MCP SERVER | NO CODE | CLOUD HOSTED

Internet Archive MCP

Access 40M+ historical media records in one chat.

Internet Archive MCP connects your AI agent to the world's largest digital library, accessing 40 million+ items in one chat session. Search everything—books, films, music, software, and historical web pages via the Wayback Machine—using natural conversation instead of complicated search forms.

A+ Quality Score 98.33/100

digital-library | wayback-machine | archival-data | metadata-search | historical-records | open-access

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/internet-archive

SCAN TO VIEW

The infrastructure that powers AI agents in the real world.

Vinkius connects AI to the world's software through secure, enterprise-grade infrastructure — enabling real-world execution at scale, built on the Model Context Protocol (MCP).

Your AI Connections Run Through Vinkius Cloud

The world's largest managed MCP catalog

Vinkius is the cloud infrastructure where AI agents connect to the software your business already runs. We handle the hosting, the security, the credentials, the uptime — you get agents that actually do things.

We operate the world's largest managed MCP catalog. Major SaaS platforms, CRMs, databases, and cloud providers — running, monitored, production-ready. This MCP server is hosted and maintained by the Vinkius Cloud for AI Agents.

> *The agent doesn't manage credentials, doesn't manage uptime, doesn't manage security. Vinkius does.*
>
> — Architecture principle

Four Pillars of the Vinkius Runtime

01 — Security by design

Credentials stay encrypted at rest via AES-256. The AI agent never touches raw keys — they're injected into a sandboxed V8 isolate at runtime. Actions are logged, and connections have an emergency kill switch.

02 — Built on MCP Fusion

This MCP server was built with **MCP Fusion**, the open-source framework (Apache 2.0) that powers the entire Vinkius catalog. Schema-as-firewall strips undeclared fields, compiled PII redaction runs at zero overhead, and cryptographic lockfiles produce git-diffable audit trails.

03 — Deterministic observability

Eight immutable metrics per endpoint: request volume, p95 latency, error rate, active connections, cost attribution. A live payload feed logs every tool call with mutation detection.

04 — Autonomous operations

Servers are deployed, monitored, and patched autonomously. New capabilities and security patches ship weekly. Zero-downtime deployments ensure continuous availability across all managed MCP servers.

AES-256	Ed25519	24h TTL	V8 Isolate
Encryption at rest	PKI vault signatures	Ephemeral session keys	Sandboxed execution

One Token. Instant Access.

Every MCP server on Vinkius is accessed through a **Connection Token**. Tokens are generated in the cloud dashboard and produce a unique MCP endpoint URL. Paste this URL into any MCP-compatible client — no SDK required.

A single token can serve **multiple AI clients simultaneously**, or you can issue separate tokens per client for granular access control. Each token tracks its own request count, last activity timestamp, and can be individually enabled or revoked.

MCP ENDPOINT

```
https://edge.vinkius.com/{token}/mcp
```

Claude · Cursor · VS Code · Windsurf · Grok · Gemini

Security Is the Architecture

Security in Vinkius is not a feature — it's the foundation of the runtime. The gateway enforces multiple independent protection layers between AI agents and third-party APIs.

01 — Ed25519 PKI Vault

Every workspace has an Ed25519 Master Key. Session keys are generated ephemerally (24h TTL) and signed by the Master Key. Credentials never leave the vault boundary.

02 — V8 Isolate Sandboxing

Tool code runs inside isolated-vm V8 isolates with 64 MB memory caps and per-request timeouts. No filesystem access, no network access except through the SSRF-guarded fetch bridge.

03 — SSRF Guard

All outbound HTTP requests are DNS-resolved and validated before execution. Private IP ranges (10.x, 172.16-31.x, 192.168.x, AWS metadata 169.254.x) are blocked at the network layer.

04 — DLP & PII Redaction

A ResponseGuard pipeline intercepts every tool response. Configurable redaction patterns strip sensitive fields (emails, SSNs, card numbers) before data reaches the AI agent.

05 — Cryptographic Audit Trail

Every request is signed into a SHA-256 hash chain with Ed25519 signatures. Events form a tamper-proof, SIEM-exportable forensic record.

06 — Honeytoken Trap System

Phantom credentials are injected into isolated environments. If a honeytoken is used outside Vinkius infrastructure, the server is quarantined instantly.

