slices SC

HORIZON 2020

H2020 - INFRAIA-2020-1

D8.6 Report on 1st Period of Virtual Access, including User Committee's assessment

Acronym	SLICES-SC
Project Title	Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies – Starting Community
Grant Agreement	101008468
Project Duration	42 Months (01/03/2021 – 31/08/2024)
Due Date	30 November 2023 (M33)
Submission Date	28 June 2024 (M40)
Authors	Serge Fdida (SU), Emile Mespoulhes (SU), Nikos Makris (UTH), Virgilios Passas (UTH), Stavroula Maglavera (UTH), Bartosz Belter (PSNC), Andrea Passarella, Raffaele Bruno (CNR), Walid Dabbous (INRIA), Barbara Tóth (SZTAKI), Kacsuk Péter (SZTAKI), Ákos Hajnal (SZTAKI), Márk Emődi (SZTAKI), Krisztián Póra (SZTAKI), Sebastian Gallenmüller (TUM), Georg Carle (TUM), Carmen Guerrero, Ignacio Berberana (IMDEA), Konstantinos Filis (OTE), Antti Pauanne (UOULU).
Reviewers	Serge Fdida (SU), Barbara Tóth (SZTAKI)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101008468. The information, documentation and figures available in this deliverable, is written by the SLICES-SC project consortium and does not necessarily reflect the views of the European Commission. The European Commission is not responsible for any use that may be made of the information contained herein.

Scientific Large-scale Infrastructure for Computing Communication Experimental Studies **Starting Communities**

www.slices-sc.eu

Executive Summary

Research data sharing and access is yet not mature in the communities covering the domains covered by SLICES. There is very little incentive to share research data for the many reasons that have been properly identified. SLICES has the potential to change the research methodology by providing a simple access to data sharing and a reward for those sharing their data.

Developing a relevant solution (MRS/SFDO) takes time thus this process is currently in progress and should be made available in its basic form for the SLICES pre-operation that will start in Q4 2024. In the meantime, acknowledging that it is a slow start, we have developed a preliminary solution, that has a value to raise awareness and kickstart the process. In order to better understand how it should develop in the future we have conducted a survey, collected the advice from the SLICES User Committee, furthermore, we provided a self-assessment of our solution and how it should develop in the future.

Virtual Access (VA) to the SLICES Research Infrastructures is an open and free access provided to the Starting Community of users and researchers. The provision of VA to the SLICES-RI is aided through datasets and tools that allow the preparatory work for experiments and even also for offline running of experiments. The Transnational Access (TA) (physical or remote access) mainly consists of high-performance computer systems that host bleeding-edge communication and networking technology equipment, testing networks and grids. The VA mainly consists of the exposure of datasets and tools of produced experimentation results.

The VA activities in SLICES-SC act complementary to the TA access, described in Deliverable D8.2 on First Report on Transnational Access, with data and tools produced either through the latter, or created and provided from the testbed operators. Access to Virtual Access services is free-of-charge, and directly findable and discoverable in the SLICES Open Data and Tools repositories (SLICES CKAN[1] and GitLab servers[2] respectively). The datasets are semantically annotated, towards promoting the migration to EOSC-II, and provided in different formats.

In this document, we describe the methodology and results of an assessment of these VA services (reported in the deliverable D8.4 that covers the first period of the project). The SLICES User Committee (UC), composed by external members of the consortium, was initially assigned the task of assessing the VA services. UC has been also involved in the evaluation process of the different open calls for experiments. The initial plan was to have the UC assessment in the final report D8.5 but during the negotiation of the amendment, it was decided to include this new deliverable D8.6 for reporting a broader picture regarding the situation of SLICES VA considering the limited amount of publicly available data and limited maturity and experience of this type of service by our community.

This is a first step in the offering of a VA for SLICES, as we learn by doing, and expect that our future plans will deliver the components to support the full research life-cycle.

[1] SLICES datasets – CKAN sever, https://slices-sc.eu/slices-datasets/ [Last accessed 27 June 2024]

[2] SLICES datasets – GitLab, https://www.slices-ri.eu/slices-gitlab/ [Last accessed 27 June 2024]

Table of contents

Acronyms

AI	Artificial Intelligence
CLI	Command Line Interface
DMI	Data Management Infrastructure
EOSC	European Open Science Cloud
ERA	European Research Area
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
HPC	High Performance Computing
ICT	Information and communication technology
IoT	Internet of Things
ML	Machine Learning
MRS	Metadata Registry System
NDA	Non-Disclosure Agreement
RI	Research Infrastructure
SFDO	SLICES Fair Digital Object
TA	Transnational Access
UC	User Committee
VA	Virtual Access

1. SLICES vision and mission, background and actions

1.1. SLICES vision and mission

The notion behind the User Committee Digital infrastructures (Beyond 5G, Cloud, Edge, wireless, IoT, HPC, etc.) are at the heart of the digital transformation of our societies, including industrial competitiveness and sovereignty. We witness an acceleration of its effect due to the fast exploitation of the infrastructure softwarization, with significant hardware support and an intense use of relevant data. The pace at which this is happening requires a quick adaptation of the skills for building the appropriate capacity. Facing international competition in Europe requires a mutualization of the effort and an aggregation of resources because the investment needed is gigantic.

