



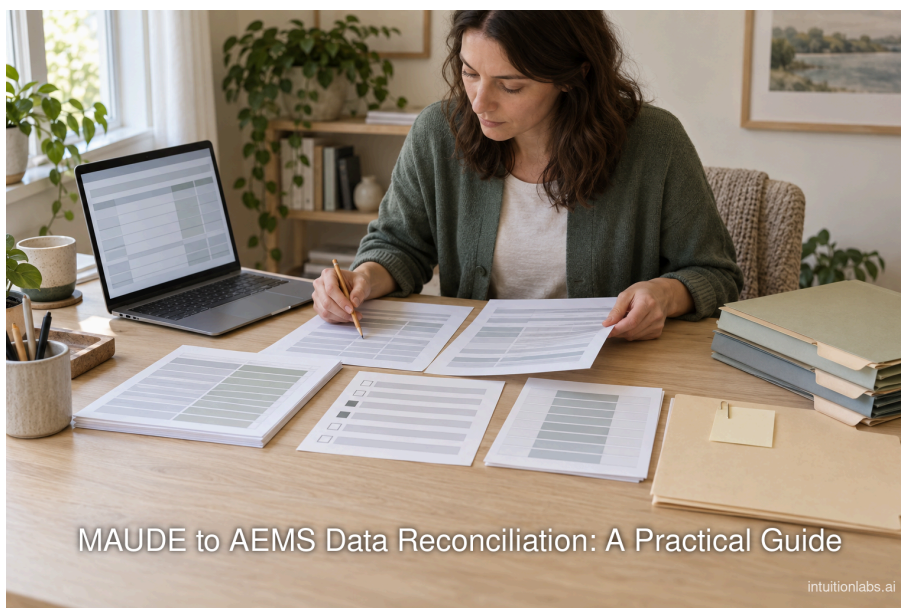
INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

MAUDE to AEMS Data Reconciliation: A Practical Guide

Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-10-13 · maude to aems data reconciliation · maude · aems · openfda · medical device reporting · postmarket surveillance · data reconciliation · medical device data quality · report identifiers

Original article: <https://intuitionlabs.ai/articles/maude-to-aems-data-reconciliation-guide>



In this article

1. Executive Summary
 2. Introduction and Background
 3. Public Sources and the AEMS Transition
 4. MDR File Families, Keys and Update Semantics
 5. Field Mapping and Device Normalization
 6. Duplicate Reports, Supplements and Analytical Limits
 7. Parsing and Operational Validation
 8. Implementation and Migration Validation
 9. Data Analysis and Evidence
 10. Implications and Future Directions
 11. Conclusion
-

Executive Summary

MAUDE to AEMS data reconciliation concerns the **Manufacturer and User Facility Device Experience (MAUDE)** database and **Adverse Event Monitoring System (AEMS)**. Its objective is controlled preservation of report identity, content, and analytical lineage. The **US Food and Drug Administration (FDA)** launched AEMS on **March 11, 2026**, with an announced end-of-May coverage target. That timetable does not independently establish completed device migration or stable device-export semantics. The finding as of October 13, 2026 is conditional: a device-specific AEMS key, export, and refresh contract was not verified from the public documentation reviewed for this guide. (www.fda.gov)

The legacy feeds require different reconciliation assumptions. The **Medical Device Reporting (MDR) report key** links MAUDE master and child records, while master records distinguish reporting sources rather than necessarily unique occurrences. (www.fda.gov) The downloaded-file documentation describes monthly updates, whereas openFDA documents weekly updates. Align the cohort and source cutoff before comparing results. (www.fda.gov) (open.fda.gov) The guide provides a source-access matrix, a conditional field crosswalk, and a control matrix covering report identities, child relationships, rejected rows, unique device identifier (UDI) availability, content changes, and date lag. Its controls are proposed implementation practices, with independently documented database and provenance mechanisms as supporting references.

The observed MDR manifest lists **1,406,457 current-year master records** and **2,835,147 narrative records**: their calculated ratio is approximately **2.016**, illustrating different grains rather than a safety rate. These are published counts, not independently parsed archives. (www.fda.gov) Historical evidence also argues for stronger reproducibility records: one study found **14 reproducible queries among 60 executable queries**, a calculated **23.3%** under its count-based definition. Matching counts does not establish matching content. (digitalcommons.library.tmc.edu) (pmc.ncbi.nlm.nih.gov)

The recommended decision is to retain an immutable baseline and observation history, apply source-specific update rules, preserve original names and identifiers alongside curated mappings, and test a frozen cohort before switching analytical inputs. The **World Wide Web Consortium (W3C)** provenance framework provides a way to describe the producing inputs and processing steps. (www.w3.org) Most importantly, reconciliation

cannot recover missing exposure denominators or turn public reports into incidence, prevalence, causality, or comparative safety rates. FDA identifies missing device-use information among these limits. The output should support descriptive trends, qualitative signal review, and traceable case assessment within a wider surveillance program. (www.fda.gov)