Emergency Kill Switch

EU AI Act Art. 14(1) Compliant

The kill switch is an **emergency halt** mechanism — not a simple toggle. When triggered, it executes three actions atomically:

01 — Server deactivated

The MCP server is immediately taken offline across the entire cluster.

02 — All tokens revoked

Every connection token is invalidated. Total lockout — reconnection blocked until new tokens are issued.

03 — WebSocket connections killed

Active connections terminated via Redis pubsub broadcast. Propagates to every runtime node in the cluster.

Full Visibility. Zero Guesswork.

The Vinkius cloud dashboard includes a full MCP Governance suite — real-time analytics and security controls for production AI operations.

Control Plane

KPI dashboard with request volume, latency, success rate, token consumption, and AI-generated operational briefings.

Agent Activity

Which AI clients are connecting, how often, and what they're doing — real-time session tracking.

FinOps

Cost tracking per tool, payload compression savings, budget optimization signals, and consumption trends.

Tool Health

Slowest and most error-prone tools, with actionable root-cause insights and performance baselines.

Firewall & DLP

PII redaction activity, sensitive data protection counters, and security event timeline.

Incident Log

Error trends, failure rates, status-code breakdowns, and forensic audit trail access.

> Get started at **cloud.vinkius.com** — connect your AI agent in under 60 seconds.

Internet Archive MCP

10 tools available
Cloud-hosted on Vinkius

This connector gives your agent access to an immense historical data vault. You don't need to learn complex query syntax or navigate endless menus; you just ask for what you want. Whether you're looking for academic papers from the 1920s, public domain films, or a snapshot of how a website looked ten years ago, your agent finds it automatically. It pulls data on everything: file formats, subject matter, and community reviews. Because Vinkius hosts this MCP, your client connects once to access this vast resource alongside thousands of other specialized tools.

It's all about natural conversation. You tell the AI what you need—a dataset from a specific decade, or the original source code for old software—and it handles the deep search and data aggregation process for you.

Core Capabilities

01 — Find content across massive collections

The agent searches pre-curated categories like Project Gutenberg ebooks or NASA images without needing to specify the collection name.

02 — Verify historical website versions

It checks if a given URL has been archived, returning the closest available snapshot date and link through the Wayback Machine.

03 — Search by specific criteria

You can narrow searches down to items created by an author or organization, or filter results only for movies, audio, or texts.

04 — Gather item details and stats

The agent retrieves full metadata—including subjects, file formats, and download statistics—for any found item.

05 — Review community feedback

It pulls user ratings and review texts from the community to help you assess an item's quality or relevance.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/internet-archive** — connect your AI agent in three steps.

01 Subscribe to this MCP in Vinkius. No API key is needed; it's a public, free resource.

02 Start your request using any MCP-compatible client (like Cursor or Claude). Simply ask the agent for historical content by topic, creator, or date range.

03 The agent executes the search and returns structured data including titles, identifiers, file formats, and direct links to download resources.

The bottom line is that you get deep access to a global digital library without writing any complex code or navigating multiple websites.

Built For

Historians, journalists, and academic researchers need this. They're tired of spending hours manually cross-referencing decades of data across dozens of different archives just to find a primary source photo or an old article.

Journalist

Uses the Wayback Machine tool to check how an opposing political group's website looked in 2018, verifying claims by finding archived versions of pages.

Academic Researcher

Searches for rare books or scientific papers from specific decades using search_by_date_range and get_item_metadata to build a comprehensive bibliography.

Content Creator

Uses the search_by_collection tool to pull public domain films or images for a video essay, ensuring they have clear usage rights information.

What Changes When You Connect

01 You get access to primary source materials. Instead of searching through limited academic databases, the search tool finds everything from old government films (Prelinger Archives) to rare scientific datasets.

02 Historical verification is instant. The wayback_availability tool lets you check any URL and instantly see if it was archived, telling you exactly when that snapshot occurred.

03 Data gathering becomes efficient. Use get_item_metadata to pull all the necessary citation info—creator, date, license, subject—before you even plan your download.

04 You can filter by format with search_by_mediatype. Need a playlist of vintage audiobooks? You limit results only to 'audio' and find them immediately.

