Large-scale EXecution for Industry & Society

Deliverable D9.4

Open Call Framework and Stakeholders Engagement on Targeted Large-Scale Pilots – First Report

Co-funded by the Horizon 2020 Framework Programme of the European Union
Grant Agreement Number 825532
ICT-11-2018-2019 (IA - Innovation Action)

DELIVERABLE ID | TITLE	D9.4 Open call Framework and Stakeholders Engagement on Targeted Large-Scale Pilots – First Report
RESPONSIBLE AUTHOR	Marc Derquennes (BLABS)
WORKPACKAGE ID | TITLE	WP9 – Impact on Targeted sectors
WORKPACKAGE LEADER	Florin Apopei (TESEO)
DATE OF DELIVERY (CONTRACTUAL)	28/03/2020 (M15)
DATE OF DELIVERY (SUBMITTED)	24/04/2020 (M16)
VERSION | STATUS	V1.1 | Final
TYPE OF DELIVERABLE	R (Report)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	BLABS, all partners
INTERNAL REVIEW	Fabien Brocheton (NUM), Stefano Serra (TESEO)

Project Coordinator: Dr. Jan Martinovič – IT4Innovations, VSB – Technical University of Ostrava
E-mail: jan.martinovic@vsb.cz, **Phone:** +420 597 329 598, **Web:** https://lexis-project.eu

DOCUMENT VERSION

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	ToC and Document creation	10/02/2020	Marc Derquennes (BLABS)
0.2	LRZ additions	27/02/2020	Stephan Hachinger (LRZ)
0.5	Partners Contributions integration	18/03/2020	All Partners
0.6	Review	01/04/2020	Fabien Brocheton (NUM), Stefano Serra (TESEO)
0.7	Update about datasets made available	06/04/2020	All Partners
0.8	Final version	15/04/2020	All Partners
0.81	Final check by the coordinator	21/04/2020	Kateřina Slaninová (IT4I)
0.9	Update according to the comments	23/04/2020	Marc Derquennes (BLABS)
1.0	Final check	24/04/2020	Kateřina Slaninová (IT4I)

GLOSSARY

ACRONYM	DESCRIPTION
ALIEN4CLOUD	Open Source platform to facilitate application management in the Cloud
BURST BUFFER	For HPC: Fast intermediate storage layer positioned between the front-end computing processes and the back-end storage systems
CDO	Chief Data Officer
CFD	Computational Fluid Dynamics
COPERNICUS PROGRAM	European Union's Earth observation programme
CPU	Central Processing Unit
DDI	Distributed Data Infrastructure
ECOMET	Provider of European meteorological data (https://www.ecomet.eu)
FAIR	Findable, Accessible, Interoperable, Reusable
FET	Finite Element Methods
GNSS	Global Navigation Satellite System
GRID FTP	Grid File Transfer Protocol (FTP Extension)
GLOBUS	Management service for research data (https://www.globus.org)
HE APPE	High-End Application Execution
HIPEAC	High Performance Embedded Architecture and Compilation
HPC	High Performance Computing
IAAS	Infrastructure As A Service
IRODS	Open Source Data Management Software
NVRAM	Non Volatile Random Access Memory
NWM	National Water Model (https://water.noaa.gov/about/nwm)
OPEN FOAM	Open source Field Operation And Manipulation: a C++ toolbox for the development of customized numerical solvers, pre-/post processing utilities for the solution of continuum mechanics problems, most prominently including Computational Fluid Dynamics (CFD)
OSCAR	Observing Systems Capability Analysis and Review
PID	Persistent Document Identifier (https://eudat.eu/services/userdoc/pids-in-eudat)
REST	REpresentational State Transfer: software architecture for Web services
RESTFUL	Web services that conform to the REST architecture
SCRATCH STORAGE	Large storage usually implemented by LUSTRE parallel filesystem or similar and available on HPC clusters
SSD	Solid State Drive

VM	Virtual Machine
WCDA	Weather and Climate Data API
WMO	World Meteorological Organization

TABLE OF PARTNERS

ACRONYM	PARTNER
Avio Aero	GE AVIO SRL
AWI	ALFRED WEGENER INSTITUT HELMHOLTZ ZENTRUM FUR POLAR UND MEERESFORSCHUNG
BLABS	BAYNCORE LABS LIMITED
Bull/Atos	BULL SAS
CEA	COMMISSARIAT A L'ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
CIMA	Centro Internazionale in Monitoraggio Ambientale - Fondazione CIMA
CYC	CYCLOPS LABS GMBH
ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
GFZ	HELMHOLTZ ZENTRUM POTSDAM DEUTSCHESGEOFORSCHUNGSZENTRUM GFZ
IT4I	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA / IT4Innovations National Supercomputing Centre
ITHACA	ASSOCIAZIONE ITHACA
LINKS	FONDAZIONE LINKS / ISTITUTO SUPERIORE MARIO BOELLA ISMB
LRZ	BAYERISCHE AKADEMIE DER WISSENSCHAFTEN / Leibniz Rechenzentrum der BAdW
NUM	NUMTECH
O24	OUTPOST 24 FRANCE
TESEO	TESEO SPA TECNOLOGIE E SISTEMI ELETTRONICI ED OTTICI

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The LEXIS (Large-scale EXecution for Industry & Society) Project is building 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, developed and build on best of breed data management solutions and advanced, distributed orchestration solutions, is augmenting them with new efficient hardware and software capabilities in the form of Data Nodes and federation, usage monitoring and accounting/billing supports to realize an innovative solution.

The consortium develops a demonstrator with a significant Open Source dimension including validation, test, and documentation. It is validated in three pilots - in the industrial and scientific sectors (Aeronautics, Earthquake and Tsunami, Weather and Climate).

On top of this validation via the 3 pilots, the consortium intends to add another level of validation via opening the LEXIS platform to various type of projects in the framework of an Open Call.

The objectives are:

- Validation of technologies developed and deployed,
- Validation of usability according to various computing scenarios,
- Validation of usability of workflows,
- Validation of usability of Data sets, data transfer and data management solutions,
- Collection of qualitative and quantitative data, feedbacks from Applicants all along the Open Call in order to refine and improve the LEXIS platform in the perspective of its final delivery at end of the project.

