MCP SERVER | NO CODE | CLOUD HOSTED

Internet Archive Metadata MCP

Analyze deep file history & item provenance

Internet Archive Metadata gives your AI client deep access to historical records. Get structured data—metadata, file lists, user reviews, collection memberships, and modification history for any item on archive.org. This MCP turns vast, unstructured public domain archives into actionable information you can query and analyze.

A+ Quality Score 98.33/100

digital-archiving | metadata-extraction | library-science | open-data | file-retrieval | historical-records

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

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 Metadata MCP

10 tools available
Cloud-hosted on Vinkius

Need to research an obscure piece of media or a niche historical record? This MCP connects your AI client directly to the Internet Archive's backend data structure. Instead of clicking through dozens of web pages just to compile facts—like checking file formats, finding out who reviewed it, or seeing how many times it was viewed—your agent handles it all in one query. You can ask for a complete item profile, pulling everything from the title and creator down to the storage location. If you're building complex knowledge tools, Vinkius makes this MCP available right alongside thousands of others, giving your AI client access to an unmatched depth of data sources. Your agent doesn't just summarize; it retrieves specific facts—from tracking changes over time to listing every single downloadable file format attached to the record.

Core Capabilities

01 — Identify item context and groupings

The MCP determines which larger collections or parent categories an item belongs to.

02 — List all available digital files

It pulls a comprehensive list of downloadable assets, detailing formats like PDF, EPUB, MP4, and their specific sizes.

03 — Extract community feedback and ratings

The MCP retrieves user reviews, including star averages and the text written by other users.

04 — Get item usage statistics

It provides access counts, showing how popular or frequently accessed the archived material is.

05 — Trace data evolution over time

The MCP tracks the modification history of an item, letting you see when and what changes were made to the record.

06 — View technical hosting details

It supplies server information regarding where the files are physically hosted.

One Click on Vinkius — From Prompt to Execution

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

01 You give your AI client an item's unique identifier (e.g., from its URL).

02 The MCP executes the necessary queries, pulling metadata, file listings, reviews, and statistics into a structured data payload.

03 Your agent receives clean, organized JSON or plain text containing all requested historical details.

The bottom line is that it turns manual web scraping into a single, programmatic query.

Built For

This MCP is for the digital historian, academic researcher, and content librarian. If your job requires cross-referencing facts from multiple sources—checking not just *what* an item is, but *how it's been used*, *where it lives*, or *who thought it was good*—you need this.

Academic Researcher

Uses the MCP to build comprehensive bibliographies on public domain media, using get_metadata and get_history to verify an item's provenance and track its evolution.

Digital Archivist

Leverages the tool to audit collections, checking get_parents for proper categorization or running get_stats to identify under-utilized assets that need promotion.

Historical Analyst

Employs it to compare different versions of a source by using get_derivatives and analyzing file listings to see what formats were available at key historical moments.

What Changes When You Connect

01 You instantly get full context on an asset. Using get_metadata ensures you don't miss the creator, license, or subject matter that defines a record.

02 Never worry about missing formats again. The ability to list all downloadable files via get_files shows you every format available—from PDF to MP3—all in one go.

03 You can gauge an item's relevance by checking community sentiment through the get_reviews tool, getting star ratings and user commentary right away.

04 Tracking changes is simple. Running the get_history function provides a clear timeline of modifications, which is critical for academic integrity and provenance research.

05 Quick checks are fast. If you only need basic item details without downloading massive amounts of data, use get_metadata_only to keep your queries light and fast.

Real-World Applications

Verifying source credibility

A student needs to cite old film footage. They ask their agent for the item's full metadata, then use get_history to see if key details (like the creator name or date) were corrected after initial upload. This verifies the source's reliability.

Optimizing digital asset libraries

A library manager wants to know which physical collections should be digitized next. They use get_collections and then check get_stats on related items, prioritizing those that have high download counts but no corresponding parent collection data.

Troubleshooting file access

A user can't open a specific file type. They prompt their agent to run the combined query for get_files and get_server_info, immediately identifying if the format is missing or if the hosting location needs updating.

Understanding content lineage

A researcher finds a derivative file but needs context. They use get_derivatives to see what was processed and then run get_parents to understand the broader thematic grouping of that content within the archive.

Patterns to Avoid

Assuming all data is available

✗ AVOID

A user sees a video title but assumes the accompanying educational material (like transcripts or PDFs) are present, leading to disappointment when only basic metadata loads.

✓ INSTEAD

Always run get_files first. This ensures you see every format attached to the item before making assumptions about what data types exist.

Ignoring content evolution

✗ AVOID

A user retrieves the initial, clean metadata but fails to check if the record has been updated since its original upload date.

✓ INSTEAD

Always include get_history in your query chain. This prevents you from using stale data and shows exactly when the item was last modified.

Over-querying basic facts

✗ AVOID

Running separate calls for title, creator, and date instead of requesting a full profile.