As a consequence, SLICES will strongly support the implementation of the European research programmes and projects, fitting the demand from the community for a more competitive research, suitable tools, and technological developments. SLICES will liaise with the major European research and industrial entities in digital infrastructures in order to engage the broader community, mobilize advanced and cutting-edge resources, enable a FAIR access to research data and act as a landmark at the best international level.

The mission and vision of SLICES is to build a strong inter-European research community to manage multiple stakeholder-driven research facilities, enabling development of innovation and richer services to EU customers. The SLICES RI acts in accordance to the objectives of the European Research Area (ERA) strengthening a European-wide scientific and technological open base for the research, institutions and businesses to circulate freely, cooperate and promote technical and scientific innovations in the European Union. Currently, most national research facilities that are being operated across Europe offer their standalone resources independently with no common substrate. Integration under the SLICES Research Infrastructure will allow to scale available resources, achieve successful sustainable operation at the European level (ESFRI) and promote infrastructure for multi-stakeholder collaboration.

SLICES will strongly contribute to the positive impact assessment of the European Research Area in the domain of digital sciences by leveraging national investments, coordinating and articulating them with the European ones, providing an open access to the various programmes and projects. Digital infrastructures represent a highly strategic domain with a strong international competition including the one from the US and China giant technology leaders. The investment needed and the human capital requires to act in a coordinated manner at the European level. SLICES will also act as a catalyst for international cooperation, including with emerging countries that also need to acquire and develop these skills with their own capital.

SLICES will encourage and support a free circulation of researchers (either physical or virtual), knowledge, data and other assets that the community will share.

In a field where research and commercial services are driven by data and AI, it is of utmost importance to provide FAIR access to research data that is not available yet and hamper research competitiveness. SLICES will act as a unique repository, providing a simple access for research data and digital objects (training methods for AI/ML, etc.) that will make European research more accessible and competitive. This will follow the open access principles whilst preserving legal, commercial or ethical considerations.

SLICES will support the virtual mobility of researchers and will spin-off a stronger collaboration at the European level, contributing to a better integration of activities through knowledge, infrastructure and data.

Finally, SLICES will engage into a customized dissemination of its activities towards its stakeholders and main target groups in order to make the investment highly valuable to researchers, students, industry and ultimately European citizens. In particular, through its Academy[3], SLICES will develop a suitable corpus of knowledge and skills that will be made openly available to its community, including support for training and hands-on.

1.2. Status and weaknesses

The ability to share research data is becoming a strong asset to support competitiveness for research and innovation. However, the maturity for sharing research data varies with domain sciences, their research practices and scientific methodologies.

It is unfortunate to realize that data sharing in the field of digital infrastructures is still at its infancy and that there exist severe obstacles for improving this situation. This is related to a lack in the management of the full research life-cycle, the wish to keep a competitive advantage by not disclosing the data as well as the effort that a researcher should consent in order to achieve this goal. The immediate impact is the difficulty to support reproducible research that hampers the trust in the research papers that the community is producing.

FAIR principles promote the reuse of data and subsequently enable improvement in reproducibility of research. This indicates that incentives should be provided to encourage the commitment to FAIR research. It means to encourage collaboration among research actors across the whole process as well as to support this methodology as a common practice. Unfortunately, as mentioned above, there exists strong bottlenecks in the domain covered by SLICES to comply with the FAIR objectives.

SLICES wants to fully endorse and adopt the Open Science and FAIR principles, acting as a catalyst to enable and foster cutting edge research, data-driven science and scientific data-sharing. Several design considerations should be taken into consideration, including:

(i) easy and open access to scientific data to facilitate further knowledge discovery and research transparency ensuring the longevity of the data and access to the wider research and innovation community;
(ii) scalable architecture to efficiently leverage a large number of storage resources to support efficient data storage and compute, including highly parallelized data workflows to support experiments;
(iii) privacy preservation methods for ensuring end-to-end security and privacy in compliance with relevant legal frameworks; and
(iv) data quality assurance methods to ensure data quality across multiple dimensions, such as accuracy, completeness and integrity, in order to improve data utility. To address this, SLICES requires to carefully design and develop efficient and scalable data management, analysis and reporting mechanisms, supported by appropriate metadata profiles to cater for access and reuse of FAIR data and services. These tools need to capture and report the entire data lineage/provenance across the data management lifecycle, while also providing the

[3] SLICES Academy, https://moocs-academy.slices-ri.eu/ [Last accessed 27 June 2024]

systematic means for secure and trustworthy interoperability of data and services ensuring the authenticity and immutability of the shared data.

The raise of Artificial Intelligence (AI) can support the ability to make the data FAIR as well as to create synthetic data artefacts, although with the objective to comply with the FAIR principles, assuming that the generated data is subject to quality assessment and solid validation.

It is important to notice that achieving even partly this objective requires a strong engagement of the community, its various stakeholders including scientific societies as well as dedicated staff and competencies.

