HORIZON 2020

H2020 - INFRAIA-2020-1

D3.1 SLICES-SC Data Management Plan

Acronym	SLICES-SC
Project Title	Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies – Starting Community
Grand Agreement	101008468
Project Duration	36 Months (01/03/2021 – 29/02/2024)
Due Date	31 August 2021 (M6)
Submission Date	31 August 2021 (M6)
Authors	Sébastien Ziegler (MI), Cédric Crettaz (MI), Adrian Quesada Rodriguez (MI), Émilie Mespoulhes (SU), Nikos Makris (UTH).
Reviewers	All partners

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

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.

Executive Summary

The objectives of the SLICES research infrastructure are to design and implement a European test platform for the ICT domain, to support large-scale experimental research by providing advanced computing, networking and storage. This research infrastructure will allow a large number of researchers to create different types of data through observations, experiments and simulations, and finally, to produce results useful for the research. The collaboration between researchers will be improved thanks to data and services provided by other testbeds supporting SLICES. Policies and procedures are essential for the regulation of the data management and publication in the context of SLICES.

The deliverable presents the operational and technical requirements provided by the scientific community, the objectives and the constraints of the proposed reference architecture, the governance, the general management, the human resources, the legal and ethical policies and regulations. This Data Management Plan (DMP) encompasses the policies and the protocols needed for an efficient governance and management of the data, included the access from external systems and infrastructures. FAIR principles are also included in this document to enhance the interoperability and the collaboration with external systems and infrastructures.

The design of the metadata is very important to enhance the interoperability and the reusability of the data. This will increase the impact of the research. The project has studied in the frame of this deliverable the metadata schema standards and proposed one of these well-known standards. The chosen metadata schema standard will be extended in function of the needs of SLICES.

The deliverable has also analysed the reference architecture, the intended services, the governance, the integration and the interoperability with other infrastructures, like EOSC, and external services related to other international testbeds.

The Data Management Plan is fully aligned with the deliverable D9.1 “POPD - Requirement No.1” solving the issues highlighted in the Ethics Summary Report (EthSR).

Table of Content

1. SLICES-SC in a nutshell

The future SLICES Research Infrastructure aims to develop and provide services related to experimentation in the context of digital sciences such as 5G, 6G, NFV, Internet of Things and cloud computing. The SLICES-SC project is currently building a community of researchers around SLICES-RI, which will offer the necessary solutions to create and manage efficiently the experiments. Among the features to be implemented by the SLICES-RI for the experimenters, the SLICES-SC project will investigate a facilitated access for the experiments, the reproducibility of the research experiments, the validation of the experiment results and finally, the publication of the results in open data access.

1.1. Objectives of WP3 and its relevant tasks

The goals of WP3 are to ease the virtual access to the data collected during the different experiments and to create the required policy and guidelines on the data management. They will ensure a correct collection of the data from the different testbeds, an efficient data management and a data protection compliant with the regulations and the standards related to the security and the privacy. A framework to facilitate the reproducibility of the experiments will be also designed in the WP3. An open data server will also be installed for the publication of open data generated during the project.

The first task of the WP3 is the task T3.1 which define the data management plan (DMP). This implies to take into account the European regulation, in particular the GDPR (General Data Protection Regulation) and the ePrivacy Directive. The FAIR principles will be also integrated in the data management plan to guarantee the sustainability of data generated in the diverse experiments.

The task T3.2 concerns the reproducibility of the results, notably through the provisioning of the data created during previous experiments. Guidelines on how to publish the experiment results will be defined to increase the scientific interactions in the different disciplines. Benchmarking will be also investigated in this task to improve the repeatability.

Finally, the task T3.3 will establish a data protection office which will monitor the application of the data management policy and the practices related to the data protection and privacy. The office will also create the guidelines concerning the Intellectual Property (IP). A CKAN server will be installed in the context of this task to publish the open datasets.

1.2. Objectives of D3.1

The deliverable D3.1 is the Data Management Plan (DMP) which should be applied by all the project partners. The current regulation like the GDPR and the ePrivacy Directive is integrated in the DMP with a strong focus on personal data protection by design and by default. To enforce the legal requirements, a data management policy is specified in the deliverable D3.1. Concrete measures are also described to allow the sharing of data used for the reproducibility of experiments.

2. Overall Structure of the Deliverable

The deliverable D3.1 is structured in several sections. The first one is the analysis of the requirements with different angles such as the user perspective, the data types and formats, the licenses, the expected size of the generated data and the interactions with other components or systems.

The second section concerns the data management policy which describes the data governance framework, the architecture and the data quality. The metadata management, the interoperability

and the analytics are also specified in the data management policy. Other data management issues and resource allocation are mentioned in this section of the deliverable D3.1, too.

The FAIR principles are described in a dedicated section. They are very important when sharing the public datasets in the research community.

A section related to the legal compliance is listing the current regulations to be taken into account when managing and sharing the data in the context of the SLICES Research Infrastructure. In particular, the General Data Protection Regulation (GDPR) and the ePrivacy Directive are explained to ensure their correct application in the SLICES-SC project.

The measures to guarantee the security and the privacy of the personal data are enumerated in a dedicated section and these measures should be put in place by all the partners of the project.

Finally, a section is taking care of the ethics aspects met in the SLICES-SC project.

3. Requirements Analysis

This section analyses end-user, technical and operational requirements that relate to the management of data within SLICES-SC. These requirements stem from analysing the requirements of scientific communities and other related stakeholders through extended surveys conducted by the consortium, and the organization of open workshops (e.g., the SLICES March 2021 Workshop) with access from a wide range of stakeholders (e.g., industry partners and academia). The data management is ensured to have legal compliance and resolve any regulation issues at a national, European and international level, while addressing interoperability with existing and future platforms.

The analysis has resulted in the identification of the user groups and their characteristics, as well as the data they manage for research. Based on the types and formats of the data, requirements have been drawn for the information model, required for storing both data and appropriate metadata to ensure their discovery. Further requirements for interoperability are shown as well as the types of intellectual property rights (e.g., licenses) that SLICES needs to support. The respective Data Management Framework analyses the appropriate procedures, protocols and services aiming to fulfil these requirements.

3.1. User Groups

SLICES aspires to provide access to state-of-the-art infrastructure and tools for deploying innovative experiments. The different identified groups of users for accessing the infrastructure/data generated over it are summarized below:

A. **Research User Group:** This user group includes *research-oriented organisations*, such as Universities, Colleges and Research Institutes, which primarily focus on the generation of new knowledge through state-of-the-art research and innovation. Such academic research groups across institutions/organisations often coordinate to form larger *research communities/networks* that collaborate to tackle more complex problems. Users within this group typically utilise institutional or cross-institutional infrastructures to experiment, collect and analyse data to discover new patterns and unveil new knowledge.
   - **Access Frequency:** High, to support research actions
   - **User Type:** Sophisticated, can perform complex operations (e.g., predictive modelling) that may require significant resources

B. **Industry User Group:** This group includes *business-oriented organisations*, such as *enterprises* and *SMEs*, but also applied research and development organisations who target the

development of products. In the case of SMEs, fast access to resources is essential in order to increase productivity and efficiency and create new opportunities in the competitive global landscape. It is also important to note that through this process, the industry organisations can acquire information about newly available technologies so that they can test them quickly. Finally, these collaborations may also be of interest to external *investor* stakeholders, such as *funders* and innovators, that are willing to fund specific endeavours.

- **Access Frequency:** Low, to support ad-hoc innovation actions
- **User Type:** Parametric, can perform complex operations using well-defined functions to draw conclusions fast

C. **Research/Industry Support User Group:** Research infrastructures are typically operated/supported by *research infrastructure managers* and technical personnel who ensure the effective and efficient operation of the infrastructure

- **Access Frequency:** Very High, to support the 24/7 operation of the platform
- **User Type:** Expert, can perform maintenance operations

D. **External Partners User Group:** This group includes users such as compliance officers, policy makers, educators and civil servants who use information to ensure compliance, enhance regulatory processes, supervise operations, drive new policies or utilise the results of science to enhance their environment and the society in general.

- **Access Frequency:** Very Low, to support a specific task
- **User Type:** Knowledgeable, can perform pre-defined operations

3.2. Types of Data

The data created and collected in the SLICES infrastructure through experiments will be of value and used for numerous IT/networking infrastructures, and will enable fast and effective implementation of new networking and big data infrastructure technologies. The datasets generated by utilising the SLICES hardware and software are summarized below:

- **Observational Data:** this type of data is collected using methods such as surveys (e.g., online questionnaires) or recording of measurements (e.g., through sensors). The data will include information related to signal or performance measurements, and network or service log data that allow for experiment evaluation and reproducibility. Such data is captured in real-time manner.
- **Experimental Data:** this type of data is created from experiments over the infrastructure and represent the evolution/effects of certain variables, trying to determine whether there is any correlation/causality. Both observational and experimental data are essential for new technology evaluation and are of interest to the wider research and industry community for making informed decisions about new technology implementations and/or improvements.
- **Simulation Data:** is generated by using computer models that simulate the operation of a real-world process or system. These may use observational data.
- **Derived Data:** involves the analysis (e.g., cleaning, transformation, summarisation, predictive modelling) of existing data, often coming from different datasets (e.g., the results of two experiments), to create a new dataset for a specific purpose. This data is required for data driven experiments and research in Artificial Intelligence or Machine Learning-driven experiments and technologies that are increasingly used in networking, telecommunication and data driven research, as well as facilitated by Industry 4.0 technologies, such as Digital Twins, Automation and Robotics.
- **Metadata:** concerns data that provides descriptors about all categories of data mentioned above. This information is essential in making the discovery of data easier and ensuring their interoperability.

The types of data are further summarized in *Table 1*. Different formats for the data are foreseen, mostly unstructured and semi-structured, thus requiring the development of a non-relational distributed databases to store them in SLICES. The data model will need to ensure openness and flexibility to accommodate for different format requirements, while in parallel, offer high performance storage, processing and retrieval operations.

Table 1 – Overview of Types of Data collected in SLICES

Data Category	Sources	Processing	Characteristics/Examples
Observational	- Surveys - Recorded measurements	No processing of data. As collected by the research tools (e.g., questionnaires, sensors).	Signal or performance measurements, network or service logs, real-time captures. Essential for technology evaluation.
Experimental	- Experiments - Interventions	Data characterises variables in an attempt to determine if there is any correlation/ causality.	Essential for technology evaluation, technology improvements.
Simulation	Simulation software environments	Implemented system models generate the data. These may be informed by the use of observational data.	Imitate the operation of a real-world process or system.
Derived	Existing datasets	The analysis (e.g., cleaning, transformation, summarisation, predictive modelling) of existing data, often coming from different datasets to create new ones.	Data driven experiments and research in Artificial Intelligence; Machine Learning-driven experiments and technologies that are increasingly used in networking, telecommunication and data driven research.
Metadata	Descriptors for data from all other categories	Processing follows rules for descriptors.	Essential in making the discovery of data easier and ensuring interoperability.

3.3. Formats of Data

Different formats of research data exist (e.g., file formats, database formats), which impact directly the ability of a system to manipulate the corresponding data and provide access to other users down the line. These formats for the SLICES case can be roughly split into two categories, open and proprietary file formats.

Open file formats[1] (e.g., csv) make their structure publicly available for others to use or implement. Such formats are essentially standardised by public authorities or international bodies with the objective to improve data interoperability. Moreover, open formats can be either text (i.e., human readable format), or binary format that is not human readable, but decodable when the specifications are known.

Proprietary file formats (e.g., rar, sas7bdat) contain a non-transparent structure with their specification not made publicly available. Software companies create such file formats for encoding the outputs of their applications, thus making it only possible to read by using the specific company software.

One of the main objectives of SLICES is to promote interoperability, thus non-proprietary, unencrypted, uncompressed, and commonly used by the research community formats should be adopted. When storing the data in open formats may lead to loss of information or structure, it is recommended to also store the data in the original proprietary format. In such cases, it is also recommended to advise users to prepare extra documentation that lists the necessary software (or provides link(s) to download it), the appropriate version and other constraints that the format entails, to further improve data reuse[2]. The library of Cornell University further suggests some file format characteristics that support long term data preservation. These include complete and open formats that are platform/vendor independent, without partial or full encryption and/or password protection. The topic of recommended formats is further explored in the Recommended Formats Statement, published by the US Library of Congress[3].

