



INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

NDC Code Normalization: 10-, 11- and 12-Digit FDA 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 · 2026-10-20 · ndc code normalization · convert 10 digit ndc to 11 digit · fda ndc directory · 12-digit ndc · ndc 4-4-2 5-3-2 5-4-1 · openfda ndc api · ndc to rxnorm mapping · ndc upc gtin conversion · ndc package code matching

Original article: <https://intuitionlabs.ai/articles/fda-ndc-directory-code-normalization-guide>



In this article

1. Executive Summary
 2. Introduction and Background
 3. Choosing the Canonical Identifier by Use Case
 4. What the FDA Directory Contains and What It Does Not
 5. Ten-Digit Segment Patterns and Eleven-Digit Padding
 6. A Deterministic Normalization Algorithm and Test Fixtures
 7. The Twelve-Digit Final Rule and Migration Horizon
 8. Product-Versus-Package Joins and Downstream Mappings
 9. Ingesting FDA and openFDA Files
 10. Data Analysis and Evidence
 11. Case Studies and Real-World Examples
 12. Implications and Future Directions
 13. Frequently Asked Questions (FAQs)
 14. Conclusion
-

Executive Summary

The National Drug Code (NDC) is a unique, three-segment number that identifies a drug's labeler, product and trade package size, and the FDA NDC Directory that publishes it is updated daily (open.fda.gov). Normalizing NDCs is hard for a structural reason. Today's FDA-assigned code has 10 digits in one of three configurations, 4-4-2, 5-3-2 or 5-4-1, while claims systems use **a fixed 11-digit 5-4-2 string** with a leading zero added to the appropriate segment (www.federalregister.gov) (www.cms.gov). FDA itself warns that a padded code such as 12345-0678-09 could be read back as 12345-678-09 or 12345-0678-9 (www.fda.gov). A syntactically valid conversion can therefore be semantically wrong.

The central recommendation of this report is to store **the raw hyphenated NDC exactly as published**, keep the segment pattern as metadata, and derive every other representation (11-digit, 12-digit, GTIN) from that record rather than from a stripped digit string. Product-level codes (labeler plus product) and package-level codes (all three segments) are separate grains, and NLM's RxNorm normalizes source NDCs to the 11-digit HIPAA format only after intake (www.nlm.nih.gov). Joins between grains, and to RxNorm or claims data, should state which grain is in play.

The migration horizon is defined. FDA's final rule makes the NDC a uniform 12-digit number with a **6-digit labeler code, a 4-digit product code and a 2-digit package code, effective March 7, 2033** (www.federalregister.gov) (www.govinfo.gov). That does not mean every payer format changes on the same day: FDA states that HIPAA code-set updates are outside the scope of FDA (www.govinfo.gov) and that it expects the HIPAA 11-digit standard to be replaced on or before the effective date (www.govinfo.gov). NCPDP's March 27, 2026 letter says its standards already support the expansion ([NCPDP](https://www.ncpdp.org)). Standards capability does not establish a payer's implementation date; confirm readiness with each trading partner.

Scale and cost figures come from FDA's own analysis. Between 240,000 and 250,000 NDC numbers are affected (www.fda.gov), and annualized industry costs are **about \$14.64 million** at a 7 percent discount rate (www.govinfo.gov). As of the 2026-10-16 export, the openFDA bulk manifest lists **138,319 product records** (api.fda.gov). This report does not publish a collision rate for padded 11-digit codes, because no snapshot-based figure was found; it gives a measurement protocol instead.

Finally, listing is not endorsement. FDA states that inclusion in the Directory does not indicate FDA verification (www.fda.gov) and does not mean a product is covered or eligible for reimbursement (www.fda.gov). Any pipeline that treats an NDC match as proof of approval, coverage or availability misreads the source.

Introduction and Background

Data engineers, regulatory operations staff and reimbursement-data teams routinely join NDC values across systems that were never designed together. A label repository stores hyphenated codes, a claims warehouse stores 11-character strings, a spreadsheet silently drops leading zeros, and a barcode scanner returns a GTIN. Each join that crosses these boundaries is an opportunity for a false match, and the failure is quiet: a wrong join returns a plausible drug.

This report examines how to ingest FDA NDC files, normalize code representations and join product and package records without creating those false matches. It answers a practical question: which canonical key, padding rule and source field should a given workflow use? IntuitionLabs is a life-sciences and AI consultancy and an adjacent advisor to this topic; it does not sell an NDC dataset or normalization product, and nothing here depends on any vendor.

Three facts frame the analysis. First, the NDC is assigned in segments, and the labeler code is assigned by FDA (www.fda.gov). Second, FDA's current regulation requires each listed finished or unfinished drug to carry a unique NDC (www.ecfr.gov). Third, the Directory is a listing database, not an approval register. The sections below therefore treat the NDC as an identifier with a defined grammar, and treat every alternate form as a derived display or exchange format.

The report covers FDA Directory files, the openFDA NDC endpoint, the 10-, 11- and 12-digit forms, and mappings to RxNorm, DailyMed and GS1 identifiers. All dates are anchored to October 2026.

Choosing the Canonical Identifier by Use Case

No single representation is right for every workflow, so the guidance below separates storage keys from exchange formats.

