



INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Orange Book RLD, Reference Standard and TE Code Crosswalk

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 · 2026-10-22 · orange book · reference listed drug · reference standard · therapeutic equivalence · te codes · anda · openfda · pharma data architecture

Original article: <https://intuitionlabs.ai/articles/orange-book-rl-d-reference-standard-te-codes>



In this article

1. [Executive Summary](#)
 2. [Introduction and Background](#)
 3. [Key Changes](#)
 4. [Product Identity and the Reference-Selection Model](#)
 5. [Implementation Considerations and Process Changes](#)
 6. [TE Codes, Product Headings and Group Membership](#)
 7. [Reference Schema and Worked Join Tests](#)
 8. [Refresh, Exception Control and Evidence Limits](#)
 9. [Data Analysis and Evidence](#)
 10. [Implications and Future Directions](#)
 11. [Conclusion](#)
-

Executive Summary

The **reference listed drug (RLD)** is the product relied upon for an abbreviated new drug application (ANDA). The **reference standard (RS)** is FDA's selected comparator for required in vivo bioequivalence studies. These are different regulatory roles, even when the same product fills both. A **therapeutic equivalence (TE) code** describes another relationship, interpreted within the relevant product context. An ingredient match cannot replace reference selection or establish an equivalence group. ([ecfr.gov](#)) Preserve complete **AB1/AB2** groups and multiple memberships. ([fda.gov](#)) FDA's **August 2026 final guidance** finalized the earlier draft. ([fda.gov](#)) Its announcement identifies an additional TE-request process for certain applications under section 505(b)(2), alongside clarifications and updates. ([federalregister.gov](#)) It does not justify describing longstanding RLD and RS distinctions as newly created rules. The final notice appeared on **August 24, 2026**. ([govinfo.gov](#)) Guidance remains nonbinding, while application requirements remain grounded in regulation and statute. ([federalregister.gov](#))

FDA's downloadable Orange Book data files use ASCII text with tilde (~) delimiters. ([fda.gov](#)) For a reproducible census, retain the exact ZIP and publish it with the counting output and member metadata. Keep the product-file and ANDA-RS relationship-list vintages separate; do not equate product-row counts with comparator-edge counts or physical supply.

The recommended implementation preserves product identities, source snapshots, explicit reference edges and TE memberships independently. JSON identifiers should remain strings; schema constraints and foreign keys should reject incomplete joins. Provenance should record where an edge came from and who verified it. This report supplies a flag matrix, source crosswalk, worked budesonide join and acceptance tests. It proposes a data-control design, not automatic comparator selection. ([rfc-editor.org](#)) ([specs.frictionlessdata.io](#)) ([w3.org](#))

Finally, TE, legal substitution, patent obligations and physical supply require separate decisions. Florida's purchaser and prescriber provisions illustrate why a national TE evaluation cannot answer every dispensing question. ANDA patent certifications are separately specified. For unresolved reference choices, obtain FDA clarification and qualified regulatory review before committing a development program. No source reviewed here establishes current comparator inventory. ([leg.state.fl.us](#)) ([ecfr.gov](#)) ([fda.gov](#))

Introduction and Background

A product master can correctly store an ingredient and still answer the wrong development question. The relevant questions are: which product supports the application, which comparator FDA selected for the study, and which products share an applicable TE evaluation? The RLD and RS definitions answer the first two questions separately. ANDA content requirements also identify the RLD by dosage form and strength, making an ingredient-only record insufficient for documenting the basis of submission. ([ecfr.gov](#))

The distinction matters when drug identifiers travel between systems. A published labeling-integration study mapped DailyMed Set IDs and DrugBank identifiers to FDA application numbers, but allowed a source identifier to map to multiple applications. That is evidence for retaining mapping cardinality, not evidence that any application-level match supplies a product-level comparator. The proposed model therefore makes every join's level explicit. ([pmc.ncbi.nlm.nih.gov](#)) ([sqlite.org](#))

