DOI: 10.26314/GARR-Conf23-proceedings-01

Digital Twin Components for Geophysical Extreme Phenomena: the example of Volcanic Hazards within the DT-GEO project

Stefano Cacciaguerra[1], Antonio Costa[1], Francesca Quareni[1], Paolo Papale[1], Flavio Cannavò[1], Arnau Folch[2], Giovanni Macedonio[1], Sara Barsotti[3]

[1]Istituto Nazionale di Geofisica e Vulcanologia (INGV), [2]Geociencias Barcelona (GEO3BCN-CSIC), [3]Icelandic Meteorological Office (IMO)

Abstract. The project Digital Twin for GEOphysical extremes-(DT-GEO) aims to use Digital Twin Components to create replicas of physical systems, serving as a virtual laboratory to study natural extreme events. The rationale is the intrinsic risks of potentially catastrophic events to anthropic activities, infrastructures, and cultural heritage. In the framework of the project, this paper describes, how the DTC workflow architecture is designed, focusing on flexibility, scalability, and maintainability, and how it is further developed. To demonstrate how ICT efforts can expand horizons in Geosciences, an application to volcanic hazard is presented taking as a case study the 2019 volcanic eruption of Raikoke (Kuril Islands).

Keywords. Digital Twin, Geoscience Data Management, Geoscience and ICT, Volcanoes and HPC

1. Introduction

Natural extreme events, like earthquakes, landslides, tsunamis, and volcanic eruptions, may involve significant risks to infrastructures, cultural heritage, and human lives, and, since ever, the scientific community is called to contribute better understand the complexity of their non-linear behavior and to mitigate their impact. On the other hand, today, major efforts are spent in Europe towards the empowering of ICT, innovation and skills development, through initiatives like European Open Science Cloud (EOSC), European Digital Innovation (EDI), and European High Performance Computing (EuroHPC - Ejarque et al., 2022).

Meanwhile, large amounts of multidisciplinary Earth system data are now available in near-real-time according to the FAIR principles (Wilkinson et al., 2016). As we enter the Exascale Era of computation, Geosciences require modeling codes able to perform over massive HPC clusters (Folch et al., 2023).

Recognizing the potential of Digital Twins Components (DTC, e.g. Fuller et al. 2020), the project Digital Twin for GEOphysical extremes-DT-GEO (Carbonell et al. 2023) has been funded under the Horizon Europe programme (2022-2025) within the European Commission initiative Destination Earth in order to develop numerical clones that accura-

Firenze 14-16 giugno 2023 - ISBN: 978-88-946629-2-4-

tely replicate the extreme geophysical phenomena. The project aims to integrate advancements from european projects, such as EOSC-synergy and eFlows4HPC (Talia et al. 2023), as well as the Center of Excellence for Exascale in Solid Earth-ChEESE (Folch et al. 2023) (Fig.1). The goal is to produce self-contained and containerized DTCs, incorporating cutting-edge codes, artificial intelligence layers, large data streams, and data assimilation methodologies. Serving as virtual replicas, the DTCs will provide a better understanding and more comprehensive analysis of extreme geophysical phenomena.
12 DTCs and their workflows are designed to be optimized on EuroHPC systems or cloud computing environments in order to address volcanic eruptions, tsunamis, earthquakes, and induced seismicity. When driven by near-real-time monitoring data, DTCs can contribute to early warning systems, estimation of potential consequences, urgent computations, and rapid post-event analysis. Alternatively, when driven by user-defined synthetic data, DTCs can draw various scenarios improving hazard assessment.
To validate their functionality, DTCs use selected Site Demonstrators (SD) with data from the EPOS Integrated Centralized Services of the European Plate Observing System (EPOS). The long-term vision is to integrate the DTCs into the Destination Earth framework, facilitating internal component coupling and promoting external integration with other digital twins. Furthermore, DTCs/SDs will be the basis for the EPOS Distributed Integrated Centralized Services (Cocco et al. 2022).

