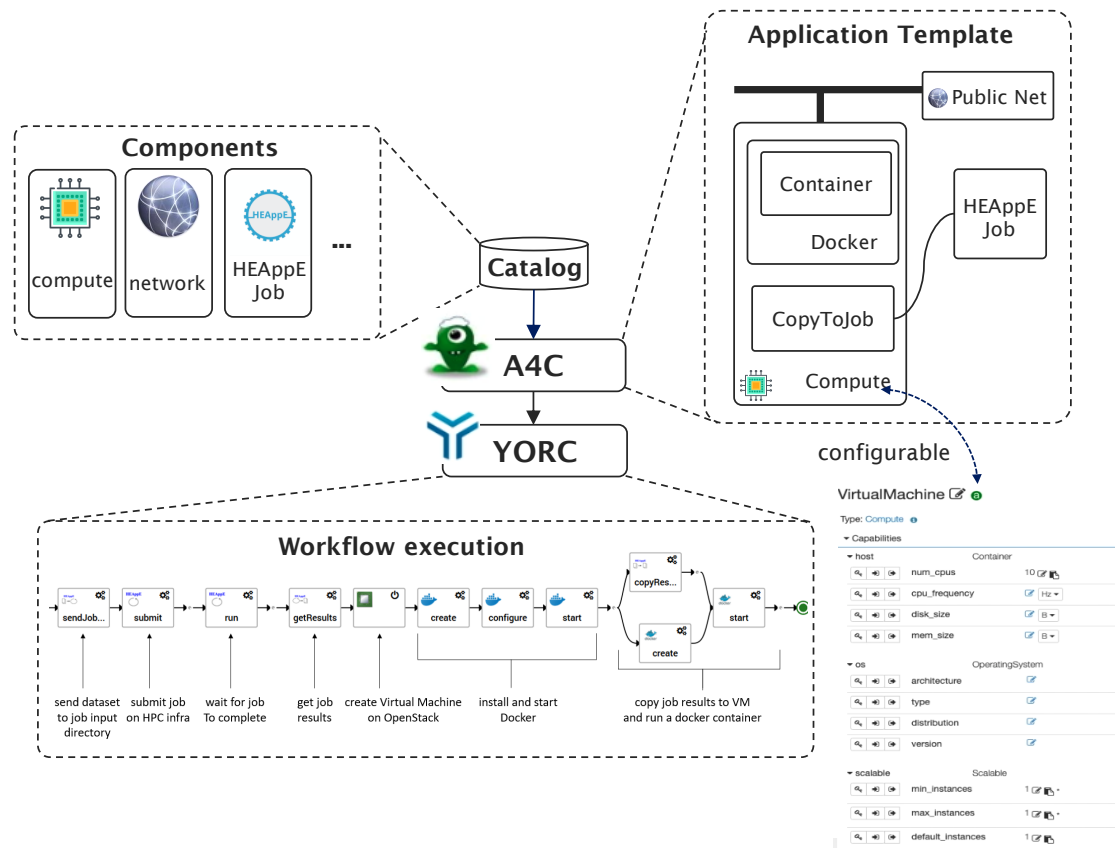
5

**Figure 2:** Defining and running a LEXIS workflow with Alien4Cloud (A4C) and Yorc: The Figure illustrates the components and steps involved (Catalog, Application Template, Workflow Description/Execution), as further discussed in the text. The application template contains configurable details on infrastructural requirements (lower/right part of the Figure), such that Yorc can place all tasks on appropriate resources.

and beyond) can be up- and downloaded. Clearly, DDI and WCDA clients can be run at any centre and also at the sites of further LEXIS partners if needed.

All LEXIS infrastructure components are strictly accessed via web APIs (HTTP-REST or similar) by both the portal (Section 2) and the LEXIS Orchestration System. Storage resources are accessed and controlled via the WCDA and the DDI APIs (cf. Section 4), and computing resources via the HEAppE and OpenStack APIs.

