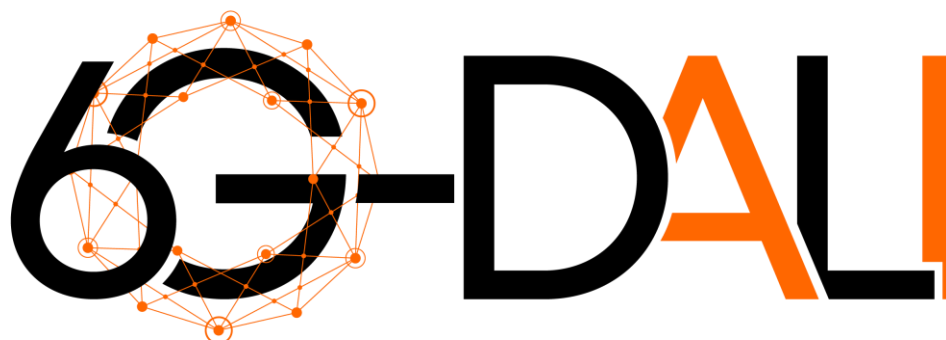




6G-DALI External Dataset Contribution Guide

Project	6G-DALI (SNS-JU)
Version	0.1 (draft)
Date	2026-05-27
Companion document	6G-DALI Metadata Application Profile

Table of Contents

Table of Contents.....	3
1 Overview	5
2 Why Contribute Your Dataset to the 6G-DALI Data Space?	6
2.1 Accelerate the Impact of Your Project Results.....	6
2.2 Improve Dataset Quality through 6G-DALI DataOps Services.....	6
2.3 Increase Dataset Value through Data Space Integration	6
2.4 Prepare Data for AI and MLOps Workflows.....	6
3 Who Can Contribute	7
4 Contribution Workflow.....	8
5 Submission Form — Metadata Fields	9
5.1 Mandatory Fields.....	9
5.2 Recommended Fields	9
5.3 Optional Fields	9
6 What Contributors Receive Back	11
7 Data Submission.....	12
7.1 Direct Upload	12
7.2 External Link.....	12
8 Validation Process.....	13
8.1 Metadata Validation	13
8.2 Data Quality Validation.....	13
8.2.1 Supported expectations.....	13
8.3 Validation Outcomes.....	14
9 Catalogue Registration	15
10 DataOps on Contributed Datasets	16
10.1 What DataOps Can Do	16
10.2 How It Works.....	17
10.3 Running a DataOps Pipeline	18
10.4 General Integration Notes	18
11 Datasets from the SLICES-RI Metadata Registry System (MRS)	19
11.1 What is MRS.....	19
11.2 Ingestion Pipeline	19
11.3 Identifiers and Cross-Linking.....	19
11.4 Metadata Mapping Summary	19
11.5 Key Properties of MRS-Ingested Datasets	20
12 Contributor Responsibilities.....	21
13 Submission Checklist	22

List of Figures

Figure 1 External Dataset Contribution Workflow	5
---	---

List of Tables

Table 1. Mandatory Fields	9
Table 2. Recommended Fields	9
Table 3. Optional Fields	9
Table 4. File Formats.....	12
Table 5. Supported expectations.....	13

Table 6. Data Cleaning and Quality Improvement.....	16
Table 7. Data Augmentation	16
Table 8.Format Conversion.....	17
Table 9. MRS Fields Translations.....	19

1 Overview

The 6G-DALI Data Space accepts dataset contributions from external parties — organisations, research groups, and individuals that are not direct partners of the 6G-DALI project but whose datasets are relevant to 5G/6G research, network slicing, AI/ML for networks, or related domains.

The objective is not simply to collect datasets into a repository, but instead, to transform isolated project datasets into interoperable, governed, high-quality, AI-ready assets that can be shared, reused, enriched, and operationalized through the first European 6G Dataspace delivered by 6G-DALI.

