slices DS

HORIZON 2020
H2020 - INFRADEV-2019-3

D4.3 Definition of the SLICES metadata profiles to support FAIR principles (initial proposal)

Acronym	SLICES-DS
Project Title	Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies – Design Study
Grand Agreement	951850
Project Duration	24 Months (01/09/2020 – 31/08/2022)
Due Date	31 August 2021 (M12)
Submission Date	31 August 2021 (M12)
Authors	Panayiotis Andreou (UCLAN), Yuri Demchenko (UvA), Émilie Mespoulhes (SU), Stavroula Maglavera (UTH), Christian Perez (INRIA), Carmen Guerrero (U3CM), Peter Van Daele (IMEC), Cédric Crettaz (MI), Loizos Christofi (eBOS).
Reviewers	All partners

Scientific Large-scale Infrastructure for Computing Communication Experimental Studies **Design Study**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951850. The information, documentation and figures available in this deliverable, is written by the SLICES-DS 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

With the advancement of technology and the plethora of electronic data being generated and available online, there is a need to ensure the longevity of such data as well as its access to the wider research and innovation community. Easy and open access to scientific data can facilitate further knowledge discovery and research transparency. Moreover, with the rapid expansion of the digital ecosystem, the use of machines to process the vast volume of available data is crucial, as humans cannot efficiently and effectively perform the relevant data processing (e.g., find, access, reuse), without additional computational support.

It is with the above in mind that the FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles were developed. The FAIR principles are intended to be used as guidelines for data producers and publishers, with regards to data management and stewardship. One important aspect that differentiates FAIR from any other related initiatives is that they move beyond traditional data processing, placing specific emphasis on automatic computation, thus considering both human-driven and machine-driven data activities. Since their publication, FAIR became widely accepted and used.

SLICES wants to fully endorse and adopt the FAIR principles, acting as a catalyst to enable and foster cutting edge research, data-driven science and scientific data-sharing. SLICES, as a research and innovation eco-system, must thus define appropriate metadata profiles to cater for access and reuse of FAIR (Findable, Accessible, Interoperable, Reusable) data and services. However, due to the problems rising from the multi/cross/inter-disciplinary nature of research SLICES is aiming to support, opting for a union of several established, domain-specific metadata schemas and vocabularies is impractical and would decrease the efficiency and effectiveness of resource discovery, access and (re)use. Instead, SLICES aims to allow its users (and interoperating platforms) to uniformly find, and access any object, such as data, services and software. To accomplish this, SLICES defines a hierarchical metadata profile scheme, where each *digital object* is first described using a select set of common metadata attributes and then according to its type, the description is extended with a set of type-specific attributes.

SLICES's hierarchical metadata profile design ensures the full support of FAIR principles at the top level of the hierarchy where carefully selected machine-readable metadata attributes allow for easy discovery of data by both humans and computer. The second and third levels improve machine and human-understandability as well as machine-actionability, allowing services to access more information and understand complex and domain-specific metadata structures to take appropriate actions. The discovered data can be then accessed using SLICES authentication and authorization mechanisms.

In what follows, we explain the initial design of the SLICES metadata profiles and show how they meet specific design objectives, such as interoperability, extensibility and adherence to the FAIR principles. These metadata profiles will be enhanced/extended throughout the lifetime of the project as the SLICES infrastructure design becomes more precise.

Table of contents

1 Introduction

Perhaps one of the most predominant definitions about metadata is "metadata is data about data", often providing information about a digital object, such as data or services, related to its contents, context, quality, structure and accessibility. As such, metadata enriches the digital object with information that improves its discovery, (re)use and management. Consequently, high-quality metadata are essential for effective interoperability, discovery, contextualisation and reuse.

There is generally a consensus that three main types of metadata exist: descriptive, structural and administrative. Descriptive metadata, such as creator and title, allow the identification and discovery of the digital object. Structural metadata elements provide information about the structure of the digital object, such as how the data elements are organized and what they are. Administrative metadata elements are used to describe the structure of a digital object for management and preservation purposes, essentially being meta-metadata.

Several metadata specifications, such as Dublin Core[1], Datacite[2], DDI[3] and ISO 19115[4], have been created to address the needs of specific research domains. Some of these standards have been most often associated with a particular file format specification used by a particular community. These standards typically include metadata elements or attributes, accompanied by their definitions and appropriate vocabularies/controlled lists for input validation. Usually, each community maintains the metadata standard and associated rules and procedures for access and discovery as well as the vocabularies used. However, problems of interoperability with different metadata standards arise when a community or system attempt to share and utilize content with external communities or systems. Well-structured metadata and appropriate crosswalks between different metadata structures play a fundamental role in achieving effective interoperability.

Interoperability problems are amplified by big data constraints and the growing demand for cross-disciplinary research that require efficient machine-actionable interoperability, which can offer seamless and precise retrieval of digital objects. It is for that reason that initiatives, such as Data Catalog Vocabulary (DCAT)[5], aim to facilitate interoperability between metadata standards and vocabularies to enable data integration and analysis. The FAIR movement is also aligned towards that goal and has given rise to new initiatives, projects and tools, such as the FAIRsFAIR project[6] and the FAIR Digital Objects Forum[7], which aim to achieve an ecosystem of FAIR infrastructures and systems.

SLICES wants to fully endorse and adopt the FAIR principles through appropriate metadata profiles that enable efficient and effective interoperability and cross-disciplinary research. This deliverable presents the initial proposal of the SLICES metadata profiles, including the rationale behind their design and structure. We first present several important metadata standards and highlight their purpose. We then overview the FAIR guiding principles and the key principles behind the design of the metadata profiles. Next, we present the SLICES metadata profiles, their specifications and show how these conform to the FAIR and the metadata principles. Finally, we present some considerations related to the transformability and management of the metadata.

[1] Dublin Core Metadata Initiative, https://dublincore.org/, [Last accessed 30 August 2021]
[2] DataCite, https://datacite.org/, [Last accessed 30 August 2021]
[3] Data Documentation Initiative, https://ddialliance.org/, [Last accessed 30 August 2021]
[4] ISO 19115-1:2014, https://www.iso.org/standard/53798.html, [Last accessed 30 August 2021]
[5] Data Catalog Vocabulary (DCAT) - Version 3, W3C Working Draft 04 May 2021 https://www.w3.org/TR/vocab-dcat-3/, [Last accessed 30 August 2021]
[6] EU project FAIRsFAIR, https://www.fairsfair.eu/, [Last accessed 30 August 2021]
[7] FAIR Digital Objects Forum, https://fairdo.org, [Last accessed 30 August 2021]

2 Metadata Standards

Research Infrastructures and other platforms, such as digital repositories, e-infrastructures and portals, utilize a plethora of dissimilar metadata standards, sometimes even unique, to facilitate data management and stewardship. These standards have different objectives related to record management and archives, providing support on multiple fronts: from aiding the archiving process to discovering, searching and preserving resources.

Several established standards are applicable for the definition of the metadata profiles of SLICES. We next provide an overview of predominant standards and highlight their objectives. A thorough guide to archival standards can be found in[8].