Overall, we conclude that SLICES adopts the following file format specifications to be used for storing research outputs:

- Open (non-proprietary) file formats
- Uncompressed, or compressed with a free/open format such as .7z
- Unencrypted, or supplemented by appropriate decryption mechanisms
- Commonly utilised by the research community (e.g., csv) or highly interoperable among diverse platforms, systems and applications (e.g., YAML, JSON, XML)
- Developed and maintained by an open standards organisation, with a well-defined inclusive process for evolution of the standard

3.4. Dataset License Types

Regarding the specific licenses for datasets produced in SLICES, the model that will be adopted is oriented towards avoiding legal challenges with data sharing. This is a multi-dimensional and complicated aspect, since different jurisdictions apply different rules and standards for different aspects of data (e.g., record values, attribute names, database model). Licences and waivers are

[1] Web Archive: Openformats.org Open vs. Proprietary formats, https://web.archive.org/web/20130219012116/http://www.openformats.org/en1 [Last Accessed 26 July 2021]
[2] Stanford Libraries: Best Practices for file formats, https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats [Last Accessed 26 July 2021]
[3] Library of Congress: Recommended Formats Statement, https://www.loc.gov/preservation/resources/rfs/ [Last Accessed 26 July 2021]

instruments that allow users to permit a second party to access and reuse data. Licenses grant permissions given that specific conditions (e.g., attribution, copyleft and intent), which are set by the data owner, are met. To avoid the complexity of drafting a license from scratch, the following standard licenses exist:

- **ODC-By:** Open Data Commons Attribution License (ODC-BY) allows re-users of the data to distribute, copy, transform, build upon and produce works using the data for any purpose. New content or new databases generated as a result of using the licensed dataset must contain a notice mentioning the use of the licensed dataset.
- **ODC-ODbL:** Open Data Commons Open Database License (ODC-ODbL) is an extension of ODC-By, since it adds more conditions. Firstly, the "copyleft" condition is added in case new databases are derived from the original database. Secondly, technological restrictions may apply only to the database or a new database derived from it, only if another copy without the restrictions is made available.
- **ODC-DbCL:** Open Data Commons Database Contents License (ODC-DbCL) removes the copyrights of the contents of a database, but it does not affect the copyright of the actual database.
- **PDDL:** Open Data Commons Public Domain Dedication and License (PDDL) is nearly identical to CC0, but instead it uses wording that is specific to database terms, while it provides a set of community norms to be associated with a database.
- **CC0:** Creative Commons Zero (CC0) allows for waiving all database rights and copy right interests to the public domain. Moreover, it can act as an irreversible loyalty free/ unconditional license for anyone who wants to use the data for any purpose.
- **CC PDM:** Creative Commons Public Domain Mark (CC PDM) acts as a tool that asserts works as already being a part of the public domain, thus allowing public works to be more easily discoverable and recognisable as public. Unlike CC0, it cannot waive work rights.
- **CC BY:** Creative Commons Attribution (CC BY) is one of the open Creative Commons licenses that is only described by a single condition, i.e., attribution. This license specifies that the re-user of the data must provide credit to the licensor when the work is distributed, displayed, performed or used to derive a new work.
- **CC BY-SA:** Creative Commons Attribution Share Alike (CC BY-SA) is an extension of CC BY, since it has an additional condition, i.e., Share Alike, which requires re-users that transform, remix or build upon the licensed dataset to distribute their contributions under the same license as the original.
- **CC BY-NC:** Creative Commons Attribution Non-Commercial (CC BY-NC) is an extension of CC BY, since it has an additional condition, i.e., Non-Commercial, which prevents re-users from using the licensed dataset for any commercial purposes.
- **CC BY-ND:** Creative Commons Attribution No-Derivatives (CC BY-ND) is a more restrictive extension of CC BY, since it has an additional condition, i.e. No-Derivatives, which prevents re-users from making additions, transformations or any type of changes to the dataset.
- **CC BY-NC-SA:** Creative Commons Attribution Non-Commercial Share Alike (CC BY-NC-SA is one of the most restrictive licenses. It allows users to share a dataset only if they provide credit, avoids using the dataset for commercial purposes and makes sure to redistribute their changes using the same license.
- **CC BY-NC-ND:** Creative Commons Attribution Non-Commercial No-Derivative (CC BY-NC-ND), yet another one of the most restrictive licenses, allows users to share a dataset if it is unmodified and not being shared for commercial reasons. No additions or transformations are permitted on the dataset.

- **CDLA-Permissive-1.0:** Permissive Version 1 is one of the Community Data License Agreement licenses and it is very similar to permissive open-source licenses. The re-users of the data can modify, adapt and share the data, as long as they provide credit, while no obligations or restrictions are imposed on derived computation results.
- **CDLA-Sharing-1.0:** Sharing Version 1 is one of the Community Data License Agreement licenses and aims to put the "copyleft" principles in use for data. Re-users are allowed to adapt, modify and share the dataset or their derived changes, but only under the CDLA-Sharing, while also giving credit to the owner of the licensed dataset. No obligations or restrictions are imposed on derived computation results.
- **OGL:** Open Government License (OGL) is a license intended for the UK public sector and cannot be used by licensors outside the UK. It is similar to CC BY since it has the attribution requirement. Moreover, derivative works and commercial uses are also allowed, while no "copyleft" condition exists. A non-commercial variant of this license also exists (NCGL).
- **Bespoke or Custom License:** Can be used when datasets contain high commercial value or when explicit responsibilities to re-users of the data need to be specified. Templates such as the Restrictive License (RL) can guide the preparation of bespoke licenses.

As it is shown in the re3data[4] statistics, some of the aforementioned license types, such as Copyrights, CC and Public Domain, dominate the license utilisation landscape. However, it is also noteworthy that most repositories allow users to specify the other license types too.

SLICES end users will need to have the ability to decide on a suitable license and attach it to their data. This should be explicitly supported by the metadata repository using a library of license types that users can choose from when constructing their metadata. In those cases where a suitable license does not exist, then **Public Domain should be provided as a default option**, but the option to select Other should also be available. Additionally, applications that query for metadata should display this information prominently, so that a second party will immediately realise that the data must be licensed prior to accessing it.

3.5. Expected Data Size

The provisioning of the testbeds in a Testbed as a Service (TaaS) format was conceived and developed several years ago in order to provide a "ready to go" environment for experimental activities. Such TaaS format provides easy access to the required communications, computing and storage resources for the experiments. Different examples of such a scheme exist, such as OneLab/FIT[5], GENI[6] and Fed4Fire/Fed4FIRE+[7], who over the years have hosted hundreds of thousands of experiments and thousands of users. However, the limitations and bottlenecks of the current Internet have called for a new design supported by recent worldwide initiatives, such as NSF-PAWR[8] and NSF-Fabric[9] in the US, but unfortunately, none of that kind in Europe.

SLICES aims to design and develop a distributed research infrastructure with next generation capabilities that will host thousands of users and their data. Our preliminary estimations include up to 5,000 users and their data, accounting for up to 50GB per user on the individual nodes and up to 1TB

[4] Re3data, Statistics, Data and Database licenses usage, https://www.re3data.org/metrics/dataLicenses [Last Accessed on 26 July 2021]
[5] OneLab, https://onelab.eu/ [Last Accessed on 03 February 2021]
[6] Geni, https://www.geni.net/ [Last Accessed on 03 February 2021]
[7] Fed4Fire+, https://www.fed4fire.eu/ [Last Accessed on 03 February 2021]
[8] Platforms for Advanced Wireless Research – PAWR, https://advancedwireless.org/ [Last Accessed on 03 February 2021]
[9] Fabric, https://fabric-testbed.net/ [Last Accessed on 03 February 2021]

on the cloud. This provides us with a preliminary estimation of 0.25PB-1PB of data storage for datacentres residing on SLICES nodes, and 5PB for the central cloud-based datacentre.

3.6. Interaction with Other Infrastructures/Systems

As the SLICES architecture relies on highly modular components, the integration with existing Next Generation Internet (NGI) European and international digital technologies will come in a standardised manner. To this aim, SLICES consortium has studied the APIs and manner of interaction with existing infrastructures and pinpointed the wide use of Network Function Virtualization (NFV) architectures across the community. SLICES will adopt similar APIs, which will allow the interaction and integration with relevant infrastructure as separate domains. This will allow the infrastructure to summon resources under a unified architecture and enable a plethora of different scenarios/use cases to be evaluated through the use of single-point-of-entry and consistent APIs across the community. The different APIs and policy enforcement across different domains is a topic that will be further analysed, along with agreements, compliance issues and technical issues that can come up during the integration process.

It is important to note that some infrastructures/systems (e.g., the European Open Science Cloud - EOSC portal) provides specific requirements for onboarding the functions that need to run. Below, we provide an indicative list of these requirements that have been taken into consideration for the design of the data management framework presented in the next section:

- Services that will be exposed or integrated into other infrastructures must be operational and provide a service that is not trivial; e.g., a metadata discovery service vs. a link to a dataset.
- Resources should provide documentation related to licensing, terms, etc.
- Data should include rich metadata to ensure effective discovery.
- Data should also adhere to the FAIR Principles (Findable, Accessible, Interoperable and Reusable).
- Appropriate procedures and processes should exist for compliance with national and international regulations (e.g., GDPR).

4. Data Management Policy

This section presents a data management policy and the data governance framework implementing this policy. The final goal is to achieve an efficient operation of the SLICES research infrastructure and the realization of the objectives defined in the project.

4.1. Data Governance Framework

The data governance framework specifies the data governance. The structure of the data governance permits to support in an effective way the operations linked to the data management. First of all, the roles are defined in a hierarchical structure. Typical roles are Data Manager, Data Protection Officer and Metadata Administrator. The tasks of the data governance consist to implement all the processes described in the different policies through standards and good practices.

All processes and systems deployed in the SLICES research infrastructure are sufficiently flexible to be updated and optimized in function of new requirements defined by the researchers. This implies a continuous monitoring of the research infrastructure through different metrics assessing at the end if the objectives given by the researchers are effectively met. This monitoring process made through the data governance framework should not only involved all the systems and the research infrastructure,

but also the people. The results of the monitoring should permit to detect early problems and to take immediate corrective actions.

Basically, the SLICES research infrastructure will be composed by different technologies and applications in the testbeds. The aim of the data governance framework is to properly exploit the datasets provided by the testbeds in line with the objectives defined in the project. The framework described in this section should also improve the value and the impact of the collected datasets. The legal compliance is also handled by the data governance framework, in particular the General Data Protection Regulation (GDPR).

4.1.1. Organizational Model

The key element for a successful data governance is a strict model for the organizational management. This model must describe all the roles inside the organization and all the responsibilities associated to the data management.

In SLICES, the highest authority concerning the data management is the Data Governance Group (DGG) and takes part of the Coordination & Management Office of SLICES. The DGG is responsible for the consultation with the Coordination & Management Office on the aspects linked to the data management, the implementation and the supervision of the data management plan.

The DGG elects every two years a Data Manager who is the coordinator of the data management team. He monitors the implementation of the policies and their practical enforcement. He ensures the good coordination between the DGG and the technical groups, including the establishment of the communication channels between the different internal and external teams. The Data Manager defines the Key Performance Indicators (KPIs) which serve to assess the data quality and the data governance. He is also responsible to report the results of the data management monitoring. Finally, the Data Manager is working together with the Data Protection Officer (DPO) for the tasks related to the data protection and the compliance to all the identified regulations.

The Core Datacenter team is coordinated by the Data Manager and manages the data management infrastructure installed in the cloud. The Core Datacenter is composed by the following people:

- The Master Data Administrator who is managing the metadata.
- The Data Administrator who is in charge of the maintenance of the data infrastructure.
- The System Administrator who is taking care of the maintenance of the cloud infrastructure where the data and the tools are stored.

Each SLICES node is operated by a Node Technical Team coordinated by the Data Manager. A Node Technical Team is composed by a data engineer and by a system engineer. The task of both engineers is to maintain the local infrastructure and the tools dedicated to the data management.

4.1.2.Roles and Responsibilities

The following table shows the roles and the responsibilities of each person concerned by the data governance framework.

Table 2 – Roles and responsibilities

Role	Appointed By	Reports to	Responsibilities
Data Manager	Data Governance Group	Data Governance Group	Policy Implementation Monitoring/Reporting Internal Communication/Training Coordination of Technical Teams External Communication Policy Monitoring/Improvement Data Lifecycle Management Data Quality Monitoring/Enhancement Change Management
Data Protection Officer	Data Governance Group	Data Manager	GDPR Strategy Implementation Compliance (GDPR, National, International) Risk/Privacy Impact Assessment/Monitoring Communication (Contact point for data subjects)
Master Data Administrator	Data Manager	Data Manager	Metadata Management Metadata Accessibility Monitoring of Data Standards FAIR Compliance
Data Administrator	Data Manager	Data Manager	Monitoring/Configuration Data Architecture/Modelling Database Security Data Quality Backup and Recovery Troubleshooting
Cloud Administrator	Data Manager	Data Manager	System Monitoring Cloud Resource Planning/ Management Tool Integration Infrastructure Security Descriptive analytics for System Operations
Data Engineer	Node Director, Data Administrator, Master Data Administrator	Node Director, Data Administrator	Metadata Accessibility Monitoring/Configuration Database Security Data Quality Backup and Recovery Troubleshooting
System Engineer	Node Director, Cloud Administrator	Node Director, Cloud Administrator	System Monitoring Resource Planning/Management Tool Integration Infrastructure Security

4.1.3. Policy Enforcement and Maintenance

The Data Governance Group and its appointed Data Manager are responsible that everybody involved in the data management will adhere to the policies and procedures used for the proper data management. In this context, the Data Manager elaborates and implements the required procedures to make the data management policies accessible. So, everybody can access them easily.

The data management plan is reviewed periodically through a dedicated procedure. This will allow to measure how the policies have positively impacted the project. On the other hand, the review of the data management will also detect issues to be quickly resolved. A review of the data management plan can serve to also to add new or updated objectives if the project has changed its strategy.

4.2. Data architecture

The data architecture should correspond to the requirements defined for correctly running the SLICES research infrastructure. Notably, the data architecture should be able to manage increased workloads at large scale. Of course, the data architecture is extendible by design; so, new technologies and processes can be implemented easily. The integration of external systems or infrastructures should be also supported by the data architecture. At the end, a good data architecture will ensure that all the objectives defined in SLICES.