Introduction and Background

MAUDE to AEMS data reconciliation means preserving the identity, history, and analytical meaning of public medical-device reports when those reports are accessed through different FDA data products. It is an evidence-preservation problem: which source record was used, what transformation occurred, and which version supported a conclusion? The World Wide Web Consortium (**W3C**) defines provenance around the entities, activities, and people producing information. (www.w3.org) (www.w3.org)

The **Manufacturer and User Facility Device Experience (MAUDE)** database and **Adverse Event Monitoring System (AEMS)** sit within a broader surveillance process. The World Health Organization (**WHO**) distinguishes manufacturer **postmarket surveillance (PMS)** from regulator market surveillance. Public reports are therefore one input into a safety program, alongside information collected through other channels. A successful public-data migration preserves that input's meaning without claiming to reproduce the entire surveillance process. (www.who.int) (www.who.int)

This report addresses safety and vigilance leaders, quality teams, analysts, and engineers who need defensible trend continuity. Its engineering perspective is adjacent to the data services IntuitionLabs describes, including integration and warehousing. The recommendations below are analytical controls, rather than claims that a consultancy operates an FDA reporting platform. (intuitionlabs.ai) The broader need for machine-readable information also appears in the Reagan-Udall Foundation's discussion of standardizing FDA information for reuse. (reaganudall.org)

The central distinction is between **source fidelity**, **semantic equivalence**, and **analytical validity**. Matching imported bytes establishes source fidelity. Matching report keys, dates, and nested records tests semantic equivalence. Neither establishes that a submitted description accurately represents an independently verified occurrence. The National Institute of Standards and Technology (NIST) digest guidance supports the first task; the UK Government Data Quality Framework distinguishes completeness from accuracy for the last. (csrc.nist.gov) (www.gov.uk)

Accordingly, the guide separates verified public documentation from proposed implementation rules. Its internal data model is deliberately source-independent. W3C provenance and quality vocabularies supply useful design patterns, but do not establish FDA requirements or prove that AEMS exports use those structures. (www.w3.org) (www.w3.org)

Public Sources and the AEMS Transition

What changed, and what remains unverified

FDA announced AEMS on **March 11, 2026** and described broad consolidation, with an end-of-May coverage target. A launch timetable is evidence of an announced plan, rather than proof that every device record and export capability subsequently became available. The separately maintained MDR files page still describes MAUDE data becoming available through AEMS during the year. (www.fda.gov) (www.fda.gov)

Research limitation as of October 13, 2026: this guide reviewed FDA's AEMS launch announcement and overview, MDR downloadable-file documentation, the public dashboard FAQ, and the openFDA device/event overview. That documentation did not establish a stable device-specific AEMS export schema, documented MDR-key equivalence, device API contract, or device refresh guarantee. Device records and exports in the live dashboard could not be independently inspected. These capabilities remain unverified in this guide; this limitation does not establish that they are absent.

The published dashboard FAQ addresses drug and therapeutic-biologic surveillance. Its daily-update and limited-export descriptions must retain that scope. They cannot supply missing medical-device semantics. (fis.fda.gov) (fis.fda.gov)

Table 1 distinguishes the access products and the evidence available for reconciliation. The MAUDE and openFDA descriptions document different refresh schedules, so identical extraction times do not imply identical content. (www.fda.gov) (open.fda.gov)

Public source	Access and documented timing	Reconciliation use and boundary
MAUDE interface	Search interface; last ten years ; monthly publication through the previous month. (www.fda.gov)	Case review and a check against the displayed current report. A displayed result is not a preserved historical snapshot.
MDR downloadable files	Pipe-delimited ZIP families; monthly updates.	Build a reproducible relational snapshot with separately retained master and child records.
openFDA device/event	JavaScript Object Notation (JSON) application programming interface (API) and bulk data; weekly updates.	Compare report keys and nested content after aligning source windows. Bulk files use the API's JSON representation. (open.fda.gov)
AEMS public dashboard	General consolidation scope includes devices; device-specific export and timing contract remains unverified here. (www.fda.gov)	Validate the actual device view and accessible export before adopting it as an interchangeable feed.

As of October 13, 2026, the openFDA overview displayed an incomplete coverage end date, so this guide does not assert a precise endpoint date range. Establish the earliest and latest relevant dates from the retained extract. (open.fda.gov)

The practical consequence is to define a **comparison cutoff** separately from extraction time. A later extraction can still contain an earlier source snapshot. Record the source's declared update or export date, the report-date filter, and the local retrieval instant as different attributes. openFDA defines its export date as file export

time; Request for Comments (RFC) 3339 timestamps represent instants rather than intervals. (open.fda.gov) (www.rfc-editor.org)

For each feed, establish the following acceptance conditions:

- **Scope:** Record which product category and historical window the source actually exposes.
- **Access:** Distinguish an interactive search, a limited export, an API response, and a complete bulk snapshot.
- **Timing:** Preserve extraction time with an explicit time-zone offset. (www.rfc-editor.org)
- **Coverage:** Test the chosen cohort using source report keys, rather than a displayed headline total.
- **Contract:** Record undocumented fields as unknown until a device-scoped definition and sample support the mapping.

