LEXIS Newsletter
June 2021

The LEXIS project has a new video

The LEXIS project has prepared a general video "**LEXIS project – platform for advanced computing, big data and cloud**". The video describes the project and informs about the LEXIS platform and summarizes its large-scale pilots. Have a look at the video and if you like it, let us know via the YouTube channel or our social media networks.

Virtual stand welcomes our guests at ISC 2021

Another means of communication, which the LEXIS project has prepared, is the virtual stand. It welcomes the guests of our virtual booth at ISC 2021. It introduces the pilots and our videos, but especially presents our posters. Take a sneak peek in our virtual stand. **And don't forget to subscribe to the LEXIS e-newsletter** and take the chance to win one of these **wonderful prizes!**

LEXIS
External Advisory Board

Introducing LEXIS External Advisory Board (EAB)

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 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.

Since the start of the project, we have organised several meetings with the EAB members e.g. during technical meetings or at the conferences, either in physical or virtual presence.

We thank the EAB board members for their valuable feedback and advice.

Now, let us introduce you the LEXIS EAB members.

Professor Fatos Xhafa

Fatos Xhafa is Full Professor at the Technical University of Catalonia (UPC), Barcelona, Spain. He received his PhD in Computer Science in 1998 from the Department of Computer Science of BarcelonaTech. He was a Visiting Professor at University of Surrey (2019/2020), Visiting Professor at Birkbeck College, University of London, UK (2009/2010) and a Research Associate at Drexel University, Philadelphia, USA (2004/2005).

Prof. Xhafa has widely published in peer reviewed international journals, conferences/workshops, book chapters and edited books and proceedings in the field (Scopus h-index 41 as of October 2020). He is awarded teaching and research merits by Spanish Ministry of Science and Education, by IEEE conference and best paper awards.

Prof. Xhafa serves as a reviewer for international journals and books from major publishers and acts as a member of Editorial Boards and as a Guest Editor of Special Issues. He is Editor in Chief of the Elsevier IoT Journal, Editor in Chief of IJGUC Journal, Inderscience, Editor in Chief of the Elsevier Book Series Intelligent Data-Centric Systems and of the Springer Book Series Lecture Notes in Data Engineering and Communication Technologies. He is a member of IEEE Communications Society, IEEE Systems, Man & Cybernetics Society and Emerging Technical Subcommittee of Internet of Things.

His research interests include parallel and distributed algorithms, massive data processing and collective intelligence, IoT and networking, P2P and Cloud-to-thing continuum computing, optimization, security and trustworthy computing, machine learning and data mining, among others. Further biographical information for Prof. Xhafa can be found at http://www.cs.upc.edu/ fatos/ and at his Scopus ORCID: http://orcid.org/0000-0001-6569-5497

Dr Alan Sill

Alan Sill holds positions at Texas Tech as Managing Director of the High Performance Computing Center, Adjunct Professor of Physics. He is also a Co-Director of Cloud and Automatic Computing Centre (CAC). Dr. Sill holds a PhD in particle physics from American University and has an extensive track record of work in scientific computing. He has published extensively in topics spanning cloud and grid computing, scientific computing, particle and nuclear physics, cosmic ray physics and radioisotope analysis.

He serves as a President of the Open Grid Forum, an international computing standards organization. He is an active member of the IEEE Standards Organization, DMTF, and other computing standards working groups, and serves either directly or as liaison for the Open Grid Forum on several national and international standards roadmap committees. He has helped to launch several large-scale cloud and grid computing projects, including the Open Science Grid (OSG), the Worldwide Large Hadron Collider Grid, Southeastern Universities Research Association SURAgrid, and Texas Internet Grid for Research and Education (TIGRE).

