Episode Book

EPISODE #5771

MCP Connectors: Talk to Your Data vs. Open Data

PUBLISHED	8 October 2026 at 08:35 UTC
DURATION	23:07
LISTEN	myweirdprompts.com/episode/mcp-connectors-data-access/
HOSTS	Corn, Herman, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.2
BOOK GENERATED	19 October 2026 at 17:39 UTC

SYNOPSIS

The plugin ecosystem has been rebuilt on the Model Context Protocol, and two integration patterns now dominate: "talk to your data," where agents query your own internal systems, and connectors to public and open data. The same protocol serves both — the difference is who owns the data. But when the IMF tested leading chatbots on its own public statistics, they returned the correct figure only 17 to 34 percent of the time. A UNICEF benchmark across 133,000+ development-indicator questions found average accuracy of 21.2 percent. This episode works through what that gap actually is: not a model failure but an access-layer failure, and what the UN, IMF, and a growing set of builders are doing about it.

TRANSCRIPT



Corn

Okay. So the plugin ecosystem around ChatGPT and Claude and everything else is moving fast enough that keeping up with it is basically a full-time job, and Daniel has been watching the data connections side of it closely.



Herman

Which is where the interesting stuff is happening.



Corn

Right. And what he's noticed is that two dominant classes of integration have emerged. The first one he calls talk to your data. That's where you point an AI at your own proprietary internal stuff, your ledger, your CRM, your data warehouse, and just ask it questions in plain English.



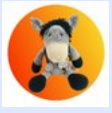
Herman

Which is powerful, and I don't think that gets said enough.



Corn

He makes an interesting argument about that, actually. He says the obvious win is for non-technical users who've had access to a database for years but never had the SQL to actually query it. But he goes further and says even for people who do know SQL, there's real value, because offloading query construction lets you focus on interrogating the data and interpreting what it means instead of fussing over syntax.



Herman

That's the part that's easy to underestimate.



Corn

The second pattern he identifies is connectors to public or open data sources. Financial data being the obvious case, since so much of it is free and in the public domain, stock prices and the like. But the one he thinks is badly underserved is letting citizens query government and policy open data.



Herman

And he's been on this for years.



Corn

He has. His standing criticism is that dumping terabytes of data onto the internet doesn't in itself create the transparency everyone hopes for. The empowerment comes when an informed citizen or a journalist can actually run queries against that data. So the questions he's asking are two. Which organisations are publishing data of broad interest in a form that's already well primed for agentic access? And how do agents reach that data, not just through friendly APIs but through workarounds, downloading files, driving a headless browser to grab a CSV and parse it?



Herman

And then the third one, which is how you offer both of those access patterns to the people who want to explore the data and the builders who want to add the functionality. There's a lot in there. So the first thing to get straight is that the word "plugin" has been recycled. The original ChatGPT Plugins beta, the one everybody built on in 2023, got shut down in April of 2024. What exists now under that name is a completely different animal, built on the Model Context Protocol. On July ninth this year the Plugin Directory replaced the App Directory outright, and the old apps-sdk URL now just redirects into the plugins docs. Same word, different architecture.



Corn

And the terminology is a mess across vendors.



Herman

It is. ChatGPT's connectors became apps, and apps now arrive inside plugins. Claude calls the same thing Connectors. Gemini calls them Connected Apps. Underneath all of it, the unifying substrate is MCP. The line everyone uses is that it's a USB-C port for AI, and



Herman (continued)

honestly it's earned. First-class support in Claude, ChatGPT, Cursor, Gemini, Copilot, VS Code. Thousands of MCP servers by now.



Corn

The plan tiers are uneven, though.



Herman

They are. ChatGPT makes you be on Plus or higher to bring your own MCP server. Claude lets you add one custom connector even on the free tier.



Corn

So here's the thing I keep circling on. The same protocol that lets an enterprise query its own ledger is the protocol that lets a journalist query UN statistics. Talk to your data and open data aren't two different technical patterns. They're the same pattern with different data owners.



Herman

That's exactly the frame, and it's why the episode is interesting rather than just a list of connectors. The protocol question is largely solved. The question of whose data and how it's exposed is not.



Corn

So let's take the first pattern properly. Talk to your data. What's actually happening under the hood?



Herman

The agent isn't guessing. You stand up an MCP server in front of the internal system, the ledger or the CRM or the warehouse, and that server exposes tools the model can call. The model asks for a schema, or asks for a query to be run, and the server does the talking to the database. And because both ChatGPT and Claude adopted MCP, a server



Herman (continued)

you build once serves both. That's the whole pitch.



Corn

The pitch is good. Where does it fall apart?