To achieve these goals, the LEXIS Project will propose to Open Call Candidates:

- Infrastructures and associated resources,
- Datasets,
- Technical Support and Training,
- Dedicated Project Managers, for each selected Application Experiment, monitored by the Open Call Board.

The LEXIS Consortium will organize the Open Call via:

- Defining the Sectors and Industries targeted by the Open Call,
- Setting up the appropriate Infrastructures and Services,
- Defining and managing the Open Call process,
- Defining the Rules for proposal and submission.

At the end of the Open Call an evaluation of the outcomes and results will be managed by the Open Call Board, both at individual project level and globally.

On top of participating to the global Dissemination effort of the LEXIS Project, the Open Call will also be the source of numerous Communications and Publications, participating to hugely increase the impact of the LEXIS Project on Research, Industry and Society.

The Open Call is also a major tool for the LEXIS Consortium in the perspective of executing its strategy for Exploitation, both during the LEXIS Project and after its official end.

Position of the deliverable in the whole project context

This deliverable aims at reporting the intermediate status of the preparation of the the Open Call framework at this stage at month 15 of the project. In addition, we also report about the various ways stakeholders are involved and engaged, or will be engaged when the Open Call will be progressively set up, developed, monitored and finally concluded.

Description of the deliverable

The deliverable D9.4 will be structured as follow:

- Introduction and Objectives: after describing the objectives we introduce what is the LEXIS proposition to the Open Call candidates including the datasets and technical support,
- Organisation: After defining the targeted sectors and industries for this Open Call, we describe infrastructures made available including computing resources and smart gateways, and finally the duration of usage for application experiments running on the proposed infrastructures,
- General process of the Open Call for Applicants,
- Rules for proposal and submission,
- How we end the Open Call,
- How we intend to communicate, disseminate, and publish results.

1 OPEN CALL - INTRODUCTION AND OBJECTIVES

The LEXIS project aims at building an advanced, geographically distributed, HPC infrastructure for Big Data analytics within three targeted pilot test beds. By proposing innovative technologies and exploiting data available from test -bed partners, LEXIS aims to generate valuable outcomes and improve the efficiency and quality of services provided to different stakeholders involved in the testbeds. The developed services will be made available to external stakeholders, with the aim of stimulating the interest of European industries and creating an ecosystem of organisations that could benefit from the implemented platform.

One of the important objectives of the LEXIS project is to create Open Calls mainly but not exclusively to the following key industrial sectors:

- Healthcare,
- Manufacturing,
- Energy.

We will target projects from various external stakeholders:

- Research organisations,
- Industry - Big players,
- SMEs, and among them start-ups,
- Other Projects from the EU funded ones (H2020 / EuroHPC) when possible and relevant.

LEXIS will address key strategic objectives that will be transversally pursued all across the work. The objective of LEXIS as well as its application is to extend demonstrators already existing in the LEXIS project by another set of application experiments selected by the LEXIS Open Call mechanism. The task will include an evaluation of the likely impact at both the initial receipt of applications, along with ongoing monitoring and support of the participants during the life of the call duration. Our plan is to support more than five application experiments which will be selected within LEXIS project Open Call.

Application experiments will leverage special workflows, knowledge and data acquired from the existing pilots. On top of that, we would like to encourage application experiments with similar, but not exactly the same, workflows as those developed by LEXIS pilots.

New applications, selected in LEXIS Open Call, will be tested on the LEXIS Platform to confirm the value and general usability of the LEXIS Hardware and Software co-design.

1.1 OBJECTIVES

1.1.1 LEXIS technologies

LEXIS will focus on creating an ecosystem for encouraging organisations such as research organisations, industries and H2020 projects to benefit from the LEXIS platform, its innovative technology and new type of integration from HPC, Big Data and Cloud computing standpoint. One of the main objectives is to improve performance of various types of computational and data analytics workflows by providing dynamic orchestration on Cloud and HPC infrastructures. The most visible advanced technology and new infrastructure and components involved on the LEXIS platform are:

- Orchestration:
  - Dynamic data-aware and hybrid **workflow** orchestration on both Cloud and HPC,
  - Remote visualization services,
  - Unified web interface for workflow specification and execution (the LEXIS **portal**),

- Federated (cross-site), **distributed data management** (the LEXIS DDI) involving optional use of I/O accelerators,
- Federated (cross-site) **Authentication & Authorization Infrastructure** (AAI),
- Secondary technological developments such as the **monitoring** and **billing** systems for the whole platform.

Validation of usability

The LEXIS concept needs to be validated to demonstrate effective operational readiness at multiple levels:

- Computing scenarios,
- Technical functionalities,
- Workflows efficiency,
- Flexibility to easily welcome new domains of application and use cases.

The Open Call will provide the opportunity to test and validate the platform outside the 3 initial Pilots and to provide the feedback needed to finalise the development of the LEXIS platform.

The selection of projects in the Open Call will be driven with this objective as a main focus.

Scenario for HPC/Big Data/Cloud convergence computing

The platform will support typical HPC workflows which can consist of multiple massive parallel tasks that can run for several days at once and several pre and post-processing tasks which would require movement of extensive data sets. Workflows in this category can have relatively smaller number of tasks, but they can be much more resource intensive compared to the other categories.

Those tasks are typical for example in mechanical engineering where Finite Element Methods (FET) and Computational Fluid Dynamics (CFD) simulations are executed in order to refine design of a particular physical part. Another field which uses this type of workflows is AI where HPC infrastructure is used to train neural networks or Big Data Analytics. These workflows often use visualisation tools to monitor the progress of the tasks and final results of the simulation. What characterizes this kind of workflow is also their hybrid aspect, in the sense where a part of the workflow may be executed on an HPC infrastructure, while another part will run on a Cloud infrastructure.

For workflows

The LEXIS orchestration service will enable the management of complex workflows, including:

- Cross computing between distributed HPC and Cloud infrastructures (hybrid workflow),
- Urgent computing-oriented workflows,
- Failover enabled workflows,
- Data aware workflows (including DDI operations).