This is a regulatory and data-architecture reference for generic-development teams. IntuitionLabs' perspective is that of an adjacent life-sciences and AI consultancy, rather than a supplier of the compared regulatory datasets. The analytical focus is evidence preservation: a reproducible selection record should distinguish observed source content from a team's interpretation. W3C's provenance model provides a useful vocabulary for recording entities, activities and responsibility. ([w3.org](#))

Related resources serve different purposes. RAPS explains ANDA terminology, while Elsevier documents Clinical Pharmacology's display conventions. Neither page should replace current FDA relationship evidence. Research databases also have distinct scopes: EMBL-EBI describes ChEMBL as a bioactivity resource. A database suitable for finding compounds need not be authoritative for selecting an ANDA comparator. ([raps.org](#)) ([elsevier.support](#)) ([ebi.ac.uk](#))

Key Changes

What FDA finalized in August 2026

The finalization should be described narrowly. The Federal Register notice identifies an additional process through which applicants of certain **505(b)(2) applications** may request TE codes. It describes other changes as minor clarifications and updates. This is relevant to evaluation-request workflows, but does not establish that all products changed codes or that every ANDA team acquired a new comparator. ([federalregister.gov](#))

The official notice and its government publication copy identify the final guidance's publication context. Neither the announcement of finalization nor a downloaded notice should become a product-specific effective date in a drug master. Proposed implementation: store a guidance-publication event separately from each observed product-state change. That preserves the difference between interpretive policy and product evidence. ([govinfo.gov](#)) ([federalregister.gov](#)) ([w3.org](#))

What was already established

FDA introduced separate RLD and RS display updates in **January 2017**. ([fda.gov](#)) Their purpose included identifying the basis of submission and the product used in required in vivo studies. Current regulatory definitions preserve the same separation. Accordingly, a migration should repair any model that treated the

flags as synonyms, without claiming the distinction originated in the latest final guidance. ([ecfr.gov](#))

The statutory ANDA framework requires evidence of bioequivalence to the listed drug, while the implementing application-content rule calls for identification of the RLD. These requirements explain why an evidence package needs more than a preferred trade name. The proposed crosswalk records the approval-basis product and study-reference product as distinct endpoints, with an explicit reason when they differ.

([uscode.house.gov](#)) ([ecfr.gov](#))

What should change operationally

Teams should review how TE evaluations are requested, represented and refreshed. The final guidance permits revisions to either TE-code letter. A system that recognizes only initial approval events will miss relevant later evaluation changes. Proposed controls should compare full code sets between snapshots and route changes to an exception queue rather than overwriting the prior interpretation. ([fda.gov](#)) ([w3.org](#))

The additional request process is not a reason to generate TE relationships from text similarity. Evidence should still establish the intended application, product and context. The schema controls below deliberately reject a missing endpoint or ambiguous mapping, consistent with foreign-key constraints and typed field validation. Their implementation is a proposed engineering policy, not an FDA-mandated database design.

([federalregister.gov](#)) ([sqlite.org](#)) ([specs.frictionlessdata.io](#))

Product Identity and the Reference-Selection Model

The model distinguishes five concepts: **application**, **product**, **product heading**, **RLD role** and **RS role**. An application contains product records; the raw data identifies strengths separately. A heading groups descriptive attributes for presentation and interpretation. RLD and RS are roles attached to identified products, rather than alternative spellings for an ingredient or an application. ([fda.gov](#)) ([ecfr.gov](#))

Preserve the source's heading and strength separately. FDA's preface uses ingredient, dosage form and route to discuss headings, with same-strength qualifications for equivalence; the final guidance's heading description includes strength. The proposed model avoids a terminology-dependent key by carrying each component independently and storing the source's own heading representation. A full TE code then qualifies membership within that context. ([fda.gov](#)) ([ecfr.gov](#))

Use the following proposed entities and keys:

- **Application:** source application type and the original application-number string.
- **Product:** application key plus the original product-number string.
- **Description:** ingredient text, dosage form, route and strength, preserved before normalization.
- **Snapshot:** retrieval timestamp, source label, content hash and parser version.
- **Role observation:** RLD flag, RS flag and marketing category for a product in a snapshot.
- **Reference edge:** a documented connection between an RS product and its RLD endpoint.
- **TE membership:** an association between a product, contextual group and complete code.

