



Wasteback Machine: a method for quantitative measurement of the archived web

David Mahoney

DOI: <https://doi.org/10.47989/ir31iConf64185>

Abstract

Introduction. Web archives are traditionally viewed as repositories of cultural memory, yet they have been theorised as computational sources for quantitative, longitudinal analysis of the web. This paper examines their potential for mapping the structural and environmental impacts of web pages, demonstrating broader applicability for web analytics research.

Method. We introduce Wasteback Machine, an open-source, extensible framework that operationalises the analytical potential of web archives. It enables reproducible, scalable measurement of page size and composition through programmatic access, structured resource extraction and mechanisms to mitigate distortions introduced during archiving and replay.

Analysis. The method is demonstrated through a case study of the United Nations Climate Change (UNFCCC) homepage, performing longitudinal analyses to capture temporal dynamics in size and compositional evolution. By situating web content within socio-technical and infrastructural contexts, the approach allows consistent comparison over time while accounting for archival limitations.

Results. Findings reveal trends in page growth, complexity and cumulative digital resource use. Despite their fragmentary nature, web archives provide sufficient fidelity to reconstruct historical practices and estimate relative environmental impacts.

Conclusion. Wasteback Machine demonstrates that web archives function as computational infrastructures, enabling rigorous, evidence-based investigation of web evolution and the environmental footprint of digital content.

Introduction

Cultural artifacts have long mediated how societies recognise themselves and envisage their futures, revealing, for instance, not only historical practices but also the environmental legacies they leave behind (Masanés, 2006). The web constitutes a pervasive and mutable medium through which contemporary culture materialises, with much of modern life instantiated online. Unlike other artifacts, the web is inherently ephemeral: websites are continually updated, supplanted, or vanish entirely, with roughly half becoming unrecognisable within a year and over 90% of web pages from the past decade irretrievably lost (Agata et al., 2014). As one of the most enduring and widely deployed internet technologies, websites remain central to human–computer interaction accounting for a measurable share of ICTs estimated contribution to global greenhouse gas (GHGs) emissions, yet their transience erases evidence of cultural change, evolving design patterns and cumulative environmental costs, limiting understanding of the long-term sustainability of digital practices.

Web archives, which preserve temporal snapshots of web pages (mementos), are valued for cultural memory and digital humanities scholarship and have long been proposed as instruments for quantitative analysis; however, no standardised methodologies currently exist (Sompel et al., 2013). This paper addresses this lacuna, and introduces Wasteback Machine, an open-source JavaScript extensible framework that renders mementos analytically tractable for systematic, reproducible, and retrospective quantitative analysis of web pages. Our method aims to realise the theoretical potential of web archives to support longitudinal studies of page size, composition and evolving web development practices. By examining these temporal dynamics, we explore the potential of web archives as computational resources for understanding the historical evolution of websites and their environmental implications.

Literature review

Environmental and social impact of ICTs

Information and communication technologies (ICTs) underpin nearly all aspects of contemporary life. Since the Paris Agreement in 2015, global internet use has risen by 28%, reaching 5.5 billion people or 68% of the world's population (ITU, 2024; UNFCCC, 2016). Despite the internet's transformative potential, *'digitalisation's promise of dematerialisation has not materialised'* (United Nations Trade and Development, 2024). The rhetoric of the cloud and the notion of information as *'weightless,'* a *'convenient fiction,'* obscure the substantial material and energetic foundations of computation (Hu, 2015; Turing, 1950).

ICTs rely on resource-intensive supply chains that consume natural and abiotic materials and emit 1.5–3.2% of global GHGs, with further growth expected from AI expansion (McGovern, 2020; United Nations Trade and Development, 2024). *'Unregulated digitalisation [also] risks leaving people behind'* (United Nations Trade and Development, 2024). 47% of the world's poorest 40% rely on mobile networks, where the growing *'bloat'* of digital media and requisite hardware constrains access, impeding progress on the Sustainable Development Goals (Nyce, 2024; GSMA, 2024; United Nations, 2015). These populations also disproportionately shoulder the deleterious effects of digitalisation, extending patterns of 18th-century *'telegraphic imperialism'* (Lahiri Choudhury, 2010).

Growth and complexity of web pages

While AI's environmental impact captures headlines, websites remain central to human–computer interaction and a locus of ICTs' deleterious effects. Page size, the total data required to load a webpage, has increased significantly despite optimisation strategies such as caching, content delivery networks, compression and more efficient media formats, reflecting the rebound effect, or Jevons paradox (Dawson, 2023; Giampietro & Mayumi, 2018; Jevons & Flux, 1965).

Since 2010, desktop web pages have grown 525% to a median size of median 2,900 KB, earning an 'F' rating (over 0.360 g CO₂e per load) according to Website Carbon (<https://websitecarbon.com>), while a recommended 'A' rating requires a web page to be under 0.079 g CO₂e (HTTP Archive, n.d.). Page size serves only as an indicator of environmental impact as not all bytes are equal: text is far less computationally intensive than CSS, which is lighter than JavaScript, both of which are far lighter than WebGL or 4K video (W3C, n.d.).

Amid this backdrop, the Sustainable Web Design (SWD) movement has gained prominence, with the W3C recently publishing a first public Draft Note in 2025 of its Web Sustainability Guidelines, which emphasises reducing page size (Mightybytes & Wholegrain Digital, n.d.; W3C, n.d.). Nevertheless, SWD remains underutilised and prevailing structural pressures driving continuous growth may negate efficiency gains. Consequently, historical analysis of page size trends is critical for elucidating long-term sustainability impacts and for identifying targeted opportunities for intervention, thereby advancing awareness of SWD practices and fostering their broader adoption.

