Large-scale EXecution for Industry & Society

Innovation Action
Information and Communications Technologies

TOPIC: HPC and Big Data enabled Large-scale Test-beds and Applications

Topic identifier: ICT-11-2018-2019
Publication date: 27 October 2017
Types of action: IA Innovation action
Model: single-stage
Opening date: 31 October 2017
Deadline: 17 April 2018 17:00:00

List of participants

Participant No	Participant organisation name	Acronym	Country
1	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA	IT4I	CZ
2	BULL SAS	Bull	FR
3	NALLATECH UK	NALL	UK
4	ISTITUTO SUPERIORE MARIO BOELLA	ISMB	IT
5	TESEO	TESEO	IT
6	COMMISSARIAT À L'ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	CEA	FR
7	LEIBNIZ SUPERCOMPUTING CENTRE OF THE BAVARIAN ACADEMY OF SCIENCES AND HUMANITIES	LRZ	DE
8	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	ECMWF	UK
9	INFORMATION TECHNOLOGY FOR HUMANITARIAN ASSISTANCE, COOPERATION AND ACTION	ITHACA	IT
10	CENTRO INTERNAZIONALE IN MONITORAGGIO AMBIENTALE – FONDAZIONE CIMA	CIMA	IT
11	AVIO AERO	Avio Aero	IT
12	HELMHOLTZ-ZENTRUM POTSDAM - DEUTSCHES GEOFORSCHUNGSZENTRUM GFZ	GFZ	DE
13	ALFRED-WEGENER-INSTITUT HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG	AWI	DE
14	OUTPOST24 FRANCE (EX SECLUDIT)	SEC	FR
15	CYCLOPS	CYC	CH
16	BAYNCORE LABS	BAY	IE
17	NUMTECH	NUM	FR

Table of Contents

Abbreviations

HPC – High-Performance Computing
HPDA – High Performance Data Analytics
GUI – Graphical User Interface
PM – Project Manager
PMO – Project Management Office
AAI – Authorization Authentification Infrastructures
CaaS – Containers as a Service
IaaS – Infrastructure as a Service
DDI – Distributed Data Infrastructure
REST – Representational State Transfer
API – Application Programming Interface
OS – Operating System
BDaaS – Big Data as a Service
CAE – Computer-Aided Engineering
NTI – New Technology Introduction
NPI – New Product Introduction
KPI – Key Performance Indications
HW – Hardware
URANS – Unsteady Reynolds Averaged Navier-Stokes
SW – Software
WP – Work Package
WCDA – Weather and Climate Data API
IoT – Internet of Things
IPR – Intellectual Property Rights
GO – Grant Office
GG – General Gathering
PTB – Project Technical Board
EIM – Exploitation and Innovation Manager
EAB – External Advisory Board
MMR – Monthly Management Report
CA – Consortium Agreement
PRACE – Partnership for Advanced Computing in Europe
CAMS – Atmosphere Monitoring Service
EFAS – European Flood Awareness System
CFD – Computational Fluid Dynamics

DTC – Design to Cost
GE – General Electric
GBX – Gearbox
ICT – Information and Communications Technology
COTS – Commercial Off-the-Shelf
BD – Big Data
CPU – Central Processing Unit
MPP - Massive Parallel Processing
TACO – Transactions on Architecture and Code Optimization
HPCaaS – HPC as a Service
GIS – Geographic Information System
CDOs – Chief Data Officers
SFC – Specific Fuel Consumption
MARS – Meteorological Archival and Retrieval System
CAGR – Compound Annual Growth Rate
SRA – Strategic Research Agenda
CoE – Center of Excellence
BDVA – Big Data Value Association
SRIA – Strategic Research and Innovation Agenda
PPP – Public Private Partnership
ITS – Intelligent Transport System
EEW – Earthquake Early Warning
DSA – Data System API
DIA – Data-Interoperability API
DTA – Data-Transfer API
TRL – Technology Readiness Level
SEEs – System Efficiency Enhancements
UC – Use case
RANS – Reynolds-averaged Navier-Stokes
LPT – Low Pressure Turbine
AI – Artificial Intelligence
ML – Machine Learning
PR – Partial Response
EW – EasyWave

1 EXCELLENCE

1.1 Objectives

Massive quantities of data are created by modern industrial and business processes, in a range of social and urban contexts. Organisations can gain enormous advantages by making use of this data, including increased productivity and resource efficiency, reduced energy use, improved service reliability, and reduced downtime, costs, product defects and time-to-market. To benefit from this data organisations need to be able to extract insight from the data, and this requires processes of generation, transformation and sharing of this data across systems, locations and organisations. The advances of modern High-Performance Computing (HPC) and Big Data (BD) technologies, provided through cloud services, are one of the most concrete opportunities for European industries to strengthen the value of their businesses. SMEs have much to gain from the cloud, as otherwise HPC solutions would be out of their reach. Cloud-based services can be feasibly procured and used on-demand[1]. To effectively implement a stable Big Data baseline for the European industrial framework, it is important to encourage different HPC/BD platforms and formats to converge towards shared architectural patterns and standard data management and representation techniques. This will support interoperability and seamless integration between platforms within multiple sectors and for various types of analytical services.

The **LEXIS (Large-scale EXecution for Industry & Society)** project aims to build 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 test-beds. 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. LEXIS will address key strategic objectives that will be transversally pursued all across the work. We aim to demonstrate the benefits of our project in the context of three industrial test-beds, leveraging modern technologies from *HPC, Big Data,* and *Cloud computing*. We will refer to Key Performance Indications (KPIs) to demonstrate success against the declared objectives within the industrial test-beds. These KPIs are relevant not only for the objectives themselves, but also for the various stakeholders involved.

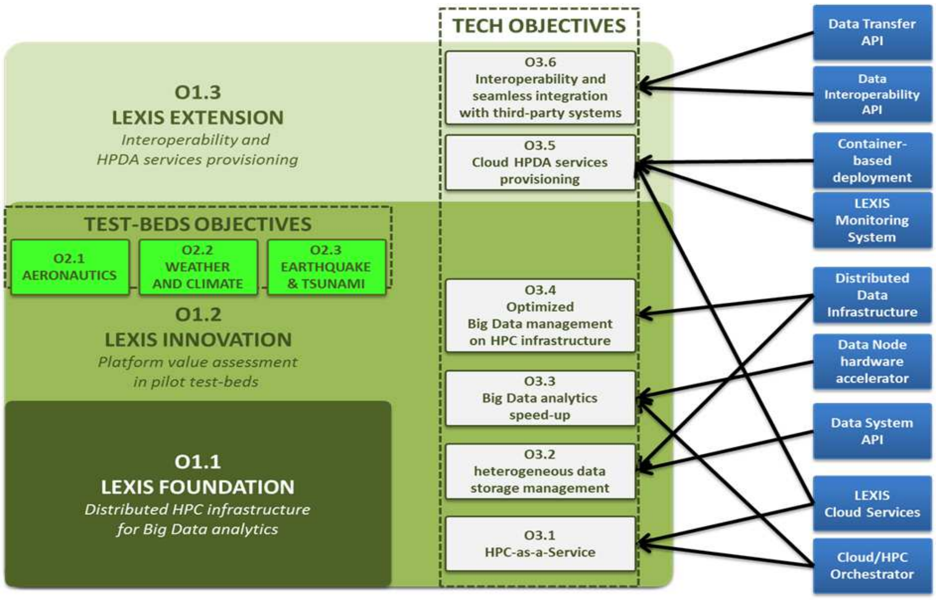

Figure 1.1a Correspondence between LEXIS objectives and proposed technological solutions

[1] European Commission, Digitising mechanical engineering: leveraging the potential of the cloud and data, version 1, May 2017

LEXIS OVERALL OBJECTIVES

O1.1) Foundation. Design and build a distributed HPC infrastructure for Big Data analytics to provide applications and services useful for the industrial sectors of the developed test-beds. This is the principal objective of LEXIS, as it represents the basis for the development of the entire project, especially the execution and validation of pilot test-beds.

O1.2) Innovation. Validate the platform within three different test-beds and **generate valuable outcomes** for test-beds stakeholders, combining data assets and innovative technological solutions. This objective aims to exploit big data assets available for the execution of test-beds and to improve the performance of analyses, both in terms of *computational time* and *results accuracy and relevance* for stakeholders. This objective depends on other objectives in the following groups; in particular O2.1 to O2.3 and O3.1 to O3.4.

O1.3) Extension. Extend the use of the LEXIS services to **external stakeholders**, developing suitable solutions for interoperability, security, resources orchestration, interaction and visualisation. This objective depends on other objectives, i.e., O3.5 and O3.6.

LEXIS TEST-BED-SPECIFIC OBJECTIVES

These objectives are relevant for the partners and external stakeholders involved in the three pilot test-beds of LEXIS: *Aeronautics*, *Weather and climate*, and *Earthquake and Tsunami*.

O2.1) Aeronautics. Reduce the running time of computer-aided engineering simulations to improve engineering productivity and design process quality. This **objective** will address simulations referred to *aircraft engines turbomachinery* and *aircraft engines rotating parts.* It will be **achieved** by investigating the industrial applicability of newly designed HPC and Big Data platforms. To accomplish this, Avio Aero has identified time-consuming data-intensive unsteady simulations of flow behavior in aeronautical *turbomachinery and rotating parts*. Specific KPIs are: reduction of simulation running time (> 20%); widening of design process quality (increased number of design loops).

O2.2) Weather and climate. Within the context of Copernicus Services, increase the timeliness and quality of prediction and analyses. Simplify the access to such services from the cloud, in order to expand the downstream markets: emergency management, sustainable food and energy production, air quality. More detailed objectives supplement the general objectives of O2.2. These are:

- *O2.2.1 Increase the timeliness and quality of emergency management services.* KPIs: (I) improve the forecast of the number and location of actual impacted areas (increase); (II) decrease the proportion of unusable imagery; (III) decrease time-to-availability of post-event imagery.
- *O2.2.2 - Quantitative assessment in terms of affected people and economic losses for forest fire and flood natural hazards.* KPIs: (I) reduction of 25% in the false alert rate for forest fire and flood; (II) reduction of 25% in the missed alert rate for forest fire and flood.
- *O2.2.3 - Sustainable food production and protection of the environment.* KPIs: A 10% improvement in the yield production estimation by agronomic model for one crop campaign.
- *O2.2.4 - Improved prediction of renewable energy production.* KPIs: A 20% improvement in the performance of object-based analysis and verification skill scores for energy production related meteorological variables as well as for predicted energy production.
- *O2.2.5 - Sustainable activity of industrial sites by limiting emission reduction action and thus economic losses.* KPIs: A 25% reduction in the number of forecasted false alerts and missing air pollution peaks over a year for an industrial site.
- *O2.2.6 - Enhanced urban air quality of life building on the integration of improved weather data.* KPIs: A 10% improvement in some air quality statistics (probably more on daily basis than on annual basis) produced by urban model.

O2.3) Earthquake and tsunami. Provide near real-time earthquake and tsunami damage/loss assessments and estimate of the tsunami inundation through simulations based on earthquake parameters, ensuring the delivery of expected consequences in time for fast response planning by emergency dispatchers. The code and interface optimisation in LEXIS will allow us to improve accuracy and computation speeds of simulations and the damage/loss assessment codes. The KPI identified to evaluate the achievement of objective O2.3 will be *time to solution* (computing time from received inputs to delivery of related response) less than or equal to 10min, and decrease by 50% the time needed to process OpenBuildingMap updates. Additional measures for improvements are: factor of speed up with unchanged settings; increase in spatial resolution; increase in area or time covered.

LEXIS TECHNOLOGY-SPECIFIC OBJECTIVES

These objectives are relevant for the architectural solutions supporting the execution of large-scale data analytic services for the test-beds.

O3.1) Providing a ready-to-be-used HPC infrastructure that offers **HPC-as-a-Service** capabilities without incurring in performance/ efficiency slowdowns. Measurable KPIs can be identified for this objective: (I) execution time of a workflow with respect to using a classical system; (II) data throughput characteristics (amount, time) per workflow; (III) total inter-site data throughput (utilisation) in the LEXIS data system per month).
O3.2) Implement a **heterogeneous data storage** management system providing simplified access to huge amounts of data. LEXIS will build a scalable Big Data storage platform, based on modern data-management technologies. Measurable KPIs can be identified for this objective O3.2: the number of data sites integrated in distributed-storage system; other typical r/w performance measures of the system, depending on data source/ sink location (other LEXIS sites).
O3.3) Speed up CPU intensive and data/memory intensive algorithms also to support real time decision making. The exploitation of innovative programming models for HPC heterogeneous infrastructures will be the key challenge for the achievement of this objective. High Performance Data Analytics (HPDA) techniques will be applied to the processing of extremely huge amounts of data, taking advantage of an innovative infrastructure that empowers different workloads. For instance, Massive Parallel Processing (MPP) will be implemented to guarantee scalability of rapidly increasing databases. The implementation of *data-locality-aware task scheduling strategies* will reduce the cost of data transfer among different sites. This will be enabled through data processing policies aimed at the optimization of tasks execution.

O3.4) Optimize data management operations and analytics algorithms that exploit the underlying infrastructures at their best to eventually extract outputs from data that help stakeholders improve their businesses. Thanks to the advanced HPC infrastructures made available by the consortium, LEXIS will be able to manage and process more data and to run more powerful algorithms during the analysis.

O3.5) Provide **simple and secure HPDA service** provisioning, through cloud technologies, for the pilot test-beds, accessible also for other users. This objective will be achieved through the implementation of different user services that will guarantee interoperability, security, resources orchestration, interaction and visualisation:

- **orchestration** of services, data and computational resources in a consistent and efficient way.

- needed services for **provisioning**, **monitoring** and **billing.**

- robust **security** mechanisms for users to easily access and demand for HPC and data resources.

O3.6) Guarantee **interoperability** with **external data sources** and seamless integration with **external systems.** LEXIS will adopt standard semi-automatic data ingestion and fusion methodologies to facilitate the integration of new sources. The semi-automatic matching between data from different sources will be guaranteed thanks to the integration based on metadata for semantic interoperability.

1.1.1 Current status and motivation

At the moment, the Big Data, HPDA and HPC landscape with its big societal impact is fragmented. In particular, the community of traditional, solver/simulation-based HPC is split from the European Data Management community, which itself again is split at least in two parts – one with focus on large amounts of data (big data, as it appears, for example in the ClimEx project[2] generating one of the largest Climate data sets of current time), the other on large amounts of (sometimes small) datasets which must be carefully procured and equipped with metadata (typically Library websites or community-data projects such as TIPTOPbase[3] or PANGAEA[4]). Bringing these two aspects of data management together with HPC and HPDA systems is essential for a European leap forward in the big data field. The BDVA SRIA[5] identifies the following challenges, in the fields of ***HPC for Big Data***, ***real-time analytics***, and ***HPDA services provisioning***, that will be addressed in LEXIS.

- *Support for industrial and business applications.* It is necessary to make use of the progress in efficient and reliable data analytics processes for advanced business applications (for LEXIS test-beds).

[2] www.climex-project.org
[3] http://cdsweb.u-strasbg.fr/tipbase/home.html
[4] www.pangaea.de
[5] Big Data Value Association, European Big Data Vaule Strategic Research and Innovation Agenda, version 4.0, October 2017.

- *Processing architectures.* The capabilities of existing systems to process streams of data and answer queries in real-time and for thousands of concurrent users are limited. Applications, for instance of Artificial Intelligence, are also required to fully exploit all the capabilities of modern and heterogeneous hardware, including parallelism and distribution to boost performance.
- *Real-time analytics.* Real-time analytics through stream processing and HPC have to be adapted to allow continuous querying over streams. A challenge in LEXIS is to have suitable techniques for real-time data streams and also integrated processing for both types of data at the same time.
- *Parallel algorithms computations.* The distribution of Big Data processing nodes poses the need for new Big-Data-specific parallelization techniques, and (at least partially) the automated distribution of tasks over clusters is a crucial element for effective stream processing.
- *HPDA services provisioning.* Taking advantage of a high-performance infrastructure that powers different workloads and starting to support workflows that actually accelerate insights and lead to improved business results for enterprises.

To pursue all these aims, LEXIS will set up a specific test-bed infrastructure connecting three major HPC centres (IT4I, ECMWF, LRZ) and providing a performant, transparent system for Distributed Data Management (LEXIS Data System). This way it will be made possible that LEXIS applications access a unified file system/data management structure. Leveraging the European Research Data Infrastructure (EUDAT) (i.e. using its software offers and connecting with it), LEXIS aims at being exactly this: a connection between HPC, HPDA and data management.

1.1.2 Correspondence between LEXIS objectives and LEXIS consortium

LEXIS OBJECTIVES	LEXIS CONSORTIUM
O1.1) Foundation. Distributed HPC infrastructure for Big Data analytics	It will be achieved thanks to assets and skills brought by consortium partners, i.e. data centres and physical servers (IT4I, ECMWF, LRZ), hardware and software technologies like accelerators (NALL), storage nodes and orchestration software (Bull), know-how on cloud services provisioning (CYC), on unified programming models for heterogeneous hardware and on database and data mining techniques (ISMB).
O1.2) Innovation. Platform value assessment in pilot test-beds	The LEXIS consortium includes relevant industrial and research partners as "owners" of the three test-beds (i.e., Avio Aero, ECMWF and CEA) and other companies and research centres that will provide their expertise in those sectors (i.e., CIMA, GFZ, ITHACA, TESEO, AWI), to appropriately manage the development of the targeted applications.
O1.3) Extension. Interoperability and HPDA services provisioning	The LEXIS consortium includes partners with the suitable know-how on cloud services for HPDA provisioning, securing, monitoring, billing (CYC, SEC, BAY), on orchestration software (Bull) and on data management (ISMB, Bull, IT4I).
O2.1) Aeronautics test-bed	Avio Aero will bring its CAE expertise in the project and will collaborate with HPC infrastructure providers (LRZ, IT4I) to optimize the execution of its algorithms.
O2.2) Weather and climate test-bed	The LEXIS consortium will develop a Weather and Climate Data API that will expose MARS products (developed and operated by ECMWF) together with in-situ observations collected by TESEO gateway and a wide range of weather and application specific products (exploiting CIMA and NUM models). ITHACA will evaluate the impact of hazard predictions on satellite acquisitions pre-triggering and on emergency mapping products improvement (ITHACA is member of the service provider consortium for the EU Copernicus EMS service since 2012, with a long-lasting experience in processing and analysing geospatial data to extract value added crisis information).
O2.3) Earthquake and Tsunami test-bed	CEA will bring its real-time dataflow expertise in the project, as well as its extensive knowledge on high performance infrastructure and code analysis tools. AWI brings its expertise on Tsunami modelling, and its use in the current *InaTEWS* indonesian warning system. GFZ brings its expertise on earthquake damage assessment and links with real-time earthquake early warning third party sources in Japan.

	ITHACA will evaluate the integration of the model outputs on the emergency mapping workflow (ITHACA is member of the consortium for the EU Copernicus EMS service since 2012 with experience in crisis information extraction from geospatial data).
O3.1) HPC-as-a-Service	LRZ, ECMWF, IT4I are HPC experts and will bring their technological assets to LEXIS. BAY will add their expertise on HPC systems and other partners are experts in cloud services and orchestration (CYC, ISMB, Bull).
O3.2) heterogeneous data storage management	LRZ has set up such a distributed storage system on the basis of iRODS in the AlpEnDAC (AlpEnDAC.eu) project. LEXIS will use modern data-management technologies (EUDAT B2SHARE/iRODS), with capabilities for organizing files as a transparently accessible (multi-site/one-view) file system.
O3.3) Big Data analytics speed-up	IT4I, LRZ and ECMWF are competent in HPC application optimisation be it memory or CPU bound. The LEXIS Data System will provide for the first time real-time access from distributed HPC/HPDA systems to distributed data libraries and repositories.
O3.4) Optimized Big Data management on HPC infrastr.	The HTTP REST API services developed within LEXIS will, among other tasks, support data search, retrieval/staging and the orchestration of these data-management tasks by interaction with the orchestration/scheduler systems.
O3.5) Cloud HPDA services provisioning	Bull will cope orchestration needs proposing the novel *Ystia Orchestrator* for hybrid HPC/Cloud systems. Additional Cloud services will be provided by: CYC (accounting and billing solution for a federated HPC/cloud infrastructure), SEC (security, authentication and authorization), ECMWF (Weather and Climate Data API), LRZ (Data System API will enable LEXIS data layer exploitation from the Cloud).
O3.6) Interoperability and seamless integration with third-party systems	LRZ has experience in this field from the *GerDI* project. (FAIR Data MGMT – Findable, Accessible, Interoperable, Reusable). ISMB has experience from the *Snowball* project (development of a Data Layer based on OASIS EDXL-DE standard for emergency data exchange among applications). A suitable module will manage the interoperability among different data sources. The module will be able to interface with multiple data formats and communication protocols, in order to support the ingestion of both standard batch data and real time streams. A common data format will be defined for each test-bed. Such format will be defined to support the common standards currently used in each industrial sector.

1.2 Relation to the work programme

This paragraph details the methodologies LEXIS adopts to address the specific challenges and objectives of the ICT11 Innovation Action.

TOPIC CHALLENGE	ACTION/CONTRIBUTION
Build industrial large-scale application test-beds that integrate and make best use of available HPC, Big Data and Cloud computing technologies	Three relevant "industrial" partners of the LEXIS consortium will manage the large-scale test-beds development (one WP for each): Avio Aero, ECMWF, CEA. They will tightly cooperate with other partners, especially those providing HPC infrastructures (IT4I, ECMWF, LRZ), but also for architecture co-design (ISMB), Cloud orchestration (Bull) and provisioning software (CYC, LRZ, BAY) and Big Data management and access APIs (LRZ, IT4I, Bull, ISMB).
“make best use of currently available HPC and data infrastructures”	The HPC layer builds on top of European top-notch HPC infrastructures at IT4I, ECMWF and LRZ and utilises these systems available in the contexts of PRACE and EGI, together with innovative Burst Buffer systems and Cloud-Based HPDA systems. As for the data layer, LEXIS will rely – as far as possible – on EUDAT[6], one of the most prominent efforts on Research Data Infrastructure in Europe. Their B2SHARE service shall be the basis for the LEXIS Data System Core (**WP3**). This approach will bring EUDAT together with a top-notch HPC/HPDA effort and thus be very valuable for both sides.
“Accelerate the pace of digitization and the innovation potential in Europe's key industry sectors”	LEXIS will dedicate a WP for each pilot test-bed, to pursue the objective of accelerating the innovation within the targeted sectors. Moreover, LEXIS will dedicate the entire **WP8** for the development of cloud services for third parties, while **WP9** will focus on improving the business processes and on involving stakeholders, by creating open calls to key industrial sectors to stimulate the LEXIS framework adoption.

[6] www.eudat.eu, H2020 #654065

1.2.1 Relation to scope a) of the Work Programme

The scope a of the Innovation Actions of the topic on HPC and Big Data enabled Large-scale Test-beds and Applications *scope* a) Innovation Actions (2018 call - deadline in April 2018) *of the Work Programme has been analysed in the following.*

SCOPE CHALLENGE	ACTION/CONTRIBUTION
"Development of large-scale HPC-enabled **industrial pilot test-beds** supporting Big Data applications and services"	We will propose three test-beds from different industrial sectors (aeronautics, earth security, weather and climate), with different kinds of analysis requirements: compute-intensive algorithms with the objective of cutting execution time; data-intensive algorithms that elaborate tons of data and are affected by scalability issues; real time analyses that need to provide responses and to trigger alerts with a minimum delay.
Combine and/or adapt existing **relevant technologies** (HPC/BD/Cloud) to handle and optimize the specific features of processing **very large data sets**	Main architectural components for BD management are described in the "LEXIS concept" section. In particular, the LEXIS data system provides a real-time synchronization over distributed data infrastructure, based on EUDAT B2SAFE (or the underlying iRODS). Data Nodes are Bull's implementation of burst buffers and represent one of the main innovation introduced by LEXIS, as data analysis applications will benefit from the integration of devices of massively parallel acceleration devices (I/O access acceleration).
The industrial pilot test-beds should handle **massive amounts** of **diverse types** of big data coming from a **multitude of players and sources**	Available data assets are described in the test-beds description of "LEXIS concept" section. Test-bed stakeholders provide the LEXIS consortium with tons of data, coming from different sources. For instance, in the weather and climate use case, conventional observations (balloons, GEOSS, Copernicus, etc.) will be fused with IoT data from different kinds of sensors (also worn by citizens), smart gateways, low-energy environmental monitoring systems, etc. Sharing of global weather and climate products is extremely challenging given the amount of data produced daily is in the order 100 terabyte.
Demonstrate how pilot test-beds will generate **innovation and large value** creation	The "Ambition" section describes the added value LEXIS intends to bring in the industrial sectors of the targeted pilot test-beds. In the Aeronautics test-bed, LEXIS will lead to (I) shortened entry into service of aero engines and (II) improved quality of engineering design, able to cope with highly demanding requirements matching market expectations. In the Weather and climate test-bed, private sector entities will be provided with prediction services with increased accuracy. In the Earthquake and tsunami, any interested tsunami warning center or civil protection agency can use the enhanced version of TsunAWI, and AWI can provide support on a non-profit basis.
Describe the **data assets available** to the test-beds and, as appropriate, the **standards** it intends to use to enable **interoperability**	Available data assets are described in the test-beds description of "LEXIS concept" section. Out of previous experience (e.g. projects GerDI and InHPC-DE of LRZ), the LEXIS data architecture strongly focuses on holding metadata describing the format of each stored dataset. On this basis, the LEXIS Data System will allow for setting up and automatizing data conversion, concentrating on the project's demonstrators/use cases as well as allowing for generic applications. Data interoperability for HPC/HPDA applications is thus a major strength of the LEXIS Data System, while towards the outside, the adherence to metadata standards (DataCite, OGC, ISO standards; OAI-PMH serving of catalogues) will allow LEXIS data sets to be interoperable and re-usable (following the FAIR principles of Research Data Management).
Pilot test-beds should also aim to provide, via the **cloud**, simple secure access and **secure service provisioning** of highly demanding data test-beds for companies and especially **SMEs**	LEXIS Cloud services include components to provide test-beds and external users with data analytics resources: User Interface for visualization, workflow scheduling, data view, and billing and monitoring. The distributed data system is augmented by APIs for data exploration/search. These APIs will simplify access to LEXIS data infrastructure for third parties. For instance, Weather and Climate Data Interchange API will provide secure and reliable access to weather-related observations and several kind of model products via the cloud.
Proposals should be led by and show strong **industrial commitment**	LEXIS consortium includes three big industrial stakeholders for the pilot test-beds: Avio Aero, CEA, ECMWF. Their role within the project is of primary importance and they will strongly participate to the related WPs. The beneficial effects on third-party industries, for each test-bed, are specified in the test-beds description of "LEXIS Concept" section.

Proposals should explain how the proposed activities will be **industrialized** and have impact on the competitiveness and **leadership of European industry**	LEXIS will dedicate **WP9** to impacts on targeted sectors in Europe. The aim is to implement actions to increase market share and accelerate the adoption, convergence and integration of HPC, Big Data and Cloud solutions, finding new possible stakeholders. We envisage a task for exploitation and sustainability to stimulate interest and investment opportunities in HPC and Big Data technologies from Public and Private bodies across Europe. A market analysis task will identify how the LEXIS platform can be positioned within the market context.
Proposals should target a **wide participation** and/or applicability and use of the targeted industrial pilot test-bed **by industrial members/**users from different countries and regions	The "Ambition" section describes the added value LEXIS intends to bring in the industrial sectors of the targeted pilot test-beds. The beneficial effects on third-party industries, are specified in the "LEXIS Concept" section. As a matter of fact, LEXIS will dedicate **WP8** to the development of cloud services for third party companies and SMEs, while **WP9** will focus on their involvement through open calls to key industrial sectors to stimulate the adoption of the LEXIS framework.
Proposals should define quantifiable outputs and impact **KPIs**, in particular related to the "Expected Impact" of the topic	Detailed KPIs have been defined under each test-bed-specific objective and also for technology-specific objectives, where appropriate. The achievement of such KPI target values will clearly have a positive impact in the considered industrial sectors, also for third party organizations and SMEs (benefits are specified in the test-beds description of "LEXIS Concept" section).

1.3 Concept and approach

1.3.1 LEXIS concept

As stated in the previous sections, LEXIS aims at building an advanced, geographically-distributed, HPC infrastructure for Big Data analytics that will support the execution of large-scale test-beds in various industrial sectors. To ease the design and implementation of the infrastructure, LEXIS will be managed by three major components: **cross-site infrastructure**, **data management** and **cloud provisioning services**. Before introducing the details about the integration among these components, it is important to better describe the three considered test-beds. Indeed, **the design and development of HPC/BD technologies in LEXIS is strongly oriented to provide a significant support to the pilot test-beds**. Therefore, the description of test-beds ***scenarios***, ***objectives***, ***requirements*** and available ***data assets***, as well as the ***benefits*** expected from the LEXIS platform will help to design a better architecture for LEXIS.

AERONAUTICS TEST- BED

SCENARIO: scope and related industrial activity. Objectives will cope with:

1. Improving speed/performance in terms of computing time, by leveraging/optimizing CAE tools integration with novel HW solutions and data structures.
2. Raising computing speed will allow to multiply number of possible design iterations.
3. Streamlining the process of analysing turbomachinery results by enhancing post-process capabilities connected to Big Data treatment.

Avio Aero will investigate and validate HPC and Big Data solutions under development by applying two specific cases taken from his products line. The introduction of fast computing systems will ease the design of current and future components, favouring a quicker adoption of new innovative concepts and implementing higher quality design, for a given set of requirements.

Objective of the analysis. Final objective of the performed numerical analyses is to develop new technologies able to produce high-quality components and systems (mainly Low-Pressure turbines, combustors and gear boxes) for both civil engines with better performance and military aviation, able to optimize energy consumption and make lighter aircraft.

DATA ASSETS available to the test-beds and related standards. Data assets, supporting the aeronautics research stream, will be represented by a bucket of CFD job simulations, running in batch mode. Typical outputting is provided through ASCII format files or visualized thru 2D mapping or trending graphs.

Expected results for the main stakeholder of the test-bed (Avio Aero). Significant cuts in CAE solvers' running time able to widen exploration regions and put up-to-date quality products on the market. Enforce Company's brand.

