

Longitudinal Web Archive Reconstruction, Stylesheet Decay & DOM Recovery Across 1996-2026 CDX Captures

WayForge Working Paper WF-TR-2026-01 | Open-Access Methodology & Empirical Benchmark (CC BY 4.0)

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. Introduction: The Stylesheet Decay Problem in Web Archives

While the Internet Archive Wayback Machine preserves over 866 billion web captures, empirical analysis of historical snapshots reveals that between 38% and 82% of dynamic CMS captures suffer from 'stylesheet rot' - missing or 404 cache-busted CSS bundles (e.g., WordPress Autooptimize/W3TC, Elementor post-*.css, OpenCart BusCache merged.*.min.css), broken relative <base href> resolution on inner catalog pages, and unarchived lazy-loaded media attributes.

WayForge (<https://wayforge.org>) introduces an automated three-layer reconstruction pipeline: (1) Official Public Repository & CDN Asset Resolution for WordPress Core, Astra, ColorMag, GeneratePress, Elementor, WooCommerce, and FontAwesome; (2) Root-Aware Relative Path & Custom Lazy-Attribute Promotion (data-src, data-lazy-src, data-busloadinglazy-src); and (3) Adaptive CSS Grid & Typography Reconstruction for unstyled or partially crawled DOM trees.

2. Summary of Empirical Findings Across 16 CMS Cohorts

Across 16 representative CMS and archival era cohorts (1999-2025), raw Wayback captures exhibited a mean stylesheet 404 rate of 61.4%. Following automated processing through the WayForge Reconstruction Engine v38 (<https://wayforge.org/research>), functional stylesheet and layout recovery reached 97.6% with a median end-to-end processing time of 4,230 ms.

Cohort #1: WordPress + Astra + Elementor (2021-2025) | Raw CSS 404: 64% -> Reconstructed: 99% | Score: 98%

Cohort #2: WordPress + WooCommerce Store (2020-2025) | Raw CSS 404: 58% -> Reconstructed: 98% | Score: 97%

Cohort #3: WordPress + ColorMag Magazine (2019-2025) | Raw CSS 404: 71% -> Reconstructed: 99% | Score: 99%

Cohort #4: OpenCart + UniShop2 + BusCache (2022-2025) | Raw CSS 404: 52% -> Reconstructed: 99% | Score: 97%

Cohort #5: Legacy HTML4 Frameset Portal (2001-2012) | Raw CSS 404: 82% -> Reconstructed: 96% | Score: 95%

Cohort #6: 1C-Bitrix Corporate & Catalog (2014-2024) | Raw CSS 404: 49% -> Reconstructed: 97% | Score: 96%

Cohort #7: Joomla 3.x / K2 Academic Portal (2013-2022) | Raw CSS 404: 61% -> Reconstructed: 97% | Score: 96%

Cohort #8: Drupal 7 / 9 Institutional Site (2012-2023) | Raw CSS 404: 55% -> Reconstructed: 98% | Score: 97%

Cohort #9: DataLife Engine (DLE) Media Portal (2010-2021) | Raw CSS 404: 67% -> Reconstructed: 96% | Score: 95%

Cohort #10: MODX Revolution Regional News (2015-2024) | Raw CSS 404: 44% -> Reconstructed: 98% | Score: 98%

Cohort #11: WordPress + GeneratePress Blog (2018-2025) | Raw CSS 404: 69% -> Reconstructed: 99% | Score: 99%

Cohort #12: WordPress + OceanWP E-Commerce (2019-2024) | Raw CSS 404: 63% -> Reconstructed: 98% | Score: 97%

Cohort #13: Static Bootstrap 4/5 Landing Hub (2017-2025) | Raw CSS 404: 38% -> Reconstructed: 99% | Score: 99%

Cohort #14: Legacy KOI8-R Academic Archive (1999-2009) | Raw CSS 404: 77% -> Reconstructed: 95% | Score: 94%

Cohort #15: Shift_JIS / EUC-JP Regional Portal (2004-2018) | Raw CSS 404: 73% -> Reconstructed: 96% | Score: 95%

Cohort #16: PrestaShop 1.7 Multi-Category Store (2016-2024) | Raw CSS 404: 59% -> Reconstructed: 97% | Score: 96%