Science Gateways in EOSC: The NEANIAS Visualisation Gateway

Eva Sciacca, Mario Raciti
INAF-Istituto Nazionale di Astrofisica,
Catania Astrophysical Observatory,
Via S Sofia 78,
95123 Catania,
Italy
Email: {eva.sciacca, mario.raciti}@inaf.it

Mel Krokos, Benjamin Kyd
School of Creative Technologies,
University of Portsmouth,
Eldon Building,
Portsmouth, PO1 2DJ
United Kingdom
Email: {mel.krokos, benjamin.kyd}@port.ac.uk

***Abstract*—The Visualization Gateway provided by NEANIAS is a novel gateway with the goal to provide onboarding to the European Open Science Cloud (EOSC), interoperability of data provided by different services and Gateways as well as the EOSC and to address some key issues regarding the visualization of big data. The uniqueness of the Visualization Gateway is enhanced by the feature-set and the deployment rationale employed on the NEANIAS Visualization Gateway.**

***Keywords*—science gateways, visualization, big data, European Open Science Cloud (EOSC)**

I. Introduction

The European Open Science Cloud (EOSC) [1] is an emerging virtual environment underpinning the handling of high complexity volumes of information generated by the big data revolution. The EOSC roadmap foresees a seamless federation of existing and future research data infrastructures, under the umbrella of a common policy to enable FAIR [2] datasets to be utilized and shared throughout entire value chains.

NEANIAS [3] is an H2020 research and innovation action addressing prototyping of new innovative services for the underwater, atmospheric and space research sectors. The action embarks from TRL 6 software solutions, to develop, validate and deliver a portfolio of cutting-edge thematic [4] as well as reusable core services maturing into TRL 8 for EOSC ecosystem onboarding.

We discuss our experiences in constructing a novel visualization gateway [5] developed within the NEANIAS core services aiming at providing a unified visualization platform in contrast to existing solutions that, although effective, they are often intrinsically tightly coupled with highly specialised communities, e.g. [9], [10].

Our core guiding principles are to prototype a general purpose cloud based visualization ecosystem to: (i) handle complex and heterogeneous datasets, (ii) support multiple visualization strategies and (iii) enable intuitive and user-friendly data exploration mechanisms.

Our system deploys JupyterHub [11] to provide a modular foundation for inserting bespoke high performance visualization services to allow users to design, construct, realise and fully validate bespoke workflows (e.g. by mixing and matching diverse visualization components to support different levels of composability and integration) to process and analyse complex datasets.

Notebook servers are spawn on the cloud, streamlining end users deployments by avoiding manual installation of software tools locally and eliminating the necessity for taking care of complex software dependencies and housekeeping mechanisms. Our science gateway is integrated seamlessly with several cloud based services e.g. for authorisation and authentication, storage allocation and data sharing, logging, monitoring and accounting, as well as open publication of resulting data products, thus making it fully compliant and ready for onboarding the emerging EOSC ecosystem.

Currently our gateway is populated with the VisIVO [12], [13] and Splotch [14], [15] services for data intensive visual discovery and is nearing its final release under the NEANIAS development infrastructure [16]. The underlying philosophy focuses on usability and tools for seamless embedding within end-user workflows and interlinking with other NEANIAS core services, e.g. for visualization of AI-powered solutions, underpinning powerful virtual reality solutions or facilitating end-user data accessibility. The gateway was prototyped using KinD [21] originally in a local setting and prior to releasing within the GARR Container Platform [22].

The web interfaces are accessible via an authentication and authorization service based on Keycloak [23] and start containerized environments that are exploited via notebooks through interactive python demonstrations. Prior to spawning a notebook, the system offers the option to mount the spawned environments to a Nextcloud [24] store via WebDAV.

Integration with accounting and logging services is realized through Filebeat [26] to pass on Kubernetes logs, integration with monitoring service is realized with Prometheus [28] to collect monitoring metrics, while linkages with Zenodo [27], towards Open Science integration, are underway. We outline lessons learned and discuss development guidelines, with pointers to future works.