- **Dublin Core[9]:** Dublin Core (or the Dublin Core Metadata Element Set) defines a set of fifteen core properties, drawn form a larger set of DCMI Metadata Terms, for describing resources. Dublin Core is formally standardised as ISO 15836, ANSI/NISO Z39.85 and IETF RFC 5013. The resources described using Dublin Core include any type of digital resources (e.g., videos, images, and web pages), but also physical resources (e.g., books, CDs, and artworks). The main uses of this standard extend from simply describing resources, to combining metadata vocabularies of different standards, as well as catering for interoperability of metadata vocabularies in the linked data cloud and semantic web implementations.
- **DataCite[10]:** DataCite provides persistent identifiers (DOIs) for research data and other research outputs. Organisations can join DataCite as members to assign DOIs to research outputs and make them discoverable to the community. DataCite then develops additional services to improve the DOI management experience, making it easier for members to connect and share their DOIs with the broader research ecosystem and to assess the use of their DOIs within that ecosystem.
- **DDI[11]:** Document, Discover and Interoperate (DDI) is a free international standard that describes data produced by surveys and other observational methodologies mainly focusing on the social, behavioural, economic and health sciences. Moreover, the standard supports multiple steps in the research data lifecycle, including conceptualisation, collection, processing, distribution, discovery and archiving.
- **ISO 19115[12] and FGDC-CSDGM[13]:** These two standards provide means for describing geospatial data. ISO 19115 defines the schema for describing geographic information, as well as information for several properties of digital geographic data, including quality, spatial and temporal aspects and spatial references, amongst others. FGDC-CSDGM stands for Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata.
- **OAIS[14]:** The Open Archival Information System (OAIS) is an international standard that targets challenges related to preserving digital resources long term. More precisely, the OAIS reference model aims at providing access guarantees for archive systems, by outlining a set of functions required for both accessing and ensuring that digital resources are effectively preserved over time.

[8] A Guide to Archival Standards, https://www.archives.org.uk/about/sections-interest-groups/archives-a-technology/news-and-events.html [Last accessed 01 February 2021]
[9] Dublin Core Metadata Initiative, https://dublincore.org/ [Last accessed 01 February 2021]
[10] DataCite, https://datacite.org/ [Last accessed 01 February 2021]
[11] Data Documentation Initiative, https://ddialliance.org/, [Last accessed 30 August 2021]
[12] ISO 19115-1:2014, https://www.iso.org/standard/53798.html, [Last accessed 30 August 2021]
[13] Federal Geographic Data Committee, https://www.fgdc.gov/metadata, [Last accessed 30 August 2021]
[14] Open Archival Information System (OAIS), http://www.oais.info/ [Last accessed 01 February 2021]

- **OpenDOAR[15]:** The Directory of Open Access Repositories (OpenDOAR) is a web-based directory that offers a list of open-access academic repositories. Based in the UK, it offers options for searching resources by locale, content and other measures. OpenDOAR is currently one of the two (2) leading open access directories worldwide.
- **AGLS Metadata[16]:** The primary objective of the Australian Government Locator Service (AGLS) Metadata standard is to facilitate the search and discovery of resources, which are supplied by the Australian government. Such resources include either digital or non-digital items.
- **AGRkMS[17]:** The Australian Government Recordkeeping Metadata Standard (AGRkMS) is based on AGLS Metadata and primarily deals with national archives. AGRkMS defines the metadata properties based on the metadata standard for record keeping (ISO 23081) that agencies in the Australian government should use when describing entities involved in their business and processes regarding record management.
- **EAD[18] and ISAD(G)[19]:** The Encoded Archival Description (EAD) is a metadata schema that is used for archiving digital resources. It allows for describing the actual content of such resources, but also its overall structure. On the other hand, the General International Standard Archival Description (ISAD(G)) is more generic in that it applies to more traditional archives, not necessarily of a digital nature. EAD is designed to be compatible with ISAD(G).
- **PREMIS[20]:** The Preservation Metadata and Implementation Standard (PREMIS) is yet another metadata standard that at its core deals with matters related to the preservation of digital resources. To this end, PREMIS defines a data model for preservation along with the corresponding data dictionary. The former comprises five distinct entities, namely, the intellectual entity, digital object, agent, rights and event.
- **MINSEQE[21]:** MINSEQE provides a description about the Minimum Information of a high-throughput nucleotide SEQuencing Experiment. This minimum information is required for unambiguous interpretation and reproduction of experimental results. The adherence to MINSEQE guidelines enhances the integration of multiple experiments in a wide variety of modalities, thus the value of high-throughput research being maximised.
- **EML[22]:** The Ecological Metadata Language (EML) is a community-maintained specification that provides a thorough vocabulary and a readable XML syntax for the purpose of documenting research data in the earth and environmental sciences. EML enables researchers to preserve and openly document and share their data and findings. Some of EML's core modules allow for (i) identifying/ citing data for describing various formats of data and research methods/ protocols, (ii) describing the structure and content of data, and (iii) annotating data with semantic vocabularies.
- **FITS[23]:** Flexible Image Transport System (FITS), currently maintained by the International Astronomical Union, is utilised for the interchange of information between observatories and for archiving. FITS is a file format that facilitates the storage, transmission and manipulation of

[15] OpenDOAR - Directory of Open Access Repositories, https://v2.sherpa.ac.uk/opendoar/ [Last accessed 01 February 2021]
[16] AGLS Metadata Standard, http://www.agls.gov.au/ [Last accessed 01 February 2021]
[17] 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]
[18] EAD: Encoded Archival Description, https://www.loc.gov/ead [Last accessed 01 February 2021]
[19] 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]
[20] PREMIS Data Dictionary for Preservation Metadata, https://www.loc.gov/standards/premis/ [Last accessed 01 February 2021]
[21] Functional Genomics Data Society, http://fged.org/projects/minseqe/, [Last accessed 30 August 2021]
[22] Ecological Metadata Language (EML), https://eml.ecoinformatics.org/, [Last accessed 30 August 2021]
[23] Library of Congress: Flexible Image Transport System (FITS), https://www.loc.gov/preservation/digital/formats/fdd/fdd000317.shtml, [Last accessed 30 August 2021]

scientific images and their associated data. Since its conception, FITS was seen as a transport format for more than still images, since it was designed to enable the unambiguous transmission of 1-D spectra, 2-D images or data cubes of three or more dimensions. Tabular data, i.e., two-dimensional data tables can also be stored in FITS.

- **MIBBI[24]:** Minimum Information for Biological and Biomedical Investigations (MIBBI) is a standard defining a set of guidelines for reporting on data extracted from biosciences. MIBBI allows a community to perform easy data verification, analysation and interpretation. If the standard's guidelines are followed, the facilitation of structured databases/ public repositories and the development of data analysis tools is achieved.

Studying the aforementioned standards, we can make the following important observations:

- **No "One-size-fits-all"**: There is no "one-size-fits-all" metadata standard to address all current and future requirements. According to the requirements of the project, SLICES needs to carefully choose the appropriate suite of metadata standards and vocabularies to best describe and provide access to its resources.
- **Domain-agnostic vs. Domain-specific:** As we can observe, some metadata standards aim to provide domain-agnostic, all-purpose formats that can be used to describe any digital object but thus lack the sufficient detail and depth to facilitate machine-actionability, while others are domain-specific, which however require substantial effort to make them interoperable with platforms operating on a different metadata format.
- **Established:** besides fulfilling its descriptive purposes and functional requirements, it is best that the metadata standard is widely adopted by the research community. An established metadata standard will be mature enough to address a comprehensive set of requirements and will be accompanied with an ecosystem of tools and services to support its operation.

These observations will serve as important guiding principles in the definition of the metadata profiles described in Section 4.

3 FAIR Guiding Principles

FAIR has been coined in 2016[25] and stands for data that meet four guiding principles: findability, accessibility, interoperability, and reusability. The principles are oriented towards enabling systems and services to combat big data obstacles, such as high generation speed, diversity, complexity and volume, by enabling them take action on the data with none or minimal human intervention.

