

IBERGRID 2022

Monday, 10 October 2022 - Thursday, 13 October 2022

Universidade do Algarve

Book of Abstracts

Contents

The SQAaaS Hackathon	1
Quality Assurance Models in the framework of EOSC-Synergy	1
From the "Proof of Concept" to EuroHPC: a user journey across the HPC landscape in Europe	2
Using Dynamic DNS for dynamic service deployment, migration and high availability	2
FedCloud client: the powerful client for EGI Federated Cloud	3
Compare computing sites using EOSC-Performance service	4
Integrated dataset placement service for scientists	4
Water Monitoring Sentinel Cloud platform (WORSICA): a new service for detecting coastline, coastal inundation areas, and inland water bodies using remote sensing.	5
Deep learning to estimate wine volume from one single-view image	6
Making Quantum Technologies more accessible: A new approach to bring the future closer	7
Coastal Digital Twins: building knowledge through numerical models and IT tools	7
FAIRness Evaluation, Validation and Advising with FAIR EVA	8
Developing a Distributed and Fault Tolerant Dataverse Architecture	9
CSCALE OpenEO Deployment	9
Helmholtz Federated IT Services: Enabling innovative software and supporting services for research	10
SSH with Federated Identities	10
jUMP: Construction of a oceanic sound propagation modelling portal for research, incorporating IT services from EOSC.	11
Serverless Services for Scientific Cloud Computing	11
EOSC-Synergy Jenkins Pipeline Library	12
Automating cloud deployment infrastructures with Kolla	13

The Software Quality as a Service (SQaaS) portal	13
Parallel FaaS on the Edge-Cloud continuum	14
User space containers with udocker	15
EPIC Cloud: a secure, GDPR-compliant, open-source cloud platform for life-science applications	15
Ozone assessment service (O3as) in action	16
COASTNET: Integrating Earth Observation and in-situ data for coastal management . . .	16
Repositories for EGI	17
FAIR data in the Photon and Neutron community	18
Automating scientific dataset management and processing using Onedata	19
Secret management service for EGI infrastructure	19
GRAPEVINE project: hiGh peRformAnce comPUting sERVICES for preVentIon and coNtrol of pEsts in fruit crops	20
Linking GitOPS towards fast innovation over BigHPC	20
PSNC role in digital transformation of the agriculture sector in Poland	21
IBERGRID status	21
The Portuguese Distributed Computing Infrastructure (INCD)	22
interTwin: An interdisciplinary Digital Twin Engine for science	22
Galicia Supercomputing Center (CESGA): Infrastructure, Research, and the Road Ahead	23
Security in the Spanish academic network	23
Digital Twins for GEOphysical extremes	23
Repositories for Cancer imaging	23
University of Algarve	23
Scientific Computing at INFN and beyond	23
Cybersecurity at the Portuguese R&E Network	24

Hackathons / 13**The SQAaaS Hackathon**

Authors: Diana María Naranjo Delgado¹; Pablo Orviz Fernández²; Jorge Gomes³; Samuel Bernardo⁴

¹ UPV

² IFCA-CSIC

³ Laboratório Física Experimental de Partículas (LIP), Portugal

⁴ LIP

Corresponding Author: orviz@ifca.unican.es

The Software Quality as a Service (SQAaaS) portal [1] has been developed during the lifespan of the EC-funded EOSC-Synergy project. One of its main contributions is to assess the quality of software projects, including the evaluation of both source code and services, through the Quality Assessment and Awarding (QAA) module. The quality criteria for source code includes static analysis, such as style compliance or unit testing coverage, and the presence of specific metadata, such as the applicable license, the definition of a code of conduct or documentation. Likewise, the quality criteria for services include its capability of being automatically deployed and further dynamic testing.

Hence, based on the presence of a code repository, the QAA iterates over the supported criteria, checking their compliance through the selection and execution of the appropriate set of open source tools. In the particular case of assessing services, the user shall provide the URL of the code repository that contains the Infrastructure as Code (IaC) files that manage the deployment of the service.

After the completion of the assessment process, a report is generated detailing every quality attribute being checked by the aforementioned tools. Each attribute or criterion is identified using a code, outlined with a small description and reported as successful or a failure. According to EOSC-Synergy standards, the quality achievements being obtained as a result of the assessment process can additionally result in the issuance of a digital badge, provided that a minimum of coverage has been reached. There are three classes of EOSC-Synergy badges: bronze, silver and gold, from lower to higher rated software.

The goal of the SQAaaS Hackathon is to improve the quality aspects of code repositories related to research software. Thus, the software developers will be challenged to obtain a higher score than the one initially obtained through the usage of the QAA module available through the SQAaaS portal. On each iteration the QAA report will provide clues to improve the overall quality rating, so that the Hackathon participant can focus on resolving those and execute the quality assessment process as many times as needed during the time specified by the Hackathon organizers. The progress (difference between final and initial criteria) obtained during the Hackathon will be best rated, followed by the highest overall badge level.

[1] <https://docs.sqaas.eosc-synergy.eu>

IBERGRID Contributions / 27**Quality Assurance Models in the framework of EOSC-Synergy**

Authors: Isabel Campos¹; Jorge Gomes²; João Pina²; Mário David²; Pablo Orviz Fernández³; Samuel Bernardo²

¹ CSIC

² LIP

³ IFCA-CSIC

Corresponding Author: david@lip.pt

This presentation gives an overview of Software and Service Quality Assurance Models as developed throughout several EU projects, and in particular EOSC-Synergy.

Quality Assurance is deemed as a way to more reliable and sustainable Software and Services.

The Quality Models outlined here are the “A set of Common Software Quality Assurance Baseline Criteria for Research Projects” (<http://hdl.handle.net/10261/160086>), and “A set of Common Service Quality Assurance Baseline Criteria for Research Projects” (<http://hdl.handle.net/10261/214441>).

The models are abstract but geared towards implementation favouring a pragmatic and systematic approach, putting emphasis on the programmatic assessment of the quality conventions.

As such, the models builds on the DevOps culture and automation, outlining a set of good practices that seek the usability and reliability of Software and Services, and meet the user expectations in terms of functional requirements.

- Quality assessment allows users and managers to have higher trust on Software and Services
- The software and related services will work as supposed.
- Give the expected results and meet their requirements.
- Furthermore, it also contributes to the maintainability, stability and sustainability of the
- Finally, it contributes to facilitating the collaboration between software developers and pr

The models are also being developed in an open way and are being exploited in several projects, in particular they constitute the basis for the Software Quality Assurance as a Service (SQaaS) being developed by EOSC-synergy, which aims to streamlines the adherence to quality best practices and make quality assurance easily accessible to software developers in research (<https://www.eosc-synergy.eu/technical-areas/quality/>).

The presentation will also highlight the software quality activities taking place in EOSC.

IBERGRID Contributions / 3

From the ”Proof of Concept” to EuroHPC: a user journey across the HPC landscape in Europe

Author: Isabel Campos¹

¹ CSIC

Corresponding Author: isabel.campos@csic.es

This presentation aims at describing the experience of a Lattice QCD medium-size research group to develop a research program involving access to large scale HPC systems in Europe: from local clusters, to national resources, and ending with EuroHPC scale.

The presentation intends to be generic enough from the application point of view, but Lattice QCD will be used as a guiding example. We will describe the level of resources required to develop a competitive simulation project from the proof of concept, development, benchmarking and scalability tests points of view. In doing so we will describe the main technical and policy challenges encountered, and discuss possible paths and ideas for improvement.

We will also describe the additional challenges posed by the fact that the research team is distributed across several countries in Europe. Those challenges go well beyond accessing CPU resources for simulation, spanning also to the data management such as distributed access and long term preservation.

IBERGRID Contributions / 6

Using Dynamic DNS for dynamic service deployment, migration and high availability

Author: Viet Tran¹

¹ *Institute of Informatics SAS Slovakia*

Corresponding Author: viet.tran@savba.sk

The Dynamic DNS service at IISAS provides a unified, federation-wide Dynamic DNS support for virtual machines in EGI Cloud infrastructure. Users can register their chosen meaningful and memorable DNS host names in given domains (e.g. my-server.vo.fedcloud.eu) and assign to public IPs of their servers.