A catalogue of workflows templates based on the LEXIS Pilots will be made available. It provides example of typical workflows addressed by the project. The intent is that Open call participants will be able to directly use or customize these workflows to implement their use cases.

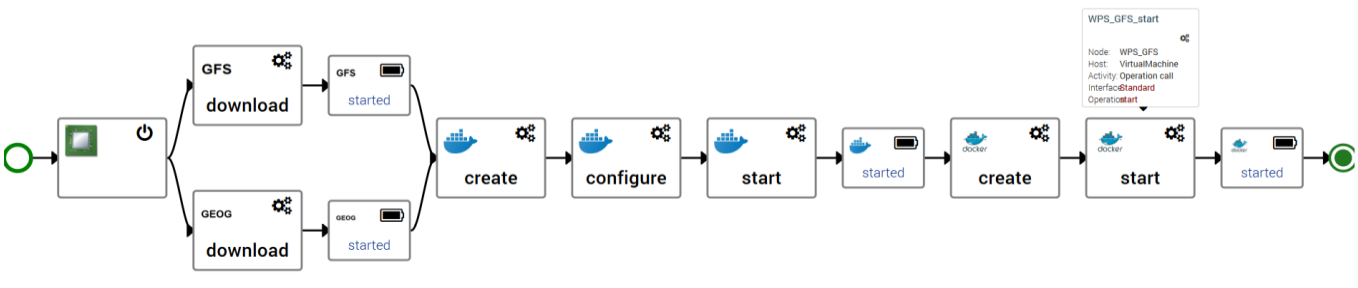

Figure 1 Workflow example in Alien4Cloud UI

For data sets, data management and data transfers

The LEXIS DDI and Weather & Climate Data API (WCDA API) are storage facilities for input, intermediate and output data sets in the context of LEXIS workflows. With the Open-Call pilots, the EUDAT-B2SAFE-/iRODS-based DDI shall be tested for its usability in its broader context. Also, the DDI APIs and the WCDA are tested.

In particular, the assessment of performance in the following use cases is crucial:

- Storage of data sets large in size and/or in file number,
- Cross-site data access,
- Data ingest and retrieval over the APIs (including WCDA), the LEXIS portal, and transfer mechanisms for large data (e.g. GridFTP, GLOBUS),
- Automated data staging by usage of the staging API,
- Assignment of PIDs and DataCite-compatible metadata to data sets in the DDI (as a step towards FAIR data).

These use cases are of such broadness that practically any proponent using LEXIS will be able to contribute testing one of them. Clearly, using standardised HTTP-REST APIs for data ingest/retrieval and staging is not so widely adopted in the HPC and Scientific Computing communities yet (in the Cloud community somewhat more). Demonstrating the added value of this modern approach to build a system and its interfaces is the challenge we will be tackling here.

Scenario for urgent computing

Disasters and crises often occur unexpectedly at anywhere and anytime. In spite of the technological and scientific advances in recent decades, it is still impossible to predict the onset and/or advancement of many natural disasters, e.g. earthquakes, tsunamis, storms, flood and flash-floods well in advance. Numerical computations can simulate the onset and/or advancement of such disasters by utilising an array of data that are commonly unknown until the zero hour, i.e. shortly before/when/after a disaster or crisis strikes, leaving no time to spare. Urgent computing (UC) is a class of computing to support such computations to enable responsible authorities to make educated decisions by providing simulated predictions of a potential extensive loss situation, e.g. the impact and required evacuation zones. It facilitates simulations to start and complete within a required time frame to support decision making for managing the affected areas to reduce casualties. UC requires computations to start in short order and complete within a stipulated deadline so as to support mitigation activities in preparation, response and recovery from an event that requires immediate attention. The result of the computations enables the relevant authorities to make swift and timely educated decisions to mitigate financial losses, manage affected areas and reduce casualties.

This Open call is aiming to attract earth science related applications that match the following characteristics:

- An urgent use case as a recurring issue, e.g. a flood, or a high impact issue, e.g. a nuclear meltdown in a reactor, which is expected to potentially result in extensive loss. The loss is expected to be mitigated with the support of time-critical computations.
- An urgent event as an occurrence at a point in space and time that can potentially create an extensive loss situation, which requires immediate attention.
- An urgent service as the act of the activities, i.e. computation, decision making and coordination work, to fulfil the function of an urgent system.
- An urgent computation as a computing activity that must start in short order, i.e. immediately or as soon as possible, to simulate an urgent event. An urgent computation is triggered by an urgent event, either the commencement or the expectation of it.
- An urgent product as the processed result of an urgent computation that can be directly used for decision making and coordination activities to support loss mitigation.
- An urgent system as a system to enable urgent computation(s) and support decision making and coordination work such that loss mitigation is possible.

For new domain & use cases applications

As indicated in introduction, LEXIS intends to target different domains of application. The main challenge will be on testing both LEXIS technology and platform on different sectors of application in order to test and improve the quality of services for the platform developed.

The objective on extending the Open Call to various application domains and use cases application will be on providing lessons learned for the capability of the LEXIS platform and technology to stimulate interest from different fields of application in industry and society context.

For opening the way to attract specific use cases and infrastructures in domains non targeted on the LEXIS project, a specific investigation will be made in first instance with parallel ICT11 projects as EVOLVE, DEEPHEALTH and CYBELE.

1.2 LEXIS PROPOSITION TO OPEN CALL CANDIDATES

In this section and its subsections, we describe what LEXIS can offer to Open Call candidates. The offerings, together making up an easy-to-use LEXIS platform for hybrid Cloud-HPC workflows, shall attract not only any customers, but in particular high-profile science and industry partners. Thus, it is essential to promote our assets in a way that clarifies the added value of using the LEXIS platform.

In Section 2.2.1, we describe the infrastructures LEXIS users can access. Section 2.2.2 discusses selected datasets present in the LEXIS platform for the existing workflows, which can have a large added value for simulations in adjacent fields. Section 2.2.3 finally discusses benefits in terms of LEXIS Technical Support & Training activities.