The table below provides an overview of the four guiding principles and their main components:

Table 1 – FAIR Guiding Principles

Code	Principle
Findable	
F1	(meta)data are assigned a globally unique and persistent identifier
F2	data are described with rich metadata (defined by R1 below)
F3	metadata clearly and explicitly include the identifier of the data they describe
F4	(meta)data are registered or indexed in a searchable resource

[24] FAIRsharing: Minimum Information for Biological and Biomedical Investigations, https://fairsharing.org/collection/MIBBI, [Last accessed 30 August 2021]
[25] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18, [Last accessed 30 August 2021]

Accessible	
A1	(meta)data are retrievable by their identifier using a standardized communications protocol
A1.1	the protocol is open, free and universally implementable
A1.2	the protocol allows for an authentication and authorization procedure, where necessary
A2	metadata are accessible, even when the data are no longer available
Interoperable	
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2	(meta)data use vocabularies that follow FAIR principles
I3	(meta)data include qualified references to other (meta)data
Reusable	
R1	meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1	(meta)data are released with a clear and accessible data usage license
R1.2	(meta)data are associated with detailed provenance
R1.3	(meta)data meet domain-relevant community standards

FAIR guiding principles, enrich the principles provided in the previous section and place particular emphasis on interoperability and reusability.

4 Metadata Management Design Principles

Metadata design and implementation are mission-critical, core activities contributing to the efficient and effective sharing and reuse of data within and outside SLICES.

As such, high-quality metadata are as important as data, and adequate resources should be allocated to cater for metadata operations. This means that appropriate protocols and procedures should be put in place on metadata operations (i.e., creation and operational workflows) ensuring proper management and stewardship. These protocols should also incorporate automated metadata production (e.g., metadata extraction) whenever possible and appropriate, decreasing both human effort and errors (e.g., use predefined lists from established vocabularies) thus further contributing to enhancing quality of data.

Metadata management should also be considered as an evolutionary process. Metadata attributes may be transformed, adapted or enhanced over time to cater for new or revised requirements. As such, metadata should be flexible and open, allowing for enrichment. Furthermore, metadata change management should be part of data governance's procedures utilizing appropriate mechanisms, such as versioning.

Several design principles should be considered to facilitate management and sharing of the research data in an intuitive and safe way, respecting the creators' rights, while ensuring compliance with open access and FAIR principles. Studying the above and given the analysis provided in Section 2 and 3, we define the following objectives for effective and efficient metadata management:

- **Flexible Hierarchical Metadata Model (Domain-agnostic and Domain-specific)**
  Since there is no "one-size-fits-all" metadata standard to address all current and future requirements, SLICES considers the utilization of a hierarchical model consisting of compulsory metadata attributes that are domain-agnostic and can describe any digital object (e.g., data, services) ensuring that it conforms to FAIR principles and beyond. Where appropriate, SLICES will support additional optional metadata attributes accompanied by their metadata model to further enhance the description of the object. Consequently, the model will comprehensively describe data objects and support a plethora of functions to query and retrieve them. The model should allow to be easily extended with new attributes or new types/categories as well as with new additional hierarchy levels.
- **Hybrid Metadata Production (automated machine generation and manual human entry)**

SLICES should streamline metadata production using appropriate tools that can automatically generate/extract metadata. Manual human-based metadata creation will also be supported using appropriate input validation workflows to improve quality of data.

- **Wide-ranging Interoperability**
  Metadata should include attributes to support different levels of interoperability: *semantic*, which will allow internal and external systems to discover and understand what the underlying object is; legal, which describes the restrictions in the data; and technical, which will enable systems to communicate effectively using appropriate catalogues and services. Furthermore, although data will be stored in a particular core metadata format, SLICES must support transformation of the data to other metadata formats, to ensure interoperability with other systems (e.g., aggregators). As such, the infrastructure should provide appropriate interoperability services to enable researchers/practitioners, content providers, funders and research administrators to collaborate or utilise an existing platform to do so.

- **Long-term Reusability**
  It is expected that through the lifetime of the SLICES project, the metadata model will evolve due to internal (e.g., support for new use cases) or external factors (e.g., introduction of new standards, communication protocols). The metadata model must incorporate appropriate elements to describe the metadata model itself, including specific attributes, such as the metadata version, and utilize services where systems or users can obtain this info. Vocabularies and Standards utilized by specific attributes should also be versioned. This will also enable mappings between different versions of the same metadata records further enhancing reusability and interoperability.

- **Enhanced Discovery**
  Discovery should be further supported with descriptors created manually (e.g., keywords) but also automatically. Automatic descriptors may come from automatic metadata data extraction (e.g., creation date, file size) but also using built in data analysis functions (e.g., term document frequency).

- **Metadata Quality Assurance**
  Metadata quality is important for ensuring reusability and interoperability with other infrastructures/platforms and applications that may want to consume data from SLICES. A Metadata Quality Assurance Framework should be put in place as part of the Data Governance Group with appropriate metrics to assess the quality of metadata, its FAIRness and other objectives.

- **Metadata Governance**
  The members of the Data Governance Group should have the knowledge and authority to make decisions on how metadata are maintained, what format is being utilised and how changes are authorised and audited.

Using the above principles, we provide the definitions for the metadata profiles in the next section.

5 Metadata Profiles

SLICES proposes a flexible hierarchical metadata model consisting of three levels. The first level consists of compulsory domain-agnostic information that can describe any digital object (e.g., data, services), ensuring that it conforms to FAIR principles and beyond. Additionally, to enhance machine actionability for specific commonly used types of digital objects, such as data, services, and software, SLICES employs a second level of compulsory metadata attributes that are type specific. Finally, the third level incorporates optional domain-specific attributes to further enhance interoperability for specific communities.

A high-level representation of the proposed model structure is illustrated in Figure 1.

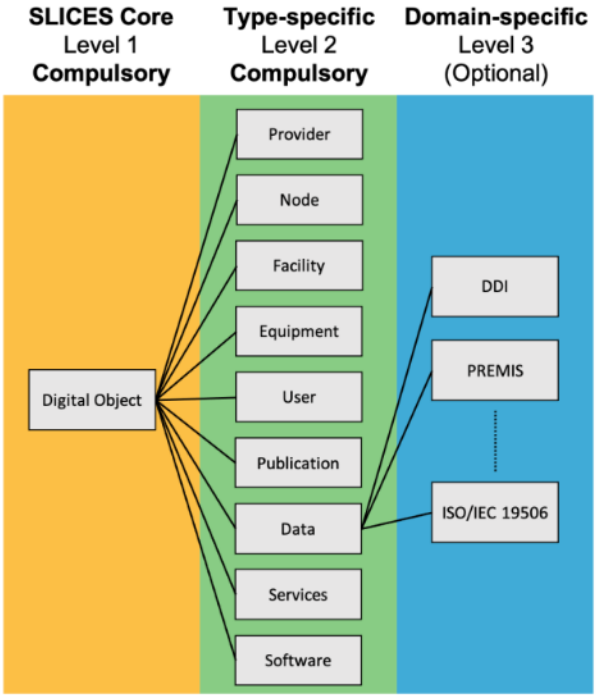

Figure 1 – SLICES Metadata Structure

Level 1 includes metadata related to the **S**LICES core **F**AIR **D**igital **O**bject, coined S-FDO. It includes basic information, such as identification, description and its resource type. Management information is also included such as version and metadata profile used. Using its type (e.g., data, service), Level 2 provides type-specific information. For example, a dataset may have start and end date, a facility may have an address. Finally, level 3, provides further domain-specific information that may be required by a specific community. For example, specific media details can be additionally provided using the PREMIS metadata format.

In the following sections, we provide more details about each level, including their attributes, definitions, types, participation and cardinality constraints.