The main features of Dynamic DNS service are as follows:

- **Simple usage:** Users should be able to register a hostname and assign it to a server (desktop, local server, virtual machine in Cloud) in two minutes. See the demonstration video in the fedcloud.eu YouTube channel for a proof.
- **Independent:** do not require additional software installation, nor additional support from Cloud providers or site administrators.
- **Universal:** hostnames could be applied for services/VMs deployed anywhere, including local servers, private Clouds, academic Clouds or commercial Clouds. That feature is very useful for service migration from local servers to Cloud or from a cloud provider to another.
- **User-friendly:** end-users can access services or virtual machines deployed in Clouds with pre-registered, meaningful and memorable hostnames instead of IP addresses.
- **Security and privacy:** with proper hostnames users are allowed to get valid SSL certificates for the services deployed in the Cloud.

One of the most interesting usages of Dynamic DNS service is service migration and high availability. The service enables easy migration of users' services from one Cloud site to another within one minute. That can be used also for designing high availability, where the service endpoint may be switched from the faulty service instance to the healthy one automatically and dynamically by monitoring.

The Dynamic DNS service is available at <https://nsupdate.fedcloud.eu/>.

IBERGRID Contributions / 4

FedCloud client: the powerful client for EGI Federated Cloud

Author: Viet Tran¹

¹ *Institute of Informatics SAS Slovakia*

Corresponding Author: viet.tran@savba.sk

The FedCloud client is a high-level Python package for a command-line client designed for interaction with the OpenStack services in the EGI infrastructure. The client can access various EGI services and can perform many tasks for users including managing access tokens, listing services, and mainly execute commands on OpenStack sites in EGI infrastructure.

The most notable features of FedCloud client are following:

- **Rich functionalities:** have wide ranges of useful commands, including checking access token, searching for services, listing sites and VOs, and interaction with OpenStack sites.

- **Simple usages:** can perform any OpenStack command on any sites with only three parameters: the site, the VO and the command.
- **Federation-wide:** Single client for all OpenStack sites and related services of EGI Cloud infrastructure. Single command may perform an action on all sites by specifying `-site ALL_SITES`.
- **Programmable:** the client is designed for using in scripts for automation or as a Python library for programming FedCloud services.

The full documentation of FedCloud client is available at <https://fedcloudclient.fedcloud.eu/>.

IBERGRID Contributions / 16

Compare computing sites using EOSC-Performance service

Authors: Borja Esteban Sanchis¹; Christophe Laures¹; Marcus Hardt¹; Valentin Kozlov¹

¹ *KIT*

Corresponding Author: borja.sanchis@kit.edu

EOSC-Performance is a platform available on the EOSC Marketplace to search and compare multiple computing sites, including those available to the EOSC community. Users can visually compare benchmarks results from a wide range of computing resources, covering cloud and HPC. Users and service providers can contribute to the platform by adding new benchmarks or uploading results for the computing resources of their interest. This upcoming autumn update includes an improved Web GUI with a number of new features: enhanced visual comparison of benchmark results, e.g. by allowing data regression analysis, data export, and a few more.

The EOSC-Performance service leverages OIDC and EGI-Check-In for authentication, a Dynamic DNS service from EGI Federated cloud, and applies the SQAaaS best practices from the EOSC-Synergy project. The service implies API-First approach to build the frontend, which is provided for users following OpenAPI version 3 specification.

The Service features, typical use cases, and highlights on the internal architecture will be presented in this contribution.

IBERGRID Contributions / 25

Integrated dataset placement service for scientists

Authors: Patrick Fuhrmann¹; Michael Schuh¹; Tim Wetzel²

Co-authors: Muhammad Aleem Sarwar¹; Thomas Beermann¹; Paul Millar¹; Johannes Reppin¹

¹ *DESY*

² *DESY IT (Research and Innovation in Scientific Computing)*

Corresponding Author: tim.wetzel@desy.de

Within the Helmholtz Association's federated IT platform HIFIS we have observed a trend towards more collaboration between different research groups and communities, both within Germany and the wider European research environment. These geographically distributed groups must manage their data so that members can access the data as needed. Meanwhile, funding agencies and scientific journals increasingly require that research data is FAIR: Findable, Accessible, Interoperable and Reusable. Of these features, Accessibility requires that computational activity can access the data; in particular, visualisation and other low-latency work may require access to data that is stored in a location separate from the compute infrastructure. In both cases, the data has to be transferred to

the computing cluster in question.

We furthermore observed that data sets tend to become very large to the extent that transferring it requires specialised tools and systems. In the talk we will outline a setup that is going to be implemented by Helmholtz Federated IT services HIFIS for the research centres of the Helmholtz Association.

The distribution of large data amounts among data centres has been practiced, e.g., in the high-energy physics community by CERN and the worldwide LHC computing grid (WLCG) for years now. More scientific communities have started to discover the need for and advantages of distributed computing. Efficient computing, however, depends on fast data access as to not waste computing time while waiting for slow I/O processes. In order to address easy sharing of data sets between different IT infrastructures, we will propose and present a data analysis setup suited for data exploration and smaller analyses. The setup comprises a Jupyterhub instance with a locally available storage element for user access. The mounted storage element itself is integrated with a mesh of other storage elements at different sites between which data can be transferred.

The mesh integration is achieved by using a combination of Rucio and FTS3, both developed at CERN for the needs of the WLCG. Rucio is integrated with the Jupyterhub instance by means of a plugin which enables searching for data sets and applying replication rules to make them available locally. Transfers started this way are executed by FTS3 asynchronously and within a reasonable time frame such that scientists are able to analyse the data without too much delay or the need to plan ahead. In the Jupyter notebook, all data sets transferred in this manner are available from variables pointing transparently to the corresponding location on the local storage system.

We argue that providing a network of storage elements for storing, replicating and transferring data between sites is a feasible and necessary way enabling researchers to share their data among each other and with the general public. There are many storage and data transfer solutions available commercially but we feel that using a system that has been developed by scientists for scientists, with full control over the data and its availability, should be the sensible option for use in the scientific community.

IBERGRID Contributions / 21

Water Monitoring Sentinel Cloud platform (WORSICA): a new service for detecting coastline, coastal inundation areas, and inland water bodies using remote sensing.

Authors: Alberto Azevedo¹; Ricardo Martins¹; André Bustorff Fortunato¹; Anabela Oliveira¹; Samuel Bernardo²

¹ *Laboratório Nacional de Engenharia Civil*

² *Laboratório de Instrumentação e Física Experimental de Partículas*

Corresponding Author: aazevedo@lnec.pt

Coastline monitoring is a significantly demanded task with substantial costs and an inherent need for computational power to process the imagery data. Therefore, LNEC has developed a new service to work with water indexes to determine water extensions in the areas of study.

Water mOnitoRing SentInel Cloud platform (WORSICA) is a new service that detects the coastline, coastal inundation areas, and the limits of inland water bodies using remote sensing. This service was developed under the scope of the H2020-EOSC-Synergy project and aimed at integrating multiple-source remote sensing and in-situ data to determine the presence of water in coastal and inland areas. It applies to various purposes, from determining flooded areas (caused by rainfall, storms, hurricanes, or tsunamis) to detecting large water leaks in major water distribution networks. It builds on components developed in national and European projects, integrated to provide a one-stop-shop service for remote sensing information, integrating data from the Copernicus satellite and drone/unmanned aerial vehicles, validated by existing online in-situ data. In addition, the WORSICA service was integrated into the European Open Science Cloud (EOSC) infrastructure and available to all European public research groups. The users can access the service through a web portal with a simple browser without additional computational costs, which are frequently needed for image satellite processing.

The present publication will present an application to the Óbidos lagoon to demonstrate some service features. This application will monitor the inlet of the Óbidos lagoon and provide an estimate of the

inter-tidal bathymetry of the area based on the sentinel-2 images tidal and wave modeling data for a specific seasonal period.