1.2.1 Access to infrastructures & related resources

LEXIS Open-Call grantees will access infrastructures at IT4I and LRZ computing centres which is likely to substantially exceed the normal size of resources they use in terms of compute power and data storage. They can use these infrastructures (cf. Section 2.2.1.2) via the convenient LEXIS Portal (cf. Section 2.2.1.1). An effort on security and data safety/protection issues (cf. Section 2.2.1.3) substantially limits the vulnerabilities of the LEXIS platform, even if it is “experimental”.

1.2.1.1 LEXIS portal & visualisation

Web portal

The LEXIS Portal will act as the main entry point to LEXIS for users who are not experienced working with the LEXIS technology stack. Participants in the Open Call will be the first external users of this portal. The LEXIS Portal will provide functionality to interact with data sets (view public data sets, upload/download private data sets), request access to resources by creating new projects, use and run workflows on a mix of HPC and Cloud technologies as well as viewing usage and accounting information pertaining to use of the system resources.

As the LEXIS portal is still new, the LEXIS team members will work with the Open Call participants more closely to understand the usage of the LEXIS Portal (resource request, dataset and workflow management), taking feedback from the usage, ergonomy and, where possible, modifying the functionality of the LEXIS Portal as necessary.

Visualisation

The 3D remote visualization consists in giving the ability for the LEXIS user to smoothly interact with their 2D or 3D high-demanding graphical applications, with neither the need of acceleration (GPU) on their client host, nor the need to download large (to extremely large) datasets to view them. With such model, the data to be viewed remains on the cloud / server side as well as the 3D rendering power (GPU devices).

Remote visualization middle-ware is necessary to encode the 3D experience being rendered on the server side into a streaming system able to accept the mouse and keyboard inputs from the user, in a way that the user cannot really perceive the difference.

At last, the LEXIS portal will make this complex mechanism transparent for the user, provided that this middleware provides a RESTful application programming interface (REST API).

Use cases requiring such technology have been clearly identified for at least two of the existing pilots (for 3D interactive CFD simulation post-processing interactive encoding of weather forecast images into 2D videos). These use cases will be modelled in the orchestrator and stored in the related forge so that they can be reused in other/new projects.

1.2.1.2 TECHNICAL RESOURCES & DOCUMENTATION

Both at IT4I and at LRZ, large-scale HPC infrastructures in the Petaflop range (with HEAppE front-end) and Cloud infrastructures (IaaS OpenStack for running VMs) with together more than 3,000 CPU cores are accessible.

On the storage side, we provide >200TB of storage altogether without large bureaucratic barriers. Besides this, our Burst-Buffer systems can be used as a very quick intermediate data store in the multi-TB range, based on SSDs and NVRAM.

1.2.1.3 SECURITY

While building the LEXIS Platform, the following concerns were the main drivers for taking into account security very seriously since the co-design:

- All Data (input datasets and output results) are not necessarily public: LEXIS platform is able to handle “sensitive data” from customers and ensure end to end “Data Security” (in-transit and at-rest),
- All source code (mainly algorithm) are not public: LEXIS platform allows to protect customer IP, by following Data Protection principles to protect source code of algorithm.

The LEXIS Platform has been built following best practice in the modern industry such as:

- Security by Design,
- Principles of least privileges.

1.2.2 Datasets

In order to drive the Open Call the LEXIS Consortium will provide various type of datasets to Applicants and will also allow some additional datasets to be added to the platform by Applicants themselves for them to be able to quickly implement their application experiments. All of this will be done in a controlled process to ensure security, respect of IP rules and compatibility with the completion status of the LEXIS platform at these stages of development prior to its final validation at the end of the project.

1.2.2.1 Aeronautics pilot

Because the software suite used within the Aeronautics pilot is either proprietary or commercial this software could not be used within the Open Call. For this reason, we prepared publicly available dataset for Computational Fluid Dynamic (CFD) simulation exploiting LEXIS technologies which will leverage on open-source CFD code. This data set will demonstrate the huge potential of LEXIS platform for CFD simulations and since the intention of the Open Call is to target a broad community, the provided data set has been prepared by using widely used open-source CFD code OpenFOAM. This will allow users in Open Call to test prepared workflows and LEXIS technologies for their own applications.

The test case provided in OpenFOAM format is the well-known benchmark of T106C turbine profile tested on a linear cascade experimental rig. Provided dataset contains not only geometrical description of the blade in term of the coordinates, but inlet boundary conditions such as total pressure, temperature and flow angle, outlet boundary conditions such as static pressure but also all necessary OpenFOAM files.

By providing OpenFOAM files, this dataset is ready-to-run on the LEXIS platform which makes it easy for user to get familiar with the LEXIS technologies and as such use easily the LEXIS platform test bed for CFD simulations.

1.2.2.2 Weather and climate pilot

A large portion of the data relevant to the weather and climate pilot is proprietary model output, used at an intermediate stage of the pilot workflows, or third-party observations (e.g. Italian Civil Protection radar and weather stations observational data, Weather Underground personal weather stations) used as inputs, which cannot be shared publicly.

However, the weather and climate pilot will make the outcomes of its workflows publicly available where possible. This includes in particular the meteorological model WRF, Continuum hydrological model, RISICO forest fire risk model and NUM urban model results.

These results will be available for selected case studies in Italy and France over selected periods. The modelling output will be mainly provided in NETCDF format. Furthermore, the main modelling tasks of the Weather and Climate pilot will adopt ad hoc virtualization techniques based on dockers and virtual machines.

These virtualization approaches will allow the formulation of simplified modelling workflows to be demonstrated to external users to get familiar with the LEXIS technologies and to address possible approaches (portal, GUIs etc) to make easier for external actors the formulation of Weather and Climate experiments.

1.2.2.3 Urban air-quality dataset

In the framework of the weather and climate pilot, urban air-quality simulations over the whole Paris are produced for a set of specific dates (year 2018). Note that the software used to produce this dataset is commercial and so cannot be used within the Open Call.

This dataset consists of hourly time-series of 2D air-quality concentrations of NO2 and PM10 at ground level. The spatial resolution is irregular with 1 m resolution close to main sources (road network, industrial sources) up to 100m resolution far away. The output points are around 1,000,000 and cover a domain of about 28 km x 28 km over Paris.