We use the following notations for participation:

R required: a value must be provided for the attribute

E required if exists: a value must be provided if it exists for the attribute

O optional: a value may be provided for the attribute

Note that an optional attribute will appear in the metadata profile but may have an empty value. In this way we ensure technical interoperability and decrease machine-based validation efforts.

We use the following notations for cardinality:

1 single occurrence

? optional single occurrence

\+ 1 to many occurrences

* 0 to many occurrences

5.1 Metadata Catalogue

The SLICES metadata catalogue presented in this initial release is the result of the analysis of the crosswalks between some of the most predominant metadata standards such as RDA metadata IG, EOSC minimum metadata set, Dublin Core, Datacite 4.3 and DCAT 2.0.

The catalogue contains the following categories of metadata fields:

- PI - Primary Information: Important information that is used to identify and describe the resource;
- MA - Management Information: Information related to the management of the S-FDO including the versioning and metadata profile used to interpret it;
- AC - Access Information: The manner and mode in which the S-FDO can be accessed;
- CL - Classification Information: Classification of the S-FDO in specific scientific domains;
- PB - Publication Information: Specific publication information such as the date of submission and acceptance;
- FN - Financial Information: Information related to the financial aspects of access for the S-FDO;
- SU - Support Information: Support information, such as helpdesk URLs, manuals, terms of use, etc.;
- US - User Information: Information related to registered users. The information access mode is by default private and only identifiers are used;
- AT - Attribution Information: Information related to the funding instruments (e.g., body, project) of the S-FDO;
- MA - Maturity Information: Fields utilized specifically with services and software and describe important information such as standards, opensource technologies and certifications;
- ST - SpatioTemporal Information: Location and Time descriptors. Locations use Geographic coordinates when applicable;
- GE - Geographic Information: Specific geographic coordinate attributes i.e., latitude, longitude and altitude;
- LA - Language Information : Language descriptors;
- DA - Dataset Information: Specific dataset descriptors, such as format, size;
- RT - Rights and terms of access: Legal information related to the rights of access, such as licences, copyrights, etc.
- SW – Software related attributes, such as repository, documentation and programming language used to develop the software.

The table 2 provides the attributes for each category and their respective Codes, Definitions and Types.

Table 2 – SLICES Metadata Catalogue

Category	Code	Attribute Name	Definition	Type
Primary Information	SLICES.PI.01	Identifier	The Identifier is a unique string that identifies a resource. Example formal identification systems include the Uniform Resource Identifier (URI), the Uniform Resource Locator (URL), the Digital Object Identifier (DOI), and the URN. Also, this can be a direct URL, or a persistent identifier, like PURL, HANDLE or other international resolution mechanisms.	String (max)
	SLICES.PI.02	InternalIdentifier	Internal Code created within SLICES which can be resolved by the SLICES portal.	String (code)
	SLICES.PI.03	AlternateIdentifier	An alternate identifier to the resource	String (max)
	SLICES.PI.04	Creator	The creator(s) of the digital object in priority order. May be a corporate/institutional or personal name.	String (max)
	SLICES.PI.05	CreatorIdentifier	Identifier(s) of the creator(s) of the digital object.	String (max)
	SLICES.PI.06	Name	A name or title by which a digital object is known.	String (200)
	SLICES.PI.07	Description	A textual description of the content. It should include information about what the digital object does and the functionality it provides.	String (max)
	SLICES.PI.08	ResourceType	The type of the digital object	List (ResourceTypes)
	SLICES.PI.09	Subject	The subject represented using key phrases, or classification codes.	String (max)
	SLICES.PI.10	Keywords	An index term, subject term, subject heading, or descriptor, in information retrieval, is a term that captures the essence of the topic of a document.	String (20)
	SLICES.PI.11	DateTimeCreated	A timestamp generated by the system.	timestamp
	SLICES.PI.12	Webpage	Webpage with information about the Resource usually hosted and maintained by the Provider.	URL
	SLICES.PI.13	Logo	Link to the logo/visual identity of the Resource. The logo will be visible at the Portal. If there is no specific logo for the object the logo of the Provider may be used.	URL
	SLICES.PR.14	Publisher	The name of the entity that holds, archives, publishes prints, distributes, releases, issues, or produces the resource. This property will be used to formulate the citation, so consider the prominence of the role.	string (max)
	SLICES.PR.15	PublicationYear	The year when the digital object was or will be made publicly available. In the case of resources such as software or dynamic data where there may be multiple releases in one year, include the dateType vocabulary.	int
Management Information	SLICES.MA.01	Version	Version of the resource.	String (20)

	SLICES.MA.02	MetadataProfile	The metadata profile used to describe this resource.	Identifier
	SLICES.MA.03	Contributor	The institution or person responsible for collecting, managing, distributing, or otherwise contributing to the development of the digital object. To supply multiple contributors, repeat this property. Examples of contributors are chair, contact group, contact person, contact for access. data collector, data curator, data manager, distributor, editor, funder. hosting institution, maintainer, producer, project leader, project manager, project member, provider, publisher, reader, registration agency, registration authority, related person, researcher, research group, review assistant, reviewer, reviewer-external, rights holder, sponsor, stats-reviewer, supervisor, translator, workpackage leader.	String (max)
	SLICES.MA.04	ContributorIdentifier	Identifier(s) of the contributor(s) of the digital object.	String (max)
Access Information	SLICES.AC.01	Access Type	The way a user can access the Resource (Remote, Physical, Virtual, etc.)	List (AccessTypes)
	SLICES.AC.02	Access Mode	Eligibility/criteria for granting access to users (excellence-based, free-conditionally, free etc.)	List (AccessModes)
Classification Information	SLICS.CL.01	ScientificDomains	The branch of science, scientific discipline that is related to the digital object.	List (ScientificDomains)
	SLICS.CL.02	ScientificSubdomains	The subbranch of science, scientific subdiscipline that is related to the digital object.	List (ScientificSubDomains)
Publication Information	SLICES.PB.01	PublicationDateSubmitted	The date the object was submitted.	date
	SLICES.PB.02	PublicationDateModified	The date(s) the object was modified.	date
	SLICES.PB.03	PublicationDateIssued	The date(s) the object was issued.	date
	SLICES.PB.04	PublicationDateAccepted	The date(s) the object was accepted.	date
	SLICES.PB.05	PublicationDateCopyrighted	The date(s) the object was copyrighted.	date
	SLICES.PB.06	PublicationType	The type of publication (e.g., book, article, conference paper)	List (PublicationTypes)
	SLICES.PB.07	JournalTitle	The title of the journal article series.	String (200)
	SLICES.PB.08	JournalVolume	The volume of the journal article series.	String (20)
	SLICES.PB.09	JournalIssue	The issue of the journal article series.	String (20)
	SLICES.PB.10	JournalPages	The pages of the journal article series.	String (20)
	SLICES.PB.11	JournalDates	The date of the journal article series.	String (100)
	SLICES.PB.12	ConferenceTitle	The title of the conference venue.	String (200)
	SLICES.PB.13	ConferenceAcronym	The acronym of the conference.	String (20)
	SLICES.PB.14	ConferenceDates	The dates of the conference	String (100)
	SLICES.PB.15	ConferenceStartDate	The start date of the conference.	date
	SLICES.PB.16	ConferenceEndDate	The end date of the conference.	date
	SLICES.PB.17	ConferenceURL	The URL of the conference.	URL
	SLICES.PB.18	ConferenceSession	The session of the conference.	String (max)
	SLICES.PB.19	ConferenceSessionPart	The part of the session of the conference.	String (20)
	SLICES.PB.20	PublicationISBN	ISBN of a book or report.	String (20)