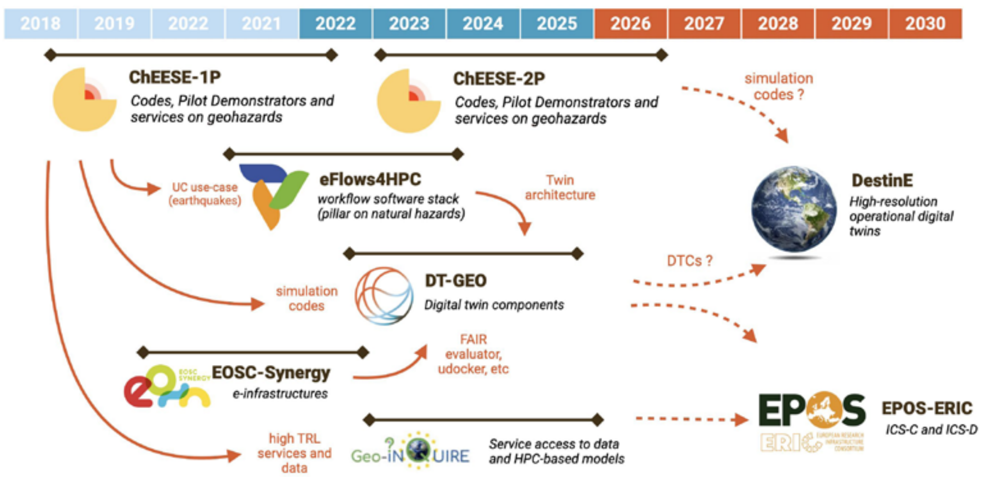


Fig. 1 Background and destination of DT-GEO in the European infrastructural research ecosystem

2. Data Architecture, Workflows and e-Infrastructures

The DTC workflow architecture should focus on flexibility, scalability, and maintainability. To identify the requirements of different DTCs/SDs, it is essential to consider the data ecosystem, such as data and metadata quality, collection methods, processing activities, and storage strategies. The required modular architecture should be adaptable to the execution across various e-Infrastructures, including HPC systems and cloud environments, within federated Research Infrastructures (RI) The architecture must also be easily exten-

dable and deployable.
To develop a DTC, DT-GEO adopts a layered software stack as shown in Fig. 2. On the top there is the physical model describing the specific geophysical phenomenon. Intermediate layers follow, managing workflows, containers, and ensuring data and metadata quality. The physical model is the main component of the workflow, which can in turn enhance it with artificial intelligence, machine learning or data assimilation techniques. This framework accesses a robust ecosystem with quality repositories for software, data, and metadata. Finally, the workflow is containerized to be executed on different e-Infrastructures.
To implement the workflows, the PyCOMPSs environment is combined with the software stack developed by the eFlows4HPC project. PyCOMPSs (Tejedor, 2017) is a Python-based framework to develop and execute parallel applications on distributed infrastructures, whereas eFlows4HPC software provides a range of libraries and frameworks for data processing, Machine Learning (ML), visualization, and workflow management. Thanks to PyCOMPSs and eFlows4HPC, developers can create scalable workflows that handle large volumes of data and are easily deployed and maintained.
To enable DT-GEO workflows on e-Infrastructures, such as FENIX (Alam et al. 2022) or EuroHPC, is required to achieve seamless access, to deploy DTCs/SDs as containers and to execute on-demand DTCs/SDs on a HPC or cloud environment. In order to implement this approach, DTCs/SDs are containerized applications downloadable from Dockerhub or other docker registries. DTCs/SDs are implemented with udocker (Gomes et al. 2018), because the containerized images can be downloaded and executed without needing root access, ensuring a secure and streamlined deployment process.