Storage and provenance

For an internal master table, keep three things per record: the raw identifier exactly as received, the segment lengths (for example 5-3-2) and the source file with its retrieval date. FDA's own file definitions publish the product NDC as the labeler and product segments separated by a hyphen (www.fda.gov), and the package NDC as all three segments separated by hyphens (www.fda.gov). The Structured Product Labeling (SPL)

implementation guide likewise treats the package item code as 10 digits excluding hyphens, divided into three hyphenated segments (www.fda.gov). Because the hyphens carry the segment boundaries, the hyphenated string is the only form that is lossless in the current 10-digit world.

Exchange with billing and claims systems

For claims and rebate workflows, the **11-digit, hyphen-less 5-4-2** string is the working convention. CMS instructions for institutional claims describe the NDC as 11 digits following the N4 qualifier (www.cms.gov), and the professional claim instructions say the same (www.cms.gov). The NUCC 1500 manual lists N4 as the qualifier for National Drug Codes (nucc.org) and tells preparers not to enter hyphens in the code (nucc.org). State drug utilization data published by CMS is reported at the 11-digit level (www.medicaid.gov). These are output formats: they should be generated at the boundary, with the pattern recorded, never used as the only stored copy.

Terminology mapping and barcodes

For RxNorm lookups, NLM normalizes source NDCs to the 11-digit HIPAA format (www.nlm.nih.gov), and its APIs accept the CMS 11-digit form or the 4-4-2, 5-3-2 and 5-4-1 forms (lhncbc.nlm.nih.gov). For scanning, GS1 US documentation states that the 10-digit NDC format is always used when embedding an NDC in a GTIN (documents.gs1us.org). The practical rule is that a GTIN is a derived key, not an NDC.

Summary of choices

- **Master key:** the raw hyphenated package NDC (three segments) plus a pattern label such as 5-3-2.
- **Product-level key:** the hyphenated labeler-product pair, kept in a separate table.
- **Claims and rebate exchange:** the derived 11-digit string, flagged as derived.
- **Future-proof key:** the derived 12-digit string, populated now only where the pattern is known.
- **Terminology join:** the RxNorm-normalized 11-digit value, with the NDC status checked before use.

What the FDA Directory Contains and What It Does Not

Grain and inclusion

The Directory contains information on active and certified finished and unfinished drugs (www.fda.gov). Entries appear only for listings that have reached their marketing start date and have not reached a marketing end date (www.fda.gov).

Two grains coexist: the product file (labeler and product segments) and the package file (all three segments). Treating them as one table produces repeated product rows and misleading counts, so they should be modeled as parent and child.

Exclusions and limits

The Directory does not include animal drugs, blood products or drugs manufactured under contract (www.fda.gov). Records can also leave the file: products with a value in EndMarketingDate are removed from the Directory (www.fda.gov), and flagged records are held out of the default product file (www.fda.gov).

Exclusion flag U identifies a listing whose certification expired for lack of listing certification ([FDA product-file definitions](#)). An ingestion job that only reads the current file will therefore see disappearance, not a reason; the reason lives in the excluded files or in a change log.

What an NDC does not mean

FDA is explicit that Directory inclusion does not indicate that FDA has verified the information provided ([www.fda.gov](#)), that assignment of an NDC does not denote FDA approval of the product ([open.fda.gov](#)), and that inclusion does not mean a product is covered or eligible for reimbursement ([www.fda.gov](#)). For openFDA, the user warning is equally direct: do not rely on openFDA to make decisions regarding medical care ([open.fda.gov](#)). In schema terms, the NDC is a listing identifier. It must never be used as a proxy for approval status, availability or payer eligibility, and a join that produces a match should be described as "listed", not "approved" or "covered".

Update cadence

Current FDA pages and the openFDA description state that the Directory is updated daily ([open.fda.gov](#)). Pipelines should still record the retrieval timestamp rather than assume a cadence, because a daily source can be fetched at any point in its cycle.

Ten-Digit Segment Patterns and Eleven-Digit Padding

The three patterns

FDA describes the NDC as a 10-digit, three-segment number ([www.fda.gov](#)) that takes one of three configurations: 4-4-2, 5-3-2 or 5-4-1 ([www.fda.gov](#)). The labeler segment identifies the company, the product segment a specific strength, dosage form and formulation, and the package segment the package size and type ([www.fda.gov](#)). The regulation constrains the combinations: a 4-digit labeler code may be combined only with a 4-digit product code and a 2-digit package code ([www.ecfr.gov](#)).

A rule that is easy to miss helps with disambiguation: a registrant with a given labeler code must use only one product-package configuration ([www.ecfr.gov](#)). That makes the pattern a property of the labeler rather than of each package, which matters for the recovery logic below. The current eCFR text of Part 207 also describes an NDC of 10 or 11 digits in three segments ([www.ecfr.gov](#)); in practice the Directory content discussed here is the 10-digit form.

Why several 11-digit forms exist

The 11-digit HIPAA string is fixed at 5-4-2 ([www.cms.gov](#)). Converting a 10-digit code means adding a leading zero to the appropriate segment ([www.federalregister.gov](#)). CMS rebate guidance spells out the three cases: add a leading zero to a 4-digit labeler code ([www.medicaid.gov](#)), to a 3-digit product code ([www.medicaid.gov](#)), and to a 1-digit package code ([www.medicaid.gov](#)).