1.3. SLICES taking action

Despite this situation, SLICES has positioned FAIR data as a strategic component of its development, activating several concurrent actions, from incentivizing the research community, discussing with the stakeholders and developing an appropriate framework to support open science and FAIR access to data. This effort started more than a year ago but requires intense discussion, coordination and development. At the same time, to start with a practical solution, we decided to offer the possibility to share raw data that is stored in our CKAN server.

SLICES started to design SFDO (SLICES Fair Digital Object), a sophisticated metadata model that will allow SLICES users and its ecosystem of platforms and systems to uniformly find, access, and use any digital object, such as data, services, tools, and software. SFDO adopts a hierarchical structure and gracefully combines domain-agnostic metadata for easy and uniform discovery of digital objects by both humans and machines and type and domain-specific information for enhancing machine-actionability, allowing internal and external services to access more complex information about the object and take appropriate actions.

As a consequence, SLICES will setup a data management framework to support the efficient and effective operation of the SLICES infrastructure. To accomplish this, the data management framework is setting its own design goals, including data governance, architecture, quality, metadata, interoperability, analytics, security and privacy.

Our preliminary estimations for SLICES include up to 5,000 users and their data, accounting for up to 50GB per user on the individual nodes and up to 1TB on the cloud. This provides us with a preliminary estimation of 0.25PB-1PB of data storage for all datacenters residing on SLICES nodes, and 5PB for the cloud-based datacenter. We therefore are setting up a preliminary storage service hosted by one of our partners before it moves to the central node as a common service when the central node will enter operation. The storage service will contribute to host our DMI (Data Management Infrastructure) and in particular our MRS (Metadata Registry System). MRS provides access and management services to SFDO using three components. First, the metadata persisted in a repository, implemented as a Postgres database. Second, a backend is responsible for exposing the repository as a REST API while providing authentication/ authorization, backwards compatibility and other functionality. Finally, a web portal is provided to facilitate human interaction with MRS. This is currently under construction and deployment and should be made available with a minimum service for our pre-operation starting Q4 2024.

Since SLICES aims to provide a pan-European experimental research platform by jointly utilizing the geographically dispersed computing, storage and networking RIs, it is highly important that the different RIs interacting in the experimental workflow are interoperable with each other. Similarly, existing research needs to be accessible and directly pluggable to SLICES services and sites. A common

interoperability framework should be adopted across the SLICES ecosystem so that different subsystems have a common understanding of resources and data/metadata are on the same page with respect to the licensing, copyright and privacy requirements.

The SLICES infrastructure is designed to ensure compatibility and integration with EOSC and existing ESFRI infrastructures, and be ready to offer advanced ICT infrastructure services to other RIs and projects, with the special focus on the FAIR data management and exchange. Therefore, it is of utmost importance to design the integration framework of SLICES with EOSC in such a way that the data exchange between SLICES and EOSC is interoperable for scientific workflow management for data storage, processing and reuse. To this end, the recommendations of the EOSC interoperability framework are considered in great detail for the design of the SLICES interoperability framework.

Finally, simply collecting and organizing the metadata can only go so far as to enable reproducibility. Recent advances in Artificial Intelligence and automation are playing an ever-greater role in the scientific process. Amongst the most significant hurdles are (i) the development of robust and flexible workflow automation systems capable of handling diverse configurations and setups across various scientific domains; and (ii) ensuring the reliability, repeatability, and reproducibility of automated experiments, accounting for factors such as equipment variations and environmental conditions which may introduce uncertainties. This is left for future development.

To summarize, despite the lack of maturity for our research community to share data, SLICES decided to add this dimension in its pre-operation phase starting in Q4 2024 as a mean to incentivize the research community to adopt this methodology by offering an “as simple access as possible” solution to data sharing and to ultimately comply with the FAIR principles. This process is ongoing and should materialize before the end of 2024. It is important to note that the same situation (lack of established attitude and maturity for sharing research data) exists at the international level in the field of Digital Infrastructures. However, we observe some initiatives that encourage data sharing and reproducibility although yet in a very primitive form (raw data for instance).

During the period covered by this deliverable, we have evangelized the value of data sharing and its benefit for our community, better understand the challenges and opportunities, prepared our future solution and initialize a first attempt to have our hands on.

2. Virtual Access (VA) in the context of SLICES

2.1. Analysis of Virtual Access in Research Infrastructures

A key element for defining the principles of RI access policy has been laid down in the European Charter for Access to Research Infrastructures[4]. The Charter promotes the harmonization of access procedures as well as the enhanced transparency of access policies adopted by RIs with the final purpose of enabling users to access the best to perform their work, wherever it might be located.

During the execution of the SLICES-SC project, as mentioned in Section 1, during the definition and initial seed of the SLICES VA services, we have addressed many challenges to broader and effective access to SLICES as a RI to be built in the coming years. Future SLICES users will need to tackle scientific and societal challenges and to underpin the future European Research Area (ERA). Within SLICES-SC

[4] European Commission, Directorate-General for Research and Innovation, *European charter of access for research infrastructures: principles and guidelines for access and related services*, Publications Office, 2016, https://data.europa.eu/doi/10.2777/524573 [Last accessed 25 June 2024]

project, we started to identify and manage some of the technical challenges of implementing open virtual access to RI.

