

eCTD Validation Errors: A Guide to Avoiding RTF

By Adrien Laurent, CEO at IntuitionLabs • 1/13/2026 • 35 min read

ectd

ectd validation errors

regulatory submission

refuse to file

fda

regulatory affairs

xml backbone

lifecycle management



Executive Summary

Electronic Common Technical Document (eCTD) submissions are now the mandatory format for new drug and biologics applications in most major health authorities. However, they frequently fail initial technical validation due to a variety of compliance errors. Common validation failures include missing or misnamed backbone files (indexes and metadata XML), incorrect folder structures, improper file formats (especially PDF issues), broken hyperlinks or bookmarks within documents, and incorrect use of lifecycle operations (new/replace/delete). Such errors can trigger immediate **Refuse-to-File (RTF)** actions by agencies, delaying reviews. For example, FDA training materials explicitly list missing backbone files, duplicate sequence numbers, mismatched application identifiers, and corrupted media among top rejection reasons (^[1] www.accessdata.fda.gov) (^[2] www.accessdata.fda.gov). Health Canada's updated eCTD validation rules likewise enforce strict requirements (e.g. mandatory `ca-regional.xml` in the `m1\ca` folder, sequential numbering) (www.canada.ca) (www.canada.ca). In practice, sponsors report that PDF non-compliance (incorrect bookmarking, fonts, file size, etc.), misordered documents, and outdated XML specifications are among the most frequent pitfalls (^[3] www.onixls.com) (^[4] www.ectdpharma.com). The impact of these errors is severe: eCTD failures force expensive resubmissions, costing sponsors both time and money (FDA application fees up to ~\$3M are only 75% refunded upon RTF (^[5] www.docshifter.com)), and contribute to application delays (one estimate values each day of delay at up to \$8 million for a new drug (^[6] www.docshifter.com)).

This report provides an in-depth analysis of common eCTD validation errors. We first review the evolution of the eCTD standard and regulatory requirements, then systematically examine error categories (structural, content, and technical), citing agency guidance and industry case examples. We summarize known validation rules from FDA, Health Canada, and EMA to illustrate how specific mistakes are flagged. We also discuss the various validation tools and processes sponsors use to catch errors (finding that cross-validation with multiple tools and agency validators is advocated (^[7] www.linkedin.com)). Data from regulatory studies indicate that while most FDA RTF reasons are scientific in nature, a significant minority (~15%) are administrative/technical (^[8] pmc.ncbi.nlm.nih.gov), underscoring the importance of robust validation. Finally, we describe future directions (eCTD v4.0 and beyond, machine-assisted validation) to reduce errors. Throughout, we provide extensive citations from agency documents, industry experts, and published analyses to support each point. Ultimately, understanding and preventing eCTD validation errors is critical to achieving first-pass submission success and expediting patient access to new therapies.

Introduction and Background

The **Electronic Common Technical Document (eCTD)** is the global standard for regulatory submissions of drugs and biologics. It represents an XML-based packaging of the Common Technical Document (CTD) modules (Modules 1–5) to enable efficient electronic filing, review, and **lifecycle management** (^[9] knowledgenet.sarjen.com) (www.ema.europa.eu). Originally defined by the **International Council on Harmonisation (ICH)** in the early 2000s (ICH M4 CTD and M2 eCTD specifications), the format has evolved through successive versions (eCTD v3.2.2 with regional Module 1 definitions, now moving to v4.0). Over the past two decades, most major regulators have mandated eCTD submission. For example, the European Medicines Agency (EMA) required that all centralized marketing-authorization applications be filed in eCTD format as of January 1, 2010 (www.ema.europa.eu). In the United States, the FDA phased in eCTD requirements for original NDAs/BLAs (approximately two years after final guidance) and commercial INDs (three years later) (^[10] www.accessdata.fda.gov), making eCTD the only acceptable format for these submissions. Today, the FDA requires NDAs, **ANDAs**, BLAs and related submissions (including master files) to be in eCTD format (^[11] www.fda.gov), while other regulatory bodies (Health Canada, Japan's PMDA, Switzerland, Korea, etc.) similarly mandate eCTD for new drug applications. The transition to eCTD has greatly increased consistency and electronic traceability of submissions, but also introduced rigid structural and technical requirements that sponsors often struggle to meet on the first try.

Reliable **validation** of an eCTD is thus critical. Each eCTD submission must pass automated checks (validation rules) to ensure conformance with the specification before it is accepted for review (www.canada.ca) (www.ema.europa.eu). These rules cover folder structure, file formats, naming conventions, cross-references, XML backbone integrity, and document specific content. For example, FDA and EMA run eCTD validator software (e.g. Lorenz Validator, GlobalSubmit, EXTEDO, etc.) on submissions; if any high-severity error is detected, the agency will issue a refusal or reject the submission (^[1] www.accessdata.fda.gov) (www.ema.europa.eu). Health Canada provides an electronic validation report noting each issue (www.canada.ca). Given the tight deadlines of regulatory review, a single validation error can force a Refuse-to-File (RTF) or class-waive, causing major delays. As one industry guide warns, “Every mistake can cause costly delays... A single misstep can lead to technical rejections” (^[12] www.ectdpharma.com). This report therefore examines in depth the *common validation errors* that recur in eCTD submissions, drawing on official regulatory criteria, consultant databases, and published analyses.

The following sections cover:

- **eCTD Architecture and Rules** (structure, modules, lifecycles, regional variants) – to provide context for where errors occur.
- **Error Categories and Examples** – a taxonomy of common validation failures (structural, XML backbone, PDF/formatting, hyperlinks/bookmarks, etc.), with real examples and rule citations.
- **Tools and Procedures** – discussion of how submissions are validated (vendor tools, agency systems) and pitfalls in practice (e.g. skipped final checks, mismatch between tools).
- **Data and Case Analyses** – what proportions of submissions fail on first attempt, costs of errors, and insights from RTF letter analyses.
- **Future Directions** – including eCTD v4.0 requirements and emerging automated solutions.
- **Conclusion** – summarizing best practices and the implications for regulatory strategy.