The consequence is that each of the three 10-digit patterns pads a different position, and the padded results look alike. FDA's background page acknowledges that, because of the conflict with the HIPAA 11-digit standard, many programs pad the product or package segments and gives the example that **12345-0678-09**

could be **12345-678-09** or **12345-0678-9** (www.fda.gov). This report's own analysis adds a second observation: even the unhyphenated 10-digit digit string is ambiguous. Both **12345-678-90** and **12345-6789-0** reduce to **1234567890**, so a spreadsheet or barcode that dropped the hyphens has already lost the pattern.

When not to convert

Do not convert when the destination can store the raw form, when the pattern is unknown, or when the digits have been through a tool that may have trimmed leading zeros. FDA says leading zeros are part of the NDC and that systems must preserve them (www.govinfo.gov). **Store NDCs as text, never as integers.**

A Deterministic Normalization Algorithm and Test Fixtures

The decision tree

The tree below is this report's synthesis of the FDA, CMS and eCFR statements cited above, not an FDA-published algorithm. It is applied in order.

- 1. Is the input hyphenated with three numeric segments?** Classify by segment lengths. Route 6-4-2 directly to step 7 for 12-digit handling (**FDA final rule**). Route 4-4-2, 5-3-2 and 5-4-1 to step 2; reject other segment lengths.
- 2. For a valid 10-digit input, derive.** Pad the labeler to 5, the product to 4 and the package to 2 for the 11-digit form. Pad the labeler to 6 for the 12-digit form.
- 3. Is the input 10 digits with no hyphens?** Reject unless the pattern is supplied from the source system. Do not guess.
- 4. Is the input 11 digits?** Enumerate candidates: a leading zero in digit 1 implies a 4-digit labeler (4-4-2); a zero in digit 6 implies a padded 3-digit product (5-3-2); a zero in digit 10 implies a padded 1-digit package (5-4-1). Because 5-digit labeler codes run from 10,000 to 99,999 (www.govinfo.gov), a leading zero is not a 5-digit labeler.
- 5. One candidate:** accept and mark the pattern as "inferred". **Zero candidates:** reject. **Two candidates:** hold as ambiguous.
- 6. Resolve ambiguity only with evidence.** Evidence means the pattern from the source record, or the labeler's single configuration observed in the Directory under the one-configuration rule. Mark the result "inferred from labeler", never "observed".
- 7. Is the input 12 digits?** Validate 6-4-2, and keep the raw value. A 12-digit value does not reveal which 10-digit pattern it came from.

Reference implementation

The functions below implement the 10-digit parsing and padding branches and enumerate candidates from an unhyphenated 11-digit input; a caller must dispatch 12-digit inputs separately to step 7. It was run against the fixtures that follow, and a randomized round-trip test confirmed that the 12-digit form always equals a leading zero prepended to the 11-digit form for existing 10-digit codes.

```
import re
CFG = {"4-4-2", "5-3-2", "5-4-1"}
```

```

def parse10(ndc):
    m = re.fullmatch(r"(\d+)-(\d+)-(\d+)", ndc)
    if not m:
        return None
    lab, prod, pkg = m.groups()
    cfg = f"{len(lab)}-{len(prod)}-{len(pkg)}"
    return (lab, prod, pkg, cfg) if cfg in CFG else None

def to11(ndc):
    r = parse10(ndc)
    return None if r is None else r[0].zfill(5) + r[1].zfill(4) + r[2].zfill(2)

def to12(ndc):
    r = parse10(ndc)
    return None if r is None else r[0].zfill(6) + r[1].zfill(4) + r[2].zfill(2)

def candidates_from11(d):
    if not re.fullmatch(r"\d{11}", d):
        return None
    lab, prod, pkg = d[:5], d[5:9], d[9:]
    out = []
    if lab[0] == "0":
        out.append(f"{lab[1:]}-{prod}-{pkg}")
    if lab[0] != "0" and prod[0] == "0":
        out.append(f"{lab}-{prod[1:]}-{pkg}")
    if lab[0] != "0" and pkg[0] == "0":
        out.append(f"{lab}-{prod}-{pkg[1:]}")
    return out

```

Test fixtures

Table 1 below lists accepted fixtures. The codes are synthetic and illustrate formatting only; they make no claim about any real product. The GTIN-12 column follows the GS1 US construction of a leading 3, the 10-digit NDC and a check digit (documents.gs1us.org), and the check digits were computed with the GS1 modulo-10 method.

Raw (hyphenated)	Pattern	11-digit derived	12-digit derived	GTIN-12 derived	Candidates recovered from 11-digit
0123-4567-89	4-4-2	00123456789	000123456789	301234567896	1 (0123-4567-89)
12345-678-90	5-3-2	12345067890	012345067890	312345678906	1 (12345-678-90)
12345-6789-0	5-4-1	12345678900	012345678900	312345678906	1 (12345-6789-0)
12345-067-89	5-3-2	12345006789	012345006789	312345067892	1 (12345-067-89)
12345-678-09	5-3-2	12345067809	012345067809	312345678098	2 (ambiguous)
12345-0678-9	5-4-1	12345067809	012345067809	312345067892	2 (ambiguous)

The paired rows illustrate syntactic ambiguity when the labeler configuration is unknown: rows two and three lose their boundaries in the same 10-digit string, while rows five and six produce the same padded strings. Each pair uses mutually exclusive configurations for labeler 12345 under **21 CFR 207.33(b)(3)**. These are alternative interpretations, not evidence of collisions between compliant assigned package NDCs.