These are proposed controls. They make the evidence needed for a switch explicit and align with W3C's distinction between a quality dimension and the procedure used to measure it. (www.w3.org) (www.w3.org)

MDR File Families, Keys and Update Semantics

Preserve the report grain and the child grains

Medical Device Reporting (MDR) is the reporting context, while **MDR REPORT KEY** is the linking identifier. FDA documents a separate master record for each reporting source and links record types through that key. Thus, a master record should not automatically be labeled a unique real-world incident. (www.fda.gov) (www.fda.gov)

A peer-reviewed MAUDE file primer describes the relational linkage and separate child identities. Verify these candidate keys against downloaded headers and uniqueness tests. (up-j-gemgem.ubiquityjournal.website) Database constraints can then enforce that contract: a primary key requires uniqueness and non-null values, while a foreign key checks that a referenced row exists. (www.postgresql.org) (www.postgresql.org)

Table 2 is a proposed reconciliation crosswalk. The legacy record relationships draw on FDA's linking documentation and the published primer. The openFDA field explorer documents the JSON representation; AEMS mappings remain expressly conditional. (up-j-gemgem.ubiquityjournal.website) (open.fda.gov)

Legacy family and grain	Identity or relationship to validate	openFDA comparison target	AEMS device mapping status
Master event: reporting-source record	MDR REPORT KEY; retain report number separately	mdr_report_key, report_number, report-level dates and flags	Key equivalence and version semantics unverified
Device: device child	Report key plus device sequence; retain device event key	device[], including source names and identifiers	Child cardinality and exported identity unverified
Patient: patient child	Report key plus patient sequence	patient[]; preserve structured child content	Public patient grain unverified

Legacy family and grain	Identity or relationship to validate	openFDA comparison target	AEMS device mapping status
Text: narrative child	MDR TEXT KEY; retain report, patient, and text-type linkage	mdr_text[]; compare type and text as well as count	Narrative identity and redaction semantics unverified
Device problems: report-code associations	Report key plus problem code, after checking repetitions	product_problems[]; verify raw-code versus description semantics	Dictionary and association mapping unverified
Patient problems: report-code associations	Report key plus patient-problem code; preserve dictionary version	Compare source patient-problem content without assuming labels are keys	Dictionary and association mapping unverified

This table is not a promise of interchangeable rows. Flattening all child arrays together can multiply the report record. The safer design stores children separately and aggregates them to the chosen analytical grain before joining summary measures. pandas provides join validation, whereas permitting many-to-many joins does not itself check uniqueness. (pandas.pydata.org) (pandas.pydata.org)

Base, add, change, current-year, and history

Treat **history** and **current-year** files as snapshot components, **add** files as incoming records, and **change** files as updates whose effect must be determined at the correct grain. Device changes can include additional children for existing masters. Consequently, a changed child record need not represent either a new report or an entirely replaced report. (www.fda.gov)

FDA's published openFDA pipeline at revision 2c152208aaa02323696d904e94a1d114e66247ec, inspected on October 13, 2026, applies precedence in the order **changed, added, existing**. It also contains specific narrative-preservation logic when changed records have empty narrative collections. This is evidence about that implementation, not a universal rule that every blank field means "retain the old value." (github.com) (github.com)

A defensible merge design follows these proposed rules:

- **Baseline:** Load compatible historical and current snapshots, retaining source file and batch identity.
- **Adds:** Compare incoming report keys against existing keys before classifying records as new.
- **Changes:** Resolve the affected master or child identity and preserve the preceding observation.
- **Precedence:** Apply a documented source-specific rule, with the FDA pipeline serving as a reference implementation.
- **Missing values:** Distinguish absent, empty, redacted, and explicitly supplied values before merging.
- **Idempotence:** Replaying the same batch should produce the same materialized state.
- **Deletion:** Treat a missing key as an exception pending scope and completeness checks, rather than assuming a deletion.