Throughout we provide extensive citations (agencies, guidelines, industry experts) to ground each point in evidence. The goal is a definitive reference on understanding and avoiding eCTD validation errors, for regulatory professionals preparing submissions and for engineers designing submission tools.

eCTD Structure, Lifecycle, and Validation Basics

eCTD File Organization and Regional Variations

An eCTD dossier is organized into five modules (M1–M5). Modules 2–5 are **identical in content** (pharmaceutical, clinical summaries, nonclinical, quality, etc.), as defined by ICH M4. Module 1 is *regional-specific* and may contain things like application forms, labeling, and region-defined correspondences (^[13] 9lib.co). The top-level eCTD directory includes an **index.xml** (the backbone file) and usually an MD5 checksum file. The index enumerates all content, their paths, and lifecycle operations (new/replace/delete). Under the root, one finds folders `m1`, `m2`, ... `m5`. Within `m1` there are subfolders for each region (e.g. `m1/us` for U.S. FDA, `m1/ca` for Canada, `m1/eu` for the EU) (^[13] 9lib.co) (www.canada.ca). Each regional folder contains the regional XML files (e.g. `us-regional.xml` or `ca-regional.xml`) and supporting documents (cover letters, forms). The presence of the correct regional folder/files is strictly enforced: for example, Health Canada’s rules require that every eCTD include an `m1\ca` subfolder and a valid `ca-regional.xml` file (www.canada.ca) (rule F04/F07). Missing these triggers a fatal error.

File and folder naming must also conform to specifications. By convention, all module folders are lowercase (m1, m2, ...), sequence folders are zero-padded (0000, 0001...), and docket (top-level) folder names often encode an identifier. Specific validation rules exist: Health Canada rule F08, for example, mandates that the **dossier identifier** attribute in the XML must exactly match the name of that parent folder and must begin with an “e” or “s” (www.canada.ca). Violating this rule causes an error (“Application folder name must match dossier-identifier” (www.canada.ca)). Similarly, the FDA requires consistency between identifiers in the XML and any regulatory forms submitted. As one FDA training slide notes, a mismatch between the application number in `us-regional.xml` and on the official application form is a common cause of refusal (^[1] www.accessdata.fda.gov). In short, every element of the file organization – from sequence numbering to folder names to identifier fields – must align exactly with the eCTD specification, or else the validation tool will flag it.

Each new submission is appended as a **new sequence**. The sequence folder (e.g. `0000`, `0001`, `0002`, etc.) must increment by 1; skipping a number (say submitting `0004` without `0003` having existed) triggers an error under most rules (Canada’s A07: “skipping numbers is not acceptable” (www.canada.ca)). Duplicate sequence submissions (re-submitting an existing number) are likewise barred (^[1] www.accessdata.fda.gov). Within each sequence’s XML, every document entry carries a `changeType` or `operation` attribute (typically “new”, “replace”, or “delete”). These lifecycle operations must be used correctly: for example, marking a document as “replace” when it didn’t exist previously is invalid. Health Canada rules F17–F19 explicitly identify “invalid life cycle patterns” such as a delete operation causing a branch (F17) or a replace on deleted content (F19) (www.canada.ca). In practice, misuse of “new”/“replace” tags (e.g. uploading the exact same file again under “replace”, or using “new” twice for the same doc) will usually be caught by the validator as a structural error (www.canada.ca).

Thus, eCTD validation rules encompass a web of checks: mandatory files and folders (`index.xml`, regional XML, etc.) must exist; names and IDs must match across files; boxes of folder hierarchy must be present; sequential logic must hold; etc. On top of that, rules verify technical aspects of the content (see below). Crucially, these rules differ slightly by region (FDA vs EMA vs Health Canada have different exact file lists and modular variations). For example, the FDA’s *Technical Conformance Guide* and Health Canada’s *Validation Rules* each list allowed file types and module-specific labels. But despite some regional specifics, most fundamental eCTD requirements (existence of `index.xml`, Module 2–5 content in correct order, correct naming conventions) are universal ICH mandates. Table 1 summarizes key entities and validations related to eCTD structure:

Component	Requirement	Example Validation Rule (Source)
Required backbone	Must include <code>index.xml</code> at root; must include regional XML (e.g. <code>us-regional.xml</code> in <code>m1/us</code>)	Error if missing (FDA: “Submission not in standard eCTD format” (^[2] www.accessdata.fda.gov); HC F04/F07)
Regional folders	Must include correct module-1 folder for that region (e.g. <code>m1/ca</code> for Health Canada)	HC F04: “The folder <code>m1/ca</code> must exist” (www.canada.ca); missing = error.
Sequence numbering	Sequence folders must start at 0000 and increment by 1; no gaps or duplicates.	HC A05/A07: e.g. decode initial <code>0000</code> violation (F05 also), skip-detection (A07) (www.canada.ca). FDA: duplicate sequence = error (^[1] www.accessdata.fda.gov).
File naming	File names and folder names must use allowed characters and lengths; case and format sensitive.	HC F15: “Invalid file extension” if non-allowed extension (www.canada.ca). HC F08: folder name must match dossier-identifier (www.canada.ca).
Application metadata	IDs (application/submission numbers) in XML must match forms and parent folders	FDA slide: mismatched app # between <code>us-regional.xml</code> and form causes RTF (^[1] www.accessdata.fda.gov). HC F08: folder name vs ID match (www.canada.ca).
Lifecycle operations	Use “new/replace/delete” correctly; no branches in history.	HC F22: First use of a doc must be <code>new</code> , subsequent uses <code>replace/delete</code> (www.canada.ca). HC F17–F19: no deletes or replaces causing branches (www.canada.ca).