Table 2 below lists inputs a pipeline should reject or hold, with the expected outcome.

Input	Expected outcome	Reason
1234506780	Reject	10 digits with no hyphens: pattern unknown
123456789	Reject	9 digits: probable lost leading zero in a spreadsheet; cannot be repaired without the source
12345678912	Reject	11 digits with no zero in positions 1, 6 or 10: not derivable from any 10-digit pattern
12345-67-890	Reject	Segment lengths 5-2-3 are not a valid configuration
1234-567-890	Reject	Segment lengths 4-3-3 are not a valid configuration
12345067809	Hold as ambiguous	Two syntactic 10-digit candidates from mutually exclusive labeler configurations; resolve with source evidence
012345067809	Validate as 12-digit; keep raw and provenance	Synthetic locally derived future representation; syntax alone does not establish assignment
012345-0678-09	Route to step 7; validate as 6-4-2	Same synthetic locally derived future representation; not evidence of an assigned native code
00123456789	Accept as inferred 4-4-2	Leading zero implies a 4-digit labeler

Table 2 separates rejected inputs, unresolved candidate interpretations and syntactically valid future representations. Hold ambiguous inputs until source evidence resolves the labeler configuration; the alternatives are not evidence of two compliant assignments coexisting. Quarantine unresolved records and report their count.

The Twelve-Digit Final Rule and Migration Horizon

What the rule says

FDA's final rule, published in the Federal Register on March 5, 2026, is effective **March 7, 2033** (www.federalregister.gov). All FDA-assigned NDCs will be 12 digits in length with 3 distinct segments: a 6-digit labeler code, a 4-digit product code and a 2-digit package code (www.govinfo.gov). FDA says the converted number is the same NDC in a different format rather than a new NDC (www.govinfo.gov). The conversion adds leading zeros to the labeler, product and package segments as needed (www.fda.gov), and, as noted above, for every existing pattern this equals a single leading zero prepended to the 11-digit form (this report's derivation, confirmed by 200,000 randomized synthetic round trips in the test harness).

The rule's rationale is labeler-code supply. The 5-digit space holds **90,000** codes, and FDA stated that it expected to run out of 5-digit labeler codes in about **10 to 15 years** (www.govinfo.gov). Adding a sixth digit is described as a **900-year supply** (www.govinfo.gov). New 6-digit labeler codes begin at **100000**, because shorter existing codes are zero-padded (www.govinfo.gov).

Timeline

Until the effective date, FDA will continue to assign 10-digit NDCs in the current formats (www.fda.gov). On the effective date FDA will assign new 12-digit NDCs, will issue only 6-digit labeler codes (www.govinfo.gov), and will convert existing 10-digit NDCs in product listings (www.govinfo.gov). FDA states it will automatically update all drug listing files to reflect the 12-digit format (www.fda.gov). The three-year labeling transition runs from **March 7, 2033 through March 6, 2036**. During that period FDA does not intend to object to continued use of 10-digit NDCs on drug labels; FDA warns that drugs labeled with a 10-digit NDC after **March 7, 2036** may be subject to regulatory action ([FDA labeling transition](#)). Products introduced into interstate commerce before the effective date may remain on the market (www.govinfo.gov).

What FDA says about the Directory during migration

FDA states that it intends to publish a mapping from each 10-digit NDC in the Directory to its 12-digit equivalent (www.govinfo.gov), and to publish both forms during the transition (www.govinfo.gov). The paired synthetic examples represent mutually exclusive labeler configurations; they do not demonstrate collisions between compliant assigned codes. For reverse conversion, FDA explains that the interpretation depends on the firm's configuration ([FDA background information](#)).

What does not change automatically

The rule does not rewrite downstream formats. FDA says HIPAA code-set updates are outside its scope (www.govinfo.gov), and that nothing in the rule governs dashes in electronically stored NDCs (www.govinfo.gov). It also states that the rule does not require a change to the NDC format used to claim or dispute historical Medicaid rebates (www.govinfo.gov). FDA expresses an expectation that the HIPAA 11-digit standard will be replaced on or before the effective date (www.govinfo.gov), and that downstream parties need systems capable of handling the new format (www.govinfo.gov). NCPDP's March 27, 2026 letter says its standards already support the expansion ([NCPDP](#)). Keep standards capability, HIPAA adoption and each trading partner's implementation date separate; confirm payer readiness.

Barcodes and the UPC question

The secondary question of converting a 12-digit NDC to a UPC cannot be answered by padding. FDA acknowledges that GTIN-12 and GTIN-14 cannot accommodate a 12-digit NDC (www.federalregister.gov) and that a GTIN cannot embed an NDC longer than 10 digits (www.federalregister.gov). GS1 US told FDA that industry will no longer be able to use the UPC-A barcode for it (www.federalregister.gov). The replacement carrier encodes the NDC with **application identifier 715** and can be read only with camera-based scanners (www.federalregister.gov). Until 2033, the 10-digit derivation still holds: a GTIN-14 example from GS1 US is the indicator digit, 03, the 10-digit NDC and a check digit (documents.gs1us.org).