In addition, he has served as general chair, program committee chair, or principal organizer for several different international conferences including the IEEE/ACM Utility and Cloud Computing (UCC) and Big Data Computing, Applications, and Technologies (BDCAT) and IEEE International Conference on Autonomic Computing (ICAC), as well as the ongoing set of workshops on Data Center Automation, Analytics, and Control (DAAC) held at the IEEE/ACM International Conference for High Performance Computing, Networking, Storage, and Analysis (Supercomputing) conference series.

Dr. Sill is a past member of the editorial board for IEEE Cloud Computing and current member of the advisory boards for the StandICT international standards coordination project and several other EU-funded projects. Further biographical information for Dr. Sill is available on EDUCAUSE.

Professor Leonard Barolli

Leonard Barolli is a Full Professor at Fukuoka Institute of Technology (FIT), Japan. He received the Ph.D. degree from Yamagata University, Japan in March 1997. From April 1997 to March 1999, he was working as a Post Doctoral Fellow Researcher of the Japan Society for the Promotion of Science (JSPS) at the Department of Electrical and Information Engineering, Yamagata University. From April 1999 to March 2002, he was a Research Associate at Department of Public Policy and Social Studies, Yamagata University. From April 2002 to March 2003, he was an Assistant Professor at the Department of Computer Science, Faculty of Engineering, Saitama Institute of Technology (SIT), Japan.

Prof. Barolli was an Editor of the Information Processing Society of Japan (IPSJ) Journal, and he is serving in the Editorial Board of many International Journals. He has been a Guest Editor for many Special Issues and engaged as PC Co-Chair and General Co-Chair of many International Conferences. Prof. Barolli is the Steering Committee Chair of the International Conferences AINA, CISIS, BWCCA, INCoS, IMIS, NBiS, EIDWT and 3PGCIC. He is also organizer of many International Workshops such as MNSA, HWISE, WAIS, TwC, BICom, EDCS, NTCCA, NGWMN, R3IC and SMECS. Prof. Barolli has served as a reviewer for many International Conferences and Journals.

He has published more than 1,100 papers in Refereed Journals, Book Chapters and International Conference Proceedings. For more than 25 years, his research has been supported by JSPS Grant-in-Aids for Scientific Research and other Japanese Foundations.

Dr Giovanni Sembenini

Giovanni Sembenini is Deputy Director of the NATO Science and Technology Organisation Centre for Maritime Research and Experimentation (STO CMRE, www.cmre.nato.int).

He holds a PhD in Aerospace Engineering and master's degrees in aeronautical engineering and in space engineering from the Politecnico of Torino (Italy). Additionally, he received a master's degree in electronics engineering of complex systems from the University of Rome 3. He is also a graduate of the NATO Accreditation Program for Managers and of the Army Staff Officer School.

In his professional career, he has served as a member of the Advisory Board of the Aerospace Testing Seminar, of the Scientific Advisory Committee of the Von Karman Institute, and as the NATO STO Systems Concepts integration panel Italian Principal Member. He has chaired and has been a member of the programme committees of several Research Task Groups, symposia and workshops of the NATO STO. He has been a member of several review boards for research projects at national, international and EU levels as well as an ethical evaluator for EU research projects and member of the steering board of various EU projects.

With his extensive research and academic experience, Dr. Sembenini has been an active reviewer for the IEEE Systems Journal. He has authored more than 30 research papers and articles in scientific journals and technical magazines. As a lecturer, he has tought graduate-level courses on Space Systems Engineering and Logistics (Politecnico di Torino) and on Climate Security (Biocampus University, Rome), to give a few examples.

The Weather and Climate Pilot

The weather and climate pilot focuses on a complex system, to provide a diverse set of forecasts concerning weather, flood, forest fire, air pollution and agriculture.

The weather and climate pilot encompasses several complex workflows each consisting of various meteorological components. These workflows will include in their final release the ingestion of conventional and unconventional observations, global weather models, regional weather models, application models and socio-economic impact models.

