A Data Management Infrastructure for Reproducible Experimental Research

Ziyad Mabrouk, Bohdan Dyshlevyy, Damien Saucez
Centre Inria d'Université Côte d'Azur
{first.last}@inria.fr

Panagiotis Hadjidemetriou, Panayiotis Andreou
University UCLan Cyprus
{PHadjidemetriou2,PGAndreou}@uclan.ac.uk

***Abstract*—Cloud, Internet, and network infrastructures are now tightly integrated, making experimental research increasingly dependent on advanced technical expertise and real-system deployments. As a result, purely theoretical evaluations are often insufficient to capture real-world behavior. This shift has driven the rise of data-driven research, where hypotheses are validated using high-quality experimental data. However, such data are frequently retained by their producers, limiting reproducibility and reuse.**

The EU SLICES-RI project addresses this challenge by providing a scientific platform that supports reproducible, large-scale experiments on state-of-the-art hardware and software and enables the systematic sharing of experimental data and metadata across the research community.

This demo presents the SLICES-RI Data Management Infrastructure and its Metadata Registry System, demonstrating how lightweight tools can be used to efficiently publish and share experimental results. This Metadata Registry System has already been adopted by other European projects, such as Converge, 6GXCEL, and 6G-Sunrise.

I. Introduction

Cloud computing, Internet services, and network infrastructures have become deeply intertwined, forming complex, distributed systems that are difficult to model accurately using purely analytical methods. Modern networked applications operate across heterogeneous environments. As a result, experimental research increasingly requires highly technical expertise to deploy, configure, and evaluate new ideas on real systems. This has led to challenges that have been widely recognized in the networking and systems research communities [1], [2].

To address this gap, data-driven approaches have gained in importance. Researchers now rely on high-quality experimental data to validate hypotheses, compare alternative designs, and support reproducible science [3]. Large-scale measurement studies, controlled testbed experiments, and trace-driven evaluations have become essential tools for understanding system performance, scalability, and reliability under realistic conditions. However, despite their importance, high-quality experimental datasets are often kept internal by the organizations that produce them, due to concerns related to confidentiality, competitive advantage, or the lack of standardized mechanisms for data publication, thereby limiting reproducibility [4].

SLICES-RI[1] is the first community-driven scientific instrument in digital sciences, aiming to support researchers and engineers in conducting experiments and generating data in real-world settings [5]. It offers experimentation services across a wide range of Digital Infrastructure technologies, including programmable radio, Edge and Cloud computing, and GPU/DPU-assisted AI model validation.

SLICES-RI is often perceived as a distributed testbed for running experiments, but SLICES-RI goes far beyond this single offering. An often underappreciated aspect is that SLICES-RI enables the publication, management, and sharing of experimental data and associated metadata through its *Metadata Registry System* (MRS) [6], even when the data are not generated within the SLICES-RI infrastructure. The MRS has been intentionally designed to be highly flexible, which is why it has already been adopted by the Converge [7], 6GXCEL [8], and 6G-Sunrise [9] European projects.

The demonstration will showcase the data and metadata system offered by SLICES-RI and illustrate how researchers can leverage it to share their data at scale while automatically complying with the FAIR principles [10], thereby fostering more reproducible research.

II. Data and metadata Management with SLICES-RI

SLICES-RI develops a state-of-the-art digital research infrastructure aligned with Europe's open science policy, addressing the critical role of metadata in ensuring reproducibility, interoperability, and the reuse of scientific results. To overcome the limitations of single or unified domain-specific metadata standards in a multidisciplinary context, SLICES-RI adopts a hierarchical metadata profile scheme that combines domain-agnostic metadata for uniform discovery with type- and domain-specific metadata to enhance machine-actionability, enabling seamless access and use of digital objects by humans and automated services.

The *Metadata Registry System (MRS)* is a central component of SLICES-RI, enabling the unified management and discovery of metadata for datasets, publications, software, tools, and internal resources while ensuring compliance with FAIR principles [10] to support findability, accessibility, interoperability, and reusability across the SLICES-RI ecosystem [11]. The MRS is organized around a hierarchical metadata model, providing flexibility and applicability across diverse research domains, as illustrated in Fig. 1.

[1]**S**cientific **L**arge **S**cale **I**nfrastructure for **C**omputing/Communication **E**xperimental **S**tudies (SLICES) - **R**esearch **I**nfrastructure (RI)

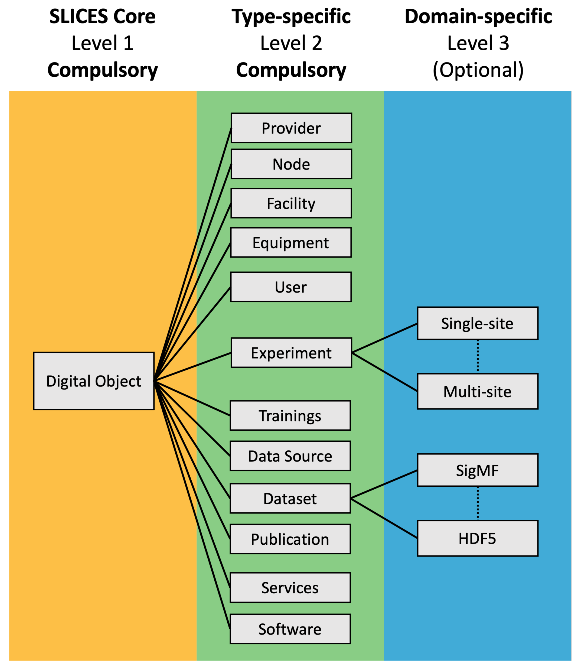

Fig. 1. SLICES-RI hierarchical metadata model.