The following figure is a preliminary proposal of data architecture and will be updated in function of new cutting-edge technologies to on-board in the SLICES research infrastructure.

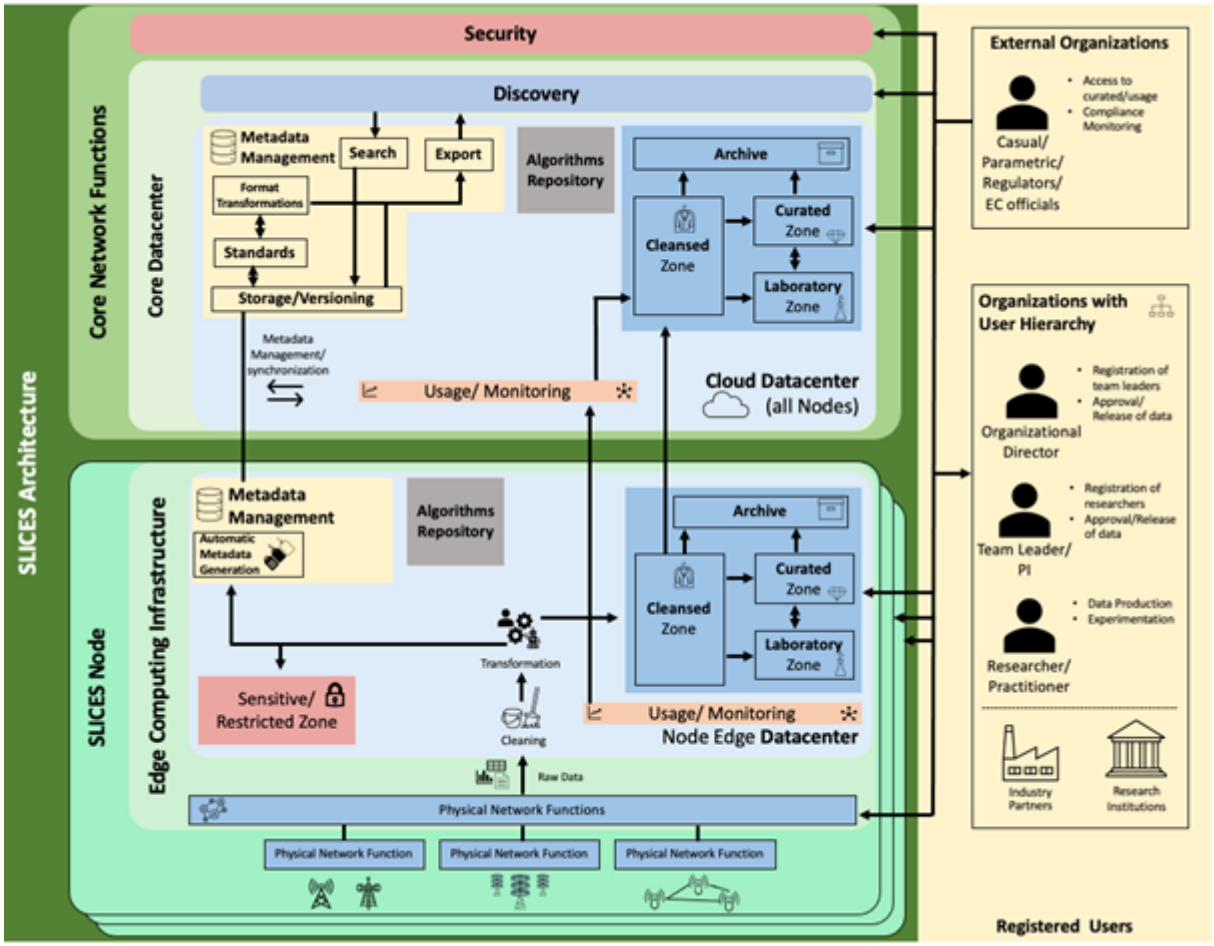


Figure 1 – Architecture

The data architecture is composed by two levels:

- Distributed datacentres located at each SLICES node. In complement to them, other sites focused on edge or core computing experimentation can be incorporated, in function of the equipment provided by the node.
- Centralised datacentres located in the cloud. They provide also the necessary connections from the distributed datacentres.

4.2.1. Node Computing Infrastructure

The experiments generate the data by accessing the virtual or physical data through physical network functions. The data are sent to the datacentre hosted in the local node. The data are processed to improve their quality. Typical treatments are cleaning and integration. A specific data model is applied to the data and metadata are added automatically or manually. The storage of the metadata is realised a database specifically dedicated to them.

The datasets containing personal or sensitive data will be tagged accordingly in the metadata and will be stored in a restricted area which the access is limited to the creators of the datasets. Of course, security mechanisms will be used to secure more the sensitive datasets. This area is represented in the figure above as “Sensitive/Restricted Zone”.

The Cleansed Zone illustrating in the previous figure saves the datasets received, but transformed by following the requirements elaborated in SLICES for the metadata. The datasets are not modified during a whole project. After their utilisation, the datasets are archived in the Archive Zone for a long period of time. The Archive Zone is also used to create data backups from all the other zones.

The next area is the Laboratory Zone is dedicated to the exploration and the analysis of the datasets. In this zone, compute-intensive operations are realised and produced new datasets. These new datasets are transferred to the Curated Zone which contains the results of aggregations, modelling or optimisations, etc.

The final area is the Usage Monitoring which measures the utilisation of the resources deployed in the research infrastructure. The measurements made in this zone permit to assess if the defined objectives of SLICES are met or not.

4.2.2.Core Datacentre Infrastructure

The Core Datacenter is based on the cloud and collect, process, store and distribute the large amount of data (Big Data). Compared to the SLICES nodes, the Core Datacenter is offering new features such as very high scalability, more efficient computations, enhanced security mechanisms. These functionalities allow SLICES to answer to the challenges encountered by the European researchers in the context of ICT.

The integration and the analysis of Big Data coming from the different SLICES nodes are using advanced components in terms of storage and computation. This will permit complex computations and calculations at high speed. The intended size of datasets to be processed is between terabytes and petabytes. Other sources of data, like external datasets, vocabularies, standards, will be also integrated to realise validations and correlations.

The metadata are playing an important role in the integration of the datasets generated by the SLICES nodes. A Discovery component will implement REST APIs exposing the metadata to the researchers. These APIs will support advanced queries, free-text searches, combinations of filters, etc. The

Discovery component will also encompass a translation engine transforming an original format of metadata in other standardised formats.

The interoperability with other systems and research infrastructures like EOSC is realised by supporting a metadata harvesting protocol like OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)[10]. The data are communicated through standard data exchange formats such as JSON, XML and YAML. Other formats could be supported in the future.

Finally, analytics will enable the discovery of patterns thanks to open-source and commercial tools supporting real-time analysis of the data. Typically, data mining, machine learning and artificial intelligence are part of these tools very useful for the researchers.

4.3. Data Quality

The data quality has different dimensions for its evaluation, such as integrity, accuracy, completeness, consistency, timeliness and believability. Processing and analysis activities give better results if the data quality is high. So, the SLICES research infrastructure should be able to detect inconsistencies and to correct them. SLICES will deploy different Data Quality Management (DQM) tools to check directly or indirectly these dimensions. In particular, the DQM tools will address these dimensions:

- Data cleaning: The noisy and missing data will be handled by the data cleaning. It includes the replacement of values by global constants or rules, the anomaly detection through confidence intervals, boxplots or IQR (interquartile range), and smoothing through regression.
- Data integration: Semantic heterogeneity and structure will be treated, notably by linking the entities based on attributes or eventually rules. The detection and the erasure of redundant information will be also executed in the data integration phase. External data such as currency conversion or weather information will be integrated, too.
- Data reduction: The objective is to reduce the datasets in terms of volume or attributes. This can achieve through correlation analysis, principal component analysis (PCA), attribute subset selection. Other techniques encompass regression, log-linear models, sampling and summarisation by aggregation or generalisation.
- Data transformation: The datasets will be transformed or consolidated through smoothing, attribute construction by feature generation for instance, summarisation, normalisation and discretisation.
- Data interpretation: The interpretation of the data concerns the graphical and tabular representations, the evaluation of tasks in function of their complexity. In particular, the tools will permit to understand the distributions of the data and their trends and detect the extreme values and the anomalies. Furthermore, complex modelling techniques will be employed to help the researchers to better understand the collected datasets. Autonomous examination, network modelling, data mining functions and modelling with soft decision trees are typical examples of complex modelling techniques.
- Data security: This will be enforced through security measures and procedures. The access control will be put in place to limit the access to the data to the authenticated and authorised researchers. Data backups and eventually data recoveries will be also undertaken in the project.

Unhappily, the DQM tools cannot ensure the completeness of the data. Typically, such tools are able to provide the percentage of missing values for each attribute or gaps in time series. At the end, only

[10] The Open Archives Initiative Protocol for Metadata Harvesting, https://openarchives.org/OAI/openarchivesprotocol.html [Last accessed 01 February 2021]

the data provider can assess if the data are complete or not. Specific descriptors can be put in the metadata to indicate the completeness of the data.

The timeliness of the data is defined by the data provider when the data is created. This implies that the timings of measurements in the physical world and the timings of the data captures are aligned and accurate. In some use cases, the timeliness cannot be fully determined, and so validated, until the first use of the data; in this situation, specific descriptors concerning the timeliness must be included in the metadata by the data provider. This permits the verification of the correct timeliness from the data provider.

The last point to take into account is the believability. The objective of the believability is to prove that the data are coming from trustworthy sources. Unhappily, the evaluation of the believability and the trustworthiness cannot be implicitly with the tools.

4.4. Metadata Management

The metadata management allows a better collaboration between the researchers, including through the sharing of data in compliance with the FAIR principles. In the context of the project, an repository for the metadata will put in place. Metadata management procedures should be elaborated to reach the following objectives:

- Reusability: Descriptors should be present in the metadata repository to clearly describe the metadata and the data. Functions to query and retrieve the data should be also existing to ensure the reusability of the data.
- Interoperability: The metadata repository should be able to transform data into particular metadata formats. This allows the interoperability with external systems. Of course, one metadata format will be selected to store the data into the repository. Furthermore, interoperability services will be provided by the research infrastructure to the different stakeholders, allowing the use of external systems.
- Data quality: The data quality is already handled in the previous section. This objective is also linked to the reusability and the interoperability.
- Governance: The Data Governance Group (DGG) will take the decisions related to the metadata, in particular the choice of metadata format, the maintenance of the metadata, the eventual changes to be brought to the metadata.

The first step is to determine which are the possible candidates for the metadata format standard which will be used in the project. These standards permit the recording, the archiving, the discovering, the searching and the preservation of resources. A list of possible metadata schema standards is given below:

- AGLS (Australian Government Locator Service) Metadata[11]: This standard originates from Australia and it is used to search and discover the digital and the non-digital resources provided by the Australian government.
- AGRkMS (Australian Government Recordkeeping Metadata Standard)[12]: This standard is based on the AGLS Metadata standard shortly described above and the ISO 23081 standard. The last standard defines the metadata for records. The Australian government and the related agencies are using the AGRkMS standard to describe the resources to be archived at the national level.

[11] AGLS Metadata Standard, http://www.agls.gov.au/ [Last accessed 01 February 2021]
[12] Australian Government Recordkeeping Metadata Standard, https://www.naa.gov.au/information-management/information-management-standards/australian-government-recordkeeping-metadata-standard [Last accessed 01 February 2021]

- EAD[13] (Encoded Archival Description) and ISAD(G) (General International Standard Archival Description)[14]: Digital resources are archived through the EAD metadata schema. It specifies the structure and the content of digital resources. ISAD(G) is targeting the traditional and non-digital resources to be archived. Both metadata schema standards are compatible between them.
- OAIS (Open Archival Information System)[15]: This standard addresses the preservation of digital resources in the long term. OAIS is defining a reference model allowing the access to archive systems. OAIS is also specifying functions to access the preserved digital resources and to ensure the long preservation of such digital resources.
- PREMIS (Preservation Metadata and Implementation Standard)[16]: This metadata standard is also intended to the preservation of digital resources. A specific data model and a corresponding data dictionary are defined in this standard for the preservation. Five categories of entities are described in the PREMIS standard: intellectual entity, digital object, agent, rights and event.
- OpenDOAR (Directory of Open Access Repositories)[17]: In fact, OpenDOAR is a directory based on the Web which lists open access academic repositories. The storage is made in United Kingdom. The researchers can search resources by locale, content and other parameters. OpenDOAR is one of the two leading open access directories in the world.
- Dublin Core[18]: This metadata standard is also named Dublin Core Metadata Element Set and defines 15 core properties describing resources. The Dublin Core and the associated vocabularies form the DCMI Metadata Terms. Formally, Dublin Core is standardised as ISO 15836, ANSI/NISO Z39.85 and IETF RFC 5013. Any kind of resources can be described like Web pages, images, videos, books, CDs, etc. Apart the description of resources, the standard permits the combination of metadata vocabularies from different standards. The standard is guaranteeing the interoperability of metadata vocabularies notably through linked data and Semantic Web.
- DataCite[19]: This standard is intended for the data and the results generated by the research. Indeed, the DataCite standard provides digital object identifiers (DOIs) assignable to the research data. An organisation can join DataCite and then, to assign DOIs to their research data. So, the research data will become discoverable by the other organisations. Additional services are provided by DataCite to improve the management of DOIs in the broad research community.
- MINSEQE (Minimum Information of a high-throughput nucleotide SEQuencing Experiment)[20]: MINSEQE provides some guidelines to improve the integration of multiple experiments in the research community. The goals of MINSEQE are to ensure a good interpretation and reproducibility of the results generated by an experiment.
- DDI (Document, Discover and Interoperate)[21]: Data produced by surveys and other methods based on observations are described through this free international standard. Furthermore, the DDI standard is managing the lifecycle of research data, notably the conceptualisation, the

[13] EAD: Encoded Archival Description, https://www.loc.gov/ead [Last accessed 01 February 2021]
[14] ISAD(G): General International Standard Archival Description - Second edition, https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition [Last accessed 01 February 2021]
[15] Open Archival Information System (OAIS), http://www.oais.info/ [Last accessed 01 February 2021]
[16] PREMIS Data Dictionary for Preservation Metadata, https://www.loc.gov/standards/premis/ [Last accessed 01 February 2021]
[17] OpenDOAR - Directory of Open Access Repositories, https://v2.sherpa.ac.uk/opendoar/ [Last accessed 01 February 2021]
[18] Dublin Core Metadata Initiative, https://dublincore.org/ [Last accessed 01 February 2021]
[19] DataCite, https://datacite.org/ [Last accessed 01 February 2021]
[20] Functional Genomics Data Society, http://fged.org/projects/minseqe/ [Last accessed 27 August 2021]
[21] Document, Discover and Interoperate, https://ddialliance.org/ [Last accessed 27 August 2021]

collection, the processing, the distribution, the discovery and the archiving of the research data. Social, behavioural, economic and health sciences are the main targets of the standard.