Fully orchestrated workflows of the weather and climate large-scale pilot in LEXIS have already reached considerable intrinsic complexity as in the case of the **WRF (Weather and Research Forecasting)** model and **RISICO (RISchio Incendi e Coordinamento / Fire Risk and Coordination)** model. WRF is a proven mesoscale numerical weather prediction system, designed to serve both operational forecasting and atmospheric research needs. It features multiple dynamical cores, a 3-dimensional variational (3DVAR) data assimilation system, and a software architecture allowing for computational parallelism and system extensibility.

RISICO is a mathematical model to support operators in forest fire prevention activities. RISICO is powered by a continuous data flow consisting of meteorological information both from weather forecasts and in situ observations. At the current stage the WRF and RISICO workflow has already realized a significant degree of convergence between cloud computing and HPC.

On the cloud computing side, it concerns the creation of a virtual machine responsible for downloading the initial and boundary conditions of global circulation model as well as geographic static data and mounting the staging area. Subsequently, the docker is installed and the WRF Preprocessing System (WPS) to prepare initial and boundary conditions is executed. Then a WRF job is created and executed on the HPC infrastructure using aforementioned initial and boundary conditions data.

The WRF results are copied back on the cloud infrastructure staging area allowing for the execution of the RISICO container, whose results are stored on the LEXIS Distributed Data Infrastructure (DDI).

Likewise, hydrological risks (Continuum model), heavy rainfall alerts (ERDS – Extreme Rainfall Detection System) and air quality (ADMS) can be assessed. With respect to previous projects (FP7 DRIHM), where only limited workflow automation was available, with LEXIS, a much broader range of applications is possible without experts' assistance and necessity to execute steps manually.

The available computing systems are optimally leveraged, so that the application models run perfectly as containers on the Cloud infrastructures at LRZ and IT4I, while WRF is a classical HPC job.
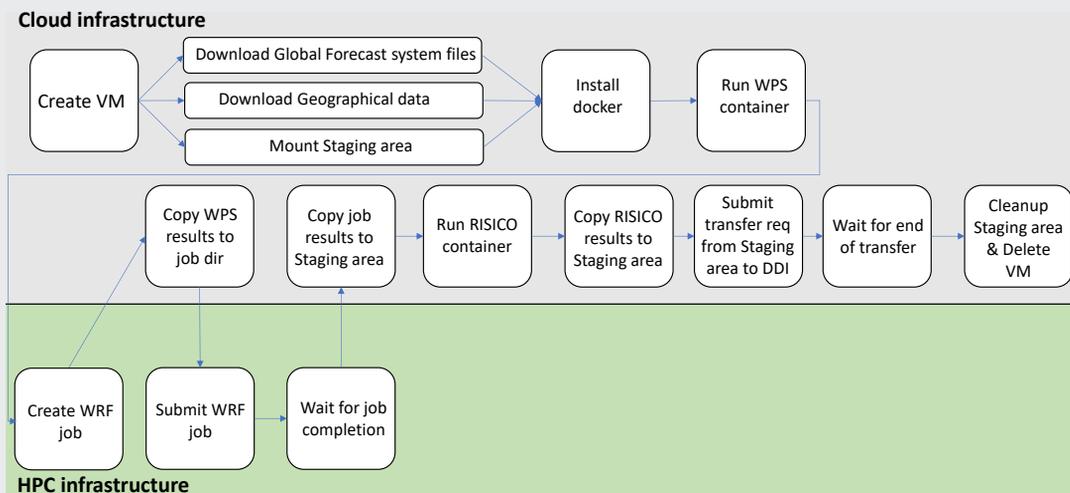
WRF-RISICO workflow from the weather and climate large-scale pilot, as presented on a SC20 conference poster (M. Hayek et al.)
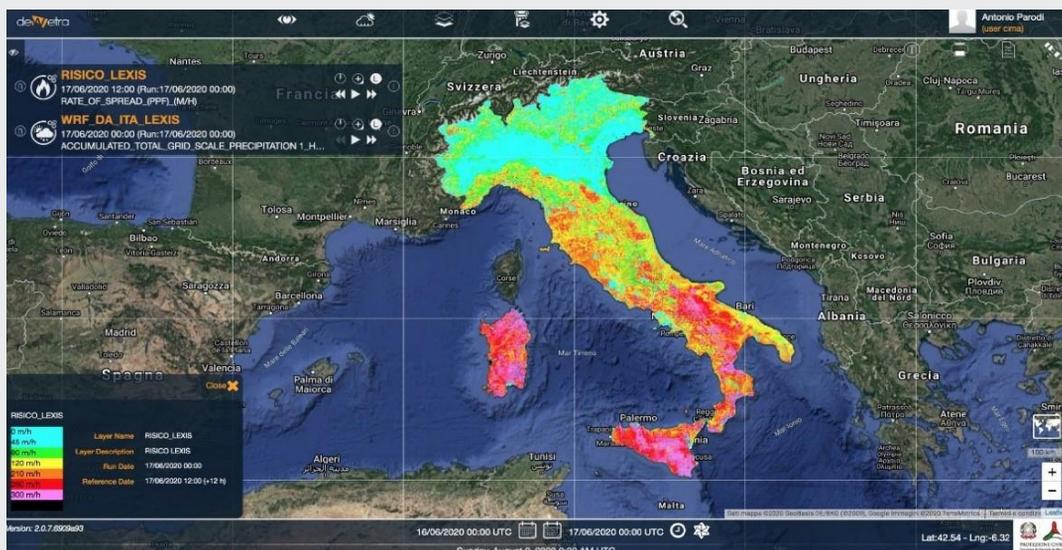
Workflow results visualized with Dewetra platform of the Italian Department of Civil Protection, with CIMA contribution