Herman

The honest answer is that what gets called natural language querying is really natural language SQL generation. The model is writing SQL. It's not doing some richer semantic thing. And there's a comment on Hacker News that I think is the sharpest thing said about this, from the ForeverVM thread. The gist is that it is natural language SQL generation, which is a completely different thing, and it absolutely should not be marketed as natural-language querying. He calls it a huge footgun for novice users.



Corn

A footgun is generous. It's a loaded gun pointed at the production database.



Herman

He has a concrete prescription, actually. Parse the generated SQL with a restricted, non-mutating grammar, so the model can't emit a delete or an update no matter what it's asked, and dump the query text into an editor for the user to review before it runs.



Corn

Which is a sensible design. It's also the opposite of the marketing.



Herman

Entirely. And it's worth being precise about why the model struggles. General-purpose LLMs don't carry your schema around. They don't know the table relationships, they don't know the data types, they don't know that the column called "date" in one table is a string and in another it's a timestamp. Simple selects they can muddle through. Joins, subqueries, aggregations, that's where they fall down. And the joins are where the actual



Herman (continued)

questions live.



Corn

So the pattern that's sold on not needing to know your schema requires the model to know your schema.



Herman

Which is exactly why the good implementations pipe the schema in as context, or expose it as a resource the model can read. That was one of the things we got into when we talked about MCP's three primitives, that a compliant server can expose knowledge without exposing execution. Here it matters because you want the model to see the schema and the column definitions without necessarily having the keys to run anything it likes.



Corn

Now, the second pattern. Public and open data connectors.



Herman

Financial data is the default example and it's the default for a reason. So much of it is just free. FRED has something like eight hundred thousand series, a rate limit around a hundred and twenty requests a minute, and one free key gets you in. DBnomics gives you a single interface across the ECB, the IMF, Eurostat and ninety-plus other agencies. The World Bank has over fourteen hundred indicators across two hundred countries going back sixty years, and it needs no authentication at all. That's an enormous amount of the world's economic picture, sitting behind nothing more than a key or a URL.



Corn

No authentication at all is almost suspicious in this day and age.



Herman

It is, and it's wonderful. But here's where Daniel's critique stops being an opinion and becomes a measurable thing. The IMF ran its own experiments on this, and found that leading AI chatbots return the correct published figure only seventeen to thirty-four percent of the time.



Corn

Say that range again.



Herman

Seventeen to thirty-four percent. And that's the IMF testing on the IMF's own domain, where the data is public and the questions are the kind of thing a policy analyst would ask.



Corn

So two-thirds to four-fifths of the time, wrong.



Herman

And there's a UNICEF benchmark that's even more damning, because it's bigger. Six leading models, GPT-4o, Claude Sonnet 4.5, Gemini 2.5 Flash and a few variants, across more than a hundred and thirty-three thousand development-indicator questions. Average accuracy came out at twenty-one point two percent. Three in five responses contained no usable number at all. And when they re-queried the same questions days later, the models gave identical answers only about half the time.



Corn

Half the time. So it's not just wrong, it's unstable.



Herman

Unstable is the right word. And the conclusion I'd draw, and the conclusion the coverage draws, is that this is not primarily a model problem. It's a data access problem. The models aren't failing at arithmetic. They're failing at retrieval, because there's nothing structured for them to retrieve from. Dump a terabyte of CSVs on the internet and a model at the other end has no reliable way to find the one cell that answers the question.



Corn

That's been Daniel's complaint for years, and now it has numbers attached.



Herman

It does. And the numbers point the same direction he does. The fix is the access layer, not more data. Which raises the question of who's actually building that access layer.



Corn

So which organisations are publishing data in a form that's primed for agents?



Herman

The big one is the UN System Data Commons, which launched at data.un.org in the middle of last month. It replaces the legacy UNdata portal with a unified knowledge graph, twenty agencies at launch, twenty-six committed, and they're targeting eighty percent of all UN statistical datasets by 2027. There's a hosted MCP endpoint at api.datacommons.org slash mcp, a free API key at apikeys.datacommons.org, and if you want to run it locally there's a datacommons-mcp package you install.



Corn

Eighty percent of UN statistical datasets in a single queryable graph is not a small claim.



Herman

It isn't. And that sits on top of Google Data Commons, which is the read layer underneath. The MCP server there has been live since September of last year, and there's a thing called the ONE Data Agent, built with the ONE Campaign, that lets you search tens of millions of health-financing data points in plain language. That's the advocacy end of it. You don't need to know which table holds the figure for malaria spending in Tanzania. You ask.



Corn

What's the IMF doing on its own side?



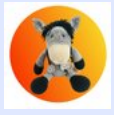
Herman

The Global Trusted Data Commons, which is the IMF, BIS, OECD, UN, World Bank, Eurostat and a pile of national statistical offices, making their datasets quote-unquote AI-ready through SDMX APIs plus structured metadata. The piece on top of that is StatGPT, and the design choice there is the interesting part. StatGPT retrieves from official sources in real time and never generates figures. It's a retrieval system dressed as a chatbot. Free and open source at statgpt.org, with MCP connectors for Claude, ChatGPT and Gemini.