IBERGRID Contributions / 1

Deep learning to estimate wine volume from one single-view image

Author: Miriam Cobo Cano¹

Co-authors: Daniel Garcia Diaz ²; Ignacio Heredia ³; Fernando Aguilar Gómez ⁴; Lara Lloret Iglesias ⁵; Begoña Bartolomé ⁶; M. Victoria Moreno-Arribas ⁶; Carmen Patricia Pérez-Matute ⁷; Silvia Yuste ⁸; Maria-José Motilva ⁸

¹ *Institute of Physics of Cantabria (IFCA) - Spanish National Research Council (CSIC)*

² *Instituto de Física de Cantabria (IFCA)*

³ *Institute of Physics of Cantabria (IFCA), CSIC-UC*

⁴ *IFCA*

⁵ *CSIC*

⁶ *Institute of Food Science Research (CIAL), CSIC*

⁷ *Center for Biomedical Research of La Rioja (CIBIR)*

⁸ *Institute of Grapevine and Wine Sciences (ICVV), CSIC*

Corresponding Author: cobocano@ifca.unican.es

The purpose of this work is to present a deep learning method to determine the volume of red wine in different types of glass liquid containers from a single-view image, with potential real life applications in diet monitoring and wine consumption studies.

Diet assessment is mainly based on Food Frequency Questionnaires (FFQs) that inquire individuals about the frequency at which they consume different food items from a predefined list. In the case of wine, estimation of its consumption through FFQs is particularly imprecise as a portion (glass of wine) consumed is standardized to 100 cc, which is not always real.

This work uses convolutional neural networks in a regression model, which efficiently estimates wine volume in almost any kind of wine glass container. This study aims to provide an automated tool for red wine volume estimation based on the proposed model, in a simple and efficient way that only requires the subject to take a photograph of the glass of wine with a mobile phone, instead of having to carry a beaker or any other instrument to perform the measurement.

To train and evaluate our system, we introduced the WineGut_BrainUp dataset 1, a new publicly available dataset of glasses of wine that contains 24305 laboratory images, including a wide range of containers, volumes of wine, backgrounds, object distances, angles and lightning, with or without calibration object. Laboratory photographs were taken indoors and outdoors at the Institute of Food Science Research (CIAL-CSIC), the Institute of Grapevine and Wine Sciences (ICVV-CSIC) and the Center for Biomedical Research of La Rioja (CIBIR).

The model was trained with a GPU Tesla V100-PCIE-32GB for 20 hrs. The model was coded using Python, Keras and TensorFlow in Ubuntu 18.04.2 LTS.

The algorithm does not require any reference object in the image, showing that in order to solve the liquid volume estimation challenge it is not needed to include a calibration object in the photograph. In contrast, this presumed deficiency can be overcome with a larger training dataset including enough photographs of all representative situations that could occur, so that the model is able to recognize the shape and size of the glass recipient containing the liquid. Our algorithm performs this task efficiently, as can be observed in the saliency maps evaluation.

Experimental results demonstrated satisfactory performance of our image-based wine measurement system, with a Mean Absolute Error of 8 mL, and a Root Mean Square Error of 11 mL, both in the test

set. Our method outperforms similar systems that were developed in the literature for related tasks. Overall, this modeling will facilitate accurate measurement of liquid volume in diet and consumption studies.

References:

1 Bartolomé, B., Moreno-Arribas, M.V., Lloret, L., Aguilar, F., Cobo, M., García, D., Heredia, I., Yuste, S., Pérez-Matute, P., Motilva, M.J. (2021) "BrainGut_WineUp laboratory images [Dataset]", DIGITAL.CSIC, <https://doi.org/10.20350/digitalCSIC/14135>

IBERGRID Contributions / 9

Making Quantum Technologies more accessible: A new approach to bring the future closer

Author: Sergio Martínez-Losa¹

Co-author: David Iñiguez Dieste²

¹ *Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza*

² *Instituto de Biocomputación y Física de Sistemas Complejos - Fundacion ARAID*

Corresponding Author: sergio.martinez@bifi.es

For the first time, Quantum Technologies and Quantum Physics are available at an understanding level that anyone can use and interpret. In the early years of the 20th century, Quantum Physics theory became more and more complicated as well as refreshing. This knowledge was only available to theoreticians and people with high-level studies, but now, it is possible to figure out how complex systems work in the nanoscale whilst, for instance, we simulate a molecule in research for a new medical drug using a Quantum Computer simulator. Quantum Technologies are now used as a tool to understand the inside Nature's processes, this fact releases new and powerful information that was hidden from us until now.

From the last years of the 20th century until now, superconductive materials based on CMOS technology, took the Quantum scene to make circuits that hold the fundamental information unit, the Qubit. It is now when Quantum Computers based on this technology begin to be available to everybody, but nevertheless, there are drawbacks that we must deal with, like the stable lifetime of the Qubits, called decoherence time. Many other technologies hold Qubits in hardware like photons, single electrons, superconductive rings, magnetic molecules, or superconductive resonators. Although there are several alternatives, we still do not have a winning technology.

We present the whole spectrum of Quantum Technologies available and ready to use right now for many knowledge fields, and how the community can take advantage of this new paradigm. There are more to come and research in areas like Quantum Computation, Quantum Machine Learning, Quantum Chemistry, Quantum Biology, and Quantum Finance. Many companies claim a Quantum Leap achievement on focused issues, but, it is still needed a complete citizen ecosystem that brings the Quantum advantages to the general audience. Thus, along with Quantum Technologies, we present the Quantum Spain initiative, which aims to bring a public use of the Quantum Computation power to every citizen, research group and company in Spain.

IBERGRID Contributions / 8

Coastal Digital Twins: building knowledge through numerical models and IT tools

Author: Anabela Oliveira¹

¹ LNEC

Corresponding Author: aoliveira@lnec.pt

Digital Twin integrates continuously, in an interactive, two-way data connection, the real and the virtual assets. They provide a virtual representation of a physical asset enabled through data and models and can be used for multiple applications such as real-time forecast, system optimization, monitoring and controlling, and support enhanced decision making. Recent tools take advantage of the huge online volume of data streams provided by satellites, IoT sensing and many real time surveillance platforms, and the availability of powerful computational resources that made process-solving, high resolution models or AI-based models possible, to build high accuracy replicas of the real world.

In this paper, the adaptation of the concept of Digital Twins is extended from the ocean to the coastal zones, handling the high non-linear physics and the complexity of monitoring these regions. The work extends from the on-demand coastal forecast framework OPENCoastS (Oliveira et al., 2020; Oliveira et al., 2021) to build a user-centered platform where multiple services, from early-warning tools to collaborative platforms, are proposed and customized to meet the users needs. As the computation effort for these services is high, integration of Coastal Digital Twins in federated computational infrastructures, such as European Open Science Cloud (EOSC) or INC4 in Portugal, is fundamental to guarantee the capacity to serve multiple users simultaneously.

References

A. Oliveira, A.B. Fortunato, M. Rodrigues, A. Azevedo, J. Rogeiro, S. Bernardo, L. Lavaud, X. Bertin, A. Nahon, G. Jesus, M. Rocha, P. Lopes, 2021. Forecasting contrasting coastal and estuarine hydrodynamics with OPENCoastS, *Environmental Modelling & Software*, Volume 143,105132, ISSN 1364-8152, <https://doi.org/10.1016/j.envsoft.2021.105132>.

A. Oliveira, A.B. Fortunato, J. Rogeiro, J. Teixeira, A. Azevedo, L. Lavaud, X. Bertin, J. Gomes, M. David, J. Pina, M. Rodrigues, P. Lopes, 2019. OPENCoastS: An open-access service for the automatic generation of coastal forecast systems, *Environmental Modelling & Software*, Volume 124, 104585, ISSN 1364-8152, <https://doi.org/10.1016/j.envsoft.2019.104585>.

IBERGRID Contributions / 7

FAIRness Evaluation, Validation and Advising with FAIR EVA

Authors: Fernando Aguilar Gómez¹; Isabel Bernal²

¹ IFCA

² URICI-CSIC

Corresponding Author: aguilarf@ifca.unican.es

Within the context of EOSC-Synergy project, which aims at expanding the capacity of the European Open Science Cloud in terms of digital infrastructures as well as promoting software and data quality, FAIR EVA (Evaluator, Validator & Advisor) has been developed. It is oriented to be scalable and customizable to adapt to different repository or data systems. This tool connects to the data service to get data and metadata information and it performs different technical test to evaluate the FAIRness status of a research product (mainly data, but it can check any available digital object).