- EML (Ecological Metadata Language)[22]: This specification is used to document research data in the context of environmental and earth sciences. An XML syntax and a vocabulary are provided by the standard to the researchers to preserve and document their research data and the related results. There are different core modules to identify and describe data, their formats, protocols and methods. The structure and the content of data are also described and the data can be annotated with semantic vocabularies.
- ISO 19115[23] and FGDC-CSDGM (Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata)[24]: The two standards are used to describe the geospatial data. First of all, ISO 19115 specifies the schema for the geographic data and also, the properties of these geographic data. Such properties are spatial references, the temporal aspect and the data quality. The FGDC-CSDGM is a well-known metadata content standard previously used in North America, but it was replaced by the ISO 19115 later.
- FITS (Flexible Image Transport System)[25]: FITS is in fact a file format maintained by the International Astronomical Union. It is used to exchange information between the observatories and also for archiving. The storage, the transmission and the manipulation of scientific images and their linked data are eased. FITS was designed as a transport format not only for still images, but also for 1-D spectra, 2-D images, data cubes with multiple dimensions or tabular data in several dimensions.
- MIBBI (Minimum Information for Biological and Biomedical Investigations)[26]: This standard defines several guidelines used for the reporting of data in the context of biosciences. The MIBBI standard facilitates the data verification, the analysis and the interpretation. This standard allows a faster building of structured databases, public repositories and data analysis tools in the research community.

An analysis made in the project has determined that Dublin Core is the most appropriate standard concerning the metadata schema. Indeed, Dublin Core is mature and already widely used in the different domains of the research. One of the largest directory of data repositories, re3data[27], is also using and supporting the Dublin Core metadata standard.

In the context of the project, some improvements of Dublin Core could be done to satisfy the requirements of SLICES. The reusability and the scalability needed to interact with external research infrastructures and external digital libraries can be improve through the automatic generation of metadata. The generation of metadata can be done for example through Machine Learning (ML).

The following table lists all the elements needed to describe the resources and their attributes in SLICES.

[22] Ecological Metadata Language (EML), https://eml.ecoinformatics.org/ [Last accessed 27 August 2021]
[23] ISO Standard (ISO 19115-1:2014), https://www.iso.org/standard/53798.html [Last accessed 27 August 2021]
[24] Federal Geographic Data Committee, https://www.fgdc.gov/metadata [Last accessed 27 August 2021]
[25] Flexible Image Transport System (FITS), https://www.loc.gov/preservation/digital/formats/fdd/fdd000317.shtml [Last accessed 27 August 2021]
[26] Minimum Information for Biological and Biomedical Investigations, https://fairsharing.org/collection/MIBBI [Last accessed 27 August 2021]
[27] Re3data, Metrics, Metadata Standards, https://www.re3data.org/metrics/metadataStandards [Last accessed 01 February 2021]

Table 3 – Metadata attributes

Category	Label	Definition	Considerations
Instantiation	Date	A point or period of time associated with an event in the lifecycle of the resource. Includes the following sub-elements: • Date Submitted • Date Issued • Date Accepted • Date Copyrighted • Date Modified	Should use the ISO8601 format. Data Modified should be used in conjunction with versioning.
	Date Available	Date that the resource became or will become available.	Appropriate security/publishing mechanisms should be set in place to ensure that no user has access to the resource before publication date.
	Format	The file format, physical medium or dimensions of the resource. Includes the following sub-elements: • Has Format • Extent • Medium	Can use a list of open formats (e.g., Format Descriptions[28] and openformats.org (accessible through Wikipedia) "Extent" can be used to validate consistency, believability and completeness.
	Identifier	An unambiguous reference to the resource within a given context. Includes the following sub-elements: • Bibliographic Citation • DOI	Tools for translating one bibliographic reference format to the other should be provided. Will produce persistent identifiers (DOIs) for research data and other research outputs.
	Language	The language of the resource.	Should use ISO 639-2.
Intellectual Property	Contributor	An entity responsible for making contributions to the resource.	
	Creator	An entity primarily responsible for creating the resource.	
	Publisher	An entity responsible for making the resource available.	

[28] Sustainability of Digital Formats, Format Descriptions, https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml [Last accessed 01 February 2021]

	Provenance	A statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.	
	Rights	Information about rights held in and over the resource. Includes the following sub-elements: • License • Access Rights • Right Holder	License should be selected from a standardised list. Access Rights should be used in conjunction with SLICES property - Privacy Level
Content	Source	A related resource from which the described resource is derived.	
	Subject	The topic of the resource.	
	Title	A name given to the resource.	
	Alternative	An alternative name for the resource.	
	Description	An account of the resource. Includes the following sub-element: • Abstract	
	Type	The nature or genre of the resource.	Recommended practice is to use a controlled vocabulary, such as the DCMI Type Vocabulary.
	Audience	A class of agents for whom the resource is intended or useful. Includes the following sub-elements: • Education Level • Mediator	A vocabulary of audiences should be compiled. The user groups defined in Section **Error! Reference source not found.**, including their subcategories, can be used.
	Instructional Method	A process used to engender knowledge, attitudes and skills that the described resource is designed to support. Includes the following sub-element: • Coverage	
	Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource or the jurisdiction under which the resource is relevant.	
	Accrual Method	The method by which items are added to a collection.	

	Accrual Periodicity	The frequency with which items are added to a collection.	Should be validated against the actual data using ML approaches to improve Data Quality – Timeliness/Believability
	Accrual Policy	The policy governing the addition of items to a collection.	Should be made compulsory, if SLICES property consent for Completeness is provided.
	Relation	A related resource. Includes the following sub-elements: • Conforms to • Has Part • Has Version • Is Version Of • Is Format Of • Is Part Of • Is Referenced By • Is Required By • Is Replaced By • References • Replaces • Requires • Source	A list of standards should be used for "Conforms to". Versioning is essential to develop correct referencing and compatibility, especially for datasets that may change over time "Is Referenced By", "References", "Is Required By" should be calculated.
SLICES-specific	Consents	Creator's consents on aspects such as Completeness and Timeliness. Currently, the following consents have been identified: • Consent for personal data contained in project • Consents of the purposes of the processing operations • Consent for Completeness of data • Consent for Timeliness of data	
	Auto	A structured descriptor added automatically by the system	This may include specific key-value properties related to various types of processing.
	Privacy Level	Privacy level zone as defined by the data management framework. Includes elements such as:	• Private: access only to creator user, overrides any other setting

		• Shared List • Access Modifier	• Shared Organisations: shared with all users of specified organisations, overrides public • Shared Users: access only to selected users, overrides other properties besides private modifier • Public: access to anyone
	Keywords	A list of keywords that can be used for user queries.	

Several standards should be used to support the properties shown above. These standards permit to exchange data with standardised codes. The standards are:

- ISO 3166: Codes representing the names of countries.
- ISO 639-3: Codes representing the names of languages.
- ISO 8601-1: Representation of date and time.
- DCMI-Period: DCMI Period Encoding Scheme.
- DCMI-Point: DCMI Point Encoding Scheme.

The chosen metadata format standard will allow the management of the data of the SLICES research infrastructure, the reuse of these data, the interoperability with external services, systems and infrastructures. Nevertheless, the evaluation of the costs to implement such metadata format standard in the SLICES research infrastructure leads to the question if it is better in terms of costs to use an existing platform able to manage research data, like CKAN, Zenodo, DSpace and Figshare. An analysis of these platforms is done in the following section and the project consortium will decide which solution is the most appropriate following factors like requirements, maintenance and costs.

4.5. Intra/Inter-operability

The interoperability is the key element to ensure the interactions with other infrastructure and systems. An efficient management of the metadata ensures the consistency between the different components of the SLICES research infrastructure, which is called intraoperability, and the interoperability with external systems or infrastructures. The previous section has suggested a metadata format standard guaranteeing the reusability of the data, an easy access to them and the intra/inter-operability. But it is not sufficient, because the intra-interoperability can be broken when inconsistences are appearing in the data. A list of typical errors and possible solutions linked to the intra-interoperability is displayed below:

- Lack of synchronisation between a node and the cloud data management: This error can be corrected through updates of the data model or format at planned maintenance time. This means that there are no data in transit during the maintenance time.

- A user is entering incorrect data in a node: Basically, a node has not detected the wrong inputs. The same solution based on updates during the maintenance time is also the good mitigation to this error.
- The metadata model is changed and the current metadata are not correctly transformed to the new metadata model: This error is more complicated to handle. A proactive solution is to seriously plan the upgrades of the metadata model with sufficient time to test and validate these changes brought by the upgrades. A more automated solution requires the use of complex data processing mechanisms to mitigate the incompatibilities between different versions of a data model.

The previous sections of this document already addressed some aspects of the interoperability. Different tools permit to improve the interoperability by providing different services to the users like the data creation, the data transfer, the search, the download and the data management based on the metadata. An analysis of different platforms dedicated to the research data management was realised based on these features and other characteristics such as the supporting community, the query and retrieval methods, the APIs. A description of each evaluated platform is given below:

Table 4 – Platform comparison

Figshare[29]	• Allows the discovery, citing, sharing and uploading of research output. • Allows for uploading up to 5GB single file of any format + 20GB of free private space. Provides a desktop uploader application. • Generates a DOI for the researcher's work/ allows for reserving a DOI before releasing the data. • Allows for collaboration through collaborative spaces/ private link sharing. • Enables the customisation of showcase portals for presenting public research outputs. • Provides reporting and statistics information at various levels, e.g., researcher or object. • Provides a REST API for automating research workflows.
Dataverse[30]	• Provides counts for web visibility, academic credit and citations. • Accepts citations for datasets and files, such as EndNote XML, RIS Format, or BibTeX. • Can expose data to other systems using a variety of metadata formats. • Provides REST APIs such as: Search API, Data Deposit API, Data Access API, Metrics API, etc. • Is discoverable by Google with adherence to Schema.ord JSON-LD. • Provides login using institutional providers or GitHub, Google etc. • Supports a Data Explorer tool for preview and analysis of tabular files. • Supports file downloads of tabular data in a variety of formats, such as TSV, RData. • Provides dataset versioning capabilities and file access control. • Allows integration with Dropbox for retrieving already uploaded files. • Can operate using a filesystem or object storage.

[29] Figshare, https://figshare.com/, [Last accessed 27 August 2021]
[30] Dataverse project, https://dataverse.org/, [Last accessed 27 August 2021]

	• Is able to pull header metadata from Astronomy FITS files.
Mendeley Data[31]	• Enables researchers to control the full lifecycle of research data, i.e., collection and discovery. • Enables the creation of projects where researchers can collaborate/ share and annotate their data. • Integrates with Dropbox, Google Drive, Box and Azure for retrieving already uploaded files. • Allows for institutions to retain data on their own servers. • Published dataset metadata are aggerated to DataCite's metadata index and to the OpenAIRE portal. • Supports the harvesting of public dataset records using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard. • Data is stored on Amazon's S3 servers and archived with Data Archiving and Network Services (DANS).
Open Science Framework[32]	• A collaboration tool/ workflow system that enables researchers to collaborate on projects and publish findings for dissemination. • Provides a centralised repository for project files, data and code, with version control capabilities. • Supports file access control for selecting which parts of a project are public or private to the team. • Enables the team to keep logs, notes and track their progress. • Offers project analytics for measuring the impact of the project using citation, downloads and project access count. • Integrates with Dropbox, GitHub or Figshare. • Can publish reports in Google Scholar, Crossref and ORCID. • Supports search integration with platforms such as Mendeley, DataCite and ZOTERO.
Zenodo[33]	• Accepts research outputs of all research fields and all file formats. • Assigns publicly available DOI to all works and supports harvesting of all content via the OAI-PMH protocol. DOI is available before publishing. • Integrates with OpenAIRE. • Enables uploading with a variety of licenses and access levels. • Citation data is sent to DataCite. • Statistics enable the tracking of visits, visitor type, country and referrer domain. • Includes citation sources such as: NASA Astrophysics Data System, DataCite and Crossref.
Code Ocean[34]	• Facilitates the creation, organisation and dissemination of computational research in a collaborative manner. • Standardises research workflows and reproduction of computational discoveries.