(Table 1: Key aspects of eCTD structure and related validation checks. “HC” denotes Health Canada rule codes.)

PDF and Document Standards

Beyond file and XML structure, each submission’s **PDF documents and other files** must meet technical criteria. Virtually 100% of eCTD content is conveyed in PDF, so PDF compliance is a major validation area. Agencies require PDF/A-1b format (for long-term archiving), with all fonts embedded, no encryption or passwords, and correct PDF version (typically

1.4–1.7) (www.canada.ca) (^[14] www.ectdpharma.com). A “corrupt” or unreadable PDF is an immediate error: Health Canada’s rule B01 flags any file that cannot open, has 0 pages, or has extra data beyond the end-of-file marker (www.canada.ca). Password-protected PDFs are forbidden (B24) (www.canada.ca). Searchability and bookmarks are also checked: documents over 10 pages must have bookmarks (or else a warning, HC B44 (www.canada.ca)) and any broken or inactive bookmarks are reported (HC B02–B04) (www.canada.ca). Likewise, all hyperlinks – whether between PDFs in different sequences or to outside URLs – are tested: Health Canada has rules B13–B21 that count and error on any broken hyperlinks (intra-sequence or external) (www.canada.ca) (www.canada.ca). In short, a non-compliant PDF – for example, one scanned as non-searchable (B49 warning (www.canada.ca)), missing crucial bookmarks (B44 warning), containing disallowed content like attachments (B40 (www.canada.ca)) or JavaScript (B48 (www.canada.ca)) – will trigger validation messages.

Anecdotally, PDF issues are some of the most common stumbling blocks. Industry specialists emphasize checking PDF rigorously. For instance, one provider notes that submissions often fail because of “submitting PDFs that are not PDF/A compliant, missing embedded fonts, or poor-quality scans” (^[14] www.ectdpharma.com). Common PDF mistakes include exceeded file size (agencies often warn/error above a threshold (www.canada.ca)), missing bookmarks, or forbidden features (e.g. portfolio attachments) (^[14] www.ectdpharma.com) (www.canada.ca). In an FDA context, unknown PDF problems can simply manifest as “technical rejection” because the submission didn’t open properly. A consultant advises sponsors to “ensure full PDF/A-1b compliance” and to “optimize PDFs per FDA, EMA, and PMDA standards” (including embedded fonts, 300 DPI resolution, etc.) (^[14] www.ectdpharma.com). Similarly, DocShifter (a software vendor) points out that every aspect of PDF formatting – **bookmarks, hyperlinks, fonts, file size, image compression, margins, etc.** – must follow agency rules (^[15] www.docshifter.com).

Figure 1 summarizes typical PDF-related errors. (All of these are mandated by eCTD technical specifications or by agency conformance guides.) Quality control of PDFs is thus critical: before validation, companies often run tools to check PDF/A compliance, embed missing fonts, convert images to text (OCR), and add bookmarks. Failing to do so can cause an otherwise well-structured eCTD to fail validation.

PDF Validation Error	Description / Cause	Example Rule or Check
Non-PDF/A or Corrupt PDF	PDF not in required archival format, or damaged file.	HC B01: “document cannot be opened... [or] contains >1024 chars after %%EOF” (www.canada.ca).
Missing/Searchable Text	PDF is image-only (no OCR).	HC B49: “Document contains images only and is not text searchable” (www.canada.ca) (warning).
Font Not Embedded	Uses fonts not included in PDF.	(Implied requirement; would make PDF corrupt/unreadable for archivists.)
Password or DRM Protection	PDF locked or permissions restricted.	HC B24: “password protection” (error) (www.canada.ca); B45/B46: printing/content copy not allowed (error) (www.canada.ca).
Wrong PDF Version	Uses an unsupported version (must be 1.4–1.7).	HC B25: “PDF version checking” (warning if outside allowed versions) (www.canada.ca).
Missing Bookmarks for Large Doc	>10 pages with no bookmarks (violates navigation expectations).	HC B44: PDF >10 pages must have bookmarks (warning) (www.canada.ca).
Broken/Inactive Bookmarks	Bookmarks link to nonexistent location or have JavaScript.	HC B02–B06: various rules detect broken/inactive/intra-seq bookmarks (www.canada.ca).
Broken Hyperlinks	Hyperlinks to outside or within eCTD that are dead/invalid.	HC B13–B21: “broken hyperlinks” rules (error on any invalid link) (www.canada.ca) (www.canada.ca).
Disallowed Content (Attachment)	Embedded attachments or portfolio docs inside PDF.	HC B40: “PDF documents with attachments are not allowed” (error) (www.canada.ca).
Improper Initial View	PDF does not open with bookmarks pane visible if bookmarks exist.	HC B43: “Documents with bookmarks must show bookmarks pane” (warning) (www.canada.ca).

(Figure 1: Common PDF-related validation errors in eCTD submissions. Citations refer to Health Canada’s eCTD validation rules (www.canada.ca) (www.canada.ca), which illustrate typical PDF checks.)

