



# Fondazione Gran Sasso Tech

## **WP2: Data-Lake Infrastructure MAPS**

Technical Design, Cloud Solution Design, Hosting Specifications

**Version:** 1.0

**Date:** 2026-03-25

**Author:** Guglielmo Celata

# 1 WP6: API and Open Data Publication

**Work Package 6:** REST API and open data catalogue according to the DCAT-AP\_IT standard

## 1.1 Overview

Development of **REST API v2** for programmatic access to SLO data and publication of open data datasets according to the European **DCAT-AP\_IT** standard.

Objectives: - Democratise access to MAPS data - Support third-party application development - Compliance with European open data standards - Integration with national open data portals

## 1.2 Timeline

- **Months:** M20-M24
- **Final deadline:** 31/03/2027

## 1.3 Planned Deliverables

### 1.3.1 D6.1 - API Specifications and OpenAPI Documentation

API design document: - OpenAPI 3.1 specifications - Authentication and authorisation - Rate limiting and quotas - Integration examples

### 1.3.2 D6.2 - API v2 Implementation

REST API implementation with: - FastAPI/Flask backend - JWT authentication - Interactive documentation (Swagger/Redoc) - Client libraries (Python, JavaScript)

### 1.3.3 D6.3 - Open Data Catalogue and DCAT-AP\_IT Metadata

Publication of open data datasets via CKAN: - **CKAN 2.11 catalogue** compliant with DCAT-AP\_IT - Complete metadata (lineage, quality, Creative Commons licences) - Automatic harvesting towards dati.gov.it - Dedicated user-friendly open data portal - Public API for programmatic download - Multi-format preview (CSV, GeoJSON, Parquet)

## 1.4 Planned API Endpoints

### 1.4.1 REST API v2 (FastAPI)

```
GET /api/v2/slo # List of Homogeneous Local Systems (SLO)
GET /api/v2/slo/{id} # SLO detail
GET /api/v2/comuni # List of municipalities with attributes
GET /api/v2/comuni/{codice} # Municipality detail
GET /api/v2/isocrone/{codice} # Accessibility isochrones
GET /api/v2/polarita # Territorial attractors
GET /api/v2/datasets # Available dataset catalogue
```

### 1.4.2 CKAN API (Open Data)

CKAN automatically exposes standard APIs for each published dataset:

```
# CKAN Action API (JSON)
GET /api/3/action/package_list           # Dataset list
GET /api/3/action/package_show?id={dataset_id} # Dataset metadata
GET /api/3/action/resource_show?id={resource_id} # Resource metadata
GET /api/3/action/datastore_search?resource_id={id} # Data query

# DCAT-AP_IT Endpoint (RDF/XML)
GET /catalog.xml                         # Complete DCAT-AP_IT catalogue
GET /dataset/{dataset_id}.xml           # Single dataset DCAT metadata

# CSW Harvesting (for dati.gov.it)
GET /csw?request=GetCapabilities
GET /csw?request=GetRecords&outputSchema=http://www.isotc211.org/2005/gmd
```

## 1.5 Publication Pipeline

Prefect automated workflow:

```
@flow(name="publish-to-ckan")
def publish_open_data():
    """
    Pubblica dataset approvati da Gold layer a CKAN

    Steps:
    1. Query gold.open_data_approved (flag published=false)
    2. Export dataset (CSV, GeoJSON, Parquet)
    3. Generate DCAT-AP_IT metadata
    4. Upload to CKAN via API
    5. Update published=true flag
    6. Notify dati.gov.it harvester
    """

    approved_datasets = get_approved_datasets()

    for dataset in approved_datasets:
        # Export da PostgreSQL
        export_dataset(dataset)

        # Genera metadata DCAT-AP_IT
        metadata = generate_dcat_metadata(
            dataset,
            license="CC-BY-4.0",
            publisher="Fondazione Gran Sasso Tech",
            temporal_coverage="2010-2025",
            spatial_coverage="Italia"
        )

        # Upload to CKAN
        ckan_client.package_create(
            name=dataset.slug,
            title=dataset.title,
```

```

        notes=dataset.description,
        extras=metadata
    )

    # Mark as published
    mark_as_published(dataset.id)

```

**Triggers:** - Manual: Dataset approval in OpenMetadata - Automatic: Weekly schedule for updated datasets

## 1.6 Dual-Catalog Architecture

The MAPS project adopts a **dual-catalog** architecture to separate internal governance from public access:

Tool	Audience	Purpose
<b>OpenMetadata</b>	Internal data operators (WP2)	Governance, lineage, data quality
<b>CKAN</b>	External public (WP6)	Open data catalogue, DCAT-AP_IT, dati.gov.it

This separation ensures: - **Contractual compliance:** CKAN satisfies WP6 D6.3 requirements out-of-the-box - **Separation of concerns:** Internal governance isolated from public access - **Standard compliance:** Native DCAT-AP\_IT, built-in dati.gov.it integration - **User experience:** Portal optimised for citizens/researchers vs. technical tool for data engineers

Workflow: Gold Layer → OpenMetadata (quality approval) → CKAN (publication) → dati.gov.it (harvesting)

## 1.7 Status

**In preparation** - Documentation will be available during the implementation phase (M20-M24)

## 1.8 Links

- ← Back to Home
- WP2 - Data Lake - Data source for API and CKAN
  - Dual-Catalog Architecture (OpenMetadata + CKAN)
- WP3 - SLO Algorithm - SLO data to expose
- WP5 - Simulator - Simulation results integration

---

**Author:** Guglielmo Celata - Fondazione Gran Sasso Tech