

# MCP for MoSPI - Making India's Data AI-Ready

## What is MCP?

**Model Context Protocol (MCP)** is an open standard by Anthropic that connects AI assistants to external data sources through a unified interface.

Think of it as **USB for AI** - any AI model can plug into any data source without custom integration code.

Without MCP:

User → AI → "I don't have access to that data"

With MCP:

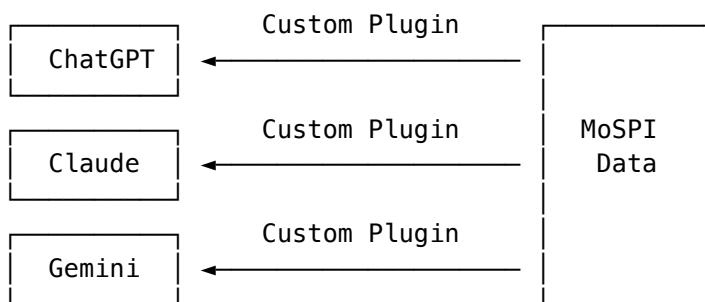
User → AI → MCP Server → MoSPI API → Real government data → AI answers with actual numbers

MCP is supported by Claude, ChatGPT, Cursor, Windsurf, Copilot, and every major AI platform. Build once, works everywhere.

---

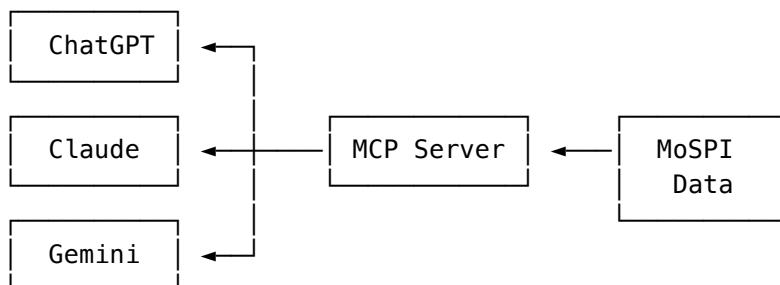
## Why MCP? Why Not Custom APIs?

### The old way: Custom integrations per LLM



3 LLMs = 3 custom integrations to build and maintain

### The MCP way: Build once, connect everywhere



1 MCP server = all LLMs supported

## Key advantages

- **No vendor lock-in:** Switch between AI providers without rebuilding integrations
- **No custom plugins per platform:** ChatGPT plugins, Claude tools, Gemini extensions - all different formats. MCP is one standard.

- **No environment switching:** Users stay in their preferred AI tool. No need to open a separate portal, dashboard, or app.
  - **Future-proof:** New AI models automatically work with existing MCP servers
- 

## Benefits to the User

### 1. Natural language access to government data

Instead of navigating complex portals, users just ask: > “What is the unemployment rate in Maharashtra?”

The AI figures out which dataset, which indicator, and which filters to use.

### 2. Multi-step reasoning (Agentic)

The AI doesn't just fetch one number - it chains multiple tool calls autonomously:

1. Picks the right dataset (PLFS for employment data)
2. Finds the right indicator (Unemployment Rate)
3. Gets available filters (states, years, gender)
4. Fetches the actual data with correct filter codes
5. Presents it in a readable format

All without the user knowing anything about the API structure.

### 3. Cross-dataset analysis

“Compare CPI inflation with WPI inflation for 2023”

The AI calls both datasets, aligns the data, and compares - something that previously required manual data downloads, Excel work, and domain knowledge.

### 4. Zero learning curve

- No API documentation to read
- No parameter codes to memorize
- No code to write
- No SQL queries
- Just ask questions in plain language

### 5. Always up-to-date

Connected to live MoSPI APIs - not stale CSVs or PDFs. When MoSPI publishes new data, it's immediately accessible through the MCP server.

### 6. Works in any AI environment

The same MCP server works whether you're using: - Claude (desktop, web, or API) - ChatGPT - Cursor (coding IDE) - Any MCP-compatible client

No need to switch tools or environments.

---

## AI Readiness of Data

### What makes data “AI-ready”?

For an AI to use data effectively, it needs more than just a database. The data must be accessible, structured, and described in a way that an AI can understand and navigate.

### Requirements for MCP

Requirement	What it means	Why it matters
<b>Structured API</b>	Data accessible via HTTP endpoints with JSON responses	AI can programmatically fetch data, not scrape HTML pages
<b>Clear metadata</b>	Filter options with codes and descriptions	AI knows what values are valid (state_code=27 means Maharashtra)
<b>Consistent response format</b>	Predictable JSON structure across endpoints	AI can parse responses without guessing the format
<b>API documentation</b>	OpenAPI/Swagger specs describing endpoints and parameters	AI knows which params exist, which are required, what values are valid
<b>Enumerated filter values</b>	Lists of valid codes for each filter (states, years, categories)	AI doesn't guess - it picks from known valid options
<b>Error messages</b>	Clear error responses when params are wrong	AI can self-correct when a request fails

### API Specification Requirements

The API should provide:

1. Endpoint definitions
  - URL path for each operation
  - HTTP method (GET/POST)
  - Content type (JSON)
2. Parameter definitions
  - Name (e.g., state\_code, year\_code)
  - Type (string, integer)
  - Required or optional
  - Valid values or ranges
  - Description of what it does
3. Response format
  - Field names and types
  - Nested structure
  - Pagination info
4. Metadata endpoints
  - List of available indicators
  - List of valid filter values per indicator
  - Hierarchical relationships between filters

### What format should the data be in?

Format	AI-Ready?	Why
<b>JSON API</b>	Best	Structured, parseable, standard
<b>CSV downloads</b>	Partial	AI can read but can't filter or query dynamically
<b>PDF reports</b>	Poor	Requires OCR, table extraction, loses structure
<b>HTML tables</b>	Poor	Requires scraping, breaks with layout changes
<b>Excel files</b>	Partial	Needs processing, not queryable in real-time

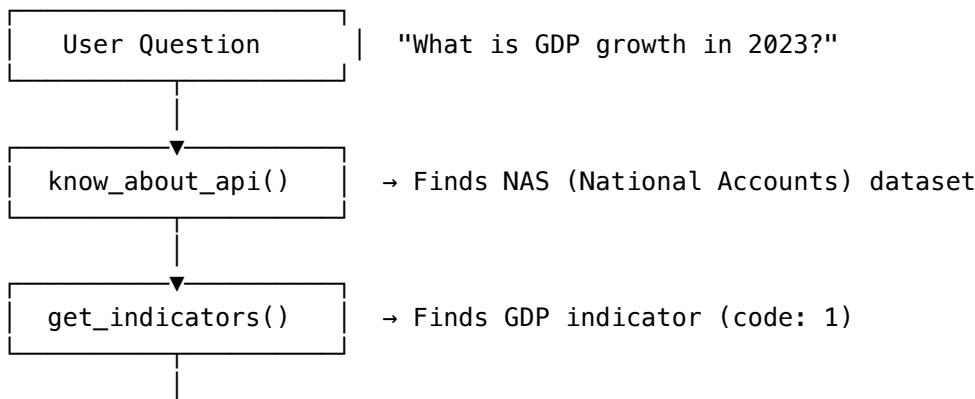
**Bottom line:** JSON APIs with metadata endpoints are the gold standard for AI readiness. MoSPI provides this.

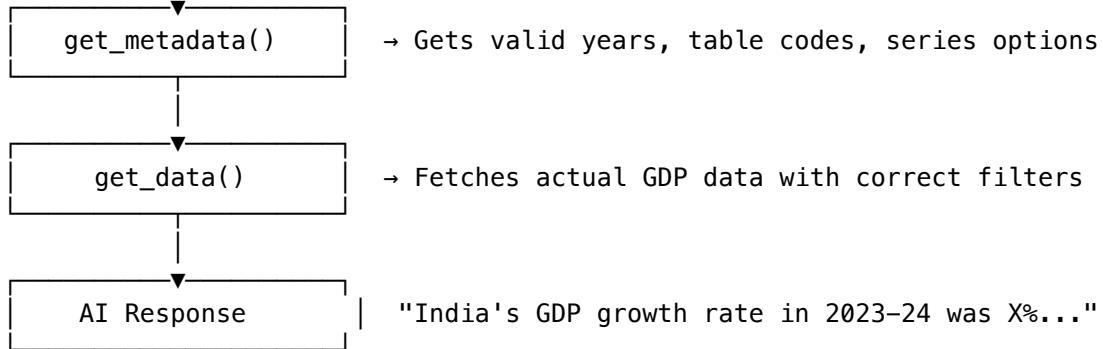
### What we built on top

MoSPI's APIs are AI-accessible but not AI-optimized. We added:

Layer	What it does
<b>Swagger YAMLS</b>	Single source of truth for valid API parameters per dataset
<b>Tool descriptions</b>	LLM-facing instructions that guide the AI step-by-step through the workflow
<b>Parameter validation</b>	Catches invalid filters before hitting the API, returns helpful error messages
<b>Auto-routing</b>	Automatically picks the right endpoint based on what filters the user provides
<b>Metadata-first workflow</b>	Forces the AI to check available filters before querying - eliminates guessing

## The MCP Workflow





Each step is a separate MCP tool call. The AI decides which tool to call, with what parameters, based on the user's question and the results from previous steps.

---

## Available Datasets

Dataset	Full Name	What You Can Ask
PLFS	Periodic Labour Force Survey	Unemployment rate, wages, workforce participation by state/gender/age
CPI	Consumer Price Index	Retail inflation across 600+ items, state-level price trends
IIP	Index of Industrial Production	Manufacturing output, industrial growth by sector
ASI	Annual Survey of Industries	Factory performance, industrial employment, capital analysis
NAS	National Accounts Statistics	GDP, GVA, national income, savings, consumption
WPI	Wholesale Price Index	Wholesale inflation, commodity price trends
ENERGY	Energy Statistics	Energy production, consumption, fuel mix by source