SLICES is providing two different types of accesses for the provided Research Infrastructure (RI): Transnational Access (TA) and Virtual Access (VA). Deliverable D8.2 clearly describes the main differences between these access types and defines the global framework for accessing the RI.

This document is focused only in the analysis and assessment of the VA activities that have taken place during the first period and reported in the deliverable D8.4. Overall, SLICES-SC develops tools regarding both access schemes, in order to provide high-quality services to the end users who access the infrastructure. By default, SLICES-SC complies with the EU guides for accessing RIs.

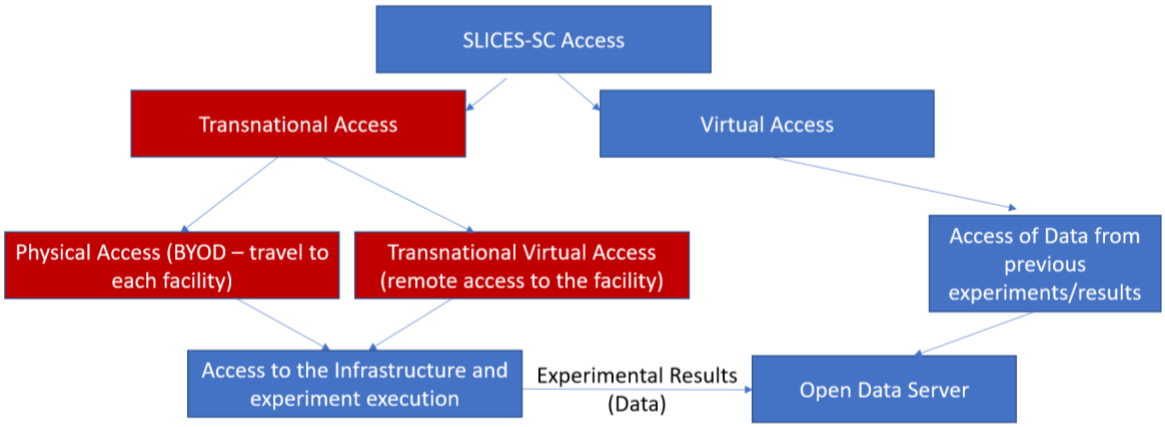


Figure 1. Definition of accesses in SLICES

Figure 1 from Deliverable D8.2 clearly shows the difference between Transnational Access and Virtual Access. Transnational Access requires the real access of the research infrastructure either physically or remotely. In the case of Virtual Access, the user does not use the research infrastructure rather uses the data that is collected during the Transnational Access experiments or provided by the SLICES testbed operators. The advantage of this form of using the SLICES RIs is that access to the data is not limited due to the capacity issues and hence any researcher can access these data freely without the competition of an application procedure. On the other hand, in the case of IT systems it is not so trivial how to generate data that could be useful in the Virtual Access mode and how to use the data produced by someone else's experiments. We have faced difficulties during the first 24 months of SLICES-SC project both in providing the data and especially in using the data.

As described in Section 1, the vision of SLICES for allowing Virtual Access is to offer a pan-European operational networking and computer infrastructure to facilitate scientific research with instrumentation and experimentation capabilities. The data and tools that will be offered through VA will assist in either facilitating easy experimentation with SLICES-RI, or adding up to the knowledge of the community through data sets, analytical results from them and tools.

SLICES has defined an initial set of categories of datasets available through the SLICES Open Data Server:

- Experiment descriptions for advanced and complex experiments with technologies like 5G/6G technologies, wireless sensor devices and IoT, cloud and AI/ML for networking.

- Analytical Datasets from experiments that have been conducted over SLICES-RI. Traces of experiments regarding network optimization, network conditions, performance related data (e.g., throughput, consumption of computing and memory resources per application) allows the community to benefit from their results, and their broader analysis from a higher number of researchers. Such data can trigger further experiments over the platform, towards validating a specific behaviour/developing new protocols and methods.
- Visualization services for directly visualizing experiment outputs on the fly/post the experimentation, enabling the deep analysis and processing of collected data.

2.2. Simple statistics

As we started our service, we did monitor usage, although yet at a low level, as we have to disseminate and exemplify the usage at the same time.

As this service is still in its infancy and SLICES-RI did not yet enter into pre-operation that includes SFDO, the current statistics on our CKAN server and GitLab are still quite low:

Table 1. CKAN server data

CKAN server data	
Number of registered users	16
Number of datasets	66

Table 2. GitLab data

GitLab data	
Number of registered users	5
Number of projects/experiments	10

We expect these numbers to scale up with the opening of SLICES-RI pre-operation that will support VA as well as promote the full research life cycle to SLICES users who want to access our service.

3. User Committee's assessment of the VA

3.1. Objectives of the User Committee

SLICES is organised with the support of a User Committee (UC) that is tailored in order to assist us in better collecting the demand (together with other instruments) as well as supporting the Open Calls and discussing with the open callers. Thus, we have maintained a regular dialogue with the User Committee during the course of SLICES-SC. In addition, the UC will be sustained during SLICES-PP.