[31] Mendeley Data, https://data.mendeley.com/, [Last accessed 27 August 2021]
[32] OSF platform, https://osf.io/, [Last accessed 27 August 2021]
[33] Zenodo platform, https://zenodo.org/, [Last accessed 27 August 2021]
[34] Code Ocean platform, https://codeocean.com/, [Last accessed 27 August 2021]

	• Provides reproducible Capsules, which are entities that contain code, data, environment setup and any associated results, while also being versioned. • Integrates with data tools such as RStudio, Jupyter, Shiny Terminal and Git. • Allows access to any size/ type of data – can generate docker environments. • Allows researchers to create and share results in easy-to-use web analytic apps. • Allows researchers to utilise public capsules from a repository that provides outputs from the global community. • Deployed on the AWS cloud with a dedicated VPC.
4TU.ResearchData[35]	• International data repository for science, engineering and design. • Enables the curation, long-term access and preservation of research datasets. • The empowering technology of this platform is Figshare, while all data is hosted by the TU Delft Library. • DataverseNL enables researchers to create project spaces. • Project spaces provide data file management, metadata and documentations, version control, collaboration tools, storage and backup. • Every dataset is provided with a DOI for linking or citing the dataset in publications. • A DOI can be reserved prior to dataset publication. • Allows researchers to find and reuse a large number of published datasets through a digital library. • Uses OPeNDAP, which is a protocol that allows the use of data from a server without the need of downloading the data files.
ANDS[36]	• The Australian National Data Service, which helps Australian researchers to publish, discover and access research outputs. • Enables access to data coming from more than 100 Australian research organisations. • The platform does not hold the actual data, but descriptors to these data. The original data is hosted by the publishing partners and contributors. • Institutions have to provide their own metadata to the Research Data Australia registry. • The four main services provided are the research data discovery portal, DOI Service (DataCite), handle service and the research vocabularies service. • Other tools include connecting and linking data, assistance with harvesting, through the utilisation of various protocols and schemas. • Data information are syndicated to global data citation indexing systems.
Dryad Digital Repository[37]	• Completely open source with code available on GitHub and is based on Stash, which is a data publication platform.

[35] 4TU.ResearchData, https://data.4tu.nl/info/en/, [Last accessed 27 August 2021]
[36] Australian National Data Service, https://www.ands.org.au/, [Last accessed 27 August 2021]
[37] Dryad Digital Repository, https://datadryad.org/stash, [Last accessed 27 August 2021]

	• The preservation of data is up to the underlying repository with which Dryad is integrated. • Each dataset has its own landing page that presents all descriptive and administrative metadata – can be also downloaded as PDF. • Integrates with any SWORD/OAI-PMH-compliant repository. • Supports DataCite and can also be configured to support other schemas. • DOIs are provided to all datasets. • Easy sign-in can be enabled via institutional providers and ORCID. • Supports Drag and Drop uploads with simple navigation. • Allows search by subject, filetype, keywords, campus, location, etc. • Allows for building relationships between datasets from other publications.
Re3data [38]	• Global data registry repository offering data from a vast range of academic disciplines. • Provided by DataCite as a service. • Provides access to datasets to researchers, publishers, funding bodies and scholarly institutions. • Indexes more than 2450 research data repositories. • Allows developers to access information via a REST API.

The maximalisation of the interoperability can be achieved by using the following aspects:

- FAIR principles: The implementation should follow the FAIR principles.
- APIs: REST APIs can efficiently expose the metadata to the users.
- Repository interoperability: The goal is to enhance the dissemination of the research data. This can be done through a metadata harvesting protocol like OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)[39].
- Flexible referencing: The transformation from a given metadata to other well-known metadata formats should be implemented.
- Serialisation: The serialisation is supporting at least JSON, YAML and XML. Other formats could be supported in the future.
- Complex querying: The metadata can be searched by the users using simple and advanced queries for free text, combinations of filters for example.
- User experience: The principles of user experience design (UXD) should be employed during the development of user interfaces. The functionality provided by the underlying infrastructure should be accessible by the users in a friendly way.

The results of the analysis show that Zenodo and Figshare are the two platforms for the research data management which are supporting all the requirements listed above. CERN is hosting Zenodo; this fact should ensure the sustainability of the platform in the long run. It is also possible to upload files from GitHub to Zenodo; this point is particularly important for the source code of the different software to be installed in the research infrastructure.

At the end, SLICES has two possibilities in terms of research data management: to design a dedicated research data management platform or use an existing one like Zenodo or Figshare. There are pros

[38] https://www.re3data.org/, [Last accessed 27 August 2021]
[39] The Open Archives Initiative Protocol for Metadata Harvesting, https://openarchives.org/OAI/openarchivesprotocol.html [Last accessed 01 February 2021]

and cons to both solutions, based on different criteria like the costs of development, the requirements of all the stakeholders. The community will adopt more easily a platform meeting all its requirements.

4.6. Analytics

The notion of analytics consists to deploy techniques related to statistics, machine learning and artificial intelligence. The results of analytics are based on the proper interpretation and the visualization of the data. One of the objectives of SLICES is to allow the researchers to discover hidden patterns in their data. Different steps are required to realise a good data analysis. First of all, the integration of different data sources permits the building of complete datasets. Generally, cleaning and pre-processing are required: a specific data model is used as reference. Different tools provide statistical methods and techniques like algorithms for predictive modelling. Different kinds of visualisations are available for the users.

To achieve this, SLICES intends to install and use several open-source and commercial tools dedicated to the analysis of the data, in particular for data mining, machine learning and artificial intelligence. Four categories of analytics will be deployed in the SLICES research infrastructure:

- Descriptive analytics: The objective of this analytics is to know what has happened and what is currently happening in the data. Standard statistical descriptors and visualisations are the main elements of the descriptive analytics, allowing at the end to recognise the distribution, the dispersion and the mode of the data.
- Diagnostic analytics: The goal of the diagnostic analytics is to determine why an event in the data has happened. It can be done through descriptive analytics to detect the irregularities in the data.
- Predictive analytics: The future data can be predicted through such analytics based the observation of the trends and the historical data. Several algorithms are used in the context of predictive analytics such as regressions, forecasting, classification, clustering, association outlier analysis and text mining.
- Prescriptive analytics: This analytics permits to select the best action if several solutions occur, in function of the resource optimisation. The prescriptive analytics employs algorithms such as optimisation, multiple-criteria decision analysis and simulation.

The various parts of the data management and the analysis can be handled by all the families of algorithms. An example is the attribution of values in a dataset with missing values; in this example, a pre-processing can consist to use a predictive model to set the missing values. To group all the algorithms used across the research infrastructure, an algorithm repository can be put in place for all the software components.

4.7. Other data management issues

This section presents important objectives to be reach by the data management infrastructure.

4.7.1. Naming Conventions

The consistency of the names given to the files is very important to retrieve the resources in the right location. The data provider is of course responsible to give a correct name to each research data files. SLICES will provide guidelines and conventions about the file naming. The initial recommendations are the following:

- Name length: A maximum length is defined to ensure the interoperability with all the systems.

- Date format: The display of the dates is done in a chronological order like the YYYYMMDD format.
- Leading zeros: The leading zeros permit to create an ascending order of numbers corresponding to the alphabetical order.
- Naming scheme: It should be consistent everywhere. Typically, the use of use spaces or punctuation symbols are prohibited. This guarantees the interoperability between different systems.
- Order: The naming scheme should permit the clear distinction of different groups of files and indicate which file is the first one.

4.7.2. File Organisation

An efficient file organisation allows a quick retrieval of a resource. SLICES will provide guidelines on file organisation which will enhance the consistency of the data structure. The initial recommendations are defined here:

- Hierarchical structure: A minimal set of folders should be included in the hierarchical structure:
  - Data: This folder contains all the inputs for data not associated to an experiment.
  - Experiments: Each experiment has its own folder for example “exp01”. An experiment folder should contain at least the following sub-folders:
    - input data: This folder contains all the data for an experiment.
    - software: There are in this folder all the software components, models and links needed for the experiments.
    - deployment: The folder contains all the steps to realise the experiment.
    - output data: The results of the experiment are stored in this folder.
  - Relationships: The references to relationships with other research data are stored here, based on the metadata specification.
  - Dissemination: In this folder, there are the material used for the dissemination like articles, press releases and presentations.
  - Miscellaneous: The folder is dedicated for all the rest.
- Folder naming: The naming of the folders should follow the naming conventions.

4.7.3. Data Storage

Basically, the SLICES research infrastructure will be distributed with new capabilities which will host thousands and their data. First estimations indicate a number of users up to 5’000. Each user will have up to 50 GB on the different nodes and up to 1 TB on the central cloud. So, the total size of the storage is between 0.25 and 1 PB in the datacentres of the nodes and 5 PB in the datacentre on the central cloud.

4.8. Resource Allocation

The sustainability of the SLICES research infrastructure is guaranteed by some constraints put to the users. For instance, the limitation of the research data storage per user, the retention period defined by the appropriate policy, the costs are some elements to be taken into account in terms of constraints imposed to the users.

5. FAIR data principles

The FAIR (Findable, Accessible, Interoperable, and Reusable) principles were designed to ease the access to the data to the research community in a sustainable manner. Indeed, the generation of data in the context of digital sciences is increasing and requires more and more automated computation to index the data correctly. The FAIR principles allow better transparency and knowledge in the research process. The data producers and publishers should follow the FAIR principles to facilitate the data-driven science and the automatic computation of the shared data.

The SLICES-SC project is applying the four FAIR principles as presented in this section.

FINDABLE

The data should be findable if the experimenter will reuse them later. To achieve this objective, several requirements should be taken into account:

1. A globally unique and persistent identifier should be assigned to the data and related metadata. SLICES-SC will use a Digital Object Identifier (DOI) when the data are uploaded.
2. Rich metadata describe the data. In SLICES-SC, the metadata will use the enhanced version of the DublinCore format.
3. The metadata contain the identifier of the data that the metadata refer to. The metadata in SLICES-SC will contain the DOI assigned to the data.
4. Metadata are stored and indexed. In SLICES-SC, a resource discovery service will permit to search the data thanks to queries containing keywords or metadata.

ACCESSIBLE

This principle ensures that the mean to access the data is correctly described. It can include of course the authentication and the authorization. Several elements will be put in action in SLICES-SC:

1. A standardized communication protocol is used to retrieve the data or the metadata through their identifier. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) will be employed in SLICES-SC. REST APIs with payloads using JSON, XML and YAML formats will be also supported.
2. The standardized communication protocol is free, open and implementable by everybody. The mentioned OAI-PMH protocol answers to this requirement. By definition, the REST APIs are also free, open and universally implementable.
3. The communication protocol offers the authentication and the authorization. In SLICES-SC, the access to non-public data will be protected by authentication and authorization. On the other hand, the metadata will be completely open and so, no authentication and authorization are required.
4. The metadata stay available, even if the data are removed. All the metadata will be stored in a dedicated data store in the SLICES infrastructure during the lifetime of the infrastructure and the project. So, the lifetime of the metadata will depend on the lifetime of the host data repository.

INTEROPERABLE

The data alone are not very useful. So, they need to be stored, analysed and processed by different applications or services. In consequence, the interoperability and the integration of the data are two important points to follow. In the context of SLICES-SC, the interoperability will be facilitated through the following topics:

1. The knowledge representation is realized through a formal, accessible, shared and broadly applicable language. As mentioned earlier, the enhanced version of the DublinCore format,

which is standardized as ISO 15836, ANSI/NISO Z39.85 and IETF RFC 5013, will be used for the metadata. Concerning the REST APIs, JSON, XML and YAML will be the reference formats. It is not excluded to translate or map them in other formats.

2. The data and the metadata should use vocabularies following the FAIR principles. In SLICES-SC, the chosen vocabularies are ISO 3166 (country codes), ISO 639-3 (language codes), ISO 8601-1 (date and time representation), DCMI-Period and DCMI-Point. Other standards compliant with FAIR principles could be added in the future.
3. Data and metadata should include qualified references to other data and metadata. In SLICES, a metadata property named "Relation" will be used to convey the references.

REUSABLE

The re-usability of the data is a key element in the FAIR principles. This means that the data and the metadata should be well-described. Several requirements should be respected:

1. Data and metadata are described with accurate and relevant attributes. In the context of SLICES-SC, the use of the enhanced version of the DublinCore format will ensure a good description and efficient discovery.
2. The data and the metadata are published with a clear data usage license. A list of licenses will be provided by SLICES-SC to the researchers when they will build their metadata. The Public Domain license will be the default license.
3. The origin of the data and metadata is provided accurately. In SLICES-SC, the experimenters should provide any changes concerning the data ownership. This is an important element to ensure the authenticity and the integrity of the published data.
4. The data and metadata follow the standards relevant to the research community. This requirement is met in SLICES by the DublinCore format which is already standardized as ISO 15836, ANSI/NISO Z39.85 and IETF RFC 5013.

