HKADH 2027 · Education University of Hong Kong, 21 January 2027

Scan for these slides

slides.frederickmadore.com

Knowledge gaps and mediated access

AI infrastructure and practice for African research data

Frédérick Madore, Jiayu Yang, Oliver Baumann, Neeraj Thandayan Viswajith, Durgesh Nandini and Mirco Schönfeld

UNIVERSITÄT BAYREUTH

fmadore · frederickmadore.com · 0000-0003-0959-2092 · frederickmadore · frederick.madore@uni-bayreuth.de

A catalogue should define its own terms

Three goals: the Cluster's agenda, taken to the catalogue

1. Say what a term means in the research we catalogue
2. Shift who describes: definitions drawn from that research
3. Keep each definition cited, dated and open to revision

The question

Can language models draft these definitions from AMIRA's own holdings, with every claim traceable to a source? Does that do better than borrowing a definition, or asking a model alone?

"It is our contention that this approach allows us to transcend the limits of thinking in fixed categories and binaries."

Rüdiger Seesemann, the Cluster's Dean, writing in 2020.
University of Bayreuth African Studies Working Papers 24

A catalogue of research data held elsewhere

The Africa Multiple Interactive Research Atlas (AMIRA) is the research-data platform of the Africa Multiple Cluster of Excellence, University of Bayreuth (Germany). The material usually stays with the researchers, archives and libraries that hold it; the metadata is shared.

Research items	3,975	Projects	106
Publications	813	Languages	28
Podcasts	43	Videos	142

AMIRA MCP server, collection overview, 16 November 2026



Why an MCP server, not just an API

An API

- Written for programmers
- Each new question needs someone to write code
- AMIRA has one, from Omeka S; the MCP server reads it

An MCP server

- Describes the same catalogue to a language model
- Named tools in plain words; the model picks one, no new code
- Instructions on connecting: how to cite, what is private
- One open standard: Claude, ChatGPT, our test harness

Who decides what

We decide which tools exist and what they return. The assistant and its provider decide what reaches the answer.

Two tools this pilot builds on

Linking keywords to Wikidata

- CAREL, led by Jiayu Yang: linking without task-specific training
- Rules settle the clear cases; a local open-source model, the ambiguous ones
- 89.8% and 90.9% accuracy on Library of Congress headings

Yang et al., JCDL 2026, doi:10.1145/3805696.3846031; CIKM 2026, doi:10.1145/3799682.3840034