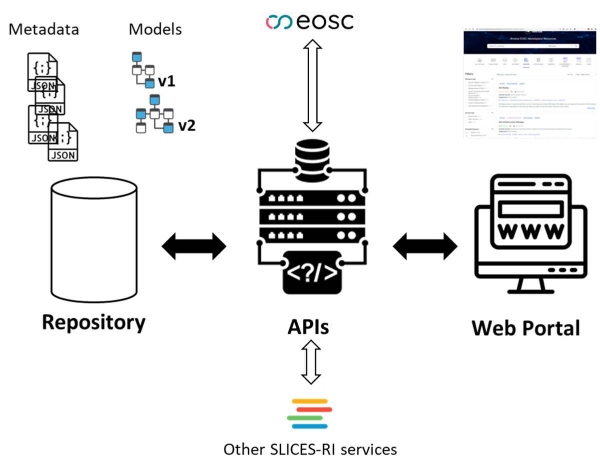

Fig. 2. Metadata Registry System (MRS) components.

MRS provides access and management services to digital objects using three components as illustrated in Fig. 2. First, the metadata are persisted in a repository implemented as a PostgreSQL database. Second, a backend is responsible for exposing the repository as a REST API, providing authentication/authorization, backward compatibility, version maintenance, and other functionality. Finally, a web portal is provided to facilitate human interaction with MRS.

The SLICES-RI MRS web portal shown in Fig. 3 is intuitive for exploring and searching data. It offers a structured, metadata-driven interface for discovering and managing resources across the infrastructure. Its advanced query engine supports field-level filtering, rich comparison operators, array-aware conditions, and structured queries over complex

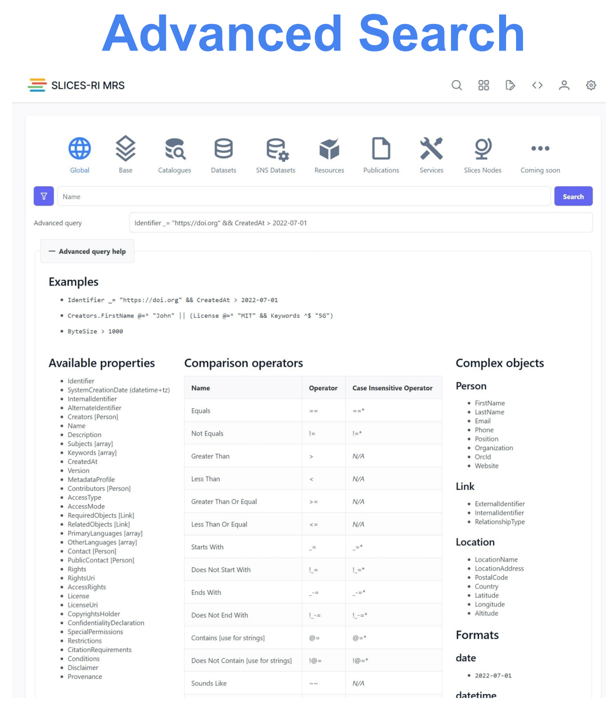

Fig. 3. Example of advanced search with the Metadata Registry System Web UI.

objects. However, it becomes cumbersome for publishing results during ongoing experiments and may introduce human error. To address this, a command-line interface (CLI) helper tool, called *dm*(Data Management), is provided [12]. This tool is implemented as a sub-command of the SLICES CLI [13](Command Line Interface). The SLICES CLI allows users to perform common SLICES-RI operations, such as publishing datasets, creating virtual machines, configuring networks, or authentication, and can be used either on testbeds or directly on users' machines.

The *dm* sub-command supports two modes: an *interactive assistant mode*, which guides the user step-by-step to publish data and metadata (e.g., *slices dm publish dataset*), and a *batch mode*, which leverages pre-filled manifests to automate metadata submission (e.g., *slices dm publish dataset –json-file metadata.json*). The tool also automatically uploads the data linked to the metadata to the SLICES-RI DMI (Data Management Infrastructure), a MinIO-based S3 service provided by SLICES-RI.