External contributors submit their datasets through the **6G-DALI Data Space submission portal**. Each submission consists of two parts:

- **Metadata** — a structured description of the dataset following the **6G-DALI Metadata Application Profile** (DCAT-AP 3.0 with 6G-DALI extensions), filled in through a guided web form.
- **Data** — the actual dataset files, uploaded through the same portal.

Upon submission, the portal immediately registers the dataset in the **Staging Catalogue** of the 6G-DALI Data Space and stores the data file in the **6G-DALI Data Repository**. The **6G-DALI validation pipeline** then runs automated metadata and data quality checks and writes the results back into the dataset's description in the Staging Catalogue. If all validations pass, the dataset is **promoted to the Community Datasets Catalogue**, where it becomes discoverable alongside 6G-DALI-native datasets and eligible for inclusion in DataOps pipelines.

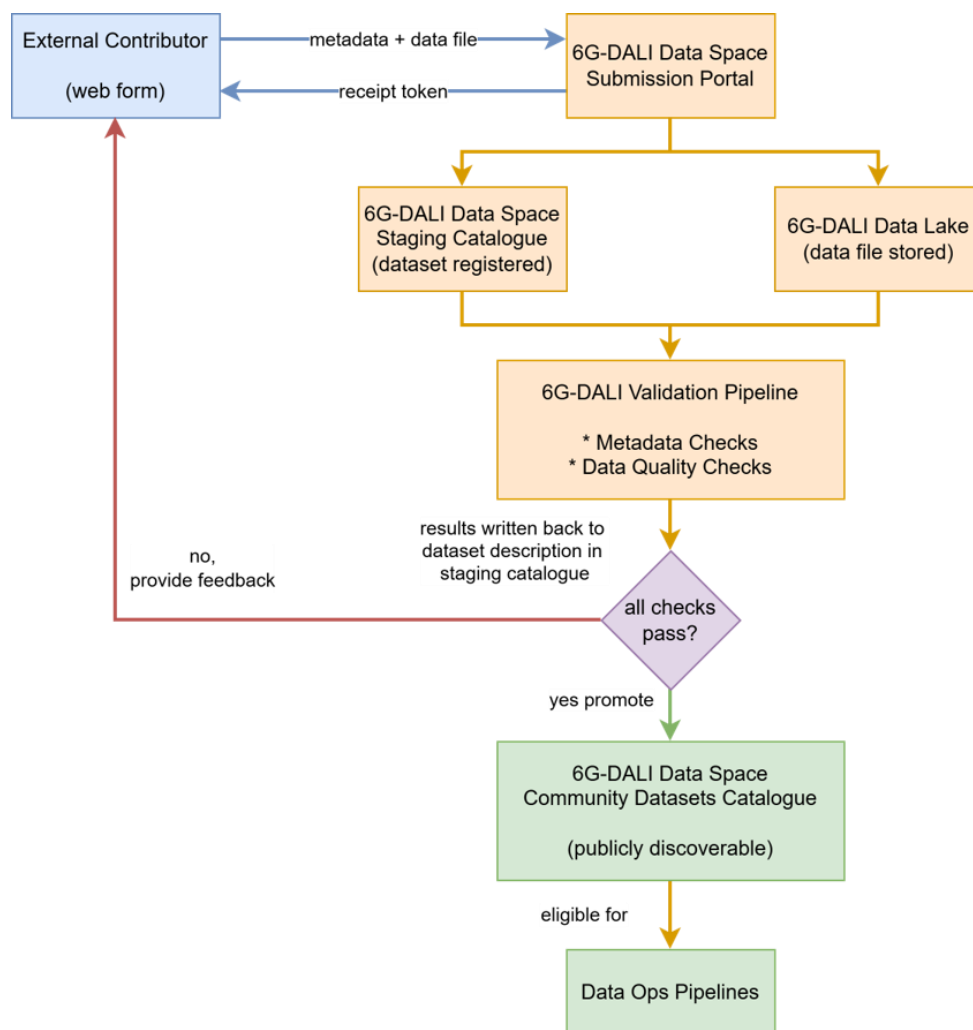


Figure 1 External Dataset Contribution Workflow

2 Why Contribute Your Dataset to the 6G-DALI Data Space?

2.1 Accelerate the Impact of Your Project Results

By contributing datasets to the 6G-DALI Data Space, SNS projects increase the visibility, discoverability, and reusability of their experimental results across the European 6G ecosystem. Rather than remaining confined to a single project repository, datasets become part of a federated 6G Data Space that is standards-compliant (International Data Spaces Association (IDSA) & GAIA-X principles), where they can be discovered and reused by:

- Other SNS EU projects
- Research infrastructures
- Open-source initiatives
- Standardization activities
- Industrial stakeholders

2.2 Improve Dataset Quality through 6G-DALI DataOps Services

Contributed datasets can be processed using the 6G-DALI DataOps framework, enabling:

- Data quality assessment
- Data cleaning
- Missing-value recovery
- Outlier detection
- Duplicate elimination
- Schema harmonization
- Metadata enrichment
- Provenance tracking

Projects receive improved and validated versions of their datasets without modifying the original source data.

2.3 Increase Dataset Value through Data Space Integration

A dataset becomes more valuable when connected with other datasets. The 6G-DALI Data Space enables:

- Cross-project dataset federation
- Semantic interoperability
- Metadata harmonization
- Cross-domain data discovery
- Dataset enrichment using contextual information from other SNS datasets

This enables new experimentation opportunities that would not be possible within isolated project repositories.

2.4 Prepare Data for AI and MLOps Workflows

Datasets published in the 6G-DALI Data Space become immediately usable by:

- AI training pipelines
- MLOps environments
