PROCEEDINGS OF SCIENCE

Wine in the Cloud, or: Smart Vineyards with a Distributed "Extreme Data Database" and Supercomputing

Piyush Harsh,[a,*] **Stephan Hachinger,**[b] **Marc Derquennes,**[c] **Andy Edmonds,**[a] **Pinar Karagoz,**[d] **Martin Golasowski,**[e] **Mohamad Hayek**[b] **and Jan Martinovič**[e]

[a] *Terraview GmbH, Technoparkstrasse 2, 8406 Winterthur, Switzerland*
[b] *Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities (BAdW), Boltzmannstr. 1, 85748 Garching b.M., Germany*
[c] *EURAXENT – French Office, 1 Rue du Grand Bois, 43360 Saint-Géron, France*
[d] *Department of Computer Engineering, Middle East Technical University, 06800 Ankara, Turkey*
[e] *IT4Innovations National supercomputing Center, VŠB – Technical University of Ostrava, Studentská 6231/1B, 708 00 Ostrava-Poruba, Czech Republic*

E-mail: piyush@terraview.gmbh

In this contribution, we sketch an application of Earth System Sciences and Cloud-/Big-Data-based IT, which shall soon leverage European supercomputing facilities: smart viticulture, as put into practice by Terraview. TerraviewOS is a smart vineyard "operating system", allowing wine cultivators to optimise irrigation, harvesting dates and measures against plant diseases. The system relies on satellite and drone imagery as well as in-situ sensors where available. Clearly, processing behind the UI is heavily based on Cloud Computing, with some Edge Computing close to the user or High-Performance/GPU Computing components. The substantial need for computing power in TerraviewOS, in particular for training AI-based models to generate derived data products, makes the further development of some of its modules a prime use case for the EU-funded Extreme Data processing project "EXA4MIND" (Horizon Europe GA No. 101092944). Two of the strongest academic supercomputing centres in Europe take part in EXA4MIND. The collaboration to evolve the "Smart Moisture Mapper" subsystem of TerraviewOS in this context is briefly sketched. Connecting database systems, High-Performance-/Cloud-Computing systems and European Data Spaces with appropriate data access and transfer mechanisms, EXA4MIND shall demonstrate competitive advantage in scientific and enterprise data analysis needed in complex applications such as TerraviewOS.

International Symposium on Grids and Clouds (ISGC) 2023,
19 - 31 March 2023
Academia Sinica Taipei, Taiwan

*Speaker

PoS(ISGC&HEPiX2023)021

1. Introduction

In an era where global warming and the respective mitigation measures are a main driver of societal change, and IT provides a large part of the solutions to our problems, environmental-science methods and IT methods have to go hand in hand. One prime example for this are Smart Agriculture and Smart Vineyards, where production is optimised for efficiency, quality and resiliency. Precision agriculture is an emerging framework that enables key decision makers - farmers, agronomists, etc. to use data from remote sensing, and connected IoT devices in optimising agricultural workflows in a timely manner. TerraviewOS[1] is an innovative tool which enables precision agriculture, targeting more specifically vineyards.

Clearly, precision agriculture applied in the right way will also help mankind to reach us UN's Sustainable Development Goals (SDGs [1]), in particular the goals number 2 ("zero hunger"), 10 ("reduced inequalities"), 12 ("responsible consumption and production"), 13 ("climate action"), and 15 ("life on land"). The potential of using IT for greater resiliency and effectiveness in particular in the agricultural sector, but also for fostering sustainable development, has long been recognised by funding agencies such as the EU, aiming to support respective "lighthouse projects" within their programmes (e.g. [2]).

With this background, Smart Agriculture and Smart Vineyards are a very topical example for what we call Environmental Computing – in the sense of supporting environmental sciences and applications with top-notch IT. Environmental Computing nowadays enables on-site applications e.g. in agriculture, building upon classical environmental and earth-system modelling concepts as well as data evaluation. Often concerned with land-use or water management, environmental computing for on-site applications profits enormously from Cloud and Edge concepts, from IoT sensors and from satellite imagery. This applies to TerraviewOS as well, with automatised, efficient data-driven workflows being key to combining the technologies just mentioned. However, beyond this, TerraviewOS can take strong advantage from GPU Computing or High-Performance Computing (HPC), for example when it comes to improving internal models.