These are proposed relational boundaries. Table Schema distinguishes physical text from logical types and relations; JSON provides string representation. Those standards support retaining identifiers before applying normalization, while composite foreign keys help preserve the entire referenced identity. None of these technical standards supplies a regulatory relationship by itself. (specs.frictionlessdata.io) (rfc-editor.org) (sqlite.org)

RxNorm illustrates the value and limit of semantic normalization. NLM creates separate clinical-drug concepts when ingredients, strengths or dose forms change. That can support terminology alignment, but this report does not equate an RxNorm concept with an Orange Book product or TE group. The proposed terminology link is supplementary, with its own source and verification state. (nlm.nih.gov) (w3.org)

ChEMBL provides a complementary design example: it registers ingredients of combinations separately and loads Orange Book products into product/formulation tables. Separating substance-level and product-level entities is useful here too. It prevents an ingredient-resolution service from silently becoming the authority for a product-reference edge. Keep normalized terminology links outside the table that asserts FDA reference relationships. (pmc.ncbi.nlm.nih.gov) (sqlite.org)

Implementation Considerations and Process Changes

Flag matrix and source hierarchy

Table 1 separates flag observations from reference decisions. The meanings of RLD and RS come from their regulatory definitions; the actions are proposed controls. Marketing categories are carried alongside the flags because an observation should retain the source's full product state. (ecfr.gov) (specs.frictionlessdata.io)

RLD flag	RS flag	Marketing context	Interpretation and proposed action
Yes	Yes	RX or OTC	The product fills both roles in this observation. Document the exact product and applicable study requirements.
Yes	No	RX or OTC	The product is an approval-basis reference, but its RLD flag does not select it as the study comparator. Seek the current RS relationship.
No	Yes	RX or OTC	The product is a selected comparator without being the RLD. Preserve the separate RLD endpoint and supporting relationship evidence.
Yes	No	DISCN	Retain the RLD observation. Review the separate withdrawal determination and comparator-selection evidence before proceeding.
No	No	RX, OTC or DISCN	The flags establish neither role for this row. Do not manufacture a reference edge from ingredient similarity.

The matrix is not an approval or availability algorithm. The withdrawal-reason determination has a separate regulatory process, and the RLD-identification requirement remains product specific. In the proposed system, every unresolved case retains its original flags and receives a review state; it never becomes a positive reference selection merely because another row shares a heading. (ecfr.gov) (specs.frictionlessdata.io)

Table 2 maps source representations to the proposed evidence fields. A source adapter should preserve both the original location and the normalized value. API documentation explicitly describes JSON conversion; physical and logical representations should therefore be checked rather than assumed identical.

(open.fda.gov) (specs.frictionlessdata.io)

Source	Fields or evidence to retain	Use and limit
FDA raw ZIP	Application type/number, product number, descriptive fields, RLD, RS, TE_Code, Type	Reproducible row census and state comparisons; flags alone do not identify every edge.
openFDA JSON	Observed nesting, original strings, boolean roles and TE-code arrays	Programmatic access; retain the response and test its adapter against the actual structure.
Official product search	Product identity, form, route, strength and displayed roles	Manual endpoint confirmation; one search result does not prove an entire portfolio crosswalk.
Official ANDA-RS PDF	Listed ANDA/product, stated RLD application, strength, page and publication label	Explicit relationship evidence; an omitted RLD product number still requires resolution.
Other terminology or vendor views	External identifier, display value, source version and mapping cardinality	Supplementary representation; never promote a display label or ingredient match into an FDA relationship.