Financial Information	SLICES.FN.01	PaymentModel	Webpage with the supported payment models and restrictions that apply to each of them.	URL
	SLICES.FN.02	Pricing	Webpage with the information on the price scheme for this Object in case the customer is charged for.	URL
Support Information	SLICES.SU.01	HelpdeskEmail	The email to ask more information from the Provider about this Resource.	Email
	SLICES.SU.02	SecurityContactEmail	The email to contact the Provider for critical security issues about this Resource.	Email
	SLICES.SU.03	ResourceOrganisation	The name (or abbreviation) of the organisation that manages or delivers the resource, or that coordinates object delivery in a federated scenario.	Identifier
	SLICES.SU.04	ResourceProviders	The name(s) (or abbreviation(s)) of organization/infrastructures/platforms that manage or deliver the object in a federated scenario.	Identifier
	SLICES.SU.05	Helpdesk Page	The URL to a webpage to ask more information from the Provider about this Resource.	URL
	SLICES.SU.06	User Manual	Link to the Resource user manual and documentation.	URL
	SLICES.SU.07	Terms Of Use	Webpage describing the rules, Resource conditions and usage policy which one must agree to abide by in order to use the Resource.	URL
	SLICES.SU.08	Privacy Policy	Link to the privacy policy applicable to the Resource.	URL
	SLICES.SU.09	Access Policy	Information about the access policies that apply.	URL
	SLICES.SU.10	Service Level	Webpage with the information about the levels of performance that a Provider is expected to deliver.	URL
	SLICES.SU.11	Training Information	Webpage to training information on the Resource.	URL
	SLICES.SU.12	Status Monitoring	Webpage with monitoring information about this Resource.	URL
	SLICES.SU.13	Maintenance	Webpage with information about planned maintenance windows for this Resource.	URL
User Information	SLICES.US.01	PrimaryContact	Primary contact for the digital object.	Identifier (User)
	SLICES.US.02	PublicContact	Public contact for the digital object	Identifier (User)
	SLICES.US.03	First Name	First Name of the Resource's main contact person/manager.	String (max 100)
	SLICES.US.04	Last Name	Last Name of the Resource's main contact person/manager.	String (max 100)
	SLICES.US.05	Email	Email of the Resource's main contact person/manager.	Email
	SLICES.US.06	Phone	Telephone of the Resource's main contact person/manager.	String (max 20)
	SLICES.US.07	Position	Position of the Resource's main contact person/manager.	String (max 100)
	SLICES.US.08	Organisation	The organisation to which the contact is affiliated.	String (max 100)
Attribution Information	SLICES.AT.01	Funding Body	Name of the funding body that supported the development and/or operation of the Resource.	List (Funding Bodies)

	SLICES.AT.02	Funding Program	Name of the funding program that supported the development and/or operation of the Resource.	List (Funding Program)
	SLICES.AT.03	Grant/Project Name	Name of the project that supported the development and/or operation of the Resource.	String (max 100)
Maturity Information	SLICES.MA.01	TechnologyReadinessLevel	The Technology Readiness Level of the Resource (to be further updated in the context of the EOSC).	List (TRLs)
	SLICES.MA.02	LifeCycleStatus	Phase of the Resource life-cycle.	List (LifeCycles)
	SLICES.MA.03	Certifications	List of certifications obtained for the Resource (including the certification body).	String (max 100)
	SLICES.MA.04	Standards	List of standards supported by the Resource.	String (max 100)
	SLICES.MA.05	OpenSourceTechnologies	List of open-source technologies supported by the Resource.	String (max 100)
	SLICES.MA.06	LastUpdate	Date of the latest update of the Resource.	Date
	SLICES.MA.07	ChangeLog	Summary of the Resource features updated from the previous version.	String (max 1000)
Spatio-Temporal Information	SLICES.ST.01	Address	The address of the resource.	String (max 1000)
	SLICES.ST.02	DateTimeStart	Object's start date (e.g., start date of data recording).	DateTime
	SLICES.ST.03	DateTimeEnd	Object's end date (e.g., end date of data recording).	DateTime
	SLICES.ST.04	Location	Combination of Latitude, Longitude and Altitude Coordinate(s).	Geography
Geographic Information	SLICES.GE.01	Latitude	Latitude Geography Coordinate.	Geography
	SLICES.GE.02	Longitude	Longitude Geography Coordinate	Georgraphy
	SLICES.GE.03	Altitude	Altitude Geography Coordinate.	Georgraphy
Language Information	SLICES.LA.01	PrimaryLanguage	The primary language of the digital object.	List (Languages)
	SLICES.LA.02	OtherLanguages	Other languages provided for the digital object.	List (Languages)
Dataset Information	SLICES.DA.01	Size	The size of the digital object in bytes.	int
	SLICES.DA.02	Duration	Duration in ISO 8601 format.	String (100)
	SLICES.DA.02	Format	The format of the digital object.	List (Formats)
	SLICES.DA.03	Medium	The material or physical carrier of the digital object.	List (Mediums)
	SLICES.DA.04	Compression format	The compression format of the distribution in which the digital object is contained in a compressed form, e.g., to reduce the size of the downloadable file (source DCAT).	List (CompressionFormats)
	SLICES.DA.05	File	Metadata about file (name, title, description, format, mimetype, type of file, persistentID, download URL, format, size, compression format, checksum, checksum algorithm, bitrate, duration encodedinf format, upload date, distribution...).	String (max)
	SLICES.DA.06	DataStandard	Information about data schema or standard of the digital object.	String (100)

Software Information	SLICES.SW.01	RepositoryURI	Repository URI for accessing the software. Can also be a URL of the web location used to download the software.	String (max)
	SLICES.SW.02	SoftwareType	Type of the software.	List (SoftwareTypes)
	SLICES.SW.03	DocumentationURI	Software documentation URI. May also include setup instructions.	String (max)
	SLICES.SW.04	ProgrammingLanguage	The programming language used to develop the software.	List (Programming Languages)
	SLICES.SW.05	ToolService	Tool/Service that can execute the software.	String (max)
Rights and terms of access	SLICES.RT.01	Rights	Rights related with the digital object.	String (max)
	SLICES.RT.02	RightsURI	A URI for the rights related with the digital object.	String (max)
	SLICES.RT.03	AccessRights	Access rights related with the digital object.	String (max)
	SLICES.RT.04	License	License related with the digital object.	String (max)
	SLICES.RT.05	LicenseURI	A URI for the licence related with the digital object.	String (max)
	SLICES.RT.06	CopyrightsHolder	Copyrights holder related with the digital object.	String (max)
	SLICES.RT.07	ConfidentialityDeclaration	Confidentiality declaration for the digital object.	String (max)
	SLICES.RT.08	SpecialPermissions	Special permissions related with the digital object.	String (max)
	SLICES.RT.09	Restrictions	Restrictions related with the digital object.	String (max)
	SLICES.RT.10	CitationRequirements	Citation requirements related with the digital object.	String (max)
	SLICES.RT.11	Conditions	Conditions related with the digital object.	String (max)
	SLICES.RT.12	Disclaimer	Disclaimer for the digital object.	String (max)

5.2 Core Digital Object Metadata (Level 1)

The SLICES core FAIR Digital Object (S-FDO) has been created by analysing predominant metadata catalogs such as RDA metadata IG, EOSC-EDMI, Dublin Core, Datacite 4.3 and DCAT 2.0. The objective is to describe a digital object adequately to facilitate FAIR principles and ensure a high degree of interoperability. The table 3 provides the definitions of the attributes.

The identifier type includes a reference to a persistent identifier such as DOI, URN and URLs. Links are tuples in the form of (identifier, relationship type, resource type), denoting the relationship to a particular resource and how it used in the context of this resource, e.g., publication A cites publication B.