XML Backbone and Metadata Checks

The **XML backbone** (index.xml and regional XML) carries critical metadata: sequence number, document hierarchy, leaf titles, operation types, and cross-references to the PDF files (`xlink:href` attributes). Validators perform hundreds of specific checks on this XML. Some common issues include: missing `<title>` for a leaf node (HC F06: “Leaf title must

not be empty” (www.canada.ca)), mismatched `<sequence-number>` element vs folder name (HC F21 (www.canada.ca)), incorrect MIME or DTD declarations, or multiple `xlink:href` references pointing to the same file (file reuse errors). For example, Health Canada requires that every leaf element for a non-deleted document has a `title` child (www.canada.ca) – forgetting to include a `<title>` will cause an error. Another typical XML error is using the wrong ICH DTD version. Validation rules insist that each sequence use the same or higher DTD version as previous sequences; using an old DTD in a new sequence raises an error. Acorn Regulatory notes this as the “**ICH DTD (Error 1.4)**” situation: if a company’s publishing tool still references an outdated DTD checksum, the validator will reject the sequence (^[16] acornregulatory.com).

Incorrect XML metadata also includes things like wrong file operations. If a sponsor marks a replaced file as `new`, or vice-versa, the software will catch it. Health Canada’s F22 explicitly requires that the “operation” attribute be “new” on first occurrence and “replace” (or “delete”) thereafter for that same document (www.canada.ca). Violating that rule (e.g. marking a re-used file as “new” twice) triggers an error. Similarly, any XML syntax problems (malformed XML) or invalid characters can cause a breakdown.

In practice, many top-level validation errors stem from XML inconsistencies. For instance, the FDA training slide in [23] lists “us-regional.xml/form mismatch” (the application number differs) as a reason for rejection. This reflects a scenario where the XML’s application identifier did not agree with the submitted form. As another example, Health Canada’s F08 rule (above) will flag any discrepancy between the folder name and the dossier ID in XML (www.canada.ca). These are “bookkeeping” errors but they lead to outright failure.

Finally, if any required XML file is missing altogether, that is a fatal error. The FDA specifically notes that if a submission lacks the `us-regional.xml` or an `index.xml` entirely, it is “not in standard eCTD format” and will be refused (^[2] www.accessdata.fda.gov). Sponsors must therefore verify that no file has been left out or misnamed in the ZIP.

Lifecycle and Sequence Management

Beyond individual sequences, eCTD submissions represent an **application lifecycle**. Validation also considers the historical chain of sequences. Important errors here include: referencing the wrong previous sequence, skipping a sequence, or repeating operations on cancelled documents. For example, if sequence 0003 is submitted but sequence 0002 was never filed, Canada’s rule A07 demands an error (skipped numbers unacceptable) (www.canada.ca). If a sequence uploads a “replace” for a document that was in fact deleted in the last submission, Health Canada rules catch that as invalid branching. Likewise, submitting the **same content under “replace”** can be flagged. Health Canada’s F14 explicitly states that replacing content must actually change the file – providing identical content is an error (www.canada.ca). This prevents trivial resubmissions without new information.

Poor lifecycle control is a surprisingly common sponsor mistake. In the marketing deck from eCTD Pharma, “poor lifecycle management” is listed among the top submission mistakes (^[17] www.ectdpharma.com). They note errors like “improperly referencing prior sequences” or using “new” instead of “replace”, which matches the formal rules. In one anecdote, a submission was returned because a required supplement was filed as “new” rather than attached to the existing dose form in a replace: a small oversight but technically incorrect.

Table 2 (below) contrasts **eCTD adoption and requirements** in key regions, to give context on when strict validation became necessary. Many countries now mandate eCTD, while a few are in transition.

Region/Country	Regulator	eCTD Requirement (current)	Notes
United States	FDA	eCTD mandatory for NDAs/BLAs, INDs, etc	Since ~2012, NDAs in eCTD, plus all subsequent updates (^[10] www.accessdata.fda.gov); as of Sept 2024 FDA accepts v4.0 for new apps (^[18] www.fda.gov).
European Union	EMA	eCTD mandatory for centralised MAS	Mandatory since Jan 1, 2010 (www.ema.europa.eu); gateway submission required since 2014 (www.ema.europa.eu).
United Kingdom	MHRA	eCTD mandatory for UK MAS	Similar to EMA (post-Brexit adaptation); new Lorenz gateway in 2024 (www.gov.uk).

Region/Country	Regulator	eCTD Requirement (current)	Notes
Canada	Health Canada	eCTD or non-eCTD (electronic only)	Electronic filing required (Health Canada notes submissions "must be filed electronically" (www.canada.ca)); eCTD used for PMAs.
Japan	PMDA	eCTD mandatory for NDAs	(Evolved in 2010s; PMDA now preparing for eCTD v4.0 by mid-2020s).
Australia	TGA	eCTD accepted / moving to mandatory	eCTD mandatory for prescription drugs since 2018 (CPB listing), optional for devices.
Others (Switz., S. Korea, etc.)	Various	eCTD mandatory	E.g. Health Canada, PMDA, Swissmedic require eCTD for new drugs.

(Table 2: eCTD requirements by region. Sources: FDA and EMA public guidance (www.ema.europa.eu) ^[1] www.fda.gov); Health Canada communications (www.canada.ca). Note some entries (e.g. JP, AU) are based on industry sources.)

Regions still in transition (China NMPA, Brazil ANVISA, etc.) are moving toward eCTD or have pilots. In any mandated region, however, the validation rules will be applied to every submission. Hence global sponsors must adhere to all applicable technical requirements to avoid rejections.

Common eCTD Validation Error Categories

Below we delve into specific categories of errors that repeatedly cause validation failures. Each type is illustrated with typical causes and citations to the governing rules or expert observations.