Field Mapping and Device Normalization

Build a crosswalk that preserves evidence

A useful **field crosswalk** records source name, data type, cardinality, business meaning, transformation, and verification status. A same-looking label is insufficient: a report identifier, a device identifier, and a device-model version answer different questions. The Fast Healthcare Interoperability Resources (**FHIR**) Device model published by Health Level Seven International (HL7) explicitly distinguishes a device business version from a resource-record version. It is an internal-design reference here, not an AEMS export specification. (hl7.org) (www.imdrrf.org)

The **unique device identifier (UDI)** requires special care. Under the US definition, its **device identifier (DI)** identifies a version or model and labeler; the **production identifier (PI)** is conditional and variable. A DI therefore cannot automatically identify a particular physical unit. (www.ecfr.gov) (www.ecfr.gov) The International Medical Device Regulators Forum (**IMDRF**) likewise explains that “unique” does not imply serialization of every production unit. (www.imdrrf.org)

Use these proposed normalization rules:

- **Namespace:** Retain identifier value, issuer, and jurisdiction together. FHIR models those attributes separately. (hl7.org) (hl7.org)
- **Capture method:** Where available, preserve whether the UDI was machine-read or entered manually. (hl7.org)
- **Unit identity:** Keep lot, serial, and model information distinct; do not infer unavailable PI values from a DI. (www.ecfr.gov) (hl7.org)
- **Legacy linkage:** Preserve the cross-reference from catalog or legacy identifiers to the UDI. IMDRF discusses the need for these mappings. (www.imdrrf.org)
- **Issuing agency:** Resolve against the appropriate scheme. GS1 describes its UDI issuing-agency role; that does not make every identifier a GS1 identifier. (www.gs1.org)
- **Multiple identifiers:** Allow multiple device identifiers where justified, rather than assuming a universal scalar field. (hl7.org)

Manufacturer names and code dictionaries

Manufacturer-name consolidation can change aggregation while leaving the underlying report unchanged. FDA's Total Product Life Cycle (**TPLC**) documentation describes simplifying and consolidating firm names. Retain the reported name, the source-provided consolidated name, and the internally curated organization identity as separate fields. (www.fda.gov) FHIR's manufacturer description accommodates a name or legally responsible entity, reinforcing the need to document which meaning a normalized field represents. (hl7.org)

Proposed organization-mapping rules should specify:

- **Reported text:** Preserve original spelling before normalizing it.
- **Alias evidence:** Store the source and rationale for each accepted alias.
- **Legal role:** Distinguish manufacturer, labeler, distributor, and reporting party where the source supports it.

- **Time validity:** Give curated mappings effective dates and preserve earlier mapping versions.
- **Review:** Queue ambiguous names rather than merging them on string similarity alone.
- **Trend stability:** Calculate whether a mapping revision moved reports between cohorts.

These rules are analyst recommendations. Their purpose is to preserve the meaning of an entity and the processing that produced its classification, using provenance rather than replacing original text with an unexplained name. (www.w3.org) (www.gov.uk)

Problem-code reconciliation also requires separate **associations** and **dictionaries**. The primer describes reaching device-problem descriptions through the code association file. IMDRF provides downloadable terminology materials; the relevant release should be retained with the report snapshot. (up-j-gemgem.ubiquityjournal.website) (www.imdrf.org) Do not identify a problem solely by its current display label. Preserve source code, coding system, release, and any mapped target code.

Before relying on a problem ZIP filename, inspect its header and actual rows. The FDA file descriptions leave some dictionary-versus-association labels ambiguous. (www.fda.gov) A code-description table and a report-code table need different keys. Enforce the tested relationship using database constraints and retain the terminology mapping as a documented transformation. (www.postgresql.org) (www.w3.org)

Duplicate Reports, Supplements and Analytical Limits

A report version is not a new occurrence

Supplemental reports update information associated with a report. FDA documents an unchanged MDR REPORT KEY after follow-up and display of the most recent information. Preserving the original receipt date alongside observation and supplement dates is therefore necessary for consistent time-series definitions. (www.fda.gov)

Maintain both a **current-state view** and an **observation history**. The current state supports case review; the history explains why a past analysis used different text or codes. Neither should be confused with a reconstructed official revision history when only periodic snapshots were observed. Record exactly which snapshots were captured and which transformations generated the analytical version. (www.w3.org) (hl7.org)

A practical duplicate and supplement decision tree is:

- **Same source key, unchanged content:** Classify as a repeated observation of the retained record.
- **Same source key, changed content:** Classify as a new observation or update of that record.
- **New child identity under an existing key:** Classify as a child addition, subject to source update semantics.
- **Different keys, shared contextual attributes:** Create a possible-related-report group for review.
- **Different keys, similar narrative:** Flag textual similarity without declaring incident identity.
- **Missing or null keys:** Quarantine identity resolution before matching.
- **Summary report:** Keep its summarized-event quantity distinct from its report count.