European supercomputing centres, in particular in the academic sector, have long recognised the importance of Environmental Computing [3]. The respective applications tend to significantly drive innovation at these centres, posing challenges beyond traditional, simulation-centric HPC, as for example *(i)* cloud-native computing services have to be offered together with computing and data-flow orchestration, and *(ii)* the centres must be able to efficiently process huge datasets in database-management systems (DBMS), often used in geoscience (e.g. via geo-information systems, GIS).

The development of TerraviewOS coincides with several EU-funded projects where supercomputing centres evolve their services in these relevant directions. Collaboration efforts on high-performance data processing can thus give TerraviewOS an edge and help the centres to orient at real requirements in their development. In particular, IT4Innovations (IT4I) National Supercomputing Center of the Czech Republic, and Leibniz Supercomputing Centre (LRZ), part of the German national supercomputing framework "GCS", will work with Terraview and further close partners to address challenge *(ii)* in the scope of the EXA4MIND[2] project ("EXtreme Analytics for

[1] https://www.terraview.co
[2] https://exa4mind.eu

(4) MINing Data spaces", 2023-2025, Horizon Europe GA No. 101092944). Evolving the "Smart Moisture Mapper" (SMM) of Terraview OS with its internal models will be a primary target of this collaboration. All this will be strongly supported by the LEXIS[3] platform from the project "Large-Scale EXecution for Industry and Society" (2019-2021, Horizon 2020 GA No. 825532, [4]), enabling easily-usable and highly efficient workflow orchestration on heterogeneous systems (Cloud, HPC and beyond), and thus addressing challenge *(i)*.

This paper gives an overview of TerraviewOS, discusses it as a typical Environmental-Computing application, and sketches development directions in the scope of the EXA4MIND project. Below, we first lay out the smart vineyard management approach of Terraview (Section 2). Then, we discuss the envisaged collaborative activity within EXA4MIND (Section 3), considering the characteristics of TerraviewOS. Finally (Section 4) we summarise our developments and ideas, and give an outlook on our upcoming activity.

2. Smart Vineyard Management Approach of Terraview – TerraviewOS

TerraviewOS is a complete vineyard management tool which utilises the power of large scale data sets with artificial intelligence (AI) through machine learning (ML) models. Before understanding the data sources and processes, we identify some of the key requirements on such a system.

2.1 Typical Requirements on a Smart Vineyard System; Functionalities

A smart vineyard management system, among other capabilities, must support the following functionalities:

- ability to store the list of vineyards along with relevant plot, planted grapes, and soil characteristics;
- ability to maintain the precise geo-locations, plot boundaries;
- ability to track all farming inputs - labour, machinery, nutrients, water usage, etc.; and
- ability to provide weather forecasts and historical trends for the vineyard plots.

In addition to these must-have features which digitise agriculture operation, additional requirements around prediction capabilities must be satisfied in order to qualify as a "smart" management tool:

- ability to predict the yield quantity from a vineyard at the end of the season;
- ability to predict the chances of certain disease manifestations;
- ability to recommend corrective actions to react on weather-related alerts as well as nutrients- and disease-related alerts; and
- ability to predict harvest dates based on evolving seasonal metrics.