The role of the User Committee was to constitute a group of experts of high esteem who represent various scientific disciplines coming from different geographical backgrounds across Europe. The members are not affiliated with any of the SLICES members thus they can act independently. They are responsible for evaluating the experiment proposals sent for the Open Calls and assessing VA services and the corresponding data.

SLICES-SC partners suggested professionals who fit the above-mentioned requirements, and then the management of SLICES-SC approached them and asked them to sign a non-disclosure agreement (NDA) if they were willing to take up the task.

The current members of the User Committee are the following experts:

Table 3. SLICES User Committee members

#	LAST NAME	FIRST NAME	INSTITUTE	TYPE OF INSTITUTE	COUNTRY	GENDER
1	**AUDOUIN**	Olivier	NOKIA	Industry	France	Male
2	**BRETON**	Vincent	Grille / IN2P3	Industry	France	Male
3	**BOUCKAERT**	Stefan	Televic Rail	Industry	Belgium	Male
4	**COUPAYE**	Thierry	Orange lab	Industry	France	Male
5	**GLABOWSKI**	Mariusz	Poznan University of Technology	University	Poland	Male
6	**KISS**	Tamas	University of Westminster	University	United Kingdom	Male
7	**MARTINI**	Barbara Martini	CNIT (National, Inter-University Consortium for Telecommunications)	Consortium of universities	Italy	Female
8	**SIPOS**	Gergely	EGI	International infrastructure	The Netherlands	Male
9	**TALIA**	Domenico	Università Della Calabria	University	Italy	Male
10	**WÄHLISCH**	Matthias	Frei University Berlin	University	Germany	Male

The table above shows the detailed information of the current members. The group is rather heterogeneous: the members come from 7 different countries; there are people from academia and industry. Unfortunately, it is not gender balanced at this stage as it was hard for us to receive positive answers from candidates. We are currently in the process of re-balancing the UC.

The members are expected to give SLICES-RI hands-on feedback about their requirements so we can fine-tune existing and design upcoming services accordingly. SLICES intends to ensure transparency, fairness and impartiality by evoking the User Committee.

3.2. Assessment procedure

To this end, within the Task 8.2 on VA, we have produced a ***survey***[5] that was reviewed and amended by partners of the consortium, targeting the SLICES User Committee representing the key SLICES-RI stakeholders. The short-term objective of this survey is to assess the current version of the SLICES VA services and discuss the challenges to be faced during the next steps in the construction and the

[5] The survey is detailed in the UC Assessment form in the Annex at the end of this document.

operation of the RI. The long-term objective of this survey is to provide a framework for future benchmarking the SLICES VA services.

During this first VA assessment, each UC member had access to the SLICES VA services and to deliverable D8.4 that includes the first report of these services. The UC members made the assessment individually, providing us the assessment in a compiled form.

3.3. UC Assessment results

The outcome of this assessment exercise is presented in this deliverable. Initial recommendations to address VA challenges are analyzed and summarized hereafter.

Some of the SLICES VA services challenges addressed during this UC assessment work are the following:

- Sustainability of the VA services
- IPR policy and data policy
- Data protection and liability
- Strengthen visibility and recognition of SLICES in national and European research systems Prioritization of cutting-edge technology developments
- Interoperability within and across RI/testbeds/nodes
- Harmonization within and across RI/testbeds/nodes
- Standardization
- FAIR principles
- Managing of multidisciplinary needs from users and testbed operators
- Environmental footprint of VA services
- SLICES-RI managing the EOSC
- Innovation (technological transfers, co-development of initiatives)
- Industry-SME collaboration on using and providing VA services
- Assessment of resources needed to provide VA services across RI/testbeds/nodes
- Training and education on VA services (upskilling of RI staff and training of users)

Regarding the evaluation of the technical criteria of the VA, the most valuable comments are regarding the relevance of SLICES VA services for researcher scientists who would like to gain access to computing/data services in different domains. There is also relevance on the scientific domain of datasets available.

When looking into the contribution of the SLICES VA services on building community and the impact on it, the main conclusion is that there is a potential impact in the aggregation and combination of VA services provided by the testbeds and nodes integrated currently in SLICES, and that may help many researchers. This confirms the general assumption that sharing research data will be highly valuable, in particular for the SLICES community, where this type of service does not exist (or is poorly represented).

The evaluation of the FAIR criteria concludes that data and tool findability and accessibility mechanisms meet the potential user requirements. The importance of the interoperability aspect of the VA services is also highlighted. A recommendation concerning this aspect is the importance of allowing the spanning applications on different sites. The reusability in SLICES is very useful and it is highly appreciated in research activities.

The metadata criteria were also included in the assessment and the UC members. It finds that the model used is very relevant and it provides users with a standard approach in understanding and exploiting available datasets. They also highlighted the promising impact on EOSC users.

The UC also emphasized major issues that the VA could face, in particular, new situations that can occur where experiments are carried out outdoor or in public spaces, where privacy is a concern. This might be found even more complex as regulation varies from continent to continent. Furthermore, other difficulties may come from data where industrial or economic issues might arise.