The API should have an independent adapter. Its example queries use **products.application_number**. In the retrieved response, **product_number** sits at the result level, while application identifiers, role booleans and TE codes appear within **products**. Blindly renaming raw columns would lose this structure. Validate the response shape before extracting or joining fields. (open.fda.gov) (api.fda.gov) (rfc-editor.org)

For reproducibility, assign a source label and retrieval time independently of API metadata. openFDA explains that **meta.last_updated** is the API's update time, rather than the date of the newest underlying record. A proposed reconciliation report should show both source snapshots and unresolved differences, without assigning either clock to every individual relationship. (open.fda.gov) (pmc.ncbi.nlm.nih.gov) (w3.org)

TE Codes, Product Headings and Group Membership

An **AB** code reflects resolved bioequivalence questions; a **B** code indicates unresolved actual or potential bioequivalence issues. **BX** means the available data are insufficient to determine therapeutic equivalence. These evaluations must be interpreted in the context of pharmaceutical equivalence and the relevant product group, rather than reduced to a shared ingredient. (fda.gov) (ecfr.gov)

Preserve **AB1**, **AB2** and other numbered codes exactly. The preface also states that some products can carry multiple TE codes. The proposed group relation is therefore many-to-many, and the complete code belongs in the membership record. A product's membership in more than one group does not authorize a transitive closure connecting every member of every neighboring group. (fda.gov) (specs.frictionlessdata.io) (sqlite.org)

A useful historical contrast is the retrieved CMS product-data form. It maps FDA AB and **AB1 through AB9** into the CMS value AB. (medicaid.gov) The form bears an expiration date of **May 31, 2024**. (medicaid.gov) It is evidence of a historical representation, not a statement of the current CMS interface or FDA assignments. Its mapping demonstrates why a downstream field may be unable to reconstruct numbered groups.

Vendor labels require the same discipline. Elsevier explains when Clinical Pharmacology indicates that a TE code does not apply. Keep such display semantics in a vendor-specific namespace. The proposed ingestion rule accepts a vendor label only as that vendor's observation; it does not translate a display state into a new FDA code without explicit source support. (elsevier.support) (specs.frictionlessdata.io) (rfc-editor.org)

A **blank TE code** is not synonymous with B or BX. FDA describes removal when a product becomes single-source, and the API supplies a query specifically for products without TE codes. Preserve missingness and its observed context. Do not fill it with a negative-equivalence code merely to satisfy a mandatory database field. (fda.gov) (open.fda.gov) (specs.frictionlessdata.io)

The key distinction is between a contextual group and a descriptive heading. Proposed membership requires a product identity, source snapshot, heading components, strength and full code. A heading match narrows a review set. A documented membership supports the specific relationship actually asserted. Pharmaceutical-equivalence definitions and normalized clinical-drug concepts do not eliminate that distinction. (ecfr.gov) (nlm.nih.gov) (sqlite.org)

Reference Schema and Worked Join Tests

Evidence-preserving edge design

The proposed reference-edge record contains source and target product keys, relationship kind, publication label, retrieval time, evidence location, original source identifiers and verification status. Add a reviewer identifier and a resolution note when an application-only endpoint is resolved to a product. W3C provenance distinguishes data entities, production activities and responsibility, while relational constraints ensure that referenced endpoints exist. (w3.org) (sqlite.org)

Use explicit states such as **unresolved**, **verified** and **superseded**. These are proposed workflow values, not FDA statuses. A product-number guess should remain unresolved even when its application exists. A foreign key can establish endpoint existence, but cannot establish the scientific or regulatory correctness of the chosen endpoint. Record the documentary reason for the join separately. (sqlite.org) (w3.org)