A Sneak Peek of the LEXIS Portal

We are glad to provide a sneak peek into the upcoming LEXIS portal web user interface. The portal is undergoing last rounds of internal validation before being released to first LEXIS Open Call shortlisted organizations. The portal UI promises unparalleled ease of use for HPC and non-HPC users alike.

The interface to manage data sets on the Distributed Data Infrastructure in the LEXIS portal – an essential component for managing input and output data of LEXIS workflows.

With its intuitive data set management interface, LEXIS users will get easy access to large sets of public datasets for experimentation, as well as ability to create their own datasets and with full control over published metadata as well as fine grained data access control possibilities.

The soon to be released portal will also allow easy creation, management, and sharing of HPC and cloud workflows among registered users. The portal abstracts all the complexity out of orchestration of a complex workload, its planning and execution over distributed HPC compute resources.

Users will be able to do large scale experiments over some of the biggest HPC centers in Europe through an ergonomic and intuitive interface.

The interface to manage data sets on the Distributed Data Infrastructure in the LEXIS portal – an essential component for managing input and output data of LEXIS workflows

Workflow progress view and control in the LEXIS portal – the interface which addresses the very core of the LEXIS platform: the orchestration system for workflows

Latest News from the Open Call 2020-2021

Deep Learning workflows boosted by the LEXIS platform: LIS2S project

In the scope of the LEXIS Open Call, the LIS2S project from Orbyta with its workflows based on Deep Learning has been selected for cooperation. For the LEXIS team, this will be an important experience in testing and validating the platform technology with **Machine Learning (ML)** workflows.

We strive to demonstrate the benefit for the user: an optimum workflow execution and easy workflow control to meet tomorrow's requirements in quality, technical leadership and user friendliness.

Deafness and hearing impairment are conditions that affect as much as 5% (about 360 million) of the world's population (WHO, International Ear Care Day, 2014). Among them, there are about 877,000 in Italy (ISTAT).

Despite the statistics, deafness has remained a socially less known condition and the difficulties it entails are mostly ignored.

LIS is the Italian sign language and Orbyta https://orbyta.it/ is using Deep Learning technologies to implement the "**LIS to Speech**"(**LIS2S**) application to solve this social problem. This software aims to facilitate communication between deaf and hard-of-hearing people without LIS knowledge. It translates LIS signs recorded on video into Italian words by reproducing them in both audible and textual form.