Involved third parties and beneficial effects for them. European Aircraft manufacturers, and related Tier1 suppliers, will take clear advantage/benefits from this Research Collaborative project since it will lead to an accelerated exposure of new break-through technologies. On the other side, Aeronautical products time-to-market should be squeezed significantly with lower development costs. Relevant information will include sharing of the key research outputs, expected in aero engines shortened entry into service and improved quality of engineering design, able to cope with highly demanding requirements matching market expectations.

WEATHER AND CLIMATE TEST-BED

SCENARIO: scope and related industrial activity. The project will facilitate complex stacks of weather-related computational models to improve the prediction of water-food-energy nexus phenomena and their associated socio-economic impacts. These computations involve multiple model layers chained together; (1) global weather and climate models, (2) regional weather models, (3) domain-specific application models (such as hydrological, drought and fire forecasts), and (4) impact models providing information for key decision and policy makers. This chain of computations requires significant volumes of data to be passed between model layers, and potentially between physical locations.

Involved third parties and beneficial effects for them. Forecast data is used by a wide range of third parties. Civil Protection Agencies, such as the Italian Civil Protection Department, will take advantage of improved prediction of natural hazards. The European Civil Protection Mechanism can potentially take advantage of pro-active triggering of satellite/aerial (manned/unmanned) observation of areas forecast to be impacted by natural hazards. This includes the proper selection of observations and data types according to the expected major damage types (e.g. wind or flood damage in the case of cyclones). Private sector entities including those in the reinsurance, food and agriculture, health, and energy sectors make use of forecast data, and can benefit from improved weather predictions. Prediction services with increased accuracy will be provided to private sector entities including in the reinsurance, food and agriculture, health, and energy sectors.

Objective of the analysis. To provide improved forecast results by making use of a large volume of unstructured observational data in both the global and regional model layers. LEXIS will also support climate simulations that are not currently feasible (i.e. in the application and impact layers) and reduce time-to-market in applications and impact estimation by leveraging improved prediction capabilities.

DATA ASSETS available to the test-beds and related standards. The test-bed will exploit Copernicus and GEOSS data services, supplemented and complemented by global and local in-situ unstructured observations; in particular we have confirmed access to a very large set of mobile phone weather data provided by Chinese mobile operator Moji. Data will be filtered and pre-processed before being assimilated to ensure quality. Moreover, each layer in the test-bed will produce its own large data set of forecast data, varying from hundreds of Terabytes for global weather models, to Megabytes at the decision maker level. Both the assimilated data and that produced by each layer, will be made available between layers, or for further analysis and industrial exploitation, through a cloud-based Weather and Climate Data API (WCDA). Data will be accessed using scientifically meaningful metadata and provided according to various applicable standards such as GRIB, BUFR, ODB, WaterML 2.0, NETCDF-CF, and SENSOR-ML.

Expected results for the main stakeholder of the test-bed (ECMWF). We expect to enable the assimilation of novel in-situ unstructured weather observations, leading to improved accuracy of global and local forecasts, especially in regions where data was previously sparse or non-existent. Further, we aim to improve the integration between model layers through the Weather and Climate Data API, from global and regional forecasting through to application and impact models. This better integration between forecasting and impact modelling will also facilitate proactive requests for rapid mapping services in case of hazard prevision.

EARTHQUAKE and TSUNAMI TEST-BED

SCENARIO: scope and related industrial activity. Provide real-time estimate of earthquake shaking, tsunami inundation, and their respective damage as soon as the key earthquake source parameters can be determined or are delivered by an Earthquake Early Warning (EEW) system.

Involved third parties and beneficial effects for them. *Earthquake Research Institute*, University of Tokyo, Japan, together with the *Japan Meteorological Agency*, Tokoy, Japan, providing EEW notifications and acting as a test case for information generated in the LEXIS framework. *InaTEWS* – Indonesian Tsunami Early Warning System. Located at *Indonesian Agency for Meteorological, Climatological and Geophysics* (BMKG), Jakarta, Indonesian. *gempa GmbH*, spin-off of GFZ, providing the tsunami decision support system TOAST to BMKG, including support, maintenance, training. Eventually: Intergovernmental Oceanographic commission of UNESCO as coordinator of the world's tsunami warning systems; civil protection agencies. TsunAWI is open source and distributed under a license derived from GPL. InaTEWS and gempa GmbH will receive training and support to integrate the real time version into their systems. Any interested tsunami warning centre or civil protection agency can use TsunAWI, and AWI can provide support on a non-profit basis.

Objective of the analysis. The final objective is to provide earthquake shaking and tsunami inundation maps that are calculated in real time in case of a tsunamigenic earthquake and a full description of the assessed damage and losses from the respective hazards. The damage/loss assessment will be constantly updated as better earthquake characterizations become available.

DATA ASSETS available to the test-beds and related standards. EEW notifications need to be integrated into the workflow for assessing the ground shaking in higher detail than EasyWave (EW) provides. The resulting shaking distributions (expressed in various quantities, e.g. peak ground acceleration, spectral acceleration) are then used as input for the damage/loss calculator of the OpenBuildMap. In TsunAWI, input data are currently read from disk. As computational meshes with bathymetry and topography require some care, they still have to be prepared in advance and provided from disk, but earthquake data must be passed to TsunAWI as an input parameter (e.g., QuakeML). TsunAWI output data: GIS formats as best suited for the warning centre. Currently, TsunAWI writes netcdf with time slices of tsunami amplitude and derived data. GIS data products (e.g. shapefile for information at coastal forecast points, raster data for visualizing tsunami propagation in the ocean) are extracted in a post processing step, however, in LEXIS, TsunAWI shall be integrated and deliver the GIS data directly. The damage/loss calculation will be performed on a building-by-building basis and deliver the key values in a probabilistic sense in the desired detail. Depending on the stakeholder's interest, any type of aggregate (e.g. by zip code, by city) can be created.

Expected results for the main stakeholder of the test-bed (CEA). As there are many parameters to decrease the computation time of TsunAWI, we are sure that we will deliver an inundation simulation that can be run in real time. In particular, a coarse resolution on a regional mesh will be feasible, and sensitivity studies will be conducted to assess the effect of each simplification. Thanks to the tuning on all levels of the workflow during LEXIS, we hope to retain most of the accuracy compared to the original data base scenarios. Furthermore, we are looking forward to extending our knowledge on code optimization. As TsunAWI can be regarded as a test-bed for numerical methods on unstructured triangular meshes, we plan to port the most promising techniques to the global ocean model FESOM. LEXIS will provide the technology to disseminate real-time damage/loss estimates in unprecendented detail to support disaster managers and relief workers. EEW providers will be able to use the OpenBuildingMap damage codes (or possibly web services) to integrate a building-by-building damage and loss assessment.

1.3.2 Overall approach and methodology

The overall architecture of LEXIS has been designed as three integrated layers: **Infrastructure Layer**, **Data Layer** and **Cloud services**.

INFRASTRUCTURE LAYER

The Infrastructure Layer will be implemented through the federation of multiple data centres (IT4I, LRZ, ECMWF). This solution allows to distribute the computation across sites when the resources available on a single site are not enough to satisfy the requirements of the application. To ensure the orchestration among these heterogeneous data centres, <u>LEXIS will introduce a novel Cloud/HPC orchestrator, to manage all sites as if they were a single data centre</u>.

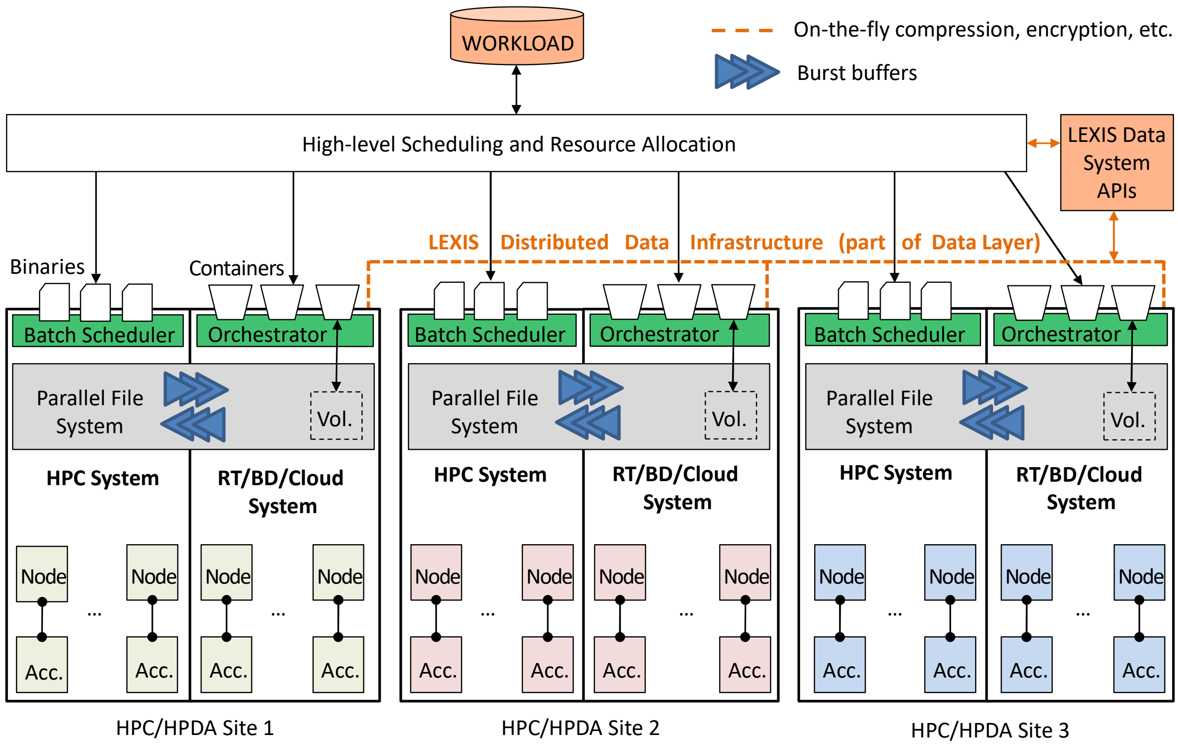

Figure 1.3.2a Architecture of Infrastructure Layer

Orchestration. This module will be responsible for scheduling the execution of tasks on different compute nodes depending on their characteristics (type of hardware) and on the requirements of the same tasks (e.g., compute intensive, data intensive, real-time forecasting). ***Ystia Orchestrator*** under Bull internal code name, is a hybrid Cloud/HPC orchestrator. It aims at supporting the whole application lifecycle. *Ystia Orchestrator* is TOSCA native to allow handling complex applications in a standard way. *Ystia Orchestrator* is also workflow-driven; this means that it does not contain any hard-coded lifecycle logic. This allows to fully customize applications behaviour and to execute custom workflows at runtime. It is designed for large-scale and is built with a tasks/stateless workers model in mind allowing to scale it horizontally easily. Finally, its comprehensive REST API and a modern CLI allow using, deploying and interacting with application at runtime. This solution targets the need for orchestration services in **O3.1**, **O3.3**, and **O3.5**.

Hardware accelerators and Data Nodes. These components represent one of the main innovation introduced by LEXIS in the contexts of the industrial test-beds. Data analysis and simulation tasks will benefit from the integration of massively parallel acceleration devices. One of them which has already been identified by LEXIS is the Burst Buffer based on Bull's data nodes solution, that requires scalability and manageability improvements. Data Nodes are Bull implementation of burst buffers. They provide I/O acceleration capabilities to applications for a subset of their data access operations when these operations are issued in burst mode, i.e. when reading or writing large amounts of data during a short period of time. This solution targets the scheduling objectives in **O3.1** and **O3.3**.

Containers. Easy application deployment will be supported by the container technology such as Docker, Singularity, LXC/LXD. This will enable ***third parties to upload applications*** to the platform and link them to the available data sets such that applications can perform their work. Linux Containers support traditional programming models and frameworks, such as OpenMP and MPI, thus allowing the creation of ***virtual large clusters***, without requiring additional hardware.This solution targets the need for provisioning of services of objective **O3.5** and **O3.6**.

DATA LAYER AND CLOUD SERVICES

The **Data Layer** will run on top of the Infrastructure Layer and includes the architectural components aimed at the collection, management, storage, retrieval and provisioning of data **Cloud services** will use APIs provided by the Data Layer (LEXIS Data System APIs) and by the Infrastructure Layer to provide users with suitable interfaces for accessing the platform and executing analytics jobs.

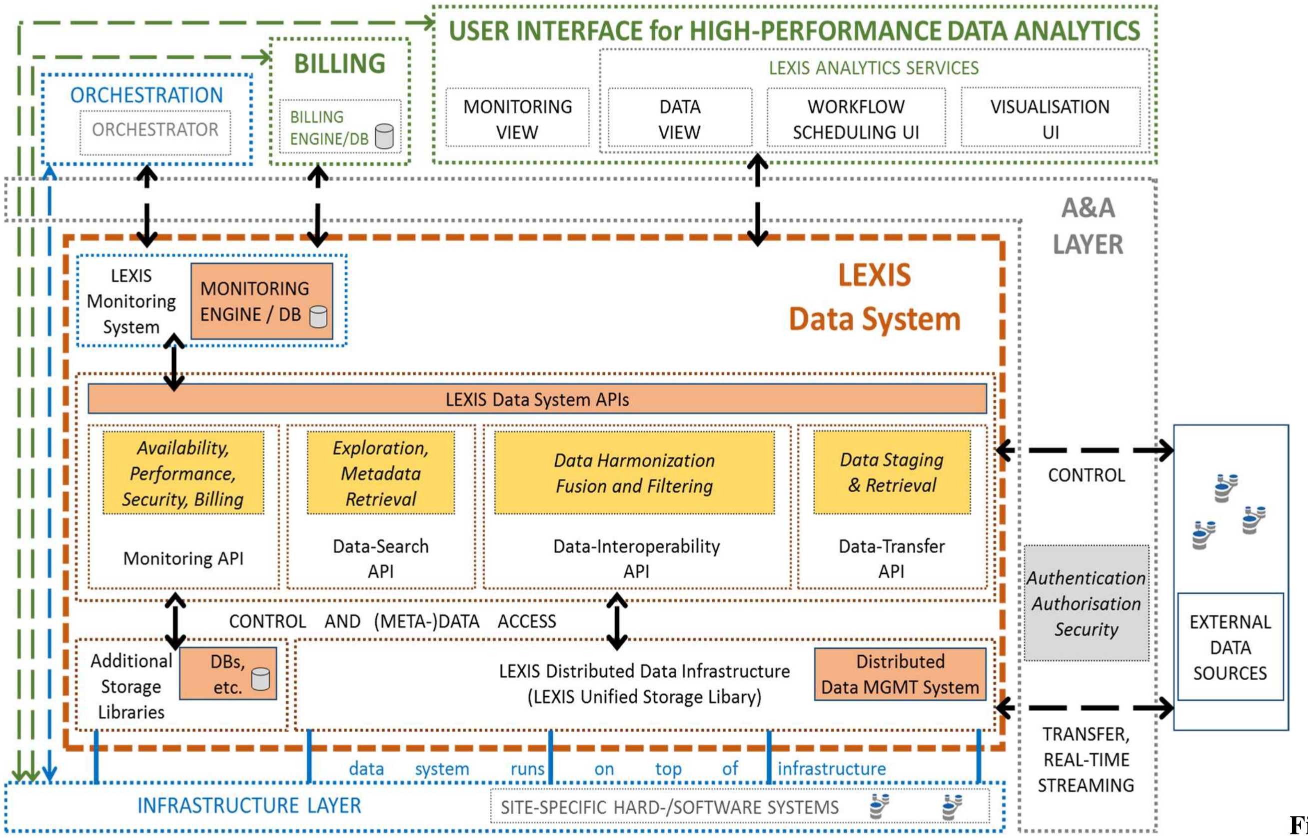

Figure 1.3.2b Architecture of Data Layer and Cloud services

<u>LEXIS Data Layer</u>

LEXIS Data Layer consists of two major components – one being a ***Distributed Data Infrastructure (DDI)*** and the other being ***Data System API (DSA)***, offered as HTTP REST APIs. In addition, the Data Layer includes monitoring functionality to support cloud services (***LEXIS Monitoring System***) and *interoperability* (***LEXIS Data Interoperability***). A common A&A infrastructure will validate users and applications for access to all these components. The main technological solutions for **LEXIS Data Layer** are better described below.

LEXIS Distributed Data Infrastructure (DDI). The DDI leverages the low-level/site-specific storage systems, connecting them with a middleware (where EUDAT B2SHARE will be used if possible/suitable). <u>The middleware makes the different storage systems into a unified virtual filesystem accessible from all sites</u>. This files system can contain extended metadata and fine-grained rules not only for A&A, but also for storage distribution or data management in general (as EUDAT is based on the widely-used iRODS).This solution targets the storage optimization objectives in **O3.2**.

LEXIS Data System APIs (DSA). The distributed data system is augmented by APIs for data exploration/search, staging, retrieval/transfer and preprocessing (aggregation, filtering, format conversion). These APIs will simplify everyday tasks in the LEXIS infrastructure and <u>allow the data system to connect to the Orchestrator and Task Schedulers as well as to user services such as a Portal</u> (**WP8**). This solution targets the need for simplified access to huge amounts of data of objective **O3.2**.

LEXIS Monitoring System. The LEXIS ***Monitoring System*** will track users access and usage of both data and computing resources and store the information into the ***Monitoring DB***. The billing system will then connect to the *Monitoring System* via a suitable ***Monitoring API***, enabling a fine-grained accounting and billing for usage or according to any other model. This solution targets the monitoring objective defined in **O3.5**.

Data Interoperability. Data interoperability function in LEXIS will be implemented through two modules of the DSA: ***Data-Interoperability API (DIA)*** and ***Data-Transfer API (DTA)***.

DTA will be interfaced with *external data sources* to <u>prepare data for the subsequent analyses</u>. It will provide *Data staging and retrieval* functionality, i.e., it will extract data from sources, supporting protocols for both batch and real-time streaming. The Data Layer will adapt to specific sources, for instance supporting protocols for IoT devices like MQTT or AMQP or being implemented through <u>a data stream-process platform like *Apache Kafka*</u>, which could be used also for the following steps. DIA will be implemented through: *Data filtering* to remove erroneous data; *Data fusion* to synchronize data from multiple sources; *Data harmonization* to apply a standard common format to data, in order to guarantee interoperability also with external system. This solution targets the interoperability objective defined in **O3.6**.

<u>LEXIS Cloud services</u>

LEXIS Cloud services include components to provide test-beds and external users with data analytics resources. They are better described below.

User Interface for HPDA. Here is where users can interact with the platform. Main user operations will be:

- Request the scheduling of an analytics application, using the ***Workflow Scheduling UI***, which will communicate with the Orchestrator and the Infrastructure Layer.
- Visualise results, using the ***visualisation UI***, that will be implemented using suitable web-based analytics reporting tools.
- Explore available data sets, using the ***data view***, that will extract data through the *Data-Search API* of the LEXIS Data Layer.

This solution targets the need for provisioning of services of objective **O3.5**.

Billing engine. The billing solution that will be developed in LEXIS will leverage the flexible, rule-based billing engine from CYC. The billing framework will process billable metrics and events monitored within both the Data Layer as well as individual Infrastructure Layers in the LEXIS federation. CYC collectors will be deployed at each federation sites which would hook into local monitoring service and send usage reports to the billing framework. The rule engine will have necessary model inserted that will allow proper transformation of usage reports into charge reports. Finally, during invoicing phase, all charge reports will be aggregated across all user-account ids belonging to the same federated user-id/billing-user-id. This solution targets the billing objective defined in **O3.5**.

Authentication, authorization and security. Authentication and authorization are keys for data security. The main missing part in existing solutions is the continuous monitoring of the system that can ensure that no improper configuration will expose assets or data to any attackers. A specific focus will be made on Data Layers security continuously monitoring that the three pillars of Data Security (Data-at-rest, data-in-transit, data-in-use) are properly configured. An additional correlation part between the data layer (and the infrastructure layer), and the user and access management part will be added to provide efficient continuous monitoring of the overall system.This solution targets the security objective defined in **O3.5**.

1.3.3 Positioning of the project

LEXIS involves a substantial amount of innovative technological solutions for HPDA and Big Data analytics provisioning. LEXIS aims to achieve an integrated pilot system demonstrator that will be assessed in the context of the three test-beds, i.e., analytics for manufacturing optimisation, weather and climate information services provisioning and real-time prediction of natural disasters effects. Piloting LEXIS on these test-beds and in federated environments will be fundamental in achieving the desired level of maturity, which we position at **Technology Readiness Level (TRL) near to 8.**

1.3.4 Linked research and innovation activities

Project name & involved partners	Short description	Relevance to the project and ICT11 scope
Mont-Blanc project (Bull, LRZ)	The third phase of the Mont-Blanc project builds upon previous FP7 projects and aims at designing a new Arm based high-end HPC platform able to deliver a new level of performance/energy ratio when executing real applications.	New type of compute nodes, with an increased memory bandwidth.
	Mont-Blanc 2020 action continues this work and aims at paving the way to the future low-power European processor for Exascale (EPI initiative).	Designs and IPs for new SoCs.
CloudDBAppliance (Bull)	It is an ongoing project and aims at producing a European Cloud Database Appliance for providing a Database as a Service able to match the predictable performance, robustness and trustworthiness provided by on-premise architectures such as those based on mainframes.	In-memory analytics engine running directly over the operational data allowing companies to act upon facts as they happen.
OPERA (ISMB, NALL, TESEO)	It is an ongoing project aiming at designing and producing low-power computing systems covering the whole spectrum of the computing continuum. By leveraging on the integration of low-power hardware accelerators and the use of smart orchestration policies, an effective interaction between Cloud edge nodes and core data center is achieved.	Orchestration and workload consolidation in heterogeneous infrastructures.

M2DC (CEA)	It is an ongoing project combining high efficiency heterogeneous hardware microserver design and production with a concept of System Efficiency Enhancements (SEEs) to improve platform efficiency and decrease TCOs.	High efficiency heterogeneous compute nodes and compilation techniques for heterogeneous targets.
AlpEnDAC (LRZ)	This subproject of the VAO-II collaboration (funding: Bavarian State Ministry of the Environment and Consumer Protection) designs and operates the "Alpine Environmental Data Analysis Centre" - a distributed RDM and simulation platform for high-altitude research facilities such as the "Schneefernerhaus" (Mt. Zugspitze, D). LRZ has been leading the tasks on simulations and on the iRODS distributed-data infrastructure connecting two Computing/Data-Analysis Centres (LRZ, DLR/WDC-RSAT) and several virtual servers (the latter for testing purposes).	The smaller scale - yet geographically distributed - iRODS system within AlpEnDAC can serve as a model for the LEXIS Distributed Data Infrastructure.
GeRDI (LRZ)	Generic Research Data Infrastructure (DFG Project Grant-No. BO818/16-1 and HA2038/6-1): national initiative on creating an interoperable, "FAIR" Research Data Management (RDM) layer, including REST-API-based functionality on cross-site data staging and workflows. GeRDI is not a project with HPC/HPDA focus, but with a focus on proper RDM.	GeRDI's openly available API concept can be built upon to create parts of LEXIS's Data Layer; at the same time LEXIS with its HPC/HPDA capabilities will be perfectly complementary. A collaboration can be set up within the respective groups at LRZ (participating in GeRDI) and then enlarged.
ComPat (LRZ)	Optimisation of Multiscale computing algorithms recognising their "patterns" (H-2020 Project ID 671564): The focus is on efficient coupling of the single-scale building blocks, fault tolerance, and energy consumption. LRZ is managing the e-Infrastructure of ComPat (WP6 / Experimental Execution Environment).	The experience in running a grid-based data-exchange and HPC-orchestration infrastructure for ComPat's special purposes can be built upon in LEXIS to avoid e.g. data-rate bottlenecks by design.
PRACE (IT4I, LRZ)	Partnership for Advanced Computing - 5IP (H-2020 Project ID 730913; continuation of PRACE-1IP/-2IP/-3IP/-4IP): This fundamental European HPC collaboration uses/runs pan-European data-exchange and AAI infrastructures for HPC on practically all European academic HPC centres. It orchestrates and disseminates HPC usage.	PRACE serves as a pan-European collaboration platform and major landmark when it comes to setting up additional functionalities as proposed in LEXIS.

1.3.5 Gender equality approach in LEXIS

The project partners represent an exceptionally strong multidisciplinary team consisting of researchers with degrees in electrical and electronics engineering, telecommunications, computer science, information technology. Of course, human users of LEXIS platform include male as well as female persons. The tests with real-life applications are very important, the consortium will take care that proper gender balance will be taken into account. Women's participation in research must be encouraged. For this reason the project aims to progressively include a larger number of female researchers when recruiting (e.g., PhD students). Gender issues have been fully considered during the proposal preparation and the aim of all the partners of the consortium is to promote equal opportunities and rights for all individuals working on the project. All partners will conceive specific incentives concerning infrastructures to make the working environment more attractive for women. Overall, we aim at creating an atmosphere within our project that all researchers feel welcomed and part of every process. In particular, the project will take care that the gender issues will be covered in workshops and seminars and contribute to changing the techno-centric image of the ICT field by involving female developers in all the activities.

1.4 Ambition

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 like aeronautics manufacturing, 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 BD/HPC platform, to eventually create large value from data (for instance, "providing easier, faster, and more accurate analytics services that generate really valuable information for stakeholders").

The industrial commitment in the LEXIS project is very strong. An entire work package is dedicated to the implementation of and is led by the respective main stakeholder (Avio Aero, CEA, CIMA+ECMWF). The main purpose for these three partners is to industrialize the results of their WP (and of the whole project) to improve not only their competitiveness within their sector, but also the leadership of European industry. For this purpose, the benefits of the project outcomes will be extended also to other stakeholders as follows.

- In the AERONAUTICS test-bed, third parties, like aircraft manufactures and their clients, will benefit of the key outputs, expected in (I) aero engines shortened entry into service and (II) improved quality of engineering design, able to cope with highly demanding requirements matching market expectations.
- In the WEATHER AND CLIMATE test-bed, prediction services with increased accuracy will be provided to private sector entities including in the reinsurance, food and agriculture, health, and energy sectors.
- In the EARTHQUAKE AND TSUNAMI test-bed, the *TsunAWI* simulator is open source. This is an advantage in terms of accessibility for several organisations. Moreover, any interested tsunami warning centre or civil protection agency can use *TsunAWI*, and AWI can provide support on a non-profit basis.

1.4.1 Progress beyond the State of the Art

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

Current state of the art in the context of aeronautics manufacturing. The CFD runs are currently hosted by internal HPC resources. The main issue is correlated to the impressive need to scale to bigger data sets, strongly limiting, in cascade, the numbers of optimization loops. ***Progress beyond the state of the art.*** Avio Aero expects a significant cut in CAE solvers' running time able to widen exploration regions and put up-to-date quality products on the market. This will be achieved by leveraging the skills and hi-tech facilities of partners involved in the development of next generation HPC hardware.

Current state of the art in the context of weather and climate test-bed. Forecast and impact models have generally run on dedicated compute clusters or HPC facilities. However, cloud systems are used for a large proportion of access to, and further processing of, large data sets. The lack of effective interoperability between HPC and cloud systems poses an obstacle to effective use of model data. Current operational systems are well suited to assimilate traditional and conventional observations. However, they are not ready to handle large volumes of unstructured observations which are becoming available from IoT and Edge-Computing technologies. ***Progress beyond the state of the art.*** The activities will be improved by new data access patterns to enable filtering and correlation algorithms that deal with the unstructured nature of novel observational data (such as filtering out mobile phones that are indoors). The various models will be extended to exchange their data output with Cloud and HPDA environments. To maximize the interoperability of the four model layers a unified data sharing Weather and Climate Data API (WCDA) will be developed. All services will use this to provide data access, and models will use it to exchange data between the model layers. Finally, performance will be improved by using in-situ data sources enabled by a new generation of data collector gateway. In the short to medium term, it is anticipated that such a gateway will play a pivotal role in gathering enormous volumes of in-situ, unstructured observational data for use in weather models (from real-time forecasts to real-time back-testing).

Current state of the art in the context of earth and tsunami models simulations. TsunAWI has until now only be used offline on HPC clusters, to calculate scenarios that are later integrated into a database. However, the simpler and faster EasyWave is already integrated as a real time code into InaTEWS. ***Progress beyond the state of the art.*** As there are many parameters to decrease the computation time of TsunAWI, we are sure that we will deliver an inundation simulation that can be run in real time. *TsunAWI* is optimized with respect to efficient cache use as far as it is possible for a 2D shallow water code, but we hardly arrive at vectorizing loops due to the unstructured mesh data structures. The parallelization is pure OpenMP and leaves room for improvement, too. Another significant improvement would be the porting of TsunAWI to GPUs or FPGAs (to be investigated). The OpenBuildingMap currently contains almost 300 million building footprints and additional information. It is updated every minute with changes from OpenStreetMap and other sources. Optimization strategies are necessary to not delay computations due to changes triggering lots of reassessments of building classifications. Likewise, data sources not providing update triggers need to be queried frequently to assure up-to-date information. Resulting damage/loss maps can be augmented with secondary information like average distance from partially destroyed homes (with expected injured people) to the closest working hospital or detailed analyses on affected schools or critical infrastructure/life lines. Such analyses will require e.g. a routing engine and time-of-day dependent whereabouts of people (home, transit, work, leisure).

HPC service orchestration. In LEXIS project, Bull will enhance Ystia Orchestrator with these main features: (I) supporting the whole application lifecycle, from deployment, scaling, monitoring, self-healing, self-scaling to application upgrade; (II) taking into account hybrid infrastructures (IaaS, HPC schedulers, CaaS).

Cross-site synchronization for distributed storage. The LEXIS data system provides a real-time synchronization solution in the sense of a distributed data infrastructure (DDI, LEXIS Unified Storage Library). This solution is to be based on EUDAT B2SAFE (or the underlying iRODS solution) provided by the EUDAT project (H2020 grant agreement #654065). Based on a central metadata registry (iCAT), the DDI provides a unified view on distributed storage systems, organized as *one* virtual file system. With respect to other state-of-the-art distributed file systems, such as Ceph FS or HDFS, EUDAT/iRODS has the advantage of being an extremely versatile, general-purpose middleware.