A distinctive property of the *dm* sub-command is its fully dynamic, runtime generation. Each time it is executed, it retrieves the current API specifications from the MRS and generates the corresponding validation logic and digital object support on the fly. This design ensures that the CLI remains synchronized with the MRS API, so any addition or modification of a digital object type is automatically supported without requiring manual updates to the tool.

III. Demo Storyboard

The demonstration is based on SLICES-RI and leverages all previously described services to showcase the complete workflow for data collection, publication, and dissemination.

For each execution of the demonstration, we will begin with a concise overview of SLICES-RI, with particular emphasis on its data and metadata management facilities. We will then provide a guided walkthrough of the MRS web-based user interface, illustrating how to discover data resources and how to create digital objects. Subsequently, we will examine the REST API specification to highlight the system's expressiveness and functional richness.

We will then proceed to create and publish a dataset using the *dm* sub-command, first in interactive mode and subsequently in batch mode.

For participants interested in reproducing the workflow, we will finally demonstrate how to install the *dm* sub-command on their own machines and how to perform authentication against the system.

This demonstration complements the study [14], accepted for publication at this conference, which leverages the SLICES-RI infrastructure to deploy 5G networks and monitor their performance. However, since data publication is of broader interest to the community, we propose a dedicated presentation of SLICES-RI data management through this demonstration.

IV. Demo setup

The demonstration relies on the SLICES-RI infrastructure; therefore, Internet connectivity is required. The firewall must allow access to TCP ports 80 and 443; access to TCP port 22 would be beneficial.

Most parts of the demo rely on Web UIs, but some command-line executions will be demonstrated; therefore, a large screen with good resolution (HDMI connection is preferred) is required, as well as a table with at least one available power outlet for the demonstration device.

Space in the booth for a SLICES-RI kakemono would be appreciated.

V. Conclusion

This demonstration presents the SLICES-RI Data Management Infrastructure and its Metadata Registry System as practical enablers of reproducible experimental research. By combining a flexible hierarchical metadata model, a programmable REST API, and a dynamically generated CLI tool, SLICES-RI provides researchers with lightweight yet robust mechanisms to publish, manage, and share experimental data and associated metadata. Through this workflow, the demo highlights how data publication can be seamlessly integrated into experimental practices while ensuring FAIR compliance and long-term reusability across research communities.

References

[1] S. Floyd and V. Paxson, "Difficulties in simulating the internet," *IEEE/ACM Transactions on networking*, vol. 9, no. 4, pp. 392–403, 2002.

[2] W. Willinger, D. Alderson, and J. C. Doyle, "Mathematics and the internet: A source of enormous confusion and great potential," *Notices of the AMS*, vol. 56, no. 5, pp. 586–599, 2009.

[3] T. Hey, S. Tansley, K. M. Tolle, *et al.*, *The fourth paradigm: data-intensive scientific discovery*, vol. 1. Microsoft research Redmond, WA, 2009.

[4] D. Saucez and L. Iannone, "Thoughts and recommendations from the acm sigcomm 2017 reproducibility workshop," *SIGCOMM Comput. Commun. Rev.*, vol. 48, p. 70–74, Apr. 2018.

[5] S. Fdida *et al.*, "Slices, a scientific instrument for the networking community," *Computer Communications*, vol. 193, pp. 189–203, 2022.

[6] Y. Demchenko, S. Gallenmüller, S. Fdida, T. Rausch, P. Andreou, and D. Saucez, "Slices data management infrastructure for reproducible experimental research on digital technologies," in *2023 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2023.

[7] CONVERGE Consortium, "Converge project." https://converge-project.eu/, 2026. Accessed: 2026-01-22. Funded under the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101094831.

[8] 6G-XCEL Consortium, "6g trans-continental edge learning (6g-xcel) project." https://www.6g-xcel.eu/, 2026. Funded by the European Union's Horizon Europe programme under Grant Agreement No. 101139194; accessed 2026-01-22.

[9] SUNRISE-6G Consortium, "Sunrise-6g: Sustainable federation of research infrastructures for scaling-up experimentation in 6g." https://sunrise6g.eu/, 2026. Accessed: 2026-01-22. Co-funded by the European Union under Grant Agreement No. 101139257 (Horizon Europe, SNS JU).

[10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, "The fair guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160018, 2016.

[11] SLICES Research Infrastructure, "Mrs (metadata registry system)." https://doc.slices-ri.eu/BasicServices/MRS/MRS.html, 2026. Accessed: 2026-01-22.

[12] SLICES Research Infrastructure, "Slices data management." https://gitlab.inria.fr/slices-ri/resources/shared/slices-cli-dm, 2026. GitLab repository (accessed 2026-01-22).

[13] SLICES Research Infrastructure, "Slices cli (command line interface) documentation." https://doc.slices-ri.eu/SupportingServices/slicescli.html, 2026. Accessed: 2026-01-22.

[14] Y. Amami, Z. Mabrouk, C. Barakat, and T. Turletti, "Toward real-time ran observability in open-source 5g systems," in *The 29th Conference on Innovation in Clouds, Internet and Networks (ICIN 2026)*, 2026.