Although it includes a generic implementation, the connection to external services as well as the list of tests to be performed can be customized developing or extending existing plugins. This make FAIR EVA flexible and ready to work within different research communities. The modular architecture and the dockerized mode facilitate scaling.

This demo will show how to deploy FAIR EVA, how it works and how it can be connected to any data system.

IBERGRID Contributions / 10

Developing a Distributed and Fault Tolerant Dataverse Architecture

Author: Zacarias Benta¹

Co-authors: César Ferreira²; Jorge Gomes³; Samuel Bernardo³; Mário David³; João Pina³; Miguel Viana³

¹ *LIP - Minho*

² *INCD*

³ *LIP*

Corresponding Author: zacarias@lip.pt

Dataverse is an open source data repository solution with increased adoption by research organizations and user communities for data sharing and preservation. Datasets stored in Dataverse are cataloged, described with metadata, and can be easily shared and downloaded. However, despite all its features, Dataverse is still missing an architecture that ensures a distributed, fault tolerant, highly available and out-of-the-box service deployment.

In this presentation we will report the efforts by the Portuguese Distributed Computing Infrastructure (INCD) to address these current limitations, by creating a dataverse deployment architecture that is easy to set-up, portable, highly available and fault tolerant.

We tackled this objective, following a DevOps approach, resorting to a wide range of open software tools such as Linux containers, source code repositories, CI/CD pipelines, keepalived in conjunction with Virtual IPs (VIPs), `pg_auto_failover` for database replication and high availability object storage as scalable data storage backend. Our solution was implemented on top of the Openstack cloud management framework, the authentication is performed by using single sign on provided by several IdPs.

This architecture, is therefore capable of providing a stable and fault tolerant Dataverse installation, while keeping a flexible enough set-up to allow for the expansion of the storage and facilitate the upgrade to new versions.

The deployment architecture is currently under testing and will be used to support a catchall data repository for the Portuguese research and academic community.

IBERGRID Contributions / 11

CSCALE OpenEO Deployment

Author: Zacarias Benta¹

Co-authors: César Ferreira²; Samuel Bernardo³; Mário David³; Jorge Gomes⁴; João Pina³; Miguel Viana³

¹ *LIP - Minho*

² *INCD*

³ *LIP*

⁴ *Laboratorio Fisica Experimental de Partículas (LIP), Portugal*

Corresponding Author: zacarias@lip.pt

The Openeo project, is a free and opensource tool to analyze and process satellite imagery. Openeo provides an API that can be used by the likes of R, python and JavaScript for interaction with the backend. To support this software the INCD infrastructure was prepared to both store and compute the data it will be fed in the context of the C-Scale project.

INCD has deployed a Kubernetes cluster on top of an OpenStack cluster to support the project use cases. The Kubernetes cluster is composed of 5 nodes with 16 cores and 32 GB of RAM. In the first approach a PVC of 12TB was created for the storage of data from the Aqua Monitor use case, this was later changed to a S3 bucket mounted on every single compute node, allowing for the sharing of data both within the cluster and the outside world. A STAC Server was also deployed

allowing the download of satellite imagery products directly to a S3 Bucket which is used by the openEO pod that runs in the cluster. Code in Python was developed to register the metadata of the downloaded products into the resto catalog stored at CESNET. In this way all the data that is stored in the S3 bucket at INCD can be accessible, findable and reusable by searching on the resto catalog at CESNET. The users that interact with this deployment at INCD do so by using a jupyter notebook that connects to the endpoint, collects all the data needed for the computation and submits a job to a spark endpoint that is available within our infrastructure. With this presentation, we aim to show the INCD OpenEO backend architecture, implementation and potential for the development of further EO products.

IBERGRID Contributions / 22

Helmholtz Federated IT Services: Enabling innovative software and supporting services for research

Author: Thomas Förster¹

Co-author: Tobias Huste¹

¹ *HZDR*

Corresponding Author: thomas.foerster@hzdr.de

The digital transformation yields one of the biggest challenges for science, economy, and society at the beginning of the 21st century. The Helmholtz association recognizes the need to address those challenges, especially regarding the value and knowledge chains “from data to knowledge til innovation.” This touches most scientific disciplines and dramatically influences the development of our society. With HIFIS, the Helmholtz Association is pursuing the goal of bringing together the enormous data treasure and the diverse, decentralized expertise of the 18 Helmholtz research centers. The talk focuses on the lessons learned and the impact of integrating RSE services, educational resources, and consulting offers since 2020 on the research community. We discuss the structure of the HIFIS platform and the service needed to provide a good and fruitful environment for working on software-related projects in a collaborative context.

IBERGRID Contributions / 32

SSH with Federated Identities

Authors: Diana Gudu¹; Marcus Hardt¹; Gabriel Zachmann¹

¹ *KIT*

Corresponding Author: hardt@kit.edu

Federated Identities are the basis on which users are identified on scientific infrastructures. Modern Authentication and Authorisation Infrastructures (AAIs) make use of Federated Identities to provide services to users based on the identity provided by their home-organisation and their membership in the community, expressed by Virtual Organisation Membership.

We present a solution to use federated identities with any modern AAI (eduTEAMS, EGI-CheckIn, B2Access, Indigo-IAM, Google, ...) to access remote resources over SSH, such as HPC machines, or VMs in the Cloud.

Our solution uses unmodified ssh client and server components, by providing a new PAM module, and an additional service on the server side for enforcing authorisation and managing of local accounts. The client side is supported with operating system dependent solutions. Supported are Linux (client and server), Windows and MacOS.

IBERGRID Contributions / 20

jUMP: Construction of a oceanic sound propagation modelling portal for research, incorporating IT services from EOSC.

Authors: Alberto Azevedo¹; Anabela Oliveira¹; Ricardo Martins¹

¹ *Laboratório Nacional de Engenharia Civil*

Corresponding Author: aazevedo@lnec.pt

The United Nations has already acknowledged anthropogenic noise's harm to marine wildlife. Therefore, the correct control and reduction of underwater noise are relevant to the intended sustainable management and protection of marine and coastal ecosystems, especially in light of the 14th development goal (conservation and sustainable use of the oceans and marine resources). Creating healthy and productive oceans is feasible by avoiding substantial detrimental consequences, enhancing their resilience, and taking steps for their restoration.

Due to maritime transportation, leisure and tourism, fishing operations, and running industrial facilities, the demand on the Portuguese coast is escalating. One harmful byproduct of these activities is underwater noise, which adversely affects species sensitive to noise and the ecosystems that support them. LNEC has created a modeling portal to simulate sound propagation in the ocean and support the monitoring activities along the Portuguese Exclusive Economic Zone to address this theme and the project "jUMP - Joint Action: A Stepping-stone for underwater noise monitoring in Portuguese waters" (EEZ).

In the current publication, the authors describe a web portal that enables users to configure sound propagation simulations on-demand, with particular configurations such as the source and receptor positions, frequency, and depth of the sound source, besides the use of European oceanographic services such as Copernicus and EMODnet, for the retrieve of oceanic stratification and bathymetry data. This study also presents some of the technologies used to construct the service. The service integrates numerous technologies and services from the European Open Science Cloud (EOSC) and will be provided without charge to the research community (e.g., Federated authentication, Workload managers, Infrastructure Managers, and computational resources). The platform, in the authors' opinion, can further study the domain of underwater sound transmission in our seas.

IBERGRID Contributions / 19

Serverless Services for Scientific Cloud Computing

Authors: Sebastián Risco¹; Vicent Giménez²; Miguel Caballer²; Germán Moltó¹

¹ *Universitat Politècnica de València*

² *UPVLC*

Corresponding Author: serisgal@i3m.upv.es

Serverless is an application development paradigm that uses Cloud-based managed services for automated resource allocation. Functions as a Service (FaaS) is a computing model that executes functions in response to events in the infrastructure to perform highly elastic computations. These are being adopted for scientific computations, to hide the operational procedures of resource allocations in the Cloud.