II. NEANIAS Core Services

The underlying philosophy of core services is to exploit synergies across diverse user communities (starting from the specific thematic clusters in NEANIAS) to establish an ecosystem of generic flavour offerings which is compliant with open science principles [17] and streamlines access to cloud resources thus paving the way for migration to EOSC [1]. Core services are being developed in distinct and complementary clusters of specific functionalities, pertaining to open science lifecycles (C1 services[1]), EOSC integration (C2 services[2]), artificial intelligence (C3 services[3]) and finally, visualization (C4 services[4]).

The core services ecosystem in NEANIAS is manifested as a service-oriented architecture supporting mechanisms for various degrees of composability and integration (vertical as well as horizontal - see section IV), fully embracing the REST paradigm [18] and FAIR principles [19]. We deploy microservices to enforce single responsibility principles [20], thus decoupling components for reusing, i.e. mixing and matching as needed, according to the intreoperability scenario on hand. REST designs are commonly deployed in EOSC for optimal offerings in terms of modifiability, performance and scalability. Finally, FAIR in core services ensures data compliant with findable, accessible, interoperable and reusable mechanisms.

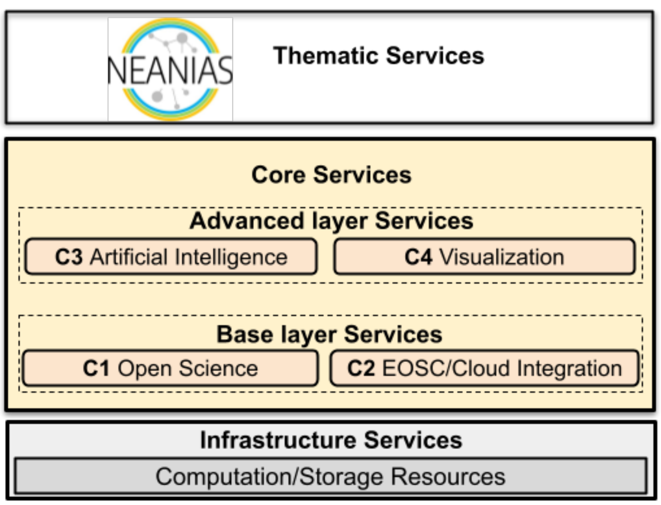

Fig. 1. NEANIAS core services ecosystem

The NEANIAS core services ecosystem is built upon a cluster of low level infrastructure services (for computation and storage cloud resources). They are structured in a hierarchical way supporting base (C1 and C2) and advanced (C3 and C4) layers (see Figure 1). Base provides tooling for FAIR compliance and integration with EOSC and cloud resources, e.g. service[5] and data[6] catalogues. Service providers can access components to support common UX designs. Several loosely coupled components are provided assembling virtual infrastructures to serve other services, facilitating integration with EOSC resources and processes and other cloud computing infrastructures, e.g. authentication and authorisation[7], log aggregation[8] and accounting[9], as well as data depositing, sharing and exploration.

The advanced core services layer supports generic service offerings built upon the base layer to be exploited across different user communities for artificial intelligence processing (C3) and advanced visualization solutions (C4). C3 services represent the elements of typical machine learning workflow lifecycles, e.g. initially a model needs to be designed and implemented, then trained served and deployed using dedicated resources. C4 services underpin a diverse range of multi-faceted visualization workflows[36], from 2D/3D spatio-temporal to composite 2D/3D visuals for complex, high dimensionality datasets, including computationally demanding cross-reality applications.

The NEANIAS core services are currently undergoing their final service release towards full EOSC onboarding [16]. The next section discusses details of our implementation of the NEANIAS visualisation gateway which is a TRL8 EOSC compliant ecosystem for realisation of complex visualization workflows.