Limitations of existing measurement approaches

Current methods for quantitative measurement of websites are largely synchronous. Website Carbon for example, calculates the emissions of individual web pages, representing a broader class of SWD tools. Björneborn and Ingwersen identify four main areas of quantitative web research: webpage content analysis, web link structure analysis, web usage analysis and web technology analysis, yet longitudinal analysis is notably absent (Björneborn & Ingwersen, 2004). This is consequential because websites are continually updated, supplanted, or vanish entirely, with roughly half becoming unrecognisable within a year and over 90% of pages from the past decade irretrievably lost (Agata et al., 2014).

Initiatives such as the HTTP Archive State of the Web report (<https://httparchive.org/reports/state-of-the-web>) and Web Almanac (<https://almanac.httparchive.org>) provide annual aggregate insights back to 2010, but do not support site-specific retrospective longitudinal analysis, leaving a critical gap in understanding historical trends. This gap also affects corporate sustainability reporting, particularly Scope 3 emissions of Greenhouse Gas Protocol (GHGP), which can represent up to 80% of an organisation's footprint (WBCSD & WRI, 2013; ITU, 2025). Limited historical visibility constrains accurate reporting, baseline setting, retrospective recalculations and targeted interventions, underscoring the need for approaches to retrospectively measure web evolution and environmental impact.

Web archives as computational sources

Web archives preserve temporal snapshots of web pages (mementos), capturing content, structure and metadata at discrete points in time (Ainsworth et al., 2014). These collections are invaluable for cultural memory as well as research in digital humanities, historical scholarship and judicial proceedings (Brügger, 2017). While many national and institutional archives exist with varying temporal and thematic coverage, the Internet Archive's Wayback Machine, launched in 1996, remains the largest and most comprehensive, storing over a trillion web pages as of October 2025 (Internet Archive Blogs, 2025). Although primarily employed for interpretive research, scholars have highlighted its potential for quantitative analyses, '*beyond the novelty of exploring*' (Brügger & Milligan, 2019; Masanés, 2006; Sompel et al., 2009; Brügger, 2017).

Despite this theoretical promise, standardised methodologies are lacking, both reflecting their emergence at the computational turn in digital humanities and the ontological status of mementos as '*re-born digital*' objects, distinct from their originals (Brügger, 2017; Brügger & Milligan, 2019; Masanés, 2006). Variations in capture completeness and technical processes, archival biases, further complicate analysis, rendering archives imperfect surrogates. This, however, represents an opportunity: by navigating these challenges through computational methods it could render web archives analytically tractable, enabling reproducible, retrospective, longitudinal and

comparative studies of the web, quantifying historical digital infrastructure, revealing design patterns and informing interventions for more sustainable web development.

Methodology

Our methodology enables quantitative measurement of mementos through a three-stage procedure: retrieval of mementos, recompositing of mementos and structured extraction, classification and measurement of resources. Each stage is designed to minimise distortions introduced by archiving and replay, a process aptly described as ‘rehydrating instant coffee’ (Kocsis and Kaufman, 2025).

Retrieval of mementos

Wasteback Machine leverages the Internet Archive’s Wayback Machine as its corpus, utilising its programmatic access through their available APIs: Wayback CDX Server API and Wayback Replay API (Internet Archive, 2025, 2013). Using the Wayback CDX Server API, Wasteback Machine identifies candidate mementos by retrieving all captures for a target URL within a specified period. When the desired datetime is unavailable, the nearest capture is selected, ensuring analyses remain temporally proximate while accommodating the inherently irregular capture rhythms of web archives.

Recompositing of memento

Preparing mementos for quantitative analysis requires mitigating distortions introduced during archiving and replay through a method of recomposing. The Internet Archive’s Wayback Machine employs a method of replay known as integrated replay, injecting toolbars, interface elements and emulation scripts for deprecated technologies such as Flash directly into the HTML (Ainsworth et al., 2014; Scott, 2020). While necessary for navigation and functionality, these integrations distort page size and composition.

The Wayback Machine Replay API supports multiple ‘endpoints’ that alter how a memento is rendered during replay by appending flags to the <datetime> in the memento URL (<https://web.archive.org/web/<datetime>/<url>>). Wasteback Machine uses the `if_` endpoint, which suppresses archive interface elements while preserving the mementos visual completeness and functionality (https://web.archive.org/web/<datetime>if_/<url>). Additional extraneous embedded resources (eURI-Ms), such as emulation scripts, are excluded by blocking URLs identified as outside the scope of the memento (Ainsworth et al., 2014).

The root HTML document returned by the `if_` endpoint still contains structural modifications. Archive linkage and replay-preserving modifications redirect resource and navigational paths to the archive, preserving temporal coherence but lengthening URLs, inflating byte size. Wasteback Machine uses the `id_` endpoint (https://web.archive.org/web/<datetime>id_/<url>) to retrieve the unmodified root HTML.

Combining the `id_` endpoint for the unmodified HTML with the `if_` endpoint for eURI-Ms, Wasteback Machine reconstructs mementos as discrete historical artefacts: encapsulated, temporally coherent representations of web pages as they were captured, preserving both structure and content ready for accurate analysis.

Extraction, classification, and measurement of resources

Wasteback Machine primary dependency is the JavaScript library JSDOM, a server-side virtual DOM, employed to traverse both API endpoints. JSDOM was selected for its ability to render legacy file types that headless browser JavaScript libraries, such as Playwright. Additionally, JSDOM facilitates integration into other projects, aligning with the research objectives of providing an extensible framework.