A compact proposed schema is:

```
product(snapshot_id, appl_type, appl_no_text, product_no_text,
        ingredient_raw, form_route_raw, strength_raw,
        rld_raw, rs_raw, marketing_type_raw, te_codes_raw)

reference_edge(edge_id, rs_snapshot_id, rs_appl_type,
              rs_appl_no_text, rs_product_no_text,
              rld_snapshot_id, rld_appl_type, rld_appl_no_text,
              rld_product_no_text, source_url, publication_label,
              retrieved_at_utc, evidence_location,
              original_endpoint_text, verification_status,
              reviewer_id, resolution_note)

te_membership(snapshot_id, appl_type, appl_no_text,
```

```
product_no_text, heading_context_id,
strength_raw, complete_te_code)
```

This schema is an original recommendation. Its composite identities, string fields and separately represented relations follow standard data-modeling capabilities, without claiming those capabilities validate reference selection. Store role flags at the product-observation level and the reference assertion at the edge level. Keep the original endpoint text even after a reviewer resolves its normalized identity. (sqlite.org) (rfc-editor.org) (specs.frictionlessdata.io) (w3.org)

Recommended required evidence fields are:

- **Original RS identity:** retain application type, number and product number.
- **Original RLD identity:** preserve exactly what the source supplies.
- **Resolved RLD identity:** store a product number only after documented resolution.
- **Relationship kind:** distinguish reference edges from TE memberships.
- **Source URL:** retain the exact location used for verification.
- **Publication label:** copy the source's stated vintage without reinterpretation.
- **Retrieval time:** record when the artifact was acquired.
- **Evidence location:** retain PDF page, row identifier or JSON path.
- **Content hash:** identify the binary artifact used for extraction.
- **Parser version:** identify the extraction implementation.
- **Review status:** distinguish an observation from an accepted interpretation.
- **Reviewer:** record responsibility for accepting the resolution.
- **Resolution note:** explain strength and product-number matching.
- **Prior observation:** link the superseded record without deleting it.
- **Exception reason:** state what prevents acceptance.
- **Scope:** identify the extraction or portfolio denominator.

These are proposed operational fields. They implement provenance and explicit constraints while retaining the original physical representation. They should be documented in the team's data dictionary and tested against actual source artifacts. (w3.org) (specs.frictionlessdata.io)

Budesonide: resolving an application-only target

The September relationship PDF explicitly lists **A215328, product 001**, budesonide rectal aerosol foam **2MG/ACTUATION**, with **RLD N205613**. (fda.gov) The list's RLD endpoint is application level. FDA's N205613 product page independently identifies product **001** with the matching form, route and strength. (accessdata.fda.gov) The approval letter additionally describes bioequivalence and therapeutic equivalence to the RLD. (accessdata.fda.gov)

The completed proposed edge is therefore RS (**A, 215328, 001**) to RLD (**N, 205613, 001**), with the PDF page and product-search record retained as separate evidence. The product number on the RLD side is a documented resolution, not a field copied from the PDF. That distinction should remain visible to a future reviewer. (fda.gov) (w3.org) (sqlite.org)

Table 3 gives proposed acceptance tests (Hypothetical Examples). These tests protect identity, cardinality and evidence lineage. They are implementation controls informed by field-validation and relationship constraints, rather than a substitute for regulatory judgment. (specs.frictionlessdata.io) (sqlite.org)

Test	Input or condition	Expected result
Identifier preservation	Application 078687; product 001	Store strings without dropping leading zeros.
Composite endpoint	Same number, different application type	Treat as a different key; never join on the number alone.
RLD product resolution	PDF supplies only an RLD application	Keep unresolved until product identity and strength are documented.
Numbered TE groups	Same heading/strength; AB1 versus AB2	Do not merge solely by stripping the suffix.
Multiple memberships	Product carries AB1 and AB2	Retain both memberships without connecting all other group members transitively.
Missing evaluation	Empty TE field	Preserve missingness; do not manufacture B or BX.
Snapshot mismatch	Edge and product sources have different vintages	Record both and require explicit reconciliation.
Endpoint deletion	Referenced product absent from a later snapshot	Retain historical evidence and open a review item.