This contribution focuses on the development of innovative serverless-based open-source software developments carried out by the GRyCAP/I3M research group at the Universitat Politècnica de València (Spain).

First, SCAR (<https://github.com/grycap/scar>), which supports Docker containers in AWS Lambda years before this functionality was natively introduced in AWS Lambda. SCAR is integrated with API

Gateway and AWS Batch to support event-driven execution of Docker-based computing applications that can execute in a hybrid approach both using a FaaS approach, in AWS Lambda, and via auto-scaled computing clusters, even with GPU support, in AWS Batch.

Second, OSCAR (<https://github.com/grycap/oscar>) provides a platform for event-driven data processing on auto-scaled Kubernetes clusters, which can be dynamically deployed across multiple Clouds using the Infrastructure Manager (<https://www.grycap.upv.es/im>). It supports both asynchronous executions via Kubernetes jobs and synchronous ones via Knative to fit different use case requirements, such as performing the inference of pre-trained AI/ML models. OSCAR can be executed in low-power devices such as Raspberry Pis, useful for edge computing, and shares the same Functions Definition Language (FDL) as SCAR and, therefore, it allows to create event-driven workflows that can span along the computing continuum.

Third, MARLA (<https://github.com/grycap/marla>) provides a completely serverless environment for MapReduce executions on top of AWS Lambda environments. MARLA uses the S3 service to receive data files to be processed via an event-driven workflow. Therefore, all the data processing is done automatically from data loading to obtaining the final result. In addition, the user can define their specific map and reduce functions thanks to the builtin python environment included in the AWS Lambda service, allowing any type of analysis on the data.

Fourth, TaSaaS (<https://github.com/grycap/TaSaaS>) is a serverless service to distribute and balance work executions among independent infrastructures. TaSaaS is focused on scientific executions with huge computational cost, which usually must be split into hundreds or thousands of smaller partitions to be computed in parallel on different nodes and computational infrastructures. With this aim, TaSaaS divides automatically the works in partitions to be executed on the available resources. It also balances the load of all partitions belonging to the same job and monitors their status to detect node failures. TaSaaS is deployed on AWS but its architecture can be extended to other cloud providers.

In summary, a range of serverless-based innovative software developments have been created which aim to simplify the adoption of this novel computational paradigm for scientific computing.

Grant PID2020-113126RB-I00 funded by MCIN/AEI/10.13039/501100011033. Project PDC2021-120844-I00 funded by MCIN/AEI/10.13039/501100011033 funded by the European Union NextGenerationEU/PRTR

IBERGRID Contributions / 26

EOSC-Synergy Jenkins Pipeline Library

Author: Samuel Bernardo¹

Co-authors: Pablo Orviz Fernández²; Diana María Naranjo Delgado³; Mário David¹; Jorge Gomes⁴; Isabel Campos⁵; Miguel Caballer⁶

¹ LIP

² IFCA-CSIC

³ UPV

⁴ Laboratório Física Experimental de Partículas (LIP), Portugal

⁵ CSIC

⁶ UPVLC

Corresponding Author: samuel@lip.pt

The Jenkins Pipeline Library (JePL) is one of the core components of the EOSC-Synergy software and services quality assurance as a service platform (SQaaS) aimed to foster the adoption of EOSC services through a quality based approach. It is a self contained component that facilitates the process of creation and execution of CI/CD pipelines.

The library facilitates the creation of Jenkins pipelines by using a YAML description to define the several stages that compose a CI/CD pipeline. The actions in the YAML configuration file are aligned with the criteria compiled in the software and service quality baselines 1, and rely on Docker Compose to orchestrate the required set of services needed during the quality assessment process. A minimal (single-stage) Jenkins CI/CD pipeline definition (Jenkinsfile) is required to dynamically compose the required set of stages defined as actions in the YAML description. This means that the use of this library does not limit the researcher to the criteria defined in the baselines, but the Jenkins Pipeline as Code structure is preserved, and thus, additional stages can be added directly in the pipeline. Once this file layout is placed in the application's source code repository, the pipelines will be automatically constructed and executed through a Jenkins CI/CD.. This approach lowers the barriers that hinder the adoption of quality-based and sustainable software and service developments in research environments, enabling automation and agility.

In the context of the EOSC-Synergy SQAaaS platform, the library is used to enable the on-demand dynamic composition of Jenkins pipelines that perform the several steps of the required environments deployment followed by the quality assurance. These steps will implement the quality validation actions defined in the EOSC-synergy software and services quality criteria. Currently the library supports Infrastructure Manager (IM) and Kubernetes (K8s) for the infrastructure and services deployments. JePL v3 is already being developed to adopt IM and K8s as configuration as code, providing an improved experience to work with infrastructures and services.

The presentation will highlight the features and capabilities of the library in practice, showing how to easily create pipelines that implement and comply with the good practices that are expected during the software lifecycle, from development to production. This is particularly relevant to developers and managers of research services both at the infrastructure and thematic levels.

1 <http://hdl.handle.net/10261/160086>

2 <https://digital.csic.es/handle/10261/214441>

IBERGRID Contributions / 24

Automating cloud deployment infrastructures with Kolla

Authors: Alejandro Feijoo Fraga¹; Javier Cacheiro¹; Carlos Fernandez Sanchez¹; Juan Villasuso¹; Ruben Diez Lazaro¹

¹ CESGA

Corresponding Author: rdiez@cesga.es

As the complexity of cloud infrastructures increases, better mechanisms for easy maintenance and deployment are needed.

Containers offer great potential in this area, promising to simplify deployment and maintenance tasks.

Considering the OpenStack ecosystem, which is the most widely used cloud solution in EGI/fedcloud, there are several projects focused in a modular deployment of OpenStack services based on containers.

One of the most promising is kolla, which uses ansible and docker containers to address the OpenStack services management and deployment.

In this work, the deployment of OpenStack using kolla at CESGA is presented. Integration in the EGI/fedcloud infrastructure is examined and technical challenges and the proposed solutions are shown.

The end result is an Infrastructure as a Service (IaaS) integrated in EGI/fedcloud easy to operate and maintain

IBERGRID Contributions / 12

The Software Quality as a Service (SQaaS) portal

Authors: Pablo Orviz Fernández¹; Diana María Naranjo Delgado²; Jorge Gomes³; Samuel Bernardo⁴

¹ IFCA-CSIC

² UPV

³ Laboratorio Física Experimental de Partículas (LIP), Portugal

⁴ LIP

Corresponding Author: orviz@ifca.unican.es

The EOSC-Synergy project features the Software Quality Assurance as a Service (SQaaS) platform that provides a toolset to bring over software-development-related mainstream practices close to researchers dealing with source code and IT services. In the present work, we will cover the SQaaS portal, which aims at fulfilling a twofold objective. On the one hand it streamlines the adoption of good practices in software and service development, in the scope of the EOSC, through the composition of continuous integration and deployment (CI/CD) workflows or pipelines. On the other hand, the SQaaS portal provides a software-quality assessment tool to promote, measure and reward quality.

The SQaaS portal encompasses both the SQaaS web and API components, and as such, its main contribution is to provide a graphical interface to realize the two aforementioned objectives. As such, the SQaaS portal streamlines both the adoption of the good software engineering practices through the Pipeline as a Service module, and gives credit to software and service developers in research that are applying those practices into their regular development life cycle through the Quality Assessment and Awarding module.

The outcome of the Pipeline as a Service module is a CI/CD pipeline in the format provided by the JePL library, yet another component of the SQaaS solution. This CI/CD pipeline is then meant to be added to the target code repository so that such code is regularly checked on each change, and thus, enforcing the good quality practices.

The outcome of the Quality Assessment and Awarding module is a report that highlights the successful achievements, according to the aforementioned good practices, and the specific areas of the software project where quality attributes could be enhanced in order to obtain a higher score. Furthermore, in case of achieving a minimum set of quality practices or criteria, the module will issue a shareable and verifiable digital badge that embeds metadata that contains all the associated data generated during the quality assessment process. The EOSC-Synergy badges provide three levels of quality: gold, silver and bronze.