[3]https://lexis-project.eu

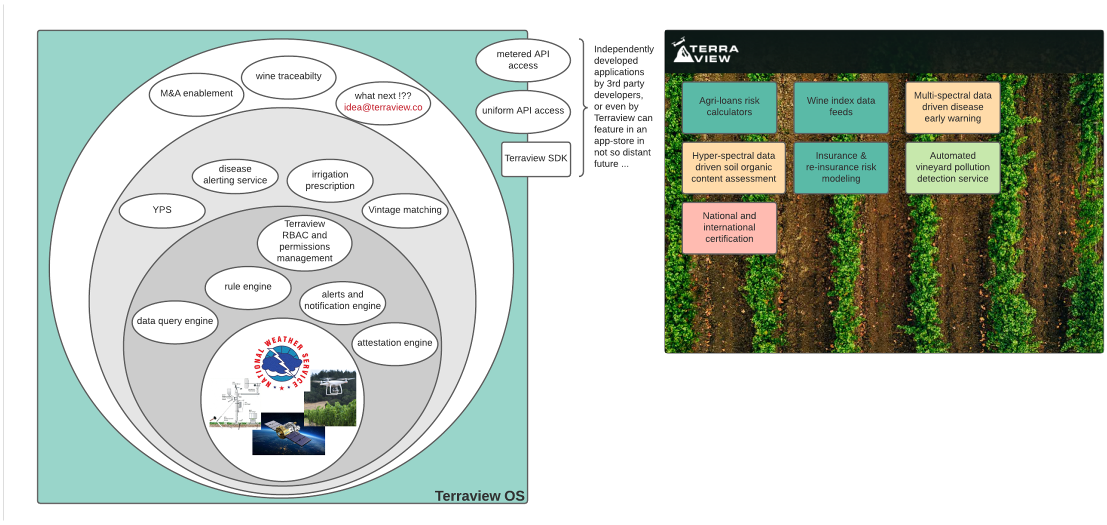

Figure 1: TerraviewOS features and structure, with inner and outer layers as in a computer OS.

One can further expand the platform requirements around environmental, social, and corporate governance and UN's SDGs [1] in the direction of water use efficiency, demanding an

- ability to manage irrigation schedules along with generation of variable rate water prescription maps; and an
- ability to map and track the soil moisture profile of an agriculture plot.

And finally, since large managed vineyards are a reality, where multiple actors are involved, there are requirements arising from practical aspects of using such a platform:

- support for multiple users with proper rights management and access control;
- ability to support data collection by ground survey teams;
- fine grained access control at each vineyard plot level; and
- differentiated view of the platform based on role.

With these requirements listed above, it is obvious that modern IT infrastructure becomes a critical enabler when dealing with massive EO data sets, and long-term weather and climate data sets. Capabilities of Cloud Computing and HPC centres become a massive asset in the realisation of such a tool.

2.2 TerraviewOS as a "Vineyard Operating System"

TerraviewOS is architected keeping the design principles of a traditional computer operating system (OS). Figure 1 illustrates the layered architecture approach in TerraviewOS where the inner layers enable the services in the outer layers.

TerraviewOS makes use of the following seven types of data sources:

- low-Earth orbit commercial and public domain satellites;