6. Compliance

The management of data, and especially the personal data, is ruled by a legal framework at both European and national level. The compliance of the project's activities with these laws is of higher importance for the partners and this is the reason why the consortium defined a series of internal rules and measures to the guarantee the compliance the European and national laws. Among these laws, and considering the project's activities' nature, two EU regulation/directive have been considered: the General Data Protection Regulation (GDPR) and the E-Privacy Directive.

6.1. General Data Protection Regulation (GDPR)

With the growing exchange of data, and especially of personal data, the EU paid, over the last years, special attention to the protection of personal data, settling a legal framework, stricter and stricter, in order to ensure its protection. Considered as a fundamental right in Article 8(1) of the Charter of Fundamental Rights of the European Union[40], the Treaty on the Functioning of the European Union[41] (TFEU) recognizes that everyone has the right to the protection of personal data concerning him or

[40] Charter of Fundamental Rights of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT [Last accessed 19 July 2021]
[41] Consolidated version of the Treaty on the Functioning of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT [Last accessed 19 July 2021]

her. It also answers an increasing demand from individuals for further protection and transparency in the treatment of their personal data.

In this context, the EU proposed a new regulation, called the General Data Protection Regulation (Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016), commonly called "GDPR". This regulation entered into force in 2016 and as a regulation, became directly binding and applicable to all the EU and EEA member states, since 25 May 2018. This regulation replaced the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data[42].

The GDPR aims to harmonise the European legal panorama in terms of personal data protection and reform of European legislation which became obsolete due to the digital explosion, the emergence of new usages and the implementation of new economic models.

In addition to the definition of the key elements and actors to be considered when it refers to personal data protection, this regulation states 5 main principles to comply with:

- **The principle of finality**: the collection and processing of personal data must be done only for a very specific, legal and legitimate purpose;
- **The principle of proportionality and relevance**: the information recorded must be relevant and strictly necessary with regard to the purpose of the data collection and processing;
- **The principle of a limited retention period**: the data collected must be stored for a limited period of time. A precise retention period must be set, depending on the type of information recorded and the purpose of the data processing;
- **The principle of security and confidentiality**: the security and confidentiality of the information stored must be guaranteed and only authorized persons must have access to the information;
- **Individuals' rights**: the right of any individual person must be safeguarded. At any time, a person can ask for the correction or erasing of his/her personal data.

6.2. E-privacy Directive

In addition to the GDPR, the E-Privacy Directive[43] is another piece of law to be considered since it concerns the processing of personal data and the protection of privacy in the electronic communications sector. Usually referred to as the "E-privacy Directive", this Directive is an amendment to the previous Directive 2002/58/EC. The aim of this Directive is to provide a legal framework to the processing of personal data and the protection of privacy including provisions on[44]:

- the security of networks and services;
- the confidentiality of communications;

[42] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31995L0046 [Last accessed 19 July 2021]

[43] Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009L0136 [Last accessed 21 July 2021]

[44] European Data Protection Supervisor: E-privacy Directive 2009/136/EC, https://edps.europa.eu/data-protection/data-protection/glossary/e_en#e-privacy_directive2009-136-ec https://edpb.europa.eu/about-edpb/about-edpb/members_en [Last accessed 21 July 2021]

- access to stored data;
- processing of traffic and location data;
- calling line identification;
- public subscriber directories; and
- unsolicited commercial communications ("spam").

Compared to the previous Directive, this amendment strengthened the protection of personal data, including a rule requiring the notification of data breaches as well as the extension of the Directive in order to cover various electronic tags, among others.

This directive is particularly important for the use of SLICES-SC website and any online activity.

6.3. National regulations

6.3.1. EU countries

The GDPR directly applies in the EU countries since 25 May 2018 without the need to be transposed into national law. However, each EU member state can decide to set up additional regulations and bodies in order to guarantee the implementation of the GDPR and to increase the level of protection of personal data, especially in the field of research activities. Data transfers between the EU countries do not require the use of transfer tools.

The EU regulation is the minimum requirement in terms of personal data protection, which means that any additional national will reinforce the EU regulation or has been voted before the GDPR. For this deliverable, as well as the D9.1, we focused on the EU legislation but we have identified the different national bodies in charge of the compliance of the GDPR and other law related to the protection of personal data. We also listed some of the laws and decrees applying at national level.

Regarding the SLICES-SC project, 9 EU countries and 1 non-EU country are represented in the consortium and as consequence 10 national law framework may, potentially, apply the project's activities. They are listed in the table below. Those entities will be our national contact points in case a question about the personal data protection is raised during the project but cannot be fully answered/covered by the project and partners' DPOs.

Table 5 – Countries in SLICES-SC

Countries represented in SLICES-SC	National Committees for Personal Data Protection[45]
Belgium	Autorité de la protection des données - Gegevensbeschermingsautoriteit (APD-GBA)
Finland	Office of the Data Protection Ombudsman
France	Commission Nationale de l'Informatique et des Libertés (CNIL)
Germany	Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit
Greece	Hellenic Data Protection Authority

[45] European Data Protection Board Members, EDPB, https://edpb.europa.eu/about-edpb/about-edpb/members_en [Last accessed 19 July 2021]

Hungary	Hungarian National Authority for Data Protection and Freedom of Information
Italy	Garante per la protezione dei dati personali
Poland	Urząd Ochrony Danych Osobowych
Spain	Agencia Española de Protección de Datos (AEPD)
Switzerland	Not member of the EDPB

The table below lists the main national laws and decrees adopted in the field of personal data protection. Other laws exist and apply to specific domains.

Table 6 - National regulations

Countries represented in SLICES-SC	List of national laws complementary to the GDPR & E-privacy Directive (transposition)
Belgium	Loi relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel / Wet betreffende de bescherming van natuurlijke personen met betrekking tot de verwerking van persoonsgegevens (30 July 2018). Federale Overheidsdienst Justitie En Federale Overheidsdienst Economie, K.M.O., Middenstand En Energie - 31 Mei 2011. - Wet houdende diverse bepalingen inzake telecommunicatie. *(Official publication: Staatsblad; Publication date: 2011-06-21; Page: 36503-36508)* Federale Overheidsdienst Economie, K.M.O., Middenstand En Energie - 14 November 2011. - Wet tot wijziging van de wet van 13 juni 2005 betreffende de elektronische communicatie wat de bereikbaarheid van de nooddiensten betreft. *(Official publication: Staatsblad; Publication date: 2011-12-02; Page: 71124-71125)* Service Public Federal Economie, P.M.E., Classes Moyennes Et Energie - 10 Juillet 2012. - Loi portant des dispositions diverses en matière de communications électroniques. *(Official publication: Staatsblad ; Number: 240 ; Publication date: 2012-07-25 ; Page: 40969-41014)*
Finland	Viestintämarkkinalaki / Kommunikationsmarknadslag (393/2003) 23/05/2003, muutettu viimeksi / senast ändring genom (363/2011) 08/04/2011. *(Official publication: Suomen Saadoskokoelma (SK) ; Number: 393/2003 ; Publication date: 2003-05-30 ; Page: 01895-01926)* Sähköisen viestinnän tietosuojalaki / Lag om dataskydd vid elektronisk kommunikation (516/2004) 16/06/2004, muutettu viimeksi / senast ändring genom (365/2011) 08/04/2011. *(Official publication: Suomen Saadoskokoelma (SK) ; Number: 516/2004 ; Publication date: 2004-06-23 ; Page: 01453-01466)*

	Laki viestintämarkkinalain 134 §:n muuttamisesta annetun lain voimaantulosäännöksen muuttamisesta / Lag om ändring av ikraftträdandebestämmelsen i lagen om ändring av 134 § i kommunikationsmarknadslagen. *(Official publication: Suomen Saadoskokoelma (SK) ; Number: 731/2010 ; Publication date: 2010-08-31 ; Page: 02496-02496)* Laki viestintämarkkinalain 134 §:n muuttamisesta / Lag om ändring av 134 § i kommunikationsmarknadslagen. *(Official publication: Suomen Saadoskokoelma (SK) ; Number: 732/2010 ; Publication date: 2010-08-31 ; Page: 02497-02498)* Tietoyhteiskuntakaari / Informationssamhällsbalk (917/2014) 07/11/2014. *(Official publication: Suomen Saadoskokoelma (SK) ; Number: 917/2014 ; Publication date: 2014-11-12)*
France	Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés. Article 5 du décret n° 2009-834 du 7 juillet 2009 portant création d'un service à compétence nationale dénommé « Agence nationale de la sécurité des systèmes d'information ». *(Official publication: Journal Officiel de la République Française (JORF) ; Publication date : 2009-07-08)* Article 17 de la loi n°2011-302 du 22 mars 2011 portant diverses dispositions d'adaptation de la législation au droit de l'Union européenne en matière de santé, de travail et de communications électroniques. *(Official publication: Journal Officiel de la République Française (JORF) ; Publication date : 2011-03-23)* Article 21 de la loi n° 2011-901 du 28 juillet 2011 tendant à améliorer le fonctionnement des maisons départementales des personnes handicapées et portant diverses dispositions relatives à la politique du handicap (1). *(Official publication: Journal Officiel de la République Française (JORF) ; Publication date: 2011-07-30)* Le titre Ier de l'ordonnance n° 2011-1012 du 24 août 2011 relative aux communications électroniques. *(Official publication: Journal Officiel de la République Française (JORF) ; Publication date : 2011-08-26)* Décret n° 2012-436 du 30 mars 2012 portant transposition du nouveau cadre réglementaire européen des communications électroniques. *(Official publication: Journal Officiel de la République Française (JORF) ; Publication date: 2012-03-31)* Décret n° 2012-488 du 13 avril 2012 modifiant les obligations des opérateurs de communications électroniques conformément au nouveau cadre réglementaire. *(Official publication: Journal Officiel de la République Française (JORF) ; Publication date: 2012-04-15)*

	Décret n° 2019-536 du 29 mai 2019 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés
Germany	Gesetz zur Änderung telekommunikationsrechtlicher Regelungen. *(Official publication: Bundesgesetzblatt Teil 1 (BGB 1) ; Number: 19 ; Publication date: 2012-10-09 ; Page: 00958-00997)* (Draft legislation) Entwurf eines Gesetzes zur Regelung des Datenschutzes und des Schutzes der Privatsphäre in der Telekommunikation und bei Telemedien (09.03.2021)
Greece	Ρυθμίσεις Ηλεκτρονικών Επικοινωνιών, Μεταφορών, Δημοσίων Έργων και άλλες διατάξεις. *(Official publication: Εφημερίς της Κυβερνήσεως (ΦΕΚ) (Τεύχος Α); Number: 82; Publication date: 2012-04-10; Page: 02131-02264).* Προστασια του ατομου απο την επεξεργασια δεδομενων προσωπικου χαρακτηρα (Νόμος 2472/1997). Προστασία δεδομένων προσωπικού χαρακτήρα και της ιδιωτικής ζωής στον τομέα των ηλεκτρονικών επικοινωνιών και τροποποίηση του ν. 2472/1997 (Νόμος 3471/2006 (Φεκ 133/Α'/28.6.2006)). Διατήρηση δεδομένων που παράγονται ή υποβάλλονται σε επεξεργασία σε συνάρτηση με την παροχή διαθέσιμων στο κοινό υπηρεσιών ηλεκτρονικών επικοινωνιών ή δημόσιων δικτύων επικοινωνιών, χρήση συστημάτων επιτήρησης με τη λήψη ή καταγραφή ήχου ή εικόνας σε δημόσιους χώρους και συναφείς διατάξεις (Νόμος Υπ' Αριθ. 3917 / 2011). Ταυτοποίηση των κατόχων και χρηστών εξοπλισμού και υπηρεσιών κινητής τηλεφωνίας και άλλες διατάξεις (Νόμος 3783/2009). Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα, μέτρα εφαρμογής του Κανονισμού (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα και ενσωμάτωση στην εθνική νομοθεσία της Οδηγίας (ΕΕ) 2016/680 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 27ης Απριλίου 2016 και άλλες διατάξεις (Νόμος 4624/2019)
Hungary	A Kormány 229/2008. (IX. 12.) Korm. rendelete az elektronikus hírközlési szolgáltatás minõségének a fogyasztók védelmével összefüggõ követelményeirõl. (Official publication: Magyar Közlöny ; Page: 14748-14754) A Kormány 180/2004. (V. 26.) Korm. rendelete az elektronikus hírközlési feladatokat ellátó szervezetek és a titkos információgyûjtésre, illetve titkos adatszerzésre felhatalmazott szervezetek együttmûködésének rendjérõl. (Official publication: Magyar Közlöny ; Page: 07246-07250)

