Synthetic CCS–MRV Dataset for OFP–OSDU Integration in Carbon Storage Reporting (SCCS-OFPOSDU)

Version 1.3 – October 2025

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

Abstract

This white paper describes the v1.3 release of the first publicly available synthetic dataset for Carbon Capture and Storage (CCS) Measurement, Reporting, and Verification (MRV). The dataset, titled SCCS-OFPOSDU, is explicitly mapped to the Open Footprint (OFP) model and the Open Subsurface Data Universe (OSDU) Well-Known Schemas (WKS) updated in 2025. It includes injection records across the capture–transport–storage–monitoring chain. The dataset supports testing, validation, education, and ESG platform integration.

Background & Motivation

Carbon Capture and Storage is essential to global net-zero goals. Regulatory frameworks such as EPA 40 CFR Part 98 Subpart RR, ISO 27916, and OGMP 2.0 mandate detailed MRV. Due to confidentiality and access limitations of real data, this dataset fills the gap with realistic synthetic values useful for prototyping digital MRV systems and testing schema integration with OFP and OSDU models.

Methods & Data Generation

The dataset was built using Python (NumPy, Pandas, Faker) with GeoPy for spatial generation, using facilities located within synthetic U.S. coordinates. All emissions and injection values are validated for internal consistency. Notable rules:

- Mass balance enforced: Injected − Produced − Loss − Leak = Net Stored
- Capture intensity mapped to standard technologies (MEA, MDEA, PSA, etc.)
- Leak events based on OGMP risk logic
- Reservoir classification influences expected CO_2 production and leak trends
- All data checked against schema completeness, additivity, and bounds

Schema Alignment

Aligned to OSDU 2025 canonical Well-Known Schemas (WKS):

- `WKS:master-data–Facility` → facility_id, location, reservoir_type
- `WKS:master-data–Asset` → capture_tech, pipeline_length, transport_mode
- `WKS:master-data–Wellbore` → injection_start, end dates, pressure/temp
- `WKS:work-product–ProductionData` → injected, produced CO_2
- `WKS:work-product–Measurement` → capture energy, loss, CH_4
- `WKS:work-product–EventMethod` → leaks, MMV methods, OGMP mapping
- `WKS:extension–EnvironmentalData` → derived net_stored_tonnes

Dataset Contents

The repository contains:
- `ccs_full_dataset_v1.0.csv`: annual, facility-level view
- `ccs_injection_daily_v1.0.csv`, `monthly_v1.0.csv`: time-series data
- `schema/schema.yaml`: JSON-style structure with validations
- `schema/schema_crosswalk.csv`: maps fields to OFP/OSDU/MRV
- `README.md`, `SCCS_MRV_dataset_whitepaper.pdf`: documentation
- `validation_summary.md`: results from schema-based QA
- `Dockerfile`, `requirements.txt`: reproducible environment

Usage & Applications

Potential use cases:
- Developing or testing CCS reporting systems
- Validating compliance with OFP or OSDU schemas
- Benchmarking AI/ML pipelines for emissions modeling
- Training in ESG disclosures or carbon data modeling
- Publishing reproducible research or simulations

Licensing & Citation

Citation:
Muktevi, S., Nagpal, Y., Thotakura, R. L., & Muktevi, J. (2025). Synthetic CCS–MRV Dataset for OFP–OSDU Integration in Carbon Storage Reporting (v1.3). Zenodo. [DOI Pending – v1.3]

Disclaimer

This dataset is entirely synthetic and does not represent any real company, site, or reservoir. It is intended solely for research, prototyping, and educational purposes.