The decision tree is an implementation proposal, not a validated automated incident-deduplication algorithm. The primer distinguishes report and event identifiers and cautions that the event identifier is not unique across master rows. A shared grouping attribute can guide review without proving that every duplicate or independent occurrence has been identified. ([up-j-gemgem.ubiquityjournal.website](#)) ([www.gov.uk](#))

Allowed descriptive uses and invalid rate calculations

Report counts cannot estimate incidence, prevalence, causality, or comparative device safety by themselves. FDA identifies missing device-use information among the reasons for these limits. An AEMS interface does not supply the missing exposure denominator merely by consolidating reports. ([www.fda.gov](#)) ([www.fda.gov](#))

Useful outputs include descriptive counts under a fixed query, narrative-review queues, coding-completeness measures, and candidate signals requiring contextual assessment. These outputs belong within a wider PMS process. WHO describes surveillance as collecting and evaluating experience, including opportunities to improve a device. ([www.who.int](#)) ([www.who.int](#))

Invalid constructions include “reports divided by database total equals device risk,” “more reports means less safe,” and “a matched narrative proves causation.” Their denominators describe a reporting system, rather than exposure in an appropriately defined population. MHRA guidance separately identifies devices placed on the market or put into service as relevant trend-report information, illustrating the additional quantity required for contextual analysis. ([www.gov.uk](#)) ([www.who.int](#))

Even an external sales or use denominator requires compatible populations, periods, geography, device definitions, and reporting rules. This report does not propose a universal conversion from public reports to risk. Method selection and justification belong to the safety team's analysis plan; MHRA's guidance explicitly addresses justification of chosen trend methods. ([www.gov.uk](#)) ([www.gov.uk](#))

Qualitative signal review remains valid when its output is framed accurately. A reviewer can identify a recurring description for further assessment, examine supplementary context, or check whether an apparent increase coincides with a coding or mapping change. Record the query and each intermediate cohort so another analyst can reproduce the selection. The reproducibility study recommends reporting subquery counts, while the quality framework emphasizes consistency across datasets. ([pmc.ncbi.nlm.nih.gov](#)) ([www.gov.uk](#))

Parsing and Operational Validation

Preserve values and evidence

For each target field, store the original value alongside the normalized value. Record whether a transformation changes punctuation, whitespace, case, coding system, or meaning. Keep unmapped fields accessible rather than silently dropping them. That design makes the processing step and its derivation inspectable, consistent with the provenance concepts in W3C's PROV family of provenance specifications. ([www.w3.org](#)) ([www.w3.org](#))

Treat a missing UDI as missing identification evidence, rather than a negative device match. openFDA explicitly notes incomplete and unvalidated UDI content. Coverage should therefore be measured for the retained cohort, separately from the proportion that can be linked to a reference record. (open.fda.gov) (www.gov.uk)

Persist both the immutable source observation and the current analytical projection. Provenance records can connect a projection to its inputs and processing steps; a digest identifies whether the retained input changed. (www.w3.org) (csrc.nist.gov)

Parsing and schema checks

The FDA page's master-schema heading says **82 fields**, while its enumeration extends through **85**. The correct response is to verify the downloaded header and row width, then version the parser against observed evidence. (www.fda.gov)

Import identifiers as strings and configure delimiter, missing-value handling, and encoding explicitly. DuckDB supports explicit schemas and text-first import, while Arrow exposes delimiter and null-token controls. Arrow's default encoding is a library default, not proof of an FDA archive's encoding. (duckdb.org) (duckdb.org) (arrow.apache.org) (arrow.apache.org) (arrow.apache.org)

Keep rejected lines visible. DuckDB can place erroneous rows in a rejects table, and pandas documents bad-line handling. Neither mechanism makes skipped data acceptable automatically. Reconcile physical input lines, accepted records, documented headers, and quarantined failures before releasing a snapshot. (duckdb.org) (pandas.pydata.org)

Handle missing identities

Nulls deserve an explicit exception path. pandas can match null join keys to other null keys, unlike ordinary Structured Query Language (SQL) semantics. A merge can consequently create apparent matched records without an actual shared identifier. Reject or segregate null keys before overlap analysis; enforce non-null report identity in the accepted projection. (pandas.pydata.org) (www.postgresql.org)

Test behaviors that can change trends

A release candidate should be tested against representative transitions: an added report, changed child, supplemented narrative, renamed manufacturer alias, changed code label, and absent record. These are proposed fixture categories, not reported production incidents. The purpose is to verify the source-specific rules and preserve the transformations that connect input to output. (www.w3.org) (www.imdrf.org)

Schema inference should also be tested against late-file variation. Arrow's incremental reader infers types from its first block, while explicit schemas avoid dependence on an initial sample. Run structural checks before changing the production parser and account for rejected lines after any configuration change. (arrow.apache.org) (duckdb.org) (duckdb.org)

Finally, distinguish dataset observation time from event or receipt dates. Use explicit offsets for local extraction timestamps, preserve date-only values without invented precision, and define every lag by its date pair. RFC 3339 addresses instants and offsets; it does not establish the business meaning of a report's date field.