The latest release of the SQaaS portal includes support for software and service deployment criteria, with the capability of issuing digital badges up to silver for software and bronze for services.

IBERGRID Contributions / 30

Parallel FaaS on the Edge-Cloud continuum

Authors: Daniele Lezzi¹; Francesc Lordan Gomis¹

¹ Barcelona Supercomputing Center

Corresponding Author: daniele.lezzi@bsc.es

The design and execution of AI applications on Edge-Cloud infrastructures require a set of programming tools to seamlessly design and partition the models and a runtime that properly assigns the different components on the available nodes, also distributing the data.

The design abstractions allow to provide high-level annotations to specify QoS constraints and code dependencies and to introduce performance parameters for the allocation of tasks to computing continuum resources and security and privacy annotations for data allocation and processing.

The runtime leverages the embedded computing resources of each node to host the execution of functions in a service manner and generates hybrid workflows, composed of atomic and continuous processing tasks, achieving distribution, parallelism and heterogeneity across edge/cloud resources transparently to the application developer.

The adoption of such a distributed model for executing the applications allows the user to concentrate on the application development and rely on the infrastructure management by the serverless platform.

The programming framework runtime parallelizes the execution of the different parts of the applications that can be invoked in a FaaS way according to the QoS constraints. The runtime is able to schedule the tasks on both edge and cloud devices, orchestrating the execution and leveraging on fault tolerance mechanisms to react to the dynamicity of the edge.

The aim of the AI-SPRINT “Artificial intelligence in Secure PRIVacy-preserving computing coNTinuum” project is to develop a platform composed of design and runtime management tools to seamlessly design, partition and operate Artificial

Intelligence (AI) applications on the Edge-Cloud continuum, providing resource efficiency, performance, data privacy, and security guarantees.

In this presentation we will demonstrate the execution of a healthcare application that is built using the design time tools and FaaS components of AI-SPRINT, namely the PyCOMPSs programming framework and the OSCAR event-driven serverless applications manager.

IBERGRID Contributions / 29

User space containers with udocker

Authors: Jorge Gomes¹; Mário David¹

¹ LIP

Corresponding Author: jorge@lip.pt

udocker enables the execution of docker containers in user space without requiring root privileges for installation or use. Udocker implements the pull, load, import and execution of containers by non-privileged users in Linux systems where docker is not available. It can be used in Linux batch systems and interactive clusters that are managed by other entities, such as grid infrastructures or externally managed batch or interactive systems. udocker does not require any type of privileges nor the deployment of services by system administrators. It can be downloaded and executed entirely by the end user.

udocker offers several execution modes exploiting system call interception, library call interception and namespaces. udocker integrates several tools to provide a subset of the docker capabilities aimed at container execution.

This presentation we will explain the udocker fundamental, how to use udocker to execute Linux containers and how to best exploit the several execution engines.

IBERGRID Contributions / 31

EPIC Cloud: a secure, GDPR-compliant, open-source cloud platform for life-science applications

Authors: Barbara Martelli¹; Davide Salomoni²

Co-authors: Vincenzo Ciaschini²; Cristina Duma²; Giusy Sergi²; Alessandro Costantini²; Diego Michelotto²; Enrico Fattibene²; Cristina Vistoli²; Andrea Chierici²; Stefano Zani²; Lorenzo Chiarelli²

¹ *INFN - CNAF*

² *INFN CNAF*

Corresponding Author: barbara.martelli@cnaif.infn.it

After the entry into force of GDPR in 2018, life-science research communities have had to deal with a brand-new regulatory framework, very restrictive in terms of cybersecurity and privacy measures to be implemented in case personal data (even if pseudonymized) are involved in the studies. In Italy, we have dealt with this new legal scenario creating EPIC Cloud (Enhanced PrIvacy and Compliance Cloud): a region of INFN Cloud with particular security measures in place. An important aspect this work is the construction of an Information Security Management System (ISMS) and its certification of conformance with the international standards ISO/IEC 27001 27017 27018. The adoption of an ISMS is important from the organizational perspective, in fact security measures must be integrated and coordinated to be effective. The certification of conformance is a mechanism which involves an independent third party (an accredited consulting firm) who verifies the status of the ISMS yearly. This is of paramount importance to guarantee users that the security measures are appropriate and actually applied. In this talk we'll describe the technical and organizational measures adopted in EPIC Cloud and will discuss pain points and achieved results of our journey towards GDPR compliance. Furthermore, we'll describe how real life-science use cases, focusing on genomic and clinica data analysis, are managed in EPIC Cloud. In particular, we'll describe Health Big Data (a ten-year project founded by the Italian Health Ministry), Harmony Alliance (an European founded by the IMI2 Joint Undertaking) and other NRRP related initiatives involving EPIC Cloud.

IBERGRID Contributions / 15

Ozone assessment service (O3as) in action

Authors: Peter Braesicke¹; Ugur Cayoglu¹; Borja Esteban Sanchis¹; Marcus Hardt¹; Tobias Kerzenmacher¹; Valentin Kozlov¹

¹ *KIT*

Corresponding Authors: valentin.kozlov@kit.edu, kerzenmacher@kit.edu

The assessment of ozone in the atmosphere is an essential task for Climate and Environment studies. The ozone assessment service (O3as) supports scientists in determining ozone trends and metrics indicating a recovery from ozone loss for different parts of the world. It also aims to aid scientists to prepare the quadrennial Scientific Assessments of Ozone Depletion. Moreover, it provides access to a comprehensive selection of ozone data for interested citizens.

The service uses a unified approach to analyse results from a large number of different chemistry-climate models, helps to harmonise the calculation of ozone trends efficiently and consistently, and produces publication-quality figures in a coherent and user-friendly way. It is one of the thematic services of the EOSC-Synergy project.

The service relies on multiple containerised components that are distributed across the cloud (orchestrated in Kubernetes) and HPC resources. It leverages the Large Scale Data Facility (LSDF) at KIT for storing the raw data. We incorporate software best practices and the Software Quality Assurance as a Service (SQaaS) approach of the EOSC-Synergy project which includes software and service testing, Continuous Integration and Continuous Delivery (CI/CD), assessment for findability, accessibility, interoperability, and reusability (FAIR) of a data repository, and service documentation.

In this contribution, we are going to present the attractiveness of the service for climate scientists and highlight the service's internal architecture and communication interface.

IBERGRID Contributions / 23

COASTNET: Integrating Earth Observation and in-situ data for coastal management

Authors: Ana Brito¹; Paola Castellanos¹; Bernardo Quintella¹; Luis da Costa¹; Paula Chainho¹; Susana França¹

¹ MARE – Marine and Environmental Sciences Centre, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisbon, Portugal.

Corresponding Author: pcoessa@fc.ul.pt

The Portuguese Coastal Monitoring Network (CoastNet) is a research infrastructure (RI) included in the Portuguese Roadmap of Research Infrastructures, designed to improve the understanding of Portuguese coastal ecosystems' functioning through the development of a remote coastal monitoring system. CoastNet aims at monitoring important coastal ecosystems of the Portuguese coast, through relevant chemical, physical and biological parameters, collected remotely in near real-time. This monitoring network is implemented through: 1) the development of an historical and near real-time satellite data centre; 2) the deployment of a set of sensors, to collect environmental and biological in-situ data, at three estuaries (Mondego, Tagus and Mira), which are almost immediately transmitted to our data centre; (3) acoustic receivers arrays deployed to track aquatic fauna movements; and 4) the compilation of an historical in-situ dataset of biological and environmental parameters. CoastNet provides a better understanding of the functioning of Portuguese coastal systems through the analysis of their temporal and spatial variability. A case study based on the Tagus estuary, covering the water continuum from the coastal zone to the inner parts of the estuary, will be presented. Gathering such scientific knowledge is expected to improve scientific production, promote socio-economic activities, such as fishing and aquaculture, as well as the conservation of endangered species. A variety of services will also be provided to a wide range of stakeholders. Additionally, the Portuguese Coastal Monitoring Network provides full and free access to the collected data sets to the scientific community, public administration and private companies, through the dedicated Geoportal, available at: <http://geoportal.coastnet.pt>.