Table 3 – SLICES core FAIR Digital Object Metadata

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier	Comments
SLICES.PI.01	Identifier	String (max)	R	1	PU	Default: Assigned by SLICES
SLICES.PI.02	InternalIdentifier	String (code)	R	1	PU	Default: Assigned by SLICES
SLICES.PI.03	AlternateIdentifier	String (max)	E	1	PU	
SLICES.PI.04	Creator	String (max)	R	+	PR	
SLICES.PI.05	CreatorIdentifier	String (max)	E	+	PR	
SLICES.PI.06	Name	String (200)	R	1	PU	
SLICES.PI.07	Description	String (max)	E	?	PU	
SLICES.PI.08	ResourceType	List (ResourceTypes)	R	1	PU	
SLICES.PI.09	Subject	String (max)	O	?	PU	Recommended best practice is to use a controlled vocabulary.
SLICES.PI.10	Keywords	String (20)	R	+	PU	
SLICES.PI.11	DateTimeCreated	DateTime	R	1	PU	Default: Assigned by SLICES
SLICES.MA.01	Version	String (20)	R	1	PU	Default: N/A
SLICES.MA.02	MetadataProfile	Identifier	R	1	PU	
SLICES.MA.03	Contributor	String (max)	E	*	PR	
SLICES.MA.04	ContributorIdentifier	String (max)	E	*	PR	
SLICES.AC.01	AccessType	List (AccessTypes)	R	1	PU	Default: Remote
SLICES.AC.02	AccessMode	List (AccessModes)	R	1	PU	Default: free
SLICES.LN.01	RequiredObjects	Link	O	*	PU	
SLICES.LN.02	RelatedObjects	Link	O	*	PU	
SLICES.LA.01	PrimaryLanguage	List (Languages)	O	*	PU	
SLICES.LA.02	OtherLanguages	List (Languages)	O	*	PU	
SLICES.US.01	Contact	Identifier (User)	O	1	PR	
SLICES.US.02	PublicContact	Identifier (User)	O	1	PU	User
SLICES.RT.01	Rights	String (max)	R	1	PU	
SLICES.RT.02	RightsURI	String (max)	O	1	PU	
SLICES.RT.03	AccessRights	String (max)	R	1	PU	
SLICES.RT.04	License	String (max)	R	1	PU	
SLICES.RT.05	LicenseURI	String (max)	O	1	PU	
SLICES.RT.06	CopyrightsHolder	String (max)	O	1	PU	
SLICES.RT.07	ConfidentialityDeclaration	String (max)	O	1	PU	
SLICES.RT.08	SpecialPermissions	String (max)	O	1	PU	
SLICES.RT.09	Restrictions	String (max)	O	1	PU	
SLICES.RT.10	CitationRequirements	String (max)	O	1	PU	
SLICES.RT.11	Conditions	String (max)	O	1	PU	
SLICES.RT.12	Disclaimer	String (max)	O	1	PU	

5.3 Type-specific Digital Object Metadata (Level 2)

This section provides the metadata profiles for specific types of digital objects. The final list of supported types will be presented in the revised version of the deliverable at the end of the project.

5.3.1 Data

This profile concerns datasets and contains specific metadata attributes, such as size, format, etc.

Table 4 – Data Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.CL.01	ScientificDomains	List (ScientificDomains)	R	+	PU
SLICES.CL.02	ScientificSubdomains	List (ScientificSubDomains)	R	+	PU
SLICES.FN.01	PaymentModel	URL	O	1	PU
SLICES.FN.02	Pricing	URL	O	1	PU
SLICES.ST.01	Address	String (max 1000)	O	1	PU
SLICES.ST.02	DateTimeStart	DateTime	O	1	PU
SLICES.ST.03	DateTimeEnd	DateTime	O	1	PU
SLICES.ST.04	Location	Geography	O	*	PU
SLICES.DA.01	Size	int	R	1	PU
SLICES.DA.02	Duration	String (100)	E	1	PU
SLICES.DA.02	Format	List (Formats)	R	1	PU
SLICES.DA.03	Medium	List (Mediums)	E	1	PU
SLICES.DA.04	Compression format	List (CompressionFormats)	E	1	PU
SLICES.DA.05	File	String (max)	R	1	PU
SLICES.DA.06	DataStandard	String (100)	E	1	PU

5.3.2 Services

This profile concerns services and contains specific metadata attributes related to service support, technology maturity and financial information.

Table 5 – Services Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.PI.12	Webpage	URL	O	1	PU
SLICES.PI.13	Logo	URL	O	1	PU
SLICES.SU.01	Helpdesk Email	Email	R	1	PU
SLICES.SU.02	Security Contact Email	Email	R	1	PR
SLICES.SU.03	Resource Organisation	Identifier	R	1	PU
SLICES.SU.04	Resource Providers	Identifier	O	*	PU
SLICES.SU.05	Helpdesk Page	URL	R	1	PU
SLICES.SU.06	User Manual	URL	R	1	PU
SLICES.SU.07	Terms Of Use	URL	R	1	PU
SLICES.SU.08	Privacy Policy	URL	R	1	PU
SLICES.SU.09	Access Policy	URL	R	1	PU
SLICES.SU.10	Service Level	URL	R	1	PU
SLICES.SU.11	Training Information	URL	R	1	PU
SLICES.SU.12	Status Monitoring	URL	R	1	PU
SLICES.SU.13	Maintenance	URL	R	1	PU
SLICS.CL.01	ScientificDomains	List (ScientificDomains)	R	+	PU
SLICS.CL.02	ScientificSubdomains	List (ScientificSubDomains)	R	+	PU
SLICES.FN.01	PaymentModel	URL	O	1	PU
SLICES.FN.02	Pricing	URL	O	1	PU

SLICES.MA.01	Technology Readiness Level	List (TRLs)	R	1	PU
SLICES.MA.02	Life Cycle Status	List (LifeCycles)	O	1	PU
SLICES.MA.03	Certifications	String (max 100)	O	*	PU
SLICES.MA.04	Standards	String (max 100)	O	*	PU
SLICES.MA.05	Open-Source Technologies	String (max 100)	O	*	PU
SLICES.MA.06	Last Update	Date	O	1	PU
SLICES.MA.07	Change Log	String (max 1000)	O	1	PU

5.3.3 Software

This profile concerns software and contains specific metadata attributes related to service support and financial information.

Table 6 – Software Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.PI.12	Webpage	URL	O	1	PU
SLICES.PI.13	Logo	URL	O	1	PU
SLICES.SU.01	Helpdesk Email	Email	R	1	PU
SLICES.SU.02	Security Contact Email	Email	R	1	PR
SLICES.SU.03	Resource Organisation	Identifier	R	1	PU
SLICES.SU.04	Resource Providers	Identifier	O	*	PU
SLICES.SU.05	Helpdesk Page	URL	R	1	PU
SLICES.SU.06	User Manual	URL	R	1	PU
SLICES.SU.07	Terms Of Use	URL	R	1	PU
SLICES.SU.08	Privacy Policy	URL	R	1	PU
SLICES.SU.09	Access Policy	URL	R	1	PU
SLICES.SU.10	Service Level	URL	R	1	PU
SLICES.SU.11	Training Information	URL	R	1	PU
SLICES.SU.12	Status Monitoring	URL	R	1	PU
SLICES.SU.13	Maintenance	URL	R	1	PU
SLICS.CL.01	ScientificDomains	List (ScientificDomains)	R	+	PU
SLICS.CL.02	ScientificSubdomains	List (ScientificSubDomains)	R	+	PU
SLICES.FN.01	PaymentModel	URL	O	1	PU
SLICES.FN.02	Pricing	URL	O	1	PU
SLICES.MA.03	Certifications	String (max 100)	O	*	PU
SLICES.MA.04	Standards	String (max 100)	O	*	PU
SLICES.MA.05	Open-Source Technologies	String (max 100)	O	*	PU
SLICES.SW.01	RepositoryURI	String (max)	R	+	PU
SLICES.SW.02	SoftwareType	List (SoftwareTypes)	R	1	PU
SLICES.SW.03	DocumentationURI	String (max)	R	+	PU
SLICES.SW.04	ProgrammingLanguage	List (Programming Languages)	R	+	PU
SLICES.SW.05	ToolService	String (max)	O	+	PU

5.3.4 Publication

This profile concerns publications and contains specific metadata attributes, such as submission date, acceptance date, publisher, etc.