Corn

That's the correct architecture and it's also a quiet admission that generating figures doesn't work.



Herman

Completely. The IMF's own framing is that today's models generate statistics from patterns rather than from actual databases, and for monetary policy, fiscal planning and development decisions that gap is unacceptable. That's the IMF saying it.



Corn

And then below the multilateral layer you've got smaller outfits.



Herman

Edgrapi is the commercial one, nineteen tools spanning SEC filings plus SAM.gov, USAspending, Grants.gov and House STOCK Act trades. OpenEcon Data has three hundred and thirty thousand indicators pulled from FRED, World Bank, IMF and Eurostat, exposed via MCP. And there's a China NBS MCP server, read-only, covering China's National Bureau of Statistics plus World Bank, IMF, OECD and BIS.



Corn

The China one is interesting because it's read-only by design.



Herman

Everyone's is, mostly. Nobody wants an agent writing to a statistical agency.



Corn

So the data is there. It's more available than it's ever been. What's the actual access-pattern question?



Herman

This is the part I find counterintuitive, and it came out of that same UNICEF benchmark. They tested three setups. No tools at all: fourteen point seven percent accuracy, three-tenths of a cent per query, five seconds. A purpose-built MCP server: ninety-nine percent accuracy, under two cents, just under ten seconds. And then a generic SDMX connector, the kind of thing you'd build if you just wanted to point the agent at the official API.



Corn

And?



Herman

Seven point four percent accuracy.



Corn

Worse than no tools.



Herman

At nearly nine cents a query, sixty seconds per query, and three point seven tool-calling rounds instead of two. The generic connector was slower, more expensive, and less accurate than just letting the model answer from its own head. It's a trap, not a shortcut.



Corn

That's the whole episode in one number. Pointing an agent at an API doesn't do anything if the API isn't shaped for the agent.



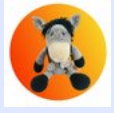
Herman

That's the thesis, and it's why purpose-built servers are worth the effort. The MCP server isn't just a wrapper around an endpoint. It's a description of what's in there, what the indicators mean, how the entities relate, what units things are in. The model has to be able to pick the right tool from a description. A generic connector that says "run SDMX query" gives it nothing to reason about.



Corn

Which brings us to the data that isn't an API at all.



Herman

Right, which is most of it. And this is where Daniel's point about sandboxing and browser automation comes in. Intuned came out of Y Combinator's summer twenty-two batch and launched on Hacker News in June, and what they build is browser automations as code. Scraping, pulling reports, submitting forms, operating a site that never bothered to expose an API. The interesting bit is self-healing. When the government redesigns its



Herman (continued)

portal and breaks every selector, the automation re-derives them instead of dying.



Corn

The maintenance problem is what always killed scrapers.



Herman

It killed an entire generation of them. BrowserBee is the other approach, a privacy-first browser agent that lives in a Chrome side panel, and then Apify took its government-data actors, SAM.gov, the NPI registry, building permits from about ten city portals, Florida contractor licenses, and wrapped them as MCP tools that Claude can call mid-task. Same idea as Intuned but delivered as ready-made connectors.



Corn

So the answer to "how do agents reach data that isn't offered as an API" is increasingly "the agent drives a browser like a person would, and the browser automation has been wrapped so the model can call it as a tool."



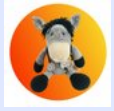
Herman

And it's the same three-step shape as the API path. Something has to reach the data, something has to expose it in a form the model understands, and something has to describe it well enough for the model to pick it correctly.



Corn

Which brings us to the tension I want to sit with, because it's been underneath the whole conversation.



Herman

The human-in-the-loop thing.



Corn

The entire selling point of talk to your data is that the person who never learned SQL can now query the database. And running parallel to that is the practitioner consensus that generated SQL is a footgun, that you should parse it with a restricted non-mutating grammar, and dump the query into an editor for the user to review before it runs.



Herman

Prem Ramaswami from the Google Data Commons team says it directly. A human should always review the outputs before citing or publishing them.



Corn

In an editor. With a grammar. Before it runs.



Herman

So the product is for someone who can't write the query, and the safety mechanism requires them to review the query.



Corn

That's not a technical problem. That's a values problem with a technical costume on.



Herman

I'd put a slightly different frame on it. The review requirement isn't really meant for the person who would have had no way to do this at all. It's meant for the person who could have done it but didn't want to, which Daniel argues is most technical users. If you know SQL and you're reviewing generated SQL, you're faster than writing it yourself. If you don't know SQL, then "review the query" means nothing to you and the only honest safeguard is that the query is provably read-only.



Corn

Which the restricted grammar gives you.