05 It saves research time. Instead of writing complex database queries, the agent handles combining criteria like creator AND date range using the powerful search tool.

Real-World Applications

Tracing website evolution for journalism

A journalist needs to prove a company's messaging changed drastically in 2015. They ask their agent to check the URL using wayback_availability, finding multiple snapshots over time and retrieving metadata on content changes.

Building a film history database

A student wants all public domain films from the 1940s. They use search_by_collection with 'prelinger' and then filter by date using search_by_date_range to narrow down the decade.

Identifying original source materials

A developer is looking for old computing software. They use search_by_mediatype set to 'software' and then get_item_metadata on a promising result to check its specific file formats and download links.

Academic literature review

A researcher needs all articles written about climate change in the 1980s. They use search_by_date_range combined with 'pubmed' (a collection) to pinpoint primary academic sources from that specific period.

Patterns to Avoid

Searching for general concepts

✗ AVOID

Asking the agent, 'Tell me about history,' or 'Find something old.' The search results will be useless and overwhelming.

✓ INSTEAD

Always use specific tools. Instead of vague terms, combine criteria: Use search_by_collection to pick a source (like 'gutenberg'), then refine with search_by_date_range for the exact decade you need.

Assuming full access to data

✗ AVOID

Just asking for a file download without checking what formats are available, leading to an error.

✓ INSTEAD

Before downloading, use get_item_files on the item ID. This shows you exactly which formats (MP3, EPUB, PDF) and sizes are ready for you.

Ignoring specialized search syntax

✗ AVOID

Writing a query like 'World War II films' when they need to find only articles about the subject. This leads to mixed media types.

✓ INSTEAD

Use specific tools or refine your broad search with field-specific parameters (e.g., title:'WWII' AND mediatype:movies) in the main search tool.

The Right Fit

Use this MCP if your job requires deep, historical data retrieval from non-standardized archives, especially when verifying web content or accessing public domain media spanning decades. If you need to check a website's history, use wayback_availability; if you only need academic papers, stick to search_by_collection with 'pubmed'. Don't use this MCP if your requirement is simple: for example, if you just want today's stock prices or the latest news headlines. For real-time, rapidly changing data, a dedicated API connection (like a live financial feed) will be more appropriate than an archive.

The pain of building research bibliographies is massive.

Today, gathering source material means clicking through specialized sites. You find a potential article, then you have to manually check its date, the author's background, and if it’s public domain before you can even download it. Then you repeat that process for fifty different sources.

With this MCP, your agent handles all those manual checks in one go. You ask for 'all scientific articles about deep sea life from the 1950-1960 period,' and the system uses search_by_collection to find the right pool of content, then searches_by_date_range to narrow it down. You get a list of vetted sources ready to cite.

Getting complete source context with get_item_metadata

Before you copy-paste anything into your report, you usually have to open the item, find the citation box, then maybe click a separate 'details' tab just to see what collection it belongs to. This is tedious and easy to miss.

Now, running get_item_metadata provides everything upfront: the title, subjects, which collections it belongs to, and who published it—all in one shot. You save yourself ten minutes of clicking through dusty web pages.

Internet Archive: 10 Tools for Archival Research

Use these specific tools to narrow down your search, check metadata, find file formats, or verify the history of a URL.

#	TOOL	DESCRIPTION
01	`search_by_collection`	Search for items within specific curated categories like Project Gutenberg ebooks or Prelinger Archives.
02	`search_by_creator`	Find all content created by a particular author, organization, or artist name.
03	`search_by_date_range`	Filter results to find content from specific historical eras or decades using start and end years.
04	`search_by_mediatype`	Limit your search to only one format, such as movies, audiobooks, images, or software.
05	`get_item_files`	List all available download formats (PDF, MP4, etc.) and file sizes for a specific item ID.
06	`get_item_metadata`	Get complete details about an item, including its title, subjects, publisher, license, and total view count.
07	`get_item_reviews`	Retrieve community reviews and star ratings to gauge how useful or well-received a specific archived item is.
08	`get_views_stats`	Measure the popularity of an item by getting its total view count and, if available, geographical breakdown.
09	`search`	Perform broad searches across all media types using complex syntax like AND or title:"query" for maximum flexibility.
10	`wayback_availability`	Check if a given URL has been archived and find the closest available snapshot date via the Wayback Machine.