(www.rfc-editor.org) (www.rfc-editor.org)

Implementation and Migration Validation

Make the comparison contract executable

Define **MAUDE AEMS reporting differences** as measurable differences in access, scope, representation, timing, or revision behavior. Do not make equality of displayed totals the acceptance criterion. W3C's Data Quality Vocabulary treats a metric as an explicit measurement procedure; pandas join indicators can distinguish keys present in either input or both. (www.w3.org) (pandas.pydata.org)

The proposed migration package should contain:

- **Inventory:** Source URLs, product scope, archive or partition names, and declared update dates.
- **Capture:** Retrieval timestamp, original bytes where available, and cryptographic digest. (csrc.nist.gov) (www.rfc-editor.org)
- **Schema:** Observed header, column order, data types, nullable fields, and parser configuration.
- **Lineage:** Source report key, child identity, source batch, transformation version, and output version. (www.w3.org)
- **Exceptions:** Rejected records, missing keys, unmatched codes, and ambiguous organizations.
- **Evidence:** Cohort filters, key-overlap results, field comparisons, reviewer decisions, and release approval.
- **Rollback:** A reproducible preceding snapshot and the configuration needed to rebuild it.

Use a frozen cohort for initial equivalence testing, then a separate moving cohort for refresh monitoring. A frozen cohort isolates representation changes; a moving cohort tests operational timing. This separation is a proposed experiment design, not an FDA threshold. The UK quality framework evaluates timeliness relative to intended use and recommends retaining quality information with archived data. (www.gov.uk) (www.gov.uk)

Reconcile identities before comparing values

First compare key sets, then child identities, then field values. Compute left-only and right-only records explicitly. A matched count cannot reveal whether those matches are the right records, and a near-identical total can conceal compensating additions and omissions. The reproducibility paper itself measures counts rather than report content, an important distinction for migration validation. (pandas.pydata.org) (pmc.ncbi.nlm.nih.gov)

Table 3 defines proposed controls and their interpretation. These measures assess the retained dataset and pipeline; they do not estimate the completeness of all underlying medical-device occurrences. W3C distinguishes the metric, its value, and the dataset or distribution being measured. (www.w3.org)

Control	Proposed measurement	Interpretation and release response
Report identity	Non-null key count; distinct keys; duplicate-key groups	Require an explicit accepted-key contract. Primary-key enforcement supports this check. (www.postgresql.org)

Control	Proposed measurement	Interpretation and release response
Child integrity	Orphan device, patient, text, and problem associations	Explain missing parents and test foreign-key relationships. (www.postgresql.org)
Key overlap	Left-only, right-only, and common key sets	Align scope and source cutoff before interpreting mismatches.
Import completeness	Accepted rows plus rejected rows versus expected input rows	A rejects table supports accounting, rather than silent loss. (duckdb.org)
Field missingness rate	Missing required-field values divided by eligible rows	Lower values indicate fewer missing values. Publish numerator, denominator, cohort, and missing-value rule.
UDI availability	Non-empty DI values divided by eligible device rows	Higher values indicate greater availability. Measure availability separately from validation or reference linkage.
Content changes	Added, changed, unchanged, and absent identities across snapshots	Use a documented digest and comparison rule. (csrc.nist.gov)
Date lag	Defined date differences, with invalid and missing pairs counted	Keep reporting delay separate from feed publication delay.
Mapping stability	Reports entering or leaving a cohort after a mapping revision	Version the mapping and preserve both classifications.
Quality lineage	Metric definition, input snapshot, computation, and result	Derived measurements should retain their derivation. (www.w3.org)

The same dashboard can show technically complete ingestion and unresolved source accuracy. A validly formatted date may still describe the wrong occurrence. A quality score that measures availability does not by itself measure accuracy. Keep **validity**, **completeness**, and **accuracy** as separate status dimensions rather than collapsing them into a universal score. (www.gov.uk) (www.gov.uk)

Data Analysis and Evidence

Published counts demonstrate different grains

The MDR manifest observed on **October 13, 2026** lists **23,636,521** historical master records through the preceding year and **1,406,457** current-year master records. It also lists **261,298** master-add records and **614,484** master-change records. These are published counts, rather than parsed archives. Test their overlap before summation. (www.fda.gov)

The current-year child counts are **1,406,216 devices**, **1,406,954 patients**, and **2,835,147 narratives**. (www.fda.gov) Calculated from those published counts, narrative rows divided by master rows equal approximately **2.016**. This is a storage-grain ratio, rather than an incidence measure, event-to-patient ratio, or proof of duplicate reports. The earlier MAUDE primer likewise described narrative observations exceeding report counts. (www.fda.gov) (up-j-gemgem.ubiquityjournal.website)

openFDA's status page observed for this report showed device/event as operational, with a last-update date of **October 7, 2026**. That date describes the endpoint's reported status. It does not establish its maximum event date, nor a reconciled difference from the legacy manifest. (open.fda.gov) Compare a defined cohort and matched cutoff rather than subtracting unmatched headline totals; the quality framework treats consistency and completeness as separate dimensions. (www.gov.uk) (www.gov.uk)