Since the goal is to reach as many people as possible, the model is based on skeletal features of hands, body and face extracted from RGB videos without the need for additional equipment, such as depth cameras or gloves with sensors.

The research has been intensive and has led to successful results that are in line with the state of the art for other sign languages. The team developed various AI models and designed different use cases of the app suitable for different types of users.

Thanks to the LEXIS team, LIS2S was successfully migrated to the Barbora cluster. The LEXIS infrastructure is used in the LIS2S project as a computational platform for ML training and to perform video processing for data augmentation. The models are repeatedly trained for improvements, upgrades, and dictionary enlargement.

LEXIS also offers the LIS2S team data storage for structured and unstructured data. Besides being a back-end for the automated workflows, this is experimentally used for real-time feedback collection, service management, loading videos into the dictionary, and for performance and usage supervision.

At the technical level, the model uses GRU recurrent neural networks and data augmentation, hyperparameter tuning, feature extraction, K-fold cross-validation, text mining, multivariate data analysis, connected network component analysis, early stopping techniques. Also, categorical Cross-Entropy loss studies have been performed.

This collaboration is a strong booster for reaching the LIS2S project's goal this year: real-time translation of entire sentences into syntactically correct Italian sentences with high reliability.

The validation of the LEXIS platform with these workflows, and its improvement based on Orbyta's feedback will be important for the LEXIS sustainability and exploitation phase. They open up possibilities for further cooperations, for example in continuous provisioning of scalable computing resources to Orbyta's data scientists, during the LEXIS exploitation phase and beyond.

Recognition of facial, body and hand features and gestures by LIS2S

Zero Trust Architecture (ZTA) within LEXIS

The Classical approach to security

Lots of companies rely mostly on perimeter-based network security, building their network as a castle with defenses on the perimeter. However, this approach is not made to cope with current ways of working, if they involve – for example – remote connectivity or interfacing to cloud services.

Such situations require opening access to internal resources/assets, and sooner or later are likely to allow an attacker to breach the perimeter and move on laterally.

LEXIS goes one step further

In the LEXIS project, we actively cope with the problem sketched above: The build-up of the LEXIS Cloud-HPC-Big Data platform has been accompanied by the development of a security concept following the "Security-By-Design" and "Zero-Trust" principles. A secure **Zero-Trust Architecture (ZTA)** has been implemented.

This has an immediate impact on LEXIS data, compute, orchestration, portal and billing services, and also on the unified **Identity and Access Management (IAM)** solution within LEXIS. This solution, based on the Open-Source product "Keycloak", is the most important component of the platform's Authorisation and Authentication Infrastructure (LEXIS AAI). It allows for authentication based on tokens (following, e.g., the OpenID Connect standard).

ZTA concept and implementation in LEXIS services

The main idea of a Zero-Trust concept is to protect assets by minimizing access possibilities and by enforcing authentication and authorization for each access request.

To give an easy example, services do not trust any other service or information source except the LEXIS AAI to check authorization for an actual access. Instead, all services verify the identity and permissions, talking to the IAM system with what they know about the user.

The figure on the next page focuses on our unified, federated IAM/AAI solution and its relationships with the main LEXIS components, to showcase LEXIS as an example for how a Zero-Trust architecture can be implemented.

The figure illustrates these relationships. It indicates that all components such as LEXIS DDI, LEXIS Orchestration Infrastructure are communicating with the LEXIS AAI to:

- validate every access token (received, e.g., from another component) which contains the identity asking to act on or access a resource,

- check the component-specific permissions of such an identity, and to

- pass an access token to any other component they need to interact with.

The solution has been fine-tuned to allow any component to only retrieve directly relevant access permissions for an identity.