- Digital Twins
- Data analytics workflows
- Federated AI experimentation

Contributors therefore gain access to a broader AI experimentation ecosystem.

3 Who Can Contribute

Any organisation or individual may submit a dataset, provided they:

- Hold the rights to share the dataset under an open or restricted licence compatible with the 6G-DALI Data Space terms.
- Can provide the mandatory metadata described in Section 5.
- Accept the 6G-DALI [Contributor Responsibilities](#).

Typical contributors include:

- Academic and research institutions running 5G/6G testbeds outside the 6G-DALI consortium.
- Standards bodies publishing reference datasets (e.g., 3GPP, ETSI).
- Industry partners contributing anonymised operational data.
- Other EU projects in the SNS-JU portfolio.

4 Contribution Workflow

1. **Register** — Create an account on the 6G-DALI Data Space portal and receive a contributor ID.
2. **Fill form** — Complete the guided metadata form (see Section 5).
3. **Submit** — Upload files or provide a link to the dataset. The portal registers the dataset in the Staging Catalogue and stores the data in the 6G-DALI Data Repository. Receive a submission receipt token.
4. **Validation** — Automated metadata and data quality checks run against the uploaded data (see Section 8).
5. **Results** — Validation results are written into the dataset's description in the Staging Catalogue. The contributor can review them via the portal.
6. **Promotion** — If all checks pass, the dataset is promoted to the Community Datasets Catalogue. The contributor is notified by e-mail with the public dataset URL.

Typical turnaround: validation completes within minutes of submission. The contributor can monitor progress using the receipt token on the portal status page.

5 Submission Form — Metadata Fields

The web form collects metadata conforming to the 6G-DALI Metadata Application Profile. Fields are grouped into mandatory, recommended, and optional tiers. The portal generates a valid, standards-compliant metadata record from the submitted values.

5.1 Mandatory Fields

Table 1. Mandatory Fields

Field	Notes
Title	Human-readable name of the dataset; accepts multiple language variants.
Description	Free-text description of content, collection context, and intended use.
Publisher	Organisation name and URL of the entity making the dataset available.
Contact point	Name and e-mail of the person responsible for the dataset.
Licence	A standard open licence (e.g., CC BY 4.0, ODbL) or a valid SPDX licence identifier.
Themes / Keywords	At least one thematic category from the EuroVoc or DCAT-AP vocabulary, plus free-text keywords.
Issued date	Date the dataset was first created or published.
Distribution format	File format of each uploaded file, e.g., CSV, HDF5, PCAP.