Recognising the inherent inconsistencies and network dependencies of web archives, the system incorporates safeguards including try-catch blocks to handle transient errors, timeouts and rate limits, logging of retrieval warnings and filtering of malformed URLs. eURI-Ms are classified using a layered approach that combines MIME type analysis, file extension heuristics and contextual inference from HTML attributes. The system distinguishes primary categories including style sheets, scripts, images, videos, audio, fonts, Flash, plugins, data and documents. When MIME information is missing or unreliable, as often occurs in early web captures, extension- and tag-based heuristics ensure consistent classification across archival eras (see Table 1).

Category	File Extensions / MIME Types
Style Sheet	css, scss, sass, less
Script	js, mjs, cjs, ts, jsx, tsx, coffee, vue
Image	png, jpg, jpeg, gif, webp, avif, bmp, ico, cur, tiff, tif, svg, apng, heic, heif, jp2, j2k, dds, ppm, pgm, pbm, hdr
Video	mp4, webm, ogv, m4v, mkv, mov, avi, flv, m2v, ts, rmvb, rm, f4v, f4p, f4a, f4b, 3gp, 3g2
Audio	mp3, wav, ogg, oga, aac, m4a, flac, opus, mid, midi, ra, ram, aif, aiff, au, m4b
Font	woff, woff2, ttf, otf, eot, pfa, pfb
Flash	swf
Plugin	jar, class, xap, unity3d, dcr, dir, cab, ocx
Data	json, xml, zip, tar, 7z, wasm, map, csv, tsv, yaml, yml, sqlite, db, db3
Document	txt, pdf, rtf, log, ini, conf

Table 1. File type categories and associated extensions used for resource classification

Text-based resources (HTML, style sheets, scripts) are stripped of archival annotations and converted to byte arrays, while the remaining binary resources are retrieved as raw byte buffers. Duplicate resources are removed to ensure accurate counts. Resource size is recorded in bytes, providing a precise metric for quantitative analysis, while completeness metrics report the proportion of successfully retrieved resources, supporting reproducibility and transparency. This approach ensures robust, reliable measurement across both early static websites and modern client-side rendered web pages.

Installation and usage

Wasteback Machine is distributed under the open-source Apache 2.0 license on GitHub (<https://github.com/overbrowsing/wasteback-machine>). The underlying methodology is implemented as a Node.js project, it can be installed as a dependency via NPM (npm install @overbrowsing/wasteback-machine) or Yarn (yarn add @overbrowsing/wasteback-machine).

The library provides two primary functions: ‘getMementos’ retrieving available mementos for a target URL within a specified timeframe and ‘getMementoSizes’ analysing individual mementos to measure page size and resource composition. By structuring the library around these two functions, Wasteback Machine supports reproducible deployment for both individual case studies and large-scale longitudinal analyses of mementos on the Wayback Machine.

The first function, ‘getMementos,’ retrieves all memento-datetime for a target URL. For example, Figure 1 illustrates fetching datetimes for the New York Times homepage (<https://nytimes.com>) between 1996 and 2025, as shown in the GitHub project documentation.



Fig. 1. Example function, ‘getMementos,’ for checking available mementos the New York Times homepage between 1996 and 2025.

The second function, ‘getMementoSizes,’ calculates a memento’s size and composition at a given datetime, producing a structured JSON object that consolidates measured data, metadata, and retrieval completeness. Optionally, it can generate a detailed inventory of the HTML document and all eURI-Ms. Table 2 summarises the primary fields and resource classification metadata. Figure 2 provides an example JSON output for the New York Times homepage from November 12th, 1996.

Field	Description
url	The canonical target URL under analysis
requestedMemento	The datetime originally specified by the user
memento	The resolved archival datetime selected via proximity
mementoURL	Link to the memento of the target URL on the Wayback Machine
sizes	Per-category byte and resource-count aggregates
completeness	Percentage of successfully retrieved resources
resources (optional)	Detailed list of all measured resources

Table 2. Output fields produced by Wasteback Machine function ‘getMementoSizes’.

2. Analyse a Specific Memento

```
import { getMementoSizes } from "@overbrowsing/wasteback-machine";

// Analyse www.nytimes.com memento from November 12, 1996
const mementoData = await getMementoSizes(
  'https://nytimes.com',
  '19961112181513',
  { includeResources: true } // optional: include full resource list
);
console.log(mementoData);
```

Example Output

```
{
  url: 'https://nytimes.com',
  requestedMemento: '19961112181513',
  memento: '19961112181513',
  mementoURL: 'https://web.archive.org/web/19961112181513/https://nytimes.com',
  sizes: {
    html: { bytes: 1653, count: 1 },
    stylesheet: { bytes: 0, count: 0 },
    script: { bytes: 0, count: 0 },
    image: { bytes: 46226, count: 2 },
    video: { bytes: 0, count: 0 },
    audio: { bytes: 0, count: 0 },
    font: { bytes: 0, count: 0 },
    flash: { bytes: 0, count: 0 },
    plugin: { bytes: 0, count: 0 },
    data: { bytes: 0, count: 0 },
    document: { bytes: 0, count: 0 },
    other: { bytes: 0, count: 0 },
    total: { bytes: 47879, count: 3 }
  },
  completeness: '100%',
  resources: [
    {
      url: 'https://web.archive.org/web/19961112181513im_/http://www.nytimes.com/index.gif',
      type: 'image',
      size: 45259
    },
    {
      url: 'https://web.archive.org/web/19961112181513im_/http://www.nytimes.com/free-images/m',
      type: 'image',
      size: 967
    }
  ]
}
```

Fig. 2. Example function, 'getMementoSizes,' analysing the New York Times homepage between 1996 and 2025.

