

Distributed Multi-Region CDX Playback & Link-Graph Continuity Benchmark (CC BY 4.0)

WayForge Technical Report WF-TR-2026-03 | High-Throughput Archival Retrieval & 301 Link-Equity Preservation

WayForge Official Platform & Cloud Archival Engine: <https://wayforge.org>

Open Academic Research Hub & Benchmark Dataset (CC BY 4.0): <https://wayforge.org/research>

Technical Documentation & Architecture Reference: <https://wayforge.org/docs>

Single-Domain Deep Archival Reconstructor (/solo): <https://wayforge.org/solo>

Multi-Region Batch CDX Fleet Engine (/batch): <https://wayforge.org/batch>

Multi-Tier Link Graph & Topology Architect (/pbn): <https://wayforge.org/pbn>

Free Archival & Web Preservation Utilities (/tools): <https://wayforge.org/tools>

Download Benchmark CSV Dataset (CC BY 4.0): <https://wayforge.org/api/research/dataset.csv>

Download BibTeX Citation Record (.bib): <https://wayforge.org/api/research/cite.bib>

1. Multi-Region CDX Playback & Sparkline Status Map Scoring

Naive archival downloaders select the most recent capture timestamp in the Wayback Machine, which for expired domains is frequently a domain-parking placeholder (Sedo, HugeDomains, GoDaddy), a registrar suspension page, or an HTTP 301/502 redirect shell. WayForge queries the __wb/sparkline HTTP status map across 1996-2026 to isolate peak 200-OK content eras and scores candidate HTML captures against a 40-signal anti-parking classifier.

2. Standalone Universal Packaging (.htaccess, Nginx & Edge Worker)

Every reconstructed archival bundle generated by WayForge (<https://wayforge.org/docs/self-hosting>) exports a zero-dependency static HTML tree paired with Apache .htaccess, Nginx nginx.vhost.conf, Cloudflare Pages _worker.js, XML sitemap, and robots.txt rules to preserve historical URL slugs without 404 link rot.

Cohort #9: DataLife Engine (DLE) Media Portal (2010-2021) - CSS Recovery: 96% | Median Latency: 4100ms

Cohort #10: MODX Revolution Regional News (2015-2024) - CSS Recovery: 98% | Median Latency: 4300ms

Cohort #11: WordPress + GeneratePress Blog (2018-2025) - CSS Recovery: 99% | Median Latency: 3600ms

Cohort #12: WordPress + OceanWP E-Commerce (2019-2024) - CSS Recovery: 98% | Median Latency: 4750ms

Cohort #13: Static Bootstrap 4/5 Landing Hub (2017-2025) - CSS Recovery: 99% | Median Latency: 2950ms

Cohort #14: Legacy KOI8-R Academic Archive (1999-2009) - CSS Recovery: 95% | Median Latency: 3100ms

Cohort #15: Shift_JIS / EUC-JP Regional Portal (2004-2018) - CSS Recovery: 96% | Median Latency: 3800ms

**Cohort #16: PrestaShop 1.7 Multi-Category Store (2016-2024) - CSS Recovery:
97% | Median Latency: 5250ms**