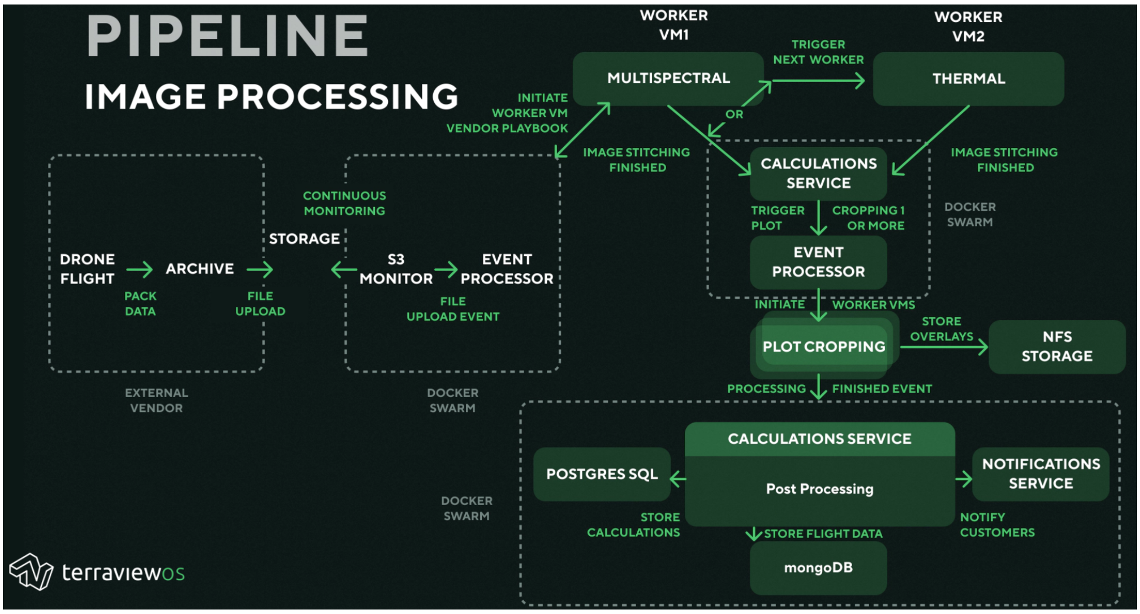

Figure 2: Image processing pipeline as an example for a workflow in TerraviewOS.

- agriculture drone survey data files, using multi-spectral and thermal payloads;
- popular ground sensors' data feeds – weather and soil moisture IoT probes;
- user provided, tracked data as entered in the Traceability module;
- regional and hyper-local weather service providers;
- data fed into the platform via companion mobile apps by scouting teams; and
- customer-related static plot metadata and soil characteristics including linked historical yield data sets.

These data sets are used to create relevant actionable insights for the decision makers such as agronomists. The TerraviewOS platform key enablers are: flexible rule-engine-driven expert system implementation, automated satellite and drone payload data processors, alerting engine, and a highly scalable data-source-agnostic data-ingestion engine backed by high-performance time series and GIS-query enabled data stores.

2.3 TerraviewOS Workflow Principles

TerraviewOS workflows are designed as event-driven processors where processing of data happens in stages, and a previous stage activates the next stage via events sent as messages over a pub-sub messaging subsystem. Figure 2 illustrates using an example of drone imaging data processing. The intermediate stages are all designed to be stateless; states are passed as inputs through a shared POSIX compliant network file system. In order to make TerraviewOS scalable, all implemented workflows within the platform generally follow these principles:

- statelessness;

- data idempotence;
- asynchronous communication via pub-sub system;
- data as input and output generally via a networked file system.

Following these principles enables the TerraviewOS platform to remain agnostic to how the input is collected, when it is collected, and when the data is ready to be further processed. Being idempotent is critical: upon failures, be they partial or total, repeated processing with same inputs must always produce the same outputs.

2.4 Context of TerraviewOS and Potential in a "Smart Moisture Mapper" (SMM)

Terraview, producing TerraviewOS, is a climate-SaaS startup working towards enabling vineyard owners and managers to globally deal with the effects of climate change. Founded in late 2019, the vision of the company has evolved multiple times. Starting with a few pilot users in Spain, the platform now counts more than 200 organisations (with more than 25'000 ha of prime vineyards) spread in all major continents. TerraviewOS with its integrated approach aims to stop the proliferation of disconnected tools, which had been becoming an impediment rather than an enabler of smart agriculture.

Lately, across the globe, water is becoming a scarce resource and policy makers are emphasising better use of water as a key sustainability goal. Globally, up to 85 % of human water consumption are typically estimated to be due to agriculture (e.g. [5]). To support a projected world population which may peak somewhat above or below 10 billion in the current century [6], agriculture production probably needs to be increased further by more than 50 % [7]. Thus, it is immediately clear why water is the centre of attention – also in the sector of precision agriculture and of TerraviewOS customers. Better water management is central to our continued survivability.