Case study

To demonstrate the capabilities of Wasteback Machine, we conducted a longitudinal analysis of the United Nations Climate Change (UNFCCC) homepage (<https://unfccc.int>) using the command line interface (CLI) demo, following the procedures documented in the GitHub repository for Wasteback Machine NPM v1.0.13. The analysis spans the earliest available Wayback Machine captures from 2000 (see Appendix Table 3).

The CLI demo integrates Wasteback Machine with CO2.js a JavaScript library for converting bytes transferred into equivalent emissions (<https://developers.thegreenwebfoundation.org/co2js/installation/>). These results reflect the impact of recomposing mementos on present-day infrastructure without renewable energy, rather than the historical conditions at the time of capture.

Figure 3 presents an output from the CLI demo, demonstrating the structured report generated in accordance with repository procedures, including total page size, composition and estimated emissions per page load.

```
wasteback-machine % node examples/demo.js www.unfccc.int 2000 10 18
Fetching mementos for https://www.unfccc.int in 2000-10-18...

Retrieved Memento:
🔗 Memento URL: https://web.archive.org/web/20001018063023/https://www.unfccc.int
✅ Completeness: 100%

Page Size Results:
📊 Data Transfer: 32.00 KB
🌐 Page CO2e: 0.010 g

Page Composition Results:
📁 HTML
Count: 1
Size: 10.59 KB (33.1%)
CO2e: 0.003 g

📁 STYLESHEET
Count: 1
Size: 0.44 KB (1.4%)
CO2e: 0.000 g

📁 IMAGE
Count: 5
Size: 20.96 KB (65.5%)
CO2e: 0.006 g
```

Fig. 3. Results from the CLI demo showing the UNFCCC homepage memento retrieved from October 18th, 2000, displayed in the terminal.

Results

The retrospective analysis of the UNFCCC homepage from 2000 to 2025 shows dramatic growth in page size and embedded resources (Appendix Table 3, Figure 4). Early pages were typically under 50 KB with fewer than ten eURI-Ms, while by 2025 the homepage exceeded 8.7 MB, a 17,720% increase, generating roughly 2.6 g CO₂e per load, far above the 525% average growth reported by HTTP Archive (HTTP Archive, n.d.). At 100,000 monthly views, this corresponds to 260 kg CO₂e per month, or 3.1 t annually, equivalent to the annual sequestration of approximately 140 mature trees (EcoTree, n.d.).

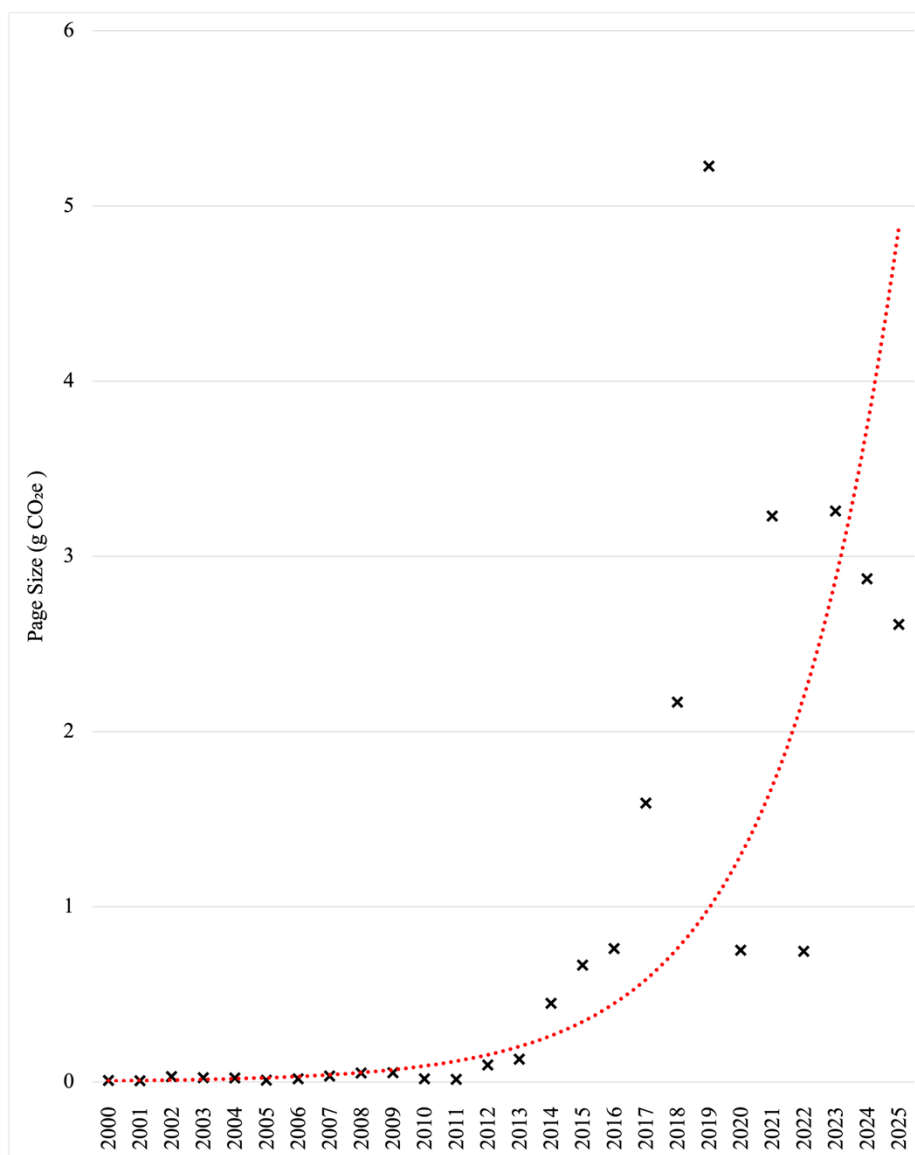


Fig. 4. Page size (g CO₂e) of UNFCCC homepage mementos (October 2000–2025).