5.2 Recommended Fields

Table 2. Recommended Fields

Field	Notes
Temporal coverage	Time interval the data spans (start date / end date).
Dataset columns	Comma-separated list of column names in the order they appear in the file. Used to enforce column presence and ordering during validation; automatically enables the ExpectColumnToExist quality check as compulsory.
Related 5G/6G frequency bands	E.g., FR1, FR2, sub-THz.
Network technology	E.g., 5G NR, 6G, Wi-Fi 7.
Data collection method	How the data was produced: measurement, simulation, emulation, crowdsourcing.
Provenance statement	Brief description of the experiment or activity that generated the data.

5.3 Optional Fields

Table 3. Optional Fields

Field	Notes
-------	-------

Identifier (DOI / accession)	An existing persistent identifier, if one is already assigned.
Related publications	URLs or DOIs of papers, reports, or related datasets.
Additional contributors	Names and roles of individuals or organisations beyond the primary publisher.
Access rights	PUBLIC, RESTRICTED, or NON_PUBLIC. Defaults to PUBLIC.
Quality expectation suite	Name of a predefined data quality rule set to apply during validation (see Section 8.2).

6 What Contributors Receive Back

For every contributed dataset, the project receives:

- **Immediate Outputs** - Persistent dataset identifier - Standards-compliant metadata record - Data quality validation report - Dataset discoverability in the Community Catalogue
- **DataOps Outputs** - depending on selected DataOps services: - Cleaned datasets - Normalized datasets - Harmonized schemas - Derived KPIs - Enriched datasets - Synthetic or augmented datasets - Alternative formats optimized for AI workflows
- **Governance Outputs** - Provenance records - Lineage tracking - Quality annotations - Metadata enrichment
- **Ecosystem Outputs** - Exposure to the SNS community - Reuse by other projects - Potential inclusion in joint experimentation campaigns

7 Data Submission

The submission portal offers two ways to provide the actual dataset. The contributor may use either or both.

7.1 Direct Upload

Upload one or more tabular data files through the portal's file uploader. Accepted formats:

Table 4. File Formats

Format	Extensions
Comma-separated values	.csv
Tab-separated values	.tsv
Apache Parquet	.parquet
Apache ORC	.orc
ZIP archive	.zip — must contain only files in the formats above

File size limit: up to **1 GB** per upload session. For larger datasets, use the external link option (Section 7.2) or contact the 6G-DALI data team at data@6gdali.eu.

7.2 External Link

Instead of uploading files, the contributor may provide a URL pointing to the dataset hosted on an external system (e.g., a repository, object storage bucket, or institutional data portal). The 6G-DALI platform will retrieve the data from that URL during the validation step.

Requirements for linked datasets:

- The URL must be publicly accessible or accompanied by access credentials supplied through the submission form.
- The file at the URL must be in one of the tabular formats listed in Section 7.1.
- The contributor is responsible for keeping the URL available for the duration of the dataset's listing in the 6G-DALI Data Space (see Section 12).

After data is received — whether by upload or retrieval — the portal:

1. Computes a checksum and records it in the metadata to guarantee file integrity.
2. Records the file size automatically.
3. Stores the data in the **6G-DALI Data Repository**.
4. Registers the dataset in the Staging Catalogue of the 6G-DALI Data Space.
5. Returns a **submission receipt token** that the contributor uses to track validation status on the portal status page.

8 Validation Process

Every submitted dataset passes through the **6G-DALI validation pipeline** before being promoted to the public catalogue. The pipeline runs two classes of checks.

