RFC-019: Documentation Pipeline

Status: Implemented **Date:** January 2026 **Author:** Derrell Piper ddp@eludom.net **Implementation:** generate-rfcs.sh

---

Abstract

This RFC specifies the documentation pipeline for the Library of Cyberspace: automated generation of canonical document formats from both Markdown and LaTeX sources, index catalogs, and future syndication feeds.

Motivation

Documentation must be:

1. **Preserved** — Multiple formats for long-term archival
2. **Accessible** — Viewable in any environment
3. **Discoverable** — Indexed for navigation
4. **Syndicated** — Subscribable for updates (future)

The pipeline automates generation of all canonical formats using the right tool for each source:

- **Markdown** → pandoc → prose documentation, RFCs
- **LaTeX** → pdflatex/latexmlc → mathematics, proofs, research papers

Computer science is math. Use the right tool for the job.

Source Formats

Format	Extension	Use Case	Pipeline
Markdown	.md	Prose, docs, RFCs	pandoc
LaTeX	.tex	Math, proofs, papers	pdflatex + latexmlc

Output Formats

Format	Extension	Purpose	From MD	From TeX
Plain Text	.txt	IETF tradition, universal	pandoc	—
HTML	.html	Web viewing	pandoc	latexmlc
PDF	.pdf	Archival, printing	xelatex	pdflatex

Plain text is not generated from LaTeX sources—math doesn't render in plaintext.

Philosophy: Text, HTML, and PDF are easy to generate and cover all use cases. Fewer formats means less wasted conversion—these three are canonical, don't add more.

Pipeline Specification

Input

Source files following the naming convention:

```
rfc-NNN-short-name.md     # Markdown source
rfc-NNN-short-name.tex    # LaTeX source
```

Where: – `NNN` – Zero-padded RFC number (000–999) – `short-name` – Lowercase, hyphenated descriptive name

The pipeline auto-detects source format by extension.

Output

For Markdown sources:

```
rfc-NNN-short-name.html    # Standalone HTML (pandoc)
rfc-NNN-short-name.pdf     # PDF (xelatex)
rfc-NNN-short-name.txt     # Plain text, 78 columns
```

For LaTeX sources:

```
rfc-NNN-short-name.html    # HTML (latexmlc)
rfc-NNN-short-name.pdf     # PDF (pdflatex)
```

Plus a navigational index:

```
index.html                 # Hypertext catalog
```

Generation Commands

Markdown Pipeline (pandoc)

```
# HTML (standalone, no external dependencies)
pandoc ${doc}.md -o ${doc}.html --standalone --metadata title=""

# PDF (XeLaTeX with monospace font for code)
pandoc ${doc}.md -o ${doc}.pdf --pdf-engine=xelatex -V mainfont="Menlo"

# Plain text (IETF-style, 78 columns)
pandoc ${doc}.md -o ${doc}.txt --to=plain --wrap=auto --columns=78
```

LaTeX Pipeline (pdflatex + latexmlc)

```
# PDF (native LaTeX — what it was made for)
pdflatex -interaction=nonstopmode ${doc}.tex
pdflatex -interaction=nonstopmode ${doc}.tex   # twice for refs

# HTML (LaTeXML — proper math rendering, used by arXiv)
latexmlc --dest=${doc}.html ${doc}.tex
```

Index Generation

The index.html catalog provides:

- RFC number and title
- Links to all four formats
- Clean, accessible HTML

Structure:

```html
<table>
  <tr>
    <td>RFC Number</td>
    <td>Title</td>
    <td>html | pdf | txt | md</td>
  </tr>
</table>
```

Publication

Local Workflow

```
# Generate all formats
./generate-rfcs.sh

# Commit to vault
seal-commit "Regenerate RFC documentation"
```

Remote Publication

Web export includes only rendered outputs—no source files:

```
# Publish to web server (HTML + PDF only)
rsync -avz --chmod=D755,F644 \
  --include='*.html' --include='*.pdf' \
  --exclude='*.md' --exclude='*.tex' --exclude='*.txt' --exclude='*.sh' \
  docs/rfc/ user@server:~/path/to/web/
```

Web export: – `*.html` – Web viewing – `*.pdf` – Download/print – `index.html` – Catalog

Stays in repo: – `*.md` – Markdown source – `*.tex` – LaTeX source – `*.txt` – Plain text (IETF tradition, always generated) – `generate-rfcs.sh` – Build script

Permission model: – Directories: 755 (world-readable, owner-writable) – Files: 644 (world-readable, owner-writable)

Future: Syndication

RSS/Atom Feeds

Future versions will generate:

```
rfc-feed.xml    # Atom feed of RFC updates
```

Feed entries will include: – RFC number and title – Publication/update date – Abstract – Links to all formats

Subscription Model

```
;; Subscribe to RFC feed
(seal-subscribe "https://example.com/cyberspace/rfc-feed.xml"
  verify-key: publisher-public)
```

Integration with Vault subscription system (RFC-006).

Implementation

generate-rfcs.sh

```
#!/bin/bash
# RFC Documentation Pipeline
```

```
RFCS=(rfc-000-declaration rfc-001-replication-layer ...)

for rfc in "${RFCS[@]}"; do
  pandoc "${rfc}.md" -o "${rfc}.html" --standalone
  pandoc "${rfc}.md" -o "${rfc}.pdf" --pdf-engine=xelatex -V mainfont="Menlo"
  pandoc "${rfc}.md" -o "${rfc}.txt" --to=plain --columns=78
done

# Generate index.html
generate_index
```

Dependencies

Tool	Version	Purpose
pandoc	2.x+	Markdown → HTML/PDF/TXT
pdflatex	TeX Live	LaTeX → PDF
xelatex	TeX Live	Markdown → PDF (via pandoc)
latexmlc	LaTeXML	LaTeX → HTML (optional)
rsync	3.x+	Publication

Note: latexmlc is optional. If not installed, LaTeX sources produce PDF only.

Security Considerations

Integrity

Generated documents inherit integrity from: – Git version control (source) – Vault signatures (releases)

Publication

Remote publication uses: – SSH key authentication – No sensitive data in documents – World-readable permissions only

---

References

1. Pandoc User's Guide

2. LaTeXML — LaTeX to XML/HTML converter
3. RFC-006 — Vault System Architecture
4. Atom Syndication Format — RFC 4287

Changelog

- **2026-01-07** — Added LaTeX pipeline for math/proofs papers
- **2026-01-06** — Initial specification

Implementation Status: Complete **Script:** generate-rfcs.sh
Source Formats: Markdown (.md), LaTeX (.tex) **Output Formats:** HTML, PDF, Plain Text (MD only)