As a concrete example, the LEXIS compute component checks the validity of an access token received through the API call (using an API call for token introspection), then it interacts with LEXIS AAI to request a specific access token for itself and information on compute access permissions for the identity in question (using an API call for getting user info).

A sound security architecture within LEXIS, beyond the ZTA

The LEXIS AAI with its IAM system deployment was designed as a unified component handling authentication and authorization in LEXIS, avoiding a fragile architecture with many small IAM solutions synchronizing with each other.

With a federated cross-data centre deployment of Keycloak, a single point of failure is avoided.

The LEXIS security architecture is not only focused on reliability and the zero-trust principle, but also implements further relevant good practices, such as limiting authorizations to a viable minimum (principle of least privilege), and proper firewalling and monitoring of network and service components.

Further information about Zero-Trust Architecture:

- csrc.nist.gov/publications/detail/sp/800-207/final

- www.nsa.gov/News-Features/Feature-Stories/Article-View/Article/2515176/nsa-issues-guidance-on-zero-trust-security-model/

- dodcio.defense.gov/Portals/0/Documents/Library/(U)ZT_RA_v1.1(U)_Mar21.pdf
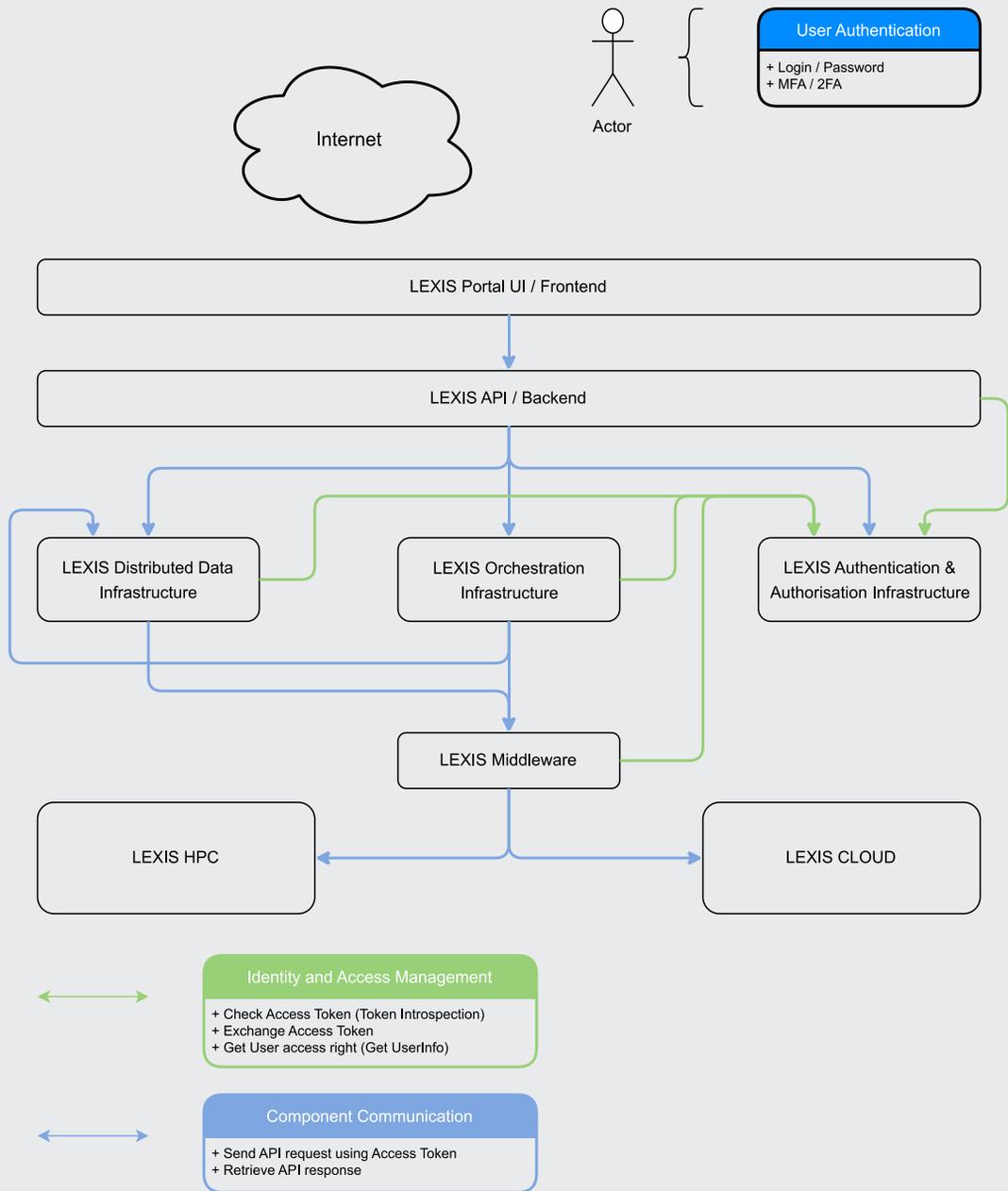
Communication with the LEXIS AAI for authentication and authorization
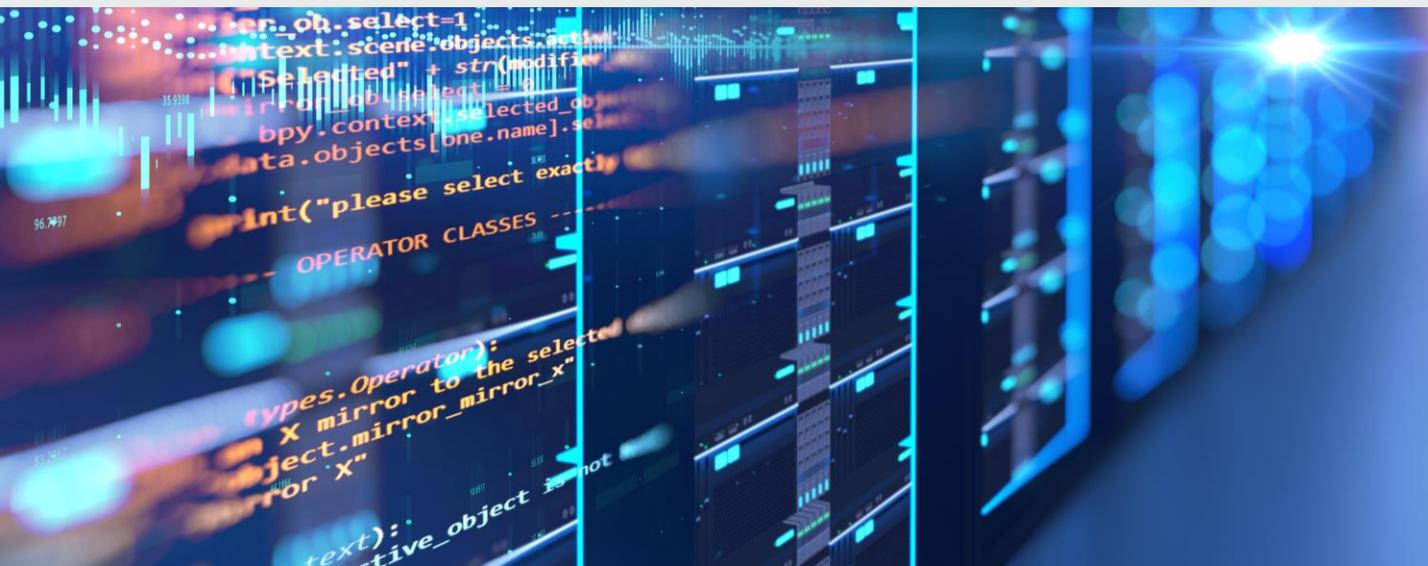
LEXIS
Activities in brief

MAY 2021

Data Week 2021
May 25–27, 2021, digital

Data Week is the gathering of the European Big Data Value and Industrial AI research and innovation community. The event follows the tradition of the Big Data Value Summit for promoting opportunities, sharing knowledge and fostering ecosystem development.