In order to optimise water usage in crops including vineyards, it is imperative to know the state of soil moisture and the moisture level variations over time. This is possible using soil moisture sensors. The main challenge is – over 90% of current TerraviewOS customers make no use of sensors. Terraview has thus embarked on a mission to gather soil moisture insights without using any ground sensors, as one important direction for further development of TerraviewOS. Soil moisture shall be derived from satellite imagery (multi-spectral/radar data) by a SMM module.

Training machine-learning models for this task will require significant computing and data resources and has to be implemented with high efficiency: the training data set for any Area of Interest combining multi-spectral and radar satellite data sets will be approximately 4TB in size, with newer data available weekly. Any model which is prepared may need to be refreshed even on a weekly basis to remain most relevant over that Area of Interest. Within EXA4MIND, partnering with LRZ and IT4Innovations as supercomputing centres provides an ideal context to solve this urgent challenge and drive the development of the SMM. This idea and the context of EXA4MIND are laid out within the next Section below.

PoS(ISGC&HEPiX2023)021

3. Leveraging supercomputing Ecosystems for TerraviewOS in the Scope of EXA4MIND Project

Here, we first discuss TerraviewOS as a typical Environmental Computing application relying on Cloud-Computing- and DBMS-based technology (Sections 3.1 and 3.2). We then sketch what the EXA4MIND consortium, including IT4Innovations and LRZ, is planning to support supercomputing with Extreme Data and DBMS (Section 3.3).

All this leads us to finally describe the first step to boost TerraviewOS via the EXA4MIND ecosystem – development of the SMM of TerraviewOS (Section 3.4).

3.1 TerraviewOS as a Cloud-Native Environmental Computing Application

TerraviewOS adopts state-of-the-art cloud-native technology as it is typical for modern applications in the industrial and SME sector, but increasingly also in scientific computing (see e.g. [8]). It uses Ansible [9] and Docker Swarm [10] to deploy and orchestrate the application in an infrastructure-as-code approach (see e.g. [11]). This implies that computing centres running parts TerraviewOS, even only for development purposes, have to provide a IaaS-Cloud and/or container-orchestration environments. If HPC or massive GPU-based resources shall be used in addition, efficient cross-system workflow orchestration is needed. All these requirements are basically met by the LEXIS platform [4, 12], which is an important basis for the developments envisaged within EXA4MIND.

Many Environmental-Computing applications have a requirement profile similar to TerraviewOS: The discipline of Environmental Computing has never been focused as much on pure HPC as the sectors of computational fluid dynamics in physics and engineering, of astrophysical radiative transfer or of molecular dynamics, just to mention a few examples. Instead, versatile smaller-scale IT machinery and systems are required. These are often also used for rich near-real-time processing workflows, producing, e.g., visualisations for various stakeholders right from actual sensor and imaging data.

3.2 TerraviewOS – Importance of Well-Adapted DBMS in Modern Environmental Computing

Recurring data-usage patterns of TerraviewOS, as also of similar applications processing GIS data or typical geoscientific data cubes, call for the usage of well-adapted DBMS. TerraviewOS processes, e.g. (cf. Section 2) GIS data, time-series data, and imagery from satellites and drones.

Besides S3- [13] and NFS-based [14] storage systems, Terraview OS thus uses, for example, PostgreSQL-based SQL databases [15, 16] and MongoDB (e.g. [17]). In addition, a planned integration of COPERNICUS/SENTINEL [18, 19] satellite data stores available in Europe and in particular at EXA4MIND partner institutes [20] poses the need of indexing these satellite data appropriately and thus making at least the metadata available in a DBMS solution. All this is to guarantee an optimum time-to-solution, in the sense that wine-growers will obtain updated imagery, statistics, alerts and strategic suggestions as fast as possible.