LEXIS users define their workflow components using Alien4Cloud and run the workflows triggering Yorc via the LEXIS Portal. The service accepts LEXIS AAI tokens and passes these on to underlying systems where necessary. Figure 2 illustrates the necessary steps: In the Alien4Cloud-based UI, pre-defined components (from the Catalog, upper/left part of Figure 2) are used to describe the characteristics of infrastructure elements used, as well as a few key action entities needed in workflows (e.g., a HEAppE job, or a copy action from the DDI to the input file location of an HPC job). With the Application Template thus created (upper/right part of Figure 2), the user can

proceed to the final step: Defining the actual workflow, i.e. the sequence of actions (e.g. start a HPC code, wait for the code to finish, save the data, run a postprocessing and visualisation code on the Cloud, and so on – cf. lower left part of Figure 2), and starting a Workflow Execution.

## 4. Distributed Data Infrastructure (DDI)

In order to run cross-datacentre workflows and safeguard LEXIS data, the LEXIS DDI was established, based on iRODS and EUDAT modules, as mentioned in the introduction. This infrastructure stores input, intermediate and output data of LEXIS workflows and provides unified access from all computing centres. It is accessed via REST-APIs built within the LEXIS project. One such API triggers data staging between the DDI and production computing infrastructures of the participating Supercomputing centres (and their attached cluster file systems) for running computations, which usually requires centre-specific transfer mechanisms and configurations. Encapsulating these mechanisms, our DDI-based data management concept for LEXIS workflows decouples from the specialities and characteristics of data centres joining the LEXIS federation. It does practically not interfere with local data management policies in place (except for the fact that the DDI clearly uses some dedicated backend storage). iRODS allows for transparent cross-centre data access (and a unified directory-tree-like view), and – with the EUDAT-B2SAFE module – also for policy-based data replication and mirroring to reduce access times. In addition, the LEXIS DDI actively supports FAIR [12] Research Data Management, as it is indispensable nowadays. For confidential corporate data, the DDI features a sophisticated rights-management concept, going hand in hand with the possibility of preparing a FAIR data publication at a later time.

Figure 3 illustrates the LEXIS DDI federation as it was built in 2019/20 (which is now extended to ICHEC, and also other European B2SAFE sites in the course of the LEXIS Open Call – cf. Section 5). The iRODS data-system middleware provides cross-centre (cross-iRODS-zone) access to data via its internal data fetching and exchange mechanisms. Manually or with EUDAT-B2SAFE, data prefetching and mirroring policies can be implemented for data within workflows or at rest, utilising the iRODS rule engine (cf. [13]). The prime example to this will be a possibility to request data geo-replication in the LEXIS Portal, in order to make data available faster and significantly increase data safety.