Other points that were stressed is related to the duty to publish a paper in open access journals or venues if the research results are provided thanks to VA. There might be situations where this is rendered difficult if not impossible for the reasons mentioned above. Part must fall into market access instead.

UC suggested that we should start providing Raw data access that could potentially populate our CKAN repository and bootstrap the process even if it does not fully follow the best practices, until our SFDO service is made available.

A final recommendation to the SLICES consortium, as a future step, is to work towards the provisioning of a set of PaaS services that may support researchers in building up complex scientific applications by composing services provided at platform level.

4. SLICES self-assessment of the VA

The SLICES-SC consortium decided to complement this analysis with a self-assessment of the VA.

4.1. Highlights

Strong expertise in different fields by individual partners

The SLICES-SC consortium is composed of 14 Partners from European Member States. Each Partner brings to the consortium its own expertise and know-how from different ICT fields, ranging from networking, through IoT, cloud computing to software development. Partners contribute to R&D activities realized through European or National projects, extending scientific achievements in these fields. Among the years, Partners have set up and maintained local research infrastructures, where scientists validated novel concepts and ideas in close-to-real, yet limited geographically testbeds.

Scientific data sets available in the consortium

As indicated in the SLICES-SC Deliverable D8.4, project partners individually developed their own methods and techniques for collecting data and sharing them with the community. Examples include data sets for AI-driven application-aware 5G networks, UE statistics in LTE networks, energy consumption data or environmental monitoring data from IoT sensors. These data are shared by the tools developed by Partners through internal policies of each Partner.

Initial set of tools available

At the project level, a set of tools have been employed to realize Virtual Access for researchers and scientists. The list includes the web site, CKAN server, visualization tools for experiments and datasets and reproducibility framework. These tools serve the community by opening access to data sets produced by the consortium and scientists, including experimenters.

4.2. Technical self-assessment

We carried out a technical self-assessment of the proposed Virtual Access tool, namely, the CKAN server, which has been deployed and operated by the project. The assessment focused on three main areas:

1. compliance with FAIR principles;
2. user experience and;
3. technical parameters.

We evaluated the CKAN implementation whether it fulfils all the FAIR requirements: findability, accessibility, interoperability, reusability, which resulted in a positive answer. From the user experience point of view, we investigated the web-based, graphical user interface (graphical elements, font sizes, navigation, intuitive usage, etc.) as well as whether there are user guides, manuals available (if needed). The results were positive here too. The technical investigation covered responsiveness (empirically), access control options, volume size limits, checking the existence of a command line interface (CLI, for accessing data programmatically), federation options, etc. Some parameters were found to be low or insufficient in the initial setup (e.g., maximum file size, overly simplified access control), which had however been fixed soon after the evaluation. We do not detail these results and findings in this document, as these are described in deliverable D3.2 in section 7.2.

4.3. Virtual Access alternatives

We investigated another tool, called Dataverse[6], which is also a popular solution in academic/educational/public sectors for storing and sharing data in the long term. We overviewed its features and found that it also complies with FAIR principles. It offers flexible metadata configurations, data citation, versioning, access control, data citation and integration options, as well as a command line interface to access the data from program codes (scripts, workflows, batch processing), which was extensively tested on a test deployment. Similarly to CKAN, there are several options for the backing up repository technology (S3, Swift, NAS, Azure blob, etc.), which can ensure redundancy (data safety) and scalability. We overviewed the Research Object concept and RO-Crate packaging, which aid reusing data by collecting raw data, metadata (context, provenance) and methods (workflows, processing codes) and packing them together. The comparison with CKAN concludes that CKAN is more suitable for storing and sharing open data (which is the objective in this project), while Dataverse is better applicable in domain-specific contexts (in specific research areas). We omit the details here; the interested reader is referred to Deliverable 3.2 Section 11.4 for a complete description.

4.4. Proposals for future improvement

Data repositories are distributed among the testbeds and not coordinated centrally

One of the weaknesses of the current status of Virtual Access in SLICES-SC is the fact that, currently, data repositories are distributed among Partners' facilities and thus they are not fully coordinated centrally. The way the local repositories are organized is described in SLICES-SC Deliverable D8.4. Access to all these repositories varies and it seems to be driven by local policies, thus limiting the impact it could achieve comparing a centrally driven Virtual Access. On the positive aspect it should be mentioned, that a critical mass has being achieved with these distributed local repositories.

[6] The Dataverse project website, https://dataverse.org/ [Last accessed 25 June 2024]

However, extra effort is required to organize and manage them centrally to fully achieve the objectives of Virtual Access.

Limited contributions/uploads to project-wide data repositories

Currently, the project-wide repositories do not work as expected. It seems that local repositories are more attractive to local researchers and scientists or that they are not yet aware of the central repository, the simple access it provides and broader exposure.