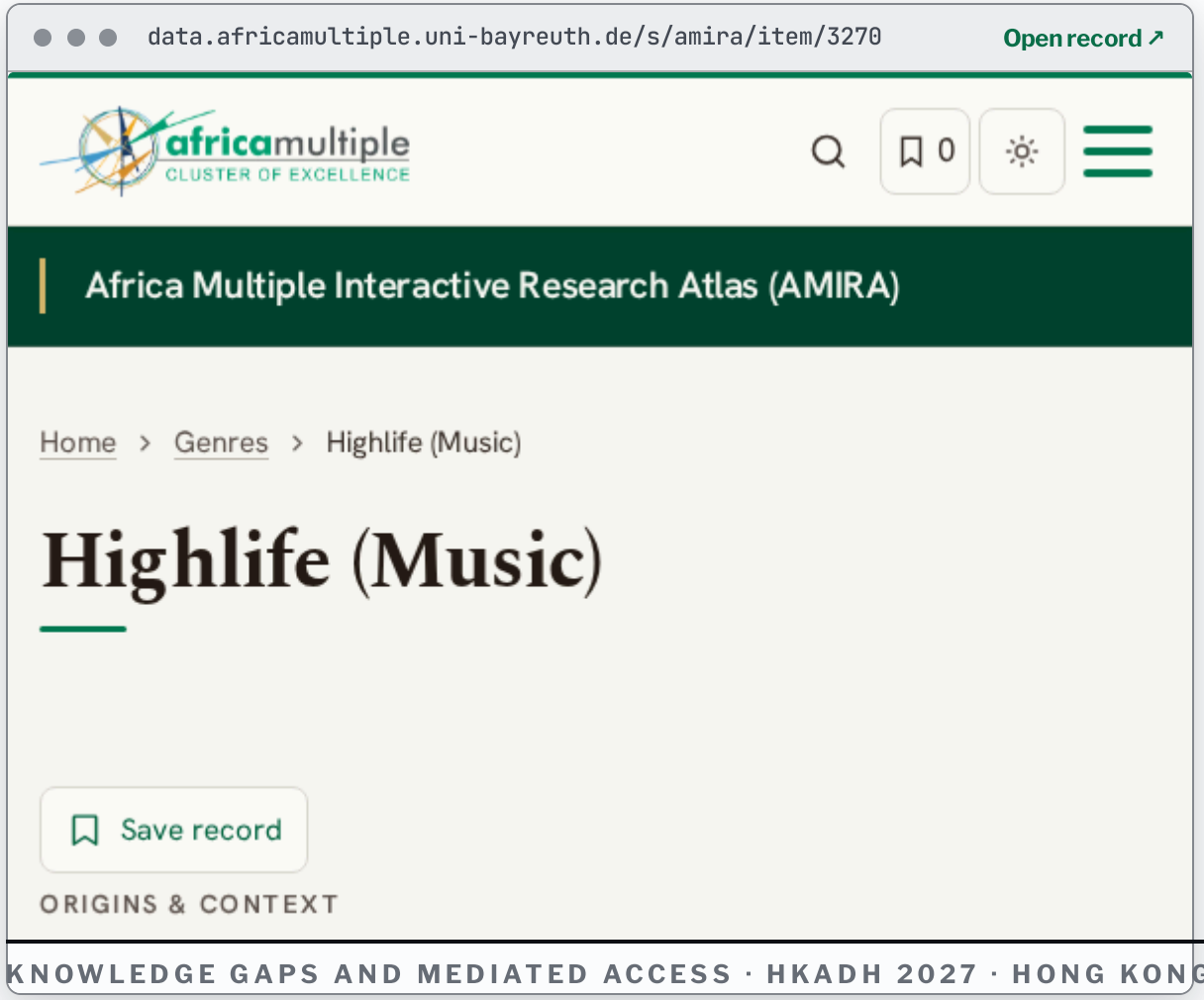
AMIRA MCP server

- The server, not the model, computes counts, facets and related records
- Curator mode: private records and full text, behind a personal token

Code on GitHub

`github.com/AM-Digital-Research-Environment/amira-mcp-server`

A label and a list, but no definition



Records

55

Definition

None

Readers also need to know what a concept means and how the research catalogued in AMIRA uses it.

What a definition would add

What highlife means in the research AMIRA holds, each claim cited to an article.

Standard vocabularies identify a concept but rarely explain it

Library of Congress

Highlife (Music)

No scope note

Mbira music

A cataloguing instruction: “Here are entered compositions not in a specific form or of a specific type for solo mbira…”

Wikidata

highlife

“musical genre originating in Ghana”

timbila, a xylophone

Listed with the alias “thumb pianos”

id.loc.gov and wikidata.org, read 16 and 17 November 2026

The sources: ILAM's journal *African Music*

The International Library of African Music (ILAM), at Rhodes University (South Africa), holds recordings, journals and photographs of music from across Africa.

Years	1954–2023	Issues	45
Articles	1,032	With full text	935

- Definitions quoted from seventy years of research
- Articles found by what they discuss
- Vocabulary dated, article by article



AMIRA, ILAM collection page and full-text map, 16 November 2026. The full text is private.

1 The gap

AMIRA · subject 2412

Highlife (Music)

Records 55

LCSH sh90004527

Definition: none
The reader gets a label and a list.

2 Four ways to fill it

A **Library of Congress heading**
No scope note for this heading.

B **Wikidata**
"musical genre originating in Ghana"

C **A small model, alone**
Writes from memory. Cites nothing.

D **The same model, reading AMIRA**
search → read → quote
10 tool calls · 5 full texts

AMIRA MCP server 935 ILAM full texts

3 A cited definition

Claude Haiku 5.5
95 words · 7 claims

Highlife is a popular dance music of urban West Africa, above all Ghana and Nigeria, that fuses African and Western musical styles...