Figures 5 and 6 present page composition, split into two panels to maintain readability given the substantial growth of post-2013 mementos. Over time, scripts, stylesheets and images all increased, with a large video first appearing in 2019 contributing over 60% of total bytes for that memento, demonstrating the impact of media. Flash usage declined, replaced by JavaScript-driven interactivity and responsive layouts. Images consistently accounted for 50–65% of page size, while HTML remained modest.

These patterns reflect the ease with which high-resolution images and videos can now be integrated into web pages, amplifying both computational load and carbon footprint. By highlighting which elements dominate resource use, these findings suggest that sustainability interventions should prioritise media optimisation through compression, modern image formats, or reducing the number of HTTP requests through selective media use. Enabled by Wasteback Machine method, this longitudinal perspective highlights both the dramatic growth in page size and the importance of SWD practices, such as setting a pages size budget and choosing a sustainable service provider.

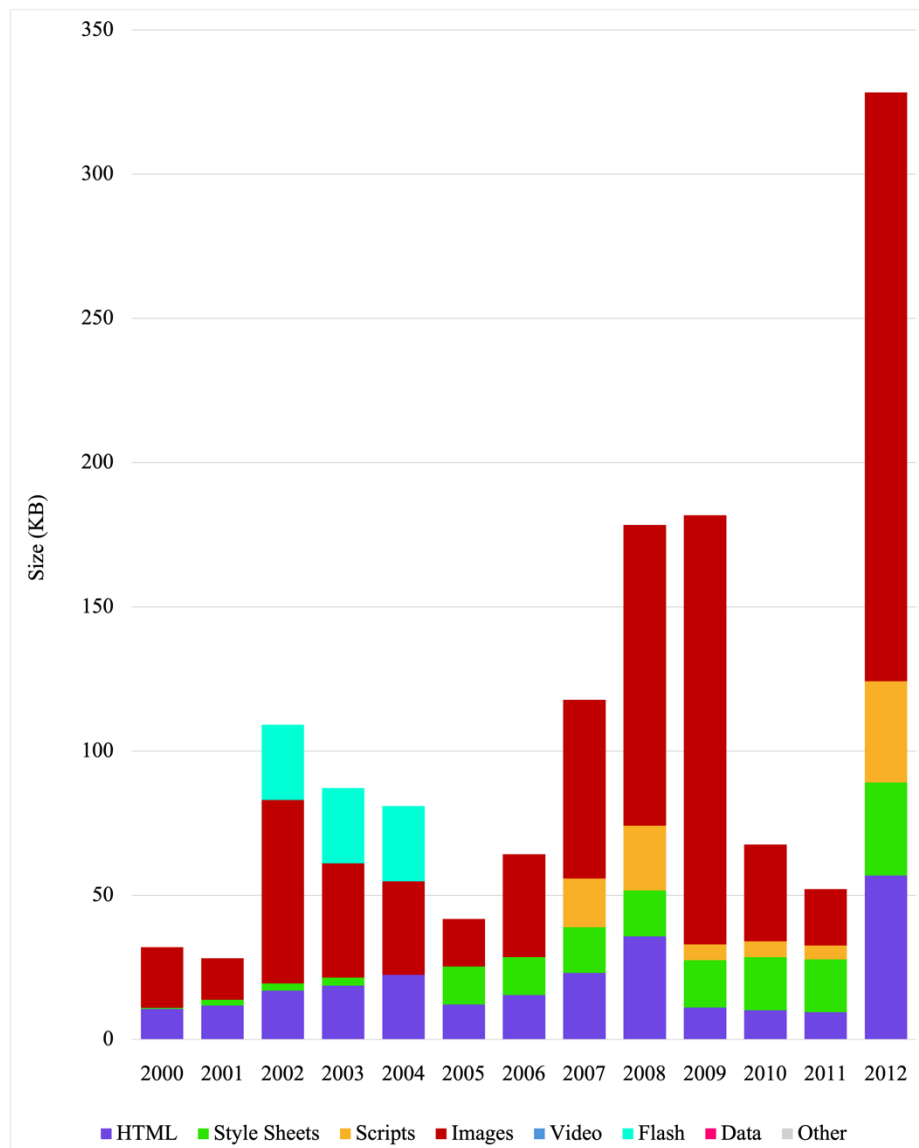


Fig. 5. Chart showing page composition (KB) of UNFCCC homepage mementos (October 2000–2012).