Product-Versus-Package Joins and Downstream Mappings

Keep grains separate

The FDA product file identifies a product by labeler and product segments, and openFDA defines the product identifier as a concatenation of the NDC product code and the SPL document identifier (open.fda.gov). The `spl_id` field identifies a particular version of a label (open.fda.gov), so it changes when the label is revised while the NDC does not. Join the corresponding FDA product and package files on **ProductID** ([FDA product-file notes](#)). For a separate product dimension, use the labeler-product pair as a business identifier only after explicit deduplication; retain the contributing ProductIDs, SPL versions and source snapshot dates so that version and snapshot selection remain auditable.

Common false-match patterns include:

- **Segment truncation:** joining a package NDC to a product-level source on its first two segments without confirming the pattern.
- **Key mismatch:** joining a derived 11-digit string to a table keyed on raw hyphenated codes.
- **Labeler as entity:** treating one labeler code as one corporate entity.

FDA notes that a manufacturer with multiple establishments can use one labeler code for all of them (www.fda.gov), so labeler code is not a facility key.

RxNorm mapping

RxNorm receives NDCs in many source formats and distributes them in RXNSAT from seven sources (www.nlm.nih.gov). Normalized NDCs follow the 11-digit, no-dashes HIPAA format (www.nlm.nih.gov), while obsolete NDCs remain with their source atoms in their original format (www.nlm.nih.gov). Normalized NDCs attach to drug products at the SBD, SCD, BPCK and GPCK levels (lhncbc.nlm.nih.gov), and a single NDC is expected to map to only one RxNorm clinical drug concept (www.nlm.nih.gov). RxNorm no longer propagates NDCs from branded to unbranded drugs (www.nlm.nih.gov).

The status endpoint matters for joins. The `getNDCStatus` function returns one of four values:

- **ACTIVE:** the NDC is active and linked to an active concept (lhncbc.nlm.nih.gov).
- **OBSOLETE:** it previously existed but is no longer active (lhncbc.nlm.nih.gov).
- **ALIEN:** it exists only in a source vocabulary other than RxNorm (lhncbc.nlm.nih.gov).
- **UNKNOWN:** it is unknown or invalid (lhncbc.nlm.nih.gov).

A pipeline should store the status next to the RxCUI and never treat a missing RxNorm match as proof that an NDC is wrong. The `getNDCs` function returns NDCs in the CMS 11-digit derivative form (lhncbc.nlm.nih.gov), and `findRxcuiById` accepts 10-digit forms including 4-4-2 (lhncbc.nlm.nih.gov). NLM publishes the full RxNorm set on the first Monday of each month (www.nlm.nih.gov), with a weekly update on Wednesdays from the MTHSPL source (www.nlm.nih.gov), and the RxNav APIs refresh after each full monthly release (lhncbc.nlm.nih.gov). openFDA records also carry an RxNorm identifier field defined as the RxNorm Concept Unique Identifier (open.fda.gov), which is convenient but should be treated as an annotation: openFDA annotates the original records with additional fields (open.fda.gov).

DailyMed and SPL

DailyMed exposes lists of all NDC codes (dailymed.nlm.nih.gov) and the NDCs for a specific set identifier (dailymed.nlm.nih.gov). Because SPL stores NDCs in hyphenated 10-digit form, a DailyMed join should use the hyphenated string and the pattern, not an 11-digit string.

Ingesting FDA and openFDA Files

Choosing a source

FDA distributes the Directory as downloadable files, including a zipped text version (www.fda.gov). openFDA provides the same underlying data as an API and as bulk JSON; each bulk file is a zipped JSON file (open.fda.gov), and users cannot simply download increments: the full file must be re-downloaded when an endpoint updates (open.fda.gov). For repeatable pipelines, the bulk file is preferable to paging the API. The API without a key allows **240 requests per minute and 1,000 per day** per IP address (open.fda.gov); with a key the daily limit is **120,000 per key** (open.fda.gov). Paging by skip works only up to a limit of about 26,000 records (open.fda.gov), well below the size of the NDC endpoint. One openFDA how-to page states a maximum limit of 100 per call (open.fda.gov); other openFDA pages state different values, so verify limits against the live service.

Source-field dictionary

Table 3 below maps the fields most relevant to normalization.

Field (FDA file / openFDA)	Grain	Meaning and caveat	Normalization rule
ProductNDC / product_ndc	Product	Labeler and product segments separated by a hyphen	Store raw text; record pattern prefix (4-4 or 5-3 or 5-4); use as a business identifier for a separately deduplicated product dimension with SPL version and snapshot provenance; join corresponding FDA files on ProductID (FDA)
NDCPackageCode / packaging.package_ndc	Package	All three segments separated by hyphens	Primary raw key; derive 11- and 12-digit forms
ProductID / product_id	Product version	Product code plus SPL document id	Join corresponding FDA product and package files on ProductID within the same snapshot; retain for version tracking, not as an NDC (FDA)
spl_id	Label version	A particular version of a label	Change-detection field
EndMarketingDate	Product	Products with a value are removed from the Directory	Close the validity interval; do not delete history

Field (FDA file / openFDA)	Grain	Meaning and caveat	Normalization rule
Listing_Record_Certified_Through / listing_expiration_date	Product	Date the listing expires if not updated or certified (open.fda.gov)	Track expiry separately from marketing end
NDC_Exclude_Flag	Product	Flagged records are excluded from the default product file	Join to excluded files to capture the reason
finished	Product	FDA does not review and approve unfinished products (open.fda.gov)	Keep the flag; never infer approval
packaging.sample	Package	Indicates a sample packaging (open.fda.gov)	Exclude samples from commercial joins
openfda.is_original_packager	Product	Whether the drug has been repackaged for distribution (open.fda.gov)	Use to separate repackager rows

Table 3 shows that no field in the Directory or openFDA carries an 11-digit or 12-digit form before 2033; those must be derived. It also shows that end date, certification date and exclusion flag are separate lifecycle fields; collapsing them into one "active" boolean loses information.

