

Beyond Keywords

*AI-Mediated Access to the Islam West Africa
Collection through an MCP Server and Agent Skill*

Frédéric Madore

Data Curator

Africa Multiple Cluster of Excellence



Scan for these slides

slides.frederickmadore.com

An open archive, built by one historian

Fifteen years of fieldwork and library work — Benin, Burkina Faso, Côte d'Ivoire, Togo — turned into an open-access collection.

- Mostly newspaper clippings and Islamic publications: magazines, brochures, tracts
- Behind it, a personal library of 30,000+ **Zotero** items
- Online at islam.zmo.de

— IN NUMBERS

DOCUMENTS **14,700+**

PAGES **68,122**

AUDIOVISUAL MINUTES **9,315**

WORDS **28.1M**

DOCUMENT TYPES **13**

REFERENCES **864**

Six countries · mostly French — also Arabic & Hausa · open access since 2023

An overview of the collection

Islam West Africa Collection

Search

EN

HOME BROWSE REFERENCES INDEX EXHIBITS VISUALISATIONS DIGITAL HUMANITIES & AI ABOUT

Islam West Africa Collection

Home

While digitisation of Islamic heritage in Africa has often prioritised manuscripts, the rich print culture of contemporary West Africa remains undervalued and at risk. The *Islam West Africa Collection* (IWAC) addresses this gap by preserving newspapers and Islamic publications: essential sources for understanding how Muslim communities have shaped — and been shaped by — the region's modern history.

Directed and developed by [Frédéric Madore](#) and hosted by the [Leibniz-Zentrum Moderner Orient \(ZMO\)](#), IWAC is a collaborative, open-access digital database. Building on the award-winning *Islam Burkina Faso Collection* launched in 2021, the repository now brings together over 14,700 items: [archival documents](#), [newspaper articles](#) (both scanned and online versions archived via the [Wayback Machine](#)), [Islamic publications](#), [audiovisual recordings](#), and [photographs](#) relating to Islam and Muslims



SCAN TO EXPLORE

islam.zmo.de

Why keyword search falls short



The faceted search of the IWAC.

Keyword search has built-in problems. I mitigated what I could:

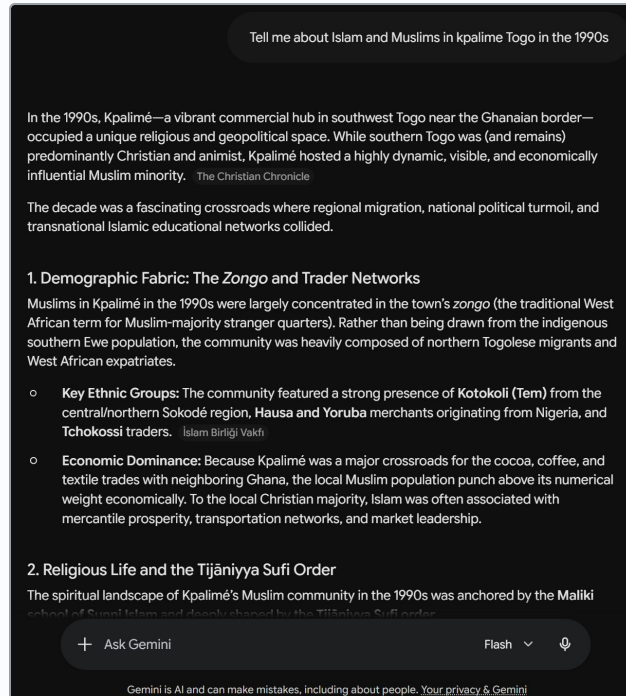
- **Spelling variants** → authority files merge 20+ forms
- **Forming a query** → autocomplete and facets guide you
- **Polysemy** → "hadj" = pilgrimage *and* the "El Hadj" honorific

THE REAL GAP

The search box assumes you already know it.

— WHY NOT JUST ASK AI?

Generic answers, dubious sources



Gemini on Islam in Kpalimé: generic and dubiously sourced. [See the chat ↗](#)

- People now ask AI directly, not a search box
- Chatbots often invent citations — worse for Africa, with far less training data
- Web search helps, but the sources are often random
- The IWAC? Crawled by AI bots, yet barely used and badly cited

— AN APPROACH

Two layers: plumbing and intelligence

THE MCP SERVER — PLUMBING

Structured, **read-only** access to the collection that any AI assistant can call.

THE CLAIM TO HOLD ONTO

The skill is where the real contribution lives. github.com/fmadore/iwac-mcp-server

THE AGENT SKILL — INTELLIGENCE

The methodology for using it *well* — a historian's method, written down.

Why a server, not just a chatbot?