III. Core Visualization Gateway

The Visualization Gateway is based on JupyterHub [11] with the goal to provide scientists with access to visualization tools such as Splotch [14], [15] and VisIVO [12], [13] which are easily accessed through the simple user interface based on JuputerHub [11]. The gateway was guided by the NEANIAS core guiding principles [29], which implement FAIR principles of the handling and metadata of the data that we employ.

A. *Gateway Architecture*

The Visualization Gateway enables visualization workflows based on the visual discovery framework [6] and accesses to the mechanisms exposed through the XR (Cross-Reality) frameworks to support virtual reality [7] and the spatial data store services [8].

Figure 2 shows the interdependencies and relationships with the base core services in the NEANIAS ecosystem. The two visual discovery services VisIVO and Splotch are exposed to the end-users via the Visualization Gateway that interlinks also with other NEANIAS core services, e.g. C3 for visualization of AI-powered solutions, Cross Realities framework to underpin powerful virtual reality solutions and Spatial Data Stores to facilitate end-user data accessibility. All services are integrated

[1] https://docs.neanias.eu/en/latest/c1-services.html#c1-services
[2] https://docs.neanias.eu/en/latest/c2-services.html#c2-services
[3] https://docs.neanias.eu/en/latest/c3-services.html#c3-services
[4] https://docs.neanias.eu/projects/c4-services/en/latest/
[5] https://catalogue.neanias.eu/home
[6] https://zenodo.org
[7] https://docs.neanias.eu/projects/aai-service/en/latest/
[8] https://docs.neanias.eu/projects/logging-service/en/latest/README.html
[9] https://docs.neanias.eu/projects/accounting-service/en/latest/README.html

with C2 services for AAI, logging, accounting, data sharing and computing access.

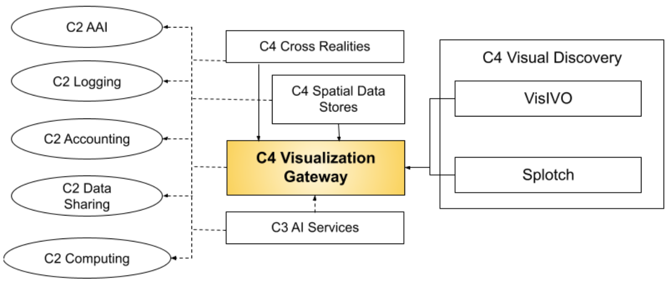

Fig. 2. NEANIAS C4 Visualization Gateway architecture diagram and integration with other NEANIAS Core services.

The gateway offers a framework for the individual applications to sit on top of to enable support for more rich data analysis, as well as providing tools to support data intensive computing for visual scientific discovery.

B. Gateway Implementation

The Visualisation Gateway is built on top of JupyterHub [11] for its use by the scientific community. It can easily spawn hubs with different environments which is especially useful for the NEANIAS use case as the Visualization Gateway includes Splotch, VisIVO and ADAM-API (the main interfaces to the Spatial Data Stores) which are containerised using Docker [30]. Due to JupyterHub's ability to be containerised so well, it can be run on technologies such as Kubernetes, which allows other applications to run alongside JupyterHub, so the Visualisation Gateway runs a FileBeat data sharing pod alongside [31]. FileBeat is a core technology of the Visualisation Gateway because it allows us to connect to the NEANIAS Logging and Accounting infrastructure which is a core requirement of the Visualisation Gateway. JupyterHub allows a user to write a script to interact with a visual discovery tool in order to visualise their data on the cloud, connecting to data sharing services like Zenodo [27] to make it easier.

The core services containerised on the Visualisation Gateway includes Splotch, VisIVO and ADAM-API as Docker containers, these docker containers are stored on GitLab with the mounting of demo data files and demo notebook files to JupyterHub through NextCloud [24] which allows them to be pulled by JupyterHub as needed to be presented to the user as a programmable user interface in the form of JupyterHub Notebooks. These services are provided by the University of Portsmouth and INAF and allow visual discovery for intense data sets. Figure 3 illustrates sample screenshots of the gateway, including the main page where users can set some server options (i.e. choose which service to use) and a notebook sample. In particular the main page allows the selection of the currently available visualization environments to start, e.g.: VD-Splotch, VD-VisIVO and DS-AdamSpace and the two different modes available to mount the NextCloud data store, i.e. default read-only mounting of a file store containing demo data and notebooks or mounting a user specified file store via WebDAV protocol.