IBERGRID Contributions / 28

Repositories for EGI

Authors: Carlos Manuel¹; Jorge Gomes¹; João Pina¹; Mário David¹; Pablo Orviz Fernández²; Samuel Bernardo¹

¹ LIP

² IFCA-CSIC

Corresponding Author: david@lip.pt

This presentation provides an overview of the architecture and implementation of the new artifacts repositories for EGI.

The EGI repositories are developed, maintained and operated by LIP and IFCA/CSIC. The new repositories will host RPMs for (RHEL and compatible distributions), DEBs (for Ubuntu and compatible distributions) and Docker images for container based services and micro-services.

The presentation will describe the architecture of the new repositories, its several components and capabilities. The architecture of the repositories is as follows:

1. "Web front-end service" containing static information about EGI UMD and CMD releases, a software catalog allowing listing and searching of all the products in the repositories, and information about how to configure the repositories. The front-end also includes proxies to a Sonatype Nexusrepo,
2. "Sonatype Nexusrepo" as the service which hosts all the artifacts (packages and docker container images),
3. "Backend service" where verification, validation and administration of a given release is performed.

The backend service is where all the actions and logic occurs in order to verify, validate and publish a release. The release manager creates a new release either manually or through submission/upload of a yaml file containing the information of a given product. He can edit the fields and verify its correctness, then he can commit the form with the yaml describing the release. Upon validation of the yaml file, the content is parsed into a json file and a series of scripts are automatically executed. The pipeline of actions are as follows in the case of RPM packages:

1. Parse json, get the list of files to download and produce list of filenames (packages).
2. Download the packages to a temporary directory.
3. RPM sign each package.
4. Verify signature of each package followed by the package verification
5. Upload each package to nexusrepo.
6. Download each package from nexusrepo.
7. Verify checksum and signature.
8. Upload json files to the front-end for publication.
9. Clean temporary directories.

IBERGRID Contributions / 17

FAIR data in the Photon and Neutron community

Authors: Juliane Marauska¹; Patrick Fuhrmann¹; Paul Millar¹

¹ DESY

Corresponding Author: patrick.fuhrmann@desy.de

In recent years, funding agencies and scientific journals have started requiring that research data is FAIR: Findable, Accessible, Interoperable and Reusable. This is not an arbitrary decision. It reflects the intrinsic value of research data if additional use were possible; for example, to support new research, as reference data, or as pedagogical material.

Although some general principals of FAIR data management adoption have been investigate, each research community is individual and has unique challenges when making their data FAIR. One important activity is to tailor general approaches to match that community while finding solutions for the community-specific aspects.

ExPaNDS is a 3½ year, EU-funded project to support the adoption of FAIR data at European Photon and Neutron (PaN) facilities. Together with PaNOSC (also EU-funded), ExPaNDS has established FAIR data as a common goal within the European PaN community. ExPaNDS has provided policies and procedures, developed software and supported software adoption, all while providing training and outreach. Such activity includes building portable analysis pipelines being used at facilities with common instrument technologies or at horizontal compute infrastructures, creating common terminology in the form of various ontologies (making data findable and reusable), writing data management policies (describing how to make data FAIR) and a self-assessment procedure (allowing progress to be marked), establishing best practice for Active DMPs and persistent identifiers (PIDs) for dataset. Together, this is laying the groundwork for a future data commons.

We will present the challenges faced by PaN facilities when adopting FAIR data along with the many achievements of ExPaNDS in tackling those challenges.

IBERGRID Contributions / 34

Automating scientific dataset management and processing using Onedata

Corresponding Author: svoboda@ics.muni.cz

Making experiment results FAIR is a known challenge. To cope with this task, we develop a system that provides an easy way to make data produced by specialized devices (such as cryo-em microscopes) available to the scientific community. We focused on making the system as easy as possible for data producers and also for users who use the datasets in their scientific computations. Our solution can be used to manage the storage of experimental data between several tiers of data storage. Beginning physical data storage of the experimental facilities where the data originates, national or scientific domain data storage services, and fast storage in computing facilities provided on both national and European levels.

The software is built on top of the global data access solution for science Onedata. Our software supports the whole process, from acquiring produced data from the device, setting up automatically all necessary Onedata parameters (access policy, metadata, ...), publishing the datasets, and archiving them in permanent storage. It implements varying policies of handling the data, e.g., expiration at the acquisition facility, archiving in multiple copies, and data publication after an embargo period. It can also export datasets to supported repositories or metadata to metadata catalogs. The life cycle of the data is defined in a YAML file which is attached to the dataset.

IBERGRID Contributions / 5

Secret management service for EGI infrastructure

Author: Viet Tran¹

¹ *Institute of Informatics SAS Slovakia*

Corresponding Author: viet.tran@savba.sk

Applications in EGI Infrastructure may need different secrets (credentials, tokens, passwords, etc.) during deployments and operations. The secrets are often stored as clear texts in configuration files or code repositories that expose security risks. Furthermore, the secrets stored in files are static and difficult to change/rotate. The secret management service for EGI Infrastructure is developed by IISAS in cooperation with INFN and CSIC to solve the issues.

The Secret management service is designed as follows:

- **Usability:** A dedicated module of FedCloud client is developed to make the service works out of the box with very simple syntax. Authentication is realized via OIDC tokens from EGI Check-in, no additional registration, no extra credentials are required.
- **Advanced features:** Built-in support for secret values from files, export/import secrets to/from files in YAML/JSON formats. Encrypting/decrypting secret values on the fly on the client side greatly improves security and trust of the service.
- **Compatibility:** The service is based on Hashicorp's Vault which is well-known in industry, with many client tools and libraries. Software for service and clients are open-sourced with strong support from communities.
- **High-availability:** Service instances are distributed on different sites, without single point of failure. A generic endpoint <https://vault.services.fedcloud.eu:8200> is dynamically assigned to a healthy instance via Dynamic DNS service.

The full documentation of the Secret management service is available at <https://vault.docs.fedcloud.eu/>.

IBERGRID Contributions / 18

GRAPEVINE project: hiGh peRformAnce comPuting sERVICES for preVentIon and coNtrol of pEsts in fruit crops

Authors: Carlos Fernandez Sanchez¹; Carlos González Muñoz²; Cecilia Grela Llerena³; Charalampos Paraskevas⁴; Dimitrios Moshou⁵; F. Javier Nieto^{None}; Francisco José Lacueva Pérez²; Francisco Landeira¹; Gorka Labata Lezaun²; Iñigo Zubizarreta Nafarrate²; Joaquín Balduque-Gil⁶; Juan J. Barriuso⁶; Konstantinos Kechagias⁴; Paraskevi Vourlioti⁷; Rafael del Hoyo Alonso²; Sergio García⁸; Stylianos Kotsopoulos^{None}; Theano Mamouka⁷; Xanthoula Eirini Pantazi⁴

¹ CESGA

² ITAINNOVA

³ Fundación Pública Galega Centro Tecnolóxico de Supercomputación de Galicia (CESGA)

⁴ AUTH

⁵ AUTH

⁶ University of Zaragoza

⁷ AgroApps

⁸ ATOS

Corresponding Author: cecilia@cesga.es

Mildew is a highly destructive disease of grapevines, it appears in all grape-growing areas of the world where there is spring and summer rainfall at temperatures above 10°C and highly affects the production.

GRAPEVINE (hiGh peRformAnce comPuting sERVICES for preVentIon and coNtrol of pEsts in fruit crops) project has the objective to improve the current processes to detect in advance the phenological grape state, the mildew and other grapevine diseases with the development of a predictive model based on Machine Learning (ML) and Deep Learning (DL) techniques. GRAPEVINE wants to improve, in the first place, the evaluation and control of mildew in wine cultivation to reduce the amount of fungicide, and the number of its treatments, to introduce sustainability criteria in agricultural production, offering higher quality agricultural products safer for consumers.

To feed these ML and DL models the use of weather forecast simulations is necessary. All these models must be deployed in a coordinated way in a daily operation to provide on time information to the farmers. The use of advanced computational and data processing services is critical for the success of the project.