The format of the dataset is NetCDF, standardized for the WCDA exchange API inside the LEXIS platform.

The dataset is freely usable by the new partners of the open call when it is an usage only inside the LEXIS platform for research or demonstration activities. No (direct or indirect) commercial use of this dataset is allowed.

1.2.2.4 Earthquake & tsunami datasets

The earthquake and tsunami datasets will be prepared around the three scenarios planned for the evaluation of the pilot, that is (i) the Padang earthquake and tsunami fictional event, (ii) the Chile 2015 earthquake and tsunami and (iii) the Nepal 2015 earthquake events. Each dataset will include meshes for the tsunami simulations, the parameters for the event (a QuakeML description file) and the exposure dataset subset, at a pre-defined resolution. The data will be in the formats used by the various tools of the pilot workflow, that is mesh elements and nodes for TsunAWI, and OpenStreetMap subsets for the exposure dataset.

A specific OpenStreetMap-based dataset and update set will be made available, to allow external participants to bench the update process of OpenBuildingMap.

1.2.2.5 External datasets

The LEXIS Team will support the Open-Call applicants with the efficient use of external datasets they need as an input for their scientific and industrial workflows. Such datasets, including datasets not yet mentioned in the Sections above, will be made available in the DDI (or, if agreed on, also in the WCDA). Ideally, to this purpose, EUDAT staging (B2STAGE), synchronisation and federation (B2SAFE) mechanisms will be used. Suitable datasets already available on EUDAT systems can be found via the B2FIND portal[1].

Once in the DDI, data are subject to the convenient staging mechanisms (DDI Staging API) within the orchestrated LEXIS workflows. An overview of them is provided within the LEXIS Portal.

As far as possible, applicants within the Open Call are expected to communicate their demands for external datasets at application time.

1.2.2.6 Data management and publication of results

The LEXIS DDI (and, as far as agreed on, the WCDA) can be used to store input, intermediate and output data of a new Pilot. A regular storage quota of 5 TB for persistent storage of a project is proposed, while up to 20 TB can be provided on negotiation (depending on availability). We note that the HPC systems in LEXIS are equipped with temporary SCRATCH storage in the range of 100s of TB per user, which can be used according to the regulations for the respective HPC system. Each project shall indicate expected storage volumes for input, output, and intermediate data in their application (response to the Open Call).

The data flow within the workflows, i.e. the provisioning of input data and the de-provisioning, copying or archiving of output data, is to be implemented via calls to the DDI APIs (mostly the staging HTTP-REST API) and the WCDA, which are normally receiving their calls directly from LEXIS workflow management.

Once output data is archived within the DDI or the WCDA, the Data Management Approach of LEXIS includes the possibility to assign persistent identifiers (B2HANDLE PIDs, DOIs) as far as interesting and feasible. This requires an obligatory provision of basic metadata (oriented mostly at the DataCite/Dublin Core standards) which may be enforced already at early stages of workflow definition. Clearly, the user having to do this is provided support by the LEXIS team.

Via the connection to EUDAT and its intrinsic geographically-distributed nature, the DDI is able to offer redundant storage, and data find-ability via the EUDAT Collaborative Data Infrastructure (B2FIND) - if the user would like his data to be public.

1.2.3 Technical support & training

Even though a great effort is being invested in developing user friendly and partially automated LEXIS user experience, we know that the power of the solution and its flexibility comes from its richness and from very advanced technology. Such a concentrate of technology is very complex and may not be understood by the candidate of the Open Call.

This means that every new LEXIS Applicant will need to be *walked through* the solution with the support of LEXIS technical experts or *professional services*. Such service may include:

- Applicant´s workflow analysis and decomposition in tasks,
- Identification of the best suited resources for each task,
- Identification of applicant-specific constraints: immediate availability of resources, resource reservation, real-time aspects, data location, national regulations, etc.,

[1] Datasets available at B2FIND portal: http://b2find.eudat.eu

- Environment set-up: as a result, one or more applicant´s workflow variants/proposals and matching set of resources translated into Alien4Cloud /TOSCA objects,
- Runtime/production/exploitation technical support (scheduling issues, stalled or crashed jobs, security management, etc.).

Technical support

Every Applicant and Project will be supported by LEXIS resources and coordinated by a dedicated Project Manager (see below), according to the conditions to be defined on a per Project basis.

Training

Training sessions will be organised for Applicants:

- Usage of the platform and its web portal access (for every Applicant),
- Project specific trainings (custom made) if relevant for the implementation of the Applicant Project.

Dedicated project manager

For each Application Experiment selected for the Open Call, the LEXIS Consortium will appoint a dedicated Project Manager who will be the coordinator and the main point of access for each Applicant.

This will ensure a smooth and efficient monitoring of each Applicant's project for better efficiency.

In addition, the Project Manager will have to manage the reporting to the LEXIS Consortium and to manage the collection of qualitative and quantitative data needed for the LEXIS quality monitoring and as such will help the Consortium to identify the potential issues and their solutions.

2 OPEN CALL – ORGANISATION

2.1 TARGETED SECTORS & INDUSTRIES

The LEXIS project has as a notable ambition to contribute to accelerate the pace of digitization and the innovation potential in Europe's key industrial and societal sectors, among them and without being exclusive:

- Aeronautics,
- Manufacturing,
- Engineering,
- Health Care & Pharmaceuticals,
- Weather-related information services (e.g., for the EU Copernicus programme),
- Civil protection.

LEXIS aims to improve the current industrial state of the art for the considered test-beds (i.e., "how data assets are currently managed, processed and exploited by stakeholders") through the proposed HPC/Big Data/Cloud platform, to eventually create large value from data (for instance, "providing easier, faster, and more accurate analytic services that generate really valuable information for stakeholders").

LEXIS will then introduce innovative technologies for the management and analysis of Big Data within the industrial and societal sectors targeted by the test-beds.

The LEXIS Consortium intends to reach out various type of Applicants:

- Big Companies,
- SMEs,
- Start-ups,
- Research Organisations,
- Other EU funded projects (when relevant).