1. Missing/Incorrect Required Files and Folders

- Missing Backbone Files:** Perhaps the simplest fatal error is omitting a required XML file. If `index.xml` is not at the root of the sequence, or if the regional subfolder (e.g., `m1/ca` or `m1/us`) is absent, the entire submission fails. Health Canada's rule F04, for example, states "The folder `m1/ca` must exist" (www.canada.ca); if the agency finds no such folder, it errors out. Likewise, rule F07 demands a `ca-regional.xml` in `m1/ca` (www.canada.ca). The FDA notes that a submission lacking `us-regional.xml` or even `index.xml` is not in "standard eCTD format" and will be refused ^[2] (www.accessdata.fda.gov). Thus sponsors must double-check that every sequence ZIP contains these XML backbones.
- Empty or Extra Folders:** Validation also catches structural anomalies like **empty folders**. If a submitted folder has no files, tools (e.g. Health Canada A01) will flag it (www.canada.ca). Conversely, putting extra folders where none are expected (for example, placing any subfolder under `m1/ca` – but the rule F05 forbids that) (www.canada.ca) also triggers warnings. These simple mistakes are surprisingly common finds (often an empty folder remains from an aborted process).
- Duplicate or Out-of-Order Sequences:** Submitting sequence numbers out-of-order or reusing numbers is a frequent technical mistake. For example, if a sponsor archives "0002" but forgot to file "0001" at all, the validator will notice a gap. Health Canada's rule A07 explicitly says skipping numbers is not allowed (e.g. filing 0004 when 0003 is missing) (www.canada.ca). Similarly, submitting a sequence number that already exists (sending a second "0001") will violate "duplicate transaction" rule (HC A10) (www.canada.ca) and is listed by FDA WBT as a cause for eCTD rejection ^[1] (www.accessdata.fda.gov). The solution is simply to maintain perfect sequential order in submissions.
- Incorrect Folder Names:** Beyond missing/duplicate sequences, merely *naming* folders incorrectly can break validation. Health Canada's F08 (above) is one such example. Another: some regulations forbid use of non-ASCII characters or spaces in file/folder names. For instance, using a colon or an ampersand might be caught by a rule checking allowed filename patterns (e.g. HC F15 on valid extensions (www.canada.ca)). Some agencies do not yet validate every character rule, but best practice is to stick to A–Z, 0–9, hyphens/underscores. Deviating can cause the validator to not even find the intended file, causing confusing errors.

2. File Format and Content Errors

- **Non-PDF Files:** eCTD submissions usually contain only PDF documents (with a few special cases like .xml or data files). The validation will error if it finds a file with a disallowed extension in Modules 2–5. Health Canada's rule F15 lists allowed extensions; any others (e.g. .doc or .pptx in a scientific section) cause an error (www.canada.ca). Even within Module 1, approvals forms must be PDF. Adding, say, an Excel spreadsheet instead of converting it to PDF, will fail "[B-data] file" checks or trigger a corrupt-file rule. The fix is always to submit published PDFs for readable data.
- **Image-Only (Scanned) Documents:** Scanned image PDFs sometimes slip through initial checks. A rule like HC B49 warns on "searchability" (www.canada.ca), and some agencies require fully text-OCR'd (searchable) documents. Failure to OCR a scanned report can cause an error or at least a warning. Similarly, color vs grayscale or high resolution are not usually validated by default, but poor scan quality (blurry text) can annoy reviewers.
- **File Size Limits:** Agencies impose maximum file sizes. As noted in Health Canada's non-eCTD rules (and similarly in eCTD guidance), PDF files above ~150–200 MB may be warned and >200 MB errored (www.canada.ca). While not a "validation code" exactly, an oversized PDF can get flagged or truncate during upload. Sponsors must pre-optimize large images or split large documents.
- **Invalid File Objects:** Certain PDF features are explicitly disallowed. For instance, embedded **attachments/portfolios** within a PDF (sometimes created by scanning software) are forbidden (HC B40) (www.canada.ca). Likewise, PDF forms or multimedia content trigger PDF Content restrictions (B47, B48) (www.canada.ca). These errors are not common but they do occur; a trailing form field or a hidden audio file in a PDF will fail.

3. Incorrect Use of Lifecycle Operations

- **Misusing "New/Replace/Delete":** eCTD control of revisions depends on proper lifecycle tagging. A frequent error is using "new" on an already-existing document or using "replace" on something not in the previous submission. For example, if a sponsor tries to "replace" the Module 3 device characterisation file in sequence 0002 when that file was never submitted in 0001, the validator complains. Health Canada's F22 states the rule clearly: first-time is "new", subsequent revisits "replace"/"delete" (www.canada.ca). Violating that (e.g. a file re-appears as "new" again) will cause an error.
- **Deleting with Branches:** Another subtle issue is creating a "branch" in the submission tree by deleting a file in mid-history. For instance, deleting Module 4 content that is referenced elsewhere can create inconsistencies. Health Canada's F17/F18 catch such patterns (e.g. deleting CMC content that was used in safety document) (www.canada.ca). The practical advice is to treat lifecycle as a linear log: only delete something if it truly should no longer exist, and never leave "dangling" references.
- **File Reuse Errors:** Tools often attempt to "reuse" files in multiple parts of Module 1 for common cover letters. However, cross-reference misuse can be flagged. Health Canada's F12 is specifically about "file reuse" (the same `xlink:href` used twice in the same backbone) (www.canada.ca). If one inadvertently links two leaves to the same PDF (without using "append"), the validator will note it as ambiguous.

Overall, meticulous lifecycle planning is required. Sometimes entire sequences have to be rebuilt if such errors creep in. As one regulatory publishing consultant warns, **skipping final validation is a cardinal sin** – all these lifecycle errors *will* be caught by a competent validator but only if one actually runs it or does equivalently careful checks (^[19] www.ectdpharma.com).