See It in Action

Real prompts you can use once this MCP is connected to your AI agent through Vinkius Cloud.

U Search for public domain films from the 1940s.

Found 12,847 films from the 1940s in the Prelinger Archives and community collections. Notable items: "It's a Wonderful Life" (1946), "The Scarlet Letter" (1934), various WWII propaganda and newsreel films, educational shorts, and home movies. Most are available for free download in MP4 and OGV formats. Would you like details on any specific title?

U Check if https://example.com has been archived.

✅ URL is archived! Closest snapshot: 20240101120000 (January 1, 2024). View the archived page at: https://web.archive.org/web/20240101120000/https://example.com. The Wayback Machine has captured this page multiple times over the years.

U Show me all NASA images available.

Found 185,000+ NASA items in the Internet Archive. The NASA collection includes: Apollo mission photographs, Hubble Space Telescope images, satellite imagery, astronaut training footage, Space Shuttle launches, ISS documentation, and planetary exploration photos. Items span from the 1960s to present day. Would you like to narrow by specific mission, decade, or image type?

Frequently Asked Questions

01 **How do I use Internet Archive MCP if I don’t know the exact name?**

You can start with a broad search using the main search tool. You just need to describe the topic, and the agent will help you refine it by date or media type.

02 **Does Internet Archive MCP handle modern websites?**

It uses the wayback_availability tool for this. If a site was online before, it checks its historical snapshots; otherwise, it won't find an archived version.

03 **Can I search for films and books at the same time using Internet Archive MCP?**

Yes. You can use the main search tool to combine criteria, like searching for 'climate change' AND limiting it by mediatype:movies or mediatype:texts.

04 **What is the best way to check a file's availability?**

Use get_item_files. This tool gives you a precise list of all available formats (PDF, EPUB, etc.) and their corresponding download links for that specific item ID.

05 **Is Internet Archive MCP only for American content?**

No, it covers global content. You can use search_by_collection to browse international libraries or use the main search tool with country-specific keywords.

Go Live in 60 Seconds

Get your connection token from **cloud.vinkius.com**, then paste the endpoint URL into any MCP-compatible client.

YOUR MCP ENDPOINT

```
https://edge.vinkius.com/[TOKEN]/mcp
```

CLIENT	WHERE TO CONFIGURE
Claude AI	Profile → Customize → Connectors → "+" → Add custom connector → Paste endpoint
Cursor	Settings → Features → MCP Servers → "+ Add New MCP Server" → Type: SSE → Paste endpoint
VS Code	Ctrl/Cmd+Shift+P → "MCP: Add Server" → add `"internet-archive": { "url": "…" }`
Windsurf	MCP Settings → `mcp_settings.json` → Add endpoint URL
ChatGPT	Settings → Tools & plugins → Add MCP server → Paste endpoint
Gemini	Extensions → Add MCP Server → Paste endpoint URL

ASK AN AI ABOUT THIS

Let your preferred AI explain this MCP server

- Ask ChatGPT
- Ask Claude
- Ask Perplexity
- Ask Gemini
- Ask Grok

READY TO CONNECT

Internet Archive is live on Vinkius Cloud.

Get your connection token, paste it into your AI agent, and start building. No SDK. No deployment. Just results.

Start at cloud.vinkius.com →

vinkius.com · support@vinkius.com

INDEPENDENT PLATFORM DISCLAIMER

Vinkius is an independent platform and is not affiliated with, endorsed by, sponsored by, verified by, or otherwise authorized by Internet Archive. All third-party trademarks, logos, and brand names are the property of their respective owners. Their use in this document is strictly for informational purposes to identify service compatibility and interoperability.

DOCUMENT INFORMATION

Generated	June 2026
MCP Server	Internet Archive MCP
Server ID	`019d75ba-dd54-717e-a982-2b18480312f5`
Platform	Vinkius Cloud for AI Agents
Endpoint	`https://edge.vinkius.com/{token}/mcp`

LICENSE & USAGE

This document is generated automatically by the Vinkius PDF Engine. Content reflects the MCP server configuration at the time of generation and may change as updates are deployed. For the most current information, visit vinkius.com/mcp/internet-archive.