The string and relationship tests can run automatically. Endpoint interpretation remains a review decision. JSON's string representation prevents numeric coercion, while schema constraints and composite foreign keys make incomplete or incompatible mappings detectable. Passing these tests establishes conformance to the proposed data contract, not an automatic authorization to buy a comparator or substitute a drug. (rfc-editor.org) (specs.frictionlessdata.io) (sqlite.org) (federalregister.gov)

Refresh, Exception Control and Evidence Limits

The proposed refresh process preserves prior states and compares role flags, full TE-code sets, marketing categories and relationship endpoints. The statute describes periodic list revisions; API update time and source publication labels remain distinct. Do not treat a successful API request as proof that the relationship PDF and downloadable products share the same effective date. (uscode.house.gov) (open.fda.gov) (pmc.ncbi.nlm.nih.gov)

Use this proposed release checklist:

- **Acquire:** retain each source artifact and its retrieval timestamp.
- **Identify:** preserve original application type, application number and product number.
- **Validate:** check schema, allowed values, missingness and composite-key uniqueness.
- **Compare:** distinguish product-state changes from newly resolved mappings.
- **Reconcile:** review source-vintage differences before accepting changed endpoints.
- **Approve:** record the reviewer, documentary evidence and unresolved questions.

- **Publish:** retain superseded observations and a readable release note.
- **Monitor:** route unexpected code, role or identity changes back to review.

These controls are consistent with provenance recording, field constraints and the practice of manual ambiguity review described by ChEMBL's authors. They are proposed internal procedures, with implementation detail chosen by the organization. A source's mapping should not be promoted into an approved reference selection merely because the extraction pipeline completed successfully. ([w3.org](#)) ([specs.frictionlessdata.io](#)) ([pmc.ncbi.nlm.nih.gov](#))

A **DISCN** observation should prompt a separate withdrawal-reason review. FDA's regulations provide for a determination and notice process before approval in relevant circumstances. Keep that determination, the observed marketing category and the comparator decision in separate fields. No individual product's safety, effectiveness or physical availability should be inferred from a discontinued-list flag alone. ([ecfr.gov](#))

For an unavailable or uncertain comparator, preserve the proposed selection and the unresolved question rather than inventing a replacement edge. Controlled correspondence addresses RLD and RS information separately when appropriate. The model should make the regulatory question reviewable; a qualified team and FDA clarification should resolve it before study commitments. A public listing does not establish stock at a supplier. ([fda.gov](#)) ([ecfr.gov](#)) ([w3.org](#))

Separate legal substitution and patent workstreams too. Florida's statute illustrates purchaser and prescriber conditions, while ANDA patent certification requirements reside in a separate application provision. The crosswalk can link to those workstreams' evidence records, but it should not expose a single combined “substitutable and clear to launch” status derived from TE flags. ([leg.state.fl.us](#)) ([ecfr.gov](#))

Data Analysis and Evidence

Snapshot scope and reproducibility

FDA updates the Orange Book downloadable data files monthly. ([fda.gov](#)) Treat the October 22, 2026 retrieval described in this report as a dated snapshot, rather than as a permanent representation of the current files.

No earlier snapshot was compared, so no product additions, removals or individual flag changes are inferred. A longitudinal analysis would require both retained snapshots.

Before publishing an exact census, make the retained ZIP available in a durable, versioned supplement with the script's output, SHA256 and member metadata. Verify all row totals, duplicate-key counts, role-flag counts and percentages against those same bytes. Record retrieval time separately from member timestamps and each source's publication label.

