

COMPLIANCE WORKING RECORD

Dataset sampling utility

Fictional sample. Requires tailoring and responsible review.

Document control

Field	Value
Document type	Dataset sampling utility
Document reference	39-dataset-sampling-utility
Jurisdiction	United States
Version	2026-10-08.dataset-utility.1
Status	Dataset review prepared
Prepared by	Customer prepared
Prepared on	October 8, 2026
Generated	October 8, 2026

Purpose and boundary

Select a reproducible sample of distinct source positions using an explicit method, seed and sample size.

Mechanical checks use customer-selected rules and exact source values. They do not authenticate records, establish a complete population, validate an identifier registry, approve an import, or determine compliance. Original files remain unchanged in this browser.

Method and exact sources

Method revision 2026-10-08.dataset-utility.1.

SHA-256 ranked sampling orders a digest for every source position and selects the lowest ranks. Systematic sampling uses $\text{floor}((\text{offset} + k) * \text{population} / \text{sampleSize})$. Source order and exact values are part of reproducibility. No representativeness or confidence claim.

Mechanical checks use customer-selected rules and exact source values. They do not authenticate records, establish a complete population, validate an identifier registry, approve an import, or determine compliance. Original files remain unchanged in this browser.

Source	Filename	Bytes	SHA-256
primary	fictional-primary.json	194	42b9af1571336e4877a7378fdd46a49e5ecaaf05e1163c5e31b6eec0a2f78dff

Frozen inputs

Field	Value
Sampling method	SHA-256 ranked
Sample size	2
Sample seed	Fictional seed 1

Review result

3 source-linked review rows. Selected: 2; Not selected: 1.

- Method revision 2026-10-08.dataset-utility.1.

- SHA-256 ranked sampling orders a digest for every source position and selects the lowest ranks. Systematic sampling uses $\text{floor}((\text{offset} + k) * \text{population} / \text{sampleSize})$. Source order and exact values are part of reproducibility. No representativeness or confidence claim.

- Selected 2 distinct source positions from 3.

Measure	Value
Review rows	3
Selected	2
Not selected	1

Source-linked checks

The formal report shows at most the first 100 review rows and 300 characters per displayed field. Complete values, result columns, sources and output are retained in the Excel and JSON files.

Reference	Result	Finding
2	Selected	SHA-256 ranked. Seed preserved in inputs.

Reference	Result	Finding
3	Not selected	SHA-256 ranked. Seed preserved in inputs.
4	Selected	SHA-256 ranked. Seed preserved in inputs.

Unresolved items and limitations

- Mechanical checks use customer-selected rules and exact source values. They do not authenticate records, establish a complete population, validate an identifier registry, approve an import, or determine compliance. Original files remain unchanged in this browser.

Limitations

- Mechanical checks use customer-selected rules and exact source values. They do not authenticate records, establish a complete population, validate an identifier registry, approve an import, or determine compliance. Original files remain unchanged in this browser.

Official sources

Source	Provision	Date	Link
RFC Editor: Common Format and MIME Type for Comma-Separated Values (CSV) Files	CSV record, field and quoting conventions. This source does not specify customer schemas, matching rules, checksum arithmetic, sampling quality or import acceptance.	October 8, 2026	https://www.rfc-editor.org/rfc/rfc4180
National Institute of Standards and Technology: Secure Hash Standard	SHA-256 digest specification for exact source-byte and delivered-file checks. It does not establish authenticity or validate the named review rules.	October 8, 2026	https://csrc.nist.gov/pubs/fips/180-4/upd1/final