HPC, Big Data, and AI Convergence Towards Exascale

Challenge and Vision

Edited by
Olivier Terzo
Jan Martinovič

First edition published 2022

ISBN: 978-1-032-00984-1 (hbk)
ISBN: 978-1-032-00991-9 (pbk)
ISBN: 978-1-003-17666-4 (ebk)

2

The LEXIS Platform for Distributed Workflow Execution and Data Management

Martin Golasowski, Jan Martinovič, Jan Křenek, Kateřina Slaninová, Marc Levrier, Piyush Harsh, Marc Derquennes, Frederic Donnat, and Olivier Terzo

DOI: 10.1201/9781003176664-2

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

2

The LEXIS Platform for Distributed Workflow Execution and Data Management

Martin Golasowski, Jan Martinovič, Jan Křenek, Kateřina Slaninová, Marc Levrier, Piyush Harsh, Marc Derquennes, Frédéric Donnat, and Olivier Terzo

CONTENTS

2.1 Motivation

Many scientific and industrial fields can benefit from easier access to computing resources whose power rapidly increases. The LEXIS platform aims to overcome the shortcomings of complicated access to a high-performance computing (HPC) system and management of terabyte-scale datasets by providing a clean and concise interface which will abstract both HPC and cloud resources using workflow and data management orchestration [1]. This chapter provides an overview of the LEXIS platform with the focus on the overall architecture and the most important concepts the platform implements from the security, accounting, and usability points of view. A market analysis is also provided to lay ground for a business case and future exploitation of the solution.

DOI: 10.1201/9781003176664-2

The platform runs workflows comprising many dependent tasks which can be executed simultaneously on various computing resources such as cloud systems or geographically distant HPC clusters. This way, the platform can make use of cloud systems deployed in the same data center as HPC cluster to significantly decrease the time required for data staging and migration while leveraging the best features provided by the cloud systems. The implemented orchestration solution is described in Chapter 5.

The platform implements a distributed data interface (DDI) [2] which leverages EUDAT [3] and iRODS technologies [4] to enable seamless and secure ingestion, staging, curation, and publication of data. The DDI also offers in-place encryption of the data to provide the highest level of security for the most sensitive data. This interface is used by the platform users to upload their datasets and download their results. It can also be used to move data between different computing centers connected to the platform, essentially moving terabytes of data by a single click. The interface is used by the orchestrator as well, to automatically move the data to the target computing resource. Detailed description of the DDI is in Chapter 4.

All APIs implemented by the LEXIS platform are based on the REST HTTP protocol making it easy to integrate it to existing systems. Using API to allocate and use a computing resource is an inherent property of cloud computing. The LEXIS platform uses this fact as inspiration to bring the similar level of usability also to the world of HPC and Big Data. Part of the platform is a web-based graphical user interface – the LEXIS Portal, which uses the APIs to expose the platform in a user-friendly way. An important part of the platform is also an accounting and billing system which is used to track resources consumed by the platform users in each connected center.

The platform provides its own authentication and authorization interface (AAI) which uses common technologies such as OpenID and JWT tokens [5]. HPC centers and operators offer hosts their own identities used to access their system. The LEXIS platform has a solution to overcome this restriction by using the HEAppE middleware [6], which is deployed in each center which provides its resources for the platform. The HEAppE middleware implements an API for job submission and control through a predefined command template. The API can be used only by authenticated LEXIS users and the middleware uses local service accounts of the HPC center to communicate with the cluster job schedulers, thus making the local HPC AAI opaque to the LEXIS Users.

2.2 Architecture (Codesign) and Interfaces

LEXIS architecture is a complex system formed from small components (blocks) and relations between them. The design approach provides loosely coupled components, which makes them easy to modify, replace, or extend without breaking the entire system.

Component versioning is necessary to maintain consistency of the component APIs. It provides control of their dependencies, where two components must use the same API in compatible versions in order to work correctly. The system defines the versioning scheme and version compatibility.

Exposing LEXIS platform as a set of APIs allows us to create for example mobile applications (Android, iOS) or integrate the platform in existing enterprise resource planning systems. These applications will use the same LEXIS back end which is currently used by the LEXIS portal user interface.

The LEXIS platform has a three-layer architecture where each of them contains its own software components. A top-level view of the architecture is in Figure 2.1. The three layers are described below.