8.1 Metadata Validation

The submitted metadata is checked for completeness and correctness:

- All mandatory fields are present and correctly formatted.
- Selected themes and licence identifiers are drawn from recognised standard vocabularies.
- Publisher and contact information follow the expected structure.
- Any 6G-DALI-specific fields use values from the 6G-DALI controlled vocabulary.

A validation report is generated. Errors block promotion to the Community Datasets Catalogue; warnings are reported to the contributor but do not block. The full report is written into the dataset's description in the Staging Catalogue so the contributor can inspect it directly through the portal.

8.2 Data Quality Validation

The pipeline validates every submitted tabular dataset using **Great Expectations**, an open-source data quality framework. Validation is driven by an expectation suite — a set of rules that the dataset must satisfy. The contributor defines the suite for their dataset through the submission form.

8.2.1 Supported expectations

Checks marked **Compulsory** are always enforced and cannot be disabled. The contributor must supply the required parameters for these checks before submitting. All other checks are optional.

The following expectations are available when configuring a suite:

Table 5. Supported expectations

Expectation	What it checks
ExpectTableRowCountToBeBetween	Total number of rows falls within a specified minimum and maximum.
ExpectTableColumnCountToBeBetween	Total number of columns falls within a specified range.
ExpectTableColumnsToMatchOrderedList	Columns appear in a specific order and with specific names.
ExpectColumnToExist	A named column is present in the dataset.
ExpectColumnValuesToNotBeNull⚠ Compulsory	A column contains no null values. This check is always enforced; the contributor must specify the column name.
ExpectColumnValuesToBeNull	A column contains only null values.
ExpectColumnValuesToBeOfType	All values in a column are of a specified data type (e.g., integer, string).
ExpectColumnValuesToBeInSet	All values in a column belong to a defined set of allowed values.
ExpectColumnValuesToNotBeInSet	No values in a column belong to a defined set of disallowed values.

ExpectColumnValuesToBeBetween	All values in a column fall within a specified numeric range.
ExpectColumnValuesToBeDateuntilParseable	All values in a column can be parsed as valid dates.

Contributors specify the expectations and their parameters in the submission form. The pipeline evaluates each expectation against the uploaded data and produces a per-expectation pass/fail result.

All results — scores, per-expectation outcomes, and the overall pass/fail status — are written back into the dataset's description in the Staging Catalogue. Contributors can review them through the portal before promotion; downstream users see the same quality summary once the dataset is in the Community Datasets Catalogue.

8.3 Validation Outcomes

In all cases the dataset remains visible in the Staging Catalogue and the results are written into its description. Promotion to the Community Datasets Catalogue depends on the outcome:

Outcome	Condition	Result
Promoted	No errors; all mandatory quality checks pass.	Dataset published in the Community Datasets Catalogue. Contributor notified by e-mail with the public dataset URL.
Promoted with warnings	No errors; some non-critical quality issues flagged.	Dataset promoted; warnings remain visible in the public record. Contributor notified.
Not promoted — metadata	One or more metadata errors found.	Dataset remains in the Staging Catalogue. Detailed report written into the description. Contributor must correct the flagged fields and resubmit.
Not promoted — data	File integrity failure or critical data quality issue (e.g., empty file, all-null columns).	Dataset remains in the Staging Catalogue. Contributor notified; resubmission required.

9 Catalogue Registration

The 6G-DALI Data Space uses two catalogues in the contribution process:

Catalogue	Purpose
Staging Catalogue	Staging area. All submitted datasets appear here immediately upon submission. Validation results are written into each dataset's description. Not publicly discoverable.
Community Datasets Catalogue	Public catalogue. Contains only datasets that have passed all validation checks. Discoverable alongside 6G-DALI-native datasets.

Each contributed dataset is assigned:

- A **persistent, globally unique identifier** (URI) within the 6G-DALI Data Space.
- A provenance record linking the dataset to the contributor's organisation.
- An “**External Contribution**” label visible in the 6G-DALI Data Space portal.

Once in the Community Datasets Catalogue, the dataset is **discoverable** through:

- The 6G-DALI Data Space portal search interface.
- The 6G-DALI Data Space API.