Data nodes. Burst Buffers solutions are available on the market today (CRAY DataWarp, DDN Infinite Memory Engine). Both solutions require dedicated (and expensive) HW appliances. The data nodes solution proposed by Bull will advance the state of the art because it can be launched on any standard compute node and/or on specialized data nodes, depending on the detailed features required by the application (standard compute nodes will provide IO acceleration while specialized data nodes will provide acceleration AND persistency). In LEXIS project, Bull will enhance Data nodes in 2 areas: management features to facilitate their integration in the LEXIS environment (ability to set up data nodes from *Ystia*); scalability of the data node solution to support the largest use cases. Acceleration and processing technologies like GPUs will also be integrated in the architecture, thanks to the expertise provided by NALL that will improve the HW acceleration capabilities provided by Bull.

Storage Unified data access layer. At the moment, the Big Data, HPDA and HPC landscape with its big societal impact is fragmented. Leveraging the European Research Data Infrastructure EUDAT, LEXIS aims at being a connection between HPC, HPDA and data management. LEXIS sets up a specific test-bed infrastructure connecting three major HPC centers (IT4I, ECMWF, LRZ) and providing a performant, transparent system for Distributed Data Management (LEXIS Data System). For the first time, it will be made possible that LEXIS applications access a unified file system/data management structure.

Access management and monitoring system. Continuous monitoring system are already available on the market. All those solutions are lacking interoperability as for instance they cannot provide continuous monitoring among different type of infrastructure. We will make improvement in this area by providing security checks related to the overall deployment on the different federated HPC infrastructure. The goal is also to provide security guidance for the dedicated environment.

2 Impact

2.1 Expected impacts

Analysts are predicting 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 increases are largely consistent; IDC are estimating 16 zettabytes of usable data in 2020[7].

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[8].

The hockey stick growth of data generation, coupled with the increase demand for organizations to extract value from this data, leads to a projected market value for the European data sector of over Eur100bn by 2020, with a significant growth rate of 15.7% per annum for the latter half of this decade[9], spanning all verticals including those represented in LEXIS – advanced engineering and weather analysis. Other reports note an expected 10% growth rate to 2020[10].

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 consortium is perfectly positioned to address significant aspects of this challenge by developing solutions leveraging its expertise in HPC, data management and cloud technologies and validating them through the pilots in **WP5-6-7**.

The LEXIS project will maximize its **impact** by engaging itself within the value chains of the verticals involved in the project (engineering design, weather forecasting) as well as providing a user-friendly onboarding mechanism which will enable third parties to experiment with and understand the LEXIS platform in the first instance, and ultimately write applications which can leverage the LEXIS data sets and deploy them to the platform.

Impact creation will be considered from the outset in LEXIS and, for this reason, **WP9** will be active from the start date of the project. The project will produce short form content in the form of white papers describing the LEXIS solution which will be circulated to stakeholders in the HPC/ Big Data/ Cloud sectors and the verticals that are being targeted by the project. The Advisory Board will be consulted early on regarding the system architecture which will be developed in **WP2** to obtain early feedback from thought leaders in the different constituencies targeted by the project and will be updated half-yearly on project progress; it is envisioned that Advisory Board members will also be able to identify opportunities for LEXIS and will be able to promote it in their circles.

In order to evaluate the project work effectiveness, a set of criteria and corresponding high-level evaluation has been devised as shown in the following (**Table 2.1a**). The project will perform the evaluation using the specified criteria as the pilot activities evolve and later when generating results from its work.

Table 2.1a. Criteria and the general evaluation related to the project impact

CRITERIA	GENERAL EVALUATION
Relevance	User needs and potential end users, added value to key sectors.

[7] Vernon Turner, John F. Gantz, David Reinsel and Stephen Minton, The digital universe of opportunities: rich data and the increasing value of the Internet of Things, Report from IDC for EMC April 2014.

[8] 3 Top Take-Aways From the Gartner Chief Data Officer Survey, https://www.gartner.com/smarterwithgartner/3-top-take-aways-from-the-gartner-chief-data-officer-survey/

[9] IDC et al., European Data Market, SMART 2013/0063, D9 – Final Report, 1 February 2017, http://datalandscape.eu/study-reports

[10] https://www.prnewswire.com/news-releases/global-big-data-market-2017-2030---76-billion-opportunities-300523816.html

Efficiency	Do LEXIS solutions contribute to an increase in the productivity of the respective sectors?
Effectiveness	What kinds of benefits have we gained by the use of Big Data? Did the project achieve the objectives it set?
Exploitable Assets from the LEXIS project for Large-scale Pilot Community	Target appropriate Stakeholder and community selected by the preliminary assessment.
Market orientation and replication	Investigate potential market and positioning strategy, time to market, and business case.
Relevant exploitable scenarios along Large-scale Pilot	Focus on exploiting LEXIS assets for commercial propositions.
Exploitation roadmap and IPR strategy	Identification of commercial and non-commercial pillar assets to be exploited in combinations with IPR strategy knowledge transfer.
Project outcome	Sustainability plan, potential adoption, product and services portfolio along Large-scale Pilot scenarios.
Value proposition for additional public and private investments	HPC, Big Data and Cloud technologies/ services partnership target.
Value chain technologies integration	HPC, Big Data and Cloud technologies/ services benchmark evaluation for stimulating business process improvements.
Multi-cross sectorial stakeholder on aeronautics/ transport pilot domain	Big Data from the field and regression testing to expedite deep learning of simulation algorithm.
Multi-cross sectorial stakeholder on weather/societal target pilot domain	Enhance and leverage traditional HPC modelling/simulation with big data generated by distributed and unstructured data.

Impact will be achieved along the lines of 4 dimensions, namely dissemination, communication, exploitation, as well as a pilot of specific activities that span across all 3 previous dimensions. LEXIS will address the Expected Impacts set out to the works programme as follows.

Impact 1: *Demonstrated increase of innovation and productivity in the main target sector of the Large Scale Pilot Action;*

In the paragraphs below, we identify how innovation and productivity will be increased in the Aeronautics, Weather/ Climate and Tsunami verticals targeted by the LEXIS Large Scale Pilots.

AERONAUTICS TEST-BED

In the aeronautical market, the demand for reduced fuel consumption is leading to new more complex engine architectures, oriented towards increasing propulsive efficiency.

Therefore, turbomachinery and other mechanical components are forced to operate in unexplored regions of their parametric design space, leading to the definition of parts that are conceptually different from their predecessors. As a consequence, designs for novel jet engine architectures frequently fail to predict component performances. One of the root causes of these problems is associated with simulation software capabilities. New, more complex physical phenomena are studied, with the effect of extremely long simulation run times for predicting performance parameters such as efficiency, compressor stability margin, wall temperatures and heat fluxes.

Efforts to move beyond current limitations involve using more realistic solutions and multi-physics approaches, able to anticipate problems typically encountered in advanced detailed design phases. This implies the adoption of computationally intensive CAE numerical solvers. From a digital perspective, the challenges are to limit as much as possible the execution running time to support, almost in real time, the design evolution and options selection. Furthermore, the big data produced as a result will require appropriate ICT solutions for rapid data access, management and post-processing tasks.

The LEXIS Project is aimed to boost and promote a step-change in the numerical investigations of aeronautical engines' critical parts, focusing in practice on reducing and optimizing overall job execution time as part of the final vision of getting real-time simulation ready. Looking to the wider scenario of European jet engine manufacturers, this change of paradigm (more physics-based computer software, real-time solutions and Big Data analysis) will facilitate a significant increase of productivity and competitiveness.

More time to dedicate to component design exploration will be a key enabler to facilitate greater focus and consequently limit aeronautical engines' Specific Fuel Consumption (SFC) and emission.

Figure 2.1b summarizes the expected trend in the number of performed simulations and gathered output data.

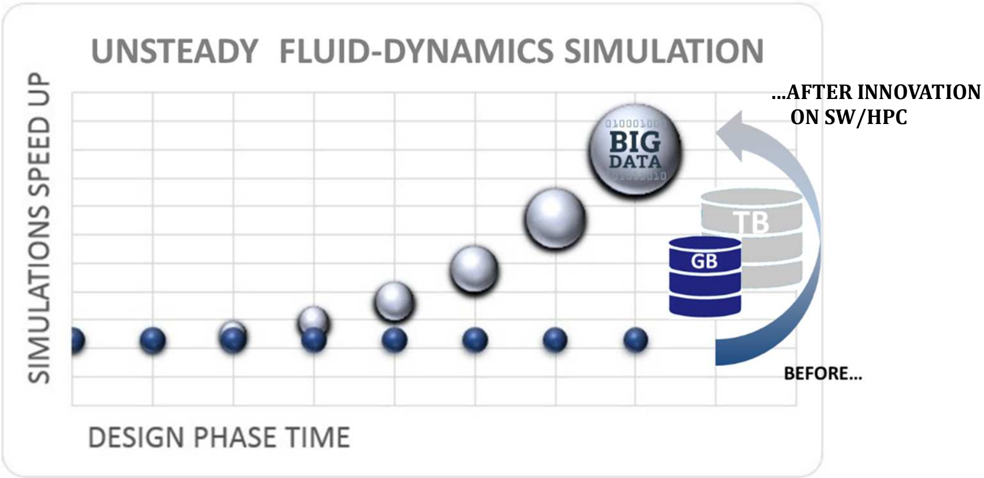

Figure 2.1b Avio Aero expected evolution in simulation runs (#) and impact on post-pro data sizing

The new capabilities noted above – more detailed physics based computer simulations, real-time solution analysis, Big Data processing, and design optimization performed by multiplying case studies – are expected to both decrease time-to-market and lead to increased innovation, in the medium-long term delivering **up to 20% reductions in engine development cost** and **0.3% reduction in engine SFC (design quality)**.

WEATHER AND CLIMATE TEST-BED

The Weather and Climate Pilot will tackle the prediction of water-food-energy nexus phenomena and related socio-economic impacts through the execution of complex, multi-leveled model stacks, including: (1) global weather & climate models, (2) regional weather models, (3) domain-specific application models (e.g. hydrological, drought, wildfire), and (4) impact models providing information for key decision/policy makers.

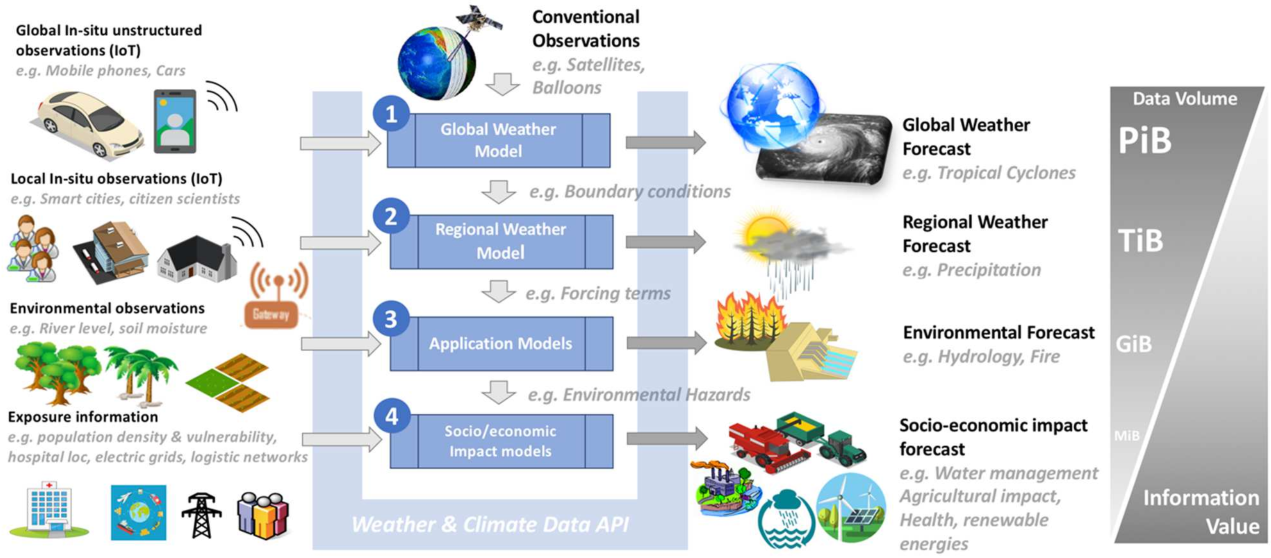

Figure 2.1c Weather & Climate Pilot Test-bed architecture, showing the in-situ unstructured observational inputs, the 4-model layer outputs and the data interchange API.

This pilot will comprise of a test-bed which will innovate weather and climate predictions in two key aspects:

1. Each model layer assimilates diverse types of weather and environment observatsion. Current operational systems are well suited to assimilate traditional observations (e.g. satellite, radio-sondes, balloons, etc) but are not ready to handle observations as provided by emerging new technologies (e.g. IoT sensors, cellphones) and Edge-Computing data sources (e.g. self-driving cars, smart-city systems). The latter require new data-analysis algorithms that deal with their unstructured nature and reliability (such as filtering mobile phones that are indoors). Moreover, data from many of these sources will often reside on heterogeneous Cloud systems, posing a data-collection problem.
2. Each model layer produces its own model output to supply subsequent layers, as well as user-ready products (e.g. atmospheric temperature, precipitation rate or river flow). Until now, most of these models have run in dedicated computer clusters or HPC facilities. We will extend these models to interchange their data output and products with Cloud and HPDA environments. Given that the daily data volumes range up to hundreds of Terabytes (global ensemble forecasts), it remains an unsolved challenge how to efficiently transfer these

massive data sets from the HPC to the Cloud systems, such that customers at the end of the data chain (e.g SME's and small research centres) can build higher value products by post-processing and design business models around these data assets.

To maximise the interoperability of the 4 model layers, both observational data and model output data will be accessed with a unified Weather and Climate Data API (WCDA), also serving as data-transfer API between the model layers. This abstraction provides the interoperability paramount to the success of this use case, as third-party users will describe their data requirements directly in scientific terms without technical overhead. With similar abstraction, the Meteorological Archival and Retrieval System (MARS) by ECMWF handles petabytes of observations and model output. To maximise the success and the impact of the test-bed, we will leverage MARS as a starting point to quickly deliver the WCDA.

EARTHQUAKE AND TSUNAMI TEST-BED

Earthquake and tsunami early warning systems disseminate warning products with estimates of e.g. expected shaking (for earthquakes) or maximum wave height at the coast and the arrival time (for tsunamis). While earthquake early warning centres determine the time and strength of shaking for distant target areas, many tsunami warning centres follow a combined approach of precomputed tsunami scenarios for the most probable earthquake sources and real-time computation with simplified numerical tsunami models. The Indonesian Tsunami Early Warning System InaTEWS for example relies on a data base of 16.000 scenarios precomputed with TsunAWI and on the fast real time model EasyWave[11]. Each TsunAWI simulation covers all Indonesian Seas and includes the flooding on land with a mesh resolution of just 50m to 250m, but it takes hours to compute, while the less accurate/detailed EasyWave model just takes a few seconds.

With TsunAWI to be accelerated within LEXIS, and connections to providers of earthquake early warning alerts to be extended, GFZ will be able to provide (near) real-time damage and loss assessments (with additional focus on critical infrastructure). With detailed earthquake source parameters available, InaTEWS will be able to disseminate maps with the estimated extent of the inundation using the speed-improved TsunaAWI code. From that, another set of damage and loss assessments will be computed by GFZ, resulting in a combined earthquake shaking and tsunami damage/loss assessment. This helps to answer questions such as: which part of the infrastructure has most probability to be affected by the earthquake or tsunami? which roads are probably blocked? where is the most damage and loss to be expected?

The pilot demonstrates a real-time processing workflow, with a combination of tools and code that both run permanently, keeping a global and dynamic exposure database up-to-date at all times (GFZ), and event triggered codes (TsunAWI, EasyWave) that run on specific events, providing near real-time assessments of the severity of natural catastrophes. These assessments are triggering further response downstream (GFZ damage/loss assessment code which will use these assessments, and ITHACA emergency mapping analyses) with products (damage/loss distribution, state of critical facilities) that are delivered to emergency responders in time for decision making and rescue dispatching. With more information about the catastrophic event and its extent becoming available, updated information will be made available to the respective entities.

This pilot will demonstrate the ability of near real-time multi-hazard (in this case earthquakes and tsunamis) and high-resolution damage/loss estimation on the building scale based on earthquake and tsunami early warning alerts. The LEXIS system will show how such systems can become feasible and foster implementation at institutions disseminating earthquake and tsunami early warnings.

Impact 2: *Increase of market share of Big Data technology providers if implemented commercially within the main target sector of the Large Scale Pilot Action;*

LEXIS comprises of 4 SMEs and 4 large enterprises who have an interest in developing their market share within the target sectors through project activities as outlined below. Further, research organizations and HPC centres involved in the project identify increasing commercial potential as Big Data demands increase.

Technological industries & SMEs. The vision of **Bull** is to design & build Insight Platforms: from intelligent machines to intelligence of things. Bull's focus on analytics & security and their applications in enterprise, the IoT/Industry 4.0/Defence/Cyber security. LEXIS will provide Bull the opportunity to develop new solutions in the HPC, Big Data analytics and Deep learning markets.

NALL is working on both the hardware design and the software side of acceleration. The company's objective is to move toward a solution approach by providing integrated designs for specific markets, such as Deep Learning. The work that will be conducted during LEXIS is a perfect opportunity for NALL to work on large-scale end-to-end use cases to integrate technology components which are being developed.

[11] https://docs.gempa.de/toast/current/apps/easywave.html

BAY as a provider of training and enabling know-how to several of the larger cluster OEMS, will use LEXIS in the enabling of the end-users of the HPC data centres it is engaged with.
TESEO vision is that collectors gateways will be an enabling building block within HPC architectures. LEXIS is going to act as innovation accelerator in TESEO hardware platform strategy for products devoted to support HPC, AI and Deep Learning.
CYC offers advanced accounting and billing solutions for heterogeneous computing environments with a strong focus on usage-based billing. Analytics work has a strong intrinsic usage component and CYC hopes to capture market share in the nascent market for accounting and billing in an analytics context. Key to this is integration with the right platforms such as LEXIS.
SEC will focus on the rapidly growing security segment pertaining to virtualization technologies, developing event correlation mechanisms to identify security incidents in distributed Large-scale platforms, and will identify how to market these innovative solutions within the verticals targeted by the pilots as well more generally to the HPC/Big Data/Cloud sector.
AERONAUTICS TEST-BED. Leveraging the newly designed HPC/Big Data/Cloud platform, the proposed aeronautics test-bed will collect/produce more effective computer-aided engineering analysis data. The Big Data technology that will be developed to handle these data, will help to analyze the collected engineering simulations data by managing, processing, and visualizing such assets in the quickest way possible. Successful demonstration of its usefulness will create new market opportunities for Big Data technology providers.
WEATHER AND EARTH TEST-BED. Assimilation of Big Data coming from unstructured in-situ observations (e.g. IoT sources) greatly improves prediction accuracy on areas more populated and potentially more exposed to natural and man-made hazards. The US National Weather Service Enterprise Analysis Report (June 2017) estimates the value of weather data to the agriculture and utilities industry to be worth $13B and we all agree that the food industry is key for Europe competitiveness. Serving through Cloud, the high-resolution NWP coming from ECMWF, and prediction at regional levels will enable end user defined post-processing workflows directly on high-resolution weather forecasts, providing greater confidence on the prevision results.
EARTHQUAKE AND TSUNAMI TEST-BED will demonstrate the ability to respect deadlines, either life-critical or environment critical, or more simply business critical, and it will offer Big Data providers new markets and possibilities to monetize data linked to those constraints. In the test-bed, real-time data availability will be demonstrated, within the limits of the underlying OS and hardware technology.

Impact 3: *Effective integration of HPC/BD/Cloud technologies in the main target sector(s) of the Large Scale Action, resulting into integrated value chains and efficient business processes of the participating organizations*;

LEXIS will lead to improved business processes in the sectors targeted by the pilots. Further, once the LEXIS solutions are integrated into the workflow of the organizations behind each of the pilots, opportunities for more comprehensive value chain integration will be explored perhaps involving also customers and business partners.
AERONAUTICS TEST-BED. Taking advantage of the newly designed HPC/Cloud, the Aeronautics test-bed proposed in this project will be mapped to more efficient business processes by speeding up the execution of sophisticated computer-aided engineering software tools, aimed to simulate the fluid-dynamic behavior of aeronautical engines' components. The expected reduction in execution time and increased capability to boost the software scalability due to improved HPC and Cloud-based technologies will allow an increase of the number of simulations that can be carried out in a specific time frame and to investigate unexplored regions of the parametric design space. The BD collected from such simulations will enable improvement of the engineering design process, thanks also to an enhanced post-processing phase designed to handle and rationalize the large amount of simulation outputs.
It is expected that new technologies integrating HPC, BD and Cloud services will accelerate the design of current and future aeronautical engines, allowing better support for the Avio Aero decision-making process. This is a fundamental step to guarantee that Avio Aero products are innovative, competitive and market appealing, with less environmental impacts for worldwide communities too.
WEATHER AND CLIMATE TEST-BED. Availability of an integrated HPC/BD/Cloud/IoT platform enables the execution of complex workflows for the improved prediction of water-food-energy nexus phenomena (flood, flash-flood, firestorms, wind, hail storms). In this way, the activities in LEXIS complement and support existing operational Copernicus services, forthcoming EUROGEOSS ones, and private ones. All in all, the agenda of this test-bed WP thus also helps to optimise industrial activities (e.g. food & renewable energy production, water resources management, hazard mitigation). For example, the **ITHACA** Emergency Mapping portfolio will be enriched by providing semi-automated analyses on the estimated impact of the exposed asset/infrastructure, integrating the outputs of flood/fire models and socio/economical models. **NUM** will use the test-bed results (high-precision weather

forecast data) to improve the accuracy of its operational system to survey and forecast air quality impacts over industrial sites (Plum'air ®) and cities (Urban'air ®). On this basis, a model to estimate health and economic impact of urban air quality maps will be developed.

EARTHQUAKE AND TSUNAMI TEST-BED. The pilot on the earthquake and tsunami test-bed will show how HPC (compute resources, real-time) and Big Data technologies (data access, data fusion, data production, real-time) can be combined. Since the pilot is inherently about reaction to events provided by sensors, the IoT aspect is absolutely prominent in it, with the InaTEWS system providing the blueprint for the sensors integration; partners in this pilot foresee extensions in the IoT area in the near future.

Impact 4: *Widening the use of and facilitating the access to advanced HPC, Big Data and Cloud infrastructures stimulating the emergence of the data economy in Europe;*

LEXIS will reduce the barrier to entry for users who wish to gain access to HPC/Big Data/Cloud systems through the activities in **WP8** which will focus on providing a portal which will enable users to easily interact with data assets provided by the platforms and deploy basic jobs in a limited context. Having performed such a validation, users can contact the LEXIS consortium to obtain deeper access to the LEXIS platform.

Supercomputing resources providers. Exascale HPC systems, as their predecessors, have been traditionally utilised to their full extent by military and academic users, with engineering companies being one of the few effective commercial users. With the advent of HDPA and Big Data, a variety of parallelisable, compute-intensive data-analysis applications are driving the industry. This is not specific to a sector such as engineering, but pertains to Biotech companies as well as to Aviation, Car-Navigation and Social-Media companies. LEXIS brings HPC and HPDA on large and extremely strong supercomputers, as they are known in traditional HPC, to industry and SMEs in all the growing data-intensive fields. It provides not only support for well-defined use cases from the proposal, but generalises the methods such that attractive offers can be made to companies. Thus, it has very strong potential to develop the participating HPC centres into computing platforms and knowledge hubs for the emerging data economy, strengthening at the same time the knowledge transfer between academia and industry.

Cloud. Tight integration between cloud and HPC centres around large data sets is non-trivial, primarily due to the challenges relating to transferring large amounts of data. The initial approach to be taken within LEXIS is to use the cloud as an entry point for the LEXIS solution, meaning that third parties can understand what data sets are available within the LEXIS platform and understand how to work with them and upload test jobs which can be run within the cloud context. Sample subsets of the data sets will be available to work with within there for test and evaluation purposes. Once third parties have demonstrated that their workload works well with the available sample data set, a small job can then be scheduled to run within the larger LEXIS platform and scheduled within the HPC environment working with the larger data set. As the project evolves, further investigation of tighter coupling will take place where larger data subsets can be placed on the cloud and compute can be subdivided between cloud and HPC resources.

General availability of LEXIS components. The technology underpinning the LEXIS platform will be available to other entities, e.g. other HPC centres; a few components on a commercial basis, more on an Open Source basis. These third parties can use it to provide a hybrid/converged platform for HPC/Big Data/Cloud markets targeting different verticals, enabling ecosystems in other verticals to evolve generally and more specifically, ecosystems around large data sets.

Impact 5: *Stimulating additional private and public target investments in HPC and Big Data technologies from industry, Member States and Associated Countries, and other sources, as referred to in the contractual arrangements of the HPC and/or the Big Data Value Public Private Partnerships.*

Big Data applications are expected to increase their compute-intensity by using new algorithms for descriptive (data aggregation and mining) and predictive (statistics and forecasting) analysis. The HPC driven approach is expected to bring an advantage for faster decision making. On the other hand, data analytics will become a key component of HPC ecosystems: (1) on output - to process results of massively parallel simulations; (2) on input – to feed numerical models with boundary conditions from large numbers of IoT sensors or results of other scientific applications. Especially in dynamic scenarios the Big Data Management approaches will enrich traditional HPC, and thus constitute a strong stimulus for economy.

Based on the latest reports[12] the global HPC server market is supposed to grow by 32.3% from 2016 to 2021, while the broader market (including middleware, applications and services) should grow at a Compound Annual Growth

[12] (1) High Performance Computing in the EU: Progress on the Implementation of the European HPC Strategy, Final Report, Contract number: 30-CE-0663100/00-22, SMART number: 2014/0021, (2) Major Trends in the Worldwide HPC Market, Presentation at HPC User Forum at HLRS, February28 -March 1, 2017, Earl Joseph and Steve Conway, (3) Hyperion - Trends in the Worldwide HPC Market, presented at ISC17, June 2017

Rate (CAGR) of 6.2%, or by 34.8%. The European HPC server market is supposed to grow at a CAGR of 6.2% or 35% in total in the same period. A new Hyperion Research study *"The Status and Prospects of European Suppliers of High Performance Computing (HPC) Products and Services*" anticipates that for European HPC market (servers, storage, software, networks) the indigenous European suppliers will grow their share from 6.4 % in 2016 to 11.3% in 2021 (from 337 million Euro in 2016 to 839 million Euro in 2021). European industrial-academic cooperations such as LEXIS are pushing this forward.

LEXIS will be closely integrated in other European initiatives such as ETP4HPC [http://www.etp4hpc.eu], EuroHPC [http://www.eurohpc.eu] and PRACE [http://www.prace-ri.eu]. These initiatives are the front runners in establishing HPC within Europe and providing resources to European researchers and industry. With IT4I and LRZ, two partners of the LEXIS consortium are members of these initiatives. They will work together with the partners to integrate the LEXIS system into national and European efforts.

The **ETP4HPC Strategic Research Agenda** (SRA) [ETP4HPCSRA] is a tool to steer the HPC research in Europe to maximize the return on investment for Europe's economy. This vision coincides with the goal of this proposal: to build a set of tools (software) that enables more efficient and easier use of the HPC and Big Data infrastructure. To be more precise, the current research areas laid out in the ETP4HPC SRA 3[13] are a) HPC system architecture, b) System software and management, c) Programming environment, d) Energy and resiliency, e) Balance compute, I/O and storage performance, f) Big Data and HPC usage models, g) Mathematics and algorithms, h) Extreme-scale demonstrators, and i) non-technical areas, such as education/training, SMEs and Start-ups.

In March 2017, initial member states signed the **EuroHPC declaration**, a collaboration between currently 15 European countries and the European Union focused on developing and supporting Exascale supercomputing by 2022/2023. The main pillars of EuroHPC are[14]:

1. HPC R&I towards development next generation of key hardware and software components and technologies including system software and applications for exascale and post exascale.
2. Infrastructure development, acquisition of world class supercomputers, data infrastructures and interconnects.
3. Support to applications and HPC related skills development.
4. Federation of national and European HPC resources through an HPC and Big Data service infrastructure facility.
5. Support to HPC, CoEs for developing, preparing and optimising HPC codes and applications for future exascale and post-exascale systems in co-design.

LEXIS will have direct impact on Pillars 1 and 2. With respect to Pillar 3 our training activities will educate new HPC users and application developers that will be able to use proposed infrastructure. Pillar 4 is also directly impacted by LEXIS, as LEXIS maximises overlap of HPC, HPDA and Cloud areas with optimised processing efficiency. Finally, Pillar 5 is addressed implementing three large test-beds on the LEXIS system.

The **Partnership for Advanced Computing in Europe (PRACE)** is a major European provider of supercomputer resources, application expertise and training. As such, it is potentially a major customer for LEXIS technologies. LEXIS will interact with PRACE to align the development of the LEXIS platform with the HPC infrastructure in the EU. To enable this, PRACE institutions will be approached to join the LEXIS supervisory board to steer the project and learn of its results.

IT4I and LRZ are members of PRACE, providing resources, organizing trainings and driving forward the PRACE initiative. With SuperMUC and its future successor and 26PFLOP/s machine SuperMUC-NG, LRZ hosts one of the seven PRACE-Tier-0 machines, giving HPC-access to European researchers and industry. With 2PFLOP/s Salomon (PRACE Tier-1) and 94 TFLOP/s Anselm machines IT4I operates the biggest supercomputing facility in the Czech Republic. In 2019 IT4I is planning to integrate a successor of the Anselm system. IT4I and LRZ will incorporate the LEXIS technologies into these infrastructures and investigate means to incorporate them into their regular PRACE training programs.

LEXIS will be closely integrated with the **Big Data Value Association (BDVA)**, the main international association for industry-driven R&I on Big Data in the European context, where ISMB is member. The *Strategic Research and Innovation Agenda* (SRIA) is a document periodically issued by BDVA, which defines the overall goals, priorities, and a R&I roadmap for the European Public Private Partnership (PPP) on Big Data Value. At least four technical priority areas of the SRIA will be addressed by LEXIS:

- *Data Management*, with the LEXIS distributed data infrastructure and data system APIs;

[13] http://www.etp4hpc.eu/sra-2017.html

[14] https://ec.europa.eu/digital-single-market/en/policies/high-performance-computing

- *Data Processing Architectures*, as LEXIS will combine both data-in-motion (real time streams) and data-at-rest in its test-beds;
- *Data Analytics*, which will be a main topic in all the three test-beds;
- *Data Visualisation and User Interaction* by means of graphical tools and easy-access UIs.