The Visualisation Gateway is deployed on the GARR cloud platform on the NEANIAS Kubernetes cluster [22] (YAML files are available on GitLab[10]) this operational dependency is managed by NEANIAS with hopes of onboarding onto the EOSC in the future (see next Section V).

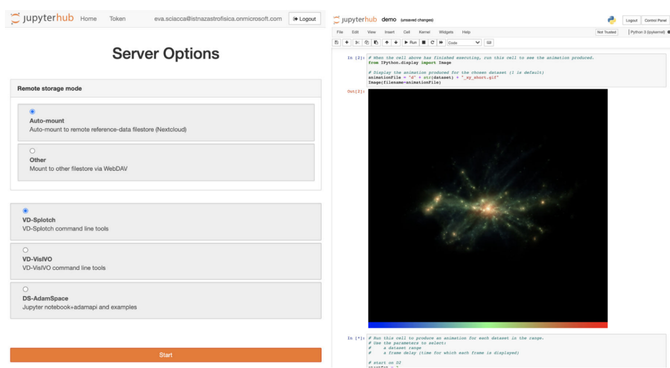

Fig. 3. The NEANIAS Visualization Gateway main page on the left and a sample notebook integrating VD-Splotch environment on the right.

C. Use Cases

The Visualization Gateway enables a number of use cases and visualization workflows, the main ones are being summarized in the following section.

Particle based visualization is enabled by the VD-Splotch environment, which makes available the suite of tools provided by Splotch for data processing and visual discovery. Splotch supports very large-scale datasets and an array of diverse parallelization models for fast, high-quality distributed volume rendering of particles, coming from numerical simulations in many file formats, e.g., smoothed particle hydrodynamics from astrophysical simulations.

Multi-dimensional visualization is enabled by the VD-VisIVO environment, which allows the visualization pipelines available within VisIVO for exploration of large-scale datasets through highly customized 3D views of multidimensional data tables from various sources.

Both use cases are enabled to connect with the Cross Realities toolkit for accessing the Data Connector Service, that is employed for retrieving, in a generic way, data coming from different sources (individual files or databases) e.g. stellar catalogues in astrophysics.

Spatial Data-store visualization is enabled by the DS-AdamSpace environment, and it access to a set of reference systems and data structures by means of data referencing, access, and retrieval services provided by the Spatial Data Store, in a standardized manner based on positional/location criteria. This is useful to navigate maps related to the planet Earth or other planetary bodies such as Mars.

[10]https://gitlab.neanias.eu/c4-service/vg

Aforementioned use cases showcase the main features by automatically importing demo data and demo interactive notebooks to test and try by each user authenticated within the platform.

IV. Related Works

Web based portals and technologies are beying developed to provide multidimensional visualization strategies on complex and heterogeneous datasets.

VisIVO Science Gateway [34] is wrapped around WS-PGRADE/gUSE [35] integrating services for processing and visualizing large-scale multi-dimensional astrophysical datasets on Distributed Computing Infrastructures.

ParaViewWeb[11], allows users to remotely connect via web browsers to a ParaView[32] server, a large scale parallel visualization software, designed for effective exploitation of high performance infrastructures.

WebViz [33] is a multi-user, client-server, visualization framework with a web-based client offering services in the cloud and is accessible via netbooks, smartphones, and other web enabled mobile devices.

The core visualization gateway, developed within NEANIAS, offers such a visual discovery platform based on JupyterHub allowing users to write and execute code, to transform, analyse, and visualize datasets by supporting several high performance visualisation services integrating seamlessly with cloud based mechanisms for authorisation and authentication, storage allocation and sharing, logging and accounting and open publication of resulting data products, thus making it ready for onboarding the emerging EOSC ecosystem.