10 DataOps on Contributed Datasets

Once published in the Community Datasets Catalogue, an externally contributed dataset is a first-class citizen of the 6G-DALI Data Space. The 6G-DALI **DataOps framework** can be applied to it to run transformations, fixes, and augmentations — producing new derived datasets that are registered in the catalogue alongside the original.

10.1 What DataOps Can Do

The 6G-DALI DataOps framework follows an **Extract-Load-Transform (ELT)** paradigm tailored to 6G experimentation data. Processing capabilities are packaged as microservices and assembled into pipeline workflows that are orchestrated by Apache Airflow and driven by the 6G Data Space catalogue. The framework handles both cold datasets (historical data already registered in the Data Space) and hot datasets (produced on demand by live testbed experiments). The following operation types are planned or under development as part of the 6G-DALI DataOps tools stack.

Table 6. Data Cleaning and Quality Improvement

Capability	Description	6G example
Missing-value imputation	Fill gaps using AI-based time-series imputation and tabular data completion algorithms, targeting at least a 30% improvement in dataset completeness relative to the raw baseline.	Reconstruct missing RSRP samples in a network measurement time-series caused by UE disconnections.
Outlier removal	Detect and remove anomalous measurements that fall outside physically plausible bounds.	Remove spurious throughput spikes introduced by testbed instrumentation errors.
Duplicate elimination	Identify and drop duplicate records to ensure uniqueness across primary key fields.	Deduplicate CSI snapshots collected independently by two overlapping monitoring probes.
Format normalisation	Standardise timestamps, column naming conventions, and value encodings across datasets from different testbeds.	Convert heterogeneous timestamp formats from KUL and EUR testbeds to a common UTC ISO-8601 encoding.
Dataset reduction and aggregation	Reduce dataset size or aggregate measurements across time windows or spatial dimensions to produce compact, analysis-ready summaries.	Aggregate per-UE RAN performance metrics into 1-minute sliding-window averages for ML model training.

Table 7. Data Augmentation

Capability	Description	6G example
Cross-dataset enrichment	Join the contributed dataset with other resources registered in the 6G Data Space to add contextual attributes.	Attach cell-tower coordinates and antenna configuration metadata to a UE measurement dataset.

Surrogate-model generation	Train surrogate models on a sampled subset of real measurements and use them to generate the remaining experimental space, reaching a user-agreed confidence threshold before synthetic data is accepted.	Extend a beamforming dataset covering 20 antenna configurations to all 120 configurations in the experimental space, avoiding exhaustive physical measurements.
Computed KPI derivation	Apply domain-specific formulas to raw columns to produce derived performance indicators.	Compute SINR, spectral efficiency, or handover success rate from raw measurement columns.

Table 8.Format Conversion

Capability	Description	6G example
Format conversion	Produce alternative file formats for downstream usability in MLOps pipelines or analytical tools.	Convert CSV to Parquet for performance-sensitive ML training, produce a compressed archive for bulk transfer.
Schema alignment	Reshape wide-format tables or restructure column layouts to match a target schema required by a downstream AI model.	Pivot a multi-UE wide table into a long-format time-series compatible with the 6G-DALI MLOps training API.

10.2 How It Works

DataOps pipelines in the 6G-DALI Data Space are registered as `dcatalog:DataService` resources and orchestrated through Apache Airflow as dynamically generated DAGs. A key design principle is **Data Space-Native Orchestration**: pipeline inputs are resolved through metadata queries against the 6G Data Space catalogue rather than hard-coded file paths. Experimenters specify what data they need (by metadata criteria) and the orchestrator resolves the actual storage location, retrieves access credentials, and injects them into the pipeline at execution time.