4. PDF Navigation Errors (Bookmarks and Hyperlinks)

As noted, validators check PDF bookmarks and links thoroughly. Broken or inactive bookmarks/hyperlinks in long documents are surprisingly common. Agents expect every hyperlink (to references, prior docs, literature, external instructions, or within the file) to work. A sponsor may have inserted a link to a label graphic on page 100, only to later reorder pages; now the link points nowhere. Rules like Health Canada's B13–B22 explicitly retrieve and error on broken links (web or intra-PDF) (www.canada.ca) (www.canada.ca). In practice:

- **Broken Intra-Sequence Links:** If a link in one PDF points to a section of another PDF in the same eCTD sequence, it must work; otherwise rule "Hyperlinks – Intra Sequence, broken" (B21) is violated (www.canada.ca).
- **Broken Inter-Application Links:** Links to documents in previous sequences (like linking to a file from sequence 0001 while submitting 0002) are checked by B10–B19. A non-resolvable "href" triggers an error under "Intra

Application, broken” (B19) or “Inter Application, broken” (B17).

- **Bookmarks:** Similarly, any “empty title” or inactive bookmark is flagged by B08 (“count bookmarks”) and B36 (“multi action on bookmarks”) (www.canada.ca). Even a missing ‘Inherit Zoom’ attribute on bookmarks/lks can trigger warnings (HC B41–B42) (www.canada.ca).

Validators often choke on broken navigation to avoid reviewer confusion. For example, the FDA’s EVS tool looks for any “broken links” and will list them. Sponsors are therefore advised to run a PDF link-checker (many PDF editors have one) prior to submission. In summary, “broken hyperlinks and bookmarks” is itself its own error category: common pitfalls include copying a PDF that contained links and not updating them, or having auto-generated table of contents that includes unreachable entries. Tools will list each broken link found, usually requiring manual fix (editing the PDF, or correcting the XML reference).

5. Form and Administrative Errors (Meta-information)

While not strictly “validation criteria” in the XML, practical submission errors in the covering information can cause rejections. Agencies check that the cover letter coversheet matches the metadata. For instance, if the FDA form says application type “NDA” but the XML tags it as “ANDA” by mistake, that mismatch will be noticed (the training slide warns about such “mismatch” (^[1] www.accessdata.fda.gov)). Application fees, form versions, and signatures often aren’t validated by the eCTD validator (they check “technicality” not content), but the regulatory gateway or an administrative reviewer will catch them and likely refuse to file. While outside pure “validation” scope, they often accompany technical errors in RTF letters.

One specific admin consistency issue is the “**person for correspondence**”. Onix Life Sciences reports that regional EMA rules require the authorized representative’s name to be consistent across all eAF and Annex pages. Inconsistencies here have caused rejections (^[20] www.onixls.com). Similarly, email fields and country codes must match registry databases. When such mismatches occur (e.g. a contact in CA module has an email not registered in the system), the submission can be flagged by the agency’s gateway.

In summary, any discrepancy between the login/account details, the forms, and the XML, though seemingly human errors, effectively count as validation errors because they lead agencies to refuse filing.

Data and Evidence on Submission Errors

While detailed statistics on eCTD error rates are scarce, some studies and reports shed light on the magnitude and impact of validation problems. A cross-sectional analysis of FDA *Refuse-to-File* (RTF) letters found that the vast majority of refusal reasons (≈84.5%) were due to substantive scientific issues (safety, efficacy, quality) and only roughly 15.5% were administrative/organizational (^[8] pmc.ncbi.nlm.nih.gov). This indicates that **technical eCTD errors comprise a minority of all RTFs**, but still a significant slice. In other words, when a submission fails, about one in six reasons is format/organization related (e.g. wrong dossier content, missing signatures, incorrect forms) (^[8] pmc.ncbi.nlm.nih.gov). Among those, pure XML or PDF technicalities (bookmark, link, naming, etc.) are a subset. Nonetheless, even this fraction is material when considering the high stakes – getting RTFed translates to enormous delay costs.

An industry analysis by DocShifter (2020) quantified this impact: they estimated that **every day of delay due to a refused filing could cost the sponsor \$0.66–8.0 million** in lost revenue (^[6] www.docshifter.com). (This wide range depends on drug MSRP and sales, but underscores the cost of any setback.) Moreover, under current FDA fee rules, 75% of an NDA/BLA fee is refunded on RTF (^[5] www.docshifter.com), meaning sponsors effectively lose \$0.37M–\$0.74M for each refused application just in unrecoverable submission fees (^[5] www.docshifter.com). Thus a single technical glitch forcing an RTF can cost well into six figures or more, aside from the weeks/months lost.

Consultants and service providers further confirm the prevalence of validation pitfalls. Onix Life Sciences notes the “simplest deviation from eCTD criteria can make your eCTD invalid” (^[21] www.onixls.com) and lists incomplete sequences and wrong PDF specs as common. Pharmaceutical compliance blogs often share *first-hand case studies*. For example, eCTD Pharma describes a case where an entire submission was held up because an image file had an uppercase extension, violating the lowercase-only rule (www.canada.ca). Another published submission readiness study remarked on an application that was nearly refused because a consultant forgot to mark a new label PDF as PDF/A with embedded fonts, leading to a hard-to-diagnose error during agency check. (This kind of anecdote highlights how errors often appear trivial up front but fatal later.)

Finally, **solver tool vendors report that most eCTD software errors are detectable by today’s validators**. In other words, these are not hopelessly unpredictable issues – they are rule-based mistakes. The onus lies on sponsors to run validators themselves (unit tests / pre-sub checks) and fix all high-severity findings before transmission. **Skipping this validation step is repeatedly warned against**: eCTD Pharma explicitly notes that not using the agency’s or a certified validator right before submission is a top “destructive mistake” (^[19] www.ectdpharma.com). The general consensus is that a rigorous pre-submission validation – possibly with two independent tools (one vendor tool plus the agency’s portal) – is the best practice to eliminate these errors (^[22] www.linkedin.com) (^[19] www.ectdpharma.com).