Table 7 – Publication Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.PR.14	Publisher	string (max)	R	1	PU
SLICES.PR.15	PublicationYear	int	R	1	PU
SLICES.PB.01	PublicationDateSubmitted	date	R	1	PU
SLICES.PB.02	PublicationDateModified	date	O	*	PU
SLICES.PB.03	PublicationDateIssued	date	O	1	PU
SLICES.PB.04	PublicationDateAccepted	date	R	1	PU
SLICES.PB.05	PublicationDateCopyrighted	date	R	1	PU
SLICES.PB.06	PublicationType	List (PublicationTypes)	R	1	PU
SLICES.PB.07	JournalTitle	String (200)	O	1	PU
SLICES.PB.08	JournalVolume	String (20)	O	1	PU
SLICES.PB.09	JournalIssue	String (20)	O	1	PU
SLICES.PB.10	JournalPages	String (20)	O	1	PU
SLICES.PB.11	JournalDates	String (100)	O	1	PU
SLICES.PB.12	ConferenceTitle	String (200)	O	1	PU
SLICES.PB.13	ConferenceAcronym	String (20)	O	1	PU
SLICES.PB.14	ConferenceDates	String (100)	O	1	PU
SLICES.PB.15	ConferenceStartDate	date	O	1	PU
SLICES.PB.16	ConferenceEndDate	date	O	1	PU
SLICES.PB.17	ConferenceURL	URL	O	1	PU
SLICES.PB.18	ConferenceSession	String (max)	O	1	PU
SLICES.PB.19	ConferenceSessionPart	String (20)	O	1	PU
SLICES.PB.20	PublicationISBN	String (20)	O	1	PU
SLICS.CL.01	ScientificDomains	List (ScientificDomains)	R	+	PU
SLICS.CL.02	ScientificSubdomains	List (ScientificSubDomains)	R	+	PU

5.3.5 Provider

This profile concerns the SLICES infrastructure provider profile and will be used for marketing purposes but also for registering SLICES in specific platforms and systems, such as EOSC.

Table 8 – Provider Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.PI.12	Webpage	URL	O	1	PU
SLICES.PI.13	Logo	URL	O	1	PU
SLICES.SU.01	Helpdesk Email	Email	Mandatory	1	PU
SLICES.SU.02	Security Contact Email	Email	Mandatory	1	PR
SLICS.CL.01	ScientificDomains	List (ScientificDomains)	R	+	PU
SLICS.CL.02	ScientificSubdomains	List (ScientificSubDomains)	R	+	PU

5.3.6 SLICES Node and SLICES Facility

This profile concerns additional information for SLICES Nodes and Facilities.

Table 9 – SLICES Node/Facility Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.PI.12	Webpage	URL	O	1	PU
SLICES.PI.13	Logo	URL	O	1	PU
SLICES.ST.04	Location	Geography	O	*	PU

5.3.7 User

This profile concerns users operating in SLICES. Other metadata profiles will use an Identifier value to a user when requiring to specify user information such as contacts.

Table 10: User Metadata Profile

Code	Attribute Name	Type	Participation	Cardinality	Access Modifier
SLICES.US.03	First Name	String (max 20)	R	1	PR
SLICES.US.04	Last Name	String (max 20)	R	1	PR
SLICES.US.05	Email	Email	R	1	PR
SLICES.US.06	Phone	String (max 20)	R	1	PR
SLICES.US.07	Position	String (max 20)	R	1	PR
SLICES.US.08	Organisation	String (max 50)	R	1	PR

6 FAIR Compliance

The following table demonstrates how the proposed SLICES metadata profiles adhere to the FAIR principles.

Table 11 – Compliance of SLICES with FAIR principles

Code	Principle	SLICES adherence to principle
Findable		
F1	(meta)data are assigned a globally unique and persistent identifier	The data will be assigned a unique identifier at the time of upload. Example identifiers are the Uniform Resource Identifier (URI), the Uniform Resource Locator (URL), the Digital Object Identifier (DOI), and the URN. Also, this can be a direct URL, or a persistent identifier, like PURL, HANDLE or other international resolution mechanisms. Internal Codes will also be created within SLICES and will be resolved by the SLICES portal.
F2	data are described with rich metadata (defined by R1 below)	Metadata will be provided according to the SLICES metadata profiles. The Core metadata profile and the type specific metadata descriptors will be compulsory. Furthermore, automatic metadata generation, e.g. using Machine Learning techniques, can be used to provide automatically metadata descriptors for increased reusability and scalability for interacting with other infrastructures and external digital libraries and their collections.
F3	metadata clearly and explicitly include the identifier of the data they describe	The data will be assigned a persistent identifier (e.g., DOI) at the time of upload. The metadata will include the identifier in a dedicated attribute.
F4	(meta)data are registered or indexed in a searchable resource	A dedicated resource discovery component will allow for searching data resources through the metadata and keywords (assigned by the user or automatically generated by the platform). The Discovery component will support simple and advanced queries that allow users to use free-text search, combine filters and drill down to the individual components of each metadata property.
Accessible		
A1	(meta)data are retrievable by their identifier using a standardized communications protocol	Metadata are retrievable using at least one metadata harvesting protocol (e.g., Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)) to streamline data dissemination. Support for REST APIs with XML, JSON and YAML capabilities for metadata (while being extendible to support other formats in the future).
A1.1	the protocol is open, free and universally implementable	OAI-PMH and REST are open, free and universally implementable, as well as supported by the vast majority of applications.
A1.2	the protocol allows for an authentication and authorization procedure, where necessary	Authentication and authorization procedures will be implemented for data that are not openly available. Metadata will be open and will not require authentication and authorization procedures.
A2	metadata are accessible, even when the data are no longer available	Metadata will be retained for the duration of the project and the lifetime of the infrastructure. In case a research data management platform is used, the lifetime will relate to the lifetime of the host repository. Metadata will be stored in a dedicated data store.
Interoperable		
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	Metadata will be provided in a consistent format with appropriate properties based on the metadata profiles presented in Section 5. The metadata and meta-metadata will be accessible through the REST APIs, which will export XML, JSON and YAML) and also will have the

		ability to translate the metadata format to others based on a mapping engine.
I2	(meta)data use vocabularies that follow FAIR principles	The following vocabularies/standards will be utilized for specific attributes contained in the metadata profiles. •ISO3166: The set of codes for the representation of names of countries. •ISO639-3: The three-letter alphabetic codes for the representation of names of languages. •ISO 8601-1: Representations for information interchange for date and time. We also anticipate more standards to be adopted when the final SLICES design is available.
I3	(meta)data include qualified references to other (meta)data	This information will be provided by the defined Links metadata category, which includes RequiredObjects and RelatedObjects, using appropriate identifiers.
Reusable		
R1	meta(data) are richly described with a plurality of accurate and relevant attributes	Metadata will be provided in a consistent format with appropriate properties based on the metadata profiles presented in Section 5. The presented hierarchical model allows for domain-agnostic, type-specific and domain specific metadata.
R1.1	(meta)data are released with a clear and accessible data usage license	Rights and Terms of Access is a category of metadata which is part of the SLICES Core metadata profile. It includes rights, licenses, access rights and other properties to describe access and usage. Data downloaded by other users will be subject to the specified license.
R1.2	(meta)data are associated with detailed provenance	Provenance is part of the SLICES Core metadata properties and requires the user to provide a statement of any changes in ownership and custody of the resource since its creation that are significant for its authenticity, integrity and interpretation.
R1.3	(meta)data meet domain-relevant community standards	SLICES Core is based on predominant metadata standards, such as Dublin Core (ranks 1st in re3data), RDA and DataCite. Metadata will be provided in a consistent format with appropriate properties based on the metadata profiles presented in Section5.