LEXIS will validate its integrated HPC/cloud/BD architecture against existing common standards and best practices so that it constitutes a standard solution for the industrial sectors of the test-beds. Adopters benefit in terms of *interoperability* with external platforms and of *seamless flow of data* with stakeholders. The active participation of LEXIS consortium in the BDVA will ease the creation of a community such adaopters. On the other hand, the input from BDVA SRIA will represent a strategic value for LEXIS development.
All three European initiatives have either direct or indirect objectives to identify benefits for the industry and SMEs. For LEXIS, such objectives are extended further to increase interest and contributions from the private and industrial sectors to HPC and Big Data. Such efforts are not only limited to the LEXIS project itself but also will leverage the partner network to guarantee a wide coverage. This allows the combination of collecting feedback as well as attracting new industries and SMEs to the project. During this process, business needs can be steadily identified and updated to define new use cases/applications for monetization.

KPI	Target
Improve the running times of the LP turbine simulations	Decrease the running time of the Unsteady Reynolds Averaged Navier-Stokes simulations by 20%
Speed up the analysis process for the gear-boxes simulations	Decrease the running time of the gear-box simulations by 20%
Decrease the time needed to produce precise data products	Decrease the time of simulation for the emergency decision process by at least 50%
Increase the speed of OpenBuildingMap updates	Make the time of the OpenBuildingMap updates 50% shorter
Increase the granularity and precision of the weather and climate forecasts	Increase the amount of the data produced by the forecasts by 50%
Increase the effectivity of distributed computation of the weather and climate models	Decrease the data accesses by 15%
Shorter time for administration of HPC resources billing and accounting	Decrease the time of the administration process at HPC centre when selling HPC resources by 50%, currently it is several days

2.2 Measures to maximise impact

a) Dissemination

The following approaches will be used to disseminate the scientific and technological innovations results (Table 2.2a) of the LEXIS project, building on the experience of the consortium partners.

- *Dissemination events (such as conferences, workshops, Ph.D. forums and tutorials):* Project participants are already active in the HPC community. The project participants will organize co-located tutorials, workshops and seminars at established conferences to support the transfer of knowledge in the international research community. Presentations of the LEXIS technology, its components, and demonstrators will be given at international conferences and exhibitions such as the European HPC Summit Week, SC, ISC, HPCA, PACT, IPDPS, EuroPar, and HiPEAC.
- *Publication of research results:* Project results and innovations will be submitted for publication in scientific journals (e.g. IEEE Trans. on Computers, IEEE Trans. on Parallel and Distributed Systems, IEEE TCAD, IEEE Design & Test and IEEE Computer, ACM Transactions on Architecture and Code Optimization (TACO), ACM Transactions on Parallel Computing, ACM TODAES, Parallel Computing (Elsevier), Journal of Systems Architecture (Elsevier), Journal of Parallel and Distributed Computing (Elsevier), International Journal of Parallel Programming (Springer), Concurrency and Computation: Practice and Experience, Wiley Software: Practice and Experience), conferences (e.g. SC, ISC, PACT, HPCA, IEEE HPEC, IEEE HiPC, EuroPar, DAC, CGO, LCTES, FDL, PPoPP, IPDPS), and workshops relevant to the topic of the project research activity. LEXIS partners will ensure dissemination of peer-reviewed scientific publications which might result from the project by using the self-archiving *"green" open access.* The published or the final peer-reviewed manuscripts will be archived by the researchers (or by their institutions) in an online repository linked from the LEXIS public project website. Submission of papers jointly written by project partners will

be encouraged. For consortium-wide publications, the *"gold" open access model* will be applied to ensure a prompt and wide dissemination of the project outcomes.

Table 2.2.a Partners' specific dissemination assets and intentions

IT4I	IT4I will participate at conferences and help with solving practical issues involved in organizing booths at relevant exhibitions. IT4I research team involved in the project will submit papers and posters to relevant conferences and publications to scientific journals. IT4I researchers will contribute to the preparation and execution of the tutorials and seminars organized to disseminate the project outcomes.
CYC	CYC will promote the work of the project within the technical community through attendance at industry events focusing on the confluence of Open Source, Big Data and Billing. CYC will promote the work of the project directly within the Zurich area which has a vibrant Data Analytics and AI/ML community by speaking at local meetups such as Big Data Developers Switzerland, and the Swiss Big Data User group [15]. CYC will also apply to speak at events such as European Big Data Value Forum and DataWorks Summit to focusing on the LEXIS accounting and billing solutions.
Bull	Bull will work on cross-dissemination with SAGE2. Bull will also take advantage of ISC High Performance trade shows to promote LEXIS.
ISMB	ISMB will work on promoting LEXIS through international conferences and communities mainly form HiPEAC main conference, HiPEAC CSW, the Heterogeneity Alliance, participation for dissemination to ISC and SC High Performance Computing, Networking & Storage, CISIS (Int. Conf. on Complex, Intelligent and Software Intensive Systems).
CEA	CEA will present the project at scientific conferences through research papers and presentations, will publish in peer-reviewed journals, and present at trade shows (Ter@tec, ISC, SuperComputing, CEA own tradeshow events). CEA will actively extract technological components out of its effort and search for interested parties to license or release it. CEA will also participate in specific workshops dedicated to the dissemination of the Earthquake and tsunami pilot results.
NALL	NALL will participate to conferences to demonstrate the work which is being conducted during LEXIS on the off-loading of applications to an FPGA, either by leveraging OpenCL or some embedded and optimized kernels. Targeted conferences are HPC events such as Super Computing and Computing Insight. NALL will also be writing additional whitepapers on how to efficiently use FPGA technologies into large HPC and Cloud deployments.
BAY	BAY will use the LEXIS platform in its training and consulting activities with its key customers – accounts include verticals in Life Sciences, Oil & Gas, Financial Services Industry, and Defence. In addition, BAY will develop bespoke training for key LEXIS users (commercial and not) in order to promote best usage of the framework. Dissemination will also include creation and promotion of hands-on training, videos and white papers.
TESEO	TESEO objective is to promote LEXIS through enhanced awareness both in National/Regional initiatives and organisms like Digital Innovation Hub and Industry 4.0 cluster, including Confindustria specific workgroup, and in Industrial specific events like European recognised Intelligent Transport Systems (ITS), and major Italian events such as Affidabilità & Tecnologie and Aeropspace and Defence. Considered resources and training are the key factor for technology adoption, TESEO will organised specific actions for vocational training (sustaining employability) which is another major EU roadmap objective. At this stage the envisioned and identified entities to leverage on are Fondazione ITS Meccatronica e Aerospazio Piemonte and ASSOCAM Camerana (IBM Big Data and AI Training Center) with the objective to train employees on these new technologies.
Avio Aero	Main results from this R&D project will be actively disseminated within the Company at two different levels, engineering and IT specialists and technology management. Internal workshops and technology conferences will trace the path to improve engineering approaches. RANS modelling could be abandoned, taking advantage of more physics-based, CPU-intensive approaches and creating unprecedented opportunities for aeronautical products investigation and optimization. Dissemination in the scientific community will be taken care of through high-end scientific journals and/or peer-reviewed international conferences, with contributions supported by the University of Florence that will participate as Avio Aero sub-contractor in LEXIS.

15 https://www.meetup.com/Big-Data-Developers-Switzerland/ and https://www.meetup.com/swiss-big-data/

GFZ	The improvements implemented within the LEXIS project will allow GFZ to disseminate the developed exposure model as web service to all major institutions working on various hazards and their respective risk. Particular partners involve the Global Earthquake Model but also regional institutions. GFZ will publish the results of the test-bed experiments in the respective peer-reviewed scientific journals and present them at scientific conferences.
AWI / CEA / GFZ /	The earthquake and tsunami pilot partners are focusing their dissemination actions on the users of the current InaTEWS system, as well as appropriate actors in the earthquake stakeholders, and intend to disseminate the results with the users of InaTEWS with workshops and trainings. In particular, BMKG, but also BPPT and BNPB, will be targeted, as well as civil protection entities (first responders) in Europe. Further extension will be planned with the addition of the University of Tokyo as third party to the project.
CIMA	CIMA will disseminate the outcome of the Weather & Climate test-bed in academic journals and in presentations and papers at conferences, as well as through CIMA's website. We plan to adopt and promote the weather and climate data access API (WCDA) in existing and future platform for the benefit of, and to spread it among, our clients.
ECMWF	ECMWF will publish its results in academic journals and in presentations and papers at conferences; publish and release the weather and climate data access API (WCDA) as open-source software; throughout the network of member states enact technology dissemination by publishing in ECMWF's newsletter and internal NWP workshops; and where appropriate, apply the developed the software infrastructure to the daily operational weather forecasting system for the benefit of our clients.
ITHACA	ITHACA will communicate the main outcomes of the Emergency Mapping related use cases in presentations at conferences and workshops, including informing the Copernicus EMS Rapid Mapping Service Provider consortium and the relevant Entrusted entity (as long as ITHACA is involved in the Service). ITHACA aims to contribute to the publication of any possible relevant project outcome in indexed journals. All the project-related information will be disseminated through ITHACA's website and social media.
LRZ	LRZ will present the project at scientific conferences through research papers and presentations and will publish in peer-reviewed journals. In addition, LRZ will promote the tools developed in LEXIS to his customers through their regular PR channels, reaching a large scientific community of HPC users and potential future applications for the LEXIS workflow. LRZ will focus its dissemination efforts on data management and data flow tools, promoting the data exchange layers of LEXIS.
SEC	SEC is a founding member of the Cloud Security Alliance and co-author of the first guidelines on cloud security. LEXIS will leverage this community to promote the projects and the results. In addition, Eurocloud, Clusir and Telecom Valley, which SEC is also a member of, will be platforms for disseminating LEXIS and foster adoption among SMEs. Through a web-site section and our social network presence (Twitter, Linkedin and Facebook), LEXIS will be given ample visibility. According to allocated budget, SEC will participate as speaker to major security events in Europe such as BlackHat Europe or InfoSecurity in order to present and promote the project, its pilots and its findings.
AWI	Besides the technology and knowledge transfer to stakeholders in Indonesia and Europe as listed above, AWI will present the project at scientific conferences and will publish in peer-reviewed journals. In addition, articles for a broader public will be published through the Earth System Knowledge Platform eskp.de. The expertise gained on optimized workflows will be promoted within AWI and the scientific community.

b) Exploitation of results

The exploitation plan is based on close collaboration of Europe's supercomputing centres and industrial partners that gives huge potential to exploit the project results through their established partnerships with international industrial partners, research institutions and universities. The main plan of the industrial partners involved in the project will be to use solutions and tools developed within the project to support the development of new products. Research partners with industry partners will make LEXIS technology available either to external industrial and end-user partners or as open-source modules to the SMEs/HPC/Academia community worldwide. Exploitation activities in **WP9** will be intensified during the last year of the project, when results will be verified by Open Call.

The ***partners*** who will share the results of the LEXIS project are the consortium members mentioned in the **Section 3.**, particularly in the **Section 3.3**. The consortium consists of companies, supercomputing centres, research institutes, and universities to cover the whole spectrum of technological innovations supported by scientific outcomes and that the project relies on, and to leverage the varying dissemination expertise and community links of the project members for maximum impact.

The more detailed ***exploitation plan*** will be created and continuously updated during the project as well as ***business and the sustainability plan*** of the LEXIS project. The first version of the exploitation and business plan will be prepared for the M15 review. The timetable with milestones is presented in **Section 3.1.1** together with work plan. However, all these strategic documents will be regularly reviewed and updated during and after the project realisation.

Initial exploitation identified for the LEXIS project

Products and Services: LEXIS technologies (hardware and software), which combine HPC/Cloud/BD state of the art technologies, integrated into the HPC infrastructures of the project partners will extend portfolio of their services towards SMEs, industry and academia and increase supercomputing centres competitiveness in relation to the cloud providers. Another mean of exploitation will be by conducting training activities towards national as well as international commercial sphere. These training activities will include the results of the project with the aim to deepen the knowledge about the project within the industrial community.

The data layer developed within LEXIS, along with the possibilities for workflow orchestration on virtualised HPC/HPDA environments, can form an integral part of supercomputer centres strategy to run its future supercomputer installations from ~2021, when new systems will be available. The whole portfolio of LEXIS Services is supposedly extremely interesting for scientists to enhance their workflows, also with methods driven by IT- or IT-heavy industry. On the other hand, we expect the LEXIS services to attract industry (especially SMEs) to use supercomputer infrastructures. In general, we expect that this fusion of the non-academic and academic HPC/HPDA world will lead to an innovation boost with HPC centres acting as a technology and knowledge hubs.

Industrial partners will exploit LEXIS technologies in many ways. LEXIS will provide them the opportunity to develop new solutions in the HPC/Big Data analytics/AI markets while enriching their existing know-hows in the related IT domains. The anticipated exploitable results arising from the project are Hybrid/converged Big Data/HPC/Cloud platform, and data management with a global orchestrator**.** Companies will use LEXIS technologies to create new products/solutions.

The technologies developed within LEXIS will provide both downstream and on-demand data access and data centric processing, in a Cloud/HPDA environment to third-party entities. This will enable to expand data services offer to include Cloud services that run close to model data and enable customers to develop their own post-processing and analysis methods.

Significant potential of Weather and Climate Large-scale Pilot lies in new data collector gateway which collect, filter and broadcast large amounts of data that then feed big data-lakes. Starting from this assessment, the weather and climate pilot is a perfect test-bed due to the variety of data to be treated and the significant volumes required by simulation models. State-of-the-art systems for weather and natural hazard prediction at regional scale are able to improve the prediction of highly impacting natural hazards (e.g. flash-flood, forest fires etc). The integrated HPC/BD/Cloud/IoT technologies can enable the delivery of such services in an operative environment. LEXIS technologies will enable companies to provide more reliable, more accurate a faster services linked to meteorological data, especially in the domain of urban and industrial air quality management and for agricultural production.

Major scope of Aeronautics Large-scale Pilot will be to increase reliability and accuracy/ speed of numerical simulations, contributing not only to reduce the cost of product development, by fulfilling performance just during the first test, but, what is more important, speeding up the innovation cycle (which is now usually based on long, expensive experimental work) by adopting new methods and software capabilities. This change of paradigm should allow European manufacturing companies to substantially increase competitiveness.

The Earthquake and Tsunami Large-scale Pilot brings a significant capability to existing warning systems, promising improved on-line results, a deadline-constrained processing workflow, and real-time translation into damage and loss estimates for communication and informed decision making. Exploitation will focus on direct transmission to clients and stakeholders of current systems: users of InaTEWS and EEW systems around the world but also stakeholders in governments and civil defence. Exploitation is facilitated by practising a short research-to-implementation path by some of the core partners of the pilot, with the mission is to bring societal advances (GFZ, AWI). Private companies, SMEs and other actors will also benefit from the real-time aspects demonstrated by the pilot.

Target group: The primary target group of LEXIS ranging from companies, through authorities and agencies to the research institutions.

Technologies provided by LEXIS will target national and international companies active in HPC, Cloud and Big Data communities as HiPEAC, BDVA, deep learning on high performance hardware recruiting from industrial sectors such as Life Sciences, Oil & Gas, Financial Services Industry, Defence etc.

Agencies and authorities targeted by LEXIS will be Tsunami warning providers, civil protection agencies, local municipalities, national and regional Civil Protection authorities, as well as international Disaster Risk Reduction Institutions (e.g. United Nations Office for Disaster Risk Reduction - UNISDR);
Academia and scientists will recruit from climate science community with HPC and data intensive focus, HPDA-heavy disciplines such as Biology, Traffic Engineering or Astrophysics.
Business plan:
HPC centres will increase the provision of compute resources through LEXIS to the national and international SMEs and larger industrial users. LEXIS technologies will extend their portfolio of services provided to the companies by creating new type of commercial and data centric services. Embracing data management and data sharing across services enables new innovative technologies based on LEXIS in areas such as HPC and Big Data analytics, hydro-meteorological prediction services, weather forecast, renewable energy trading market and security.
Following key aspects are already considered for commercialization and will be extended during the project:

- Better cost efficiency to competing solutions;
- Consulting services for better integration of data, models, algorithms, etc.;
- Improved security (in technical and legal terms);
- Commercializing HPC infrastructure and services by opening them to industry and SMEs;
- Creation of new products and solutions based on existing data and models (e.g. by combination of data and technology or by exploiting synergies between partners);
- Open Data as driver for new services (make sharing data easier and more efficient);
- Extend solutions to new markets and users; a steady dialog with potential users is required for this;
- Feasibility & case studies to demonstrate the economic potential;
- Training & consulting services, and raising awareness about LEXIS across Europe; project partners provide such services and support users to adapt LEXIS technologies;
- Synergies of different domains and industries of different sizes;
- Accounting and billing policies/infrastructure.

For each aspect of LEXIS and potential marketization a market and cost benefit analysis will be carried out. It will contain at least:

- Executive summary, which briefly describes the intent of commercialization;
- A mission statement defining the targeted market, the product or service delivered, and a differentiation in terms of competition and cost/benefit categorization, including a SWOT analysis;
- Description of the business case in detail with analysis of the ecosystem/environment;
- Marketing plan to ensure reaching the right target group and delivering the services/products needed;
- Operations and financial plans to specify the development lifecycle of the products/services in terms of reasonable a time to market and financial window.

c) Communication activities

The LEXIS communications strategy identifies different constituencies who may be interested in the work of the project along with specific communications mechanisms tailored to these constituencies. These are detailed in the table below.

Constituency	Targeted Communications Activities
The general public	• YouTube videos will be created summarising the main benefits of the project and why the partners are involved in the project. • Basic collateral (project flyers, basic slide deck etc) will be developed which captures the key points of the project. • Partners will use their press contacts to promote the work of project in the local and national press. • As a number of the partners have close ties to educational institutions, it is straightforward for the project team to be present at local science fairs etc.
The HPC sector	• Content will be developed specifically for the HPC sector, detailing the LEXIS data management solution; a white paper is envisaged. • Content will be developed specifically for the HPC sector, detailing the LEXIS Cloud/Big Data/HPC integration solution; a white paper is envisaged. • As the project team is well represented within the HPC community and will be in attendance at key industry events (e.g. Supercomputer), the team will

	be able to publicize the LEXIS solution to both operators and users in the sector.
The analytics/ML sector	• Screencasts will be developed showing how to interact with/use the platform to interact with different data sets. • Screencasts will be developed showing how to develop and deploy a simple application which can leverage the LEXIS platform. • Attendance at key industry events targeting large data sets and advanced use of data analytics platforms.
Industry verticals interested in weather	• A white paper explaining how the LEXIS platform can be used within the weather industry will be developed. • The project will be demonstrated at a key weather industry event.
Industry verticals interested in engineering solutions	• A white paper explaining how the LEXIS platform can be used within the engineering industry will be developed. • The project will be demonstrated at a key engineering industry event.
Industry verticals interested in tsunami data	• A white paper explaining how the LEXIS platform can be used within the tsunami sector will be developed. • The project will be demonstrated at a key tsunami sector event.

In addition to the above targeted communications activities, the standard project communications tools and mechanisms will be maintained including:

- Project website.
- Github repository for Open Source software.
- Slideshare repository for project presentations.
- Twitter feed for interacting with others and publishing content relating to the project.
- Blog for publishing content relating to the project.
- YouTube channel for project related videos.

In addition, the External Advisory Board will be leveraged both as a feedback mechanism to determine whether the messages being generated by the project resonate, but also as a channel for qualified communication with other industry thought leaders in the constituencies relevant to LEXIS.

Dissemination and communication activities	Description	KPI
Project website	LEXIS website will be created for both public and private use and will play an important role in disseminating project information to the wider industry.	Bounce rate <30 sec
Social media campaign	Be social nowadays is important and strategic for entering the market and spread messages to stakeholders or end users. For this reason LEXIS will have the principal social media and will be monitored the results constantly.	Followers >250 Comments > 100
Newsletter	Creating a newsletter is an important tool for a continuous communication about the project evolutions and improvements.	>100 of users subscribed the newsletter or Induced click
Organization of workshops	A direct workshops organised directly from the consortium is an important method for showing out the project's outcome.	3 of workshops organised directly by the project due to the project life.
Participation at exhibitions, fairs, seminars, workshops, or conferences	The first immediate tool for communication and dissemination is attending .events/workshops/conferences/seminars	At least 3 per each partner/30 months

d) Leveraging Regional and National Initiatives

Europe, countries and regions are extremely active with initiatives for capturing the Society 4.0 opportunity (Industry 4.0 is too limiting the view angle of the technology disruption we are going through). An important objective then is to engage stakeholders and initiatives and leverage on them in order to broaden the awareness and benefit from integrating the project in the larger picture covered by such initiatives.

Notwithstanding that data may change and be integrated in the life of the project we already envision to leverage on two major pillars of the regional and national initiative: Industry Driven and Human Resource driven. Industry drive are mostly based around the Digital Innovation Hubs and Industry 4.0 cluster, with a significant involvement of Confindustria. Human Resource is another key driver in regional and national initiatives especially about vocational training (sustaining employability). As of today we estimate to target specific actions, coherent with EU roadmap, towards Fondazione ITS Meccatronica e Aerospazio Piemonte and ASSOCAM Camerana (IBM Big Data and AI Training Center).

The Italian national and regional initiatives will be the test-bed for this action which then, properly tuned and adapted, could be extended to other countries, in particular those represented in the consortium. It has to be evaluated in due course of the project the best way to scale up such leverage in order to enhance the impact towards country and regional initiatives.

e) IPR and knowledge management

General IPR Principles

The consortium partners will endeavour to ensure that as much of the project work is open and usable by others as possible; not all the work of the project will be open, however, due to commercial considerations of some of the industrial partners. The non-profit making entities will make their work open and available, including software, documentation and training material.

The Open Source outputs include:

- an improved and enhanced software consisting of contributions to core library, utilities and top-level solvers;
- a central database of test cases including performance and benchmarking results;
- documentation;
- training material.

Software license: When Open Source outputs are being developed, the project will use the Apache 2.0 Open Source License where possible; when working with existing software technologies however, it may be required to use a license already applicable to that technology.

Patents: This project is unlikely to produce patents as its core deliverables are not patentable algorithms and software. However, new innovative products resulting from the use of the improved software can potentially lead to patent applications e.g. by the industrial partners participating in this project. Possibilities for this will be assessed and the commercial partners will write an exploitation plan during the project (see **Task 9.2**).

IPR Management: Innovation Manager will coordinate all activities and negotiations surrounding IPR issues. Before the project starts, the Consortium members will sign a formal **Consortium Agreement (CA)** in which roles, responsibilities and mutual obligations will be defined. The CA will be based on the MCARD template (developed by Digital Europe), and will codify the governance of the project as described above and other legal and IP issues. The purpose of the CA is to establish a legal framework for the project 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 project and any other matters of the consortium's interest. The CA will include in detail: (1) **IP Ownership** - Results shall be owned by the project partner carrying out the work leading to such Results. If any Results are created jointly by at least two project partners and it is not possible to distinguish between the contribution of each of the project partners, such work will be jointly owned by the contributing project partners. The same shall apply if, in the course of carrying out work on the project, an invention is made having two or more contributing parties contributing to it, and it is not possible to separate the individual contributions.

Any joint Results, including inventions and all related patent applications and patents shall be jointly owned by the contributing parties. In order to further the competitiveness of the EU market, and to enhance exploitation of the Consortium Results, each contributing party shall have full own freedom of action to exploit the joint IP as it wishes, and further the goals of the consortium. To promote this effort the contributing party will have full own consideration regarding their use of such joint Results and will be able to exploit the joint IP without the need to account in any way to the other joint contributor(s). (2) **Access Rights to Background and Results -** the project partners will grant each other and their affiliated companies, royalty-free Access Rights to their Background and Results for the execution of the project. This will offer the Partners the possibility to execute the project to the best of their ability, without being hindered by administrative issues. (3) **Consortium Organization** - the internal organization of the

Consortium, its governance structure, decision-making processes, reporting mechanisms, controls, penalties and management arrangements; (4) **Fund distribution** - arrangements for the distribution of funds among participants and among activities; (5) **IPR Management during the project** – will be defined by explicit rules concerning IP ownership, access rights to any Background and Results for the execution of the project and the protection of Intellectual Property Rights (IPRs) and confidential information before the project starts. (6) **Transfer of Results** - As Results are owned by the project partner carrying out the work leading to such Results, each project partner shall have the right to transfer Results to their affiliated companies without prior notification to the other project partners, while always protecting and assuring the Access Rights of the other project partners. Such use of Results will encourage competitiveness of the EU market by creating broader uses of the Results and opening up the markets for the Consortium's Results in all markets. (7) **Open Source and Standards** - Some of the project partners may be either using Open Source code in their deliverables or contributing their deliverables to the Open Source communities. Alternatively, some of the partners may be contributing to Standards, be they open standards or other.

e) Data Management Plan

The project will generate data in a wide range of levels of detail. Most data will be associated with results that may have a potential for commercial or industrial protection and therefore cannot be made accessible for verification and re-use in general due to intellectual property protection measures presented in the previous section. However, the decision concerning the publication of data will be made by the decision-making bodies of the consortium. Maintenance and preservation of the data will be guaranteed by the partners of the consortium, under the responsibility of the Project Coordination. More information on Management of Data will be described in the Data Management Plan. The Data Management Plan will describe the type of the data will be generated or gathered during the project, the standard that will be used, the ways how the data will be exploited for verification of reuse, and how the data will be preserved.

3 Implementation

3.1 Work plan — Work packages, deliverables and milestones

3.1.1 Work packages inter-relation

The LEXIS project will run for 30 months and is divided into 9 work packages (WPs) as shown in **Figure 3.1.1a. WP1** will be responsible for the management of the project and quality assurance of all the outputs. Innovative tools and architecture for the federated combined HPC and Cloud technologies with focus on the Big Data analytics will be designed in the technical WPs (**WP2** – **WP4**). Several improvements to the current technologies will be needed to effectively leverage the data access in the federated heterogeneous environment, and the necessary innovations will be identified and designed in **WP2**. This architecture design will provide the hardware basis for the LEXIS Data System developed in **WP3** and the federated orchestrator developed in **WP4**. The data layer developed in **WP3** will implement tools for effective HPDA on the HPC clusters. Additionally, it will provide tools for data exchange and management, and will create a catalogue of datasets for the test-beds coming from the **WP5** – **WP7** pilots. Since the project is focused on the federated system of the HPC centres, an orchestration tool that is able to optimally distribute workload to the HPC/Cloud will be designed in **WP4**. **WP4** will also consider the security of the whole system, communication protocols and access to the data layer. Finally, tools will be implemented in **WP4** for monitoring jobs run by the orchestrator. LEXIS will include three pilots, each with their own **WP** (**WP5** – **WP7**). In these pilots, novel technologies and the architecture created in **WP2** – **WP4** will be tested. These pilots will provide feedback to the technical **WPs** to ensure that the final architecture and tools will be able to support a range of use cases and create a complex test-bed environment for the future users. Pilots will also provide datasets for the test-bed which will be incorporated into **WP3**. In **WP8** a portal will be developed for the provisioning of the test-bed services to third parties. This will include a web-based interface for the selection of datasets, a framework for easy deployment of applications to the infrastructure and the creation of accounting and billing policies. The impact of the LEXIS project on the market will analysed on an ongoing basis to ensure the success and relevance of the final system. This analysis will be main part of the **WP9**. **Table 3.1a** summarizes the work packages.

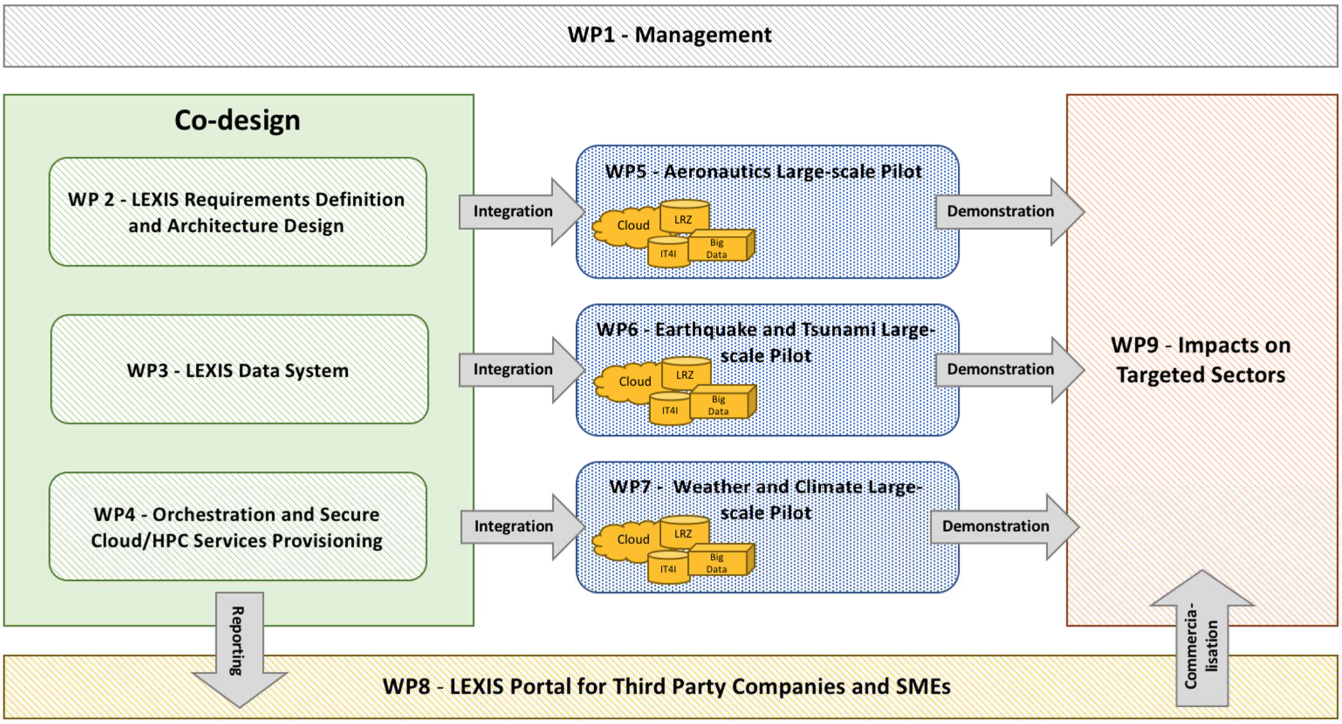

Figure 3.1.1a Work Packages Inter-relation in LEXIS

Timing of the activities divided into particular tasks in each WP are shown in **Figure 3.1.1b** in Gantt chart. It also includes deliverables and milestones.