When a pipeline consumes a contributed dataset and produces output, the following happens automatically:

1. The output is stored in the 6G-DALI Data Repository as a new derived dataset.
2. A new `dcatalog:Dataset` record is registered in the Community Datasets Catalogue, linked to the original contributed dataset via a `prov:wasDerivedFrom` provenance relationship.
3. The pipeline that produced it is recorded as `prov:wasGeneratedBy`, preserving full lineage.
4. Any quality annotations produced during the pipeline run are attached to the derived dataset record as `dqv:QualityMeasurement` entries.

The resulting provenance graph enables data lineage visualisation — tracing derivation chains from raw testbed measurements to cleaned and augmented analytical datasets. The original contributed dataset is never modified. All DataOps operations produce new versioned records, keeping the source data intact and auditable.

10.3 Running a DataOps Pipeline

External contributors whose datasets have been published in the Community Datasets Catalogue can run DataOps pipelines directly through the **6G-DALI DataOps Dashboard**. The dashboard presents the available Data Processing Services, allows configuration of pipeline parameters, and provides real-time monitoring of pipeline execution status and logs.

To run a pipeline:

1. Log in to the 6G-DALI Data Space portal and navigate to the DataOps Dashboard.
2. Select the contributed dataset by its identifier from the Community Datasets Catalogue.
3. Choose the desired Data Processing Service (transformation, augmentation, format conversion, etc.).
4. Configure the pipeline parameters and any reference datasets required for augmentation.
5. Submit the pipeline. The dashboard displays execution progress and notifies you when the derived dataset is available in the catalogue.

10.4 General Integration Notes

- Only datasets in the Community Datasets Catalogue are eligible as pipeline inputs. Datasets still in the Staging Catalogue cannot be referenced by pipelines.
- Within the DataOps Dashboard, contributors can use their own published datasets and any other dataset marked PUBLIC in the Community Datasets Catalogue as pipeline inputs.
- Quality results produced by pipeline runs are written back to the dataset's entry in the Community Datasets Catalogue, keeping the quality picture current for all users.

11 Datasets from the SLICES-RI Metadata Registry System (MRS)

In addition to the manual submission workflow described in this guide, the 6G-DALI Data Space ingests datasets automatically from the **SLICES-RI Metadata Registry System (MRS)** — the centralised metadata catalogue of the **SLICES Research Infrastructure**. This provides a second, automated pathway through which external datasets enter the 6G-DALI catalogue without requiring contributors to fill in the submission form.

11.1 What is MRS

The SLICES-RI MRS is a FAIR-compliant metadata registry that manages datasets, publications, software, and other digital objects produced within the SLICES infrastructure. Metadata is stored in a versioned JSON profile (`Sllices_v0_2`) organised in three levels: domain-agnostic mandatory fields (identifier, name, description, version), type-specific fields for datasets (format, byte size, temporal coverage, access mode, licence), and optional domain-specific extensions. Dataset content is stored separately and accessed via pre-signed download URLs returned by the MRS API.

11.2 Ingestion Pipeline

An automated ingestion pipeline — the AF-MRS Pipeline — runs on a scheduled basis and synchronises MRS dataset records into the 6G-DALI Data Space catalogue. For each record it processes, the pipeline:

1. **Fetches** the dataset record from the MRS REST API, including all metadata fields and the pre-signed download URL for the data content.
2. **Transforms** the MRS JSON record into a DCAT-AP 3.0 RDF document following the 6G-DALI Metadata Application Profile. The mapping covers all core MRS fields: title, description, creators, licence, format, compression, byte size, temporal coverage, spatial coverage (country and coordinates), access rights, provenance, and keywords. Fields absent from MRS but required by the 6G-DALI profile (such as `dali:snsProjectName` set to "SLICES-RI" and `dali:fairCompliant` set to `true`) are added automatically during transformation.
3. **Validates** the resulting RDF document against the 6G-DALI SHACL shapes before registration, ensuring that all mandatory fields are populated and correctly formatted.
4. **Registers** the record in the 6G-DALI Data Space catalogue using an idempotent PUT operation, so that re-running the pipeline on an already-registered record performs an update rather than creating a duplicate.