Reproduce the measurement

The following code operates on a locally retained ZIP artifact; it does not select references or infer edges. Read identifier fields as text, count full composite keys and report duplicates before attempting joins. Physical-versus-logical representation and composite identity are explicit parts of the proposed method. ([specs.frictionlessdata.io](#)) ([rfc-editor.org](#)) ([sqlite.org](#))

```

import csv, hashlib, io, zipfile
from collections import Counter
from pathlib import Path

blob = Path("orange-book.zip").read_bytes()
archive = zipfile.ZipFile(io.BytesIO(blob))
member = next(n for n in archive.namelist()
               if n.lower().endswith("products.txt"))
text = archive.read(member).decode("utf-8-sig")
rows = list(csv.DictReader(io.StringIO(text), delimiter=chr(126)))
keys = [(r["Appl_Type"], r["Appl_No"], r["Product_No"])
        for r in rows]
flags = Counter((r["RLD"], r["RS"], r["Type"]) for r in rows)
print(hashlib.sha256(blob).hexdigest())
print(len(rows), len(set(keys)), len(keys) - len(set(keys)))
print(sorted(flags.items()))

```

Coverage metrics without false precision

Define **unresolved-edge rate** as unresolved candidate edges divided by all candidate edges in the declared extraction scope. Define **reviewed-selection coverage** as reviewed reference selections divided by required selections in a user-supplied portfolio. Report numerator, denominator, scope and review date. These are proposed management metrics; neither is a comparator-selection rule or a cost-saving estimate. ([w3.org](#)) ([specs.frictionlessdata.io](#))

For the worked-example scope, unresolved edges are **0/1 (0%)** after endpoint resolution. This measures only the demonstrated edge, not whole-list completeness. ([fda.gov](#)) ([w3.org](#)) This run did not enumerate and review the entire relationship PDF into a crosswalk. It therefore reports no whole-list unresolved-edge rate. No portfolio was supplied, so portfolio coverage is **not calculated**, rather than reported as zero or complete. Missing denominators should remain explicit in the proposed evidence model. ([w3.org](#)) ([rfc-editor.org](#))

Implications and Future Directions

The useful output is a defensible selection record, not the largest possible ingredient graph. A proposed architecture should let a reviewer reconstruct an endpoint resolution, inspect the source snapshot and see which decision remains open. That makes provenance a practical review capability. It also preserves distinctions between regulatory definitions, internal workflow states and source representations. ([w3.org](#)) ([ecfr.gov](#))

Historical datasets can support longitudinal research, but their scope must remain visible. NBER's historical resource digitized patent and exclusivity tables for **1985-2016**, using hand entry or parsing of historical PDFs. ([nber.org](#)) Its user guide discusses methodological choices. Such a resource is complementary to current product-reference evidence, rather than a substitute for the latest role observations or relationship list. ([nber.org](#))

Cross-dataset integration should remain explicit about cardinality and timing. The published labeling study permits multiple application mappings, and ChEMBL describes both automated checks and manual ambiguity review. The proposed next step is an auditable reconciliation layer that preserves alternative candidate mappings until evidence resolves them. Do not let chemical or clinical terminology normalization erase application/product identity. ([pmc.ncbi.nlm.nih.gov](https://pubmed.ncbi.nlm.nih.gov/))

For downstream analytics, publish the full source-code representation alongside any simplified display. The historical CMS AB mapping and Elsevier's documented presentation semantics show why display-oriented data must retain its namespace. A normalized clinical-drug concept can aid navigation, while the reference edge and TE membership remain separately sourced assertions. That separation is the durable design decision. ([medicaid.gov](https://www.medicare.gov/)) ([elsevier.support](https://www.elsevier.com/health/science-direct)) ([nlm.nih.gov](https://www.nlm.nih.gov/)) (specs.frictionlessdata.io)

Conclusion

An RLD-to-RS-to-TE crosswalk should answer distinct questions with distinct evidence. The RLD identifies the product relied upon for the application; the RS identifies FDA's selected comparator for required in vivo studies. Application content requirements reinforce the need to identify form and strength. A product master should preserve these roles without assigning either one from a shared ingredient alone. ([ecfr.gov](https://www.ecfr.gov/))