V. EOSC Interoperability and Onboarding

The Visualization Gateway developed within NEANIAS results in accomplishing different aspects of technical interoperability recommendations, as defined in the EOSC Interoperability Framework [37] and in the European Interoperability Framework, namely:

- Use **open specifications**. This has been accomplished by standards and protocols implemented within the gateway such as OAuth 2.0 for authorization, JSON for logging metrics.
- Define a common **security and privacy framework** and establish processes to ensure secure and trustworthy data exchange. The integration with NEANIAS AAI ensure protected access across communities.
- **Easy access to data sources** available in different formats. This is allowed by integrating NEANIAS Data Sharing service to handle demo data and notebooks to be available across communities, and to tools (e.g. VD-VisIVO and VD-Splotch) enabling the usage of these data.
- There should be a clear **EOSC PID policy**, accommodating any appropriate PID usage. This will be made available by finalizing the integration with Zenodo.

[11]ParaViewWeb:https://kitware.github.io/paraviewweb/

These aspects have been accomplished thanks to the different types of compositions and integration implemented to ensure links with both EOSC Core and EOSC Exchange services that are the fundamental building blocks in EOSC[38].

A. Integrations with EOSC-Core

The EOSC-Core provides the functionality that is required to enable open science practices to occur across domains and countries according to the EOSC interoperability framework.

The Visualization Gateway has been composed with resources available within the EOSC-Core to make it interoperable in EOSC (e.g. NEANIAS AAI). Furthermore, the gateway is enriched with horizontal added value services, such as cloud computing services, providing additional features and easy/elastic/on-demand access to EOSC resources (e.g., the GARR container computing service).

B. Integrations with EOSC-Exchange

EOSC-Exchange is a digital marketplace that builds on the EOSC-Core to offer a progressively growing set of services. Those services are expected to exploit FAIR data and encourage its reuse by publicly funded researchers. It is expected that services, such as those that store, preserve or transfer research data as well as those that compute against it, will be made available via EOSC-Exchange.

The Visualization Gateway is interoperable horizontally with other resources from different scientific clusters by its integration with the NEANIAS Data Sharing service which allows to make data and products interoperable between different services. Finally, the gateway is able to compose scientific resources to create added value solutions to handle complex scientific workflows (e.g. data from the XR-toolkit are reached and visualized using VD-VisIVO within the gateway).

C. EOSC Onboarding

The aforementioned activities will allow to proceed with the process of onboarding the gateway to the EOSC Portal[12] provided by University of Portsmouth as resource provider and INAF as collaborator. The process will be facilitated by employing the NEANIAS Service Catalogue[13] that already contains all the needed information and metadata[14].

VI. Conclusion and Future Works

The NEANIAS project addresses the prototyping of new innovative solutions in EOSC, driving the co-design, delivery, and integration of thematic and core services, derived from state-of-the-art research assets and practices in three sectors (underwater, atmospheric and space research) to create a collaborative ecosystem promoting Open Science principles and practices and contributing to the effective materialization of the EOSC landscape services.

As a cornerstone of scientific discovery, data visualization is crucial for unlocking new discoveries and generating knowledge. The big data paradigm sweeping in almost

[12]EOSC Portal: https://eosc-portal.eu/

[13]NEANIAS Service Catalogue: https://catalogue.neanias.eu/home

[14]https://catalogue.neanias.eu/service/inaf.visualization_gateway_vg

every scientific discipline requires innovative and efficient data visualization solutions. In this work we presented the NEANIAS core visualization gateway service tailored for visual discovery, and integrating with cross-reality services and spatial data stores.

The development of NEANIAS services is being finalized and will be completed by the end of 2022. Till the final delivery validation sessions to gather user feedback and suggestions as well as webinar events are planned to further engage the scientific community and enlarge the visualization wokflows and use cases currently available.