Incremental sync is supported: the pipeline filters MRS records by modification timestamp, retrieving only records changed since the last successful run.

11.3 Identifiers and Cross-Linking

Each MRS-ingested dataset is assigned a stable identifier in the 6G-DALI catalogue of the form `sllices-<MRS-UUID>`, distinguishing it from manually submitted and 6G-DALI-native records. The original MRS UUID is preserved in an `adms:Identifier` node with scheme agency "SLICES-RI MRS", and a `dct:source` back-link points to the authoritative MRS record URL — enabling unambiguous cross-system lookup in both directions.

If the MRS record also carries a DOI (DataCite identifier), it is propagated as a second `adms:Identifier` entry with scheme agency "DataCite".

11.4 Metadata Mapping Summary

The table below summarises how the key MRS fields are translated into the 6G-DALI DCAT-AP profile:

Table 9. MRS Fields Translations

MRS Field	6G-DALI DCAT-AP predicate
identifier	<code>dct:identifier</code> , <code>adms:identifier</code> (cross-link)
name	<code>dct:title</code>

description	dct:description
version	adms:version
createdAt	dct:issued
dateTimeStart/ dateTimeEnd	dct:temporal → dcat:startDate / dcat:endDate
creators	dct:creator → foaf:Person
locations	dct:spatial(country URI), schema:geo (coordinates)
accessType	dct:accessRights
licenseUri / license	dct:license
format/byteSize/ compressionFormat	dcat:distribution properties
keywords+ scientificSubdomains	dcat:keyword
scientificDomains	dcat:theme
provenance	dct:provenance

11.5 Key Properties of MRS-Ingested Datasets

- **MRS is the source of truth.** The 6G-DALI catalogue holds a read replica enriched with 6G-DALI-specific extensions. Authoritative changes to the metadata must be made in MRS first and will be reflected in the 6G-DALI catalogue on the next scheduled sync.
- **Once registered, MRS-sourced datasets are first-class citizens.** They appear in the Community Datasets Catalogue alongside manually contributed and 6G-DALI-native datasets, and are eligible for DataOps pipelines (see Section 10).
- **Data access follows MRS access controls.** The dcat:accessURL in the 6G-DALI record points back to the MRS record page. Pre-signed download URLs are time-limited and not stored in the catalogue; users requiring direct data access should retrieve a fresh URL from the MRS API or the [SLICES-RI portal](#).

12 Contributor Responsibilities

By submitting a dataset, the contributor confirms that:

1. **Ownership and rights** — they hold the necessary rights to share the dataset under the stated licence and that sharing does not violate any third-party agreements, personal data regulations (GDPR), or export control rules.
2. **Accuracy** — the metadata provided is accurate and complete to the best of their knowledge.
3. **Notification of changes** — they will notify the 6G-DALI data team at data@6gdali.eu if the dataset is retracted, significantly updated, or if the licence terms change.
4. **No malicious content** — the uploaded files do not contain malware, obfuscated scripts, or content designed to harm downstream consumers or the 6G-DALI infrastructure.
5. **Persistent availability** — for metadata-only submissions that reference an external hosting location, the contributor will maintain availability at that URL for at least 3 years from the publication date.

The 6G-DALI project reserves the right to remove or delist any dataset that is found to violate these responsibilities, without prior notice.

13 Submission Checklist

Use this checklist before submitting:

- ☐ I have created an account on the 6G-DALI Data Space portal and have a contributor ID.
- ☐ I have filled in all **mandatory** metadata fields (Section 5.1).
- ☐ I have selected a valid open licence.
- ☐ I have assigned at least one thematic category and at least two keywords.
- ☐ My dataset files are in one of the accepted formats (Section 7).
- ☐ I have confirmed I hold the rights to share this dataset under the stated licence.
- ☐ I have noted my **submission receipt token** to track validation status.
- ☐ I understand that the dataset will be publicly visible in the 6G-DALI Data Space once validation passes (unless I set access rights to RESTRICTED).