Update and change-detection logic

Keyed by the raw package NDC plus an effective-date interval, the change table should record first-seen and last-seen dates, a hash of each mapped field, marketing start and end dates, listing expiry, the exclude flag and the retrieval timestamp. The openFDA manifest provides an export date and a record count (api.fda.gov), which make a natural snapshot identifier. A suggested loop is:

- **Fetch:** download the full file and record the manifest export date.
- **Join FDA files:** pair product and package files from the same retrieval snapshot and join on ProductID (FDA); retain SPL-version provenance before deduplicating by the labeler-product business identifier.
- **Hash:** compute a row hash over all mapped fields, with the raw NDC as the key.
- **Diff:** classify rows as new, changed, absent or reappeared.
- **Interpret absence:** treat absence as "not in the current file", and look for an end date or exclusion flag before concluding that a product was withdrawn.
- **Derive:** regenerate 11- and 12-digit forms only for rows whose raw code or pattern changed.
- **Alert:** flag any row whose derived 11-digit value now matches a different raw code.

Downloadable mapping template

The following header row is a minimal mapping template, with one row per package NDC.

```
raw_package_ndc,pattern,product_ndc,ndc11_derived,ndc12_derived,pattern_basis,rxcul,rxnorm_status,source_file,source_export_date,first_seen,last_seen
```

The `pattern_basis` column takes the values observed, inferred or ambiguous, so consumers that need certainty can filter on it.

Data Analysis and Evidence

Scale of the identifier space

As of the 2026-10-16 export date, the openFDA bulk manifest lists **138,319** records for the NDC file (api.fda.gov). That figure counts product-level records, not package NDCs, and openFDA carries an unvalidated-data disclaimer. For comparison, FDA's final regulatory impact analysis, using registration and listing records, puts the count of NDC numbers affected by the rule at **240,000 to 250,000** (www.fda.gov), and the 2022 proposal's paperwork section cited about **288,000** listed drug packages (www.govinfo.gov). These numbers differ because they count different things and different dates: products versus packages, a listing-database extract versus the public Directory, and 2022 versus 2025 and 2026 vintages. They should not be combined into a single trend.

Cost and affected parties

FDA's analysis estimates annualized costs of **\$14.64 million**, with a range from \$7.64 million to \$22.79 million at a 7 percent discount rate over ten years (www.govinfo.gov), and a total present value of **\$102.85 million** (www.govinfo.gov). These are regulator estimates of industry compliance cost, not of any single payer's or software owner's cost.

The collision question

FDA's own impact analysis identifies the collision risk in 11-digit form: the issue would arise if FDA began to assign 11-digit NDCs that are indistinguishable from the HIPAA standard form (www.fda.gov). That concerns a hypothetical native 11-digit assignment, which the final rule avoids. The ambiguity in this report, one 11-digit string mapping back to two 10-digit codes, is documented by FDA with a worked example.

This report does not report an empirical collision rate. A proposed data-quality protocol is to retain a dated package-file snapshot and validate segment lengths and the one-configuration-per-labeler requirement in **21 CFR 207.33(b)(3)** before deriving keys. Separately group derived 11-digit values by distinct raw package NDC and investigate duplicates for invalid configurations, repeated versions or mixed snapshots. Report unresolved observed duplicates and the denominator with exclusions disclosed; do not count the synthetic alternatives above as collisions between compliant assignments.

Mapping coverage evidence

Published mapping studies show that even well-formed NDCs do not always link to terminologies. A peer-reviewed study of claims-based NDC mapping concluded that incomplete mappings are the norm rather than the exception (pmc.ncbi.nlm.nih.gov). A preprint analyzing Medicaid pharmacy claims reported **809,484,945**

claims and **126,604** unique NDCs ([arxiv.org](#)), and found that **59.4 percent** of unique NDCs were linked to an ATC class ([arxiv.org](#)). The preprint's peer-review status could not be confirmed, and its figures describe one dataset and one pipeline. They indicate that an unmatched NDC is a normal event to design for.

Case Studies and Real-World Examples

ProPublica's public linkage repository

ProPublica's Rx Inspector data repository is a documented example of a public-source join that records its uncertainty. The methodology filters the FDA product list to generics using ANDA numbers ([www.propublica.org](#)) and, for repackaged or relabeled drugs, traces back to the original label using the source NDC code ([www.propublica.org](#)). The repository documents a linkage method field, noting that some links come from address matching ([github.com](#)). Its README states that a three-segment NDC can be searched by the first two segments ([github.com](#)), a product-level key convention that matches the grain separation recommended here. It also warns that the data is not perfect ([github.com](#)), that the output is likely an undercount ([www.propublica.org](#)), and gives an update date of November 2025 ([github.com](#)). This report uses the repository only for its linkage design and draws no conclusions about any named establishment. For an NDC pipeline the transferable practices are concrete: keep the product-level key distinct from the package-level key, store how each link was made as a column rather than a footnote, tie every output to a dated source snapshot, and publish known gaps beside the data. The mapping template in the previous section applies the same discipline through its `pattern_basis` and `source_export_date` columns.