3.3 EXA4MIND: Integrating Extreme Data & DBMS with European Supercomputing and Data Spaces

Coming at a time where TerraviewOS is rapidly advancing, the EXA4MIND project will build an ecosystem around an "Extreme Data Database" (EDD), which integrates various (SQL and noSQL-based, cf. e.g. [21]) DBMS and data store solutions (e.g. iRODS/EUDAT-B2SAFE [22, 23]) and will use advanced querying and indexing techniques as well as common REST APIs. Together with efficient data-ingest mechanisms, pre-processors and modules for post-processing, visualisation and interactive query, an efficient platform for Extreme Data analytics and processing is developed. These ideas, which also involve distributed set-ups of storage systems and/or DBMS, extend on the concept of the Distributed Data Infrastructure of LEXIS [24].

Thus, EXA4MIND effectively incorporates database ecosystems and their rich application in hybrid Cloud/HPC workflows for AI, Big/Extreme Data and simulation. Efficient staging and data-loading mechanisms will for the first time enable a standardised usage pattern for DBMS on HPC/supercomputing clusters, where IT4Innovations and LRZ are providing ideal systems for testing and development. EXA4MIND and its EDD will provide REST APIs and mechanisms for making DBMS data available to HPC or Cloud applications within orchestrated workflows. Furthermore, the EDD will be equipped with connectivity to data-management systems on the EU level (in particular EUDAT-B2SAFE [23]) and European Data Spaces. FAIR [25] data publication will be supported as well.

3.4 Supporting TerraviewOS's SMM with EXA4MIND, and further EXA4MIND/EDD Utilisation

One key functionality considered in TerraviewOS development (cf. Section 2) is the ability to recommend irrigation based on weather data (and prediction), the grapevine characteristics and soil moisture. To support this, the SMM module (see Section 2.4) shall enable TerraviewOS to derive soil moisture data purely from satellite (multispectral/radar) data, i.e. independent of actual ground sensors which are often unavailable. At the core of the SMM, a ML/AI model for doing this mapping is needed, which must be adequately trained.

The SMM will thus be ideal as one of the first use cases for the EXA4MIND platform. COPERNICUS/SENTINEL [18, 19] and USGS [26] earth observation data have to be made available for computation, as well as weather datasets from the European Centre for Medium-Range Weather Forecasts (ECMWF [27]) and local soil-moisture sensor time series of vineyards. On a European scope, additional data from the DestinE programme [28] in general can be interesting for this application. We are currently mapping the respective storage and data-access requirements to the DBMS and storage systems envisaged within EXA4MIND's EDD. Clearly, the DBMS used already now in TerraviewOS (Section 3.2) are prime candidates for implementation in the EDD. In a co-development approach, we then aim at significantly boosting SMM development with fast and automatised training workflows (including periodic data acquisition). This shall demonstrate the immediate benefits of usage of the EXA4MIND and LEXIS ecosystems for TerraviewOS and many other kinds of complex applications. The significant supercomputing and data-storage resources we have at hand will be an ideal basis for this endeavour.

Starting from this and other pilot use cases within EXA4MIND, we will develop the ability to efficiently map data management requirements from various disciplines onto the EDD ecosystem. This also opens up the possibility of integrating further TerraviewOS workflows with EXA4MIND and its EDD.

4. Summary and Outlook

In this contribution we have presented the smart vineyard operating system TerraviewOS and envisaged developments on this system in the context of the Horizon Europe Project EXA4MIND.

TerraviewOS, which we have argued is a typical Environmental Computing application for on-site usage, addresses the need for precision agriculture in the field of vinery. Systems as this will be essential for pursuing the UN Sustainable Development Goals [1] – in particular those related to water and food – in the context of climate change and a still-growing world population. They will save water by precise irrigation, and make farming still more efficient. Complex Environmental Computing systems such as TerraviewOS usually depend on flexible Cloud-Computing infrastructure and data systems, as well as proper workflow control.

Building on technology from EUDAT and on advanced orchestration, the LEXIS platform has been enabling automatised cross-system workflows involving HPC and Cloud-Computing clusters at supercomputing centres [4]. However, it has not yet specifically addressed the challenge of including DBMS in such workflows, and thus of enabling supercomputing involving the most efficient DBMS with Big or Extreme Data. This is the focus of EXA4MIND, where an "Extreme Data Database" (EDD) – integrating various Database Management and Data Management Systems – will be built and enabled to be used with supercomputers and various data sources. Moreover, data exchange between the EDD and European data ecosystems will be facilitated, including FAIR data handling.