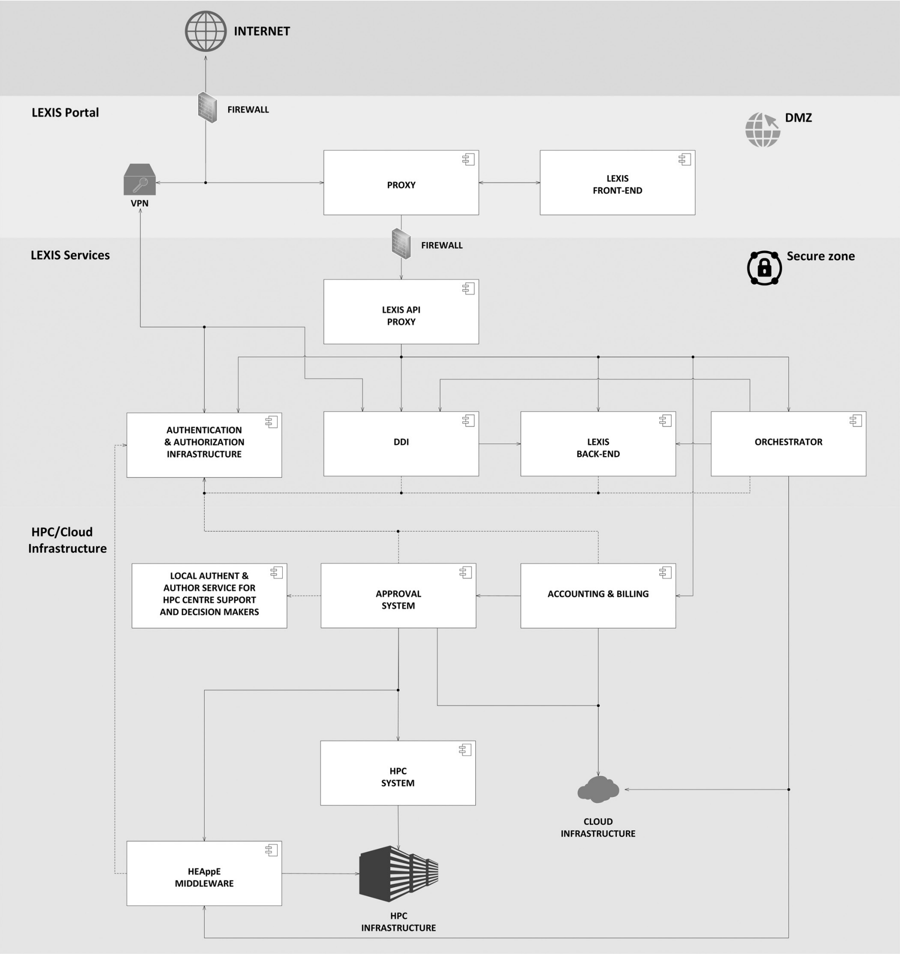


FIGURE 2.1
High-level architecture of the LEXIS platform with three-layer separation.

The LEXIS Portal Layer, providing easy access to the LEXIS platform for pilots and possible external users, contains the front-end (portal) which is the main entry point to the LEXIS platform. It is designed primarily for the users who do not necessarily have deep experience with HPC or cloud environments, so they can have an easy way to access the most advanced capabilities and features of the platform.

The LEXIS services layer lies in the middle between the portal and infrastructure. It includes federated security infrastructure (authentication and authorization infrastructure), data management (distributed data infrastructure (DDI)), and orchestration services (orchestrator).

Authentication and authorization infrastructure (AAI) is based on Keycloak which is an open-source solution. This specific core component of the architecture is in charge of handling the authentication of the LEXIS users or processes and of the authorization according to the role assignment and RBAC matrix.

The DDI component is a distributed data storage, providing unified access to data from all participating institutes and computing resources. The DDI provides several interfaces for data access, including REST API, native iRODS protocol, or GridFTP [7].

The orchestration service represents one of the key technological pillars of the LEXIS platform. It provides the features that enable LEXIS users to run their workflows using federated resources available in one or more HPC service providers (starting with IT4I and LRZ). The LEXIS orchestration service is based on the integration of several modules into a functional component. Specifically, the service architecture uses Alien4Cloud (A4C) [8] as the front end of the Yorc [9] orchestration service. The architecture also contains a monitoring module and a business logic module providing dynamic placement capabilities for the workflow tasks. Architecturally speaking, orchestration service API is exposed through an API module.

The LEXIS back-end API is the main entry point for the LEXIS portal; in principle, it facilitates API-based access to LEXIS functionality from other clients/tools. The LEXIS portal acts as a proxy to the back-end services and hence the endpoints offered by the LEXIS portal can be seen primarily as an aggregation of the endpoints offered by the other services (UserOrg service, Alien4Cloud interface, and DDI interface).

The HPC/cloud infrastructure layer focuses on the interactions among HPC and cloud hardware systems to provide the computing power and data storage space to the upper layers. It is implemented as a federation of multiple HPC providers and data centers with the help of the HEAppE middleware [14].

The accounting and billing module provides a periodical status report of used core hours at HPC infrastructure and credits at cloud infrastructure. Solutions include system collectors which collect information from specific resources (HPC or cloud) for specific LEXIS projects. For more information, please refer to Section 2.4.

2.3 Security

Since the very beginning and the design phase of the LEXIS project, security has been more than an important topic for the following two main reasons. The first one is the rise of cybersecurity incidents during the past decade, mainly due to a growing hacking activity for profit. The second one is the importance of intellectual property, data, and results for all private companies (SMEs to big enterprises) that are the targeted customer for the LEXIS platform.

It is worth mentioning that it is quite impossible nowadays to provide or add security at the end of a project due to increased complexity in projects and all technologies that may be used. Including security during the design phase and all along the project is not optional.

Due to the high security standards of all LEXIS partners, we have targeted a high level of security by aiming at the most advanced security concept and principle (at the time the project started) such as a "no trust" concept and "security by design" principle [12].

In a few words, the security concepts and principles can be summarized in the following actions:

- limiting the features and applications running in the platform to only what is required and restricting access to them for all users and processes to minimize the attack surface area;
- putting in place proper identity and access control mechanisms such as reducing privileges and permissions to bare-minimum mandatory access to follow least-privileges principles;
- validating identity and access in all layers and components of the LEXIS platform, adding auditing capabilities and a proper segregation of duties to follow a defensive approach to security; and
- putting in place proper default configurations that do not expose applications or data and do not grant access to resources to follow secure defaults.

Starting from the infrastructure codesign where several layers have been created to properly describe security zones for the segregation of duties concept, both HPC centers (IT4I and LRZ) have been interconnected using site-to-site VPN providing a default secure communication channel between them as shown in Figure 2.2.

Once the main HPC centers are interconnected, each HPC center is an equal deployment of the LEXIS platform to make handling security easier and avoid having different security policy according to the HPC center. A high-level view of the LEXIS platform in each HPC center is in Figure 2.3. The different security zones described in the diagram are a *one-to-one* mapping