- Smith 1962 · 3 quotes ✓
- Emielu 2013 · 1 quote ✓
- Otchere 2016 · 2 quotes ✓
- Jackson 2012 · 1 quote ✓

7 of 7 claims carry a quote found in the text the model read.

4 Review

Code
Is each quote in the text the model read?

Three judge models, blind
Which of two definitions goes on the record?

Researchers (planned)
Would you put this on the record?

Highlife, defined from AMIRA's sources

Highlife is a popular dance music of urban West Africa, above all Ghana and Nigeria, that fuses African and Western musical styles. Writers describe it as a compromise between African traditional forms and Western music, mostly sung in many vernaculars with simple, binary-form melodies. It emerged as a national rather than ethnic music in the 1950s.

Claim	Quote found in the text the model read	Source
1	"all represent this compromise between African traditional forms and Western music."	Smith 1962
2	"The highlife, primarily vocal music, is sung in the various vernaculars."	Smith 1962
3	"Highlife emerged in the 1950s as a national, rather than ethnic or regional-based music."	Emielu 2013

7 of 7 claims carry a quote found in the text the model read.

Claude Haiku 5.5, route D, run of 16 November 2026. Sources: articles in African Music (ILAM), linked to their AMIRA records.

Four models reading AMIRA on highlife

Model	Origin given	Quotes found	Est. cost (USD)
Claude Haiku 5.5	Debated: Konkomba singing bands, several streams	7 of 7	0.014
GPT-6 Luna	Ghana, then across West Africa	3 of 3	0.014
Gemini 3.5 Flash-Lite	Local forms with brass band, jazz and ballroom	3 of 3	0.016
Mistral Small 4	Local forms with ballroom, brass band and Latin American music	1 of 5	0.029

Route D, one run per model, 16 November 2026. Costs estimated from list prices. Tool calls and what each definition adds: annexe.

One English name, three drums

Talking drum

57

articles

One Wikidata item. Reading AMIRA, all three usable definitions name members: Yoruba dùndún, Akan atumpan.

Dùndún, Yoruba (Nigeria)

26

articles

No Wikidata item. The Yoruba label “Dundun” sits on the Mande dunun, another drum.

Donno or dondo, Akan (Ghana)

9

articles with “donno”

No Wikidata item. Alone, only one model of four placed it in Ghana.

Does one model keep the three drums apart?

Verdicts “distinct”: model alone 3 of 12; reading full texts 8 of 12; passages only 0 of 12, each naming the dondo merged with a Yoruba drum.

Articles: full-text matches in African Music (ILAM). Wikidata read 17 November 2026. Definitions: four models, one run per cell, 17 November 2026. Open verdicts, no answer key: Claude Opus 5.5, GPT-6 Astra, Gemini 4 Argon, 17 November 2026; uncalibrated; agreed on 11 of 12 sets.

Reading AMIRA won where the models knew least

Blind verdicts, uncalibrated	Model alone	Reading AMIRA
Known well: highlife, timbila	18 of 24	6 of 24
Three drums: dondo, dùndún, talking drum	6 of 33	27 of 33
The Cluster's concepts: multiplicity, relationality, reflexivity	9 of 36	27 of 36
All eight concepts	**33 of 93**	**60 of 93**

The judges know what the models know

They score from their own training, the knowledge the model alone draws on. No researcher has calibrated them.

Claude Opus 5.5, GPT-6 Astra and Gemini 4 Argon chose between each author model's two definitions, shown as text only. Definitions 16 and 17 November, verdicts 17 November 2026; one run per cell. All three judges agreed on 20 of 31 pairs. Talking drum: 9 verdicts, one grounded answer unusable.

Wikidata, extended but never contradicted

Wikidata: multiplicity

"philosophical assertion that there is more than one geo-historical trajectory, developed by Edmund Husserl, Henri Bergson, and Gilles Deleuze"

Reading AMIRA: GPT-6 Luna

"Multiplicity is a way of understanding African and African diasporic life-worlds as heterogeneous, coexisting, and mutually influential rather than as one uniform object or fixed category. …"

57 verdicts by the same judges

The grounded text extends or nuances Wikidata in every verdict; it never contradicts it, and never simply agrees. Multiplicity: another sense in 12 of 12. Talking drum: a category, not one hourglass drum, in 9 of 9.

