ai + history collaboratory

Session 3

Exploring MCP as an Interface with Historical Databases

24 February 2026 | 4:00 pm UK

today's agenda

1	**Quick introductions** *Members + two guests*	**5 min**
2	**What is MCP? Three models** *Typology, ecosystem, who's building*	**10 min**
3	**Dan Cohen — loose coupling** *Northeastern University Library + screenshots*	**5 min**
4	**Tom Scheinfeldt — reference interview** *Sourcery for ArchivesSpace — demo*	**10 min**
5	**Ra-MCP — what we built & found** *Five-phase methodology, Solr testing — demo*	**20 min**
6	**IWAC + use cases + wrap-up** *Frédérick Madore's collection, your databases, next steps*	**10 min**

our group

members

Gavin Beinart-Smollan
Maurice Brenner
David Brown
Abi Cunningham
Marc Eagle
Jacob Forward
Colin Greenstreet (convenor)
Lucas Haasis
Sam Kaislaniemi
Thiago Krause
Oren Okhovat
Mark L. Thompson

Cambridge, East Anglia, Groningen, NYU, Trinity College Dublin, Wayne State, Western Kentucky, Yale, York

session 3 guests

Frédérick Madore
University of Bayreuth — Data Curator, IWAC
Islam West Africa Collection; AI + historical sources

Jiayu Yang
University of Bayreuth — Data Curator

Oliver Baumann
University of Bayreuth — Data Curator

Folami Kolade
Temple University — PhD Student
Historical and business research methodologies

what is MCP? three architectural models

Model Context Protocol — a standard that defines how AI models request and receive information from external data sources. Think of it as a universal plug.

A. SOURCE-ORIENTED

One server per archive or database

Analogy: *A dedicated assistant for each library you visit*

Examples: Riksarkivet, Gallica, Northeastern Library, MarineLives

B. TASK-ORIENTED

One server per research function across sources

Analogy: *A research assistant who can search many libraries*

Examples: Paper Search (arXiv + PubMed + Scholar), Multi-source academic

C. PROTOCOL-ORIENTED

One server per data standard or protocol

Analogy: *A translator who speaks every library's language*

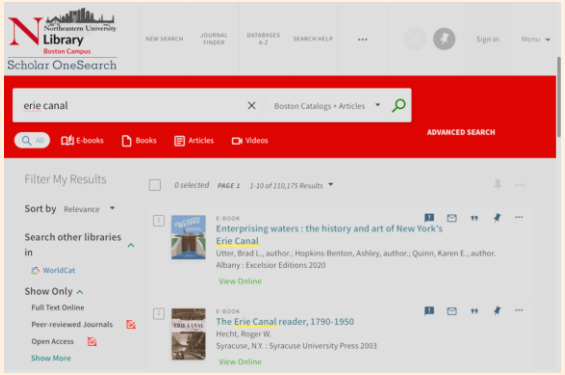
Examples: SPARQL (Wikidata), IIIF (manuscript images), OAI-PMH (metadata)

Who's building? Mostly solo developers. Institutions provide APIs but not MCP servers — with notable exceptions: Riksarkivet (AIRA), Northeastern Library (Cohen), UConn (Scheinfeldt).

Dan Cohen: loose coupling in practice

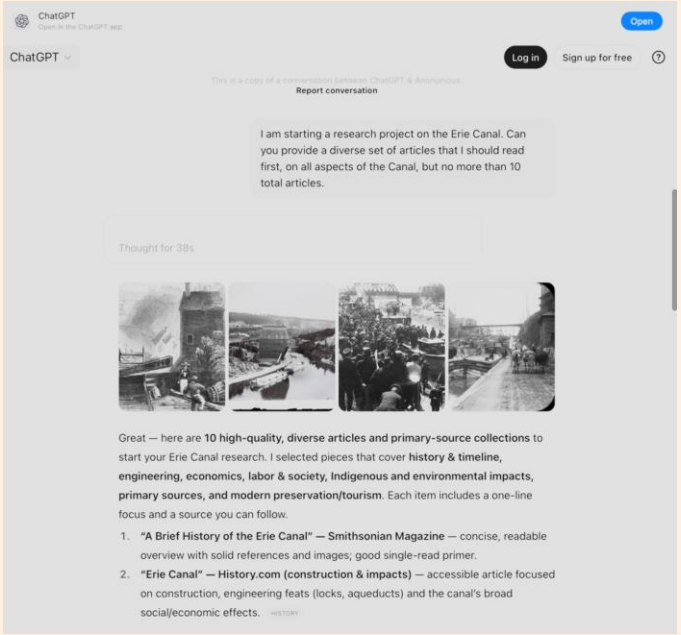
Institutions retain control while providing structured AI access — a dial, not a switch.