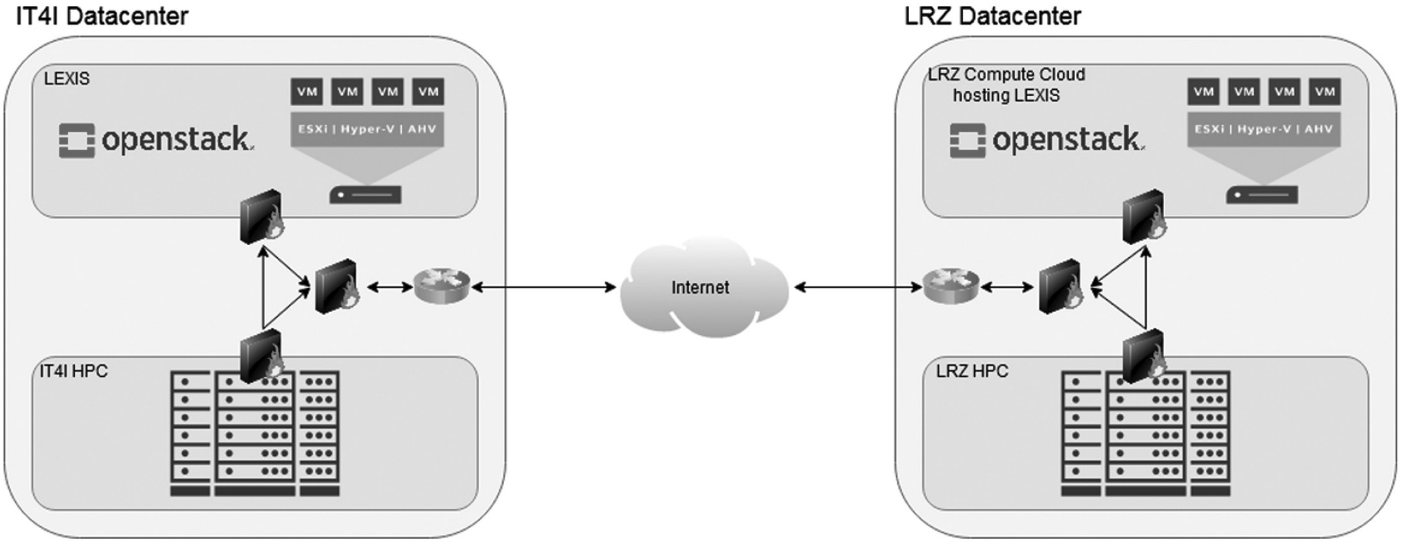


FIGURE 2.2
Federation between data centers with separation of concerns and firewalls.

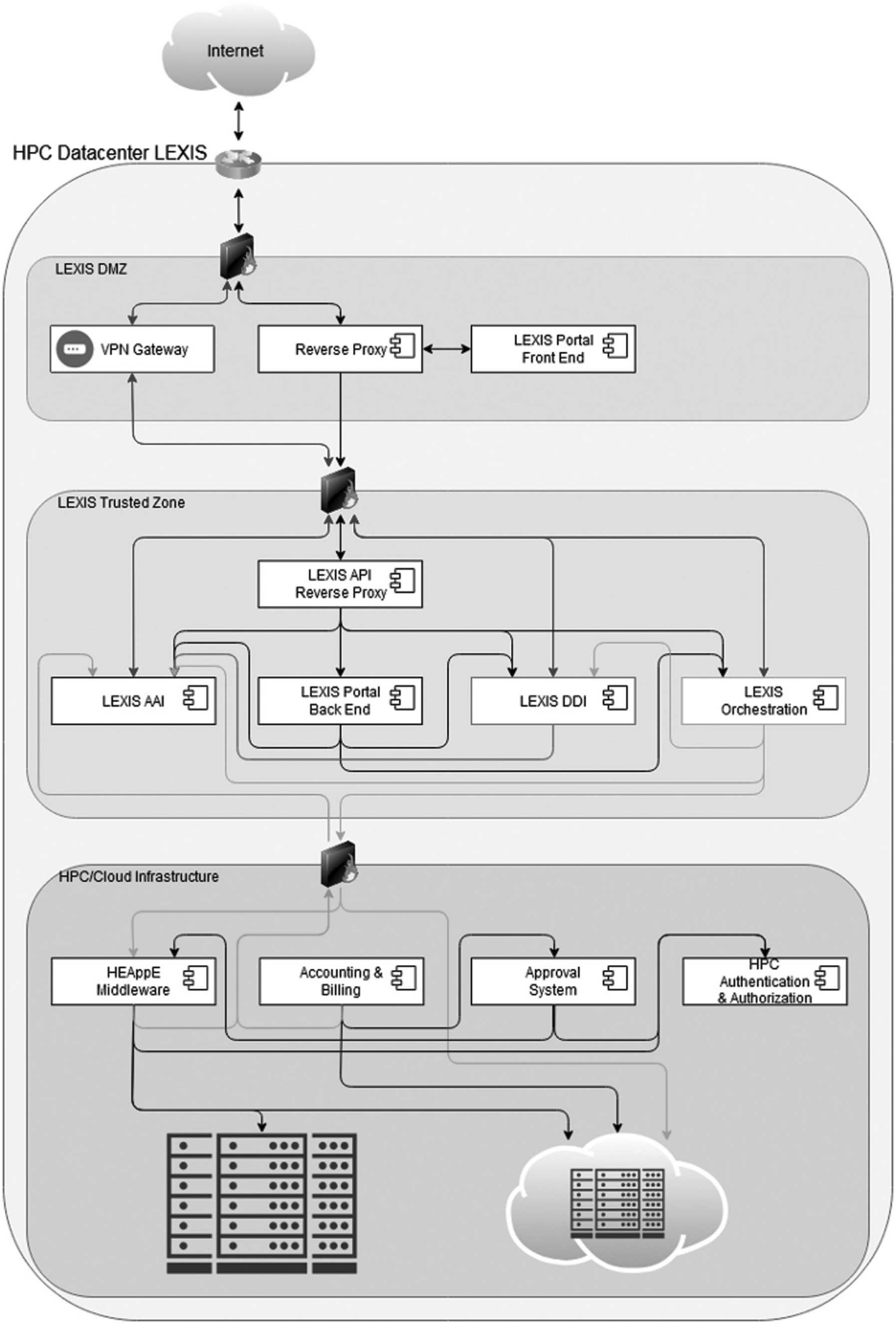


FIGURE 2.3
Security view on high-level architecture of the LEXIS platform.

with the platform architecture services layers in respect to proper segregation of duties. One additional benefit of such deployment is that this can be replicated in another HPC center and connected with site-to-site VPN.

As a broad topic, security also encompasses building resilient and highly available systems for instance. In this respect, the core components such as LEXIS AAI are deployed on different clusters in each HPC center (allowing also better scalability) that are interconnected through secure communication channels and synchronized. The HEAppE middleware then provides a separation between the LEXIS AAI and the local HPC center AAI solution. A more detailed description can be found in Chapter 5 (Section 5.2).