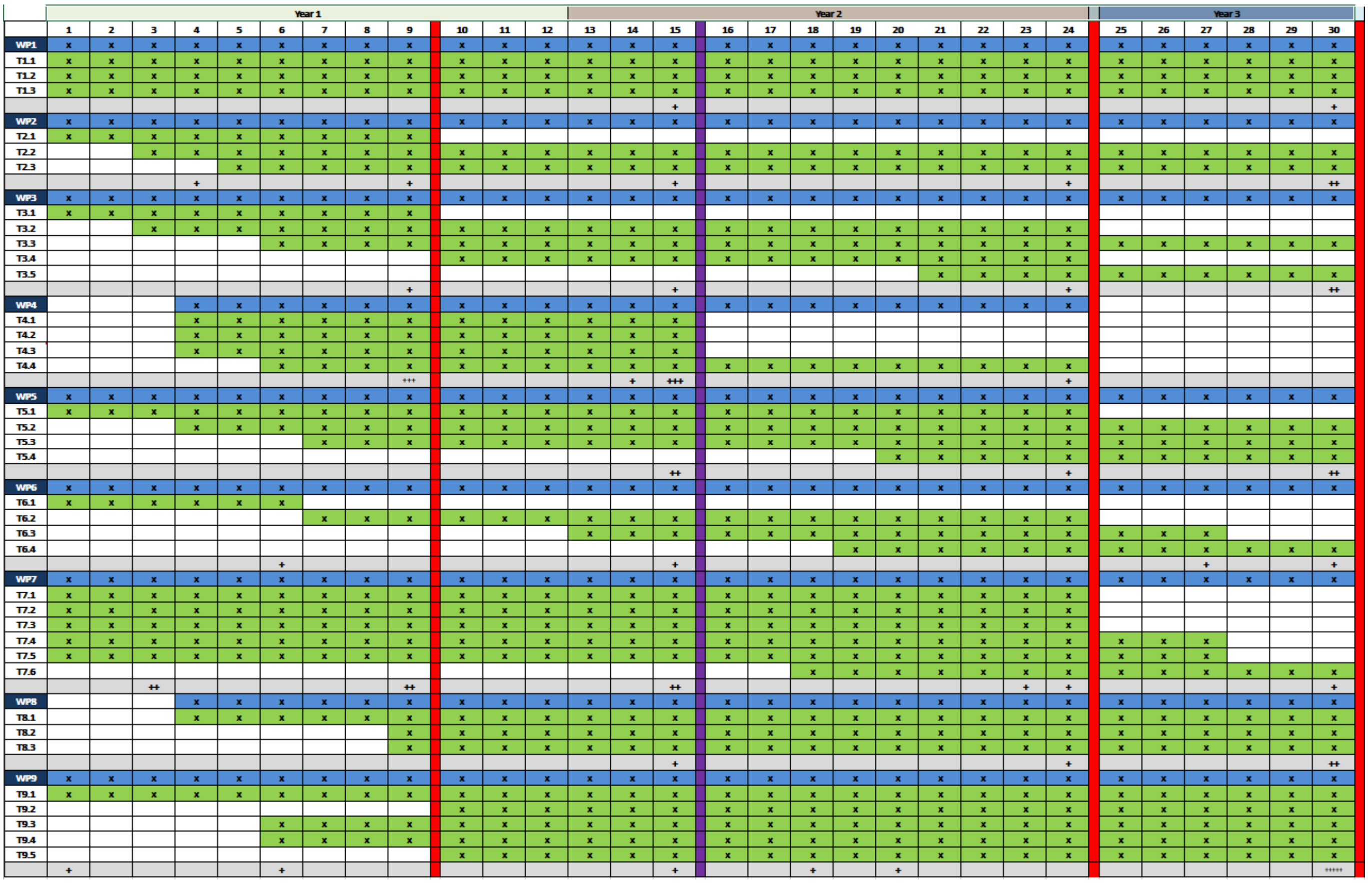

Figure 3.1.1b LEXIS Gantt Chart

Table 3.1a: List of work packages

WP No	Work Package Title	Lead Particip ant No	Lead Participant Short Name	Person-Months	Start Month	End month
WP1	Management	1	IT4I	65	M1	M30
WP2	LEXIS Requirements Definition and Architecture Design	2	Bull	175	M1	M30
WP3	LEXIS Data System	7	LRZ	233	M1	M30
WP4	Orchestration and Secure Cloud/ HPC Services Provisioning	4	ISMB	229	M4	M24
WP5	Aeronautics Large-scale Pilot	11	Avio Aero	116	M1	M30
WP6	Earthquake and Tsunami Large-scale Pilot	6	CEA	225	M1	M30
WP7	Weather and Climate Large-scale Pilot	10	CIMA	212	M1	M30
WP8	LEXIS Portal for Third Party Companies and SMEs	15	CYC	118	M4	M30
WP9	Impacts on Targeted Sectors	5	TESEO	129	M1	M30
			Total PMs	1502		

Work package number	1	Start Date or Starting Event					M1										
Work package title	**Management**																
Participant number	**1**	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Short name of participant	**IT4I**	Bull	NALL	ISMB	TESEO	CEA	LRZ	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	**18**	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Objectives

The main objectives of this work package are to:

- manage the project and coordinate the activities with special attention to the proposed use cases and to the foreseen pilot exploitation;
- ensure that the project reaches milestones and generates deliverables in a timely manner according to the plan;
- ensure that appropriate communication infrastructure and relevant tools are in place to facilitate the work and cooperation between the partners (shared workspace, software repositories, group chat mechanisms, mailing list, etc);
- ensure that the project outputs (documentation, prototypes, software) have the expected excellent quality by the appropriate review mechanisms;
- support efficient coordination among all the project WPs;
- track the consumption of the project resources throughout the project;
- run project meetings in an effective and efficient manner;
- organize the General Gathering (GG) meetings as necessary;
- provide timely communication with the European Commission.

The administration of the project including regular reporting, consortium meetings, and submission of deliverables is an integral part of this work package.

Role of participants:

IT4I as the leader of the project will coordinate this work package including all the tasks. All partners contribute to this work package.

Description of work

WP1 is responsible for the overall management of the project and ensuring successful delivery of the project. It consists of three tasks, which will run for the whole duration of the project. All partners will have modest contributions to the three tasks with most of the effort belonging to the project coordinator. The effort of the other partners is for their respective reporting and for contributing to the overall quality assurance process (reviewing deliverables, papers, and software).

Task 1.1 Project Management and Administration

Lead [**IT4I**], Contributors [**ALL**] – Period [**M1-M30**]

The management of the project and the definition of priorities and goals for the project is the responsibility of the coordinator that will ensure that the work assigned to the different work package teams will progress according to the schedule and within the allocated budget. The day to day operation and the implementation of the project goals will be done through a Grant Office (GO) headed by a designated experienced Project Manager located within the Coordinator and supported by the administrative and financial staff of the Coordinator led by the Administrative Manager. The progress will be monitored on a quarterly basis in conjunction with meetings or teleconferences of Project Technical Board (PTB). Exceptional conditions or problems will be handled by the GG including risk assessment and mitigation to ensure rapid resolution or corrective action, if necessary, together with the European Commission. Detailed description of the managing procedures and decision making are provided in **Section 3.2.1**.

Task 1.2 Quality Assurance

Lead [**IT4I**], Contributors [**ALL**] – Period [**M1-M30**]

Project results and formal deliverables will undergo a quality control process of internal project verification and document reviews prior to their timely submission to the European Commission. At least 2-3 internal reviewers will be assigned to review documents prior to the approval by the Project Management Board and submission to the European Commission. Software quality will also be assured by using project software management processes and tools as described in **Section 3.2.1**. Prior to each of the major software releases of the LEXIS platform, a systematic review of documentation and test coverage of all software modules will be conducted – a partner who has not been involved in the module development will review the work of the other partners highlighting where improvements can be made prior to release.

Task 1.3 Financial and Legal Management

Lead [**IT4I**], Contributors [**ALL**]– Period [**M1-M30**]

The financial and legal aspects of the project will be maintained within this task. The financial status reflecting actual versus planned effort and actual vs. planned expenditures will be maintained within the project on a quarterly basis internally and audited financial reports will be submitted to the Commission as required. GO will distribute the funds provided by the European Commission to the partners according to the Grant Agreement.

Management reports will be prepared and submitted to the European Commission. The coordinator will be responsible for the submission of the formal deliverables as defined by the contract. The coordinator will prepare and organize the annual project reviews.

Deliverables

D1.1 Project periodic report 1 (**R, CO**) (**M15**) [**IT4I**]

D1.2 Project periodic report 2 (**R, CO**) (**M30**) [**IT4I**]

Work package number	**2**	**Start Date or Starting Event**						**M1**									
Work package title	**LEXIS Requirements Definition and Architecture Design**																
Participant number	1	**2**	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Short name of participant	IT4I	**Bull**	NALL	ISMB	TESEO	CEA	LRZ	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	17	**17**	5	19	10	8	18	9	8	10	8	11	6	10	3	7	9

Objectives

The main objective of this WP is to define the LEXIS infrastructure according to the needs and technical requirements of the users and use cases, as determined in this and other WPs. First, an initial evaluation of the large-scale pilots' needs and the existing hardware useful to the LEXIS project will be done. One of the tasks will be the enhancement of key LEXIS infrastructure technologies to allow easy use and running HPC/ Cloud/ BD applications at lower costs. Finally, a harmonized development of the entire LEXIS infrastructure will be performed in close cooperation with all the relevant stakeholders through periodical workshops.

Consequently, the following objectives will be addressed:

1. Provide a comprehensive description of the infrastructural needs raised by our target pilots.
2. Provide a comprehensive view of potential technical solutions (HPC/ Cloud/ BD) matching the pilots needs.
3. Organize a global architecture and co-design activity, including all the stakeholders, to identify the most suited technical blocks and the needed enhancements.

4. Validate the LEXIS infrastructure as well as its benchmark.

Role of participants:

Bull is the coordinator of this work package; **ALL** will contribute to the LEXIS infrastructure specification and design while **Bull** will lead the data and infrastructure related user needs description, **ISMB** for the co-design of the LEXIS infrastructure. Finally, **LRZ** will lead the infrastructure set-up, roll-outs, validation and benchmarking of the LEXIS infrastructure together with **IT4I** and relevant partners.

Description of work

WP2 will provide the infrastructure requirements and co-design of pilot test-beds infrastructure for large-scale applications which need HPC/ Cloud/ BD technologies.

Task 2.1 Infrastructure Evaluation and Key Technology Identification for LEXIS

Lead [**Bull**], Contributors [**ALL**] – Period [**M1-M9**]

In this task, an initial evaluation of the existing hardware and data-management-system infrastructure at the HPC sites will be performed. The evaluation will focus on the following points:

- existing data management systems, data-transfer interfaces;
- existing Authorization and Authentication Infrastructures (AAI);
- existing HPC/ HPDA systems with the potential for use in LEXIS;
- resilience of existing hardware with respect to data loss due to hardware failure or power outage.

The evaluation of the infrastructure will be followed by the deployment of the initial LEXIS software solution on the existing infrastructure and the evaluation of its performance and bottlenecks. This will lead to the identification of key technologies, both software and hardware, that will improve the LEXIS user experience.

We expect that Burst Buffer hardware will be based on large SMP servers and rely heavily on memory size, performance and volatility. Thus, we will investigate how new memory technologies such as NVMe PCIe based storage or NVDIMMs would improve the performance of the LEXIS architecture.

The key software technology is the orchestrator with the following main features: a) supporting the whole application lifecycle, from deployment, scaling, monitoring, self-healing, self-scaling to application upgrade, b) working with the hybrid infrastructures (IaaS, HPC schedulers, CaaS).

This will be complemented with a survey among the LEXIS partners, mainly the ones providing the test-beds, clarifying the users' data management needs. The evaluation will focus on:

- typical data set volumes and patterns, typical data life-cycle patterns;
- typical throughput requirements;
- typical data-visualization requirements.

This work will lead to an identification of the survey techniques (e.g. questionnaires) that will be used to approach potential users to identify their data-management needs and how these can be addressed by current infrastructure of the HPC/ HPDA centres.

Task 2.2 Co-design and Lesson Learned

Lead [**ISMB**], Contributors [**ALL**] – Period [**M3-M30**]

The project will produce a set of technical specifications defining an overall LEXIS architecture, covering both pure technical and infrastructure aspects as well as data set characterisation and data management aspects. It is important that the diverse requirements of a range of users and stakeholders are brought together with the evaluation of existing solutions and other technical constraints. A co-design process will be followed to harmonise the inputs generated in the other tasks of this WP and in other relevant WPs. As a co-design process, this will involve many iterations with other tasks and work packages.

An intermediate report of the architecture will be delivered at M15 (**D2.3**), including lessons learned from the co-design process during the first half of the project. The final LEXIS architecture and the results of the assessment within the test-beds validations will be delivered at M30 (**D2.7**).

All partners will actively contribute to this task. Periodic workshops will be organised to bring together all the relevant stakeholders.

Task 2.3 Infrastructure Set-up, Roll-outs, Validation and Benchmarking

Lead [**LRZ**], Contributors [**IT4I, Bull, NALL, ISMB, TESEO, ECMWF, GFZ, SEC, BAY**] – Period [**M5-M30**]

Within LEXIS, specialized hardware is to be deployed as a Burst Buffer and Bridging device between HPC and Cloud-HPDA facilities at each participating compute centre. This hardware will be set up at **IT4I** and **LRZ** within a running HPC/ HPDA infrastructure, requiring substantial modifications within the HPC centres to interface the new hardware efficiently with the existing systems. M5-M9 will be devoted to preparing and planning the respective modifications including switch systems etc., high-bandwidth connectivity to existing storage systems, installation of storage-access software e.g. for GPFS, connection to the site-specific water/ air cooling infrastructure. After this, the system installation may proceed, with the plan to begin hardware installation by M10, and systems running in test mode by M12. The systems will then be integrated in site monitoring and maintenance efforts, and benchmarks will be run every 6 months to measure data-throughput rates to the HPC/ HPDA facilities at each computing centre. With stabilised operation expected from M15, connectivity and software optimisation efforts will help to increase efficiency and bring a measurable advantage from the usage of the additional systems with respect to the original site infrastructure. This effort will be rounded up by software (OS, standard libraries for simulations) updates and continuous maintenance of the specialised systems, including the mitigation of problems within the experimental infrastructure.

Deliverables

D2.1 Pilots needs/Infrastructure evaluation report (**R, PU**) (**M4**) [**Bull**]
D2.2 Key parts of LEXIS technology deployed on existing infrastructure and key technologies specification (**R, PU**) (**M9**) [**Bull**]
D2.3 Report of LEXIS technology deployment - intermediate co-design (**R, PU**) (**M15**) [**ISMB**]
D2.4 Report of LEXIS technology deployment - updated test-beds infrastructure (**R, PU**) (**M24**) [**LRZ**]
D2.5 Final assessment of the co-designed LEXIS architecture (**R, PU**) (**M30**) [**ISMB**]
D2.6 Infrastructure validation and assessment report (**R, PU**) (**M30**) [**LRZ**]

Work package number	**3**	**Start Date or Starting Event**							**M1**								
Work package title	**LEXIS Data System**																
Participant number	1	2	3	4	5	6	**7**	8	9	10	11	12	13	14	15	16	17
Short name of participant	IT4I	Bull	NALL	ISMB	TESEO	CEA	**LRZ**	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	38	39	0	31	7	13	**24**	20	1	11	4	17	0	21	2	5	0

Objectives

The LEXIS Data System, set up in this WP, is a data-access and data-exchange layer with unprecedented efficiency and advanced functionality. It will connect the participating HPC and HPDA sites together and provide the basis for a consistent data-management process in the LEXIS framework, as required for Big Data analytics and HPDA. Through an easy-to-use and secure REST interface, it will offer high-performance, resilient and monitored data staging functionality.

The detailed objectives of this WP are:

1. Objective: Develop the infrastructure for a unified data access layer (data storage).
2. Objective: Provide data management/staging/retrieval and task scheduling interfaces and data middleware/ backend components to enable HPDA.
3. Objective: Manage, monitor and optimize the data flow between data storage and HPC resources.

These objectives are reached by integrating with the European research-data management infrastructure EUDAT (www.eudat.eu, H2020 #654065), using their modules as a basis of the project-specific system. The co-design of components with the partners from **WP2** and **WP4** will then enable us to offer a tightly integrated and consistent solution. This LEXIS Data System will build on the infrastructure layer and work together with the high-level scheduling and resource allocation mechanisms of LEXIS, as well as the low-level schedulers and data-exchange mechanisms at individual HPC/ HPDA sites. Adapted interfacing to these sites will be necessary to match the different requirements in terms of hardware, software stack, access policy and local scheduling. In addition, management and monitoring systems will be developed and implemented, enabling HPDA workflows and continuous optimization towards optimum performance.

Role of participants:

LRZ will coordinate **WP3**. **Bull** leads the set-up of site-specific (hardware/ software) solutions for storage (**Task 3.1**). All HPC/ HPDA sites (**LRZ, IT4I, ECMWF**) contribute here as well as other partners with specific expertise (**ISMB, Avio Aero, SEC**). The central component integrating the local Storage, HPC and HPDA solutions, is the LEXIS Data System Core. It consists of data staging/retrieval and task-scheduling interfaces and the necessary middleware components, with significant development contributions by **IT4I, LRZ, Bull, ISMB, CEA and GFZ** (**Tasks 3.2/ 3.3**, leads: **IT4I** and **LRZ**). **CYC** leads the installation of an integrated monitoring system (**Task 3.4**), collecting performance and usage data for optimisation and billing, respectively. As considerable competence on system-specific monitoring solutions is needed in this effort, **ISMB** contributes substantially as well as **IT4I, Bull, CEA and LRZ.** Finally, **ECMWF** leads the efforts on data flow optimisation (**Task 3.5**). Here, experts from **CEA** as well as all HPC/ HPDA site partners (**IT4I, ECMWF, LRZ**) contribute. The rest of the partners will contribute on the appropriate tasks they are involved in to support main WP partners.

Description of work

In **Task 3.1,** existing storage and data-transfer solutions on each site are extended such that first use cases may be run locally on the sites. The resulting storage systems will be low-level building blocks of the LEXIS Data System. **Tasks 3.2** and **3.3** integrate all building blocks, developing and deploying a data-transfer and task-scheduling infrastructure. With **Task 3.4** we monitor the infrastructure in terms of usage, performance and security, which is the basis for performance optimization of data transfers (**Task 3.5**) as well as billing and security-patching activities within LEXIS.

Task 3.1 Set-up of Site Specific Basic Experimental Storage Solutions

Lead [**Bull**], Contributors [**IT4I, ISMB, LRZ, ECMWF, Avio Aero, SEC**] – Period [**M1-M9**]

Initially, the project partners will install storage solutions that match their existing infrastructure (hardware and/ or middleware for storage access) and have the potential to act as data storage and exchange solutions for the LEXIS Data System. The main goal of this task is to gain experience with local solutions, which then serve as building blocks and blueprints for the large-scale data infrastructure (**Task 3.2**). Close collaboration between the HPC site experts and the other infrastructure work packages of LEXIS will be ensured through frequent online and face-to-face meetings to ensure the development of optimal local solutions (high-throughput, easy-access).

Task 3.2 LEXIS Data System Core I: Storage Library, Distributed-Data Infrastructure (DDI) with Fundamental API Components

Lead [**IT4I**], Contributors [**Bull, ISMB, TESEO, CEA, LRZ, ECMWF, CIMA, GFZ, SEC, BAY**]– Period [**M3-M24**]

The LEXIS Data System unifies the storage solutions from **Task 3.1** (site storage libraries), combining them within a performant, imply-to-use and scalable service layer. In this task, the first part of the LEXIS Data System will be implemented, a common DDI. This will be capable of holding data and metadata, and it will expose itself as a common file system, integrated with the LEXIS AAI (**WP4 Task 4.3**). This is augmented by APIs (on HTTP REST or similar basis) for data search/ collection from various sources including the DDI (Data-Search API), and for data preprocessing (format conversion, data fusion). In the first three months (M3-M6), different possible low-level solutions for the DDI, such as EUDAT's B2SAFE service (or the underlying iRODS solution) are evaluated. During and after the installation phase (M6-M12) of the DDI, the APIs mentioned above are developed (M10-M24, possibly leveraging EUDAT's B2FIND).

Task 3.3 LEXIS Data System Core II: Interfaces for HPC/ HPDA with Focus on Data Staging/ Retrieval and Task Scheduling

Lead [**LRZ**], Contributors [**IT4I, Bull, ISMB, CEA, ECMWF, CIMA, Avio Aero, GFZ, SEC, BAY**]– Period [**M6-M30**]

The next-generation services to be developed in this task cover the high-performance staging of data to the HPC/ HPDA systems involved in LEXIS, and the retrieval of data from them (Data-Transfer REST API). LEXIS hard- and software solutions (Burst Buffer etc.) are integrated in this service. **Task 3.3** develops the service including the API (M6-M24) and integrates it with the task-scheduling/orchestration solutions in LEXIS (M9-M30), so that these are enabled to control and schedule the data flow for the use cases. Besides allowing for convenient triggering and control of HPC/ HPDA workflows, the system developed by **Task 3.3** will provide Monitoring interfaces (Monitoring API) necessary for **Task 3.4**.

Task 3.4 Layout and Installation of Monitoring and Data Collection System

Lead [**IT4I**], Contributors [**Bull, ISMB, TESEO, CEA, LRZ, ECMWF, CIMA, CYC**]– Period [**M10-M24**]

Closely related to the data system architecture itself, we will design a monitoring and data-collection system based on industry-standard workflow/ IT-infrastructure monitoring tools such as ICINGA. This system will monitor the availability of significant infrastructure components, their security status (open ports, suspicious activity), and the actual data flows/ code-execution times within the LEXIS Data System. Thus, it not only serves classical monitoring purposes, but allows for performance profiling/ optimization (**Task 3.5**) as well as reporting and billing (**WP8**).

Task 3.5 Data Flow Optimization

Lead [**ECMWF**], Contributors [**IT4I, Bull, ISMB, ITHACA, GFZ, SEC**] – Period [**M21-M30**]

Based on the profiling data continuously collected by the monitoring system (set up in **Task 3.4**), a multidisciplinary effort will be started to optimize data flows (data transfer rates, scheduling so that load imbalances are avoided). This includes operational research on data-flow optimization (the results of which will be directly fed back to orchestration/ schedulers in collaboration with **WP4**), and possibly also on optimizations in hardware and network management. The general performance goal for the LEXIS Data System is such that e.g. in the Tsunami use case (**WP6**), tasks can be (re-)scheduled in near real time, and all tasks have transparent, quasi-immediate access to incoming input data for the calculations.

Deliverables

D3.1 Local storage solutions (**DEM, CO**) (**M9**) [**Bull**]
D3.2 Mid-term infrastructure (**R, PU; OTHER, PU/CO**) (**M15**) [**LRZ**]
D3.3 Monitoring system (**OTHER, PU**) (**M24**) [**IT4I**]
D3.4 LEXIS Data System Core (**OTHER, PU/CO**) (**M30**) [**LRZ**]
D3.5 Data Flow optimisation (**R, PU**) (**M30**) [**ECMWF**]

Work package number	**4**	**Start Date or Starting Event**						**M4**									
Work package title	**Orchestration and Secure Cloud/ HPC Services Provisioning**																
Participant number	1	2	3	**4**	5	6	7	8	9	10	11	12	13	14	15	16	17
Short name of participant	IT4I	Bull	NALL	**ISMB**	TESEO	CEA	LRZ	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	32	44	2	**40**	3	7	22	12	3	10	7	14	0	24	0	9	0

Objectives

The main objectives of this WP are:

1. Analysis and definition of policies for data access, management and processing to be integrated into the orchestrator.
2. Analysis and definition of metrics, criteria and policies concerning the exposition of services to the end-user to drive the allocation of HPC processing resources.
3. Analysis and definition of security and infrastructural requirements needed to ensure fast processing of huge data sets through the innovative analytics workflows.
4. Integration of infrastructural resource management and allocation policies into an orchestration system, providing optimised workflow balancing capabilities.

Role of participants:

ISMB will coordinate the effort and activities carried out in **WP4**. It will also be the leader of activities carried out within the **Task 4.1**, thanks to its knowledge and expertise in data analytics techniques. All the partners providing HPC/ HPDA infrastructures (**IT4I**, **LRZ**, **ECMWF**) will be active on the tasks included in **WP4**, specifically on the activities related to **Task 4.1**. Their role as infrastructure providers will be of help in defining the most appropriate resources orchestration policies, along with infrastructural requirements and constraints. Contributors to this task will be **CEA**, **LRZ**, **IT4I**, **CIMA** and **Avio Aero**, **IT4I**, **ECMWF** and **LRZ** will also take part in **Task 4.2** along with **ISMB** and **Bull**. Here, the task leader **ISMB** will lead the effort in analysing and defining the set of policies related to the exposition of HPC-as-a-Services to the end-users. Similarly, in **Task 4.3**, **SEC** will lead the analysis and definition of requirements and constraints at infrastructural and application level which are related to security aspects. Contributors to such activities will also be **IT4I**, **LRZ**, **ECMWF**, **Bull** partners. Finally, **Bull** will be the leader of the activities carried out in **Task 4.4** and aimed at integrating orchestration policies and resource allocation algorithms into a working orchestration layer. **ISMB**, **CEA**, **SEC**, **CYC**, **IT4I**, **LRZ**, **ECMWF** partners will actively contribute to the integration of proper management policies taking into account infrastructural requirements and constraints. The rest of the partners will contribute on the appropriate tasks they are involved in to support main WP partners.

Description of work

The LEXIS architecture enables the effective operation and advanced functionality provided in the data and infrastructural layers, by means of the integration of an innovative orchestration system. The massive amount of data fed into HPC infrastructures and the deployment of analytics for processing them requires an adequate orchestration system, which is aimed at balancing the workload in such a way that all the running jobs experience a fast execution. Infrastructural requirements and constraints will be analysed and defined. Similarly, end-user requirements and constraints will be considered, also to manage the accounting and billing process. Requirements concerning the way analytic services and workflows are exposed to the end-user will be taken into account. Big Data and related HPDA represent a crucial aspect in LEXIS. Thus, data analytics requirements and the definition of analytic services composed into proper workflows will be studied. Finally, security aspects will be taken into account and related policies and constraints will be defined and integrated into the orchestrator. The outcome of such activities will be a fully integrated orchestration layer able to allocate infrastructural resources according to the data analytics application demands, infrastructure and end-user requirements, by means of the application of proper policies. To guarantee the effective operation of the orchestrator, workload balancing optimisation techniques will be put in place.

Task 4.1 Data Access Priority and Processing Policy for Big Data Analytics

Lead [**ISMB**], Contributors [**IT4I, Bull, TESEO, CEA, LRZ, ECMWF, ITHACA, CIMA, Avio Aero, GFZ, SEC, BAY**] – Period [**M4-M15**]

This task aims to investigate the data-related criteria and metrics (e.g., the volume, the location, etc.) used to drive orchestration for different components of the data layer. A multi-tenant processing environment is characterized by concurrent accesses to data, which need to be scheduled and regulated by appropriate and secure policies. Therefore, data access priorities will be implemented for different types of analytic tasks, with the aim of optimizing the processing schedule across different analytical tasks. On the other hand, data processing policies will be defined to optimize the processing of single tasks. For instance, according to the requirements of the analysis (e.g., real-time constraints), to the type of algorithm (data-intensive, compute-intensive) and to the available resources, data processing policies will determine where to deploy the task (based on data locality) and if data should be reallocated across nodes. This task will also cope with the orchestration and synchronization of data ingestion and fusion from different sources (see **Task 3.2**). For instance, for real time analytics it is necessary to define timings of data streams ingestion, buffering policies, sequencing of transformation operations, parallelization of different operations, etc. Data sharing among different sites and among different analytical services will follow the authorization agreements among infrastructure providers and the security requirements of **Task 4.3**.

Task 4.2 Orchestration Policies in Federated Cloud/ HPC Environments for BDaaS

Lead **[ISMB]**, Contributors [**IT4I, Bull, LRZ, ECMWF, SEC**] – Period [**M4-M15**]

This task aims to investigate the set of user-level criteria and metrics used to drive the allocation of resources in the federated HPC/ Cloud environment. Such criteria and metrics will be at the basis of definition of policies used to manage the accounting and billing process for the end-users, as well as the creation of the management interface exposed to the end-user. Objectives for this task are:

- analysis and definition of end-user level criteria and metrics for driving resource allocation;
- definition of requirements and constraints at the infrastructure layer concerning end-user interaction;
- analysis and definition of the accounting and billing mechanisms.