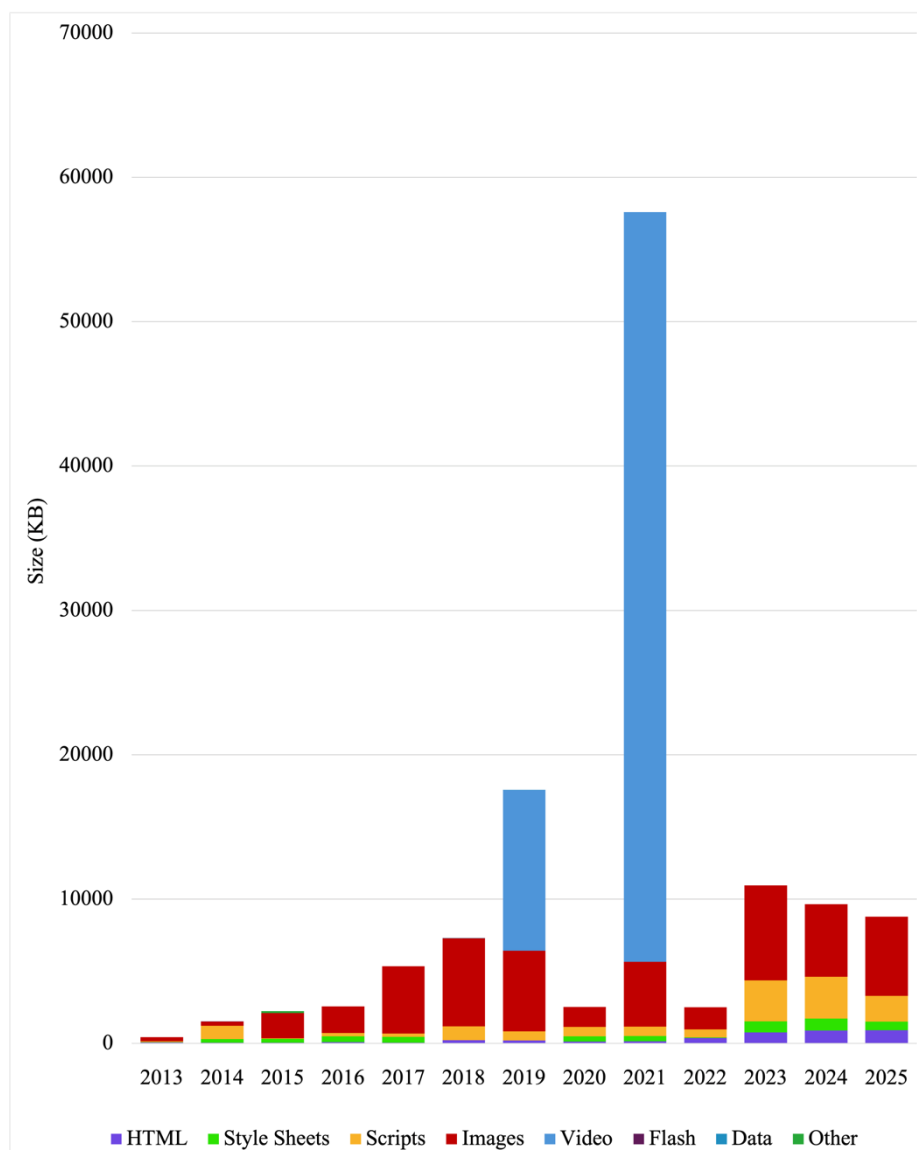


Fig. 6. Chart showing page composition (KB) of UNFCCC homepage mementos (October 2013–2025).

Discussion and analysis

Web archives as computational sources

The theoretical positioning of this work builds on long-standing discussions in the digital humanities. Mementos occupy a distinctive ontological position as ‘*re-born digital*’ objects: static snapshots of a medium defined by continuous change, shaped both by the original web and the technical and curatorial decisions of archiving institutions (Brügger, 2017). Unlike live web pages (‘*born digital*’ objects), they simultaneously reflect web content and the processes that preserve it, determining what is captured, how it is represented and what is absent. This recursive nature introduces methodological complexity and sensitivity, requiring careful contextualisation for quantitative analysis.

Wasteback Machine addresses these challenges by integrating digital humanities insights with computational methods, operationalised through recompositing mechanisms that minimise archival and replay distortions. This paper establishes a framework for reproducible and temporally coherent quantitative measurement of mementos. For analyses of longitudinal trends

in page size and composition, web archives remain among the few reliable sources, and Wasteback Machine demonstrates that, despite ontological and technical complexities, archived web content can support systematic and reproducible quantitative web research.

Limitations

Despite Wasteback Machine's capabilities, interpretation of memento-derived measurements is constrained by several factors. Modern web pages increasingly rely on dynamic scripts, lazy loading, client-side rendering and single-page application frameworks, which may be only partially captured depending on crawler behaviour. Early mementos are sometimes fragmentary, while later captures benefit from higher fidelity, introducing temporal variability (Ainsworth et al., 2015).

Furthermore, mementos often represent first-load states, excluding caching effects that reduce live data transfer and variability in crawler technologies, regional coverage and archival scope further limits cross-site comparability. Crucially, web archives do not capture if a memento was originally hosted using a sustainable service provider, precluding accurate direct conversion of page size into equivalent emissions.

Like fragmentary cultural artefacts, mementos retain sufficient structure to support analysis, but measurements must be interpreted as indicative of historical web practices rather than definitive metrics, requiring researchers to situate findings within the socio-technical and temporal context of the web and recognise that metrics reflect past archival, infrastructural and web development conventions.

Future research

Wasteback Machine opens multiple avenues for further investigation. Retrospective analyses could extend the temporal coverage of datasets such as the HTTP Archive State of the Web report, supporting more comprehensive reconstructions of historical web evolution and addressing gaps in measures of page size and composition.

Future research could examine how these trends affect the sustainability, management, and long-term preservation of web archives, which must continue to be maintained and refreshed. In Web Archiving, Masanès outlines a vision for a global grid of interconnected archives forming a unified navigable space akin to the live web, expanding temporal coverage and strengthening longitudinal fidelity (Masanés, 2006). Yet variation in capture standards, replay environments, and archival technologies complicates such integration. Ongoing initiatives mapping these technical differences offer a pathway for addressing these challenges and may be essential for realising an interoperable archival grid.

Further work will explore coupling Wasteback Machine measurements with emissions-estimation frameworks, including the ISO Software Carbon Intensity model, translating historical byte counts into approximate energy use and carbon emissions while accounting for temporal changes in computational efficiency (ISO, n.d.). Finally, a public-facing web application of Wasteback Machine is in development to increase accessibility, providing researchers, educators, and practitioners with a direct interface to engage with web archives and the underlying methodology.