Historical evidence: reporting units and reproducibility

A peer-reviewed reporting-trends study retained **15,775,718 reports** spanning **1991 through 2022**. Its restricted cohort beginning in **2005** contained **15,182,755 reports**. These are historical study populations, not current platform totals. (pmc.ncbi.nlm.nih.gov) (pmc.ncbi.nlm.nih.gov)

For **2022**, that study described **3,135,501 represented events** from **2,945,026 submissions**. The analyst calculation is **190,475 more represented events**, approximately **6.47%** above the submission count. This illustrates different counting units within that analysis, without establishing unique incidents or a population risk rate. (pmc.ncbi.nlm.nih.gov) Its historical cohort also contained a **96.5%** manufacturer-report share, demonstrating why source mix must accompany descriptive trend interpretation. (pmc.ncbi.nlm.nih.gov)

A separate reproducibility study identified **523 articles**, of which **336** contained queries and **60** had executable queries; **14** of those executable queries reproduced the published counts under the study's definition.

(digitalcommons.library.tmc.edu) (digitalcommons.library.tmc.edu) Calculated as **14 divided by 60**, this is **23.3%**. The study's coefficient interval, **0.95 to 1.05**, is a research definition, rather than a recommended FDA migration tolerance. It compares report counts and does not establish content identity.

(digitalcommons.library.tmc.edu) (digitalcommons.library.tmc.edu) (pmc.ncbi.nlm.nih.gov)

The practical lesson is to retain the query, source snapshot, preprocessing, and intermediate counts together. Count reproducibility is useful evidence, but matching the keys and relevant fields tests a stronger form of equivalence. W3C provenance captures derivation, and its quality vocabulary supports measurements derived from other measurements. (www.w3.org) (www.w3.org)

Compute field missingness, UDI availability, and lag transparently

For a retained cohort, define the **field missingness rate** for each required field as missing values divided by eligible records, and **UDI availability** as non-empty DI values divided by eligible device records. Lower field missingness rates indicate fewer missing values; higher UDI availability indicates greater availability. Declare the missing-value rule first: blank strings, omitted properties, redacted values, and unparseable values may require separate categories. Library controls for null tokens and string preservation support that distinction. (arrow.apache.org) (pandas.pydata.org)

Report **reference linkage** separately from UDI availability. A populated DI, a syntactically acceptable DI, and a verified reference match are different outcomes. The regulatory definition gives DI a model-and-labeler meaning; it does not make a populated field proof of accurate submission. (www.ecfr.gov) (www.gov.uk)

For **date-lag distributions**, distinguish FDA receipt minus awareness, FDA receipt minus event date, and extraction minus declared source-update time. Publish the valid-pair count, missing pairs, negative differences, and chosen quantiles. The reporting-trends study used receipt minus awareness and specified exclusions for

its lag analysis. Its choices are study methods to assess, rather than universal replacements for local business definitions. ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/)) ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/))

Current field missingness rates, UDI availability, deletion counts, and AEMS device lag require retained extracts. Quality measurements should identify their dataset, procedure, and result, and be archived with the evidence. (www.w3.org) (www.gov.uk)

Implications and Future Directions

The key migration decision is whether a verified AEMS device extract can become an analytical input under a documented contract. It should remain a parallel candidate while identity, child cardinality, revision behavior, scope, and timing are unresolved. This is an implementation recommendation informed by IMDRF's emphasis on cardinality and editable-field rules, rather than a claim that FDA requires a particular reconciliation architecture. (www.imdrf.org) (www.imdrf.org)

For software product owners, a source adapter should translate representation while preserving source meaning. A stable internal report model can then carry multiple observations and identifiers without encoding one interface's assumptions throughout the application. HL7's distinction between business and resource versions, and its support for multiple UDI carriers, are useful design references. They establish neither AEMS compatibility nor FDA adoption of FHIR for these exports. (hl7.org) (hl7.org)

For safety teams, a time-series break should be reviewed against source changes before interpreting it as a change in device experience. Candidate causes include scope, refresh timing, terminology release, organization mapping, and chosen counting grain. Preserve both the pre-change and post-change classifications so the effect can be inspected. The quality framework's consistency dimension and W3C's measurement provenance support this evidence-oriented approach. (www.gov.uk) (www.w3.org)

For an adjacent consultancy perspective, IntuitionLabs describes connecting AI to authoritative sources with permissions, citations, and evaluation. Applied here, that principle favors traceable source observations and reviewable transformations. It does not confer privileged FDA access or independent confirmation of device-export behavior. (intuitionlabs.ai) Automated extraction or narrative classification should be evaluated against the defined review task, with the original report and derivation available to the reviewer. (www.w3.org)

