Synthetic UHS–MRV Dataset for OFP–OSDU Integration in Hydrogen Storage Reporting (SUHS-OFPOSDU)

Authors: Sreekanth Muktevi, Yogesh Nagpal, Rajesh Leela Thotakura, Jyotsna Muktevi

Abstract

This white paper describes the first publicly available synthetic dataset for Underground Hydrogen Storage (UHS) Measurement, Reporting, and Verification (MRV). The dataset, titled SUHS-OFPOSDU, is explicitly mapped to the Open Footprint (OFP) model and the Open Subsurface Data Universe (OSDU) schemas. It includes annual, monthly, and daily injection and withdrawal records across the storage–transport–monitoring chain. The dataset enables interoperability testing, standards-based validation, and academic research, while protecting confidentiality by providing entirely synthetic but realistic values.

Background & Motivation

Hydrogen is already widely used in industry for refining, ammonia/fertilizer production, methanol, and chemicals, with emerging pilots in steel, power, and mobility. To enable large-scale adoption of hydrogen as a clean energy carrier, underground storage is essential to balance seasonal demand, provide grid flexibility, and ensure supply continuity. Salt caverns are commercially used for hydrogen storage today, while depleted gas fields and aquifers are still in pilot or feasibility stage. Real-world UHS data is largely confidential. To address this gap, we developed a synthetic but realistic dataset that mirrors expected industry practices, allowing open testing, ESG reporting validation, and schema integration with OFP and OSDU.

Methods & Data Generation

The dataset was generated using Python with libraries including NumPy, Pandas, and Faker, ensuring reproducibility. Geospatial coordinates were created with GeoPy, bounded within the U.S. mainland, and explicitly marked as synthetic. Key design rules included:

- Mass balance enforced: Net = Injected − Withdrawn − Losses (≥0)
- Reservoir-specific logic: Salt cavern (low losses, high cycling, cushion gas constant); Depleted gas (moderate losses, possible methane co-production); Aquifers (higher uncertainty)
- Compression energy intensity validated against ranges of 0.1–0.5 MWh/t-cycle
- QA checks validated additivity (daily = monthly = annual), leak logic, geo bounds, and schema completeness

Schema Alignment

The dataset has been explicitly mapped to the OFP emissions model and OSDU Well-Known Schemas (WKS) and Well-Known Entities (WKE):

- WKS:master-data–Facility → facility_id, country_code, lat, lon, reservoir_type
- WKS:master-data–Asset → pipeline_length_km, transport_mode
- WKE:Wellbore / Well → well_id, injection_start_date, injection_end_date, avg_reservoir_pressure_MPa, avg_reservoir_temp_C
- WKE:ProductionData → h2_injected_tonnes, h2_withdrawn_tonnes, methane_co_produced_tonnes
- WKE:Measurement → h2_cushion_gas_tonnes, compression_energy_MWh, h2_losses_tonnes
- WKE:Event / Method → leak_events_count, leak_mass_tonnes, mmv_methods, ogmp_source_category
- Proposed EnvironmentalData extension → h2_net_stored_tonnes

Dataset Contents

The SUHS-OFPOSDU package includes:

- uhs_full_dataset_v1.0.csv – Facility-level annual summary (10 facilities, QA-checked)
- uhs_daily_v1.0.csv – Daily injection/withdrawal records with seasonality
- uhs_monthly_v1.0.csv – Monthly roll-up from daily
- schema/schema.yaml – Dataset schema definition
- schema/schema_crosswalk.csv – Field-to-OFP/OSDU/MRV mapping
- docs/NOTES.md – Methodology and QA notes
- LICENSE.txt – License (CC BY 4.0)
- CITATION.cff – Citation metadata
- README.md – User guide and authorship

Usage & Applications

This dataset is intended for use in:

- UHS MRV reporting system prototyping
- Data model integration testing between OFP and OSDU
- Benchmarking analytics pipelines for hydrogen storage monitoring
- Academic research, teaching, and training

Licensing & Citation

Citation:
Muktevi, S., Nagpal, Y., Thotakura, R. L., & Muktevi, J. (2025). Synthetic UHS–MRV Dataset for OFP–OSDU Integration in Hydrogen Storage Reporting (SUHS-OFPOSDU) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.TBD

Disclaimer

This dataset is entirely synthetic and does not represent any real company, facility, or reservoir data. Salt cavern examples reflect commercial practice, while depleted gas and aquifer cases are based on pilot and feasibility scenarios. It is provided solely for educational, research, and development purposes.