4.5. Conclusions and recommendations for the following period

Take a more centralized approach to Virtual Access – concentrate on what is the offer of SLICES-SC as a whole, not provide individual contributions

It seems necessary that SLICES-SC should take more coordinated effort towards Virtual Access. The tools available at the project level should be proposed and promoted among the Partners. The coordinated approach would increase availability of scientific data sets to research communities at the European level and contribute to the goals of SLICES Research Infrastructure as a whole. It fully corresponds to our initiative to design, deploy and operate MRS/SFDO. We expect that its availability will incentivize researchers to share their data.

Use information collected from local facilities to enhance Virtual Access with new data and tools

It's evident that Partners of SLICES-SC gathered expertise, skills and knowledge in data collection and sharing. However, in the following period the consortium should concentrate on how to use the outcomes developed so far by individual Partners to enhance the portfolio of SLICES-SC Virtual Access (e.g. by grouping the results by thematical domains and provide tools and services to share scientific data sets in more homogeneous and coordinated way). We believe that SFDO/MRS will provide a first attempt to support the FAIR principles and expose its value to our community.

Continue the development of tools allowing VA

The tools proposed so far are good candidates to realize the goals of SLICES-SC Virtual Access. CKAN, GitHub and other tools, in particular data management and reproducibility frameworks should be further developed to allow all SLICES-SC Partners to collect and share scientific data derived from experiments with research communities. This will be pursued in MRS/SFDO, including the provision of the storage service.

Enhance SLICES-SC data repositories with data from experiments and measurements collected in the following period

SLICES-SC should further explore efficient ways to collect and populate data from scientific experiments. It should be considered to engage experimenters selected in Open Calls to contribute to scientific data sets. Moreover, Partners should more actively report in project-wide data repositories the results of experiments handled during the next period, to offer more complete and exhaustive scientific data to end users. Once MRS/SFDO is made available and SLICES will enter into pre-operation, we expect that these conditions will leverage and accelerate research data sharing.

5. Conclusions

Research data sharing and access is not yet mature in the research communities that SLICES targets as its main user base. There is still very little incentive to share research data for the many reasons that have been properly identified. SLICES has the potential to change the research methodology by providing a simple access to data sharing and a reward for those sharing their data.

Developing a relevant solution (MSR/SFDO) takes time but this process is in progress and should be made available in its basic form for the SLICES pre-operation that will start in Q4 2024. In the meantime, and although it is a slow start, we have developed a preliminary solution, that has a value to raise awareness and start to fill the process. We have conducted a survey, collected the advice from the SLICES User Committee as well as we provided a self-assessment of our solution and how it should develop in the future.

The SLICES UC has performed an initial assessment of the VA services reported in the previous deliverable D8.4. This deliverable D8.6 is an extension of the previous one (D8.4) with the inclusion of the UC assessment exercise and results.

The UC members had access to the VA services and have had the opportunity to evaluate them across several criteria: relevance of the SLICES VA services and the impact in the community (EOSC, researchers, testbed operators, Industry). The UC highlighted the relevance of the FAIR principles achieved on the SLICES VA services, giving special emphasis on the challenges of the interoperability and reusability. There are also positive comments on the potential impact and relevance on the joint contribution of the SLICES partners on VA services that will help much scientific community to improve their research. They also stressed the challenges and issues that our community might face.

In addition, we decided to conduct a self-assessment on the current status of the VA services, identifying its weakness and strengths. This self-assessment exercise aims to identify the lessons learned on starting to build the SLICES VA services that will constitute the seed for the definition and implementation of the future services during the future implementation and operation phases.

We found this first step highly informative, valuable and fruitful to propose a relevant solution to our community.

Annex: User Committee assessment form

USER COMMITTEE MEMBER

A. First name, family name, position, institution

B. Which main perspective(s) does you/your institution bring to this assessment?

- Potential user of RIs from Academia
- Potential user of RIs from Industry
- Potential access provider to external users
- Technological RI partner from Academia
- Technological RI partner from Industry
- Other (please specify)

C. Which category (-ies) of user are you/your institution representing in this assessment?

- Researchers.
  - Researchers from country members of the RI
  - European Researchers
  - International Researchers
- PhD Students
  - PhD Students from country members of the RI
  - European PhD Students
  - International PhD Students
- Public authorities
  - Country members of the RI
  - European
  - International
- Other Users from public/non-profit initiatives
- Users from private sector
  - SMEs
  - Industry
  - Services
- Other: please specify

EVALUATION OF TECHNICAL CRITERIA

D. Added value of the SLICES VA services

E. Technical relevance of the SLICES VA services

F. Contribution to the SLICES building community and potential impact

EVALUATION OF FAIR CRITERIA

G. Findability criteria

- *The data can be searchable and discoverable online*
- *Deposit of the data in a suitable repository that is recognized in the scientific field*
- *A metadata description of the data is provided using a standard scheme*
- *A persistent and unique identifier (DOI) is a key aspect of making the data FAIR*

H. Accessibility criteria

- *Clear information on how to get access to the data*
- *FAIR data does not necessarily mean open data, if necessary, it can be accessed using authentication process*
- *Appropriate license of the data help to clarify what others can and can´t do with the shared data*

I. Interoperability criteria

- *The objective is to facilitate to make the data reusable and being integrated with other data*
- *The data can be utilized by applications or workflows for analysis, storage, processing and long-time usability*