We present the GRAPEVINE project, the models, the software to orchestrate all these services (including data management activities and monitoring) and how the use of Cloud compute services of the European Open Science Cloud in the EGI-ACE project have provided a platform for the development of this innovative service for the farmers. This paper shows how the proposed architecture had a positive impact on the usage of the computational resources and how users can benefit from advanced infrastructures with lower effort and required know-how

IBERGRID Contributions / 87

Linking GitOPS towards fast innovation over BigHPC

Corresponding Author: samuel@lip.pt

The BigHPC project is bringing together innovative solutions to improve the monitoring of heterogeneous HPC infrastructures and applications, the deployment of applications and the management of HPC computational and storage resources. It aims as well to alleviate the current storage performance bottleneck of HPC services. In order to keep all development tasks tracking in a common path, some good practices are needed to get a shorter development life cycle and provide continuous delivery and deployment with software quality. GitOPS is a way to implement the continuous deployment and software quality best practices. Deploy faster and more often, easy and fast error recovery, easier credential management, well documented deliveries with complete history of every change made to the system and share knowledge between teams with great commit messages. These are some examples of the advantages in adopting a GitOPS framework solution. As a result, everybody would be capable of reproducing the thought process of changing infrastructure and also easily find examples on how to set up new systems. In this work we are creating the git workflow being adopted for application development and the tools that we are joining together to answer the three components of GitOPS: infrastructure as code, merging changes together and deployment automation. We will show the technical capabilities and advantages of using this approach, keeping developers focused on the continuous development of the software and infrastructure management with the IT operations team, adopting good practices and pursuing fast innovation delivery.

IBERGRID Plenary Session / 33

PSNC role in digital transformation of the agriculture sector in Poland

Corresponding Author: marcinp@man.poznan.pl

Agriculture is currently undergoing a rapid digital transformation from mechanization to automation and the use of artificial intelligence or online decision support systems (e.g. in fertilization, integrated pest management). Agriculture 4.0 technologies are currently being introduced - smart farms based on a very high degree of automation and robotization. The concept of Agriculture 4.0 is based primarily on the analysis and processing of large amounts of data collected using the Internet of Things, drones, robots or satellite data, which are often used by artificial intelligence to make decisions. Therefore, an infrastructure that supplies data on an ongoing basis is indispensable. Also one of the key aspects with a large number of data sources is to ensure data interoperability. PSNC has been working for several years with key partners in agriculture to support digital transformation in agriculture in Poland. One of the key initiatives in this area is the eDwin project, which built a network of agrometeorological stations throughout Poland and prepared and made available to farmers an online decision support platform in integrated plant protection in the form of a Virtual Farm. The AI4EOSC project will use the eDWIN infrastructure to develop improved predictive models for plant protection based on AI models. In the Demeter project, the PSNC collaborated on the development of the AIM - Agriculture Information Model, and on the integration of agriculture and beekeeping. As part of the Agrobank PSNC initiative, it has developed a support system for selecting seed varieties. As part of the ICOS project, PSNC is working together to create a new generation of agricultural robots that fully communicate with the smartphonearms infrastructure. All these activities aim to create an integrated advisory platform and digital environment for the farmer.

IBERGRID Plenary Session / 35

IBERGRID status

Corresponding Author: isabel.campos@csic.es

IBERGRID was created under the agreement of Scientific and Technological cooperation initially

signed by Spain and Portugal in 2003. This presentation will provide an overview of the IBERGRID infrastructure status and related activities.

IBERGRID Plenary Session / 36

The Portuguese Distributed Computing Infrastructure (INCD)

Corresponding Author: jorge@lip.pt

The INCD is a digital infrastructure supported by the Portuguese roadmap of research infrastructures of strategic interest (RNIE). INCD aims at the establishment, continuous development and operation of a digital infrastructure, whose object is to provide advanced computing and data-oriented services to the national scientific and academic community. INCD provides computing and data oriented services to the national scientific and academic community in all areas of knowledge. The infrastructure is especially dedicated to provide scientific computing services, supporting researchers and their participation in national and international projects. The INCD services are federated in similar international infrastructures with which it shares computing resources for the benefit of projects of national and international relevance. INCD is the main Portuguese computing infrastructure in IBERGRID. This presentation will provide an overview of the INCD status and activities.

IBERGRID Plenary Session / 44

interTwin: An interdisciplinary Digital Twin Engine for science

Corresponding Author: malgorzata.krakowian@egi.eu

interTwin co-designs and implements the prototype of an interdisciplinary Digital Twin Engine (DTE), an open source platform that provides generic and tailored software components for modelling and simulation to integrate application-specific Digital Twins (DTs). Its specifications and implementation are based on a co-designed conceptual model - the DTE blueprint architecture - guided by the principles of open standards and interoperability. The ambition is to develop a common approach to the implementation of DTs that is applicable across the whole spectrum of scientific disciplines and beyond to facilitate developments and collaboration.

Co-design involves DT use cases for High energy physics, Radio astronomy, Astroparticle physics, Climate research, and

Environmental monitoring, whose complex requirements are expected to significantly advance the state of the art of modelling and simulation using heterogeneous distributed digital infrastructures, advanced workflow composition, real-time data management and processing, quality and uncertainty tracing of models, data fusion and analytics. As a result, a consolidation of software technologies supporting research will emerge.

The validation of the technology with multiple infrastructure facilities, will boost the accessibility of users to technological capacity and the support of AI uptake in research. interTwin builds on the capacities of experts from pan-European research infrastructures and the long tail of science, an open source community of technology providers that will deliver TRL 6/7 capabilities to implement the interdisciplinary DTE, experts of the European Centre of Excellence in Exascale Computing, and infrastructure providers from the EGI Federation, PRACE and EuroHPC supporting data and compute intensive science. interTwin key exploitable results will be continually co-developed and aligned with the contribution of external initiatives such as Destination Earth, EOSC, EuroGEO and EU data spaces.

IBERGRID Plenary Session / 54

Galicia Supercomputing Center (CESGA): Infrastructure, Research, and the Road Ahead

Corresponding Author: lorosa@cesga.es

Overview of new projects and activities at the Galicia Supercomputing Center (CESGA).

IBERGRID Plenary Session / 39

Security in the Spanish academic network

Corresponding Author: antonio.fuentes@rediris.es

Overview of security in the Spanish academic network.

IBERGRID Plenary Session / 43

Digital Twins for GEophysical extremes

Corresponding Author: afolch@geo3bcn.csic.es

With present computational capabilities and data volumes entering the Exascale Era, digital twins of the Earth system will be able to mimic the different system components (atmosphere, ocean, land, lithosphere) with unrivaled precision, providing analyses, forecasts, and what if scenarios for natural hazards and resources from their genesis phases and across their temporal and spatial scales.

IBERGRID Plenary Session / 53

Repositories for Cancer imaging

Corresponding Author: iblanque@dsic.upv.es

New activities in data repositories for cancer imaging.

IBERGRID Plenary Session / 55

University of Algarve

Corresponding Author: atmaia@ualg.pt

Compute intensive research in life sciences at the University of Algarve and related infrastructure support.

IBERGRID Plenary Session / 41

Scientific Computing at INFN and beyond

Corresponding Author: davide.salomoni@cnaif.infn.it

In the past couple years, INFN has been working toward a major restructuring of its organization for the provisioning and evolution of its computing services. This process overlapped with the opportunities brought forward by the COVID-19 National Recovery and Resilience Plan, which offered significant funding to build a national infrastructure for science and industry. INFN will have a leading role in these initiatives. As a result, in the next few months a complete revisiting of the INFN big data distributed infrastructure will be carried on. In this talk, I will highlight the results and prospects of these activities, describing what we have done so far and how we intend to continue to innovate and integrate multiple scientific use cases in a federated cloud data lake.

IBERGRID Plenary Session / 40

Cybersecurity at the Portuguese R&E Network

Corresponding Author: cfriacas@fccn.pt

RCTS CERT is the incident response team for the Portuguese Research & Education Network, operating since 2000. The talk will describe the current cybersecurity services available to RCTS members, the cooperation at different levels and also the most recently faced challenges.