Analysts are predicting an explosive growth in data generation over the coming years which is driving greatly increased data processing everywhere – from right out at the network edge, to inside the largest Data Centres.

Numbers vary, depending on the source, but the orders of magnitude are largely consistent. For example, IDC are estimating 16 zetabytes of usable data in 2020. Data is becoming a key asset of organizations although the real value derives from unlocking intelligence from the data.

Chief Data Officers (CDOs) have an increasingly important role in organizations and are reporting more directly to business leaders which are more and more engaged directly in driving this business change.

As an effect, CDOs target interestingly shifts from managing data and ensuring it is secure to identifying new revenue opportunities based on the organization's data. Key to generating business value is extracting knowledge and actionable intelligence from mostly unstructured data; something that is increasingly difficult to do, as data sets get larger and larger, and structured as well as unstructured data have to be managed and synthesized. The LEXIS project will validate developed solutions leveraging its expertise in HPC, data management and cloud technologies through application experiments, with a perspective to help various type of stakeholders to take advantage of the state-of-the-art developments introduced by the LEXIS Platform.

The Open Call is the opportunity to do so within the targeted sectors and type of Applicants.

2.2 INFRASTRUCTURES AVAILABLE FOR THE OPEN CALL

2.2.1 Computing resources

European computing resources of world class level are made available through the LEXIS Open Call. The intended usage within the Open Call is an *experimental usage with the focus on testing the LEXIS platform* in order to validate and optimise the platform's functionality. This means, that scientific production usage is *generally not in the scope*.

The new pilots can request a usage of the LEXIS platform within the following limits:

- Normalised core hours - minimum: 100,000 core hours
- Normalised core hours - maximum: 1,000,000 core hours

A definition of Normalised core hours (normalisation factors which relate the "raw" core hours to the normalised values on each computing facility within the LEXIS platform) will be published with the open call. This definition may be subject to minor adjustments in the course of the project.

2.2.1.1 LRZ compute cloud, LINUX cluster & SuperMUC-NG

At LRZ, LEXIS Customers have the access to HPC and Cloud-Computing systems (in shared usage mode with other users). The systems relevant for LEXIS are:

- LRZ Linux Cluster - about 1 PFlop/s, different segments (predominantly Intel Haswell CPUs), SLURM scheduler, parallel jobs with up to 1,792 cores on largest (Haswell-CPU-based) segment
- SuperMUC-NG - 26 PFlop/s, Intel Skylake, SLURM scheduler, OmniPath, jobs can scale up to 147,456 cores — access to machine via special proposal
- DGX-1 - 170 TFlop/s (FP16 Peak), NVIDIA P100 GPUs x 8, NVLink
- LRZ Compute Cloud - IaaS Cloud with >3000 Intel Skylake cores (and a few GPUs) for the user to dynamically start Virtual Machines (VMs). Typical VMs have 4 GB of RAM per CPU core and a disk space of 20 GB

For further documentation see:

- https://doku.lrz.de/display/PUBLIC/High+Performance+Computing
- https://doku.lrz.de/display/PUBLIC/Compute+Cloud

2.2.1.2 IT4I infrastructure & HPC clusters

The operational infrastructure at IT4I is the following:

- Barbora HPC cluster - 848 TFlop/s, Intel Cascade Lake, NVIDIA V100, InfiniBand HDR, PBS scheduler
- Salomon HPC cluster - 2011 TFlop/s, Intel Haswell, Xeon Phi, InfiniBand QDR, PBS scheduler
- DGX-2 - 520 PFlop/s (FP16), 130 TFlop/s (FP64), NVIDIA V100 GPUs x 16, NVLink
- Anselm HPC cluster (decommission in Q4 2020) - 94 TFlop/s, Intel Sandy Bridge, InfiniBand QDR, PBS scheduler

For further documentation see:

- https://docs.it4i.cz

Furthermore, IT4I has procured an experimental infrastructure which will be used to operate a small OpenStack Cloud which will be provided to the call applicants as a resource.

2.2.2 TESEO smart gateway

The smart gateway is mainly composed of three parts:

- Data receiving module,
- Data elaboration,
- Data sharing.

Data receiving module is based on a LoRa module able to collect the data from the different sensor node in the network. The core of the data elaboration part is a STM32 CPU based on an ARM CORTEX M4, and the main operations are data correction and data elaboration. The data sharing part is devoted to the data parsing & formatting and data upload to the cloud, this part currently can be done via 3 different ways:

- WiFi connection,
- LAN connection,
- Narrow band connection.

The smart gateway in LEXIS provides a platform able to collect and send the data to a central database for the use of LEXIS weather and climate pilot. For the Open call, the smart gateway can easily collect and send the data as actually is doing for the weather and climate pilot where the smart gateway is collecting several data as is temperature, humidity, pressure, wind speed, wind direction and precipitation intensity; these data are pre-validated directly by the smart gateway and shared via TCP/IP protocol to an external database which than can be exploited for further services.

TESEO expects to broaden the LEXIS project know-how thanks to additional real users´ experience using the smart gateway, so hard data from the field.

When required by the selected Applicant, TESEO will make available up to 10 Smart Gateways for the Open Call.

2.3 DURATION OF USAGE OF THE LEXIS PLATFORM

The experimental activities of the Open Call pilots will end 2 months before the end of the LEXIS project, while Exploitation, Communication and Dissemination activities related to these Application Experiments have to be managed as long as the project is running. Special regulations referring to Computing and Storage resources are reported below.

- Special usage-duration limits applying to computing resources: computations of an open-call pilot will end when the limit of core-hours allocated to that pilot is achieved (cf. Section 3.2.1).

- Special usage-duration limits applying to storage resources: preservation of data, if not guaranteed otherwise (e.g. by direct communication with a computing centre in LEXIS), will only be guaranteed until the end of the LEXIS project. In emergency cases and for limited amounts of data (<10TB), longer preservation may be agreed on a case by case basis.

3 OPEN CALL GENERAL PROCESS

NOTE: in the present report – early April 2020 - the assumption is that the Corona virus crisis has no impact on the time-line and organisation of the Open Call. This may change significantly depending on the development and outcomes of the Corona virus situation.