✓ INSTEAD

Use get_metadata. This single call pulls all core descriptive fields (title, creator, date, description) in one clean batch, making your process faster.

The Right Fit

You should use this MCP if your task involves deep data retrieval and context building around publicly archived media or documents. If you need to know *what* the item is (get_metadata), *how it's categorized* (get_collections/get_parents), *who liked it* (get_reviews), or *if it has changed* (get_history), this MCP is necessary. Don't use it if your goal is simple search—for that, basic keyword searching works fine. Also, don't run it if you only need a single piece of data; for instance, if you just want the file list, get_files is more efficient than running the full get_metadata call.

Use this MCP when multiple discrete pieces of historical information must be compiled into one cohesive answer. If your workflow involves verifying provenance or auditing asset metadata, stick with this tool.

The Pain of Manual Archival Research

Today, researching a single item from the Internet Archive means bouncing between multiple tabs: checking the general details page for title and creator. Then you have to click through another menu just to list all available formats—PDF, MP4, EPUB. After that, you might open yet *another* section to read what users thought of it, forcing you to copy-paste data points across a dozen different documents.

With this MCP, your agent handles the entire process automatically. You ask for an item's profile, and the tool returns one comprehensive payload containing everything: the core metadata, every file type available, the user reviews, and usage statistics. The messy clicking stops; you just get the facts.

Get Complete Item Profiles with Internet Archive Metadata MCP

The biggest manual time sinks disappear instantly. You no longer have to manually check if an item belongs to multiple collections or track down its modification history using separate web searches.

Now, you get a single, definitive data source that tells you the full story of an archived asset—from its initial upload details to every subsequent change and where it lives on the server.

Internet Archive Metadata: 10 Tools

Use these tools to query specific aspects of an archived item, whether it's the community reviews, parent collections, or the file formats available for download.

#	TOOL	DESCRIPTION
01	`get_collections`	This tool shows all the specific collections an item belongs to, giving you its structural context.
02	`get_derivatives`	It lists automatically processed versions of the original upload, such as optimized thumbnails or OCR text files.
03	`get_files`	This tool retrieves a list of every single downloadable file format available for that item.
04	`get_metadata`	It fetches the complete, core data about an item: creator, date, subjects, description, and license details.
05	`get_history`	This tool tracks every recorded change to an item over time, providing a full audit trail of modifications.
06	`get_metadata_only`	Use this when you only need the basic descriptive data about the item without pulling file lists or reviews.
07	`get_parents`	It reveals the higher-level categorization structure, showing which broader parent collections the item falls under.
08	`get_reviews`	This tool pulls community ratings and review text from users who have viewed the material.
09	`get_server_info`	It provides technical details on where the item's files are stored, useful for diagnosing access issues.
10	`get_stats`	This tool returns key usage metrics, including download counts and general access statistics.

See It in Action

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

U Get metadata for item big_buck_bunny.

big_buck_bunny — Creator: Blender Foundation. Date: 2008. Type: Animation/Short film. Duration: 10 minutes. Collection: Community Video. Available in MP4, OGV, and archival formats.

U List all files for item gutenberg_etext1.

Found 8 files: gutenberg_etext1.txt (plain text, 1.2 MB), gutenberg_etext1.epub (EPUB, 800 KB), gutenberg_etext1.mobi (MOBI, 900 KB), gutenberg_etext1.pdf (PDF, 2.1 MB), and various metadata files.

U Get reviews for item nasa_apollo11.

Found 23 reviews. Average rating: 4.8/5 stars. Top review from user 'spacefan42': "Incredible historical footage. The quality restoration is remarkable. A must-watch for anyone interested in space exploration."

Frequently Asked Questions

01 **How do I use Internet Archive Metadata MCP to find all file types for a record?**

Run get_files. This tool specifically lists every format available, whether it's plain text, an EPUB book, or a high-res MP4 video.

02 **Can Internet Archive Metadata MCP track if item details were changed over time?**

Yes, use get_history. It provides a modification timeline, letting you see exactly when the record was updated and what changes were made to it.

03 **Do I need to run all tools for full metadata on Internet Archive Metadata MCP?**

No. For basic facts, use get_metadata. If you also want community opinion, you'll need to supplement that by running get_reviews.

04 **What is the difference between get_metadata and get_metadata_only?**

get_metadata provides a comprehensive profile including files and reviews. get_metadata_only runs a lighter query, giving you just the core descriptive fields for faster lookups.

05 **How do I find out which collections an item belongs to using Internet Archive Metadata MCP?**

Use get_collections. This tool explicitly lists all the various groups or categories that contain the specific archived item.

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-metadata": { "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 Metadata 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

DOCUMENT INFORMATION

Generated	June 2026
MCP Server	Internet Archive Metadata MCP
Server ID	`019d75b6-42c2-72b2-8ea5-1a937a6a255e`
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-metadata.