MCP SERVER | NO CODE | CLOUD HOSTED

# data.world MCP for AI Agents

## Discover and Govern Enterprise Data Assets Using the Catalog

data.world connects your AI agent directly to an enterprise data catalog, letting you discover and govern organizational data assets through conversation. You can search across all available datasets and projects, retrieve detailed metadata, track project progress, and list saved SQL or SPARQL queries without ever leaving your chat interface.

A+ Quality Score 100/100

data-catalog | metadata-management | data-governance | collaborative-data | data-discovery | sql-queries

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/dataworld

SCAN TO VIEW

# The connectivity layer between AI and the world's software.

Vinkius sits between AI and every application. All communication passes through Vinkius Cloud via the Model Context Protocol (MCP) — with governance, observability, and security at every layer.

# Your AI Connections Run Through Vinkius Cloud

The world's largest managed MCP catalog

Vinkius is the connectivity layer where AI connects 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.

# data.world MCP

10 tools available
Cloud-hosted on Vinkius

Need to find a dataset but don't know where it lives? This MCP connects your AI agent straight into data.world, the enterprise catalog platform. You stop clicking through dozens of dashboards just to check if that metric exists. Instead, you ask your agent directly: 'What datasets track global emissions?' The agent searches the entire index for relevant assets and projects across your organization's data landscape.

It's more than just searching; it lets you manage governance too. You can review project status, list who owns an asset, or check historical activity logs to see when a dataset was last updated. When working with Vinkius, connecting this MCP gives your agent instant access to the full data catalog, making data discovery part of the conversation itself.

It means you spend zero time on manual metadata gathering and all your time analyzing what the data actually says.

## Core Capabilities

**01 — Search the entire data catalog**

Find datasets or projects by searching titles, descriptions, or tags across the whole organization's assets.

**02 — Retrieve detailed dataset metadata**

Get specific details for any asset, including field definitions, associated tags, and licensing information.

**03 — List owned datasets and projects**

Generate lists of all data assets or active projects you manage on the platform.

**04 — Access saved queries and insights**

List documented findings, visualizations, or specific SQL/SPARQL query definitions linked to a project or dataset.

**05 — Monitor data governance activity**

Review recent platform activity logs, including updates to datasets or changes in collection membership.

## One Click on Vinkius — From Prompt to Execution

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

01 Connect the data.world MCP to your AI client and authorize it using your API token.

02 Tell your agent what you need—for example, 'Show me all datasets related to Q3 sales' or 'What is the status of Project Phoenix?'

03 The agent executes the necessary lookup tools, returning structured metadata like field definitions, project members, and asset lists directly in the chat.

The bottom line is that your AI client uses this MCP to treat data discovery and governance tasks as simple conversations, not multi-step UI workflows.

## Built For

This tool is for anyone drowning in metadata. If you spend time clicking between spreadsheets, project dashboards, and the asset catalog just to confirm who owns a column or if that data exists, this MCP saves your day. It's built for people whose job requires knowing *where* the right data lives.

**Data Scientist**

You use it constantly to quickly find datasets and retrieve query definitions needed for a new analysis, saving hours of searching.

**Data Steward**

You monitor project status and audit data assets via chat commands, ensuring compliance and tracking ownership without manual dashboard checks.

**Knowledge Manager**

You explore organizational collections and documented insights during the initial stages of a new data planning effort.

## What Changes When You Connect

01 Instead of digging through UIs, you can use the `search_catalog` tool to instantly find data assets by title or tag. Your agent handles the index search so you don't.

02 The MCP lets your agent get detailed metadata using `get_dataset_details`. You immediately see field definitions and licensing info without navigating to a separate asset page.

03 Need to know project status? Calling `list_project_insights` means your AI client pulls all documented findings right into the chat, keeping your context clear.

04 You can track governance history by running `list_recent_activity`. This gives you an immediate audit trail of who changed what and when.

05 The agent supports structured queries. You don't just get a list; using `list_dataset_queries` provides the actual SQL or SPARQL code definitions, ready for review.

# Real-World Applications

**Figuring out who owns that weird dataset**

A data steward needs to audit an asset. Instead of guessing which department owns it, they ask their agent to run `get_project_details` or check the resource owners via the API call for project information.

**Building a report based on old findings**

A knowledge manager needs to understand historical data patterns. They prompt their agent to use `list_project_insights` for a specific project, instantly surfacing documented findings from previous teams.

**Comparing multiple datasets across projects**

A data scientist wants to see if 'Sales' data is tracked in three different places. They ask their agent to use `search_catalog` and then run `list_my_datasets` to compare the metadata of all available versions.

# Patterns to Avoid

### Trying to copy/paste metadata manually

✗ AVOID

A user opens the dataset dashboard, copies the license type, then has to open a separate project page and paste it into an email. It's slow and error-prone.

✓ INSTEAD

Use your agent to call `get_dataset_details` within this MCP. The agent retrieves all necessary fields, licenses, tags, and definitions in one step, letting you copy the final output directly.

### Only searching by keywords

✗ AVOID

The user types 'sales' but misses that the relevant dataset is tagged as 'revenue metrics'. The search fails because it was too vague.

✓ INSTEAD

Use `search_catalog` with specific criteria. Your agent lets you combine natural language intent ('all sales data') with structured filters (by tag or owner) to ensure precision.

### Confusing project scope with dataset scope

✗ AVOID

The user sees a list of datasets and assumes they are all connected, but forgets that the project needs specific inputs. They don't know what resources are required.

✓ INSTEAD

Use `get_project_details` to see exactly which datasets are linked to a project. This confirms the boundaries and dependencies before you start building.