Last but not least, a network security operation center (NSOC) [15] is in charge of providing monitoring, alerting, and auditing capabilities for the LEXIS platform. This security system is based on two independent systems in each HPC center that collect the same metrics from the LEXIS platform and all of its components to properly detect any security incident (security breach, data filtration, etc.). The reason to have two different systems relies on two facts:

- each HPC center has some specific sensitive information that is collected and cannot be exported to any other place; and
- avoid single point of failure for the NSOC: the aim is to have some independent system that can also monitor and audit each other so that any part of the LEXIS platform is redundant and properly monitored and audited.

It is worth mentioning that the infrastructure security is complemented by a software component security. The LEXIS teams put in place a continuous integration pipeline that allows to shift left security testing for all the software components of the LEXIS platform. In addition to that, a security assessment will be performed on the LEXIS platform on a regular basis aiming at automating security testing.

2.4 Accounting and Billing

The European SME ecosystem is getting substantial support from the European Commission for commercialization of cutting-edge innovations developed within Europe. There is an emerging convergence trend with HPC and cloud resources utilization within Europe. For European SMEs to start engaging more organically with European HPC centers, a sound accounting and billing best practices are a must. The accounting and billing capabilities being developed as a part of the LEXIS project is critical to ensure the long-term financial viability of the LEXIS platform. The main requirements of the LEXIS accounting and invoicing engine are:

- ability to track resources consumed from both traditional HPC as well as cloud services offered by an HPC operator;
- ability to support a per account depleting virtual credit pool;
- ability to support separate pricing plans to offer differentiated pricing to different class of LEXIS users;
- ability to support on-demand running cost and usage consumption reporting to enable users to track their cost and resource utilizations; and
- extensibility – the ability of the accounting tool to support new products and services which can be created in the future by the HPC operator.

Figure 2.4 shows the overall architecture of the Cyclops accounting and billing engine [10] that satisfies all the above-mentioned requirements. A brief description of key components includes:

- Collector processes: periodic agents which track various resources, either HPC or cloud, and report consumed metrics to the Cyclops core, if a new service is offered by the HPC operator in future, a detached collection strategy enables the operator to develop a new collector process to bring the new service within the ambit of the Cyclops accounting and billing process.
- Plan management service: allows for registration and management of stock-keeping units (SKUs) as well as association of unit pricing rules with the defined SKU element. This service also allows grouping of various pricing rules, together with discounting criteria under one of more plans. This service enables an HPC operator to offer different service plans to different classes of customers.
- Customer data store: this service maintains the relationship between LEXIS customers, and the service plans with which they are associated. This service also allows association of different service instance identities with a customer account which allows Cyclops to aggregate multiple HPC/cloud service accounting data with the correct customer accounts.
- Credit management service: this module allows virtual credits to be added to the customer's account, which also allows other LEXIS services to query existing credit for a customer in order to take a decision whether a particular service is to be rendered to the customer or denied where allocated credits have run out.

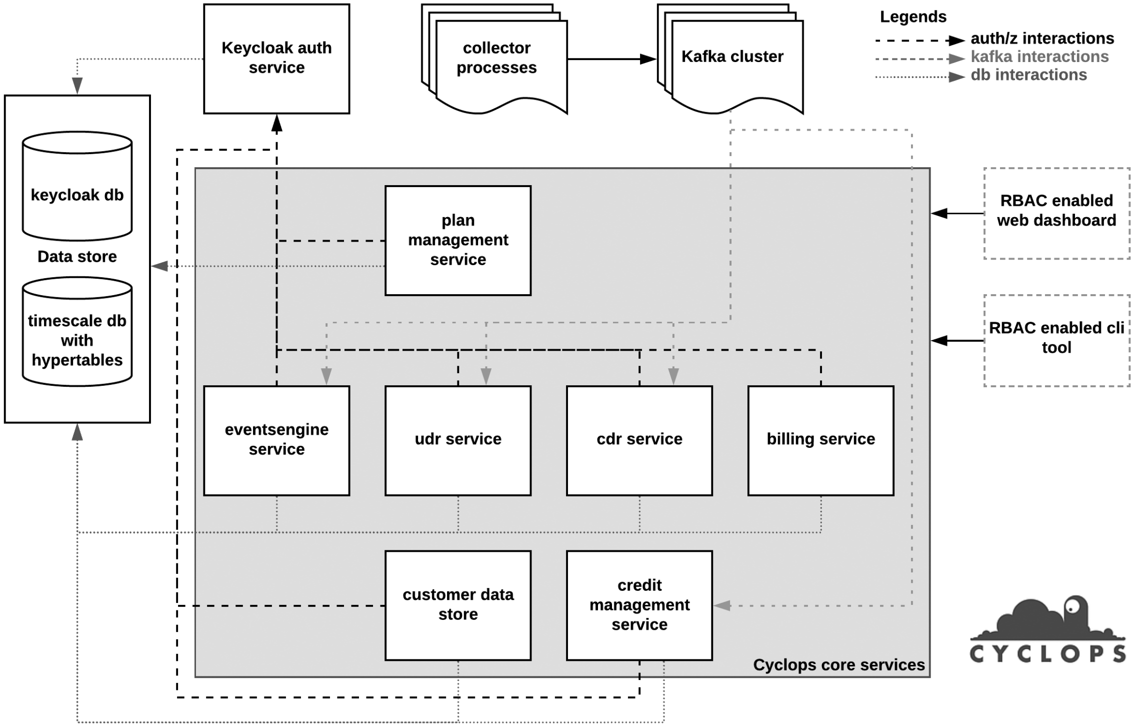


FIGURE 2.4
Architecture of the Cyclops accounting and billing system

- Events engine: this service intelligently tracks the life cycle state changes of an HPC resource. This service enables accounting of a resource based on the time interval a resource is in a given state.
- UDR, CDR services: usage record, and charge record services perform periodic aggregation of usage reports into a consolidated usage and charge reports. Conversion of usage into cost depends on the plans associated with a customer account. Using APIs of these services, the LEXIS platform displays running usage and costs information to the users.
- Billing service: depending on the customer's configured billing period (daily, weekly, monthly, quarterly, semi-annually, or annually), this service generates invoice objects while taking into account any allowed discounts depending on the linked plan with the customer account.