7 Metadata Transformation

Metadata transformation should be provided by a service that provides mappings between attributes in different metadata standards, enabling translation from one format to another when exchanging data. The service should enable the SLICES metadata format to be transformed, initially to a select set of predominant metadata formats, such as Dublin Core, DataCite, DDI and ISO19115, to further enhance interoperability. Additionally, it should be able to transform the metadata attributes to specific application profiles that are used by different platforms and services, such as search engines (e.g., Google Scholar, Microsoft Academics) and metadata aggregators. The work presented in[26] provides crosswalks among the most commonly used metadata schemes and guidelines to describe digital objects in Open Science, including RDA metadata IG recommendation of the metadata element set, EOSC Pilot - EDMI metadata set, Dublin CORE Metadata Terms, Datacite 4.3 metadata schema, DCAT 2.0 metadata schema and DCAT 2.0 application profile, EUDAT B2Find metadata recommendation, OpenAIRE Guidelines for Data Archives and many more.

[26] Milan Ojsteršek, Crosswalk of most used metadata schemes and guidelines for metadata interoperability, January 5, 2021, https://doi.org/10.5281/zenodo.4420116, [Last accessed 30 August 2021]

The following table presents a subset of the mapping for three attributes, identifier, creator and title, amongst five metadata formats, RDA, EOSC-EDMI, Dublin Core, Datacite and DCAT 2.0.

Table 12 – Example Crosswalk Mapping between different metadata formats

Property	RDA metadata IG	EOSC - EDMI	Dublin Core	Datacite 4.3	DCAT 2.0
Identifier	Unique Identifier Location (URL)	identifier url accessUrl accessInterface	dct:identifier	Identifier (M) Identifier Type	dct:identifier
Creator	Originator (organisation(s) / person(s))	creator	dct:creator	Creator (M) creatorName nameType givenName familyName nameIdentifier (R) nameIdentifierScheme schemeURI affiliation affiliationIdentifier affiliationIdentifierScheme schemeURI	dct:creator
Title		name	dct:title dct:alternative	Title (M) TitleType	dct:title

The final version of the transformation framework will be made available by the end of the project, when the final version of the SLICES Interoperability Framework and the final metadata profiles will be presented.

8 Metadata Documentation (Meta-Metadata)

The SLICES metadata profiles, including their definitions, types, taxonomies, classification schemes, and vocabulary schemes should be made publicly available and with access via appropriate APIs to improve machine-actionability. All information should be established serialization formats (e.g., YAML, JSON, XML), so that they can be used in external platforms or systems.

Versioning information should be incorporated in the documentation to enable mappings between different versions of the same metadata records further enhancing reusability and interoperability.

9 Ethical Considerations on Metadata

When referring to metadata, we are considering data about the data itself that is available, either purposefully or by the system such that the creator of the data did not explicitly approve the metadata. Metadata may be categorised as one of two types: 1) the “non-visible” data generated automatically, e.g., by the software application, or, 2) author-created data, easily visible to the author and others. In the first case, “non-visible” metadata can often reveal information about the creators (persons/organisations), versioning information, drafting/editorial information, etc. that may violate

ethical and privacy guidelines if made available to external parties. Given the existence of such metadata, it is important that mining is handled with care by individuals or systems such that ethical violations are prevented. In terms of transparency, it is important that data creators can view all metadata, including automatically generated metadata, and can erase the metadata, if they wish.

There are concerns that as technology advances metadata transparency may not be completely possible. However, in terms of ethics and privacy considerations, it is important that any metadata attached to a piece of data can be viewed, edited and deleted by the data owner. Moreover, the data owner should be permitted to hide such metadata from others if necessary.

In terms of open data, however, we are dealing with data that is available for anyone to access, use and share. In the case we are dealing with open research data, then if metadata contains sensitive information about individuals or organisations, appropriate controls should be in place to safeguard such personal/sensitive data. In terms of author-generated metadata, we must remember that open data is findable, accessible, interoperable and reusable (according to the FAIR principles), and therefore datasets and software must be published with appropriate metadata such that they can be found, accessed and used.

Moreover, in terms of reusability, appropriate guidelines must be followed to address potential issues that may hinder FAIR data. For example, repository interoperability must be accomplishable, without compromising privacy or anonymity of the data, or data types and formats may hinder researchers from reusing the data, especially in inter-disciplinary contexts. Better metadata can help finding and reusing data, and so can the use of similar formatting standards across disciplines. Overall, we may identify ethical issues with regards to open and reusable data[27] that relate to the following concerns: 1) the level of data trustworthiness and reliability, 2) the data anonymity, privacy and confidentiality, 3) data authorship and licensing. The first concern and second concerns cannot be resolved simply by the use of metadata, but metadata that is made available without the intent of the data creator can cause ethical issues relating to these concerns. The third concern is to ensure that metadata exists such that recognition of authorship is possible when reusing others' data and that information on licencing is available because it is essential when reusing research data.

For better and FAIRer management that considers the ethics of promoting open data, while being responsible about ethical and privacy violations, data inventories can be used where metadata explicitly refers to software and dataset contents, source, licensing and similar useful information. Data repositories aid in supporting research integrity by controlling the data storage and release. This can only be done if the data's sources and processing are known, and metadata can be clearly defined to support such research integrity guidelines[28].

[27] Bote, Juanjo and Termens, Miquel (2019). Reusing Data: Technical and Ethical Challenges, DESIDOC Journal of Library & Information Technology, v. 39, pp. 329 – 337, doi: 10.14429/djlit.39.6.14807
[28] Libby Bishop (2017). Big data and data sharing: Ethical issues. UK Data Service, UK Data Archive.

10 Conclusion

This deliverable presents the initial design of the SLICES metadata profiles, including the rationale behind their design and structure. The proposed design complies fully with the FAIR principles but also adopts important guiding principles enabling efficient and effective interoperability and cross-disciplinary research.

SLICES proposes the utilization of a hierarchical model consisting of compulsory metadata attributes that are domain-agnostic and can describe any digital object (e.g., data, services) ensuring that it conforms to FAIR principles and beyond. Where appropriate, SLICES will support additional optional metadata attributes accompanied by their metadata model to further enhance the description of the object. The model comprehensively describes data objects and also allows to be easily extended with new attributes or new types/categories as well as with new additional hierarchy levels.

The initial design describes the SLICES core FAIR Digital Object (S-FDO), which has been created by analysing predominant metadata catalogs with the objective to describe a digital object adequately to facilitate FAIR principles and ensure a high degree of interoperability. We then present important type-specific metadata categories, such as data, publication, software and services.

In the future versions of this deliverable, the metadata profiles will be enhanced/extended as the SLICES infrastructure design becomes more precise.


slicesDS