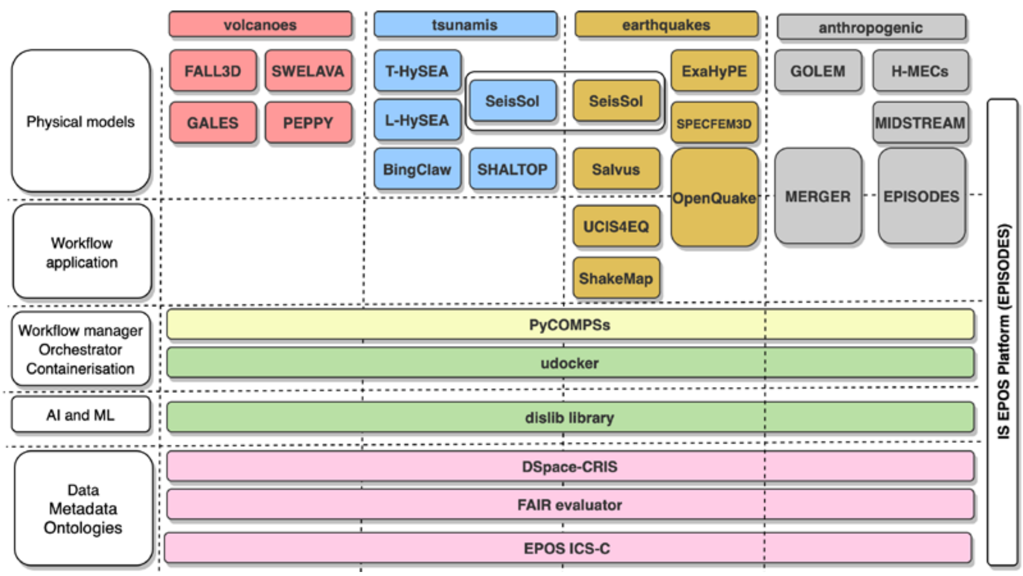


Fig. 2
DT-GEO layered software stack

3. Application to volcanic phenomena: a case study

Volcanoes encompass many different domains, from deep regions of magma storage below and within volcanoes, to dykes and conduits, surrounding rocks, the Earth surface, and the atmosphere, with non-linear coupling. From each domain a stream of high-quality data (accessible e.g. through the EPOS RI) is produced, which is precious to base research, and feeds risk analyses and civil protection activities, including early warning systems and emergency planning.

The first goal within DT-GEO is the definition of the requirements, specifications, and interoperability of the volcano related components, including software engineering tools, HPC resources, and data lakes. Once the first step is achieved, four DTCs will be developed and implemented for volcanic unrest, volcanic ash clouds and tephra ground accumulation, lava flows, and volcanic gas dispersal.

Here we present as a case study the diffusion in the atmosphere of a volcanic ash column. Fig. 3 shows an aerial photograph of an ash column, and a sketch of the characteristics of the phenomenon, highlighting the observed data.

Modeling volcanic ash concentrations relies on the measurement of various physical parameters that are affected by errors. To enhance the accuracy of assessments in an operational context, the project utilizes geostationary satellite observations (including atmospheric radiance, VIS/IR, and UV/IR with shadow), aerial visual observations, and ground-based measurements like radar, lidar, and VIS/IR cameras. The integration of the observed data into numerical models is able to significantly improve the accuracy of assessments.

As sketched in Fig. 4, the DTC replicates the phenomenon of volcanic ash dispersion in atmosphere by feeding the existing input observations (from ECMWF, ESA, INGV/EPOS, and IMO) to a data assimilation tool based on an ensemble approach, which combines the FALL3D dispersion physical model (Folch et al. 2009) with the Parallel Data Assimilation Framework (PDAF). The FALL3D+PDAF system is designed to operate in parallel and support online-coupled data assimilation, and can be effectively integrated into operational workflows thanks to HPC capabilities.

The spatial distribution of column mass, according to Mingari et al. (2022), is shown in Fig. 5a,18 hours after the start of the eruption, according to the considered real state, after which synthetic observations are created (Fig.5b). Figs. 5c and 5d show the distribution of column mass estimated from the physical numerical model, and the result of the DTC analysis with data assimilation.

In both cases, the ash cloud is transported eastward by upper-level winds. However, the results obtained without assimilation exhibits a broader spatial distribution compared to the real state due to ensemble dispersion. Furthermore, the modeling without assimilation fails to reproduce the peak mass position of the column occurring in the northern region of the cloud.

In contrast, the analyzed mass load field closely approaches the real state after a few assimilation cycles (Fig.5d).

