Technical Specification
EDI Analysis Unique Configurations
Version 0.3 • Status: Approved • Date: April 2nd 2024

Document Control

Field	Value
Author	Greg Tolman
Reviewers	Kandy R and Alicia L
Approvers	Jack F
Related Documents	Sample EDI docs

1. Purpose and Scope

1.1 Purpose

This specification defines the design for an EDI Analysis of unique document values among all trading partners. The report processes a configured range of historical EDI documents and produces an inventory of envelope and transaction-set characteristics.

The report is needed to support configuration of the new SAP OMS and middleware systems.

1.2 In Scope

- Parsing of ANSI X12 EDI documents at the ISA, GS, ST, and N1 loop levels.
- Aggregation of envelope metadata into a grouped inventory record set.
- Capture of distinct N1-segment usage per group.
- Production of a report file in .CSV format.

1.3 Out of Scope

- EDI compliance validation (e.g., 997/999 acknowledgement generation).
- Transformation of EDI to canonical or downstream formats.
- Long-term storage or indexing of parsed EDI content beyond report generation.
- Real-time or streaming processing; this is a batch report.

2. Definitions and References

2.1 Terminology

Term	Definition
ISA	Interchange Control Header segment; the outermost X12 envelope.
GS	Functional Group Header segment; nested inside ISA.
ST	Transaction Set Header segment; nested inside GS. Identifies the transaction set type (e.g., 850, 856).
N1 loop	Name segment loop containing party identification (N101 = entity identifier code, N102 = name, N103/N104 = identifier qualifier and ID).
Sending System	Logical name of the upstream system that produced the EDI file. Derived via filename (see 4.2).
Region	Logical region/business unit attribution. Derived via filename (see 4.2).
Group	A unique combination of the grouping-key fields defined in 5.2; one output row per group.

2.2 References

- ANSI ASC X12 standard (envelope structure: ISA/IEA, GS/GE, ST/SE).
- Trading partner configuration requirements.

3. Inputs

3.1 EDI Source Files

Attribute	Specification
Format	ANSI X12 EDI
Source location	Archive file server
File naming	Filename is used to determine system and region.
Date range	Process all EDI documents received between January - March 2024
Encoding	ASCII / UTF-8. Files with other encodings converted by code and if still error out will be put in an error list for research and resolution.
Volume (estimated)	Approximately 500,000 documents

3.2 Runtime Parameters

Parameter	Type	Required	Description
start_date	date	Yes	Inclusive lower bound for file selection.
end_date	date	Yes	Inclusive upper bound for file selection.
source_path	string	Yes	Root location of EDI files to scan.
output_path	string	Yes	Destination for the generated report.
output_format	enum	Yes	Report will be in .csv format

4. Derived Fields

Three fields are not generated from the EDI document.

4.1 Sending System

The Sending System is resolved from the filename.

4.2 Region

Region is a property of the Sending System, not from the EDI document and is captured from the filename.

4.3 Count

Count is generated from the number of files that match all other field values except System, Region, N1 Types and N1 Segments.

5. Processing Logic

5.1 High-Level Flow

1. Enumerate candidate files in source_path that fall within the configured date range.
2. For each file: read the ISA segment to discover delimiters, then parse all envelope, group, transaction set, and N1 segments.
3. Aggregate all contributions by grouping key (5.2).
4. N1 fields are treated differently, they are concatenated to the record when all other values match.
5. Generate a log noting details of any file skipped.

5.2 Grouping Key

One output row is produced per unique combination of the following fields:

#	Field	Source	Notes
1	Transaction Set	ST01	Three-character code e.g., 850, 856, 810
2	ISA Sender Qualifier	ISA05	Two-character qualifier code
3	ISA Sender ID	ISA06	15 characters
4	ISA Receiver Qualifier	ISA07	Two-character qualifier code
5	ISA Receiver ID	ISA08	15 characters
6	ISA Element Separator	ISA11	1 character
7	ISA Version	ISA12	e.g., 00401, 00501
8	GS Sender ID	GS02	Alphanumeric
9	GS Receiver ID	GS03	Alphanumeric

#	Field	Source	Notes
10	GS Version	GS08	e.g., 004010, 005010UCS

5.3 Output

Attribute	Specification
Default format	CSV with UTF-8 encoding, header row included.
File name	edi_report_unique_configurations.csv
Sort order	Sending System, Region, Count, Transaction Set, ISA Sender Qualifier, ISA Sender ID, ISA Receiver Qualifier, ISA Receiver ID, ISA Separator, ISA Version, GS Sender ID, GS Receiver ID, GS Version, Transaction Set, N1 Types, N1 Segments. Transaction Set is output in two separate columns at the request of the team, one near the beginning of the record and again near the end.
Empty result	If no files are found in range, an empty file with header row is produced and a warning is logged.

6. Error Handling and Logging

- **File-level errors.** Files that cannot be opened, are empty, or do not begin with a valid ISA segment are logged and moved to a directory for later research and resolution; processing continues with remaining files.
- **Envelope-level errors.** Malformed envelopes (e.g., missing IEA, mismatched control numbers) are logged with file name and ISA control number. Partial contents already parsed and loaded to the database will simply save some compute later.
- **Run summary.** Each run produces a log containing: files scanned, files processed, files skipped, total ST count and run duration.

7. Non-Functional Requirements

Category	Requirement
Performance	Process files from January - March 2024 (3.1).
Idempotency	Re-running with the same parameters and the same input set produces a byte-identical output (except log timestamp).
Determinism	Output row order is fully determined by the output script but can be sorted easily.
Memory	The source process files will be batched.
Observability	Processing logs written to allow for easy monitoring.
Security	EDI files may contain sensitive information. Output is restricted to envelope, set and N1 metadata; other content is not retained.

8. Acceptance Criteria

1. Given a known fixture set of EDI files spanning multiple senders, receivers, ISA versions, GS versions, and transaction sets, the report produces exactly one row per unique grouping key as defined in 5.2.
2. transaction_count for each group equals the number of ST segments in the fixture set for that group.
3. Every unique N1 type and N1 segment string appears exactly once in the output for its group.
4. Files with malformed envelopes are skipped with a logged reason and do not abort the run.
5. Re-running with identical parameters and inputs produces an output file byte-identical to the prior run (excluding file timestamps).