Task 4.3 Security Access, Authentication and Interoperability for Federated HPC Infrastructures

Lead [**SEC**], Contributors [**IT4I, Bull, ISMB, CEA, LRZ, ECMWF, ITHACA, CIMA, Avio Aero, GFZ**] – Period [**M4-M15**] This task aims to analyse and define all the constraints, solutions and aspects regarding security, authentication, interoperability, etc. Specific aspects investigated and analysed in this task are: ● secure access to data; ● authentication and authorization policies for data access by users and services; ● authentication and authorization policies for computing resources access by users and services. A larger number of site-specific adaptation modules (e.g. connectors to local authentication systems) will have to be implemented. With the end of M15, a unique set of credentials (e.g. a X.509 certificate valid on one participating site) per user will be sufficient for utilizing significant parts of LEXIS HPDA/ HPC/ BD infrastructure. A major focus in this unified-access layer must be on IT Security, for which a risk assessment and action plan is presented at M15. **Task 4.4 Integration of the Overall HPC/ Cloud Orchestration System** Lead [**Bull**], Contributors [**IT4I, NALL, ISMB, CEA, ECMWF, CIMA, GFZ, SEC, BAY**] – Period [**M6-M24**] This task is primarily concerned with the integration and implementation of the orchestration solution, by considering data, user and infrastructure constraints and requirements for the allocation and meta-scheduling. We also consider the analysis of a checkpointing solution, as well as solutions for data sharing and federation. In order to keep resource usage under control and avoid slowdowns of running processes, it is necessary to optimise the entire workload running on the underlying infrastructure. **Task 4.4** will also analyse and integrate optimization techniques for providing workload balancing within the federated Cloud/ HPC infrastructures. A checkpointing mechanism will also be studied and integrated to make the entire LEXIS architecture reliable and robust against failure and infrastructure outages. Specific objectives addressed are: ● definition of infrastructural requirements for each analytics service and workflow; ● definition of analytics services and analytics workflows (composed by multiple analytics services) and their deployment strategies; ● integration of **Task 4.1**, **Task 4.2** and **Task 4.3** policies, metrics and criteria into an orchestration layer for an optimal management of workflows.
Deliverables
D4.1 Analysis of mechanisms for securing federated infrastructures (**R, PU**) (**M9**) [**SEC**] D4.2 Design and implementation of the HPC-federated orchestration system – intermediate (**R, PU; OTHER, PU**) (**M14**) [**Bull**] D4.3 Definition of data access priority, analytics policies and security assessment (**R, PU**) (**M15**) [**ISMB**] D4.4 Definition of workload management policies in federated Cloud/ HPC environments (**R, PU**) (**M15**) [**CYC**] D4.5 Definition of mechanisms for securing federated infrastructures (**R, PU**) (**M15**) [**SEC**] D4.6 Design and implementation of the HPC-federated orchestration system – final (**R, PU; DEM, PU**) (**M24**) [**Bull**] D4.7 Centralized AAI: Coverage of all significant Systems (**R, PU**) (**M24**) [**SEC**]

Work package number	**5**	**Start Date or Starting Event**	**M1**													
Work package title	**Aeronautics Large-scale Pilot**															
Participant number	1	2	3	4	5	6	7	8	9	10	**11**	12	13	14	15	16	17
Short name of participant	IT4I	Bull	NALL	ISMB	TESEO	CEA	LRZ	ECMWF	ITHACA	CIMA	**AvioAero**	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	26	5	5	17	4	0	16	0	0	0	**43**	0	0	0	0	0	0

Objectives
The main objectives of this WP: The aim of this WP is to investigate the **industrial applicability** of HPC/ Cloud and Big Data platform development efforts, carried out in previous WPs, to sophisticated Computer-Aided Engineering (CAE) tools, whose goal is focused to examine complex flows behaviour in aeronautical engines' components. To accomplish assigned targets, **Avio Aero**, with the support of the academic partners and computing research centres, has identified some aeronautical use cases that will run on both state-of-art and enhanced HPC platforms aimed to investigate/ certificate how simulations running time is evolving thanks to new technologies applied in Advanced Computing Systems. A significant impact in terms of time reduction is expected: this is a fundamental step to improve **engineering productivity in** New Technology Introduction (NTI)/ New Product Introduction (NPI) Aeronautical processes. **Expected Outcome and Detailed Objectives** The expected outcomes of the WP are: ● critical comparison of simulation results employed in the design process of turbomachinery components, (pace and complexity) obtained with the new HPC infrastructure with respect to the current one, staying as a reference datum; ● assessment of new HPC solutions in terms of benefits, expected in terms of computational costs and time-to-market advantages, and also limitations, if any; ● final validation of innovative solutions (mainly computational running time) implemented in HPC system and developed in previous WPs. **Role of participants:** **Avio Aero** selecting and carrying out the test cases to be used as the representative of aeronautical needs. **IT4I** and **LRZ** contributing to define the hardware (HW) and software (SW) requisites of the HPC and Big Data platform; providing the hardware infrastructure of HPC and Big Data platform and the computing hours both needed to run the CAE simulations which the Turbomachinery and Rotating parts use cases are based on; contributing to finally summarize achievements, review Aeronautics use cases' KPIs and estimate expected impact on Aeronautical Industries and market. **Bull** contributing to define the hardware and software requisites of the HPC and Big Data platform; newly designing or improving the hardware infrastructure of HPC and Big Data platform; contributing to finally summarize achievements, review Aeronautics use cases' KPIs and estimate expected impact on Aeronautical Industries and market. **TESEO** contributing to define the hardware and software requisites of the HPC and Big Data platform. **ISMB** contributing to define the hardware and software requisites of the HPC and Big Data platform; contributing to finally summarize achievements, review Aeronautics use cases' KPIs and estimate expected impact on Aeronautical Industries and market. **NALL** contributing to define the hardware and software requisites of the HPC and Big Data platform; improving the HW acceleration capabilities provided by **Bull**. The rest of the partners will contribute on the appropriate tasks they are involved in to support main WP partners.
Description of work
WP5 will deliver evidence, based on Aeronautical use cases, of the achievements gathered in the present Research Project, as part of the final vision of getting real-time simulation ready at an industrial level. Since data exchange and management will be a relevant matter in this work package, the project team will jointly work to assure reliable service operations and secure data access in all conditions. **Task 5.1 HW/SW Integration Requirements** Lead [**Avio Aero**], Contributors [**IT4, Bull, NALL, ISMB, TESEO, LRZ**] – Period [**M1-M24**] This task aims to define the hardware and software requisites of the HPC/ Cloud and Big Data platform needed to run the CAE simulations which the Turbomachinery and Rotating parts use cases (refer to following **Task 5.2** and **Task 5.3**) rely on. While these two use cases will be moved forward during the **WP5** implementation, the needed hardware and software requirements will be identified and reviewed in this task, specifically: ● at the beginning, in order to determine how the available state-of-art technological platform meets the software prerequisites of aeronautics use cases' simulations, and how this platform and the used CAE tools may (separately or both) evolve and be integrated to speed up the running performances;

- then, during the development stages, the adopted CAE tools (and any of their own enhanced versions) will be checked again to understand how they meet the newly designed or improved hardware platform.

Task 5.2 Turbomachinery Use Case Set-up and Run

Lead [**NALL**], Contributors [**IT4I, ISMB, LRZ, Avio Aero**] – Period [**M4-M30**]

Avio Aero has gathered significant experience in the past decades in the turbomachinery design, development, testing and manufacturing. Among different product lines, unsteady fluid-dynamic simulations of an entire LP turbine (with multiple rows interacting each other) can represent one of the best-in-class use case that can be selectable. This use case needs computer-intensive time resources. The goal is to carry a detailed performance assessment of the analysed configuration.

The simulation tackles moderately low, transitional Reynolds numbers flow regimes and faces critical phenomena like flow transition from laminar to turbulent occurring over the air foils, separation effects at blades' trailing edge, producing downstream strong mixing and complex wake structures impacting on neighbouring rows.

To limit the computing running time as much as possible, it will be important to enable the scalability of these URANS simulations (Unsteady Reynolds Averaged Navier-Stokes) to very large-scale HPC (thousand cores) so as to achieve important contraction in execution time and, at the same time, improve quality of the analyses, thanks to a dedicated quick post-processing of massive amounts (TB) of simulation outputs.
Different milestones are envisaged to measure benefits and/ or limitations. The former milestone will take place at the beginning of the project using current state of the art IT systems. Tests will be replicated later to accompany, drive and measure ongoing HPC developments.
Engineering activity will include CAE solver set-up, installation and run. Finally, post-processing of output data will take place.

Task 5.3 Rotating Parts Use Case Set-up and Run

Lead [**ISMB**], Contributors [**IT4I, LRZ, Avio Aero**] – Period [**M7-M30**]

Beyond standard CFD simulated products, today new frontiers and challenges are coming to deal with complex flow fields realized in mechanical parts rotating in presence of air and oil. Today this kind of simulation is at the leading edge of numerical technology.

Among possible examples that could be chosen, Gear-boxes represent a clean-cut case of this need.
The need to design gearboxes capable to withstand high transmission efficiency, poses the challenge to predict and simulate the flow field operating inside with greater precision. Differently from automotive applications, Aeronautics products are usually cooled and lubricated by oil jets instead of splash lubrication. The combination of jet lubrication and high tangential speeds precludes the possibility to neglect interactions between the liquid and gaseous phase. This represents the main challenge from a numerical point of view.
Simulations of these phenomena require a large amount of computing time and need to revise design outcomes in a detailed way several times (Big Data produced) before finding a configuration fully fitting the operative requirements. An integration with advanced HW solutions is envisaged to support and speed up the analysis process.
Engineering activity will foresee solver set-up, installation and run in different steps, firstly to fix the reference datum and then investigate to support the HPC development program.

Task 5.4 Final Validation Assessment

Lead [**Avio Aero**], Contributors [**IT4I, Bull, ISMB, LRZ**] – Period [**M20-M30**]

A final validation assessment will take place to conclude the Aeronautical Package and check if the newly designed LEXIS platform meets the Aeronautics use cases requirements by using:

- fully-evolved simulation CAE solvers, which will be achieved thanks to the technical contribution of University of Florence for the turbomachinery use case;
- best-in class HPC/ Cloud and Big Data solutions identified as output in previous WPs.

Merging these two key contributors should allow to achieve the challenging target of running time reduction (-20%) needed for allowing a break-through in engineering design processes. The expected reduction in execution time and increased capability to boost the software scalability due to improved HPC and Cloud-based technologies will allow an increase of the number of simulations that can be carried out in a specific time frame and to investigate unexplored regions of the parametric design space. The task will finally summarize such achievements, review Aeronautics use cases' KPIs and sort out expected impact on Aeronautical Industries and market.

Deliverables

D5.1 Turbomachinery use case: analysis of results run on state-of-art HPC System (SW model configuration set-up, HW features, execution running time and post processing data amount characteristics) (**R, CO**) (**M15**) [**NALL**]
D5.2 Rotating parts use case: analysis of results run on state-of-art HPC System (SW model configuration set-up, HW features, execution running time and post processing data amount characteristics) (**R, CO**) (**M15**) [**ISMB**]

D5.3 HW/SW integration optimization supporting Aeronautical use cases (**R,PU**) (**M24**) [**Avio Aero**]
D5.4 Avio Aero use cases: critical review to highlight benefits and limits by operating on advanced HPC Solutions (**R, CO**) (**M30**) [**Avio Aero**]
D5.5 Avio Aero use cases: review of Aeronautics use cases' KPIs and expected impact on Aeronautical market (**R, PU**) (**M30**) [**Avio Aero**]

Work package number	**6**	**Start Date or Starting Event**											**M1**				
Work package title	**Earthquake and Tsunami Large-scale Pilot**																
Participant number	1	2	3	4	5	**6**	7	8	9	10	11	12	13	14	15	16	17
Short name of participant	IT4I	Bull	NALL	ISMB	TESEO	**CEA**	LRZ	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	17	12	2	8	4	**55**	16	0	32	6	0	25	32	10	0	6	0

Objectives

This work package is concerned with the establishment and successful completion of an earthquake and tsunami pilot test-bed demonstrating the integration of value-creating innovations in the project, with a core focus on event-triggered/ time constraint earthquake and tsunami processing with multiple data sources, coordinated processing workflows and outputs. As such, the work package will specify the pilot coordinated workflow in detail, one or more scenarios describing this workflow, then focus on software improvement and integration with the project innovations (the LEXIS architecture, data layer and orchestration, and real-time constraints) to be completed with pilot demonstrations, with results disseminated to the pilot prospective clients.

The main objectives of this WP are to:

1. Define and build the complete earthquake and tsunami workflow for the pilot.
2. Improve components of the pilot workflow, using knowledge and architecture improvements.
3. Integrate the pilot workflow on the test-bed(s).
4. Run, test and evaluate the pilot.

This work package presents a specific compute flows that must provide results by pre-specified deadlines to answer to emergencies and disaster response deadlines, both for warnings and operational deployment of response units. It has a strong level of integration, mixing pre-existing data sets, simulations under strict time constraints, on-line and off-line computations, and output options and data products that can be merged and reused by other work packages, as well as graphical output needs. A particular notion of that work package is the fact that it aims at a real-time workflow with codes that are too heavy for on-line processing (i.e. not amenable to making the computation fast enough to meet deadlines by using enough).

Role of participants:

In this work package, **CEA** will work on real-time orchestration and workflow arbitration, study and evolution of the codes of the use case, study of hardware acceleration possibilities (hybrid computation runtime). **AWI** will work on Tsunami simulation use case and code (TsunAWI), InaTEWS tsunami warning system, TsunAWI acceleration. **GFZ** will work on earthquake ground-shaking distributions for disaster scenarios, building vulnerability classifications and damage/ loss assessment for disaster response through its global and dynamic OpenBuildingMap (OBM). **ITHACA** will take care of emergency mapping products based on the outputs of the aforementioned Tsunami and damage/ loss assessment models. **LRZ** and **IT4I** will provide test-beds infrastructures and the access to the LEXIS Data System. **Bull, NALL** will provide support with the LEXIS architecture and HW as well as providing guidance for offloading activities such as code porting and kernels optimization. **TESEO** and **BAY** will be providing support with the testing and test reports. **ISMB** will provide orchestrator tool knowledge and help with integration of created tools into LEXIS technology. **SEC** will provide security solutions for the integration with the LEXIS technology.

Description of work

WP6 will build baselines from the use cases proposed by the partners (establish KPIs), transmit requirements on the baselines, build or integrate or configure solutions to accelerate the use case baselines, measure KPIs over the new implementations, and transfer to original use case owner (KPI as an adoption metric).

Task 6.1 Baseline Scenarios, Measurements and Requirements

Lead [**GFZ**], Contributors [**IT4I, Bull, NALL, ISMB, TESEO, ITHACA, CIMA, CEA, AWI, SEC, BAY**] – Period [**M1-M6**]

This task will define baseline scenarios, establish requirements (data, processing) and collect the current state/ measurements for those scenarios. It will also liaise with the test-beds to establish the target of the pilot and with third parties to integrate input data requirements. Earthquake rupture and shaking scenarios will be provided through OpenQuake as input for the tsunami modeling and assessment for earthquake- and tsunami-induced damage. Then a baseline scenario for the TsunAWI code will be defined alongside required data product. Then the building vulnerability-classification schema will be created and possibilities of integrating the earthquake and tsunami alerts into operational emergency mapping workflows, with the main goal to improve the impact of emergency mapping products in the emergency management domain, will be analysed. The result will be the description of the complete workflow of the pilot, from sensors, simulations and databases to output. In case a real major earthquake/ tsunami disaster occurs during the project phase, this event will be rebuilt as an additional scenario for result testing against ground truth.

Task 6.2 Pilot Development/ Evolution

Lead [**AWI**], Contributors [**IT4I, Bull, ISMB, CEA, LRZ, ITHACA, CIMA, GFZ**] – Period [**M7-M24**]

This task will focus on implementing the pilot components, both on the level of each component itself and on the level of the LEXIS architecture components. Intermediate releases of those components will be used by **Task 6.3**, with a first milestone at M15. Main focus will be work on the real-time requirements and task decomposition of the overall workflow that includes OpenQuake, TsunAWI, machine learning algorithms and EasyWave, as well as the OpenBuildingMap building damage estimation code, and integration of this workflow into LEXIS technology. Codes provided by the partners, such as TsunAWI and real-time building vulnerability classification algorithm, will be enhanced by integration of the new open data providers and optimizations for the LEXIS technology.

Task 6.3 Pilot Integration on Test-beds

Lead [**CEA**], Contributors [**IT4I, Bull, NALL, ISMB, LRZ, ITHACA, CIMA, GFZ, AWI, SEC**] – Period [**M13-M27**]

In this task, tools created in the **Task 6.2** will be integrated with the test-beds to be able to run the baseline scenarios. Results from other WPs in the infrastructure (data, tasks) will be used at that stage. The first step will be to establish proper access to the test-beds resources, before a first integration effort in line with **Task 6.2** results (M15). In an agile fashion, intermediate versions from **Task 6.2** will be integrated, and updated until the final releases of **Task 6.2** components (M24). By M24, the LEXIS architecture will also be available for the final integration of the pilot.

Task 6.4 Measurements on Test-beds

Lead [**ITHACA**], Contributors [**IT4I, Bull, ISMB, CEA, LRZ, CIMA, GFZ, AWI, SEC, BAY**] – Period [**M19-M30**]

In this task, results from the scenarios will be collected and analyzed; an overlap between **Task 6.3** and **Task 6.4** will allow for collections of results from intermediate versions of the integration, if any. The impact on the integrated emergency mapping operational workflow and real-time requirements will be evaluated. Additionally, test metrics and damage/ loss assessment evaluations for the scenarios will be provided.

Deliverables

D6.1 Baseline scenarios and requirements (**R, PU; Technical Diagram, PU**) (**M6**) [**GFZ**]
D6.2 Pilot improvements: solutions adopted (**R, PU**) (**M15**) [**AWI**]

D6.3 Pilot improvements: evaluation of software development (**R, PU**) (**M27**) [**CEA**]
D6.4 Pilot results: improved scenarios measurements and evaluation (**R, PU**) (**M30**) [**ITHACA**]

Work package number	**7**	**Start Date or Starting Event**							**M1**								
Work package title	**Weather and Climate Large-scale Pilot**																
Participant number	1	2	3	4	5	6	7	8	9	**10**	11	12	13	14	15	16	17
Short name of participant	IT4	Bull	NALL	ISMB	TESEO	CEA	LRZ	ECMWF	ITHACA	**CIMA**	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	5	9	0	7	24	0	10	31	32	**47**	0	0	0	10	3	6	28

Objectives

The Weather & Climate Use Case focuses on a complex system, to provide a diverse set of forecasts: weather, flood, fire, energy, air pollution. This system will exploit Copernicus and GEOSS data services, supplemented and complemented by global and local in-situ unstructured observations. Data will be filtered and pre-processed before being assimilated to ensure quality.
Moreover, each layer in the use case will produce its own large data set of forecast data, varying from hundreds of Terabytes for global weather models, to Megabytes at the decision maker level.
Both the assimilated data and data produced by each layer will be made available between layers, or for further analysis and industrial exploitation, through a cloud-based Weather and Climate Data API (WCDA).
The system will be applied to several test cases, spanning all the available forecasts, to fine tune the system and demonstrate the innovative value.

Role of participants:

In this work package, **CIMA** will work on regional weather & climate modelling, fire, flood, energy applications and on impact modelling. **ECMWF** will focus on global weather & climate modelling, data management and sharing. **ITHACA** will contribute on emergency mapping products and impact estimation. **NUM** will work on atmospheric dispersion of emissions, agricultural applications, and impact estimation. **TESEO** will develop a gateway for collection, filtering and transmission of in-situ observations. **LRZ** will support the test-bed with technical issues (data-transport, target HPC/ HPDA system integration). **IT4I** will provide support with test-bed integration and data layer exploitation. The rest of the partners will contribute on the appropriate tasks they are involved in to support main WP partners.

Description of work

WP7 will exploit Project achievements to deliver a system for prediction of water-food-energy nexus phenomena and their associated socio-economic impacts. These computations involve multiple model layers chained together:

1. Global weather and climate models.
2. Regional weather models.
3. Domain-specific application models (such as hydrological, drought and fire forecasts).
4. Impact models providing information for key decision and policy makers.

This WP will solve 2 main challenges in the deployment of such integrated systems:

1. Assimilation of diverse kinds of weather environmental observations; both "conventional" and new in-situ, unstructured observations coming from sensor-enabled devices in the Internet of Things (IoT). These latter observations will require new data access patterns to enable filtering and correlation algorithms.
2. Each model layer produces its own model output which is consumed by subsequent layers, as well as providing products directly to users. Depending on the layer, the data volumes range from Megabytes (from impact models for decision makers) to Petabytes (from global ensemble forecasts), and most of these models run in dedicated HPC facilities. This project will approach the challenge of serving model output data and products with Cloud and HPDA environments.

The integrated system will be validated by 6 test cases:

1. Enrichment of the Portfolio of Rapid Mapping analyses and increase of timeliness.
2. Quantitative assessment (affected people and economic losses) for forest fire and flood natural hazard.
3. Sustainable food production and protection of the environment.
4. Improved prediction of solar and wind energy production.
5. Sustainable activity of industrial sites with more efficient forecasted atmospheric impacts.
6. Urban air quality of life by improving quality and robustness of local air quality maps over cities.

Task 7.1 Develop Weather and Climate Data API

Lead [**ECMWF**], Contributors [**Bull, ISMB, TESEO, ITHACA, CIMA, SEC, CYC**] – Period [**M1-M24**]

In this task we will develop the Weather and Climate Data API (WCDA), which will provide a common access to structured and in-situ observations, as well as access to model output. This API will also be available as a mechanism for interfacing different layers in the applications stack. Data will be accessed using scientifically meaningful metadata, and provided according to various applicable standards such as GRIB, BUFR, ODB, WaterML 2.0, NETCDF-CF, SENSOR-ML.

This task will investigate the best technology to deliver high resolution model data output to Cloud/ HPDA systems for user defined post-processing. This task will make use of the LEXIS Data System technology provided by **WP3**.

Task 7.2 Global Weather and Climate: from Global in-situ Unstructured Observations to Forecast Products on the Cloud

Lead [**ECMWF**], Contributors [**ITHACA, CIMA**] – Period [**M1-M24**]

This task will develop tools and infrastructure to analyse, filter and pre-process the in-situ unstructured observations and make them available in a technically and scientifically meaningful manner to ECMWFs global NWP model (the IFS). A mechanism will also be developed to provide high-resolution NWP model output to Cloud environments, particularly Copernicus Climate Data Store and DIAS.

Task 7.3 Regional Weather and Climate: Assimilation of Local in-situ Unstructured Observations in High-Resolution Downscaling of Global Forecast

Lead [**CIMA**], Contributors [**Bull, TESEO, LRZ, ITHACA, NUM**] – Period [**M1-M24**]

Improvements to the predictive capability of regional weather models, executed on HPC facilities, will be explored by supplementing the assimilation of "conventional" observations (e.g balloons, remote sensing observations from Copernicus and GEOSS) with local in situ observations such as those labelled as IoT); e.g. smart cities, citizen scientists, low-energy and low-cost environmental monitoring systems, and smart gateways.

Multiple Terabytes of output data are produced by these models, and techniques and infrastructure for their efficient storage and sharing between HPC, Cloud and HPDA systems will be investigated.

Task 7.4 Cloud-Based Domain Specific Application Modelling, Forced by Regional Forecasts and Environmental Observations

Lead [**NUM**], Contributors [**Bull, TESEO, LRZ, ITHACA, CIMA, BAY**] – Period [**M1-M27**]

A wide range of application models are run on HPC and Cloud computing facilities. These include yield forecast, wind and solar energy prediction, hydrologic/ hydraulics models, forest fire models, urban air quality models, and impact models for industrial activities. A range of input data can be used to provide forcing to these models. This task will investigate the provision, assimilation and impact of data from regional forecasts, environmental observations (e.g. land surface temperature and soil moisture) and satellite-based data (e.g. from Copernicus and GEOSS).

Task 7.5 Cloud-Based Socio-Economic Impact Modelling Based on Exposure Information and Environmental Forecasts

Lead [**ITHACA**], Contributors [**Bull, LRZ, CIMA, NUM**] – Period [**M1-M27**]

This model will demonstrate the use of impact models to provide reliable quantitative assessments of socio-economic impacts by using exposure information supplemented by model output from environmental forecasts.

The Emergency Mapping portfolio will be enriched, and its timeliness improved, by exploiting the outputs of **Task 7.4** models and integrating the outputs of the impact models, for example by evaluating the impact of pro-active triggering of satellite/ aerial (manned/ unmanned) acquisitions covering forecasted affected areas and producing estimated impact maps to support the decision-making process. Furthermore, the thematic accuracy of the generated information will be improved.

Task 7.6 Full Test-beds Integration, from Global Models to Socio-Economic Impact

Lead [**CIMA**], Contributors [**IT4I, Bull, ISMB, TESEO, LRZ, ECMWF, ITHACA, SEC, NUM**] – Period [**M18-M30**]

Integrating a chain of diverse modelling steps on a range of HPC, Cloud and HPDA systems in an efficient and reliable manner is a significant challenge. This ranges through the full spectrum of data-related activities from quality control of unstructured in-situ observations, through sharing and transferring module output between layers in a timely and efficient manner, to making data available to users and for post-processing on Cloud and HPDA systems.

This task will involve the proper handling of model interoperability and data exchange across the entire model stack through interactions with the LEXIS Data System technology developed in **WP3**, the integration of a means to address and access data through scientifically meaningful metadata, and the use of applicable standards (such as GRIB, BUFR, ODB, WaterML 2.0, NETCDF-CF and SENSOR-ML).

Deliverables

D7.1 Architectural requirements and system design for interchange of weather & climate model output between HPC and Cloud environments (**R, PU**) (**M3**) [**ECMWF**]
D7.2 Architectural requirements and system design for interchange of in-situ unstructured weather & environmental observations (**R, PU) (M3**) [**CIMA**]
D7.3 Design of a smart gateway for collecting, pre-processing and transmitting in-situ observations (**R, CO**) (**M9**) [**TESEO**]
D7.4 Deployment of first test-bed infrastructure components in HPC/ Cloud (**R, PU**) (**M9**) [**NUM**]
D7.5 First release and test-bed deployment of Weather and Climate Data API for both in-situ unstructured observations and model data output (**R, PU**) (**M15**) [**ECMWF**]
D7.6 Deployment of test-bed infrastructure components and adoption of Weather and Climate Data API for model layer interoperability (**R, PU**) (**M15**) [**CIMA**]
D7.7 Field deployment of a smart gateway for collecting, pre-processing and transmitting in-situ observations (**DEM, PU**) (**M23**) [**TESEO**]
D7.8 Final deployment of test-bed infrastructure components with full interoperable model layers (**R, PU**) (**M24**) [**NUM**]
D7.9 Final report (KPI included) on demonstration and validation of the Weather & Climate test-bed applied to selected cases (**R, PU**) (**M30**) [**ITHACA**]

Work package number	**8**	**Start Date or Starting Event**							**M4**								
Work package title	**LEXIS Portal for Third Party Companies and SMEs**																
Participant number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	**15**	16	17
Short name of participant	IT4I	Bull	NALL	ISMB	TESEO	CEA	LRZ	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	**CYC**	BAY	NUM
PMs per participant:	24	14	0	6	6	5	17	6	0	0	0	0	0	7	**19**	14	0

Objectives

WP8 will provide a portal which will make use of the LEXIS infrastructure and LEXIS Data System to enable third party organizations such as companies and SMEs to obtain access to both the data sets available within the platform as well as to the distributed HPC capabilities of the platform. The system will provide functionality enabling third parties to:

- understand what data sets are available;

- understand the structure of the data sets and perform some basic exploration of the data;
- deploy and validate applications which use the data set(s);
- estimate the cost of deploying the application to the available federated infrastructure;
- deploy the application to the LEXIS infrastructure;
- monitor the state/ progress of the application;
- obtain access to the resulting output as appropriate;
- review accounting information relating to their usage of the platform.

WP8 will leverage the APIs provided by both **WP3** and **WP4** to support data access and application deployment respectively.

Role of participants:

In this work package, **CYC** will focus on the development of accounting and billing functionality. **IT4I** will work on design and implementation of the reporting system as well as assisting **CYC** with correct policy development from a provider perspective, **LRZ** will work on the data management frontend which will connect to LEXIS Distributed Data Infrastructure, **Bull** will help in the development of LEXIS portal from hardware perspective as well as contribute to Ystia orchestrator, **ISMB** will give technical support with respect to monitoring tools and the orchestrator, **SEC** will provide security advice ensuring the use of the correct communication protocols especially while handling third party user data, **TESEO**, **CEA,** and **ECMWF** will help in the handling of datasets from the data catalogue that will be created in **WP3**, **BAY** will provide feedback on the features from a point of view of a training and consultancy expert.

Description of work

WP8 is composed of three main tasks which together will facilitate the development of a comprehensive tool which can enable third parties to access the LEXIS platform. The three tasks will use a common development web framework to enable the tools developed by each task to be integrated to provide a single entry point and a uniform user experience while working with the LEXIS platform. **Task 8.1** focuses on enabling third parties to understand and work with the available data sets and allows to familiarize the users with the datasets structure and content. **Task 8.2** focuses on enabling third parties to upload applications to the platform, through container technology such as Docker or Singularity and link them to the available data sets so that applications can perform their work. This task will also enable the users to view the status of the work by providing easy access to status and log information. **Task 8.3** will provide accounting and billing supports, enabling users to estimate the cost of large, compute intensive workloads based on tariffing schemes of the resource owners as well as to see a final bill when the work has been completed.

WP8 will follow the overall project software development processes to support rapid feature development but will also have three specific releases throughout the lifetime of the project. These releases will each have proper documentation for the capabilities developed, be packaged so that other users can download and work with the functionality (e.g. in Docker containers with associated Docker-compose files) and have sufficient test coverage. Each of these releases will comprise of a packaged software prototype and a report describing the architecture of the system and how it can be used.

Task 8.1 Develop Functionality to Provide Access to Data Sets

Lead [**LRZ**], Contributors [**IT4I, Bull, ISMB, TESEO, CEA, ECMWF, SEC, CYC**] – Period [**M4-M30**]

This task will develop functionality to enable third parties to understand the data sets available within the LEXIS platform. This will involve integrating with the Data Layer to present the set of available data sets, including specifics for each data set such as terms of use, scope of data set, data structures and formats, means to use the data, available libraries for working with the data, links to documentation, accounting information related to use of data sets etc. The task will also provide an interface which will enable the third party user to perform basic actions on a limited subset of the data in order to gain understanding of how to work with the data set. The task will also provide a means to view and access result sets as necessary. In technical terms, **Task 8.1** develops an easy-to-use web GUI interface to the HTTP REST APIs (the Data-Search API in particular) of the LEXIS Data System, as well as to the APIs offered by the DDI of LEXIS.