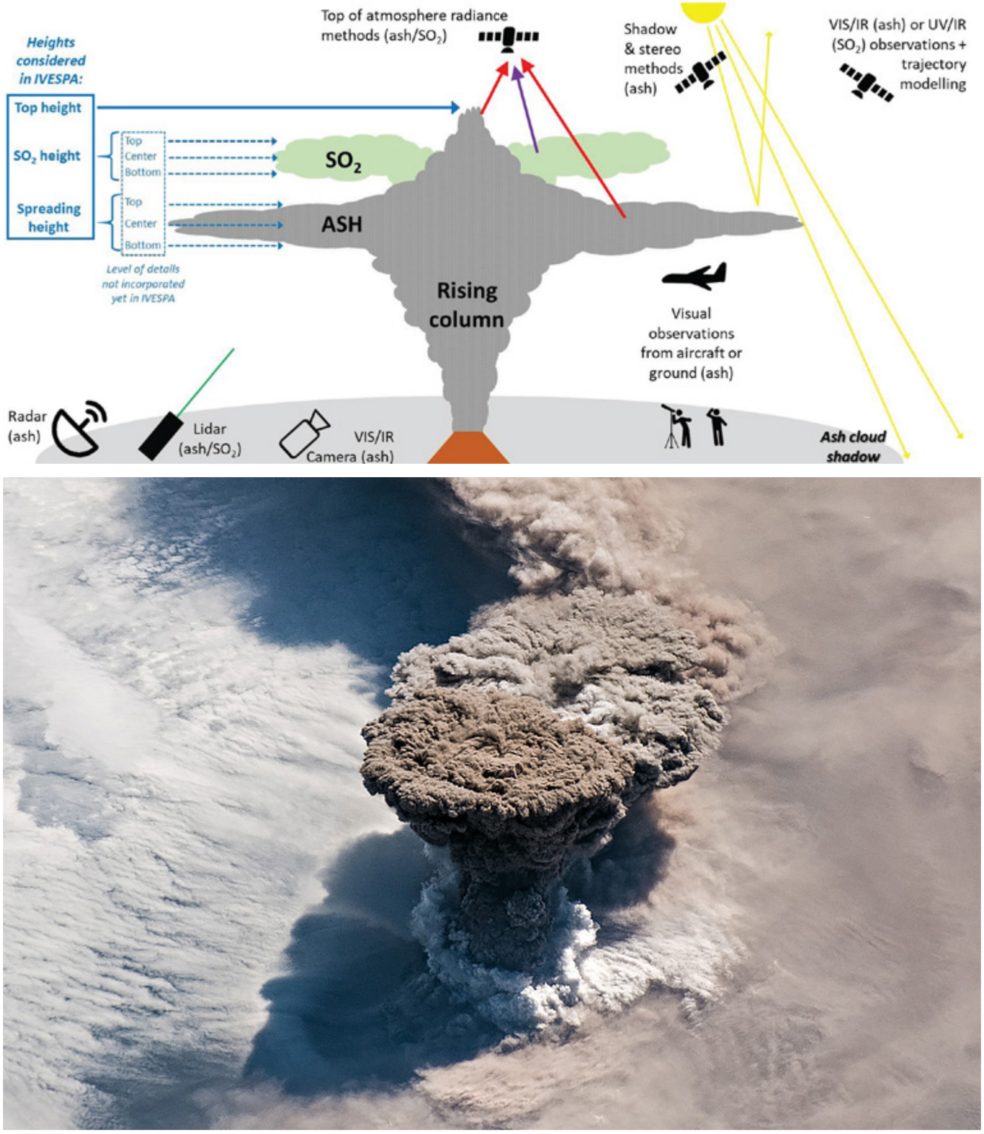


Fig. 3
Top: Aerial photograph of an ash column generated by a volcanic eruption (Raikoke, Kuril Islands) after Global Volcanism Project (2019). Bottom: sketch highlighting the observed data (bottom) after Aubry et al. (2021)

