

SEC Financial Statement Reconstruction Engine

Technical Handoff & System Specification Document

Purpose: This document defines the architecture, processing logic, validation standards, and scalability requirements for building a reusable SEC financial statement reconstruction engine using the official SEC Financial Statement Data Sets (XBRL structured data).

1. Executive Objective

Develop a production-grade, reusable system that:

- Ingests structured SEC Financial Statement Data Sets (TSV format)
- Reconstructs complete financial statements exactly as presented in SEC filings
- Is first implemented for a single filing (one quarter)
- Scales to additional filings without structural code changes
- Avoids hardcoded, company-specific logic

2. Official Data Sources

Primary Data Repository:

<https://www.sec.gov/dera/data/financial-statement-data-sets>

Technical Documentation:

<https://www.sec.gov/files/aqfsn.pdf>

Each quarterly archive contains structured TSV files used for reconstruction.

3. Dataset Components

Each quarterly zip includes:

sub.txt – Filing metadata (CIK, ADSH, form, filing date, period)
pre.txt – Presentation structure (line order, hierarchy, indentation)
num.txt – Numeric fact values
tag.txt – Tag definitions and metadata

These files collectively define structure, values, and context.

4. Initial Build Scope (Strict Constraint)

The first implementation must:

- Target ONE filing only (single ADSH)
- Avoid mixing years or periods
- Avoid combining multiple filings
- Serve as the reusable base template

5. System Architecture – Processing Flow

Step 1: Identify Filing

- Filter sub.txt by CIK and form type (10-Q / 10-K)
- Extract ADSH and reporting period

Step 2: Extract Structure

- Filter pre.txt by ADSH and statement type (BS, IS, CF, EQ, CI)
- Preserve presentation order and indentation

Step 3: Extract Numeric Facts

- Filter num.txt using ADSH, tag, version, ddate, qtrs
- Select correct reporting period values

Step 4: Map Structure to Values

- Inject numeric values into ordered structure
- Preserve blank rows where applicable
- Do not reorder rows

6. Statement Coverage

The engine must equally support:

- Balance Sheet (BS)
- Income Statement (IS)
- Cash Flow (CF)
- Statement of Equity (EQ)
- Comprehensive Income (CI)

No statement-specific hardcoding permitted.

7. Sign & Formatting Logic

- Apply rule-based financial sign handling
- Display negative values using parentheses
- Handle cash outflows, dividends, repurchases appropriately
- Avoid blind sign inversion

8. Structural Integrity Requirements

The final output must:

- Preserve line order from pre.txt
- Maintain indentation hierarchy
- Retain section headers
- Preserve subtotal and total placement
- Handle repeated tags appearing in multiple sections

9. Output Specification

Output file: SEC_FULL_STRUCTURED.xlsx

Separate sheets:

- BS
- IS
- CF
- EQ
- CI

Each sheet must align numerically and structurally with the SEC filing.

10. Validation Standards

Numerical Validation:

- All values present
- No duplicates
- No missing subtotals
- Correct reporting period

Structural Validation:

- Identical line ordering
- Accurate indentation
- Headers preserved
- Totals correctly positioned

11. Scalability Requirement

The same engine must execute successfully for:

- Different quarters
- Different years
- Different companies

Without rewriting structural logic.

12. Final Deliverable Goal

Deliver a reusable SEC financial statement reconstruction engine driven exclusively by structured dataset logic, capable of automated multi-filing execution after initial validation.