An MCP server connects an AI assistant to one outside system. Like an app to your accounts:

EMAIL

GOOGLE CALENDAR

GOOGLE DRIVE

SLACK

GITHUB



IWAC

RAG

Retrieval-augmented generation. Proven for GLAM, but it chunks records, and retrieves by a similarity the user never sees.

THE ANALOGY

Like IIIF for images: a shared interface, without handing over the originals. Build the server once — Claude, ChatGPT, or open-source models all connect.



MCP

Ingests **nothing**. The data stays put; the server holds the logic; the model just asks and answers.

modelcontextprotocol.io

What is a skill.md?

A skill is a **methods handbook** for a **machine reader** — loaded only when a request matches its purpose.

- Competence by **reading, not retraining**
- Versioned, reviewed, refined — a prompt is typed once and lost
- An open standard across agents — no vendor lock-in, like MCP
- **Born in coding agents** — the format fits any domain

```
my-skill/  
├─ SKILL.md           # Required: metadata + instructions  
├─ scripts/          # Optional: executable code  
├─ references/        # Optional: documentation  
├─ assets/            # Optional: templates, resources  
└─ ...                # Any additional files or directories
```

The standard layout of an agent skill — SKILL.md is the only required file.

— WHAT YOU CAN ASK IT

The IWAC MCP server

"Newspaper articles on Islam in Abidjan, in the 1990s"

ARTICLES PLACE · PERIOD

"Islamic publications on secularism in Togo and Benin"

PUBLICATIONS SUBJECT · COUNTRY

"Academic references on Muslim women in Burkina Faso"

REFERENCES SUBJECT · COUNTRY

One plain-language request; the model picks the tools and combines the filters.

22 TOOLS · 6 SUBSETS

- **Read-only:** search, full text, authority records, sentiment, stats
- **Verifiable:** every call a lookup, no AI at query time
- **Traceable:** every result links to its canonical record
- **Two ways in:** one-click for Claude Desktop, a hosted URL for ChatGPT

Search by meaning, not words

1 · ASK

Your question

« *Islam et laïcité au Burkina Faso* »

Plain language — and any language works.



2 · EMBED

Turned into coordinates

An embedding model (Gemini) maps the text to **768 numbers** — one point in a space of meaning.

laïcité → [0.07, -0.12, 0.93, ...]

Every article was mapped the same way, in advance.



3 · MATCH

Nearest points win

Cosine similarity ranks all **12,000+** articles by how close they sit — closest in meaning, not in wording.

WHY IT MATTERS

Finds the right articles even when they never say « *laïcité* » — secularism, Franco-Arab schools, family-code debates. The one tool that runs a model live, just to turn your question into coordinates.

Keyword vs. meaning — in the slide itself

🔍 *laïcité au Burkina Faso*

KEYWORD

BY MEANING

2 DOCUMENTS · EXACT-STRING MATCH ONLY

« ... le débat sur la **laïcité** au Burkina Faso ... »

L'Observateur Paalga · 1994

« ... la **laïcité** de l'État et les écoles ... »

Sidwaya · 2002

THE BLIND SPOT

Dozens of articles are *about* secularism but never type the word — schooling, family-code debates, the Arabic press. Keyword search cannot see them. Click **By meaning**.

A five-phase research method

Two depths, chosen up front:

Brief DEFAULT

a quick scan and a few close reads

Extended

the full five-phase analysis below

- 01 **Scoping** — what does the collection even hold for this question, before any keyword?
- 02 **Systematic searching** — French + transliteration variants (Tabaski = Eid al-Adha); log every search, including null results
- 03 **Deep reading** — full text and sentiment, article by article

- 04 **Triangulation** — cross-check articles, publications, references, index entries
- 05 **Synthesis** — findings with source attribution and confidence grades

THE POINT

Twenty-two tools give reach. The "judgement" lives in the skill.

— WHAT A SERVER NEVER COULD

Discipline, disclosure — and a new role

- **Linguistic discipline** — search the French corpus in French, whatever language the question is in
- **Bias disclosure** — every synthesis states the skew: francophone tilt, missing *Arabissants*, thin Niger and Nigeria
- **Evidential discipline** — claims tagged primary / secondary / AI-derived; *absence of evidence* ≠ *evidence of absence*

THE ROLE SHIFT

Cataloguer, then API designer — now **author of research methods a machine can follow**. The server took a weekend; the skill is where the expertise lives.