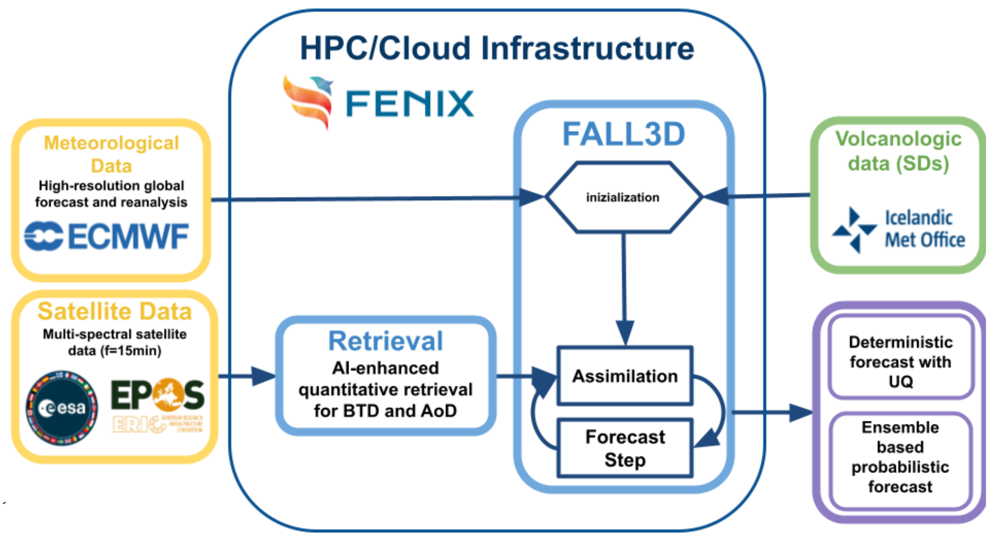


Fig. 4
Sketch of the DTC for volcanic ash showing the data input sources and the assimilation and forecast loop process

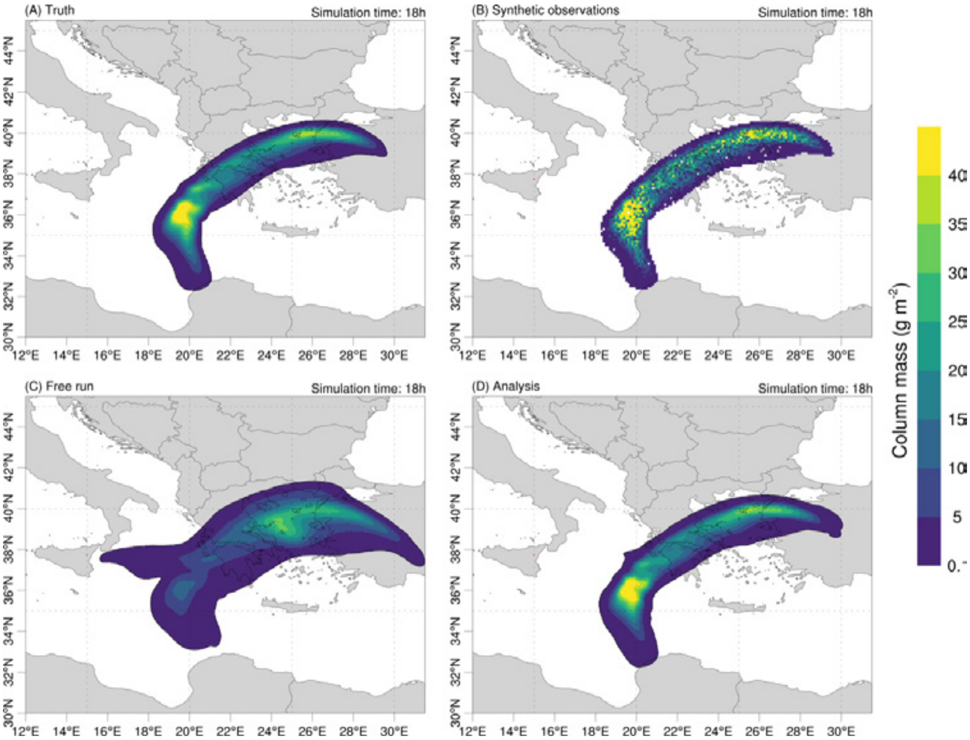


Fig. 5
Spatial distribution of ash mass loading for the DTC experiment at t=18 h after eruption start. From the true state (a) synthetic observations (b) are generated. Panel (c) shows the results from a run without assimilation, while in panel (d) the result of the DTC analysis with data assimilation are displayed (after Mingari et al. 2022)

4. Conclusions

This paper delves into the design guidelines for DT-GEO's data-driven workflows, with a spotlight on DTC implementation, and presents a case study on volcanic hazards. The case study provides a first example of the potential impact of DTCs/SDs in Geosciences. DT-GEO plans to test 3 additional SDs to consolidate these findings. However, the research team is prepared to switch to a different SD if a new eruption occurs, so as not to miss the chance to observe and monitor the evolution of the phenomenon in real time.

It's crucial to note that, while these events can have a significant impact on human activities, they are sporadic in nature. Capturing their behavior as it happens is crucial for knowledge advancement. If such observations are not possible, we can still rely on DTCs/SDs, which then become the only virtual laboratory available to us.

Finally, we plan to access the pre-exascale supercomputer Leonardo, hosted by the CINECA consortium at the Big Data Technopole in Bologna, Italy, and owned by EuroHPC JU. This will empower us to run interactive and massively scalable GPU-enabled workflows (Turisini et al. 2023).