Conclusion

Web archives constitute computational instruments capable of supporting systematic, longitudinal analysis of the web. Despite their ontological status as 're-born digital' materials and the attendant methodological complexities, mementos, like fragmentary historical artefacts, retain sufficient sources to permit investigation of temporal dynamics in web development.

Wasteback Machine operationalises this potential, providing a reproducible and extensible framework for quantifying memento size and composition. By situating web pages within their historical, infrastructural and socio-technical contexts, this approach illuminates the evolution of

web practices and the cumulative environmental footprint of the web, demonstrating how web archives can be utilised for evidence-based research.

About the author

David Mahoney is a PhD candidate at The University of Edinburgh's Institute for Design Informatics. His principal research interest is reducing the environmental impact of the internet and improving accessibility, aligning technological potential with environmental stewardship. He is the founder of Overbrowsing, an applied research group focused on advancing sustainable web practices. He can be contact at d.f.mahoney@sms.ed.ac.uk

References

- Agata, T., Miyata, Y., Ishita, E., Ikeuchi, A., & Ueda, S. (2014). Life span of web pages: A survey of 10 million pages collected in 2001. IEEE/ACM Joint Conference on Digital Libraries, 463–464. <https://doi.org/10.1109/JCDL.2014.6970226>
- Ainsworth, S. G., Nelson, M. L., & Sompel, H. V. de. (2014). A Framework for Evaluation of Composite Memento Temporal Coherence (arXiv:1402.0928). arXiv. <https://doi.org/10.48550/arXiv.1402.0928>
- Ainsworth, S. G., Nelson, M. L., & Van De Sompel, H. (2015). Only One Out of Five Archived Web Pages Existed as Presented. Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15, 257–266. <https://doi.org/10.1145/2700171.2791044>
- Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. Journal of the American Society for Information Science and Technology, 55(14), 1216–1227. <https://doi.org/10.1002/asi.20077>
- Brügger, N. (2017). The Web as History. UCL Press. <https://doi.org/10.14324/111.9781911307563>
- Brügger, N., & Milligan, I. (2019). The SAGE Handbook of Web History. SAGE Publications Ltd. <https://doi.org/10.4135/9781526470546>
- Dawson, A. (2023). The Carbon Impact of Web Standards. <https://web.archive.org/web/20250912172751/https://websitesustainability.com/cache/files/research23.pdf>
- EcoTree. (n.d.). How much CO2 does a tree absorb? Let's get carbon curious! EcoTree. Retrieved 24 April 2025, from <https://web.archive.org/web/20250805095905/https://ecotree.green/en/how-much-co2-does-a-tree-absorb>
- Giampietro, M., & Mayumi, K. (2018). Unraveling the Complexity of the Jevons Paradox: The Link Between Innovation, Efficiency, and Sustainability. Frontiers in Energy Research, 6, 26. <https://doi.org/10.3389/fenrg.2018.00026>
- GSMA. (2024). The Mobile Economy 2024. GSMA. <https://web.archive.org/web/20250906132723/https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2024/02/260224-The-Mobile-Economy-2024.pdf>
- HTTP Archive. (n.d.). HTTP Archive: Page Weight. Retrieved 4 November 2025, from <https://httparchive.org/reports/page-weight?start=earliest&end=latest&view=list>

- Hu, T.-H. (2015). *A Prehistory of the Cloud*. The MIT Press.
<https://doi.org/10.7551/mitpress/9780262029513.001.0001>
- Internet Archive. (2025). Help: Using the Wayback Machine. In Wikipedia.
https://web.archive.org/web/20250915082515/https://en.wikipedia.org/w/index.php?title=Help:Using_the_Wayback_Machine&oldid=1275603720
- Internet Archive. (2013). Wayback CDX Server API - BETA – Internet Archive Developer Portal.
<https://archive.org/developers/wayback-cdx-server.html>
- Internet Archive Blogs. (2025, November 21). Celebrating 1 Trillion Web Pages Archived | Internet Archive Blogs. <https://blog.archive.org/trillion/>
- ISO. (n.d.). BS ISO/IEC 21031:2024. ISO. Retrieved 23 September 2025, from
<https://www.iso.org/standard/86612.html>
- ITU. (2025). Greening Digital Companies 2025: Monitoring emissions and climate commitments.
<https://www.itu.int:443/en/ITU-D/Environment/Pages/Publications/GDC-25.aspx>
- ITU. (2024, December 11). Statistics: Individuals using the Internet.
<https://web.archive.org/web/20250912085250/https://www.itu.int/en/ITU-D/Statistics/pages/stat/default.aspx>
- Jevons, W. S., & Flux, A. W. (1965). *The Coal Question; an inquiry concerning the progress of the Nation, and the probable exhaustion of our coal-mines*. New York, A. M. Kelley.
<https://archive.org/details/coalquestionani00jevooog>
- Lahiri Choudhury, D. K. (2010). *Telegraphic imperialism: crisis and panic in the Indian Empire, c.1830*. Palgrave Macmillan.
- Masanés, J. (2006). *Web Archiving*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-46332-0>
- McGovern, G. (2020). *World Wide Waste: how digital is killing our planet - and what we can do about it* (First edition). Silver Beach Publishing.
- Mightybytes, & Wholegrain Digital. (n.d.). *Sustainable Web Design*. Sustainable Web Design. Retrieved 12 September 2025, from
<https://web.archive.org/web/20250912091249/https://sustainablewebdesign.org/>
- Nyce, C. M. (2024, January 15). The Internet Is Being Ruined by Bloated Junk. *The Atlantic*.
<https://web.archive.org/web/20250406081749/https://www.theatlantic.com/technology/archive/2024/01/long-youtube-videos-tiktok/677130/>
- Scott, J. (2020, November 19). Flash Animations Live Forever at the Internet Archive | Internet Archive Blogs. <https://blog.archive.org/2020/11/19/flash-animations-live-forever-at-the-internet-archive/>
- Sompel, H. V. de, Nelson, M. L., Sanderson, R., Balakireva, L. L., Ainsworth, S., & Shankar, H. (2009). *Memento: Time Travel for the Web* (arXiv:0911.1112). arXiv.
<https://doi.org/10.48550/arXiv.0911.1112>
- Sompel, H. V. de, Nelson, M., & Sanderson, R. (2013). HTTP Framework for Time-Based Access to Resource States – Memento (Request for Comments RFC 7089). Internet Engineering Task Force. <https://doi.org/10.17487/RFC7089>