3.1 OPEN CALL TIMELINE

Due to the technical developments in progress, the Open Call will be officially launched at the end of Q2 2020, with the intent to use and leverage the major event of ISC 2020 as a platform to communicate and raise awareness and interest in the entire European ecosystem (Industry and Research)

Prior to this, the LEXIS Consortium will start to advertise the future Open Call as soon as in May 2020 in specific targeted events as their audience are composed of very well informed and experienced potential future users of the LEXIS Project.

As a consequence, the Open Call will be:

- Launched: end of Q2 2020
- Selection of Applicants: planned in Q2 & Q3 2020
- Implementation of selected Application Experiments: from September 2020 until early Q2 2021
- Evaluated and ended 2 months prior to the end of the LEXIS Project: in Summer 2021

3.2 WHO CAN BECOME AN APPLICANT?

The Open Call is open to any organisation, European based, needing for a project:

- Scalable computing intensive services,
- Data management for significant datasets,
- Cloud access and monitoring.

The precise and detailed criteria will be published in Q2 2020 prior to the official launch of the Open Call

Participants can be:

- Academics,
- Research organisations (private or public),
- Blue Chip companies,
- SMEs,
- Start-ups.

3.3 REQUIREMENTS TO BE FULFILLED

Due to the today's stage of progress, specifically on the technical side of the LEXIS platform, the exact requirements (mostly technical in effect), will be communicated and published at a later stage, just prior to the official launch of the Open Call at the end of Q2 2020.

4 RULES FOR PROPOSAL AND SUBMISSION

At the time of the present report (early April 2020), the LEXIS platform is working on developing the components needed for making the Open Call actionable. From technical point of view, we need to keep in mind at this precise stage:

- A major part of the rules for proposal and submission are directly linked to the final technical characteristics and processes of each component of the LEXIS platform.
- Being only at month 15 of the project, only one part of the final technical characteristics is precisely known, because the consortium is building first the architecture and infrastructure components. The layers dedicated to manage the interactions with future users are less advanced in terms of development and as a consequence some important factors impacting the rules for Open Call will be clearly defined later on, most of them in Q2 2020 only.

As a consequence, the Rules for Proposal and Submission will be fully detailed at a later stage, just prior to the official launch of the Open Call at end of Q2 2020.

4.1 STAGES

Proposals for a new Application Experiment in LEXIS follow a single stage process, by submitting a complete application form. The proposal template will be made available on-line on the LEXIS website.

The new Application Experiments invited by the LEXIS platform will have to use this template to formalise their wish to participate following the invitation by the LEXIS Consortium.

2 channels will be managed to recruit applicants:

- **Process by Invitation from LEXIS:** in this process, the LEXIS Consortium will select potential applicants among the networks of the LEXIS partners and invite the pre-selected ones to participate to the Open Call. Having advanced knowledge of the projects and research programs of these organisations, the LEXIS Consortium Members are ideally placed to detect projects of high interest due to their nature to be included in this Open Call phase.
- **Process by External Submission:** following a multi-level campaign (Web, social media, conference, emailing, booths at events, field actions by Consortium members, etc.) to advertise the LEXIS Project and its Open Call, organisations are able to identify among their projects which ones are valuable to be included in this Open Call. Following the guidance of the on-line published rules and application process, these organisations can submit their projects.

4.2 PROPOSAL LANGUAGE

The application and the related documents have to be written in English (US English, UK English, International English). Any document or application submitted in any other language will be rejected.

4.3 SUBMISSION OF PROPOSALS

The 2 parallel processes to recruit Applicants will be managed as following:

- **By invitation:** Following the invitation and the subsequent application form, the process will be set-up with the applicants depending on the particular circumstances linked to the invitation send by LEXIS Consortium to the Applicant.
- **External applicants**: Proposals will be submitted electronically as per instructions that will be detailed on the LEXIS Project website when published in Q2 2020.

Applicants will be able to submit, resubmit and amend existing applications until the deadline of the Open Call, as per instructions to be published on the LEXIS website.

4.4 ACKNOWLEDGEMENT OF RECEIPT

Applicants will receive a formal acknowledgement of reception by email. This acknowledgement will not be in any way a validation or acceptance of the Application

4.5 RE-OPENING OF THE OPEN CALL

In the case that the evaluation process concludes with insufficient number of accepted or validated proposals, the LEXIS Consortium reserves the right to re-open the Call.

4.6 PROPOSALS EVALUATION AND SELECTION

Essential to the whole process of selecting the Application Experiments and to proceed to the continuous evaluation and monitoring of results, outcomes and feedbacks, an Open Call Board will be set up by the LEXIS Consortium.

The members of the Open Call Board will be appointed in Q2 2020, prior to the launch of the Open Call.

The Open Call Board will:

- Define the Rules for the selection of Application Experiments,
- Monitor the selection process,
- Appoint the Project Managers dedicated to each selected Application Experiment,
- Monitor the successful implementation of the whole set of application experiments and their progress,
- Monitor with the Project Managers the collection and interpretation of data, and results from the experiments,
- Define the evaluation rules and criteria,
- Proceed to the final evaluation of the Open Call: feedbacks, qualitative and quantitative outcomes & results and final report.

4.7 OPEN CALL – AUP / AGREEMENTS FOR THIRD PARTIES (APPLICANTS)

To access to the LEXIS platform, Open Call Applicants will have to sign rules governing this access and will also have to comply with these rules.

These rules govern in particular the possibilities of using:

- The computing capacity of the infrastructure,
- The use of software,
- The conditions of access to the infrastructure.

For simplicity of using existing set of rules, partners will have to sign the rules of those supercomputing centres, where they will perform their calculations.

4.8 OPEN CALL – IPR MANAGEMENT

IPR management applied in the Open Call will follow the rules described in the LEXIS project proposal as well as in the Grant Agreement signed by all members of LEXIS project. General IPR principle of LEXIS project is to ensure as much as possible that the project work is open and usable by others.