An implementation example

An AWS prescription-validation sample is useful as implementation context and not as an authority on FDA rules. The sample stores the medication name directly for quick access ([aws.amazon.com](#)) rather than keying on NDC, so it does not demonstrate NDC normalization at all, and its closing discussion points readers to public sources that offer free, standardized data but need more internal processing and maintenance ([aws.amazon.com](#)).

Implications and Future Directions

For data engineers

The long runway is not a reason to defer design. Systems should add a pattern column and a derived-form column now, because retrofitting the pattern after hyphens are lost is impossible for historical records. When FDA publishes its 10-to-12-digit mapping, it should be loaded as a reference table and compared with locally derived values as a test.

For regulatory operations

Listing teams do not need to resubmit listings only because of the format change, as FDA converts the listing files. Labeling and barcode teams face a different task: carriers that encode the NDC with application identifier 715 need camera-based scanners ([www.federalregister.gov](#)), and the UPC-A path is closed for the new format ([www.federalregister.gov](#)).

For reimbursement and market-access teams

Claims instructions in current CMS manuals describe 11-digit values (www.cms.gov), the state invoice definitions require zero-padded fields (www.medicaid.gov), and the State Drug Utilization Data are published at the 11-digit level (www.cms.hhs.gov). CMS notes that it may not publish a payment limit or crosswalk for all drugs that are reported by manufacturers (www.cms.gov), so a missing crosswalk row is not evidence of an invalid NDC. An NCPDP white paper observes that without an NDC11 on the claim the state cannot identify the product for rebate collection (www.ncdp.org). NCPDP reports that its standards already support the expansion (NCPDP); teams should track HIPAA adoption and confirm each payer's implementation date.

Migration checklist

- **Inventory:** list every system that stores, hashes or indexes an NDC, and its current format.
- **Preserve:** store raw text with hyphens; add the pattern; stop storing NDCs as integers.
- **Derive:** implement 11- and 12-digit generation from the raw code, with unit tests from the fixtures above.
- **Quarantine:** route reject and ambiguous cases to a review queue and count them per load.
- **Confirm:** ask each trading partner which format it will accept and when.
- **Labeling:** plan updates for March 7, 2033 through March 6, 2036; FDA warns of possible regulatory action for drugs labeled with a 10-digit NDC after March 7, 2036 (FDA). Track payer implementation dates separately.
- **Barcodes:** check scanners against the AI 715 requirement.
- **Document:** record the pattern basis and source date on every derived value.

Organizations that want an independent review of an NDC normalization design may consult a consultancy such as IntuitionLabs, which advises on life-sciences data and AI work; the choice of canonical key remains a business and compliance decision for the data owner, and regulatory or legal questions should go to qualified professionals.

Frequently Asked Questions (FAQs)

How is a 10-digit NDC converted to 11 digits?

Take the hyphenated code, identify the pattern, and add one leading zero to the short segment: the labeler for 4-4-2, the product for 5-3-2, or the package for 5-4-1. CMS rebate guidance gives the same three cases (www.medicaid.gov). Without hyphens, the conversion is unsafe.

Can an 11-digit NDC be converted back to 10 digits?

Only sometimes. A leading zero in the first digit implies a 4-digit labeler; leading zeros in both the product and package positions can yield two syntactic candidates. FDA's 12345-0678-09 example depends on the firm's configuration (FDA); those alternatives represent mutually exclusive configurations. Use the source pattern when available.

What is the 12-digit format and when does it apply?

It is a 6-4-2 structure effective March 7, 2033 (www.federalregister.gov). Existing codes convert by adding leading zeros.

Is the NDC the same as a UPC?

No. A GTIN-12 is built from a leading 3, the 10-digit NDC and a check digit (documents.gs1us.org), and a 12-digit NDC cannot be embedded in it (www.federalregister.gov).

Where can NDC directory data and an API be found?

FDA provides downloadable Directory files, and openFDA provides an API and bulk JSON (open.fda.gov), with rate limits that depend on whether a key is used.

How is an NDC mapped to RxNorm?

Normalize to 11 digits and query RxNorm, which stores normalized NDCs in that form (www.nlm.nih.gov); check the NDC status before relying on the result.

Does an NDC prove a drug is approved, covered or available?

No. FDA says Directory inclusion is not verification (www.fda.gov) and does not mean reimbursement eligibility.

Conclusion

NDC normalization is an exercise in preserving information. The 10-digit FDA code carries its segment boundaries in its hyphens, and every other common form either discards those boundaries (the 11-digit string and the GTIN) or standardizes them after the fact (the 12-digit form). Pipelines that keep the raw hyphenated code, record the pattern, treat product and package as separate grains, and derive other forms only at the edge can reproduce FDA's conversions and can explain every join they make.