The power of these systems shall be leveraged to complete an AI-based "Smart Moisture Mapper" (SMM) for TerraviewOS, able to derive soil moisture from satellite data where (as often) ground sensor data are unavailable. The SMM can then be used for irrigation recommendations based on earth observation data and weather data/predictions, with water as a scarce resource in mind, as mentioned above. Building this system and making it ready for every-day usage involves a demanding data-driven training workflow for the model, with sensor data and satellite data, which has to be repeated for different Areas of Interest and times. Orchestrating the workflow and collaboratively optimising it for efficient execution, using excellent back-end systems, will advance TerraviewOS and help the EXA4MIND project to develop at the same time.

The case presented demonstrates how top-notch Environmental Computing applications challenge European supercomputing centres and data experts, and how they in the end help to co-develop IT systems. Currently, we are preparing the implementation of the SMM within EXA4MIND, and of other first-round use cases within the project. On the long term, a much improved support for data-intensive and complex workflows at Europe's supercomputing centres shall enable a broad range of applications, with the opportunity to also run further modules of TerraviewOS.

Acknowledgements

Parts of this work and part of the contributing authors are funded/co-funded by the EU's Horizon Europe Research and Innovation programme under grant agreement N° 101092944 (EXA4MIND – EXtreme Analyitcs for MINing Data spaces). Various computing and data facilities at IT4Innovations and LRZ as well as tools and services from EUDAT have been and will be utilised for the work described, and we thank all our colleagues responsible of these services for support.

References

[1] UN Resolution, *Transforming our world : the 2030 Agenda for Sustainable Development, Issued in GAOR, 70th sess., Suppl. no. 49.* (2015), https://digitallibrary.un.org/record/3923923.

[2] EU Publications Office, "Smart XG, last-mile and edge solutions for remote farming, forestry and rural areas | Programme | Horizon | CORDIS | European Commission." https://cordis.europa.eu/programme/id/HORIZON_HORIZON-CL6-2021-COMMUNITIES-01-03, Cited 19 Mar 2023, 2022.

[3] M. Heikkurinen, M. Schiffers and D. Kranzlmüller, *Environmental Computing 1.0: The Dawn of a concept*, in *Proceedings of International Symposium on Grids and Clouds 2015 — PoS(ISGC2015)*, vol. 239, 030, 2016, DOI.

[4] A. Scionti et al., *HPC, Cloud and Big-Data Convergent Architectures: The LEXIS Approach.*, in *CISIS 2019*, L. Barolli, F. Hussain and M. Ikeda, eds., Advances in Intelligent Systems and Computing, vol. 993, pp. 200–212, Springer, Cham, 2020, DOI.

[5] P.H. Gleick, *Water use*, *Annual Review of Environment and Resources* **28** (2003) 275, DOI.

[6] United Nations, Department of Economic and Social Affairs, Population Division, "World Population Prospects 2022, Online Edition." https://population.un.org/wpp/Download/Standard/MostUsed/ and https://population.un.org/wpp/Download/Files/1_Indicators(Standard)/EXCEL_FILES/1_General/WPP2022_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT_REV1.xlsx, Cited 16 Apr 2023, 2022.

[7] M. van Dijk, T. Morley, M. Rau and Y. Saghai, *A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050*, *Nature Food* **2** (2021) 494, DOI.

[8] C.L. Gentemann, C. Holdgraf, R. Abernathey, D. Crichton, J. Colliander, E.J. Kearns et al., *Science storms the cloud*, *AGU Advances* **2** (2021) e2020AV000354, DOI.

[9] Red Hat, Inc., "Ansible is Simple IT Automation." https://www.ansible.com, Cited 16 Apr 2023, 2023.