LIBRARY SEARCH



110,175 results — overwhelming

CHATGPT



Popular web articles — no depth

CLAUDE + MCP

Curated scholarly articles, linked

Northeastern University Library MCP Server + Claude plugin (October 2025)

"We want to foreground the expressive works of human beings — the articles, books, documents, and works of art, rather than the AI's digests of these objects."

Tom Scheinfeldt: the reference interview

Sourcery for ArchivesSpace — a conversational AI archivist

Iterative search refinement

When initial terms fail, suggest alternatives: 'Try trade instead of merchants'

Transparent scope explanation

Explains what the collection contains and what it doesn't

Graceful escalation

'Connect with Archivist' button — escalate to a human when AI reaches its limits

Conversational guidance

Guides researchers through unfamiliar collections with knowledge of structure

→ LIVE DEMO

ra-mcp: what we built

Adapting Scheinfeldt's reference interview into a formalised academic research methodology

1. **Scoping**
   Map archives, identify digitised holdings, locate published finding aids
2. **Term Mapping**
   Build controlled vocabulary in Swedish with wildcards, fuzzy search, historical variants
3. **Deep Reading**
   Browse full page transcriptions, record reference codes, note bildvisaren links
4. **Triangulation**
   Cross-reference with secondary literature, finding aids, and web sources
5. **Synthesis**
   Structured output with three-tier confidence grading and source-type tagging

Beyond Scheinfeldt

1. Confidence grading (strong / moderate / weak)
2. Null results as evidence
3. Survivorship bias awareness
4. Triangulation against secondary literature
5. Growing reference library

→ **DEMO: switch to Claude Desktop Chat**

ra-mcp: what we found

From a single exploratory session: Swedish maritime jurisdiction compared with the English High Court of Admiralty

Sjörätt discovered

Local maritime courts in port towns — Skanör, Helsingborg, Landskrona — found in 1780s-1800s Göta hovrätt appeal records. First-instance courts; no separate admiralty appellate chain.

Strong: direct archival attestation

Sjölagen mapped

153 page-level hits for the codified Maritime Code. Named divisions: Seamen's Code, Ship Charter Code, Bottomry Code, Insurance Code. Fundamentally different from English common law tradition.

Strong: abundant references

Admiralty = military, not judicial

Amiralitetskollegium (1634-1791) held entirely at Krigsarkivet (War Archives). Navy administration, not courts. No Amiralitetsrätt found anywhere in civilian archives.

Moderate: clear archival provenance

The key comparative finding

England: one centralised national specialist court (HCA), operating under Roman civil law — anomalous in a common law country.
Sweden: local Sjörätter in port towns, feeding into general appellate hovrätter, governed by a codified statutory Maritime Code (Sjölagen) — standard civil law.
Prize jurisdiction: not resolved. Zero archival hits. The most significant open question — where were Swedish prize cases adjudicated?

From: RA_MCP_Swedish_Maritime_Jurisdiction_Research_Memo_v1.1 (17 Feb 2026) — a collaborative human-AI research session using Claude Cowork + ra-mcp

IWAC + your collections

Islam West Africa Collection

Frédérick Madore, University of Bayreuth

- 14,500+ documents on Islam in West Africa
- Omeka S platform with REST API
- ~96% French; also Hausa, Arabic, Dendi
- OCR on all documents; IIIF-compliant
- 4,600-entry thematic index
- Existing AI pipelines (OCR correction, NER, HTR)
- GitHub: fmadore/Islam-West-Africa-Collection

Could the ra-mcp methodology transfer?

Your collections, your questions

- What databases or archives do you work with day-to-day?
- Does your institution provide an API? Is it stable?
- Which architectural model (source / task / protocol) fits your research?
- Where could an MCP server change how you interact with your collections?
- What would you want to build — or see built — by June?

pre-reading & resources

1. "What is the Model Context Protocol (MCP)" — the official introduction
   modelcontextprotocol.io
2. Anthropic, "Claude can now connect to your world" (May 2025)
   anthropic.com
3. Dan Cohen, "AI and Libraries, Archives, and Museums, Loosely Coupled" (Aug 2025)
   newsletter.dancohen.org
4. Dan Cohen, "The Library's New Entryway" (Oct 2025)
   newsletter.dancohen.org
5. Tom Scheinfeldt, "Making an AI Frontend for ArchivesSpace" (Nov 2025)
   foundhistory.org
6. ProfessionalWiki, "Let AI access your wiki with MCP" (Fall 2025)
   semantic-mediawiki.org

Distributed memo: MCP Server Architectures for Archives, Historical Databases, and Libraries (v1.5, 17 Feb 2026) *[update link after upload to repository]*