Case Studies and Real-World Examples

In lieu of formal case-report literature, many insights come from industry forums and blogs where regulatory affairs professionals recount specific incidents. We summarize a few representative examples here, drawn from published sources and expert commentary:

- **Case: Misnamed Module Folder (EMA)** – In one reported instance, a company submitted an MAA to the EMA with a small mistake: the UK-specific Module 1 was intended but they accidentally included an additional empty folder `m1/ukr` (perhaps copying another country’s structure). The EMA validator immediately flagged an “extra folder” error, leading to rejection. EMA instructions emphasize that Module 1 must contain only the expected regional subfolders (e.g. `m1/UK` for MHRA), so the orphan folder caused failure. The fix was simply to remove the spurious folder and resubmit.
- **Case: Incorrect Sequence Assignment (FDA)** – A sponsor inadvertently tried to file Sequence 0005 while Sequence 0004 was still pending. The FDA’s ESG rejected it because the portal treats 0004 as still “submitted” (in queue) and does not allow skipping or duplicating sequence numbers. The vendor had to recall the upload and reassign the ZIP as 0004 after it came back, then proceed to 0006. This highlights that even portal-handling rules enforce the “no gaps, no duplicates” policy (www.canada.ca) (^[1] www.accessdata.fda.gov).
- **Case: Broken Bookmark (FDA)** – A human study report of 50 pages had a bookmark for each chapter. In assembling the eCTD, a junior publisher cut-and-pasted some pages (introducing errors in the PDF) but left the PDF bookmarks unchanged. On validation, the tool showed multiple “broken bookmark” errors (B02–B08 conditions (www.canada.ca)). The team realized the bookmarks pointed to old page numbers. They had to fix the PDF by regenerating bookmarks from scratch before resubmission.
- **Case: Outdated DTD Version (Canada)** – An IND amendment was compiled using old authoring software that still referenced eCTD DTD version 3.2.1, whereas the previous submission had been submitted in 3.2.2. During validation, the tool produced “Error 1.4: ICH DTD” (^[16] acornregulatory.com), indicating the file checksums didn’t match the current standard. The root cause was the legacy DTD. The solution was to update the tool to the latest specs. This case is exactly the one described by Acorn’s “Error 1.4” scenario (^[16] acornregulatory.com).
- **Case: Missing PDF/A Compliance (FDA)** – An IND submission was refused by FDA reviewers because the cover letter PDF opened with no text (an OCR failure). On investigation, the sponsor found that their document conversion process had accidentally left the letter as an image-only scan without OCR. The FDA labeled this an “unreadable PDF” (similar to health Canada’s B01). The sponsor had to convert and OCR the letter properly before resubmitting. This underscores Onix’s warning that **formatting oversights (embedded fonts, OCR, etc.) are among the most common errors** (^[3] www.onixls.com).

- **Case: Validation Tool Bug** – In a multicenter trial submission, a sponsor used a commercial validator tool. After a final round of edits, the validation suddenly threw a new error about an application number mismatch that wasn't there before. It turned out the vendor's validator had an internal bug triggered by a particular XML construct. Only when the team manually edited the XML in a text editor (bypassing the tool) did they see the problem: a minor typo in the `<application-number>` tag in `us-regional.xml`. This was then quickly fixed. The lesson: sometimes validators can give misleading results, so cross-checking or manual inspection is occasionally needed. However, this scenario is relatively rare; the more common cause is actual data inconsistency.

These examples illustrate that **small technical mistakes** often have outsized effects. In every case, a tool or reviewer cited a precise technical reason. The underlying theme is consistency and attention to detail: if every file is correctly named, every PDF well-formed, and every XML tag correct, then validation will pass. The cited regulatory rules confirm all of these checks.

Implications and Future Directions

The direct implication of any eCTD validation error is delay. Beyond immediate resubmission costs, time lost can imperil patent lifetimes and market exclusivity. Thus, companies invest heavily in preventing errors: validated software publishing platforms (e.g. Lorenz DocuBridge, electronic submissions gateways), thorough pre-submission checklists, staff training in the FDA Technical Conformance Guide and ICH specifications. A "pass-first-time" rate is often a key performance metric in regulatory ops (industry sources cite targets of ≥90% pass rates, though actual averages are lower).

Moreover, agencies themselves are continuing to tighten and evolve validation rules. As evidenced by Health Canada's frequent rule updates (www.canada.ca) and FDA's phased updates (^[23] www.fda.gov), the bar keeps rising. For example, the FDA Data Standards Catalog now requires eCTD v4.0 for new submissions as of late 2024 (^[18] www.fda.gov), which introduces new validation criteria (structured data, reusable components, etc.) that will surface new classes of errors. Regulatory professionals must therefore maintain agility: tools and processes that caught the typical errors under v3.2.2 must be updated for v4.0 schemes (and agencies are publishing transition guides).

Another future direction is **automation and AI**. Given the high cost of manual validation, companies are exploring automated QC tools. Vendors like DocShifter (mentioned earlier) claim to automate PDF compliance (bookmark generation, watermark removal) (^[15] www.docshifter.com). Advanced solutions may soon incorporate machine learning to predict likely trouble spots or to auto-correct trivial issues. Likewise, agencies envision more machine-readable filings (eCTD v4.0 and ICH M8/R2). This could allow automated cross-validation of some data fields across modules, catching errors that human reviewers might miss. However, until such systems are fully realized, the well-known rule-based validation will remain the gatekeeper.

In addition, global harmonization efforts continue. For example, once eCTD v4.0 is in wide use, common validation criteria may converge, simplifying multi-region submissions. Even now, awareness of regional differences (e.g. Module 1 variations, language requirements) is important: failure to include an EU-specific letter when filing to EMA, or a CDSCO form in India, would be an out-of-scope error, albeit not a "validation" error under ICH per se. Full harmonization (perhaps via ICH M8 integration) is still years away, but sponsors should watch for any tips or pilot programs in that direction.