The fundamental concept of the DDI being accessed via APIs reflects in the staging mechanism to/from the computing facilities within LEXIS. Building on EUDAT-B2STAGE and other back-end transfer mechanisms (SCP, SFTP, ...), we have implemented an API receiving data-staging requests and pushing them into a Celery/RabbitMQ-based queuing system [21, 22], where they are processed asynchronously. The requests contain AAI-token data, source and target systems/paths and data set specification (including an unique ID identifying the data set in LEXIS). The result of a request (ongoing, executed, failed) can be checked by querying a separate, dedicated API endpoint. The API concept makes it easy to, e.g., optimise data-transfer mechanisms while keeping the interface to the orchestrator stable. At the moment, mechanisms to include the LEXIS Data Node / Burst Buffer servers (cf. Introduction; two machines per centre at IT4I and LRZ) in the data flow are developed. The Burst Buffers feature novel memory modules with large capacity (Intel Optane DC NVDIMMs) and NVMe SSDs, where data can not only be buffered, but also be converted,
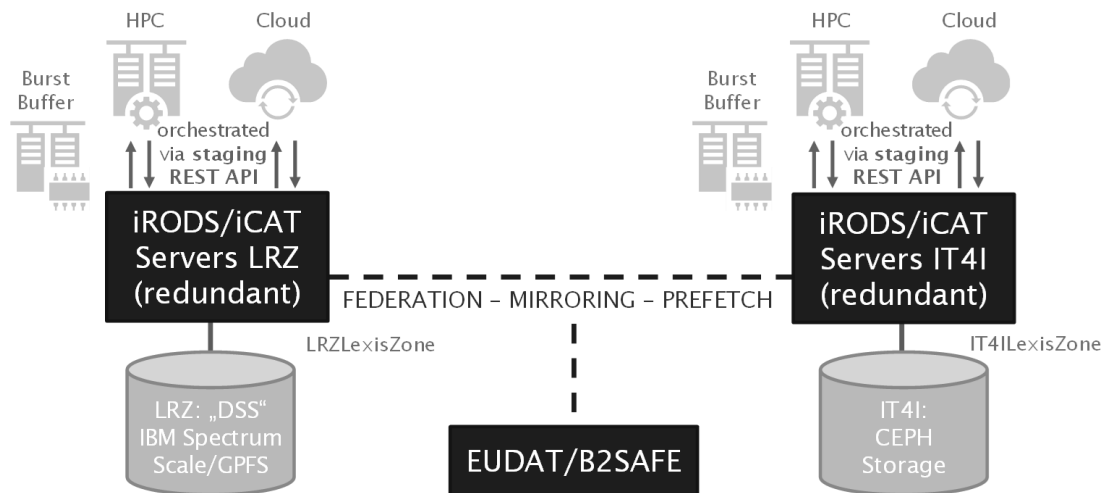
**Figure 3:** Federation of IT4I and LRZ within the LEXIS DDI, based on the iRODS data-system middleware and EUDAT-B2SAFE. The back-end systems for data storage at LRZ and IT4I are cluster file systems (IBM Spectrum Scale / ex GPFS [19] and Ceph [20]) with suitable hardware behind. Each centre runs a so-called iRODS zone with its own, independent metadata catalogue (iCAT / iRODS "provider server"), and data are transparently accessible via the inter-zone federation. Data staging to cloud and HPC systems is controlled via the staging API of the DDI, with Burst Buffers providing temporary, cache-like storage (see text).

compressed, encrypted or decrypted in place. Thus, encrypted handling of data on the DDI can be implemented for industry-oriented use cases.

The FAIR data management functionalities of the DDI focus on PID assignment with EUDAT-B2HANDLE [23] and assignment of DataCite-compatible metadata (cf. [24, 25]) in the builtin metadata store for each file/directory ("data object"/"collection" in iRODS terminology) in iRODS (cf. [13]). These guarantee data findability and unambiguous data set identification within LEXIS. Users can search for their data via a full-text search over selected metadata fields in the LEXIS Portal (which, to this purpose, queries a data discovery API), and thus immediately experience the usefulness of proper metadata enrichment. Prospectively, we aim for data discovery via the EUDAT-B2FIND [26] search portal, and – on user request – for the assignment of DOIs to data in addition to the B2HANDLE PIDs already assigned.

## 5. LEXIS Open Call

In 2020 and 2021, the LEXIS project has been inviting further test users in order to validate the LEXIS platform within realistic usage scenarios. Thus, it has launched an Open Call, to which test users can respond on the LEXIS project web site [1], and which is disseminated via various media. These users will be guided through the platform by one collaborative "project manager" on the LEXIS side each. The call is part of an extensive benchmarking, validation and KPI

evaluation strategy which the project runs in its last year (2021), as a basis for sustainability and future enhancement concepts. The Open Call is not only used to attract academic, industrial, and SME customers, but also to launch collaborations with compute- and data-intensive EU projects as platform users. As a result, e.g., a collaboration with the CompBioMed project (H2020 Grant Agreements no. 675451, 823712) has been initiated.

## 6. Summary and Outlook

In our contribution we presented the LEXIS project, building a novel platform for running HPC-/Cloud-based computing and Big Data workflows, in a co-design approach with three Pilot use cases (aeronautics, earthquake/tsunami and weather and climate simulations). With a sophisticated, but easy-to-use Orchestration and FAIR Research Data Management system, we address the needs of industrial engineering and academic science communities alike. The LEXIS Portal as a "one stop shop" allows users to manage data and workflows, and to view their resource usage, or, as commercial users, also their billing data.