Task 8.2 Develop Functionality to Support Deployment and Monitoring of Application

Lead [**IT4I**], Contributors [**Bull, TESEO, LRZ, SEC, BAY**] – Period [**M9-M30**]

Task 8.2 will focus on the complex problem of deploying given applications to the LEXIS platform and leverage the available data sets. Functionality will be developed to enable an application to be uploaded to the platform. Using the facilities/ APIs of the LEXIS DDI (**WP3 Task 3.2** and **Task 3.3**), data sets can be associated with applications so that when they are deployed the data set will be available to the application. Easy application deployment should be supported by the container technology such as Docker or Singularity. Additionally, users will have the possibility to test their application on reduced datasets or do quick online data preprocessing through the Cloud service and use HPC service for computationally heavy parts of the applications. Functionality developed within this task will enable the application to be deployed directly to the LEXIS platform for production: this will require integration with the orchestration capabilities developed in **WP4**. The task will also query the **WP4** workload management functions to show the state of the application.

Task 8.3 Develop Functionality to Provide Accounting and Billing Supports

Lead [**CYC]** Contributor [**IT4I, Bull, CEA, BAY**] – Period [**M9-M30**]

Accounting and billing supports will leverage the monitoring capabilities to be developed in **WP3** and **WP4**. More specifically, the accounting and billing functions will obtain the LEXIS platform usage information for all users of the system, apply the accounting and billing policies to determine the accounting implications for all users of the system. This task will enable users of the system to estimate the cost of their use of the system before deploying their application by providing some basic information relating to their anticipated use of the platform. The specifics will depend on the way the application is described: for applications comprised of VMs, basic calculations consisting of expected number and size of VMs and expected job duration can be used to estimate costs associated with compute activities; this can be combined with costs associated with accessing the data to obtain an estimate of the costs of the application. **Task 8.3** will also devise accounting rules for the diverse sets of accounting related metering information captured in **WP3** and **WP4** in conjunction with the envisaged potential future service offerings of the HPC centres which will be implemented in the flexible Cyclops Billing Rule Engine. The performance of the rule set operating on the rule engine will be assessed for ingestion of HPC scale usage data. The accounting information will be formatted as reports which reflect the costs incurred by the third parties working on the LEXIS platform. The rule set will deal with exception handling in cases such as the job being terminated prematurely.

Deliverables

D8.1 First Release of the Lexis Portal (**R, PU; DEM, PU**) (**M15**) [**LRZ**]
D8.2 Second Release of the Lexis Portal (**R, PU; DEM, PU**) (**M24**) [**IT4I**]
D8.3 Final Release of the Lexis Portal (**R, PU; DEM, PU**) (**M30**) [**LRZ**]
D8.4 Roadmap for further development of the LEXIS Portal (**R, PU**) (**M30**) [**CYC**]

Work package number	**9**	**Start Date or Starting Event**							**M1**								
Work package title	**Impacts on Targeted Sectors**																
Participant number	1	2	3	4	**5**	6	7	8	9	10	11	12	13	14	15	16	17
Short name of participant	IT4I	Bull	NALL	ISMB	**TESEO**	CEA	LRZ	ECMWF	ITHACA	CIMA	AvioAero	GFZ	AWI	SEC	CYC	BAY	NUM
PMs per participant:	14	10	3	4	**34**	5	4	5	5	5	4	4	2	3	6	19	2

Objectives

The main objectives of this WP are:

1. Definition of the strategy for achieving impact through dissemination, communication and exploitation, guiding all the outreach activities of the project and beyond.
2. Increasing Market share impact and adoption in Europe of HPC, Big Data and Cloud effective convergence and integrations finding new possible stakeholders.
3. Analysing the productivity and Business Process Improvement on targeted Large-scale Pilots.
4. Monitoring and report metrics/ criteria for measuring success of the outcome of the Project.

Role of participants:

ISMB will lead dissemination activities and ensure that the dissemination plan will be properly executed. **BAY** will lead exploitation and sustainability activities and will address potential interested parties. **ALL** partners will contribute to the dissemination and exploitation activities and work towards increasing the LEXIS project impact.

CYC will be responsible for the market analysis of the Big Data and HPC/ Cloud platforms and identification of the LEXIS technology position in this market. In **Task 9.4** contributors, led by **TESEO,** will execute Intellectual Property Rights (IPR) strategy and business process analyses. **Task 9.5** contributors will help **BAY** with the creation of the Open Calls and the industrial partners training.
Description of work
The overall objective of this work package is to maximise the project's impact, communication, and dissemination. The objective is to develop an effective impact strategy for covering all the aspects involved in the project and find the best way to exploit the project results. **Task 9.1 Execution of LEXIS Communication and Dissemination Plan** Lead [**ISMB**], Contributors [**ALL**] – Period [**M1-M30**] This task will ensure that the Communication and Dissemination plan described in **Section 2.2** will be executed. It will monitor communication and dissemination activity within the project, tracking each partner's plans for communication and dissemination as well as working with partners to identify joint communication and dissemination opportunities. Regular reports will be prepared for project management detailing communication and dissemination progresses. This task will also monitor the impact of the communication and dissemination activity in terms of social media attention, website traction, paper download, paper citations etc. Communication and Dissemination are on the same stage of Technological aspects because of the importance to spread out results, it is mandatory for creating interest in the project and engaging stakeholders. To achieve this goal it is primarily crucial to identify the right tools to be used and then the right Key Performance Indicators (from now KPI) and the related metrics to measure the success of what has been planned. **Task 9.2 Exploitation and Sustainability, HPC and Big Data Technologies Investments Stimulation, Private and Public partnership** Lead [**BAY**], Contributors [**ALL**] – Period [**M10-M30**] The objective of this task is to stimulate interest and investment opportunities from Public and Private bodies across Europe. As a full set of results is available and project technologies become sufficiently 'market-ready', **BAY** will pro-actively seek to identify and recruit interested parties as part of the ongoing sustainability effort – leading to a continuous use of the technology beyond the life of the project. Detailed Exploitation plans of partners are described in **Section 2.2**. **Task 9.3 Market Analysis of Converged HPC, Big Data and Cloud: LEXIS Positioning** Lead [**CYC**], Contributors [**IT4I, Bull, TESEO, ITHACA, CIMA, Avio Aero, BAY**]– Period [**M6-M30**] This task will review the major Big Data and Cloud platforms and how their markets are growing both during the project and after the project. This will include enterprise platforms that run on-premise as well as cloud hosted platforms. These offerings will be compared and contrasted with the HPC sector, highlighting scenarios which are more suited to classical on-premise or cloud solutions as well as those which are more suited to the HPC context. The analysis will then identify opportunities for a converged approach which can provide the benefits of each approach. The analysis will also consider the growth of ecosystems around specific large data sets and how converged solutions can relate to such ecosystems. This task will identify how the LEXIS platform can be positioned within the market context outlined and provide recommendations for enabling enterprises and SMEs in different sectors to leverage the LEXIS platform to obtain greater productivity: the initial focus will be on the sectors relating to the pilots, but consideration will also be given to other important sectors. **Task 9.4 Productivity and Business Process Improvement on targeted Large-scale Pilot, Replication, Innovation and IPR Management** Lead [**TESEO**], Contributors [**IT4I, CEA, ECMWF, ITHACA, CIMA, Avio Aero, GFZ**] – Period [**M6-M30**] The objective of this task led by **TESEO** is to execute the IPR strategy and demonstrate how using HPC and Big Data technologies can really improve the productivity and facilitate replication of Large-scale Pilots by integrating value chains in the industries´ Business Process. **Task 9.5 Open Call Stimulating Project Framework Adoption and Stakeholders Engagement on Targeted Large-scale Pilot** Lead [**BAY**], Contributors [**IT4, ISMB, TESEO, CEA, LRZ, ITHACA, CIMA, GFZ, SEC**] – Period [**M10-M30**] The objective of this task led by **BAY** is to create Open Calls to key industrial sectors (e.g. healthcare, manufacturing, energy). 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.

BAY will provide suitable training and close co-operative assistance to each of the participants involved in the Large-scale Pilots.
Deliverables
D9.1 Impact KPI and metrics achievement report and plan – first version (**R, PU**) (**M3**) [**TESEO**] D9.2 IPR and Data Management approach **(R, PU**) (**M6**) [**TESEO**] D9.3 Impact KPI and metrics achievement report and plan – intermediate version (**R, PU**) (**M15**) [**TESEO**] D9.4 Open Call Framework and Stakeholders engagement on targeted Large-scale Pilots – first report (**R, PU**) (**M15**) [**BAY**] D9.5 Market analysis of converged HPC, Big Data and Cloud Ecosystems in Europe (**R, PU**) (**M18**) [**CYC**] D9.6 Report on IPR Management (**R, PU**) (**M20**) [**TESEO**] D9.7 Impact on Productivity and Business Process Improvement in Aeronautics (**R, PU**) (**M30**) [**Avio Aero**] D9.8 Impact on Productivity and Business Process Improvement for Earth (**R, PU**) (**M30**) [**CEA**] D9.9 Impact on Productivity and Business Process Improvement for Weather & Climate (**R, PU**) (**M30**) [**CIMA**] D9.10 Impact KPI and metrics achievement report and plan – final version (**R, PU**) (**M30**) [**TESEO**] D9.11 Updated market analysis of converged HPC, Big Data and Cloud Ecosystems in Europe and positioning of the LEXIS Platform (**R, PU**) (**M30**) [**CYC**] D9.12 Open Call Framework and Stakeholders engagement on targeted Large-scale Pilots – final report (**R, PU**) (**M30**) [**BAY**]

The following table list the scheduled deliverables.

Table 3.1c: List of Deliverables:

Del.	Deliverable name	WP	Short name of lead participant	Type	Diss. level	Deliver y date
1.1	Project periodic report 1	1	IT4I	R	CO	M15
1.2	Project periodic report 2	1	IT4I	R	CO	M30
2.1	Pilots needs / Infrastructure Evaluation Report	2	Bull	R	PU	M4
2.2	Key parts of LEXIS technology deployed on existing infrastructure and key technologies specification	2	Bull	R	PU	M9
2.3	Report of LEXIS technology deployment - intermediate co-design	2	ISMB	R	PU	M15
2.4	Report of LEXIS technology deployment - updated test-beds infrastructure	2	LRZ	R	PU	M24
2.5	Final assessment of the co-designed LEXIS architecture	2	ISMB	R	PU	M30
2.6	Infrastructure validation and assessment report	2	LRZ	R	PU	M30
3.1	Local Storage Solutions	3	Bull	DEM	CO	M9
3.2	Mid-term Infrastructure	3	LRZ	R; OTHER	PU; PU/CO	M15
3.3	Monitoring System	3	IT4I	OTHER	PU	M24
3.4	Lexis Data System Core	3	LRZ	OTHER	PU/CO	M30
3.5	Data Flow Optimisation	3	ECMWF	R	PU	M30
4.1	Analysis of mechanisms for securing federated infrastructure	4	SEC	R	PU	M9
4.2	Design and implementation of the HPC-federated orchestration system – intermediate	4	Bull	R; OTHER	PU; PU	M14
4.3	Definition of data access priority, analytics policies and security assessment	4	ISMB	R	PU	M15
4.4	Definition of workload management policies in federated Cloud/ HPC environments	4	CYC	R	PU	M15
4.5	Definition of mechanisms for securing federated infrastructures	4	SEC	R	PU	M15
4.6	Design and implementation of the HPC-federated orchestration system – final	4	Bull	R; DEM	PU; PU	M24
4.7	Centralized AAI: Coverage of all significant Systems	4	SEC	R	PU	M24
5.1	Turbomachinery use case: analysis of results run on state-of-art HPC System (SW model configuration set-up, HW features, execution running time and post processing data amount characteristics)	5	NALL	R	CO	M15

5.2	Rotating parts use case: analysis of results run on state-of-art HPC System (SW model configuration set-up, HW features, execution running time and post processing data amount characteristics).	5	ISMB	R	CO	M15
5.3	HW/SW integration optimization supporting Aeronautical use cases	5	Avio Aero	R	PU	M24
5.4	Avio Aero use cases: critical review to highlight benefits and limits by operating on advanced HPC Solutions.	5	Avio Aero	R	CO	M30
5.5	Avio Aero use cases: review of Aeronautics use cases' KPIs and expected impact on Aeronautical market.	5	Avio Aero	R	PU	M30
6.1	Baseline scenarios and requirements	6	GFZ	R, Technical Diagram	PU; PU	M6
6.2	Pilot improvements: solutions adopted	6	AWI	R	PU	M15
6.3	Pilot improvements: evaluation of software development	6	CEA	R	PU	M27
6.4	Pilot results: improved scenarios measurements and evaluation	6	ITHACA	R	PU	M30
7.1	Architectural requirements and system design for interchange of weather & climate model output between HPC and Cloud environments	7	ECMWF	R	PU	M3
7.2	Architectural requirements and system design for interchange of in-situ unstructured weather & environmental observations	7	CIMA	R	PU	M3
7.3	Design of a smart gateway for collecting, pre-processing and transmitting in-situ observations	7	TESEO	R	CO	M9
7.4	Deployment of first test-bed infrastructure components in HPC/ Cloud	7	NUM	R	PU	M9
7.5	First release and test-bed deployment of Weather and Climate Data Interchange for both in-situ unstructured observations and model data output	7	ECMWF	R	PU	M15
7.6	Deployment of test-bed infrastructure components and adoption of Weather and Climate Data Interchange for model layer interoperability	7	CIMA	R	PU	M15
7.7	Field deployment of a smart gateway for collecting, pre-processing and transmitting in-situ observations	7	TESEO	DEM	PU	M23
7.8	Final deployment of test-bed infrastructure components with full interoperable model layers	7	NUM	R	PU	M24
7.9	Final report (KPI included) on demonstration and validation of the Weather & Climate test-bed applied to selected cases	7	ITHACA	R	PU	M30
8.1	First Release of the Lexis Portal	8	LRZ	DEM, R	PU; PU	M15
8.2	Second Release of the Lexis Portal	8	IT4I	DEM, R	PU; PU	M24
8.3	Final Release of the Lexis Portal	8	LRZ	DEM, R	PU; PU	M30
8.4	Roadmap for further development of the LEXIS Portal	8	CYC	R	PU	M30
9.1	Impact KPI and metrics achievement report and plan – first version	9	TESEO	R	PU	M3
9.2	IPR and Data Management approach	9	TESEO	R	PU	M6
9.3	Impact KPI and metrics achievement report and plan – intermediate version	9	TESEO	R	PU	M15
9.4	Open Call Framework and Stakeholders engagement on targeted Large-scale Pilots – first report	9	BAY	R	PU	M15
9.5	Market analysis of converged HPC, Big Data and Cloud Ecosystems in Europe	9	CYC	R	PU	M18
9.6	Report on IPR Management	9	TESEO	R	PU	M20
9.7	Impact on Productivity and Business Process Improvement in Aeronautics	9	Avio Aero	R	PU	M30
9.8	Impact on Productivity and Business Process Improvement for Earth	9	CEA	R	PU	M30
9.9	Impact on Productivity and Business Process Improvement for Weather & Climate	9	CIMA	R	PU	M30
9.10	Impact KPI and metrics achievement report and plan – final version	9	TESEO	R	PU	M30
9.11	Updated market analysis of converged HPC, Big Data and Cloud Ecosystems in Europe and positioning of the LEXIS Platform	9	CYC	R	PU	M30
9.12	Open Call Framework and Stakeholders engagement on targeted Large-scale Pilots – final	9	BAY	R	PU	M30

3.2 Management structure, milestones and procedures

3.1.1 Current status and motivation

The LEXIS project will implement a management organisation in order to ensure the achievement of the project goals on time, on budget and with high quality as well as to guarantee the effective collaboration of the 17 partners during 30 months. Consequently, the project must be flexible to be able to apply different solutions quickly and address difficulties when they arise. Given the size of the project and the flexibility required, we have chosen the following set of management structures and procedures.

The management structure of the project is designed to ensure that the project objectives are reached according to the work plan. This is achieved by assigning clear responsibilities to the management bodies of the project and by close collaboration. The management structure is described in **Figure 3.2.1a.**

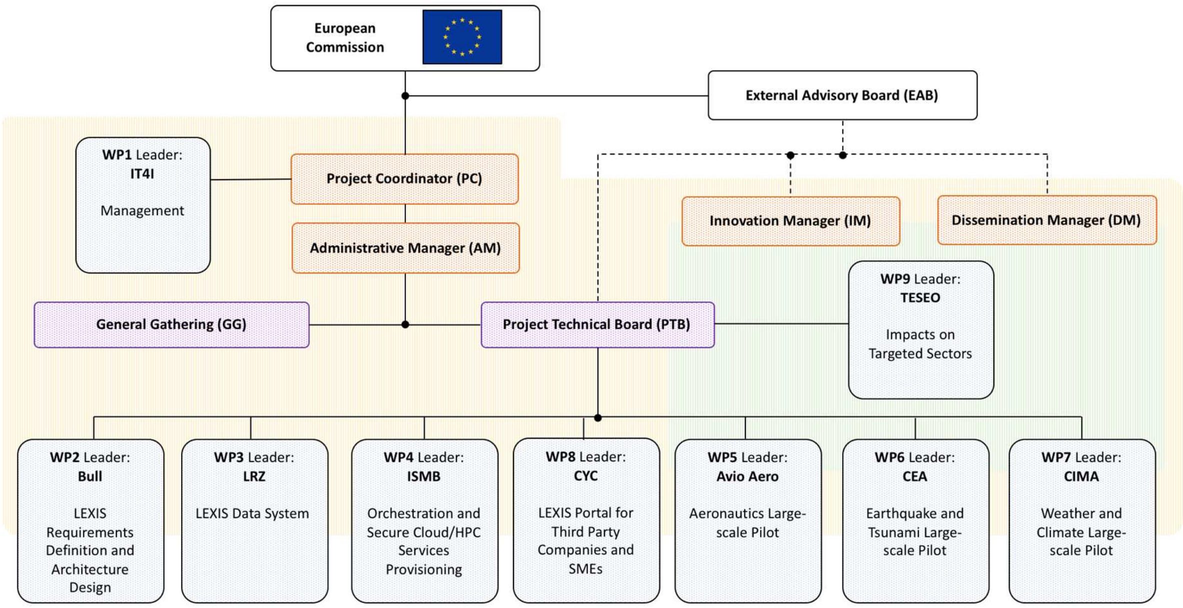

Figure 3.2.1a Organizational structure

The following bodies will be established to manage the project for inter-communication as well as for compliance to the funding agency, the European Commission.

The project coordination will be supported by the **Grant Office (GO)** at IT4I with respect to administrative, legal and financial aspects. As early as 2008, VSB-TUO established the Project Support Centre to support national, and later in 2010, also international project management and administration. GO as former part of the Project Support Centre supported the development of this proposal.

General Gathering

The **General Gathering (GG)** will be established as the decision-making and management body of the highest importance w.r.t to the project's direction that makes decisions on issues that transcend the work packages, such as evolution of the consortium, finances, changes to the project or intellectual property rights. This general gathering is represented by the responsible Principal Investigator from each partner institution and chaired by the Project Coordinator.

The GG shall convene twice a year (at least once face-to-face) or at any time when needed on request of one of the members. The GG shall prepare the meetings, propose decisions and prepare the agenda. The GG has the responsibility to:

- implement and execute the decisions;
- monitor the proper execution of the project;
- assess the risks of the project;
- initiate, coordinate and organize the work packages;
- support the coordinator in preparing the progress reports to the EC project officer and review meetings with him/her.

Project Technical Board

The **Project Technical Board (PTB)** consists of the work package leaders and co-leaders. It ensures the coherence of the work undertaken in this project. The PTB is responsible for production and documentation of the technical work in the work packages. The members ensure that deadlines are met and that dependencies are respected by the collaborators. They have the responsibility to identify potential problems or necessary changes in the work plan and alert the GG accordingly. The work package leader of **WP1** Management chairs the PTB.

WP-Leaders

The WP-Leaders are responsible for the overall management and progress of their work package. They will make decisions on the content and general direction of their work package, within the boundaries of the work plan. WP-Leaders have the authority to decide on the resolution of delays and set-backs within their work package but will do so in close cooperation with the other partners active in their work package, and the leaders of the related WPs. WP members will meet (or call) at least every quarter.

The WP-Leaders will execute the decisions of the GG, being responsible for a high quality preparation and execution of the project. WP-Leaders have the responsibility to:

- initiate, coordinate and organize their work packages, including the timely delivery of reports and WP-results;
- ensure revision of the quality of the outputs from the work packages by reviewing the outputs (deliverables) by at least 2-3 reviewers;
- collect information on the progress of the project monthly, examine that information to assess the compliance of the project to the work plan;
- in case of major deviations from the plan that have an impact on the work packages and/or main objective of the project (e.g. failure of a milestone, >20% budget change), inform the GG, and propose possible fall-back scenarios;
- formulate an implementation plan for the activities within the WP for the next period;
- support the coordinator in preparing related data and deliverables;
- alert the coordinator in case of any delay in the performance of the work package.

Project Coordinator

The **Project Coordinator (PC)** will facilitate/chair the GG in its decision making process. The PC can also advise WP leaders on their day-to-day decisions, and advise the General Gathering on the structures and procedures within EU-projects. PC will be the intermediary between the partners and the European Commission and will perform all tasks assigned in the Grant Agreement and the Consortium Agreement, including:

- monitoring the compliance of the project partners to their obligations - ensuring that the project technical objectives are met and that technical decisions are taken in time;
- collecting, reviewing and submitting reports and other deliverables (including financial statements and related certification) to the EC;
- monitoring of the technical quality of the WP output (e.g. deliverables and internal reports), and ensuring the consistency of the output;
- preparing the agenda of GG meetings, chairing the meetings when relevant, writing the minutes of these meetings and monitoring the implementation of decisions taken at these meetings;
- ensuring effective internal project communication;
- administering the EC financial contribution and other obligations forthcoming of the Grant or the Consortium Agreements;
- monitoring the progress on a regular basis;
- organising regular (telephone) conferences with the WP-leaders and representatives of the other partners and with EAB;
- managing risks and avoiding delays.

Project Coordinator is ***Jan Martinovič,*** **Ph.D.** (IT4I), the Head of the Advanced Data Analysis and Simulations Lab at IT4I. He has extensive experience leading substantial R&D activities, being the head of a lab comprising of 32 FTE staff. He has also coordinated various contracted research activities in cooperation with international and national partners such as Microsoft Corporation USA or ArcelorMittal Frýdek-Místek a.s. Czech Republic as well as in coordination of technical issues in several national projects. He is currently a leading partner of the two H2020-FETHPC-2014 projects ANTAREX and ExCAPE at IT4I.

PC will be supported by the **Administrative Manager**, who will be responsible for administrative and financial tasks arising in connection with the project:

- Timely collection and preparation of financial reports, including collection of financial statements from the beneficiaries for transmission to the EC;
- Financial management of the project, such as:
    - Supervision of the project budget and preparation of the distribution of payments,
    - Providing administrative support for the project participants,
    - Checking of the individual financial statements,
    - Preparation of summary financial reports,
    - Obtaining audit certificates.

Administrative Manager is ***Anna Němcová*** (IT4I). The coordinator will be supported by a lawyer who will participate in the preparation and maintaining of the Consortium Agreement as well as any legal and contractual matters.

Innovation Manager

The **Innovation Manager (IM)** will assist the PC in all issues related to the innovation, providing the project with periodical innovation assessments to be used to secure the maintenance and expansion of the innovation-based project's competitive advantage. IM will take the lead in the development of a general exploitation strategy, including the related business cases. This plan will be started once the definition of the expected results is set, and will be updated and completed during the exploitation meetings. All partners will participate in the exploitation meetings where a methodology to reach consensus on exploitation issues will be defined and an exploitation plan will be drawn up to ensure that all results are covered and internal agreement is given on the selection of exploitable results and ways to achieve them. A strong policy and agreement on IPR will be implemented during the project in order to be agile in further exploitation of the project's results. The Innovation Manager is ***Florin Ionut Apopei*** (TESEO).

Dissemination Manager

The **Dissemination Manager (DM)** will undertake the responsibility to outline and execute a communication plan starting from the beginning of the project activities, to control its execution along the duration of the project. He will work in close contact with the IM and with the PC. The Dissemination Manager is ***Olivier Terzo*** (ISMB).

External Advisory Board

The **External Advisory Board (EAB)** consists of representatives of external research institutions or industry not involved in the project Consortium, serving as a gateway between the project and potential markets as well as other stakeholders like industrial partners, end users and regulatory bodies. The role of EAB is twofold: (1) to give an informed opinion on the work of the project from the perspective of a thought leader in their domain and in particular to highlight any weaknesses or deficiencies in the work which could reduce potential impact in their domain(s) and (2) to help promote the project to their constituencies as appropriate. The EAB will discuss the progress and results of the project with other members of the EAB and/or with Partners. Then the EAB will give strategic advice to ensure that the project pursues its objectives in an adequate manner, addresses the needs of relevant European industries as well as reacts on recent trends in this domain and supports the user uptake of the project results.

The annex of this proposal document contains the commitment letters of the stakeholders willing to support our project as members of EAB:

- Fatos Xhafa, professor at Department of Computer Science, Polytechnic University of Catalonia, Spain (BarcelonaTech). His current research interest include parallel and distributed algorithms, combinatorial optimization, approximation and meta-heuristics, networking systems, distributed programming, Grid and P2P computing. His research is supported by Research Projects from Spain, EU and NSF/USA.
- Leonard Barolli Ph.D., professor at the Department of Information and Communication Engineering of Faculty of Information Engineering at Fukuoka Institute of Technology (FIT), Japan. Dr. Barolli was an Editor of Information Processing Society of Japan (IPSJ), he is engaged as a Program Committee (PC) Member in many International Conferences. Dr. Barolli has published more than 200 papers in Referred Journals, Main field of research are Quality of Service (QoS), traffic control mechanisms (policing, routing, congestion control, Connection Admission Control (CAC)), intelligent algorithms (fuzzy theory, genetic algorithms, neural networks), network protocols, agent-based systems, grid and Internet computing.
- Giovanni Sembenini, board member of The Aerospace Testing Seminar, and member of the Scientific Advisor Committee of the von Karman Institute for Fluid Dynamics. He has an extensive R&D projects experience being "Head of Section for NATO and non-European countries Research & Technology programs" in the Italian Ministry of Defense Secretariat General of Defense; he also is the NATO Science and Technology Organisation Systems, Concepts and Integration (SCI) Panel Principal Member as well as Italian NATO Science and Technological Organisation National Coordinator.
- Alan Sill is Senior Director of the High Performance Computing Center and Adjunct Professor of Physics at Texas Tech University. He also holds positions as Co-Director for the multi-university US National Science Foundation Cloud and Autonomic Computing Industry/University Cooperative Research Center and Visiting Professor of Distributed Computing at University of Derby, UK. He is an author on nearly 500 peer-reviewed publications covering topics from particle and nuclear physics to scientific computation to cloud and grid computing.

3.2.1 Managing procedures and decision making

Decision making

The consortium partners have the intention of working together in good harmony, and take decisions with the consent of all partners, after a discussion of the facts and possible paths to follow.

The GG takes all major decisions regarding the project, such as:

- Major changes in the project implementation plan;
- Major changes to the consortium budget (> 20% of original plan in a category);
- Acceptance/suspension of partners;
- Review and/or amendment of the Consortium Agreement;
- Selection of WP-Leaders;
- Taking decisions about intellectual property rights.

Each partner in the project will have one vote in GG. Other voting procedures will be further detailed in the Consortium Agreement. Decisions in GG are in general taken with a two-third majority, except where the Consortium Agreement explicitly states otherwise. All decisions will be taken either at regular or extraordinary meetings or through email and/or teleconference voting.

Communication flow

A project intranet or a similar shared storage will be set up for internal use to form a repository of documents and other information that is relevant to the implementation of the project, such as minutes of meeting , including financial and technical progress reports, a list of decisions taken and a to-do list and updated contact list, that will be made available to all partners.

Furthermore, informal communication will be carried out when needed by use of common internet communication tools, including Skype, e-Mail, shared documents and others. Regular online meetings will be scheduled on a monthly/quarterly basis to discuss the progress of the project, including deliverables and dissemination activities. Additional face to face meetings may be called by the WP-Leaders or upon request by the minimum of a third of the partners.

A public website for external communication will be hosted by the consortium, which will disseminate the project progress and will contain all public domain documentation.

Conflict resolution

The conflict resolution process will be based on the general principle that decisions are made at the lowest of competence in all areas of the project activities. For all needs and for decisions arising in the project, a consensus decision will be sought. When conflicts arise during the execution of the project, they will be resolved according to the following simple process: (1) They will first be addressed within the relevant WP through discussion chaired by the WP-Leader. (2) If this fails, the issue will be presented by the WP-Leader to the TB depending on the nature of the problem. (3) If an agreement is not reached it will be presented to the GG. Each member (Consortium Partner) will have a single vote. The matter will be resolved with a two-third majority. (4) Any conflicts that cannot be resolved through the principles above will be handled according to the CA rules.

Reporting Process

The reporting process will be divided in two categories:

Internal (within Consortium) – every partner will deliver a Monthly Management Report (MMR) to the PC. The MMR will contain a description of the technical and management project work done, listing effective time spent on the project. The MMR will included difficulties, milestones and deliverables (or contributions to deliverables in case of joint deliverables) that have been reached, patents, publications, travel and visits. The PC will be able to track both overall budget burn rates, as well as identify potential problems with individual organizations as a result of this report.

External (to the European Commission) – The PC will coordinate and consolidate quarterly, annual and final activities, and management and financial reports which need to be submitted to the EC. When needed, audit certificates (according to H2020 rules) will be provided by each partner. These certificates will be prepared by a external auditor selected by the partners, and certifying that the costs incurred during the period meet the conditions required by the grant agreement.

Consortium Agreement

The consortium members will conclude a Consortium Agreement (CA) which will be signed before the project starts (before the signature of Grant Agreement). The CA will include the internal management guidelines, consortium organization, fund distribution, management of background and foreground including IPR management during and after the project, and transfer of the LEXIS project results. CA and knowledge management is described in **Section 2.2 b)** in more details.

Quality Management

The LEXIS project will ensure a high level of quality of its output - both documents and other deliverables. Peer review will be an integral part of the Quality Assurance (QA) process: all deliverable reports will undergo peer review by members of the project team that are not specifically involved in the deliverable, with feedback being used to improve the overall report quality. WP leaders will also be responsible for the review of all deliverables produced by their WPs. Standard document quality review processes specifying timelines and specific mechanisms to obtain feedback will be formalized by the Project Management Team at the start of the project and communicated to the project team at the project kick-off meeting and stored on the project internal wiki.