The next publication-time verification should seek device-scoped documentation, a readable sample export, identifier definitions, refresh metadata, and evidence of supplement handling. Keep the resulting quality statement attached to the specific source distribution and processing version. W3C's quality vocabulary permits dataset-specific measurements and joint tracking of their provenance. (www.w3.org) (www.w3.org)

A public-data pipeline remains part of a wider safety process. WHO's distinction between manufacturer and regulator surveillance helps keep responsibilities clear: a consolidated interface improves access only to the information it actually exposes. It does not eliminate the need for device expertise, contextual evidence, and a justified analysis plan. (www.who.int) (www.gov.uk)

Conclusion

MAUDE to AEMS reconciliation should preserve report identity, child-record structure, observed revisions, source timing, and the transformations behind every analytical cohort. The appropriate acceptance evidence combines keys and content, rather than relying on similar totals. Provenance describes the path from input to output, while quality metrics document what was tested on the retained distribution. (www.w3.org) (www.w3.org)

The practical sequence is to retain a baseline, verify the parser and grain, apply source-specific update rules, compare a frozen cohort, and document exceptions before changing the production source. Names, problem codes, and UDI values should retain their original form alongside curated mappings. Tested database constraints and explicit join validation help prevent identity assumptions from becoming silent counting errors. (www.postgresql.org) (pandas.pydata.org)

AEMS device-specific field and refresh semantics remain publication-time verification work in this report. Unknowns should stay visible in the crosswalk until device-scoped documentation and accessible samples support them. A migration can be useful without prematurely claiming equivalence.

Finally, defensible lineage does not convert passive reports into incidence, prevalence, causal findings, or comparative safety rates. Public reporting remains one input into surveillance and professional case assessment. Preserve what the data can support, identify what it cannot establish, and keep the wider evidence and method visible to the safety team. (www.fda.gov) (www.who.int)

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. <https://www.fda.gov/news-events/press-announcements/fda-launches-new-adverse-event-look-tool>
2. <https://www.fda.gov/medical-devices/medical-device-reporting-mdr-how-report-medical-device-problems/mdr-data-files>
3. <https://open.fda.gov/apis/device/event/>
4. https://digitalcommons.library.tmc.edu/uthshis_docs/723/
5. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11636620/>
6. <https://www.w3.org/TR/prov-overview/>
7. <https://www.fda.gov/medical-devices/mandatory-reporting-requirements-manufacturers-importers-and-device-user-facilities/about-manufacturer-and-user-facility-device-experience-maude-database>
8. <https://www.who.int/publications/i/item/9789240015319>
9. <https://intuitionlabs.ai/>
10. <https://reaganudall.org/sites/default/files/2024-07/Improving-Access-to-Publicly-Available-FDA-Information-Final.pdf>
11. <https://csrc.nist.gov/pubs/fips/180-4/upd1/final>
12. <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>
13. <https://www.w3.org/TR/vocab-dqv/>
14. <https://fis.fda.gov/extensions/FPD-FAQ/FPD-FAQ.html>
15. <https://open.fda.gov/apis/downloads/>

16. <https://www.fda.gov/safety/fda-adverse-event-monitoring-system-aems>
17. <https://open.fda.gov/apis/download-api-fields/>
18. <https://www.rfc-editor.org/rfc/rfc3339.html>
19. <https://up-j-gemgem.ubiquityjournal.website/articles/10.5334/egems.221>
20. <https://www.postgresql.org/docs/current/ddl-constraints.html>
21. <https://open.fda.gov/apis/device/event/searchable-fields/>
22. <https://pandas.pydata.org/docs/reference/api/pandas.merge.html>
23. <https://github.com/FDA/openfda/blob/2c152208aaa02323696d904e94a1d114e66247ec/openfda/maude/pipeline.py>
24. <https://hl7.org/fhir/R5/device-definitions.html>
25. <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-190321-udi-sag.pdf>
26. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-830/subpart-A/section-830.3>
27. <https://www.gs1.org/industries/healthcare/udi>
28. <https://www.fda.gov/about-fda/cdrh-transparency/total-product-life-cycle-tplc-data-sources-and-disclaimers>
29. <https://www.imdrf.org/working-groups/adverse-event-terminology/imdrf-adverse-event-terminology-web-browser>
30. <https://www.gov.uk/government/publications/medical-devices-post-market-surveillance-requirements/vigilance-reporting-requirements>
31. <https://duckdb.org/docs/current/data/csv/overview>
32. <https://arrow.apache.org/docs/python/csv.html>
33. https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html
34. <https://open.fda.gov/about/status/>
35. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11755539/>

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore [AI enablement](#), [pharma and biotech software development](#), [data engineering and BI](#), and [Veeva services](#). [Contact IntuitionLabs](#) to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: <https://intuitionlabs.ai>

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.