The proposed implementation retains original string identifiers, composite product keys, source-specific observations, explicit reference edges and contextual TE memberships. Schema validation can identify missing values or incompatible keys, while provenance records the documentary basis and review responsibility. These controls support review; they do not turn a valid database relationship into an automatic development decision. ([rfc-editor.org](https://www.rfc-editor.org/)) ([sqlite.org](https://www.sqlite.org/)) (specs.frictionlessdata.io) ([w3.org](https://www.w3.org/))

Teams should carry that separation through their decision process. Source publication, retrieval and observation times need their own fields. Historical research datasets, terminology services and display systems remain useful when their scopes are explicit. The final guidance's request-process clarification belongs in regulatory procedures, while unresolved reference choices remain documented review questions. ([nber.org](https://www.nber.org/)) ([nlm.nih.gov](https://www.nlm.nih.gov/)) ([federalregister.gov](https://www.federalregister.gov/))

A defensible record lets the next reviewer trace the chosen product, understand the relationship claimed and identify what still requires confirmation. Keep substitution conditions and patent certifications in their own workstreams. The resulting graph is an evidence-backed reference resource whose limits remain visible to the people making the regulatory and development decisions. (leg.state.fl.us) ([ecfr.gov](https://www.ecfr.gov/)) ([w3.org](https://www.w3.org/))

Source links

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

1. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-314/subpart-A/section-314.3>
2. <https://www.fda.gov/drugs/development-approval-process-drugs/orange-book-preface>
3. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/evaluation-therapeutic-equivalence>
4. <https://www.federalregister.gov/documents/2026/08/24/2026-17215/evaluation-of-therapeutic-equivalence-guidance-for-industry-availability>

5. <https://www.govinfo.gov/content/pkg/FR-2026-08-24/pdf/2026-17215.pdf>
6. <https://www.fda.gov/drugs/drug-approvals-and-databases/orange-book-data-files>
7. <https://www.rfc-editor.org/info/rfc8259>
8. <https://specs.frictionlessdata.io/table-schema>
9. <https://www.w3.org/TR/prov-overview>
10. https://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&Search_String=440&URL=0400-0499%2F0465%2FSections%2F0465.025.html
11. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-314/subpart-C/section-314.94>
12. <https://www.fda.gov/media/164111/download>
13. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8222600>
14. <https://www.sqlite.org/foreignkeys.html>
15. <https://www.raps.org/resource/fda-clarifies-reference-listed-drugs-other-anda-te.html>
16. <https://www.elsevier.support/ckpharma/answer/what-are-orange-book-therapeutic-equivalence-ratings-in-the-drug-product-information>
17. <https://www.ebi.ac.uk/about/teams/chemical-biology-resources>
18. <https://www.fda.gov/drugs/drug-approvals-and-databases/fda-introduces-reference-standard-data-updates-orange-book>
19. <https://uscode.house.gov/us/usc/119-121/t21/s355>
20. <https://www.fda.gov/media/160054/download>
21. <https://www.nlm.nih.gov/research/umls/rxnorm/sourcereleasedocs/rxnorm.html>
22. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12516679>
23. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-314/subpart-D/section-314.161>
24. <https://open.fda.gov/apis/drug/orangebook>
25. <https://open.fda.gov/apis/drug/orangebook/example-api-queries>
26. https://api.fda.gov/drug/orangebook.json?search=products.application_number:078687&limit=1
27. <https://open.fda.gov/apis/anatomy-of-a-response>
28. <https://www.medicaid.gov/medicaid/prescription-drugs/downloads/mdp-cms-form-367c-product-data-11.pdf>
29. <https://www.fda.gov/media/103879/download>
30. https://www.accessdata.fda.gov/scripts/cder/ob/results_product.cfm?Appl_No=205613&Appl_Type=N
31. https://www.accessdata.fda.gov/drugsatfda_docs/appletter/2023/215328Orig1s000ltr.pdf
32. <https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations-orange-book>
33. <https://www.nber.org/research/data/orange-book-patent-and-exclusivity-data-1985-2016>
34. <https://www.nber.org/papers/w30628>

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.