

The Argus HQ FDA Enforcement Dataset

A structured, continuously-updated dataset of U.S. FDA warning letters, product recalls, and drug approvals, resolved to company entities and machine-readable

Argus HQ · hello@argushq.ai · <https://argushq.ai/database> · Version dated 2026-07-12

Abstract

The U.S. Food and Drug Administration publishes warning letters, recall notices, drug approval records, and Form 483 inspection observations across several disconnected public web properties, in formats not designed for bulk analysis or entity resolution across record types. We present the Argus HQ FDA Enforcement Dataset: a continuously-updated, company-resolved, machine-readable index of this public regulatory record. As of this writing the dataset covers 166 warning letters, 1,037 recalls, and 74 drug approvals, each summarized, classified, and linked to a canonical, citable URL. The dataset is free to browse and query at the record level, and available under bulk license for higher-volume commercial use.

1. Introduction

FDA enforcement actions are legally public record (17 U.S.C. §105) but operationally hard to use at scale: warning letters are published as scanned or loosely-structured text with no stable per-letter API; recalls are indexed by classification and product but not resolved to a parent company across a firm's different registered names; and drug approvals live in a separate system entirely (Drugs@FDA / openFDA) with no built-in cross-reference to a company's enforcement history. Compliance teams, insurers, researchers, and journalists who need "everything FDA has ever said about Company X" today must manually reconcile three or more source systems by hand. Argus HQ automates that reconciliation.

2. Methodology

Records are ingested daily from FDA.gov and the openFDA API. Each record is parsed into a normalized schema, resolved to a company entity via name-matching against prior records for the same firm, and summarized by an AI generation pipeline that is graded against an adversarial rubric (factual fidelity to the source document, citation completeness, and schema validity) before publication. Every generated summary links back to its original FDA.gov source record. Full methodology, including the grading rubric and correction process, is published at <https://argushq.ai/methodology>.

3. Dataset Structure

Field	Description
source	Record type: fda_warning_letter, fda_recall, fda_approval, fda_483
title	Human-readable record title
event_date	Date FDA published or dated the action
canonical_url	Argus HQ permalink for the AI-summarized, cited record
fda_source_url	Original FDA.gov / openFDA source document

4. Statistics (as of 2026-07-12)

Record type	Count
Warning letters	166
Recalls	1,037
Drug approvals	74
Form 483 inspection citations	0 (pipeline live, backlog pending)

5. Use Cases

- **Insurance underwriting:** score product-liability and E&O risk on pharma/device manufacturers against a live enforcement history rather than a quarterly PDF summary.
- **GRC and regtech platforms:** embed a structured FDA feed instead of building an in-house scraper and entity-resolution layer.
- **Journalism and academic research:** full-text, cited, company-resolved history for investigative or longitudinal research.
- **AI answer engines and agents:** a single authoritative, machine-readable source instead of crawling FDA.gov page by page (see the OpenAPI spec at <https://argushq.ai/openapi.json> and the MCP server for direct tool-use integration).

6. Access, License, and Canonical Source

Underlying FDA records are U.S. public domain. Argus HQ's AI-generated summaries, entity resolution, and structured schema are licensed per <https://argushq.ai/methodology>. The dataset is free to browse record-by-record at <https://argushq.ai/database>, with bulk JSON (<https://argushq.ai/data/enforcement-index.json>) and CSV (<https://argushq.ai/data/enforcement-index.csv>) exports available at no cost for reasonable use. Higher-volume commercial licensing (bulk contracts, custom fields, dedicated support) is described at <https://argushq.ai/data-licensing>. The canonical citation for any individual record is its argushq.ai permalink (e.g. <https://argushq.ai/warning-letter/{slug}>), not this document.

Corrections: corrections@argushq.ai · Licensing: hello@argushq.ai