Acknowledgment

The research leading to these results has received funding from the European Commission's Horizon 2020 research and innovation programme under the grant agreement No. 863448 (NEANIAS).

References

[1] https://www.eosc- portal.eu
[2] https://ec.europa.eu/info/sites/info/files/turning fair into reality 0.pdf
[3] https://www.neanias.eu/
[4] https://catalogue.neanias.eu
[5] https://docs.neanias.eu/projects/c4-services/en/latest/services/vg.html
[6] https://docs.neanias.eu/projects/c4-services/en/latest/services/vd.html
[7] https://docs.neanias.eu/projects/c4-services/en/latest/services/xr.html
[8] https://docs.neanias.eu/projects/c4-services/en/latest/services/ds.html
[9] https://www.digitalrocksportal.org
[10] https://openmsi.nersc.gov/openmsi/client/
[11] https://jupyter.org/hub
[12] https://github.com/VisIVOTeam/VisIVOServer
[13] https://docs.neanias.eu/projects/c4-services/en/latest/services/vd.html#vd-visivo
[14] https://github.com/splotchviz/splotch
[15] https://docs.neanias.eu/projects/c4-services/en/latest/services/vd.html#vd-splotch
[16] Krokos, M., et al.: NEANIAS Deliverable: D6.8 Core services software release report #3. Tech. rep. H2020 NEANIAS Project (2022)
[17] Sciacca, E., et al.: NEANIAS Deliverable D6.4 Core Services Architecture, Design Principles and Specifications (update). Tech. rep. H2020 NEANIAS Project (2021)
[18] Feng, X., Shen, J., Fan, Y.: Rest: An alternative to rpc for web services architecture. In: 2009 First International Con- ference on Future Information Networks, pp. 7–10. IEEE (2009)
[19] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Apple- ton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al: The fair guiding principles for scientific data management and stewardship. Scientific Data 3(1), 1–9 (2016)
[20] Martin, R.: Clean Architecture: A Craftsman's Guide to Software Structure and Design. Prentice Hall (2018)
[21] https://kind.sigs.k8s.io
[22] https://cloud.garr.it
[23] https://www.keycloak.org
[24] https://nextcloud.com/
[25] http://www.webdav.org/
[26] https://www.elastic.co/beats/filebeat
[27] https://zenodo.org/
[28] https://prometheus.io/
[29] https://www.neanias.eu/images/neanias/Deliverables/D61_M6Architecture-Specs2020-05-13.pdf
[30] docker.com/
[31] https://www.elastic.co/beats/filebeat
[32] Ahrens, J., Geveci, B., Law, C. (2005). Paraview: An end-user tool for large data visualization. The visualization handbook, 717(8).
[33] Zhou, Y., Weiss, R. M., McArthur, E., Sanchez, D., Yao, X., Yuen, D., et al. (2013). Webviz: A web-based collaborative interactive visualization system for large-scale data sets. In GPU Solutions to Multi-scale Problems in Science and Engineering (pp. 587-606). Springer, Berlin, Heidelberg.
[34] Sciacca, E., Bandieramonte, M., Becciani, U., Costa, A., Krokos, M., Massimino, P., et al. (2013, June). VisIVO Science Gateway: a Collaborative Environment for the Astrophysics Community. In IWSG.
[35] Kacsuk, P., Farkas, Z., Kozlovszky, M., Hermann, G., Balasko, A., Karoczkai, K., Marton, I. (2012). WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities. Journal of Grid Computing, 10(4), 601-630.
[36] Sciacca, Eva, et al. Scientific Visualization on the Cloud: the NEANIAS Services towards EOSC Integration. Journal of Grid Computing 20.1 (2022): 1-18.
[37] European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K., et al., EOSC interoperability framework : report from the EOSC Executive Board Working Groups FAIR and Architecture, Publications Office, 2021, https://data.europa.eu/doi/10.2777/620649
[38] Barbarossa, Emanuele, et al. "Solutions for a sustainable EOSC: a FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group." (2020): 43.