At the Data Week 2021 event, co-organised by DAIRO/BDVA and the EUHubs4Data project, **LEXIS offered a session on "Future challenges in IoT, AI, and convergence of HPC & Cloud & Big Data" on May 26, 2021.** The session gathered ICT-11 projects currently working on these topics, including LEXIS.

After talks from all projects, where Stephan Hachinger gave an overview about LEXIS, a panel discussion helped us develop technological visions. We built upon ideas of the BDVA Strategic Research and Innovation Agenda (SRIA), starting from basic data, workflow management to the actual applications involving IoT and AI, on the way towards a sound Cloud-/HPC-based ecosystem.

The talk and the discussion sessions were led by **Jan Martinovič** and **Stephan Hachinger**, with **Marc Levrier** and **Martin Golasowski** from the LEXIS project as panelists. With total 104 attendees, our session was one of the most visited at the Data Week 2021.

Agenda:
https://www.big-data-value.eu/dw21-agenda/

JUNE 2021

iRODS Virtual User Group Meeting 2021
June 8–11, 2021, digital

On June 9th, 2021, **Martin Golasowski (IT4I)** and **Mohamad Hayek (LRZ)** of LEXIS WP3 gave a tech talk on the LEXIS Distributed Data Infrastructure at the iRODS UGM 2021. They discussed details of the redundant setup of DDI components, and of the DDI APIs for data staging and beyond.

The talk, including contributions of **Rubén J. García-Hernández (LRZ)** on OpenID-Connect authentication for iRODS, won the **iRODS UGM Best Student Paper Award** and triggered a number of fruitful technical discussions.

Event: https://irods.org/ugm2021/

FORUM TERATEC 2021
June 22–24, 2021, digital

This online event, which brought together the best international experts in the field of HPC, simulations and Big Data, offered the registered users a series of workshops, keynote and plenary sessions, and a virtual exhibition.

The LEXIS project took this opportunity to create **a virtual booth**, where visitors found relevant information about the project, had a look at the technical poster or watched the new video. Additionally, **Marc Levrier (ATOS)** presented the LEXIS project at the **Europa Village (live session) on June 22, 2021**.

Event: https://teratec.eu/gb/forum/index.html

ISC High Performance 2021
June 24 – July 2, 2021, digital

ISC 2021 is shaping tomorrow, and LEXIS will contribute to drive this process.

The ISC 2021, one of the biggest HPC events this year is just around the corner. The event offers excellent opportunities for digital technology enthusiasts from research, academia and industry to keep up with state-of-the-art developments and trends in HPC, machine learning and High-Performance Data Analytics (HPDA).

The LEXIS project showcases itself in a virtual booth at this particular event. We offer the visitors possibilities to get to know the LEXIS project, our technological focuses on HPC/Cloud/Big Data and to find out more about our collaborations with other projects or finally about interesting articles in our blog posts or the e-newsletters.

Stop by our booth at ISC 2021 and win wonderful prizes!

Event: https://www.isc-hpc.com/

JULY 2021

ASITA Academy 2021
July 1–2–9–16–23, 2021, digital

Paola Mazzoglio (Politecnico di Torino - ITHACA) of LEXIS WP7 will present a work titled **"The WRF-ERDS workflow in the November 2020 Calabria flood event"** at the ASITA Academy conference on July 9th.

The work, contributed by co-authors from the LEXIS project (Mazzoglio et al, 2021) compares the WRF-ERDS workflow result, run on LEXIS HPC/Cloud computing resources with rain gauge data and with the daily bulletins issued by the Italian Civil Protection Department, demonstrating improved forecasting accuracy for extreme rainfall events.

Event:
https://www.asita.it/en/home-english/