In order to support HPC resources accounting, a dedicated HEAppE collector has been developed which reports the amount of CPU core hours consumed by an HPC scheduled task. Cyclops existing collectors for OpenStack – server collector, floating IP collector, object and block storage collectors enable tracking of both HPC and OpenStack services used by LEXIS users.

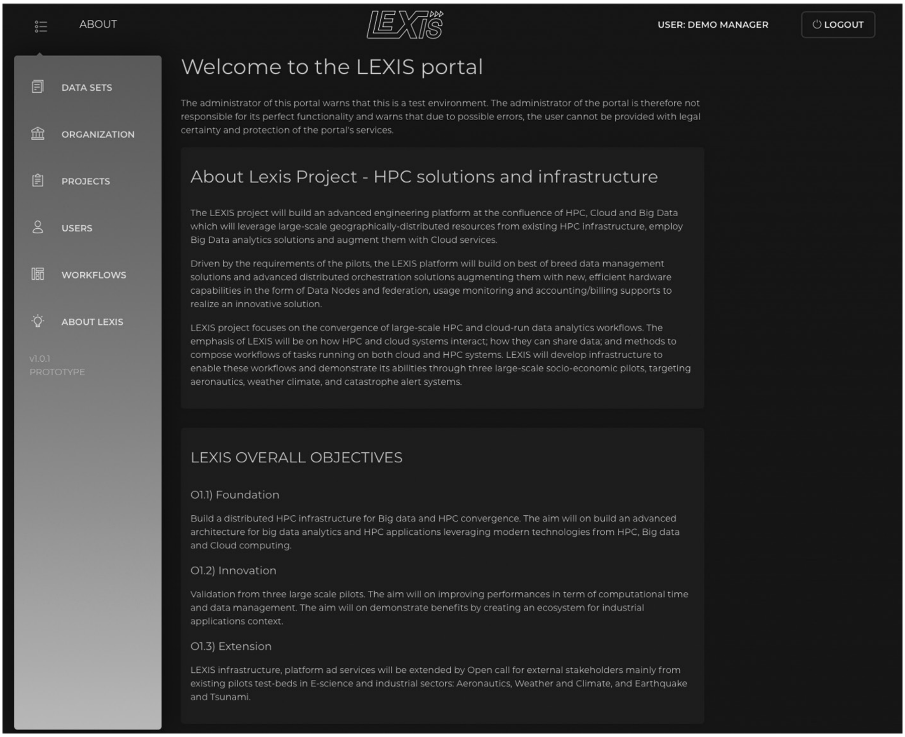
Cyclops services coordinate with LEXIS back-end services (see architecture diagram in Figure 2.1) an establishment of linkages between HPC and cloud project identities and customer accounts.

The entire LEXIS accounting and billing workflow supported by Cyclops is aided by the RESTful interface of the framework, enabling fine-grained access control, as well as ease of integration with the rest of LEXIS services.

Flexible collector development aided by availability of a collector template enables future readiness of the LEXIS billing systems. Billing scenarios to enable invoicing of various levels of technical support hours, future services such as container run times, managed application runtime environment (like PaaS for HPC) and similar forward-looking services can be easily supported by Cyclops accounting and billing platform – as long as a sensible SKU nomenclature, linked pricing rule, and respective collector services can be created.

2.5 Easy Access to HPC/Cloud through a Specialized Web Portal

Access to HPC resources and services has been predominantly used by research centers and academic institutions. The LEXIS portal makes for easy,



Welcome to the LEXIS portal

The administrator of this portal warns that this is a test environment. The administrator of the portal is therefore not responsible for its perfect functionality and warns that due to possible errors, the user cannot be provided with legal certainty and protection of the portal's services.

About Lexis Project - HPC solutions and infrastructure

The LEXIS project will build an advanced engineering platform at the confluence of HPC, Cloud and Big Data which will leverage large-scale geographically-distributed resources from existing HPC infrastructure, employ Big Data analytics solutions and augment them with Cloud services.

Driven by the requirements of the pilots, the LEXIS platform will build on best of breed data management solutions and advanced distributed orchestration solutions augmenting them with new, efficient hardware capabilities in the form of Data Nodes and federation, usage monitoring and accounting/billing supports to realize an innovative solution.

LEXIS project focuses on the convergence of large-scale HPC and cloud-run data analytics workflows. The emphasis of LEXIS will be on how HPC and cloud systems interact; how they can share data; and methods to compose workflows of tasks running on both cloud and HPC systems. LEXIS will develop infrastructure to enable these workflows and demonstrate its abilities through three large-scale socio-economic pilots, targeting aeronautics, weather climate, and catastrophe alert systems.

LEXIS OVERALL OBJECTIVES

O1.1) Foundation

Build a distributed HPC infrastructure for Big data and HPC convergence. The aim will on build an advanced architecture for big data analytics and HPC applications leveraging modern technologies from HPC, Big data and Cloud computing.

O1.2) Innovation

Validation from three large scale pilots. The aim will on improving performances in term of computational time and data management. The aim will on demonstrate benefits by creating an ecosystem for industrial applications context.

O1.3) Extension

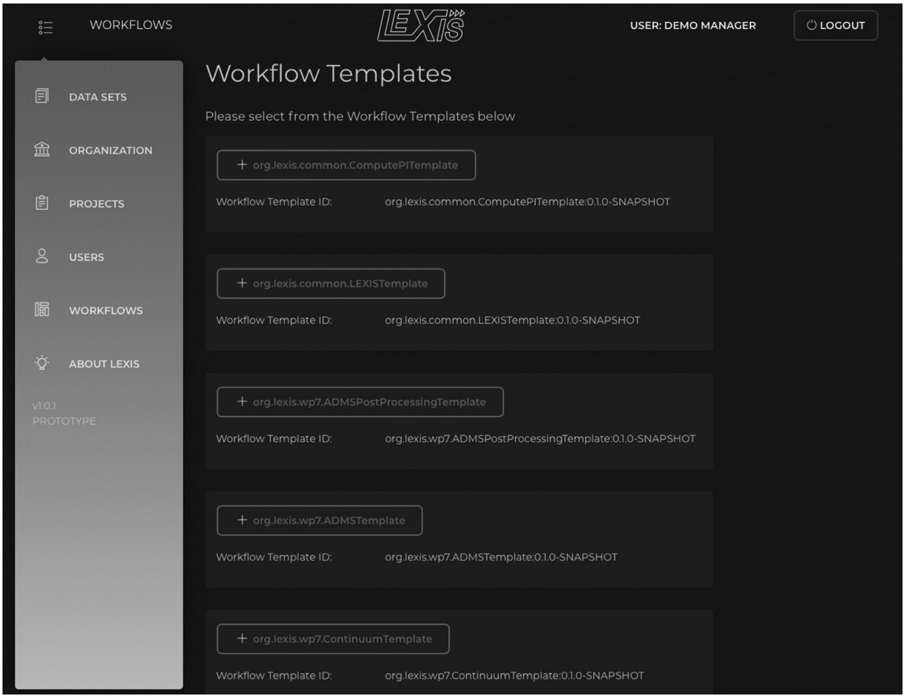
LEXIS infrastructure, platform ad services will be extended by Open call for external stakeholders mainly from existing pilots test-beds in E-science and industrial sectors: Aeronautics, Weather and Climate, and Earthquake and Tsunami.