Acknowledgement

We acknowledge the Project DT-GEO: Digital Twin for GEOphysical extremes - funded by Horizon Europe under the grant agreement No 101058129, https://cordis.europa.eu/project/id/101058129

Bibliography

Alam, S. R., Bartolome, J., Carpene, M., Happonen, K., Lafoucriere, J. C., Pleiter, D. (2022), Fenix: a Pan-European federation of supercomputing and cloud e-infrastructure services. Commun. ACM, 65(4), pp 46-47.

Aubry, T. J., Engwell, S. L., Bonadonna, C.,. Carazzo, Scollo S., , Van Eaton, A. R., Taylor I.A., Jessop D., Eychenne J., Gouhier M., Mastin L.G., Wallace K.L., Biass S., Bursik M., Grainger R.G., Jellinek M., Schmidt A. (2023), The Independent Volcanic Eruption Source Parameter Archive (IVESPA, version 1.0): A new observational database to support explosive eruptive column model validation and development, J. Volcanol. Geoth. Res., 417, 107295

Carbonell R., Folch A., Costa A., Orlecka-Sikora B., Lanucara P., Løvholt F., Macias J., Brune S., Gabriel A.-A. Barsotti, S., Behrens, J., Gomes J., Schmittbuhl J., Freda C., Kocot J., Giardini D., Afanasiev M., Galves H., Badia R. (2023), Digital Twinning of Geophysical Extreme Phenomena (DT-GEO), EGU General Assembly 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-5674

Cocco, M., Freda, C., Atakan, K., Bailo, D., Contell, K. S., Lange, O., Michalek, J. (2022). The EPOS Research Infrastructure: a federated approach to integrate solid Earth science data and services. Ann. Geoph., 65(2), DM208.

Costa, A., Macedonio, G., and Folch, A. (2006), A three-dimensional Eulerian model for transport and deposition of volcanic ashes, Earth Planet. Sc. Lett., 241, pp. 634–647

Ejarque J. , Badia R. M. , Albertin L. , Aloisio G. , Baglione E. , Becerra Y. , Boschert S. , Berlin J. R. D'Anca A. , Elia D. , Exertier F. , Fiore S. , Flich, J, Folch A., ..., Volpe, M. (2022), Enabling dynamic and intelligent workflows for HPC, data analytics, and AI convergence, Future Generation Computer Systems, 134, pp. 414-429

Folch, A., Costa, A., Macedonio, G. (2009), FALL3D: A Computational Model for Transport and Deposition of Volcanic Ash, Comput. Geosci., 35, pp. 1334–1342

Folch A., Abril C. , Afanasiev M., Amati G., Bader M., Badia R.M., Bayraktar H.B., Barsotti S., Basili R., Bernardi F., Boehm C., Brizuela B., Brogi F., Cabrera E., Casarotti E., Castro M.J., Cerminara M., Cirella A., Cheptsov A., Conejero J., ..., Wössner U., (2023), The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase, Future Generation Computer Systems, 146, pp. 47-61,

Fuller A., Fan Z., Day C., Barlow C. (2020), Digital twin: Enabling technologies, challenges and open research. IEEE access, 8, pp. 108952-108971.

Gomes J., Bagnaschi E., Campos I., David M., Alves L., Martins J., ... Orviz P. (2018), Enabling rootless Linux Containers in multi-user environments: the udocker tool, Comput Phys Commun., 232, pp. 84-97.

Global Volcanism Program (2019), Report on Raikoke (Russia), in Crafford, A.E., Venzke, E., (eds.), Bulletin of the Global Volcanism Network, 44 (8), Smithsonian Institution (USA)

Mingari L., Folch A., Prata A. T., Pardini F., Macedonio G., Costa A. (2022), Data assimilation of volcanic aerosol observations using FALL3D+PDAF. Atmospheric Chem. Phys., 22(3), pp 1773-1792.

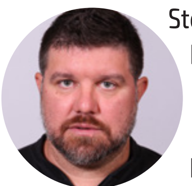
Talia D., Trunfio P. (2023), Urgent Computing for Protecting People From Natural Disasters, Computer, 56 (04), pp. 131-134.