Wikidata Q3860307, read 17 November 2026. Route D definition, 16 November 2026. Open comparison by the three judges, 17 November 2026; one run per cell.

Passages fit a small model's window, at a price

Grounded runs	Full texts allowed	Passages only
Largest single call, median (tokens)	46,133	13,658
Runs whose largest call fits 32k	8 of 32	16 of 16
Est. cost, same 16 runs (USD)	0.377	0.121
Blind panel's choice, 45 verdicts	38	7

The price

With passages only, all four models merged the Akan dondo with a Yoruba "dondon": a short passage does not carry enough context to tell them apart.

Provider token counts. Full texts 16–17 November; passages (no whole texts, at most 12 tool calls) and panel 17 November 2026. One run per cell; panel uncalibrated; no open-weight model yet.

Three limits of the pilot

The judges know what the models know

They score from their own training, as the model alone writes from it, and no researcher has calibrated them. One marked down a 1962 spelling the model had quoted faithfully.

Keyword search finds a term only as written

Another language or spelling is missed unless the record carries it. Retrieval by meaning is not built yet.

Older sources, older words

6 to 7

uses of “tribal” per 10,000 words, 1957 and 1962

Under 1 from 2007 to 2023, apart from one sparse year. Spellings date too.

“Tribal”: AMIRA MCP server, full-text map of 657 African Music articles, 16 November 2026. Every result in this talk: one run per cell.

What comes next in the pilot

1

About 50 concepts

Chosen by family: lamellophones, xylophones, drums, popular genres. Three runs per cell.

→

2

A local open-weight model

On the passages variant, which fits a 32k window: a 27-billion-parameter model, run locally.

→

3

Review by researchers

Subject specialists assess the drafts and calibrate the judges.

“Would you put this on the record?”

→

4

Retrieval by meaning

Behind the same server, so other languages and spellings are found.

Thank you

Reading the archive pays where it knows more than the models.

Frédérick Madore, Jiayu Yang, Oliver Baumann, Neeraj Thandayan Viswajith, Durgesh Nandini and Mirco Schönfeld

github.com/AM-Digital-Research-Environment

Africa Multiple · Digital Solutions

UNIVERSITÄT BAYREUTH

Scan for these slides

slides.frederickmadore.com

Appendix

Held back for questions.

What happens when someone asks

1

A person asks

In Claude, ChatGPT or another assistant: “What does AMIRA hold on highlife?”

→

2

The assistant picks a tool

From the tools’ descriptions: search, count, read a record.

→

3

The server answers

Records, counts and passages from the catalogue, each linked to its page.

→

4

The assistant writes

In its own words, ideally citing the records it received.

Who decides what

We decide which tools exist and what they return. The assistant and its provider decide what reaches the answer.

Four models reading AMIRA on highlife

Model	Tool calls	Full texts read	Claims with a quote found	Est. cost, list prices (USD)
Claude Haiku 5.5	10	5	7 of 7	0.014
GPT-6 Luna	18	6	3 of 3	0.014
Gemini 3.5 Flash-Lite	6	3	3 of 3	0.016
Mistral Small 4	10	6	1 of 5	0.029

Read with care

One run per model: anecdotes, not measurements.

Mistral Small 4: three quotes over the 25-word limit, one not found in the text it read.

Four definitions of highlife, read side by side

All four: an urban popular music of West Africa, chiefly Ghana and Nigeria, mixing African and Western forms.

Model	Origin	What it adds	Quotes found
Claude Haiku 5.5	Debated: Konkomba singing bands, several streams	Vernaculars; a national music by the 1950s	7 of 7
GPT-6 Luna	Ghana, then across West Africa	Variation by place and period	3 of 3
Gemini 3.5 Flash-Lite	Local forms with brass band, jazz and ballroom	“Intertribal” music; love and social commentary	3 of 3
Mistral Small 4	Local forms with ballroom, brass band and Latin American music	Harmony in thirds or sixths; mid-twentieth century	1 of 5

Route D, one run per model, 16 November 2026. Definitions of 80 to 102 words.