iwac-mcp - SKILL.md

```
---
name: iwac-mcp
description: |
  Structured academic research workflow for the Islam West Africa Collection (IWAC) MCP server.
  Use this skill when:
  - Conducting research queries through the IWAC MCP server (iwac-mcp-server tools)
  - Investigating questions about Islam and Muslims in West Africa using IWAC data
  - Performing systematic searches across IWAC articles, publications, index, references, or documents
  - Analyzing sentiment or temporal patterns in West African press coverage
  - Comparing coverage across countries, newspapers, or time periods
  - Building structured research outputs with source attribution and confidence grading
  This skill provides a five-phase research methodology, search strategy guidance for francophone sources, bias awareness, and documentation conventions. It complements the
  iwac-data skill (data schemas + Omeka S API).
---

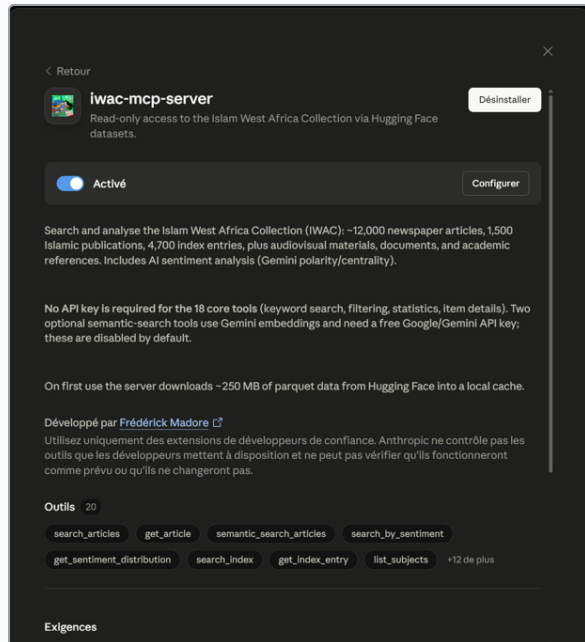
# IWAC MCP Research Workflow

Structured methodology for academic research using the IWAC MCP server's 22 tools (20 core + 2 optional semantic). Adapted from ALA-compliant archival research practices.
Applies to server **v0.6.0+** – all matching is accent- and case-insensitive, and result objects use short English keys (`id`, `date`, `polarity`, `centrality`,
`subjectivity`, `description_ai`, `url`).
```

↓ Scroll to read the whole file — 190 lines, open source at
github.com/fmadore/iwac-mcp-server

— LIVE DEMO

From a plain question to cited sources



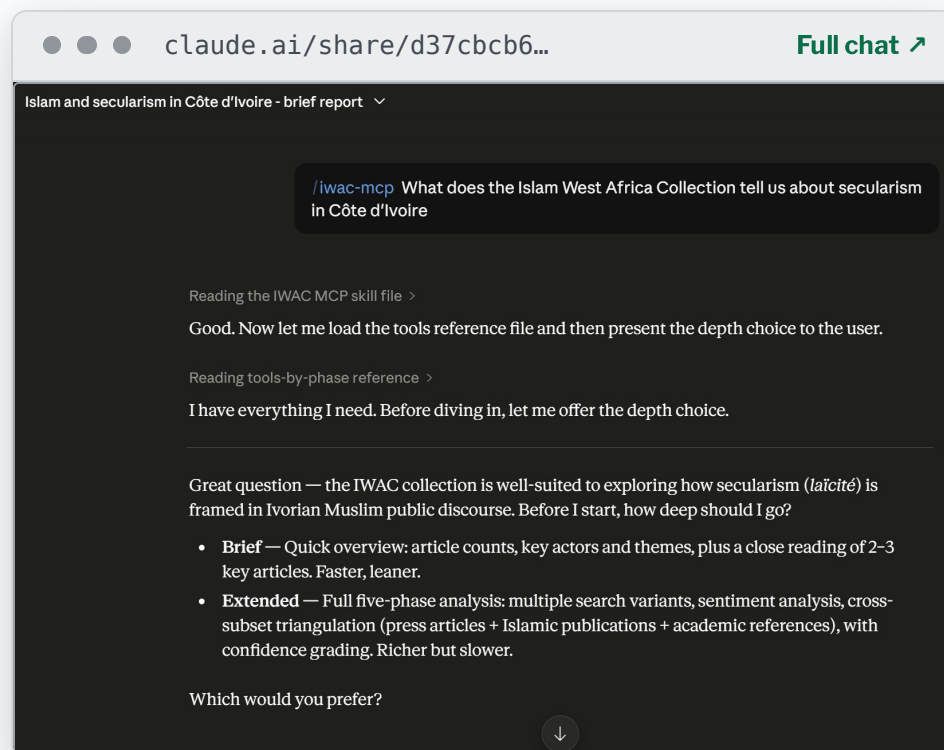
Installed as a Claude Desktop extension — read-only, no API key for the core tools.

One real question, and the skill runs the method live.