The 2033 horizon is clear for FDA-assigned codes and uncertain for downstream formats. FDA has fixed the 6-4-2 structure and the date, and has stated an intention to publish a mapping, while payer-side standards follow separate processes that teams must track with their own trading partners. Until then, 10-digit, 11-digit and 12-digit forms will coexist in the same data estate, and the failure mode to engineer against is the plausible but wrong match.

Three practical rules summarize the report. Store text, never integers, and keep the hyphens. Treat any conversion that cannot be reversed with certainty as derived, and label its basis as observed, inferred or ambiguous. Describe every positive match as a listing match only: an NDC is a listing identifier, not a statement of approval, coverage or availability.

Source links

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

1. <https://open.fda.gov/apis/drug/ndc>

2. <https://www.federalregister.gov/documents/2026/03/05/2026-04368/revising-the-national-drug-code-format-and-drug-label-barcode-requirements>
3. <https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/clm104c25.pdf>
4. <https://www.fda.gov/drugs/development-approval-process-drugs/national-drug-code-database-background-information>
5. <https://www.nlm.nih.gov/research/umls/rxnorm/overview.html>
6. <https://www.govinfo.gov/content/pkg/FR-2026-03-05/pdf/2026-04368.pdf>
7. https://standards.ncdpd.org/Standards/media/Learn/Correspondence/2026/20260327_NCPDP-Letter-RE-Impending-FDA-NDC-Format-Expansion-and-Systemic-ImpactDC-Format-Expansion-and-Systemic-Impact.pdf
8. <https://www.fda.gov/media/191419/download?attachment>
9. <https://api.fda.gov/download.json>
10. <https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory>
11. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-C/part-207>
12. <https://www.fda.gov/drugs/drug-approvals-and-databases/ndc-product-file-definitions>
13. <https://www.fda.gov/drugs/drug-approvals-and-databases/ndc-package-file-definitions>
14. <https://www.fda.gov/media/84201/download>
15. <https://www.cms.gov/regulations-and-guidance/guidance/manuals/downloads/clm104c26.pdf>
16. https://nucc.org/images/stories/PDF/1500_claim_form_instruction_manual_2024_07-v12.pdf
17. <https://www.medicaid.gov/medicaid/prescription-drugs/state-drug-utilization-data/state-drug-utilization-data-faq>
18. <https://lhncbc.nlm.nih.gov/RxNav/APIs/api-RxNorm.getNDCStatus.html>
19. <https://documents.gs1us.org/adobe/assets/deliver/urn:aaid:aem:5b606108-5439-439f-a20a-a0081aeace07/Frequently-Asked-Questions-by-the-Pharmaceutical-Industry-in-Preparing-for-the-US-DSCSA.pdf>
20. <https://www.cms.gov/regulations-and-guidance/guidance/transmittals/downloads/r1496cp.pdf>
21. <https://www.medicaid.gov/sites/default/files/2023-02/ndra-ref-guide.pdf>
22. <https://www.govinfo.gov/content/pkg/FR-2026-03-05/html/2026-04368.htm>
23. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-C/part-207/section-207.33>
24. <https://www.fda.gov/drugs/electronic-drug-registration-and-listing-system-edrls/national-drug-code-format>
25. https://open.fda.gov/fields/drugndc_reference.pdf
26. <https://www.fda.gov/drugs/drug-registration-and-listing-instructions/how-request-labeler-code>
27. <https://www.nlm.nih.gov/research/umls/rxnorm/docs/techdoc.html>
28. <https://lhncbc.nlm.nih.gov/RxNav/APIs/api-RxNorm.getNDCs.html>
29. <https://lhncbc.nlm.nih.gov/RxNav/APIs/api-RxNorm.findRxcuiById.html>
30. <https://lhncbc.nlm.nih.gov/RxNav/APIs>
31. https://dailymed.nlm.nih.gov/dailymed/webservices-help/v2/ndcs_api.cfm
32. <https://dailymed.nlm.nih.gov/dailymed/app-support-web-services.cfm>
33. <https://open.fda.gov/apis/downloads>
34. <https://open.fda.gov/apis/authentication>
35. <https://open.fda.gov/apis/paging>
36. <https://open.fda.gov/apis/drug/ndc/how-to-use-the-endpoint>

37. <https://www.govinfo.gov/content/pkg/FR-2022-07-25/html/2022-15414.htm>
38. <https://pmc.ncbi.nlm.nih.gov/articles/PMC5001754>
39. <https://arxiv.org/html/2404.01514v1>
40. <https://www.propublica.org/article/rx-inspector-fda-generic-drug-tool-methodology>
41. <https://github.com/propublica/rx-inspector-data>
42. <https://aws.amazon.com/blogs/database/implement-prescription-validation-using-amazon-bedrock-and-amazon-dynamodb>
43. <https://www.medicaid.gov/medicaid/prescription-drugs/downloads/cms-r-144-state-invoice-data-definitions-jul-2021.pdf>
44. <https://www.cms.hhs.gov/files/document/medicaid-state-drug-utilization-data-field-descriptions.pdf>
45. <https://www.cms.gov/medicare/payment/part-b-drugs/asp-pricing-files>
46. <https://www.ncpdp.org/NCPDP/media/pdf/WhitePaper/Medicaid-Drug-Rebate-Program-Challenges-Across-the-Industry.pdf?ext=.pdf>

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.