J. Reusability criteria

- *To maximize, the potential reuse of the data*
- *The data and related metadata are well-described so they can be replicated and/or combined in different settings*
- *Contextual frameworks for re-users to understand the data and how it came out*
- *Description of the research methodology employed to create the data*

EVALUATION OF VA METADATA CRITERIA

K. Relevance on the SLICES VA metadata[7]

L. Potential impact on EOSC ecosystem

EVALUATION OF VA TOOLS CRITERIA

M. Relevance on the SLICES VA Tools[8]

N. Potential impact on EOSC ecosystem

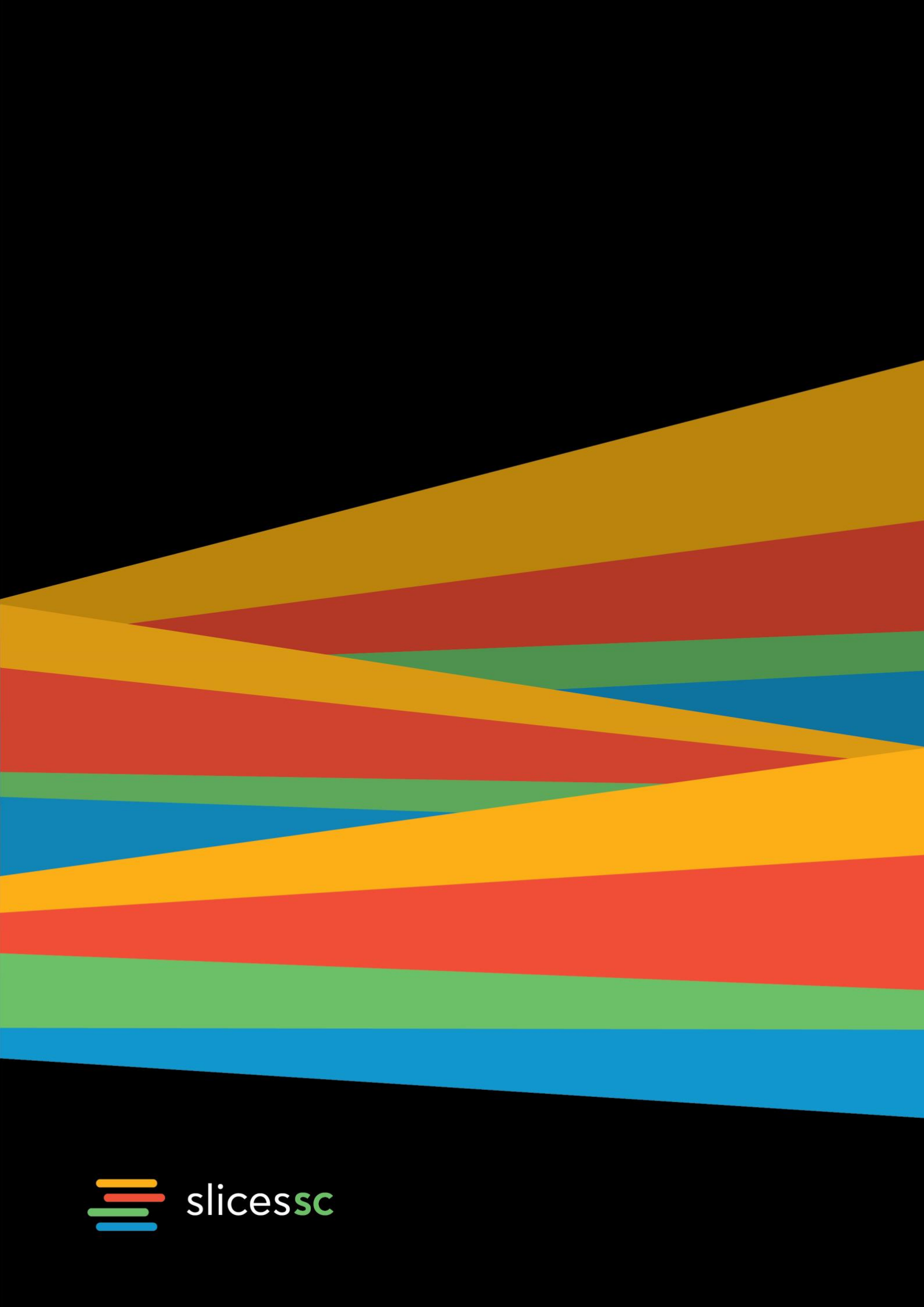
SUMMARY OF THE ASSESSMENT AND RECOMMENDATIONS

O. Summary

P. Recommendations

[7] SLICES community within the SLICES-SC project, and targeting the EOSC catalogue/marketplace, is proposing a metadata profile for EOSC in the deliverable D4.5, section 5.3, and in the deliverable D4.3. The virtual access metadata follows this already defined profile based in Dublin core metadata schema. Dublin Core, DCMI Schemas, https://www.dublincore.org/schemas/ [Last accessed 25 June 2024]

[8] SLICES Software and Tools repository: https://gitlab.distantaccess.com/slices [Last accessed 25 June 2024]


slicessc