FIGURE 2.5
Screenshot of the LEXIS portal landing page

streamlined access to European SMEs removing many of the older process friction points. The design of the portal takes concrete steps to specifically address the needs of a European SME entity:

- self on-boarding by organizations and user management;
- simplified resource requisition and approval process;
- access to large inventory of popular HPC workflow templates;
- intuitive workflow execution management, post-execution results access;
- real-time usage and cost tracking; and
- fine-grained access control capabilities in line with needs of privacy and data safety best practices

Figure 2.5 shows the LEXIS portal components. The portal allows an organization to access a rich set of public datasets, as well as manage their internal datasets, stored fully encrypted for additional safety, if needed. The datasets



Workflow Templates

Please select from the Workflow Templates below

+ org.lexis.common.ComputePITemplate

Workflow Template ID: org.lexis.common.ComputePITemplate:0.1.0-SNAPSHOT

+ org.lexis.common.LEXISTemplate

Workflow Template ID: org.lexis.common.LEXISTemplate:0.1.0-SNAPSHOT

+ org.lexis.wp7.ADMSPostProcessingTemplate

Workflow Template ID: org.lexis.wp7.ADMSPostProcessingTemplate:0.1.0-SNAPSHOT

+ org.lexis.wp7.ADMSTemplate

Workflow Template ID: org.lexis.wp7.ADMSTemplate:0.1.0-SNAPSHOT

+ org.lexis.wp7.ContinuumTemplate

Workflow Template ID: org.lexis.wp7.ContinuumTemplate:0.1.0-SNAPSHOT

FIGURE 2.6
Screenshot of the LEXIS portal with workflow templates listing

form an essential component in workflow execution. The following access levels are configurable for every dataset in LEXIS:

- user-only – access is limited to the dataset creator only;
- project-wide – access is allowed to all members of a project; and
- public-access – the access is allowed to everyone with valid LEXIS portal credentials.

Figure 2.6 shows the LEXIS workflow template registry page. LEXIS community and participating HPC centers continually refreshes the registry adding more workflow templates covering a wider set of HPC workload use cases. A user, with the correct access rights, can create a workflow instance from the template, or may even define their own workflow if none of the available templates fits the use case. During workflow instantiation, if needed, appropriate datasets can be linked as input or output targets of the workflow.

The LEXIS portal allows the user to see visually at what stage of execution the processing is as presented in Figure 2.7. Depending on how the workflow stages are defined, the tooltip can show the user workflow stage relevant information.

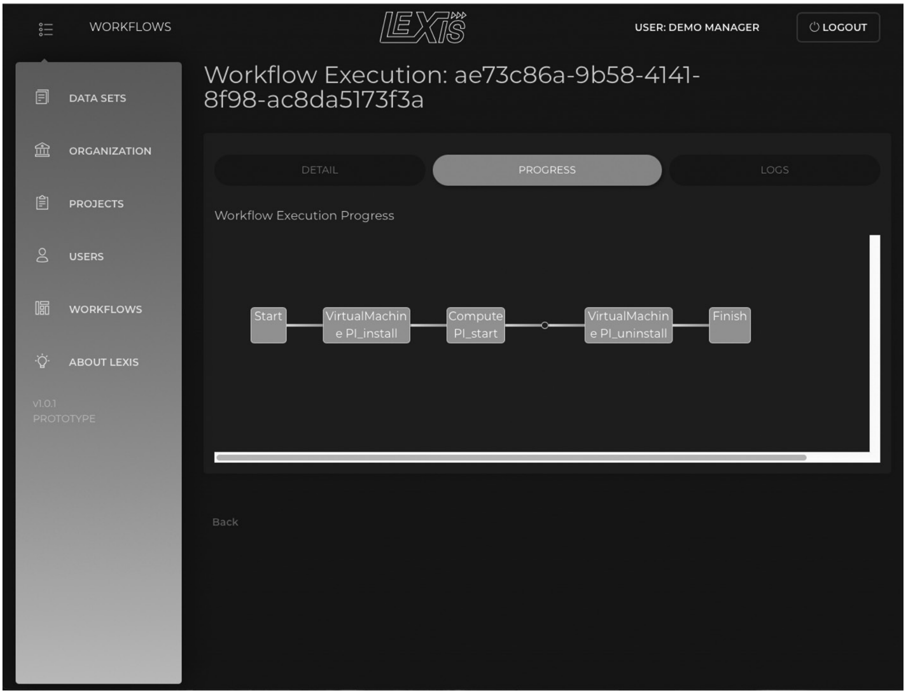


FIGURE 2.7
Screenshot of the LEXIS portal page with detail of the workflow execution

The portal also enables the user to request allocation of relevant HPC/cloud resources to a project at participating HPC centers. Once made, the request is sent to the approval subsystem of the target HPC operator. The HPC operator follows their center specific process for approval, but using a unified interface, notifies the LEXIS user of the status of the request.

The portal supports a role-based feature access by its users. The following roles are supported by the portal: LEXIS administrator; organization manager; organization user; LEXIS support user; organization financial manager; organization license manager; LEXIS project manager.

For detailed description of access rights available with the above-mentioned roles, see Section 2.3.