In setting up the platform, care was taken of security and data privacy aspects, but also of data sharing and collaboration opportunities with other European projects. In line with this, tools provided or endorsed by the EUDAT Collaborative Data Infrastructure (EUDAT CDI) were chosen for building the LEXIS Distributed Data Infrastructure. Leveraging these tools, a future data federation with other EUDAT-based data sites is envisaged.

Currently, the LEXIS platform is extensively validated through an Open Call for test users, and through technical benchmarking. With refinements, a service management strategy (documentation, technical support ticketing system, LEXIS monitoring system), and the clear goal to drive convergence between HPC and Cloud Computing, but also convergence and collaboration between industrial and academic research, we are making the LEXIS platform sustainable and future-proof.

Further details on LEXIS and the systems discussed above are published in a series of conference contributions and book chapters (e.g. [5, 27–29]). Open software and documents developed in the project will be disseminated via GitHub, Zenodo (dedicated sections: [30, 31]) and the official EU project web pages [32].

## Acknowledgements

## References

[1] LEXIS consortium, "Lexis Project - High Performance Computing (HPC) in Europe." https://lexis-project.eu, Cited 27 Apr 2021, 2021.

[2] Bull/Atos, "Ystia Suite." https://ystia.github.io, Cited 6 Nov 2020, 2020.

[3] A. Brogi, J. Soldani and P. Wang, *TOSCA in a Nutshell: Promises and Perspectives*, in *Service-Oriented and Cloud Computing*, M. Villari, W. Zimmermann and K.-K. Lau, eds., (Berlin, Heidelberg), pp. 171–186, Springer, 2014, DOI.

[4] FastConnect, Bull/Atos, "Alien 4 Cloud." http://alien4cloud.github.io, Cited 6 Nov 2020, 2020.

[5] V. Svaton, J. Martinovič, J. Krenek, T. Esch and P. Tomancak, *HPC-as-a-Service via HEAppE Platform*, in *CISIS 2019*, L. Barolli, F.K. Hussain and M. Ikeda, eds., Advances in Intelligent Systems and Computing, vol. 993, (Cham), pp. 280–293, Springer, 2019, DOI.

[6] V. Svaton, "Home · Wiki · ADAS / HEAppE / Middleware · GitLab." http://heappe.eu, Cited 6 Nov 2020, 2020.

[7] JBoss (Red Hat Inc.), Keycloak Community, "Keycloak." https://www.keycloak.org/, Cited 6 Nov 2020, 2020.

[8] J. Shiers, *The Worldwide LHC Computing Grid (worldwide LCG)*, *Computer Physics Communications* **177** (2007) 219.

[9] E. Imamagic and T. Ferrari, *EGI grid middleware and distributed computing infrastructures integration*, in *Proceedings of the International Symposium on Grids and Clouds (ISGC) 2013 – PoS*, vol. 179, 010, 2014, DOI.

[10] D. Lecarpentier, P. Wittenburg, W. Elbers, A. Michelini, R. Kanso, P.V. Coveney et al., *EUDAT: A New Cross-Disciplinary Data Infrastructure for Science*, *Int. J. Digit. Curation* **8** (2013) 279.

[11] EUDAT Collaborative Data Infrastructure, "EUDAT - Research Data Services, Expertise & Technology Solutions." https://www.eudat.eu, Cited 6 Nov 2020, 2020.

[12] M.D. Wilkinson et al., *The FAIR Guiding Principles for scientific data management and stewardship*, *Scientific data* **3** (2016).

[13] H. Xu, T. Russell, J. Coposky, A. Rajasekar, R. Moore, A. de Torcy et al., *iRODS primer 2: Integrated Rule-Oriented Data System*, Morgan & Claypool Publishers, Williston, VT (2017), DOI.

[14] W. Allcock, J. Bresnahan, R. Kettimuthu and M. Link, *The Globus Striped GridFTP Framework and Server*, in *SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*, pp. 54–54, 2005, DOI.