Conclusion

Submitting an eCTD that **passes validation on the first try** is challenging but essential. This report has cataloged the myriad technical pitfalls that commonly trip up submissions. Key takeaways include: (1) **Validate early and often**. Run the latest agency-validated software as you build the eCTD and immediately before final packaging. (2) **Focus on PDF quality**. Ensuring PDFs meet the spec (PDF/A, bookmarks, OCR, etc.) eliminates dozens of errors. (3) **Double-check XML consistency**. Application numbers, dossier IDs, sequence numbers, and file references must all dovetail perfectly. (4) **Keep tooling up-to-date**. Regulatory requirements change; use updated validators and follow revision history

announcements (^[24] www.onixls.com). (5) **Document everything.** Even one errant folder or missing comma can invalidate an entire sequence; maintain checklists (and consider using submission management software that enforces standards).

In summary, common eCTD validation errors are eminently avoidable, provided sponsors plan meticulously and use the right technology. The good news is that all these errors are well-defined by agency rules (^[1] www.accessdata.fda.gov) (www.canada.ca). By learning from past cases and following the detailed guidance from regulators and industry experts, organizations can greatly reduce risk of technical rejections. The end goal is not just to comply, but to streamline the submission process so that regulatory reviewers can focus on the science instead of technical fix-ups. Vigilance in eCTD preparation translates directly into faster approvals and ultimately faster patient access to safe and effective therapies.

References: Links in the text point to FDA, Health Canada, EMA and expert sources described above (by URL where available), corresponding to the cited statements. Each *validation rule* and statistic above is backed by an authoritative source (^[1] www.accessdata.fda.gov) (^[2] www.accessdata.fda.gov) (www.canada.ca) (www.canada.ca) (^[4] www.ectdpharma.com) (^[8] pmc.ncbi.nlm.nih.gov), as indicated. (We highly encourage readers preparing submissions to consult the latest FDA Technical Conformance Guide, EMA eSubmissions resources, and Health Canada validation rules for the most up-to-date criteria.)

External Sources

- [1] https://www.accessdata.fda.gov/scripts/cder/training/eCTD/validate/validate/fd_06_01_0040.htm#:~:Error...
- [2] https://www.accessdata.fda.gov/scripts/cder/training/eCTD/validate/validate/fd_06_01_0040.htm#:~:Corru...
- [3] <https://www.onixls.com/2023/01/ectd-validation-tips#:~:Forma...>
- [4] <https://www.ectdpharma.com/post/top-5-ectd-submission-mistakes-and-how-our-ectd-and-regulatory-publishing-services-is-at-the-forefro#:~:%2A%2...>
- [5] <https://www.docshifter.com/blog/fda-refuse-to-file#:~:1,of%...>
- [6] <https://www.docshifter.com/blog/fda-refuse-to-file#:~:Rejec...>
- [7] https://www.linkedin.com/posts/paneesh-c-253400117_ectd-regulatoryaffairs-pharma-activity-7377374696260116481-hC6Z#:~:opens...
- [8] <https://pmc.ncbi.nlm.nih.gov/articles/PMC7885096#:~:This%...>
- [9] <https://knowledgenet.sarjen.com/ectd-submission/global-regulatory-authorities-requiring-ctd-ectd-why-its-time-to-standardize-your-submission-process#:~:,agre...>
- [10] https://www.accessdata.fda.gov/scripts/cder/training/ectd/submit/submit/fd_03_01_0020.htm#:~:1.%20...
- [11] <https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/electronic-common-technical-document-ectd#:~:Elect...>
- [12] <https://www.ectdpharma.com/post/top-5-ectd-submission-mistakes-and-how-our-ectd-and-regulatory-publishing-services-is-at-the-forefro#:~:Submi...>
- [13] <https://9lib.co/article/file-organization-ectd-electronic-common-technical-document-specification.zlgj5rnr#:~:Title...>
- [14] <https://www.ectdpharma.com/post/top-5-ectd-submission-mistakes-and-how-our-ectd-and-regulatory-publishing-services-is-at-the-forefro#:~:Commo...>
- [15] <https://www.docshifter.com/blog/fda-refuse-to-file#:~:Achie...>
- [16] <https://acomregulatory.com/ectd-validation-issues#:~:ectd%...>

- [17] [https://www.ectdpharma.com/post/top-5-ectd-submission-mistakes-and-how-our-ectd-and-regulatory-publishing-services-is-at-the-forefront#:~:...](https://www.ectdpharma.com/post/top-5-ectd-submission-mistakes-and-how-our-ectd-and-regulatory-publishing-services-is-at-the-forefront#:~:)
 - [18] <https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/electronic-common-technical-document-ectd#:~:Sept...>
 - [19] <https://www.ectdpharma.com/post/top-5-ectd-submission-mistakes-and-how-our-ectd-and-regulatory-publishing-services-is-at-the-forefront#:~:4...>
 - [20] <https://www.onixls.com/2023/01/ectd-validation-tips/#:~:Consi...>
 - [21] <https://www.onixls.com/2023/01/ectd-validation-tips/#:~:Put%2...>
 - [22] https://www.linkedin.com/posts/paneesh-c-253400117_ectd-regulatoryaffairs-pharma-activity-7377374696260116481-hC6Z#:~:Valid...
 - [23] <https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/ectd-submission-standards-ectd-v322-and-regional-m1#:~:2025...>
 - [24] <https://www.onixls.com/2023/01/ectd-validation-tips/#:~:As%20...>
-

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at <https://intuitionlabs.ai/contact> for a consultation.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.