This approach will be applied for Open Call as well. However, due to commercial considerations of some industrial partners, not all the work carried out during the Open Call will be open. All the activities regarding IPR issues in the Open Call will be coordinated by the Innovation Manager of the LEXIS project.

Before the Open Call project launches, the LEXIS Project partners will define rules governing access to the LEXIS platform in order to minimize any internal issues within the consortium related to the work, IP ownership, Confidential Information, Access Rights to Background and Foreground IP for the duration of the Open Call.

The Open Call Applicants can identify and agree (in writing) on the background IP for the action. 'Background' means any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights that the Open Call Applicant would be using during the Open Call.

Some of the Results as well as data sets, for example workflows, reports, generated during the Open Call will be made publicly available as an Open Source.

5 END OF PROJECT

5.1 EVALUATION OF OPEN CALL PROJECTS

The Project Managers in collaboration with the LEXIS Management and the Open Call Board will monitor a continuous evaluation process all along the Open Call duration.

Ultimately, at the end of each application experiment, the Open Call Board will have a fully documented evaluation report made by the Project Managers.

Each Application Experiments will have an "End of Project" evaluation made and discussed with the Applicants.

Each individual End-of-Project Report will be then communicated to the LEXIS management team and the Open Call Board for final evaluation.

The detailed and global evaluations of the Open Call will be produced in 2 versions, to respect the confidentiality rules and the IP management rules:

- 1 version (confidential) for internal usage (EU, LEXIS Consortium, Open Call Board, Applicants when relevant),
- 1 version (public) for external communication purposes (when relevant).

5.2 COLLECTION OF FEEDBACK FOR EACH APPLICATION EXPERIMENT & QUALITY MANAGEMENT

All along the duration of the Open Call, each selected project will be monitored by the Project Managers in coordination with the Open Call Board with a process for collecting qualitative and quantitative data for assessing:

- User experience in using the LEXIS platform (Portal, Infrastructure, Data management, Visualisation, etc.),
- Effectiveness of infrastructure management,
- Effectiveness of orchestration,
- Effectiveness of Data management, Data Nodes, Cross site synchronisation for distributed storage,
- Security both from External User and LEXIS Platform administration perspectives.

This collection of data, qualitative and quantitative, will help the LEXIS Consortium to assess quality issues and to address them as soon as possible in preparation of the final delivery of the LEXIS Project.

6 COMMUNICATION & DISSEMINATION, PUBLICATION OF RESULTS

The Open Call must generate a significant amount of various type of Communication and Dissemination actions.

It is a major tool to significantly increase awareness about the LEXIS Project and platform to attract interest of both the Research and the Industry communities.

The ultimate goal of the LEXIS platform is to provide these communities with an efficient way to face the challenges from modern digitalisation and computing era in a complex and converged environment (HPC, Big Data, Artificial Intelligence, Cloud based infrastructures and/or accesses, etc.). As a consequence, it is strategic to use the opportunities generated by using Real Use Cases and Application Experiments on the LEXIS platform Open Call to demonstrate the effectiveness, the relevance and the efficiency for the benefits of our research and modern economies in the EU.

These actions will be organised around 3 major phases:

- Phase 1: Launching the Open Call (awareness, recruiting applicants, Exploitation advertisement and Teasing)
- Phase 2: Along the duration of the Open Call (News, results sharing, monitoring, networks animation, teasing for Exploitation)
- Phase 3: End of the Open Call (results, outcomes, teasing for Exploitation)

For each of the above mentioned phases, we will act in 2 categories of actions, using the tools and strategies developed by the LEXIS Consortium:

- Dissemination,
- Communication.

The global perspective of the Open Call itself and the Dissemination and Communication efforts developed around it is to significantly prepare and support the Exploitation plan for the LEXIS Project, itself focusing on 2 phases of Exploitation:

- During the LEXIS project,
- After the end of the LEXIS project.

In doing so, we will significantly increase the reach and the impact of LEXIS by capitalising on and leveraging:

- The efforts structured and deployed by the LEXIS Consortium for the LEXIS Project itself, and for its own benefit during and after the end of the project,
- The individual partners of the LEXIS Consortium by allowing them to use the communication, dissemination tools and outcomes for their own network in their respective industry and research communities, to their own benefits,
- The possibility for Applicants to largely communicate on the success of using the LEXIS platform for their projects.

The LEXIS Consortium will action every tool, channel build, developed and monitored for and on behalf of the LEXIS project, including among them:

- Scientific and other Publications,
- Conferences, Workshops, Trainings, Events and Tradeshows,
- Website, Newsletters, Blogs and SEO best practices,
- Marketing materials,
- Social Networks (from the LEXIS Project itself and from Consortium Members),
- Field actions by the operational teams of each partner in the LEXIS project (Sales, Marketing, Communication),
- Corporate, EU and National bodies relays when possible.

7 CONCLUSION

In the LEXIS Project, the Open Call provides a unique opportunity for the LEXIS Consortium to achieve multiple objectives:

- Validating technologies developed and the platform as a whole,
- Testing in real conditions by implementing real life workloads or Application Experiments from third parties,
- Collecting precious qualitative and quantitative data for improving and fine tuning the platform before the end of the LEXIS Project,
- Implementing the strategy decided by the Consortium for Exploitation, before the end of the project and after,
- Generating a significant number of publications and communications, thus significantly improving the impact of the project on the research, industry, and society fields.

The Framework designed for this Open Call guarantees the optimal path for reaching these objectives, and its implementation is made possible by the strong engagement of every stakeholder in the LEXIS Consortium.

Each Member of the Consortium will have to be part of this Open Call, at multiple levels (technology development, assistance and support to applicants, project management, communication, dissemination, exploitation, advertising of the Call and recruitment of Applicants and Application Experiments, analysis of feedback and management of subsequent improvements to the LEXIS platform, etc.).

The next steps will allow the execution of this Open Call:

- Setting up the Open Call Board,
- Definition of Rules and Evaluation criteria for publication and advertisement,
- Recruitment of Applicants and selection of the most appropriate Application Experiments.

As a consequence, the Open Call will significantly participate to the final preparation of the Exploitation part of the LEXIS project and help increase the LEXIS Project Impact on its targeted sectors.