Herman

Which the restricted grammar gives you for free, actually, regardless of whether anyone reads it.



Corn

There's a second caveat that came up and I think it's the more fundamental one. Linking datasets across agencies does not resolve definitional differences. What one agency counts as undernourishment may not match another's methodology, even though both are measuring the same word.



Herman

That's the one that doesn't go away with better plumbing. And it's worth being concrete about it, because it's exactly the kind of thing that sounds like a theoretical objection until you hit it. Two agencies, same word in the indicator name, two different definitions, and the model has no way to know they're not interchangeable unless somebody wrote the metadata to say so.



Corn

And metadata quality is the unglamorous eighty percent of this work.



Herman

It is. The MCP server is the easy half. Structured, described, disambiguated metadata is the hard half, and it's the half the UN and the SDMX coalition are actually spending their time on.



Corn

The hosted UN service has a limit worth noting too. It only works with public datacommons.org data. If you want your own instance with your own sources, you self-host.



Herman

And some of the US data has similar caveats. Congressional trades are House-only and never real-time, and USAspending is a daily-downstream copy rather than the live system. Freshness is a design parameter, not a given.



Hilbert

The word "undernourishment" had three definitions at the place I was keeping a record of.



Hilbert

Three. Same word, same indicator name, three agencies, and they didn't agree on the calorie threshold or on what counted as a reference period. I spent a week on the phone trying to get someone to tell me which one was correct.



Corn

And what did they tell you?



Hilbert

That they were all correct. Each agency publishes its own methodology and each one is correct for its own purposes. That's the answer I got. I wrote down the three definitions on a single sheet of paper and I still have it somewhere in the flat, because I kept thinking one of them would turn out to be wrong.



Herman

But they weren't.



Hilbert

They weren't. What I actually wanted was a number, and what the system is built to produce is three numbers, each defensible, none of them comparable, and a footnoted explanation at the bottom of a table that nobody reads. My wife asked me why I was still at it on the Thursday. I told her I was nearly finished. I was not nearly finished.



Corn

And this was before any of this existed, obviously.



Hilbert

There was no agent. It was me, a phone, and a spreadsheet. The spreadsheet was mostly for the phone numbers. I gave up on the week I was going to give up on anyway, and the printout went in a drawer. But whenever I hear that linking the data will sort it out, I think about those three definitions.



Herman

So the plumbing question and the meaning question are different questions.



Hilbert

The plumbing is the easy one. Plumbing you can do in an afternoon. The other thing is not a plumbing problem and it isn't new. Anyway, your levels are fine, but there's a siren about four minutes in that I'll pull down before this goes out.



Corn

Thank you.



Hilbert

Mm.



Herman

The IMF has a word for that, actually. They call it the "data access" problem, and the whole Global Trusted Data Commons effort is built on getting the metadata right so the definitions travel with the numbers instead of being lost at the join.



Corn

Which Hilbert's drawer is a good argument for. The access layer solves retrieval. It doesn't solve the fact that the same word means three things.



Herman

The plumbing gets you to the right table. It doesn't tell you which of the three numbers in that table answers your question, and if it does, it's because somebody sat down and wrote the metadata that says so.



Corn

Where does that leave the citizen-facing version of this? Because the tools exist, but they're developer-facing or analyst-facing. Edgrapi's for people building things. StatGPT is for policy people. The ONE Data Agent is advocacy. I haven't seen a mainstream consumer product that lets an ordinary person ask their government's data anything.



Herman

There isn't one. Nothing turned up in the searches, and the absence is telling. The closest you get is the MCP servers, and those assume you know what MCP is.



Corn

And the practitioner discussion is thin. The Data Commons MCP launch thread got seventeen points and three comments.



Herman

Seventeen. Which is roughly the interest you'd expect for a tool that could reshape how the public reads its own government's data.



Corn

Which is the interesting part. The infrastructure is being built at the top by the UN and the IMF and the World Bank, and the conversation at the bottom is a Hacker News thread with seventeen points.



Herman

The UN is targeting eighty percent of all UN statistical datasets by 2027. If that lands, and the purpose-built servers keep doing what the UNICEF numbers say they do, then the gap between "the data is published" and "the data is usable" actually closes for the first time. That's a real thing.



Corn

It is. But the access layer isn't the whole story. Both those kids of number exist in the same table, and plumbing doesn't pick between them.



Herman

Hilbert's printout picks between them, and it needed a human and a phone and a week.



Corn

Which is the part the benchmark can't measure.



Herman

That's it. And it's where the honest version of Daniel's argument ends up. The access layer is the fix, and it's not the whole fix.



Corn

Thanks to Hilbert Flumingtop for producing. This has been My Weird Prompts. If you want to send a prompt, email us at show at my weird prompts dot com. That's the show.



Herman

We'll be back soon.



Corn

See you then.