	2010. évi LXXXII. törvény a médiát és a hírközlést szabályozó egyes törvények módosításáról. *(Official publication: Magyar Közlöny ; Page: 22319-22357)* 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról. *(Official publication: Magyar Közlöny ; Publication date: 1001-01-01 ; Page: 25449-25485)* 2011. évi CVII. törvény egyes elektronikus hírközlési tárgyú törvények módosításáról. *(Official publication: Magyar Közlöny ; Publication date: 1001-01-01 ; Page: 25189-25248)* A Nemzeti Média- és Hírközlési Hatóság elnökének 4/2012. (I. 24.) NMHH rendelete a nyilvános elektronikus hírközlési szolgáltatáshoz kapcsolódó adatvédelmi és titoktartási kötelezettségre, az adatkezelés és a titokvédelem különleges feltételeire, a hálózatok és a szolgáltatások biztonságára és integritására, a forgalmi és számlázási adatok kezelésére, valamint az azonosítókijelzésre és hívásátirányításra vonatkozó szabályokról. *(Official publication: Magyar Közlöny ; Publication date: 1001-01-01 ; Page: 00513-00518)*
Italy	Decreto Legislativo 10 agosto 2018, n. 101 Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonche' alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati) Modifiche al Decreto Legislativo 30 giugno 2003, n. 196, recante codice in materia di protezione dei dati personali in attuazione delle direttive 2009/136/CE, in materia di trattamento dei dati personali e tutela della vita privata nel settore delle comunicazioni elettroniche, e 2009/140/CE in materia di reti e servizi di comunicazione elettronica e del regolamento (CE) n. 2006/2004 sulla cooperazione tra le autorità nazionali responsabili dell'esecuzione della normativa a tutela dei consumatori. *(Official publication: Gazzetta Ufficiale della Repubblica Italiana ; Number: 126 ; Publication date: 2012-05-31)*
Poland	Ustawa z dnia 29 sierpnia 1997 r. o ochronie danych osobowych. *(Official publication: Dziennik Ustaw ; Number: Dz. U. z 2002 r. Nr 101)* Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 29 kwietnia 2004 r. wsprawie dokumentacji przetwarzania danych osobowych oraz warunków technicznych i organizacyjnych, jakim powinny odpowiadać urządzenia i systemy informatyczne służące do przetwarzania danych osobowych. *(Official publication: Dziennik Ustaw ; Number: Nr 100, poz. 1024*)

Spain	Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Real Decreto 726/2011, de 20 de mayo, por el que se modifica el Reglamento sobre las condiciones para la prestación de servicios de comunicaciones electrónicas, el servicio universal y la protección de los usuarios, aprobado por Real Decreto 424/2005, de 15 de abril. *(Official publication: Boletín Oficial del Estado (B.O.E) ; Number: 123/2011 ; Publication date: 2011-05-24 ; Page: 51433-51453)*

6.3.2. Non-EU country: Switzerland

In addition to the 10 EU countries, the consortium is represented, through MI and IoT Lab, in non-EU country where the GDPR does not apply: Switzerland.

It is important to remind that, in this field, the Swiss law is similar to the EU one and European Commission[46] recognised that Switzerland has an adequate level of data protection, and for which transfers of personal data do not require supervision by transfer tools in accordance with the Commission Decision of 26 July 2000[47].

In Switzerland, the protection of personal data is established by Article 13 of the Federal Constitution[48]. This article states that "(1) every person has the right to privacy in their private and family life and in their home, and in relation to their mail and telecommunications and (2) every person has the right to be protected against the misuse of their personal data".

Another important text is the Federal Act on Data Protection (FADP)[49], in force since 1 July 1993. The corresponding ordinance (DPO) regulates the details[50]. Other laws at a cantonal level contain numerous provisions which further specify the requirements associated to personal data protection. A new version of the FADP has been also defined to better align the Swiss dispositions with those contained in the GDPR. It was approved by the Swiss Federal Parliament on 27 September 2020 and is expected to come into force in 2022.

The authority in charge of the application of the personal data protection legislation is the Federal Data Protection and Information Commissioner[51]. This entity will be our national contact point for Switzerland in case a question about the personal data protection is raised during the project in relation with activities in Switzerland but cannot be fully answered/covered by the project and Swiss partners' DPOs.

[46] https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en [Last accessed 19 July 2021]
[47] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000D0518 [Last accessed 19 July2021]
[48] Federal Constitution of the Swiss Confederation, of 18 April 1999, https://www.fedlex.admin.ch/eli/cc/1999/404/en#a13 [Last accessed 19 July 2021]
[49] Swiss Federal Act on Data Protection (FADP) of 19 June 1992, https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en [Last accessed 19 July 2021]
[50] Ordinance to the Federal Act on Data Protection (DPO), of 14 June 1993, https://www.fedlex.admin.ch/eli/cc/1993/1962_1962_1962/en [Last accessed 19 July 2021]
[51] Federal Data Protection and Information Commissioner https://www.edoeb.admin.ch/edoeb/fr/home.html [Last accessed 19 July 2021]

6.4. Other Regulations and Sources

In addition to the GPDR, the E-privacy Directive and the national legislation, other regulations have been identified and are relevant for this DMP:

- Directive 2001/29/EC
  - Harmonisation of certain aspects of copyright and related rights in the information society.
- Directive 2000/31/EC
  - Certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

As a H2020 project, the project's personal data protection policy will be ruled also by the recommendations and provisions defined by the H2020 programme. For that, three main documents have been identified:

- **Grant Agreement**. Article 39.2: Processing of personal data by the beneficiaries.
  - *"The beneficiaries must process personal data under the Agreement in compliance with applicable EU and national law on data protection (including authorisations or notification requirements). The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement. The beneficiaries must inform the personnel whose personal data are collected and processed by the Commission. For this purpose, they must provide them with the privacy statement(s), before transmitting their data to the Commission.".*
- **Consortium Agreement** (based on the DESCA model). Article 11.3: Personal Data
  - *"The Parties agree that, except their personnel contact details required for the Project, they will not (I) make available any Personal Data to other Parties or (ii) process any Personal Data on behalf of other Parties. The personnel contact details required for the Project are solely name and communication addresses (such as phone number, email, chat-IDs) provided by the employer, as well as unavailability information for a certain amount of time without the reason for unavailability. If a particular Party requires that any Personal Data, other than the Personal Data referred above, controlled by this Party has to be processed by another Party, the controlling Party will, before any Personal Data is made available, make sure to enter into all data processing agreements and obtain all consents required in accordance with applicable data protection laws. The exchange and processing of Parties' contact details do not require a specific data processing agreement, and shall be made in accordance with applicable legislation, and only for the purposes of communication within the Project. Nevertheless, in case each Party provides or otherwise makes Personal Data available to any other Party, such Party ("Contributor") represents that, as per applicable data protection laws: (i) it has the authority to disclose the any Personal Data or information, if any, which it provides to the Parties under this Consortium Agreement; (ii) where legally required and relevant, it has a legal ground to provide the Personal Data and information and shall comply with the data protection laws; and (iii) there is no restriction in place that would prevent any such other Party from using the Personal Data and information for the purpose of this Project and the exploitation thereof."*

- **Ethics and data protection**[52]. This is the H2020 guideline about ethics requirements in the field of data protection.
    - One of the important topics to highlight is that in the course of a collaborative project, the project may have joint data controllers and the responsibilities of the partners must be defined in an agreement available to data subjects and provide them with a single point of contact (see section 2.2).

In addition to above-mentioned regulations, other sources have been analysed and considered for the definition of the SLICES-SC personal data policy, such as:

- **European Data Protection Board**: "it is an independent European body, which contributes to the application of data protection rules throughout the European Union, and promotes cooperation between the EU's data protection authorities. The EDPB is composed of representatives of the EU national data protection authorities (national Supervisory Authorities), and the European Data Protection Supervisor (EDPS)"[53]. The EDPB and its guidelines further clarify the legal framework surrounding personal data protection. It also provides frequent publications, guidance, recommendations, etc. in that field, not only at EU level but also national information.
- **European Code of Conduct for Research Integrity**: promoted by ALLEA[54], representing more than 50 academies from over 40 EU and non-EU countries, this code "serves the European research community as a framework for self-regulation across all scientific and scholarly disciplines and for all research settings". It provides with recommendations, for instance, on Data Practices and Management as good research practices.
- **GDPR.eu**: it is the outcome of an EU-funded project, coordinated by the Swiss company PROTON TECHNOLOGIES AG. This website provides many information and resources related to the GDPR compliance, such as compliance checklists, forms and templates, etc. The SLICES-SC website privacy policy as well as the Right the Erasure Request form have been inspired by the templates proposed on this website and are available in the D9.1.

7. Data Security and Protection of Personal Data

This section is presenting the principles and measures to guarantee the data protection and security in the context of the SLICES-SC project. The enforcement of security and privacy is realised through well-defined data protection and security policies. The usage of the components of the research infrastructure, including tools, applications and services, must follow the policies defined in the project. Different sources of personal data have been already identified in the initial phase of the project. The open calls and the SLICES-SC Web portal are the two main sources of data. Personal data will be also be collected in other SLICES-SC activities, such as the workshop, summer schools, training, mobilities, etc. This is requiring specific policies and measures to protect the data. The datasets generated by the experiments will be also protected in the case they are containing personal data.

[52] Ethics and data protection, European Commission, 14 November 2018, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf [Last accessed 19 July 2021]
[53] European Data Protection Board: Who we are, EDPB, https://edpb.europa.eu/about-edpb/about-edpb/who-we-are_en [Last accessed 19 July 2021]
[54] ALLEA, All European Academies, https://allea.org/allea-in-brief [Last accessed 19 July 2021]

First of all, an Acceptable Use policy is intended to describe the rules and the guidelines on the manner to use the SLICES research infrastructure in conformity with the current laws, including the GDPR for the privacy. The Acceptable Use policy includes the consequences of a violation of the defined rules and guidelines; typically, a suspension of an account or the launch of a legal action are two possible consequences of a violation of the Acceptable Use possible.

The security and trust management policies aim to specify the processes needed to correctly enforce the data security and protection in the SLICES research infrastructure. Different stakeholders generating personal data should be taken into account, namely the experimenters, the research infrastructure providers and managers. Each stakeholder should be trustable in particular the testbed providers towards the researchers. For instance, the creation of certificates for each SLICES node is useful to ensure the trustworthiness of testbed provider. The data will be only exchanged if the certificates are recognised and valid. Of course, data encryption is applied during the transmissions of data between the different components of the SLICES research infrastructure, but also with the researchers. In this context, the access to the SLICES research infrastructure is implemented by a SLICES-SC Web portal. In consequence, the data used by the SLICES-SC Web portal will be encrypted to avoid any interceptions by third parties. Basically, the usage of HTTPS will be of course mandatory in the Web portal.

At the storage level, the data will be anonymized. Any personal identifiers contained in the data will be randomly generated or hashed. In case of personal data used to prove the correction implementation of the SLICES-SC project and there are no means to anonymize them, the personal data will be saved until five years after the official end of the project. Examples of such personal data are videos, interviews, presentations, etc. When personal data are considered as unnecessary for the project, they will be erased immediately.

The Web portal will also publish a privacy policy and a cookies policy as requested by the GDPR. The privacy policy will describe how the data are managed in conformity with the GDPR. The cookies policy will list all the cookies encountered in the Web portal and their purposes.

The open calls will also be handled with care for the data protection as the related data contain personal data. Typically, the exchanged data will be encrypted and also minimized. The data collection will be limited to a minimal level which ensures that the open calls are running smoothly and in full respect of the regulation, in particular the GDPR.

The data concerning the administration of the SLICES-SC project will be stored in a secure way and of course, will not be shared outside the project consortium. This concretely means that the storage of any data should be done uniquely through tools, applications or services hosted exclusively by the partners. The access to these personal data is limited to a few partners and will deleted when they will become useless. The personal data will be anonymized and encrypted as much as possible.

In summary, the principles of privacy by design and by default will be applied during all the duration of the project. The principle of data minimization will be applied during all the duration of the project, Concretely, the data collection will be limited to the minimal amount of personal data and the storage of these personal data will not last for a longer period than it is required by the SLICES-SC project. In the same way, the best practices and standards related to protection and security of personal data will strictly enforce by each component used in the context of the SLICES-SC project.

The recording of personal data processing activities is a very important topic and is under the responsibility of each data controller. A such record contains the following information:

- Name and contact details of the data controller
- Purposes of the data processing

- Description of the categories of data subjects
- Description of the categories of personal data
- Categories of recipients which will receive the data, including non-EU countries and international organisations
- Retention period for each category of data
- General description of the security measures

A data management processing form is available in the first annex of this document to get the required information from the project partners and other eventual stakeholders involved in the SLICES-SC project.

8. Ethical Aspects

This section mentions the ethical aspects linked to SLICES-SC project. These aspects are furthered detailed and explained in the D9.1 "POPD - Requirement No.1", due at M6, which provides an answer to the questions raised in the Ethics Summary Report prepared by the EC. This document presents, among others, the description of the technical and organisational measures implemented to safeguard the protection of personal data, as well as information on the informed consent procedure and the templates of the informed consent form and the information sheets.

Regarding the activities undertaken in SLICES-SC, they are following all the ethical standards, guidelines and principles. Of course, all the international, European and national laws are fully respected. The SLICES-SC project is applying the principles defined in the All European Academies (ALLEA) European Code of Conduct for Research Integrity. The goal of this code of conduct is to define all the principles of an ethic, efficient and robust research. The code of conduct takes into account external changes such as political, social and technological changes which interfere with the research environment.

The ALLEA European Code of Conduct defines four main ethical principles which are explained below.

RELIABILITY

The reliability ensures that the research methodology and the associated analysis are consistent. The credibility of the resources involved in the research process is also a part of this principle.

HONESTY

This ethical principle guarantees that the development, the review, the reporting and the communication of the research are made fairly, transparently, completely and without any biases.

RESPECT

The respect concerns the behaviour of the researchers towards the other researchers, but also with external stakeholders such as end-users, societal groups, etc. The environment and also the cultural heritage are also two important elements to take into account in this ethical principle.

ACCOUNTABILITY

The accountability principle consists to the organisation and the management of the research, including the training, the monitoring and the impacts such as the publications.