Tejedor E., Becerra Y., Alomar G., Queralt A., Badia R. M., Torres J., ... Labarta, J. (2017), PyCOMPSs: Parallel computational workflows in Python. Int J High Perform Comput Appl., 31(1), pp. 66-82.

Turisini, M., Amati, G., Cestari, M. (2023). LEONARDO: A Pan-European Pre-Exascale Supercomputer for HPC and AI Applications, preprint arXiv:2307.16885.

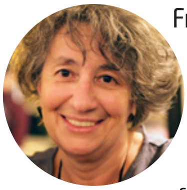
Wilkinson MD., Dumontier M., Aalbersberg IJ., Appleton G., Axton M., Baak A., Blomberg N,. Boiten J.W., da Silva Santos L.B., Bourne P.E., Bouwman J., Brookes A.J., Clark T., Crosas M., Dillo I., ..., Mons B. (2016), The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data, 3: 160018.

Autori



Stefano Cacciaguerra stefano.cacciaguerra@ingv.it
PhD in Computer Science, he serves as Primo Tecnologo at the Istituto Nazionale di Geofisica e Vulcanologia. He designs and implements ICT infrastructure like HPC clusters, cloud architecture and data banks in order to support geophysical and volcanological applications. He is involved in projects DT-GEO and ChEESE-2P. He represents INGV in Italian Computing and Data Infrastructure. He develops ICT solutions for the EMSO Western Ionian Facility and manages EMSO's ICT initiatives in Bologna.

Antonio Costa antonio.costa@ingv.it
He is Senior Researcher at INGV Bologna, where he was also Director (2019-2022). Got the PhD at the University of Bologna (2004). He was Researcher at University of Bristol (2005-2006), where he was also honorary researcher (2007-2010). He worked at University of Reading (2011-2012). He visited the University of Tokyo (2015) and Munich University (2019). For his scientific achievements was awarded the IAVCEI Wager Medalist in 2013 and from 2021 is Member of Academia Europaea.



Francesca Quareni francesca.quareni@ingv.it

Degree and PhD in Physics, she is presently Senior Researcher at the Istituto Nazionale di Geofisica e Vulcanologia, after been employed as Post-Doc Researcher in Geophysics at Arizona State Univ. (USA), and as Associate Professor in Physics of Volcanism at Osservatorio Vesuviano in Naples. From 2013 to 2019 she served as Director of Bologna Branch of INGV. At present, she is member of the Board of Directors of the BigData Association, and represents INGV in the EOSC Association.

Flavio Cannavò flavio.cannavo@ingv.it

He is a Researcher at INGV Osservatorio Etneo in Catania, Italy. He earned both his Computer Engineering degree and Ph.D. in Complex Systems Engineering from the University of Catania. Flavio's expertise revolves around ground deformation modeling in volcanic and tectonic regions. He employs advanced statistical and machine learning techniques to enhance the understanding of Earth's dynamics, making his contributions to the field of Earth sciences and disaster management.

Arnau Folch afolch@geo3bcn.csic.es

He is a research Professor at the Geociencias Barcelona (GEO3BCN-CSIC). Degree in Physics and PhD in Applied Mathematics, he authors 120+ scientific peer-reviewed publications and has participated in 40+ Spanish and European competitive research projects. Coordinator of the EuroHPC Center of Excellence for Exascale in Solid Earth (ChEESE) and the DT-GEO project for digital twins in geophysical extremes. Member of the Strategic Advisory Board of the Destination Earth initiative.

Giovanni Macedonio giovanni.macedonio@ingv.it

He is Research Director at INGV Naples, where he was also Director (2001-2017). He graduated in physics in 1984. He was Researcher at CNR in Pisa (1988-1998). Since 2018, he has been a member of the Grandi Rischi Commission, volcanic risk sector: the Commission of the Civil Protection Department to support decisions in the event of a volcanic crisis. For his scientific achievements was awarded the IAVCEI Wager Medalist in 1998 and from 2019 is Member of Academia Europaea.

Sara Barsotti sara@vedur.is

She is Coordinator of Volcanic Hazards at the Icelandic Meteorological Office and the leader of the State Volcano Observatory team. She obtained her Ph.D at the University of Bologna (2006) after a Ms.D. in Physics at University of Pisa (2002). In the period 2006-2013 she worked as Researcher on contract at INGV, Sezione di Pisa (Italy). Dr Barsotti is and has been involved in several
European and International collaborations and projects.