- Turing, A. M. (1950). I.—Computing Machinery and Intelligence. *Mind*, LIX(236), 433–460.
<https://doi.org/10.1093/mind/LIX.236.433>
- UNFCCC (Ed.). (2016). Report of the Conference of the Parties on its 21st session, held in Paris from 30 November to 13 December 2015: addendum. UN.
<https://digitallibrary.un.org/record/831052>
- United Nations. (2015). Transforming our world: the 2030 Agenda for Sustainable Development. UN. <https://digitallibrary.un.org/record/1654217>
- United Nations Trade and Development. (2024). Digital Economy Report 2024: Shaping an Environmentally Sustainable and Inclusive Digital Future. United Nations.
<https://doi.org/10.18356/9789213589779>
- W3C. (n.d.). Web Sustainability Guidelines (WSG). Retrieved 8 December 2025, from <https://www.w3.org/TR/web-sustainability-guidelines/#set-goals-based-on-performance-and-energy-impact>
- WBCSD, & WRI. (2013). GHG Protocol Technical Guidance for Calculating Scope 3 Emissions. World Resources Institute & World Business Council for Sustainable Development.
https://web.archive.org/web/20250906182341/https://ghgprotocol.org/sites/default/files/standards/Scope3_Calculation_Guidance_0.pdf

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).

Appendix

Date	Memento URL
2000	https://web.archive.org/web/20001018063023/https://www.unfccc.int
2001	https://web.archive.org/web/20010722064142/https://www.unfccc.int
2002	https://web.archive.org/web/20021014163344/https://www.unfccc.int
2003	https://web.archive.org/web/20031016042617/https://www.unfccc.int
2004	https://web.archive.org/web/20041009100909/https://www.unfccc.int
2005	https://web.archive.org/web/20051016023644/https://www.unfccc.int/2860.php
2006	https://web.archive.org/web/20061017213005/https://www.unfccc.int/2860.php
2007	https://web.archive.org/web/20071016100745/https://www.unfccc.int/2860.php
2008	https://web.archive.org/web/20081017084911/https://www.unfccc.int/2860.php
2009	https://web.archive.org/web/20090912204304/https://www.unfccc.int
2010	https://web.archive.org/web/20100828125929/https://www.unfccc.int
2011	https://web.archive.org/web/20111022050807/https://www.unfccc.int
2012	https://web.archive.org/web/20121015093140/https://www.unfccc.int/2860.php
2013	https://web.archive.org/web/20131018200408/https://www.unfccc.int/2860.php
2014	https://web.archive.org/web/20141017053200/https://www.unfccc.int/2860.php
2015	https://web.archive.org/web/20151016054246/https://www.newsroom.unfccc.int
2016	https://web.archive.org/web/20161015194814/https://www.newsroom.unfccc.int
2017	https://web.archive.org/web/20171018023837/https://www.newsroom.unfccc.int
2018	https://web.archive.org/web/20181018073838/https://www.unfccc.int
2019	https://web.archive.org/web/20191019181609/https://www.unfccc.int
2020	https://web.archive.org/web/20201017040902/https://www.unfccc.int
2021	https://web.archive.org/web/20211018144835/https://www.unfccc.int
2022	https://web.archive.org/web/20221019023812/https://www.unfccc.int
2023	https://web.archive.org/web/20230910043835/https://www.unfccc.int
2024	https://web.archive.org/web/20240801005030/https://www.unfccc.int
2025	https://web.archive.org/web/20250101043332/https://www.unfccc.int

Table 3. URLs of UNFCCC homepage mementos (October 2000–2025).

Date	Size (KB)	Size (g CO ₂ e)	Completeness (%)
2000	32.00	0.010	100
2001	28.18	0.008	88
2002	109.16	0.033	100
2003	87.16	0.026	100
2004	80.92	0.024	100
2005	41.81	0.012	100
2006	64.21	0.019	100
2007	117.74	0.035	94
2008	178.40	0.053	96
2009	181.80	0.054	100
2010	49.23	0.015	100
2011	52.19	0.016	100
2012	328.36	0.098	100
2013	438.40	0.131	97
2014	1514.32	0.451	98
2015	2245.79	0.669	100
2016	2561.50	0.763	97
2017	5350.75	1.593	71
2018	7283.90	2.169	88
2019	17558.95	5.229	95
2020	2528.22	0.753	96
2021	10852.15	3.232	96
2022	2507.20	0.747	93
2023	10951.32	3.261	95
2024	9646.13	2.873	95
2025	8778.07	2.614	95

Table 4. Page size and recompositing completeness of UNFCCC homepage mementos (October 2000–2025).