## The Right Fit

Use this MCP if your workflow relies on knowing *where* data lives, who owns it, or what its specific definitions are. If you need an agent to confirm compliance details (license type, tags) or track the lineage of a metric across multiple projects, this is essential. Don't use it if all you need is simple text generation based on external knowledge; then your AI client's native capabilities suffice. You should also look at dedicated data visualization tools for actual graphing, but this MCP is perfect for the *metadata* work—for listing datasets ( `list_my_datasets` ), checking project scope ( `get_project_details` ), or retrieving code definitions ( `list_dataset_queries` ).

## data.world MCP: Solving Data Discovery Pain Points in Metadata Management

Today, finding a single piece of data is a manual nightmare. You open the internal wiki, then check the project dashboard, then log into the asset catalog, searching by vague keywords and hoping you remember which department owns the correct version. This process wastes hours just confirming metadata.

With this MCP, your agent handles that entire sequence. You simply ask, 'What is the most current dataset for Q3 revenue?' The agent uses the `search_catalog` tool to query across all available sources and provides a ranked list of options instantly, letting you proceed with confidence.

## data.world MCP: Improving Data Governance in Collaborative Platforms

Manual governance is reactive. You only realize data is missing or outdated when a report breaks. Tracking ownership, monitoring changes, and validating permissions requires constant manual checks across multiple dashboards.

This MCP shifts you to being proactive. You can ask the agent to run `list_recent_activity` or check project membership via `get_project_details`. This gives you an immediate, conversational view of who's involved and what has changed, turning governance from a chore into part of your daily workflow.

# data.world: 10 Tools for Metadata Management and Data Discovery

Use these tools to search the data catalog, retrieve project details, list datasets you own, and access technical metadata like field definitions and saved queries.

| # | TOOL | DESCRIPTION |
|---|---|---|
| 01 | `get_dataset_details` | Retrieves comprehensive details for one dataset, including its field definitions, tags, and license information. |
| 02 | `get_my_profile` | Fetches metadata about the currently authenticated user, showing your display name and account permissions. |
| 03 | `get_project_details` | Gets detailed information on a specific project, including its members, objective, and linked resources status. |
| 04 | `list_recent_activity` | Returns a log of recent actions across your data.world account, such as dataset updates or new contributions. |
| 05 | `list_my_collections` | Lists all organizational collections you manage, providing their descriptions and item counts. |
| 06 | `list_dataset_queries` | Generates a list of saved SQL or SPARQL query definitions associated with a specific dataset. |
| 07 | `list_my_datasets` | Provides a catalog listing of all datasets owned by your account, showing visibility and last modification dates. |
| 08 | `list_project_insights` | Lists documented findings, visualizations, or insights that have been published within a specific project workspace. |
| 09 | `list_my_projects` | Returns a summary of all data projects you own, including the linked datasets and your role (owner/contributor). |
| 10 | `search_catalog` | Searches across the entire data.world platform for relevant datasets or projects using full-text search. |

# See It in Action

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

**U** I need to analyze the scope for our new climate model. Can you list all relevant datasets?

**Climate Data Assets Found:**

• **NASA GISS Surface Temperature** (Tags: Climate, Global)
Details:* Tracks global temperature anomalies. Licensed under CC BY 4.0.
• **UN Emissions Stats** (Tags: Energy, Emissions)
Details:* Annual CO2 emissions by sector. Last updated: 2024-10-01

Do you want the specific field definitions for either of these?

**U** What was the last activity on 'Marketing Analytics' project?

Here's a quick look at recent changes for that project:

• **2026-06-11:** User Jane Doe updated the `Customer_Churn` dataset definition. (Type: Dataset Update)
• **2026-06-10:** System added 'Q3 Report' to the project's collection.
• **2026-05-28:** John Smith uploaded a new insight on regional spending patterns. (Type: Insight Added)

Looks like activity is steady!

U List the saved queries for our sales data.

I found 3 saved query definitions attached to the `Sales Transactions` dataset:

1. **Monthly Revenue Recap:** (Language: SQL) Calculates total revenue grouped by month.
2. **Top Performing Regions:** (Language: SPARQL) Identifies regions with highest sales volume.
3. **Q4 Churn Analysis:** (Language: SQL) Determines customer churn rates for the last quarter.

Let me know if you want to see the full definition for any of these.

# Frequently Asked Questions

01 **How does the data.world MCP help me find specific datasets?**

The data.world MCP allows your agent to search across all assets using full-text search, letting you pinpoint exactly which dataset exists without knowing its exact location or owner. It's like having a map of every piece of data in the company.

02 **Can I use this MCP to check project status and ownership?**

Yes. The agent can run tools to get detailed information on any specific project, showing who is a member, what the objective is, and which resources are linked. This saves you from having to open multiple dashboards.

03 **Is this data.world MCP good for data governance?**

It's excellent for governance because it lets your agent list recent activity logs, showing who modified an asset or project and when. You get a clear audit trail without doing manual checks.

04 **What if I need to see the code used in saved queries?**

The MCP can list all saved SQL or SPARQL query definitions for any dataset you specify. It retrieves the actual language and metadata, so you know exactly how the data is being processed.

05 **Do I need to be a data scientist to use data.world with this MCP?**

No. While it's powerful for analysts, any role that needs to find or manage corporate information can benefit. It simplifies the process of discovery and validation regardless of your job title.

# 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 `"dataworld": { "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

# data.world 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 data.world. 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 | data.world MCP |
| Server ID | `019d7582-8029-733d-b83b-f2998aa0d5ff` |
| 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/dataworld.