- Scopes the collection, then searches it in French
- Reads the strongest hits in full
- Answers with every claim linked to its IWAC record

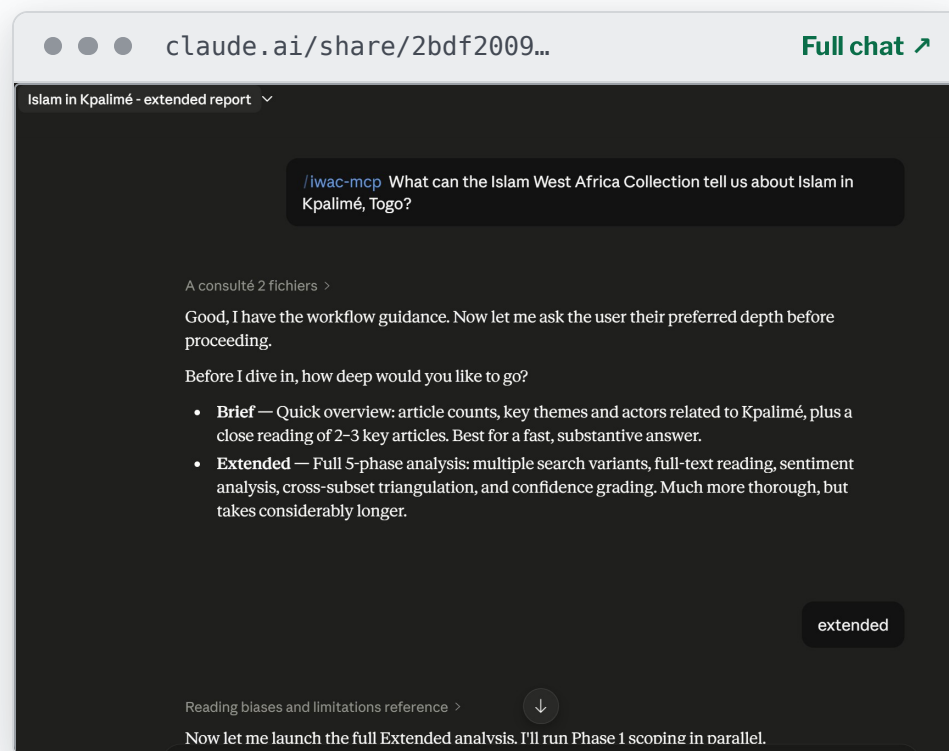
↓ Two saved runs below — press ↓

Brief — the default depth



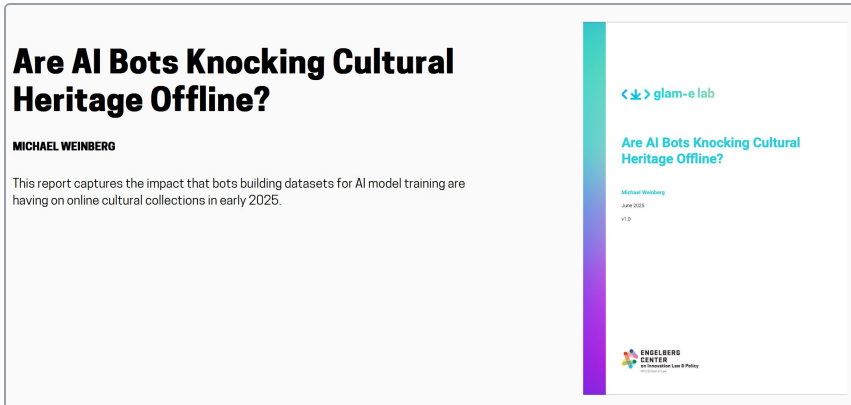
Claude Sonnet 4.6 · high effort

Extended — the full method



The same question we put to Gemini earlier, now answered from the IWAC.

Whose infrastructure? African collections and AI extraction



39 of 43 heritage institutions hit by AI-bot traffic spikes (Weinberg, GLAM-E Lab, 2025); Wikimedia: **65%** of its costliest traffic is bots. [Read it ↗](#)

- The default is **extraction** — public institutions and open-access labour subsidising commercial AI
- **MCP, a third path** — the collection stays put; the institution decides what the AI sees

YÉKÚ'S PARADOX · 2026

Un-digitised, an African archive is invisible to AI; digitised, it's exposed to scraping — the incomplete corpus becomes the *single story*, the colonial library rebuilt in the training set. [Read the essay ↗](#)

Conclusion

A new way in for a **wider public** — but no replacement for archival research. It finds what you ask for, not what you weren't looking for: the unexpected find that slow reading turns up.

Try it: connect Claude or ChatGPT to the collection — how-to on the **IWAC AI-access page**.

Next: a small open-source model on the server itself, answering on the website — no install, no account. github.com/fmadore/iwac-mcp-server



Scan for these slides

slides.frederickmadore.com

 [fmadore](#)  frederickmadore.com  [0000-0003-0959-2092](#)  [frederickmadore](#)  frederick.madore@uni-bayreuth.de