In SLICES-SC, the data collection and management will follow these ethical aspects as well as the applicable regulations. Well-defined processes and methods will ensure that the ethical principles mentioned above will be respected by each person involved in the project. The stakeholders do not only consist of the consortium partners, but also external people like the research subjects and the

participants to studies or surveys, workshops, trainings, for example. One of the goals of the ethical principles is to avoid fabrication, falsification and plagiarism of data.

In case of data collection involving any human being, the project will inform the participants about the purpose of the data collection, their access, their format, the retention period and the rights what the participants have legally, in particular with the General Data Protection Regulation (GDPR). This implied to provide each time to the participants a consent form for the data collection with all the required information. The data processing will strictly follow the GDPR.

Ethical risks will be handled by the SLICES-SC project in line with the Ethics and Data Protection report published by the European Commission. To minimize the risks, the project will not collect data related to specific types of personal data, namely:

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Genetic, biometric and health data
- Sex life or sexual orientation
- Trade union membership

No data collection and processing will be undertaken by the project for data concerning special data subjects implying specific ethical risks. Indeed, children, vulnerable people or people who have not provided an explicit consent to participate are in this category of special data subjects.

There are other risks which are mentioned below:

- Complexity of personal data processing, in particular in case of data processing at large scale.
- Techniques on data collection or processing which are invasive in terms of privacy. This includes the Artificial Intelligence (AI).
- The data transfer outside the European countries.

To manage these ethical risks, SLICES-SC will monitor the data collection and the data management to identify the risks. When an ethical risk is detected during an operation on the collected data, an analysis is immediately undertaken and the following three steps are executed:

- Overview of the processes concerning the data collection and the operations on the data.
- Identification and analysis of the ethical risks or issues.
- Explanation on how to address concretely the ethical risks or issues.

All these steps will be documented and reported to the relevant subjects, including the data subjects, the funding agencies and the data protection supervisory authorities. This process is fully aligned with the Ethics and Data Protection report defined by the European Commission.

9. Conclusion

The Data Management Plan (DMP) has described the required policies and procedures used for an efficient management and publication of the research data. The different sorts of requirements were analysed. At the same time, the governance and the management of the research data were presented. All the measures will improve the interoperability and the collaboration towards external systems or infrastructures. The different aspects linked to the metadata were also explained in this document. Finally, the DMP will serve as a reference document in the design, implementation and operation phases of the SLICES research infrastructure.

10. Annex A: Data Management Processing Form

This annex presents a data management processing form which should be completed by all the data providers, typically the researchers. It will permit to submit electronically data in compliance with the data management plan. This form should be accompanied by the completed and signed data processing agreement which is presented in the Annex B of this deliverable.

Identification/Instantiation		
Internal ID	*Generated by the resource manager, i.e., DOI*	
External IDs	*Other identifiers for the resource (e.g., links, bibliographic citations)*	
Privacy/Access level	Open	◯
	SLICES Node level	◯ *(select node from list)*
	Shared	◯ *(provide a list of one or more organizations from a list)*
	Private	◯
Version		

Content	
Creator	*Organization/Person Name*
Creator ID	*Identifier used to recognize creator, e.g., ORCID, DAI, LinkedIn*
Title	
Alternative Title	
Description	
Subject	
Keyword(s)	*Multiple-selection from a list (e.g., frequent keywords) or free text*
Language (s)	*Multiple-selection from a list*
Duration *(if applicable)*	*Selection from date/time pickers*
Location(s) *(if applicable)*	*(Ideally) multiple-selection from hierarchical location lists*
Funder(s) *(if applicable)*	*Multiple-selection from a list (e.g., known funding authorities) or free text*
Publishers(s) *(if applicable)*	

Date	
Create date	*Automatically generated from the system*
Date Submitted	*Date of submission of the resource*
Date Issued	*Date of formal issuance of the resource*
Date Accepted	*Date of acceptance of the resource*
Date Copyrighted	*Date of copyright of the resource*
Date Modified	*Date on which the resource was changed*
Availability of the resource	*Minimum date that the resource should become available*
Expiration of the resource	*Maximum date that the resource should be available*

Relationships (for each relationship)	
Relation Type	*Selected from a list*

Reference to resource	*e.g., DOI*
Description	*e.g., uses external dataset for specific purposes*

Rights Management	
Provide any right(s) that are related to the resource:	*e.g., link to terms (in list format)*
Provide any License(s) that are related to the resource	*specific licenses that apply to the resource*

Resource Characteristics (to be completed for each resource)		
What is the resource type?	Collection/Project	◯
	Single resource	◯
	Part of a collection	◯ *If yes, select collection/project from a list*
What is the measurement type (if any)?	Observational	☐
	Experimental	☐
	Simulation	☐
	Derived	☐
	Other: (please specify)	☐
What is the format of the resource	Open file format	◯
	Proprietary file format	◯ *If yes, provide additional information below*
	Proprietary file format details	*e.g., link to required software to access the resource*
Specify the size of your resource	*Automatically assessed by the system. Large files may require different upload processing.*	
Specify any special requirements for the resource	Computationally intensive	◯*Yes* ◯*No* *e.g., if yes, provide requirements*
	Storage intensive	◯*Yes* ◯*No* *e.g., if yes, provide requirements*
	Network intensive	◯*Yes* ◯*No* *e.g., if yes, provide requirements*
Provide any other characteristics for the resource	*(key, value) pairs, where key is selected from a list, e.g., (Software code, python), (Tabular data, csv)*	

Compliance/Data Quality		
Does the resource contain:	Personal data	☐
	Sensitive data	☐
	Data subject to license	☐
	Derived	☐
	Do you verify the completeness of the data?	◯*Yes* ◯*No* ◯*N/A*

Provide appropriate consents for the resource:	Do you verify the timeliness of the data?	◯*Yes* ◯*No* ◯*N/A*	
	Have you obtained appropriate consents for the use/processing of personal data contained in the resource?	◯*Yes* ◯*No* ◯*N/A* *If yes, provide description and/or link to resource*	
	If you are using external resources, have you obtained appropriate licenses/rights to use them?	◯*Yes* ◯*No* ◯*N/A* *If yes, provide description and/or link to licenses/rights*	
	Additional Consents (please specify)	◯*Yes* ◯*No* ◯*N/A*	
Data Quality Assurance	*Do you verify the quality of the data during data collection?*	◯*Yes* ◯*No* ◯*N/A* *Select yes if the instruments used for data collection provide quality assurances*	
	Are the data provided in raw format (i.e., no pre-processing has been performed on the data)?	◯*Yes* ◯*No* ◯*N/A* *Select yes if the instruments used for data collection provide quality assurances*	
	Have you used the recommended directory structure?	◯*Yes* ◯*No* *If no has been selected, please describe the structure*	
	Have you used the recommended naming conventions?	◯*Yes* ◯*No* *If no has been selected, please describe the naming convention*	
	How will the data be documented?	*(key, value) pairs, where key is selected from a list,* *e.g., (configuration, link to read.me), (jupyter notebook, link to notebook)*	
	How will versioning be managed?	No versioning, new resource will overwrite the previous	☐
		Automatic numbering/Date-Time/Version number in the structure (directory/filename)	☐
		"Track changes" feature in software	☐ *Specify software and method*
		Dedicated version control software:	☐

			Specify software and method
		Other	☐ *Specify method*
	Data Security	*Have any measures been taken to secure the data*	*(key, value) pairs, where key is selected from a list, e.g., (anonymization, details), (encryption, technique)*

Data Security (to be completed for each resource)		
What is the nature of any security requirements?	Brief summary of requirements, or a link to where they are specified.	
Have any measures been taken to ensure security?	List potential risks	
What is the measurement type (if any)?	Observational	☐
	Experimental	☐
	Simulation	☐
	Derived	☐
	Other: (please specify)	☐
What is the format of the resource	Open file format	◯
	Proprietary file format	◯ *If yes, provide additional information below*
	Proprietary file format details	*e.g., link to required software to access the resource*
Specify the size of your resource	*Automatically assessed by the system. Large files may require different upload processing.*	
Specify any special requirements for the resource	Computationally intensive	◯*Yes* ◯*No* *e.g., if yes, provide requirements*
	Storage intensive	◯*Yes* ◯*No* *e.g., if yes, provide requirements*
	Network intensive	◯*Yes* ◯*No* *e.g., if yes, provide requirements*
Provide any other characteristics for the resource	*(key, value) pairs, where key is selected from a list, e.g., (Software code, python), (Tabular data, csv)*	

11. Annex B: Data Processing Agreement

This annex presents the data processing agreement to be signed by the data provider or subject before any data processing by a legal entity of the project.

Agreement to process data in the SLICES-SC project

Name:

Home organisation:

Email address:

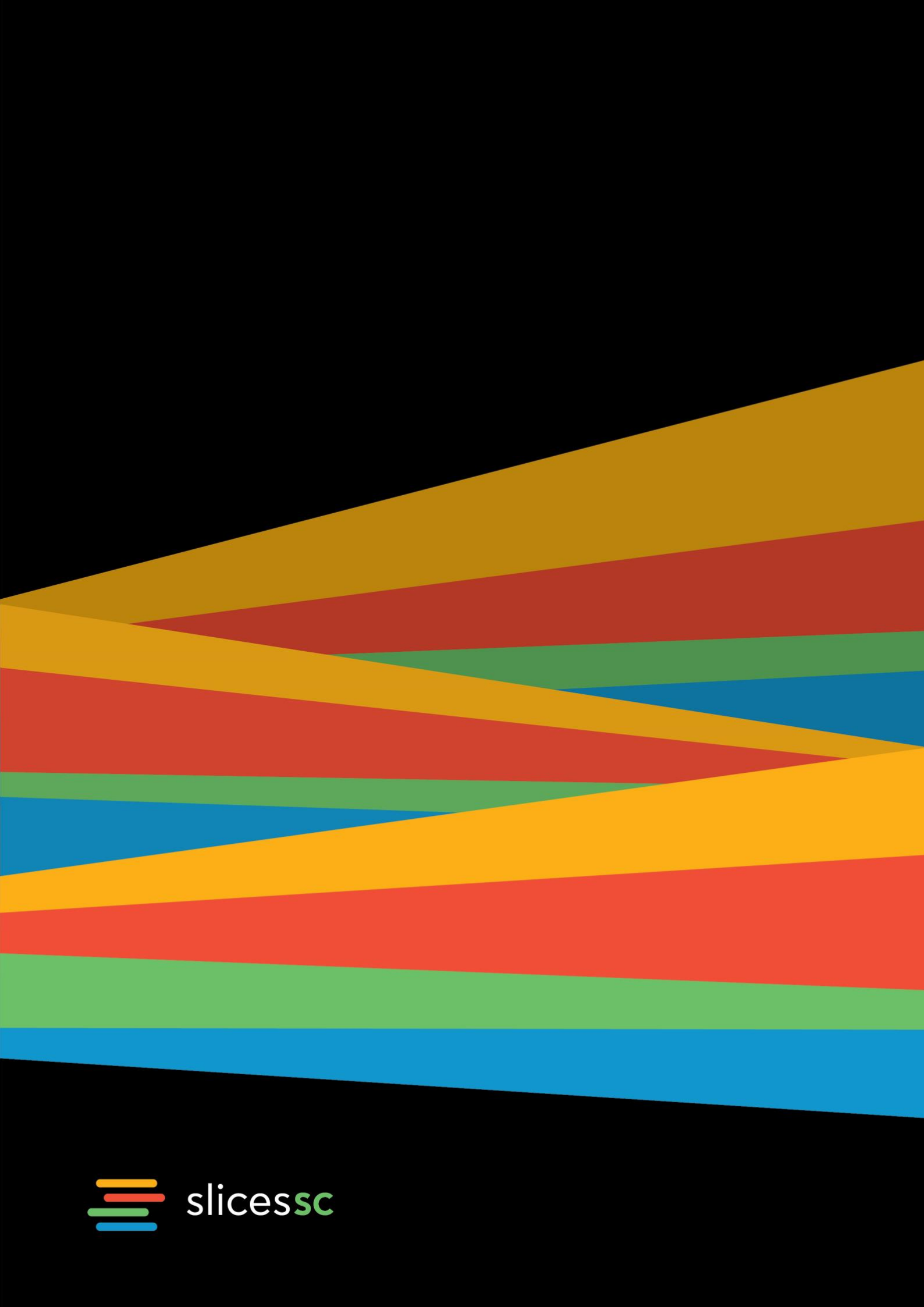
Phone number:

SLICES-SC is an EU-funded project aspiring to foster the community of researchers around the SLICES Research Infrastructure (SLICES-RI), create and strengthen necessary links with relevant industrial stakeholders for the exploitation of the infrastructure, advance existing methods for research reproducibility and experiment repeatability, and design and deploy the necessary solutions for providing SLICES-RI with an easy to access scheme for users from different disciplines. A set of detailed research activities has been designed to materialize these efforts in tools for providing transnational (remote and physical) access to the facility, as well as virtual access to the data produced over the facilities. The respective networking activities of the project aspire in fostering the community around these infrastructures, as well as open up to new disciplines and industrial stakeholders. For further information: https://slices-sc.eu/

Date:

Data to be processed by the SLICES-SC project: *internal ID and eventually external IDs from the data management processing form*

I agree ☐ / I don't agree ☐ to the process of the above-mentioned data in the SLICES-SC project, following the rules and guidelines described in the Data Management Plan and the related policies.


slicessc