2.6 Market Analysis

The rapid evolving market targeted by the LEXIS project and the emergence of new critical factors such as the new EU strategic plan for digital sovereignty, GAIA-X [11], the deployment of 5G, the COVID crisis, and the geopolitical battles for the supply of key technologies (i.e. China vs USA), have all impacted the way the LEXIS project has managed its market positioning and its targeted impacts. It has also reinforced the relevance of the LEXIS project in the EU landscape of today and tomorrow.

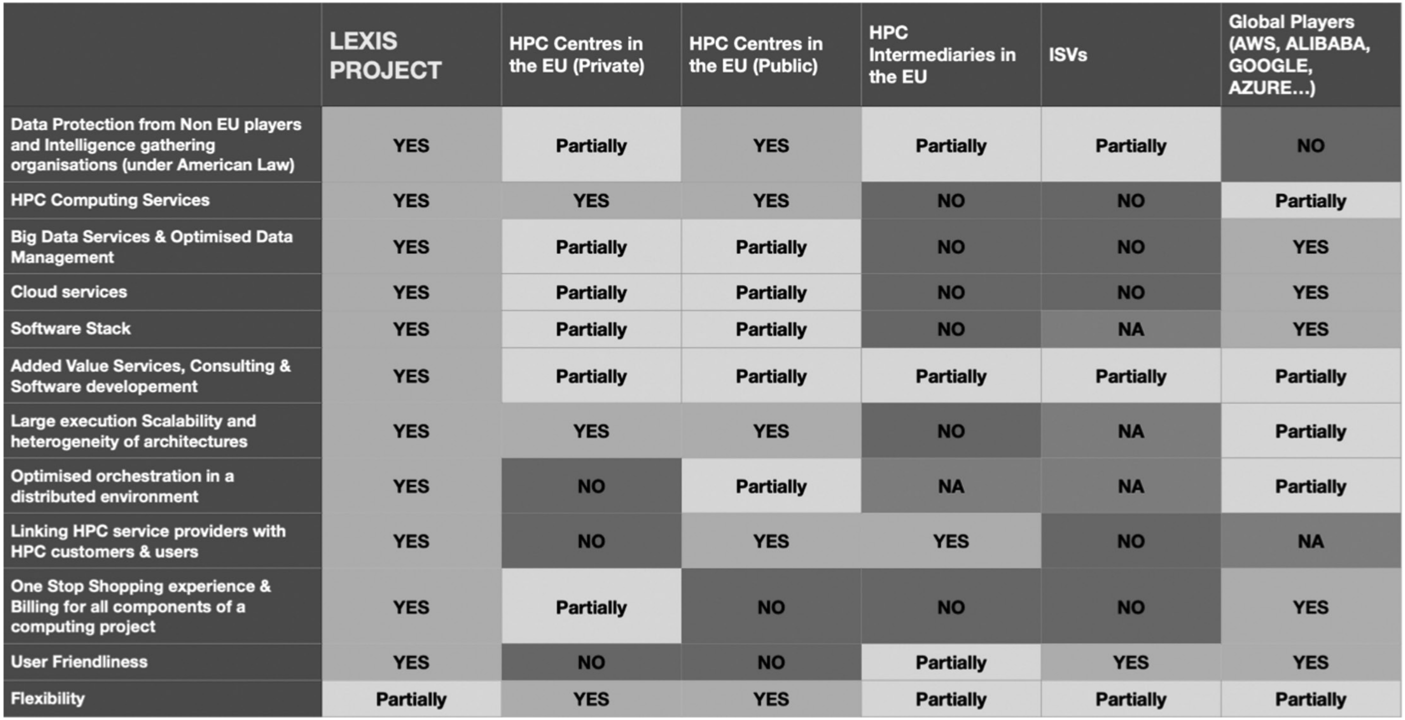
The global trend is clearly to develop a cloud-computing approach for HPC, big data, HPDA, AI integrating de facto the IoT and edge growing ecosystems as part of an end-to-end computing continuum. In addition, the EU wants to push organizations of all kinds to adopt practices, values, and standards aligned with its strategic goals of independence, security, and sovereignty.

All of this was taken in consideration and integrated when designing the LEXIS platform and its services. It was done in addition to the primary objective of the LEXIS project: to overcome the shortcomings of the excessively complicated access to HPC systems, large-scale datasets management, that appears to be a major barrier to the global adoption of these capabilities and technologies by companies (large, SMEs, start-ups) and even a very significant proportion of researchers, to the detriment of European competitiveness.

Instead of trying to directly compete with the dominant players (all from the USA or China), LEXIS has taken a different approach to the market and is building its offers by capitalizing on the best assets available today and in the future and developing complementary state-of-the art additional technologies to procure a clear and valuable differentiation. The key factors of LEXIS market positioning are as follows:

- ease of access;
- scaling of datasets;
- scaling and flexibility of computing, able to welcome additional new infrastructures including exascale-level ones;
- accompanying users to remove major hurdles for non-HPC professionals and to reduce TCOs, operations costs, reactivity;
- preserving key interests in IP, security, and confidentiality within the new framework and global interests of the EU;
- ability to interconnect with other infrastructures via APIs (i.e. AWS, Azure, Google) for flexibility and adaptability to various scenarios; and
- preparing the communities of researchers and companies to use the next level (exascale) in the near future

As a consequence, the positioning is visualized in Figure 2.8 [12].