[10] Docker Inc., "Swarm mode overview | Docker Documentation." https://docs.docker.com/engine/swarm, Cited 16 Apr 2023, 2023.

[11] D.A. Dias, P. Mestre and C. Serôdio, *Infrastructure-as-Code for Scientific Computing Environments*, in *CENTRIC 2019 – The Twelfth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services.*, pp. 7–10, IARIA, ISBN: 978-1-61208-754-2, 2019.

[12] S. Hachinger, M. Golasowski, J. Martinovič, M. Hayek, R.J. García-Hernández, K. Slaninová et al., *Leveraging high-performance computing and cloud computing with unified big-data workflows: The lexis project*, in *Technologies and Applications for Big Data Value*, E. Curry, S. Auer, A.J. Berre, A. Metzger, M.S. Perez and S. Zillner, eds., pp. 159–180, Springer, Cham (2022), DOI.

[13] Amazon Web Services, Inc. and affiliates, "Cloud Object Storage – Amazon S3 – Amazon Web Services." https://aws.amazon.com/de/s3/, Cited 23 Mar 2023, 2023.

[14] T. Haynes and D. Noveck, "Network File System (NFS) Version 4 Protocol." RFC 7530, Mar., 2015. DOI.

[15] M. Stonebraker and L.A. Rowe, *The Design of POSTGRES*, *SIGMOD Rec.* **15** (1986) 340–355, DOI.

[16] The PostgreSQL Global Development Group, "PostgreSQL: Documentation: 15: 1. What is PostgreSQL?." https://www.postgresql.org/docs/current/intro-whatis.html, Cited 30 Mar 2023, 2023.

[17] K. Chodorow and M. Dirolf, *MongoDB - The Definitive Guide: Powerful and Scalable Data Storage.*, O'Reilly & Associates, Sebastopol (CA) (2010).

[18] European Commission / Copernicus, "Homepage | Copernicus." https://www.copernicus.eu, Cited 16 Apr 2023, 2023.

[19] European Space Agency (ESA), "Sentinel Online - ESA - Sentinel Online." https://sentinels.copernicus.eu, Cited 16 Apr 2023, 2023.

[20] C.M. Albrecht, "Open Source Science for Large-Scale Data Mining in Earth Observation." https://zenodo.org/record/6315122, Feb., 2022. DOI.

[21] R. Cattell, *Scalable sql and nosql data stores*, *SIGMOD Rec.* **39** (2011) 12–27, DOI.

[22] H. Xu, T. Russell, J. Coposky, A. Rajasekar, R. Moore, A. de Torcy et al., *iRODS primer 2: Integrated Rule-Oriented Data System*, Morgan & Claypool Publishers, Williston, VT (2017), DOI.

[23] EUDAT Collaborative Data Infrastructure, "B2SAFE-EUDAT." https://www.eudat.eu/services/b2safe, Cited 30 Mar 2023, 2023.

PoS(ISGC&HEPiX2023)021

[24] J. Munke, M. Hayek, M. Golasowski, R.J. García-Hernández, F. Donnat, C. Koch-Hofer et al., *Data system and data management in a federation of hpc/cloud centers*, in *HPC, Big Data, and AI Convergence Towards Exascale*, pp. 59–77, CRC Press, Boca Raton (FL) (2022), DOI.

[25] M.D. Wilkinson et al., *The FAIR Guiding Principles for scientific data management and stewardship*, *Scientific Data* **3** (2016), DOI.

[26] M.C. Rabbitt, *The United States Geological Survey, 1879-1989*, U.S. Geological Survey circular ; 1050, U.S. G.P.O. (1989).

[27] European Centre for Medium-Range Weather Forecasts, "ECMWF | Advancing global NWP through international collaboration." https://www.ecmwf.int/, Cited 16 Apr 2023, 2023.

[28] European Commission, "Destination Earth | Shaping Europe's digital future." https://digital-strategy.ec.europa.eu/en/policies/destination-earth, Cited 16 Apr 2023, 2023.