[15] Google and Go Contributors, "The Go Programming Language Specification – The Go Programming Language." https://golang.org/ref/spec, Cited 6 Nov 2020, Jan, 2020.

[16] Facebook Inc. and React contributors, "React – A JavaScript library for building user interfaces." http://reactjs.org, Cited 6 Nov 2020, 2020.

[17] Cyclops Labs GmbH, "it-accounting-and-billing | Cyclops Labs GmbH." https://www.cyclops-labs.io/page/it-accounting-and-billing, Cited 26 Apr 2021, 2020.

[18] P.T. Kirstein, *European international academic networking: A 20 year perspective*, in *One step ahead, The 20th Trans European Research and Education Networking Conference, June 7-10, 2004, Rhodes, Greece, Selected Papers*, D. Chadwick et al., eds., TERENA, 2004.

[19] F.B. Schmuck and R.L. Haskin, *GPFS: A Shared-Disk File System for Large Computing Clusters*, in *Proceedings of the Conference on File and Storage Technologies*, FAST '02, (US), p. 231–244, USENIX Association, 2002.

[20] S.A. Weil, S.A. Brandt, E.L. Miller, D.D.E. Long and C. Maltzahn, *Ceph: A Scalable, High-Performance Distributed File System*, in *Proceedings of the 7th Symposium on Operating Systems Design and Implementation*, OSDI '06, (US), p. 307–320, USENIX Association, 2006.

[21] Celery, "Celery: Distributed Task Queue." http://www.celeryproject.org, Cited 1 Apr 2020, 2020.

[22] Pivotal Software, "Messaging that just works - RabbitMQ." https://www.rabbitmq.com, Cited 1 Apr 2020, 2020.

[23] EUDAT Collaborative Data Infrastructure, "B2HANDLE-EUDAT." https://www.eudat.eu/services/b2handle, Cited 6 Nov 2020, 2020.

[24] J. Starr and A. Gastl, *isCitedBy: A Metadata Scheme for DataCite*, *D-Lib Magazine* **17** (2011), DOI.

[25] DataCite Metadata Working Group, DataCite e.V., "DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4." http://schema.datacite.org/meta/kernel-4.4, 2021. DOI.

[26] EUDAT Collaborative Data Infrastructure, "B2FIND-EUDAT." https://www.eudat.eu/services/b2find, Cited 6 Nov 2020, 2020.

[27] A. Scionti et al., *HPC, Cloud and Big-Data Convergent Architectures: The LEXIS Approach*, in *CISIS 2019*, L. Barolli, F.K. Hussain and M. Ikeda, eds., Advances in Intelligent Systems and Computing, vol. 993, (Cham), pp. 200–212, Springer, 2019, DOI.

[28] A. Parodi et al., *LEXIS Weather and Climate Large-Scale Pilot*, in *CISIS 2020*, L. Barolli, A. Poniszewska-Maranda and T. Enokido, eds., Advances in Intelligent Systems and Computing, vol. 1194, (Cham), pp. 267–277, Springer, 2020, DOI.

[29] T. Goubier, J. Martinovič, P. Dubrulle, L. Ganne, S. Louise, T. Martinovič et al., *Real-Time Model of Computation over HPC/Cloud Orchestration - The LEXIS Approach*, in *CISIS 2020*, L. Barolli, A. Poniszewska-Maranda and T. Enokido, eds., Advances in Intelligent Systems and Computing, vol. 1194, (Cham), pp. 255–266, Springer, 2020, DOI.

[30] LEXIS consortium, "GitHub – LEXIS: Large Scale Execution for Industry & Society." https://github.com/lexis-project, Cited 6 Apr 2021, 2021.

[31] LEXIS consortium, "Zenodo Community – LEXIS project." https://zenodo.org/communities/lexis, Cited 6 Apr 2021, 2020.

[32] European Commission, "Large-scale EXecution for Industry & Society | LEXIS Project | H2020 | CORDIS | European Commission." https://cordis.europa.eu/project/id/825532/results, Cited 27 Apr 2021, 2021.

PoS(ISGC2021)004