The software will also be an important output of the project and processes will need to be implemented to ensure good quality software outputs. The project team differentiates between three types of software:

1. Proprietary software of commercial interest to one of the companies participating in the project which may be extended/enhanced during the project.
2. Software developed jointly throughout the course of the project which has strategic relevance for one or more of the project members.
3. Open Source software developed by the project team.

For the first, the commercial organization in question will maintain ownership of the code base and will be responsible for ensuring that the software is of good quality according to internal processes within the organization. For the second, the software will be developed jointly and hence project level tools and processes are required; its ownership will be covered by the CA and the software will not be published unless agreed by the partners involved. For the third, software will be developed in the open with project members committing to a public repository hosted by the project co-ordinator.

For the collaboratively developed software (2 and 3 above), standard best practices for distributed, collaborative development using best of breed tooling will be used with the project co-ordinator assuming responsibility for hosting the project tooling. Monthly sprints will be performed starting from month 4, after the system architecture has been developed. A hosted gitlab deployment will be used for the management of all software; it will be regularly backed up by the IT4I team. The Gitlab Issues[16] functionality will be used to manage tasks within a sprint. Further, having the work hosted on Gitlab in this manner will provide the project co-ordinator with a straightforward view of the activity of the project. Gitlab supports Continuous Integration processes which will be used to ensure that code commits do not break the existing functionality: this will be used from month 6 when basic integration of some of the project components will be possible. Testing will be considered from the outset employing standard test harnesses for unit tests - e.g. JUnit for java based testing, tox/unittest for python testing and go test for go based work. Test coverage will also be assessed using, e.g. Jacoco for java based work. For the open source software, releases will be periodically published to github on an automatic basis.

The LEXIS project is also aiming at a high level of quality of its documents and other deliverables to fulfil the goals and milestones of the project, e.g. by partner peer-review and feedback by the EAB. Accordingly, the project will implement a formal review procedure to ensure continuous quality of its documents and other deliverables, e.g. by having draft deliverables available a few weeks before the deadline for the internal review by consortium and EAB. WP leaders will be responsible for the revision of deliverables (with the possibility of delegation) before their submission to the EC. At least 2-3 reviewers will revise the quality of the documents. The approval process can be specified in detail in the Consortium Agreement. The following list of milestones (**Table 3.2.1a**) has been identified for easier tracking of the progress of the project towards the expected results.

Meetings	Period	Country	Main items to be discussed
Kick-off meeting (Face to Face)	M1	CZ	Overall objectives and Administrative/ Financial procedures
Progress meeting 1 (audio/video conference)			Progress Meeting + WP meeting
Progress meeting 2 (Face to Face)	M8	IT	Progress Meeting + WP meeting
Progress meeting 3 (audio/video conference)			Progress Meeting + WP meeting
Progress meeting 4 (Face to Face)	M15	FR	Progress Meeting + WP meeting
Progress meeting (audio/video conference)			Progress Meeting + WP meeting
Progress meeting (Face to Face)	M23	DE	Progress Meeting + WP meeting
Progress meeting (audio/video conference)			Progress Meeting + WP meeting
Final meeting (Face to Face)	M31	CZ	Final review

[16] https://about.gitlab.com/features/issueboard/

Table 3.2.1a: List of milestones

Milestone number	Milestone name	Related work package(s)	Estimated date	Means of verification
1	Awareness of project objectives and data management plan	WP9	M6	This milestone is reached with data management and communication plans for the entire project incl. demonstrators being available. Furthermore, a project website was created and media channels are set up.
2	Alpha version of LEXIS technologies	WP2, WP3, WP4, WP8	M9	Key technologies specification for LEXIS based on the evaluation of the needs. Local storage solutions for local test runs of Pilot simulations have been set up. The Distributed Data Infrastructure design is clear and its implementation is at 50% overall progress.
3	Awareness of intermediate project foreground	WP1, WP9	M15	With this milestone, the first/interim report, dissemination and exploitation plans are available.
4	Beta version of LEXIS technologies	WP2, WP3, WP4, WP8	M15	An intermediate report of the architecture will be delivered. The LEXIS Distributed Data infrastructure is implemented and can be used. The first versions of most LEXIS Data System APIs are available and begin to be utilised by Pilots. In particular, first versions of the APIs interacting with orchestration/task-scheduling are available; integration of the LEXIS data system with these components has begun.
5	The first pilots to test-beds integration	WP5, WP6, WP7	M15	Execution of the first release of Weather&Climate pilot on the test-beds, with preliminary model integration Integration of the initial release of the earthquake and tsunami pilot software and workflow on the test-beds has started on the LEXIS technologies that are already deployed (Milestones 2 and 4). The integration between software and hardware requisites both needed to run on LEXIS infrastructure the Aeronautical Turbomachinery and Rotating simulations' pilots has already started and is continuing. The first version of aeronautics large-scale pilots will be available and the related computer-aided engineering simulations begin to be run on LEXIS HPC/cloud infrastructures, so to compute and collect the first simulation data to be handled.
6	Final version of LEXIS technologies	WP2, WP3, WP4, WP8	M24	Test-beds infrastructure is updated. The Lexis Data System is completely developed (Distributed Data Infrastructure & APIs) and utilised by Pilots. Also, the infrastructure monitoring/ data collection system is installed
7	Awareness of final project foreground	WP1, WP9	M30	At this milestone, the final report, dissemination and exploitation plans are available.

8	Final integration and LEXIS technologies validation	WP2, WP3, WP4, WP8	M30	The co-designed LEXIS architecture is assessed and validated. The LEXIS data system is fully integrated with the task-scheduling/orchestration solution. The system has been benchmarked and bottlenecks have been removed/mitigated
9	Finalised and validated Pilots	WP5, WP6, WP7, WP9	M30	Execution of the final release of the fully integrated Weather&Climate pilot, based on WCDA, and verification against KPIs Final results of the earthquake and tsunami pilot are collected, evaluated and interpreted on the LEXIS technologies of Milestone 8. A final validation assessment will take place to check if the newly designed LEXIS technologies meet the aeronautics pilots requirements. Specifically, the aeronautics pilots will be validated based on the impact that will be measured in terms of reduced execution time for simulation data computing and post-processing.

3.2.2 Risk Management

The LEXIS project has defined clear procedures for the management of all risks and issues which are identified at the beginning of the project and arise over its course. All project participants and in particular the Project Coordinator and the WP-Leaders will be responsible for raising any material or perceived risk as part of the normal reporting. The agile methodology defined for the LEXIS planning will be an important approach for the risk and losses mitigation during the execution of the project. For many risks in the project the lessons learned can be used to cope correctly with the possible errors. In **Table 3.2.2a** the levels of probability are defined as follows: Impossible or remote under normal conditions, low - under normal conditions, medium - 50/50 chance, high - greater than 50% chance. The expertise of the partners in their own field has been shared in the consortium to evaluate the major risks and avoid them. The GG will ensure that all necessary actions will be undertaken to minimise risks and to drive all necessary counter measures. The major risks identified in the LEXIS project are described in **Table 3.2.2a**.

Table 3.2.2a: Critical risks for implementation

WPs involved	Description of risk	Probability	Proposed risk-mitigation measures
WP7, WP6	**Data quality and accuracy will be low**	Low	Existing methodologies for evaluating data quality and accuracy will be used in the project. These methods are able to detect all kinds of errors or inaccuracies and consequently fix them.
WP2, WP5, WP6, WP7	**HPC infrastructure has a downtime or is inaccessible**	Low	IT4I's and LRZ`s HPC facilities have redundant electrical sources and network connections. In the case of downtime or critical failure, the LEXIS framework is able to relocate submitted simulations to other supported infrastructure.
WP2, WP3, WP5, WP6, WP7	**Data storage infrastructure has a downtime or is inaccessible**	Low	Data storage is synchronised on multiple geographical sites, with redundant power and network connections. Backup data storage infrastructure can be operated and synchronized with the main infrastructure.
WP5, WP6, WP7	**Supercomputer resources are insufficient**	Low	If supercomputer resources are insufficient, we will ask for further resources through PRACE.
WP1-WP9	**Key member of staff or partner will leave the consortium**	Low	Redistribute tasks to other partners within the consortium or find a new partner.
WP1-WP9	**Unforeseen unavailability of qualified staff**	Low	Reassigning responsibilities among the staff of the involved departments. Esp. within the administrative departments dealing with contracts, finance, and legal matters backup plans are in place.
WP1	**Lack of response by individual partner**	Medium	Continuous progress monitoring of work by regular reports. In case of expected delay, early escalation at coordinator and decision on corrective actions. If

	preventing timely delivery of project results to EC		needed, redistributing of resources and manpower or alternative approaches will be considered to reach deadlines.
WP1, WP5, WP9	**Lack of the necessary expertise or skills to support the legal needs of the project**	Low	There is budget foreseen for external legal advice.
WP4, WP8	**No technical solutions for specific common services**	Medium	Continue monitoring and implement alternative services.
WP7	**Potential overlap of data-system effort with international (EUDAT) and national (InHPC-DE, GeRDI) projects**	Medium	Risk is turned into advantage by the LEXIS Data System building upon these infrastructures, instead of re-duplicating effort, as far as they are ready (collaboration partners in these projects are known / the projects are open).
WP9	**Dissemination of the project results is not sufficient to create impact**	Low	A dedicated work package for dissemination, exploitation and communication strategies will plan and execute this.
WP1, WP9	**Lack of communication or consensus within the consortium**	Low	The project coordinator and the members of the General Gathering have the necessary skills to resolve conflicts by adequate negotiation. Keeping close contact within the consortium by regular telephone conferences.
WP1	**The scope of the project is too ambitious for the given person months**	Medium	The project consortium will ensure through proper requirement analysis and planning that the objectives can be achieved.
WP1	**Project costs are in danger of exceeding budgeted costs**	Medium	Project management is to establish monitoring of resource consumption and to come up with plans to achieve the goals in a more resource efficient manner.
WP2, WP3	**Specialised hardware (Data Nodes) comes with excessive acquisition/installation/usage costs**	Medium	Our cost numbers are calculated so that at least a basic version of the optimised hardware can be acquired. Integration and basic problem resolution is – in essential parts – taken care of e.g. in **WP2**, so as to avoid surprises.
WP2, WP3, WP4, WP8	**Firewalling and other site-specific peculiarities prevent implementation of LEXIS Data System (i.e. hinder certain data flows)**	High	In HPC, such problems are very usual. However, judging from our ample experience with infrastructure projects (e.g. grid-based H2020 project ComPAT, grant #671564), in large parts these can be resolved by setting up (virtual) gateway machines with decent network connectivity. In addition, the Data Nodes (Burst Buffers) acquired within the scope of LEXIS can serve as high-performance gateway machines if needed. An appropriate network connectivity solution has been included in the financial planning for these nodes.

3.3 Consortium as a whole

3.3.1 Consortium Operational Capacity

The LEXIS Consortium is a balanced mix of complementary companies, supercomputing centres, research institutes, and universities to cover the spectrum of scientific and technological innovation that the project relies on, and to leverage the varying dissemination expertise and community links of the different project members for maximum impact. It is composed of seventeen (17) partners, geographically distributed in seven (7) European countries. LEXIS is built on top of infrastructure and knowledge provided by three supercomputing research centres: **LRZ**, one of Germany's three National Supercomputing Centres (Tier-0), **IT4I**, the Czech National Tier-1 Supercomputing Centre, and **ECMWF**, which provides world-leading global weather forecasts in a time-critical HPC environment. Industry is represented by four companies (**Bull**, **NALL**, **TESEO** and **Avio Aero**) and four SME partners are involved in LEXIS as well (**SEC**, **CYC**, **BAY** and **NUM**). The innovation potential of the LEXIS project draws - besides technical foundations - fundamentally from the scientific expertise provided by six research organisations: **ISMB**, **CEA**, **ITHACA**, **CIMA**, **GFZ** and **AWI**.

The roles of each participant in the Consortium as well as their assigned tasks and commitments are clearly identified in **Section 4** as well as their relevant previous experience and competences. Globally, the participants complement each other in terms of background, knowledge and competences and they collectively constitute a Consortium capable of fulfilling the project objectives.

The combined knowledge of all partners allows the project to progress quickly and to identify a final solution that equally reflects the requirements from both users and supercomputing centres.

Table 3.3.1a Partners involvement in the project implementation

Partner	Country	WP Leading	Task leading
IT4I	CZ	WP1	Task 1.1, Task 1.2, Task 1.3, Task 3.2, Task 3.4, Task 8.2
Bull	FR	WP2	Task 2.1, Task 3.1, Task 4.4
NALL	UK		Task 5.2
ISMB	IT	WP4	Task 2.2, Task 4.1, Task 4.2, Task 5.3, Task 9.1
TESEO	IT	WP9	Task 9.4
CEA	FR	WP6	Task 6.3
LRZ	DE	WP3	Task 2.3, Task 3.3, Task 8.1
ECMWF	UK		Task 3.5, Task 7.1, Task 7.2
ITHACA	IT		Task 6.4, Task 7.5
CIMA	IT	WP7	Task 7.3, Task 7.6
Avio Aero	IT	WP5	Task 5.1, Task 5.4
GFZ	DE		Task 6.1
AWI	DE		Task 6.2
SEC	FR		Task 4.3
CYC	CH	WP8	Task 8.3, Task 9.3
BAY	IE		Task 9.2, Task 9.5
NUM	FR		Task 7.4

3.3.2 Partner's multidisciplinary skills

Table 3.3.2a Multidisciplinary skills provided by partners

Partner	Type	Value to the Consortium
IT4I	RO-SCC	IT4I is the strategic research infrastructure in the Czech Republic. The centre is providing state-of-the-art technology and expertise in HPC/HPDA/HTC and makes it available for Czech and international research teams from academia and industry. IT4Innovations research activities are distributed into five laboratories: Parallel Algorithms Research Lab., Advanced Data Analysis and Simulations Lab., Modelling for nanotechnologies lab, Big Data Analysis Lab. and Infrastructure Research Lab. IT4Innovations currently operates two supercomputers, Anselm (94TFLOPS), and Salomon (2PFLOPS).
Bull	PB	Bull is the European leader in HPC and has commercialized a complete open exascale-class supercomputer offer containing Hardware and Software technologies. One of Bull strategies is around analytics, identified as one of the pillars of growth of the group and called Codex. Atos Codex is a set of fully integrated analytics solutions, leveraging an open, powerful and secure platform that allows organizations to quickly derive value from their data with a real return on investment.
NALL	PB	NALL is a leading supplier of FPGA accelerated computing solutions. For 20+ years, NALL has provided hardware, software and design services to enable customer's success in applications including high performance computing, network processing, near storage acceleration and real-time embedded computing. NALL is also focusing on integrated solutions such as Deep learning accelerators for Binarized Neural Networks.
ISMB	RO	Istituto Superiore Mario Boella (ISMB) is a research and innovation centre operating in the Information and Communication Technologies (ICT) domain. Founded in 2000 by Compagnia di San Paolo and Politecnico di Torino, today ISMB relies on the technological and process competences of around 150 researchers working in close cooperation with companies, academia and public administration. The Advance Computing and Electromagnetics ACE Research area will be involved, major Expertise in Cloud Computing infrastructures, scalability, interoperability and portability applications; Expertise on accelerating application on FPGA architecture; Expertise in large EU Projects Management.
TESEO	PB	TESEO has 40 years track record in design and manufacturing of low power and fibre optic innovative hardware matched with a significant hardware and software integration capability. Skills highly valued and recognized by major European players within Aerospace and transportation industry
CEA	RO	CEA has significant experience in real-time high performance embedded computing, and the integration of accelerated hardware components at the server architecture level. CEA has extensive technology transfer expertise, and is involved in numerous European initiatives and projects.
LRZ	RO-SCC	With its HR policy, aiming at optimising the collaboration with its customers, LRZ employs more Computational Scientists with domain-science backgrounds than with actual computer science background. In the proposed LRZ team for LEXIS, scientists with background in geophysics, astrophysics, biology, and computational science, as well as an engineer (electrical engineering/IT) are involved to support LEXIS's pilots/use cases with maximum efficiency.
ECMWF	RO-SCC	ECMWF has significant experience in producing world-leading global weather forecasts in a time-critical HPC environment, and runs two services as part of the EU's Copernicus programme; the *Atmosphere Monitoring Service* (CAMS) and the *Climate Change Service* (C3S). ECMWF also hosts the computational systems for the European Flood Awareness System (EFAS) and manages an extremely large persistent archive of meteorological data (MARS).
ITHACA	RO	ITHACA has a long-lasting experience in the framework of emergency management within the EU Copernicus programme, especially in processing and analysing remote sensing imagery to extract value added information, ITHACA will be therefore mainly involved in the pilots Weather & Climate and Earthquake and Tsunami, assessing the outputs of the models to identify how these can be integrated in operational emergency mapping workflows.
CIMA	RO	CIMA is a private non-profit research organization committed to promote the study, scientific research, technological development and higher education in engineering and environmental sciences to improve civil protection, public health and the preservation of aquatic and terrestrial ecosystems. CIMA Research Foundation is part of the Italian National Civil Protection System Network of Excellence within the fields of floods and forest fires, with special focus on early warning systems. Its founding institutions are the National Civil Protection Department of Italy, the University of Genova, the

		Government of the Region of Liguria, and the Administration of the Province of Savona. CIMA has long experience in design, deploying and executing complex modelling workflows in the environmental sector.
Avio Aero	PB	Avio Aero has long experience in design, manufacturing and maintaining components and systems for civil and military aviation. We provide our customers with innovative technology solutions to quickly respond to the endless changes required by the market: additive manufacturing, rapid prototyping, as well as technologies dedicated to the production of mechanical transmissions, turbines and combustors. Our challenge is to develop new technologies for applications on architectures that can reduce energy consumption, make aircraft engines lighter, and provide better performance. Through continuous investment in research and development and a consolidated network of relationships with major Universities and international Research Centers, we have developed a globally acknowledged technological and manufacturing excellence. In this scenario of big challenges, digitalization of the company came out as a natural consequence of maintaining competence and services at the highest possible level and leverage experts in different digital technology domains, leading to innovative solutions as HPC, Cloud, IoT and Big Data.
GFZ	RO	GFZ is one of the leading scientific institutions in the field of natural hazards. The participating group at GFZ is well connected with partner institutions in the area of earthquake hazard and risk and together extending their activities into the area of big-data and citizen science.
AWI	RO	The tsunami modelling group at AWI is experienced in technology and knowledge transfer, and it is part of the scientific computing group at the computing center that bridges between scientific users and IT infrastructure, with a focus on HPC and data workflows.
SEC	SME	SEC has built its recognized security expertise in Cloud, Big Data, Container and Virtualization. Founding members of CSA in 2009, SEC appears in several major publications or events such as Forbes or ACM/SAC conference "A Security Analysis of Amazon's Elastic Compute Cloud Service"
CYC	SME	CYC is an innovative cloud based accounting and billing provider with a focus on complex, usage based accounting arising in enterprise and cloud contexts. It has a deep understanding of the base technologies underpinning cloud systems and a good understanding of the HPC world. It also has concrete software development and applications operations skills.
BAY	SME	Recognised within Europe as a key provider of training and consultancy. Expertise includes Code Modernisation, HPC development, Performance Benchmarking and Artificial Intelligence.
NUM	SME	NUM has a strong expertise on meteorology and atmospheric modelling, especially in regards to local air quality (impact on health, …), as well as an experience to use HPC cluster (internally or by cloud)

3.4 Resources to be committed

The resources required for performing this project have been accurately defined. As shown in **Table 3.4a** the total PMs of the LEXIS project is **1 502** spread over 2.5 years for a total EC requested contribution of cca **€12 218 000**. The following table indicates the number of person/months over the whole duration of the planned work, for each work package and for each participant.

The budget shows a balanced project, both in terms of effort and costs (see **Figure 3.4a**), with efficient utilization of the resources; yet sufficient to efforts required to complete the required tasks successfully. The project's work plan has been broken down into work packages, tasks and individual work elements for each task. The consortium has been very carefully put together so that there is no overlap and duplication of the resources. The distribution of the effort of SMEs demonstrates that LEXIS is has significant share of effort in this project (16%) while a reasonable amount of resources is devoted to the project management activities as well as the impact creation and business planning. Travel costs represent 3% of the total budget and allow attending the three GG Meetings foreseen each year in addition to the review meeting and a limited number of ad hoc-technical meetings or workshops.

Table 3.4a: Summary of staff effort

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	Total PM
IT4I	**18**	17	38	32	26	17	5	24	14	191
Bull	3	**17**	39	44	5	12	9	14	10	153
NALL	2	5	0	2	5	2	0	0	3	19
ISMB	3	19	31	**40**	17	8	7	6	4	135
TESEO	3	10	7	3	4	4	24	6	**34**	95
CEA	3	8	13	7	0	**55**	0	5	5	96
LRZ	3	18	**24**	22	16	16	10	17	4	130
ECMWF	3	9	20	12	0	0	31	6	5	86
ITHACA	3	8	1	3	0	32	32	0	5	84
CIMA	3	10	11	10	0	6	**47**	0	5	92
Avio Aero	3	8	4	7	**43**	0	0	0	4	69
GFZ	3	11	17	14	0	25	0	0	4	74
AWI	3	6	0	0	0	32	0	0	2	43
SEC	3	10	21	24	0	10	10	7	3	88
CYC	3	3	2	0	0	0	3	**19**	6	36
BAY	3	7	5	9	0	6	6	14	19	69
NUM	3	9	0	0	0	0	28	0	2	42
Total PM	**65**	**175**	**233**	**229**	**116**	**225**	**212**	**118**	**129**	**1502**

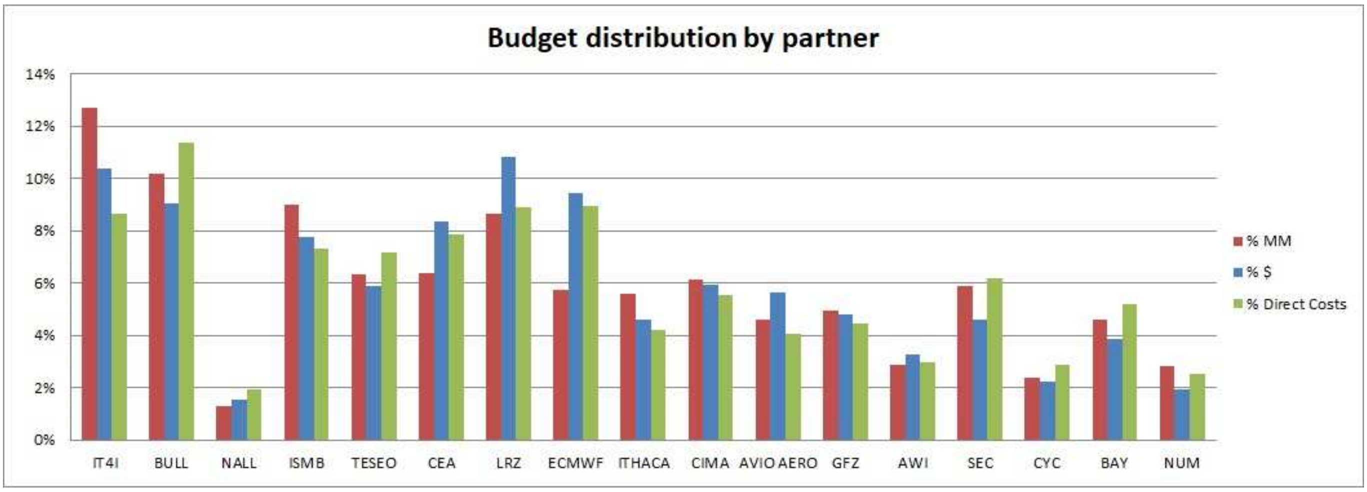

Figure 3.4a - Budget and Effort Distribution

Equipment and consumables represent 6.5% of the total budget and are related to the purchase of the equipment devices needed for the project research activities and showcase the results of the project. In **Figure3.4b** the balanced distribution of effort among the WPs.

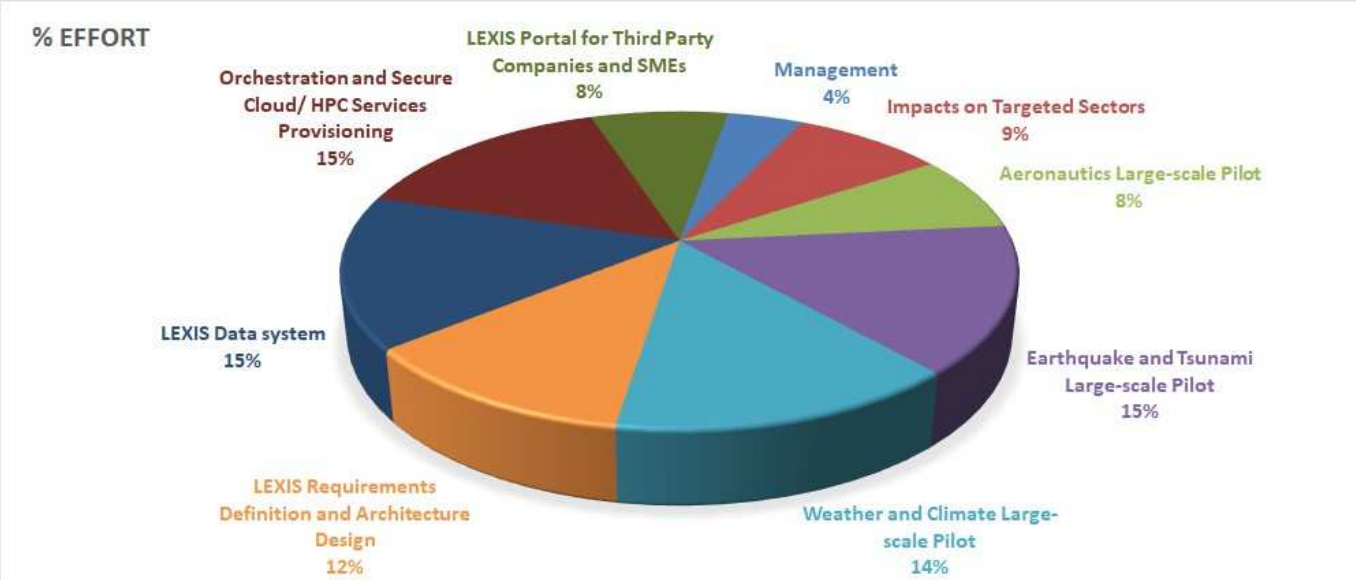

Figure 3.4b – Work Package Effort Distribution

Table 3.4b: 'Other direct cost' items (travel, equipment, other goods and services, large research infrastructure)

Participant 1 - IT4I	**Cost (€)**	**Justification**
Travel	32 000	Large-scale and geographically distributed project; 4 participants: sum is equiv. to: 20k€ total for project/WP/project-review meetings (0.7-1k€ per meeting and person; 5-7 meetings); additional 6k€ for participation in intl. conferences & other dissem. actions (0.7-1.5k€ per journey per person, 1-2 journeys)
Equipment	95 000	Rack with data nodes (2 nodes, each 30k€), these Burst Buffers form an integral part of LEXIS's strategy to accelerate workflows and must be attached to high-performance network switches (100GE, €20k) and local storage (€15k)
Other goods and services	10 000 10 000 6 000	Organisation of F2F meeting, subsistence for members of EAB Booths Audits
Total	**150 000**	

Participant 2 - Bull	**Cost (€)**	**Justification**
Travel	28 000	Travel costs
Equipment	100 000	For 2 data nodes servers based on Bull Sequana S series
Other goods and services	5 000	Audit, organisation of F2F meeting
Total	133 000	

Participant 5 - TESEO	**Cost (€)**	**Justification**
Travel	20 000	Budget allocated for travelling, meetings, workshops and events.
Equipment	80 000	Budget allocated for materials for gateway prototypes to be installed in situ.
Other goods and services	3 000	Budget allocated for digital communication, experts support (website and social media development)
Total	103 000	

Participant 7 - LRZ	Cost (€)	Justification
Travel	26 000	Large-scale and geographically distributed project; 4 LRZ participants: sum is equiv. to: 20k€ total for project/WP/project-review meetings (0.7-1k€ per meeting and person; 5-7 meetings); additional 6k€ for participation in intl. conferences & other dissem. actions (0.7-1.5k€ per journey per person, 1-2 journeys)
Equipment	95 000	Rack with data nodes (2 nodes, each 30k€), these Burst Buffers form an integral part of LEXIS's strategy to accelerate workflows and must be attached to high-performance network switches (100GE, €20k) and local storage (€15k)
Other goods and services	49 000 6 000	€44k predicted energy costs (10kW for essential data nodes mentioned above; 18ct/kWh with climate factor 1.4, about 2yrs running time); € 5,000 housing rent installation and other costs, organisation of F2F meeting Audit
Total	**176 000**	

Participant 11 - Avio Aero	Cost (€)	Justification
Travel	10 000	2 travels per year (1 person)
Equipment	206 000	This cost refers to the purchase expense needed to buy (and including the fee due to the first year of license maintenance and technical support) a software asset that will be properly selected to carry out the task “5.2 Rotating parts use case set-up and run” in WP5 “Large Scale Pilot Aeronautics Use Case”. This software shall be a Commercial Off-the-Shelf (COTS) Computational Fluid Dynamics (CFD) innovative application based on an advanced numerical method allowing engineers to model and simulate large-scale problems related to fluid-dynamic flow field of mechanical parts rotating in mixed air-oil space, dealing with complex boundary geometries. It shall support multi-core CPU computing (OpenMP and MPI), GPU computing using GPU boards, and multi-GPU computing, taking advantage of parallelization to shorten calculations and handle large-scale computing. Depending on the reliability of results that will be obtained in task “5.2 Rotating parts use case set-up and run”, its field of application may be extended to other computer-aided simulation tasks in the framework of aeronautical engineering, and also be exploited and adopted by other big aeronautical companies. The investigation and optimization of aeronautical products are expected in this way to vamp and raise at unpredictable levels of design opportunities.
Other goods and services	80 000 10 000	This cost includes 80 000€ for the license maintenance and technical support fee related to the software previously described in section “Equipment” and 10 000€ for submitting, to the European Commission, the financial costs certification (FCS) that will be in charge of an external Auditing company.
Total	**306 000**	