

LB Solvent Workflow integration v2

Solvent Pipeline & SQL View

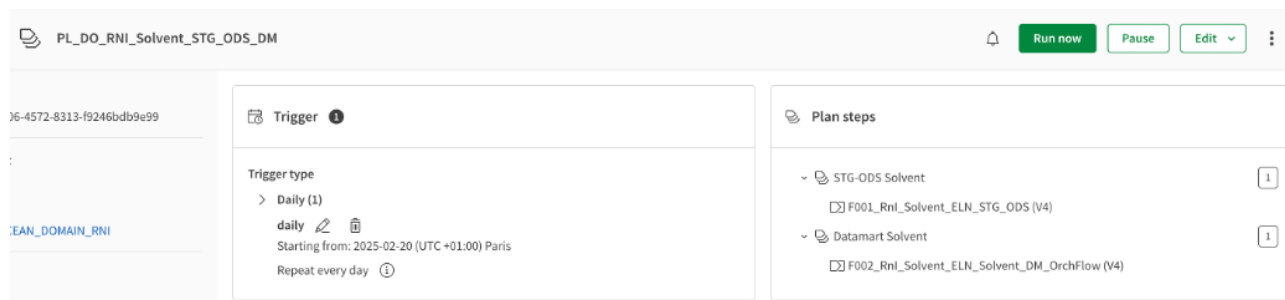
Useful links:

- ELN Solvent folder list : \\ASEW1PTLNDENG01.oxo.priv\data\PROD\DATA_OCEAN_DOMAIN_RNI\Input\ELNsubFolder.xlsx
- Data model ODS: [Luci - Solvent ODS](#)
- Data model DM & facts: [Luci - Solvent DM](#)
- Mapping doc (DMs & facts): [Luci - Solvent Mapping \(DM & Facts\)](#)

Layer 1: Ingestion & Processing

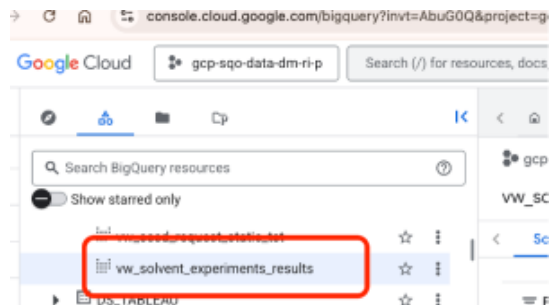
We run a daily scheduled pipeline that pulls data from the ELN and loads it through the usual Staging , ODS, DM flow. If anything comes in with a weird format or breaks validation rules, the pipeline sends error notification emails to the stakeholders.

You can find the full setup in **TMC Project**: [WS_DATA_OCEAN_DOMAIN_RNI](#).



Layer 2: Solvent SQL View

On top of the ODS tables, we built a second layer SQL view that consolidates all solvent-related information. **Sydata**lab is plugged to this [view](#).



1. Purpose

The goal of the view is to stitch together several ODS tables and output a clean consolidated dataset describing solvent experiments.

It combines:

- Experiment metadata (project, product, operator, date, setup fields)
- Ingredient & solute details (substance, batch, CAS, SMILES, concentration)
- Formulation information (solute weights, solvent volumes)
- Solvent characteristics (inventory lookups, blend values like dD/dP/dH, TEB, etc.)
- Test conditions & results (score, temperature, duration, observations)
- HSPIP-style outputs (D, P, H, Tot, Fit, Inside/Outside, etc.)

The final output is essentially: **one row per experiment x solute x test group**, enriched with everything above.

2. Source Tables

All tables live in:

Project: gcp-sqo-data-dm-ri-p

Dataset: ODS

- ODS_SOLVENT_0000_F001_F_D_datalab_experiment
- ODS_SOLVENT_0000_F001_F_D_datalab_Ingredients
- ODS_SOLVENT_0000_F001_F_D_Formulation
- ODS_SOLVENT_0000_F001_F_D_datalab_Solvent_inventory
- ODS_SOLVENT_0000_F001_F_D_datalab_Result
- ODS_SOLVENT_0000_F001_F_D_Test_groups
- ODS_SOLVENT_0000_F001_F_D_datalab_Tests
- ODS_SOLVENT_0000_F001_F_D_datalab_Sample
- ODS_SOLVENT_0000_F001_F_D_datalab_Formulation_batch

3. Output Structure (High-Level)

A few examples of how columns map to the logic:

- **Experiment fields:** EXPERIMENT_ID, PROJECT_, DATE_, OPERATOR, setup flags from SB_0010
- **Solute/ingredient fields:** name, batch, CAS, SMILES, concentration SB_0020
- **Formulation:** solute weight + unit, solvent volume + unit SB_0030
- **Solvent properties:** CAS, SMILES, formula, H-values, TEB, MPT, GHS, etc. SB_0040
- **Condition/results:** duration, temperature, score, observation SB_0050
- **HSPIP metrics:** D, P, H, Tot, R, Fit, Inside/Outside... SB_0053

The CTE names above match those used in the script.

4. Breakdown of Each CTE

SB_0010_SOLVENT_EXPERIMENTS

Pulls the latest metadata per experiment using a ROW_NUMBER() over meta executions.

All the usual experiment attributes are pivoted into columns.

Output: one row per experiment.

SB_0020_SOLVENT_INGREDIENTS

Extracts ingredient-level details: substance, batch, CAS, SMILES, concentration.

Filters out empty ingredients.

Output: one row per ingredient per experiment (mainly solutes).

SB_0030_SOLVENT_FORMULATIONS

Builds formulation fields by pivoting amounts, units, and volumes.

Identifies solute vs. solvent based on Ingredient_ID.

Output: one row per experiment with solute weight + solvent volume info.

SB_0040_SOLVENT_CHARACTERISTICS

Matches ingredient names to solvent inventory entries.

Includes logic to handle synonym lists (split + UNNEST).

Blend solvents get their dD/dP/dH values extracted separately.

Also ensures we only take the latest solvent inventory metadata.

Output: enriched solvent properties.

SB_0050_SOLVENT_CONDITION_RESULT_V2

Combines test-group results and test conditions:

- Pulls the latest meta execution for result tables
- Pivots result values (Score, Observation)
- Links Samples Batches Ingredients
- Combines with test-conditions (Temperature, Duration)

Output: one row per experiment x test group with score/conditions.

SB_0053_SOLVENT_RESULTS_RS_HSPIP

Extracts HSPIP-type test groups only.

Two pivots: one for conditions (temperature, inside/outside, fit type), one for results (D, D, P, etc.).