	LEXIS PROJECT	HPC Centres in the EU (Private)	HPC Centres in the EU (Public)	HPC Intermediaries in the EU	ISVs	Global Players (AWS, ALIBABA, GOOGLE, AZURE...)
Data Protection from Non EU players and Intelligence gathering organisations (under American Law)	YES	Partially	YES	Partially	Partially	NO
HPC Computing Services	YES	YES	YES	NO	NO	Partially
Big Data Services & Optimised Data Management	YES	Partially	Partially	NO	NO	YES
Cloud services	YES	Partially	Partially	NO	NO	YES
Software Stack	YES	Partially	Partially	NO	NA	YES
Added Value Services, Consulting & Software developement	YES	Partially	Partially	Partially	Partially	Partially
Large execution Scalability and heterogeneity of architectures	YES	YES	YES	NO	NA	Partially
Optimised orchestration in a distributed environment	YES	NO	Partially	NA	NA	Partially
Linking HPC service providers with HPC customers & users	YES	NO	YES	YES	NO	NA
One Stop Shopping experience & Billing for all components of a computing project	YES	Partially	NO	NO	NO	YES
User Friendliness	YES	NO	NO	Partially	YES	YES
Flexibility	Partially	YES	YES	Partially	Partially	Partially

FIGURE 2.8
Positioning of the LEXIS platform on the market

2.6.1 LEXIS Project Impact

The impacts of the LEXIS project can be put into two categories. The first one is for infrastructure and data suppliers. Most HPC infrastructures in Europe are primarily used for academic and public research, with only 20% of their capacities available for the private sector. As an average, less than 4% among the 20% available are really used, with the exceptions of HLRS in Germany and CINECA in Italy (but the 20% of CINECA are mostly used by only one client) [16]. By removing the adoption barriers, the LEXIS project will allow the participating infrastructures to welcome many more large, small, and medium companies including start-ups, hence improving dramatically the ROI and TCO for each infrastructure, without major additional investment.

HPC centers will be able to welcome projects beyond their own computing capabilities, by using capabilities available in other participating infrastructures and still keeping the full management of the relationships they have with their clients/users.

Each HPC center will quickly demonstrate a major social impact by allowing economic actors to largely benefit from being empowered to use these capabilities.

Data providers will securely value their data much more by having new users in position to fully extract value from these data (previously they were unable to access computing resources and services in such an easy and affordable way).

The second market positioning category targets end users, researchers, and private companies, irrespective of industry or service type. Finding computing and big data resources is today limited in choice, with the very dominant overseas players managing more than 85% of the European market for cloud-based computing resources and storage. Working with these resources requires significant budgets, specialized skills, and in addition, once you have started with these providers you are trapped due to the difficulties and costs attached to any migration you would like to manage. The LEXIS project allows users and potential new users to access very easily the computing and data resources they need, in compliance with the EU's digital sovereignty vision, the required security, no strings attached, still keeping the flexibility to interconnect with other platforms if necessary.

- Data sources of all sizes can be accessed, connected, or uploaded, including future GAIA-X data spaces, opening the door to unlimited potential use cases.
- One-stop shopping experience for end users even for those with no HPC experience.
- Ability to innovate faster, better, and to increase competitiveness, decrease time to market, or increase the number of R&D cycles at a pace not seen before

It is believed that the LEXIS project's approach to the market will bring a real, valuable, and credible alternative to other solutions today dominating the market (all of which are under US and China control). This solution provided by the LEXIS platform is the start in setting up effective and impactful measures for the benefit of the European economy and research and its independence and sovereignty.

Acknowledgment

This chapter was supported by the LEXIS project, funded by the EU's Horizon 2020 Research and Innovation Programme (2014–2020) under grant agreement No. 825532.

References

[1] Scionti, A., Martinovic, J., Terzo, et al. 2019. HPC, cloud and big-data convergent architectures: The lexis approach. In *Conference on Complex, Intelligent, and Software Intensive Systems* (pp. 200–212). Springer, Cham.

[2] Hachinger, S., Golasowski, M., Martinovič, J. et al. 2021. Leveraging High-Performance-Computing and Cloud Computing with Unified Big-Data Workflows: The LEXIS Project. In E. Curry et al. (eds.) *Technologies and Applications for Big Data Value*. Springer, Cham.

[3] EUDAT Ltd. 2020. EUDAT – Collaborative Data Infrastructure. www.eudat.eu (accessed Apr. 6, 2021).

[4] Xu, H., Russell, T., Coposky, J. et al. 2017. *iRODS Primer 2: Integrated Rule-Oriented Data System*. Williston: Morgan & Claypool. https://doi.org/10.2200/S00760ED1V01Y201702ICR057.

[5] Jones, M., Bradley, J., and Sakimura, N. 2015. RFC7519–JSON Web Token (JWT). IETF.

[6] HEAppE Middleware. 2021. High-End Application Execution Middleware. https://heappe.eu (accessed Apr. 19, 2021).

[7] Allcock, W., Bresnahan, J., Kettimuthu, R., and Link., M. 2005. The Globus Striped GridFTP Framework and Server. In *SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing.* Seattle. https://doi.org/10.1109/SC.2005.72

[8] Bull Atos. 2021. Alien 4 Cloud. http://alien4cloud.github.io/ (accessed Apr. 6, 2021).

[9] Bull Atos. 2021. Ystia Suite. https://ystia.github.io (accessed Apr. 6, 2021).

[10] CYCLOPS Labs, Cloud financial intelligence. www.cyclops-labs.io (accessed Apr. 2, 2021).

[11] Braud, A., Fromentoux, G., Radier, B., and Le Grand, O. 2021. The Road to European Digital Sovereignty with Gaia-X and IDSA. *IEEE Network*, *35*(2), 4–5.

[12] LEXIS Project deliverable 9.5: Market Analysis of Converged HPC, Big Data and Cloud Ecosystems in Europe. https://lexis-project.eu/web/wp-content/uploads/2020/08/LEXIS_Deliverable_D9.5.pdf (accessed Apr. 2, 2021).

[13] Ross, R. S., McEvilley, M., and Oren, J. C. 2018. Systems security engineering: Considerations for a multidisciplinary approach in the engineering of trustworthy secure systems [including updates as of Mar. 1, 2018].

[14] Svaton, V., Martinovic, J., Krenek, J., Esch, T., and Tomancak, P. 2019. HPC-as-a-Service via HEAppE Platform. In Conference on Complex, Intelligent, and Software Intensive Systems (pp. 280–293). Springer, Cham.

[15] Vielberth, M., Böhm, F., Fichtinger, I., and Pernul, G. 2020. Security Operations Center: A Systematic Study and Open Challenges. In IEEE Access, vol. 8, pp. 227756–227779, doi: 10.1109/ACCESS.2020.3045514.

[16] Gigler, B., Casorati, A., and Verbeek, A. 2018. Financing the future of supercomputing. Innovation Finance Advisory. *European Investment Bank* www.eib.org/attachments/pj/financing_the_future_of_supercomputing_en.pdf (accessed Apr. 19, 2021).