Same name, different country

AI agents, human oversight and entity disambiguation in the Islam West Africa Collection

Scan for these slides

slides.frederickmadore.com

Frédérick Madore

Researcher and Data Curator

Africa Multiple Cluster of Excellence, University of Bayreuth

UNIVERSITÄT BAYREUTH

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

One record, four men

Item 1314

Mamadou Bamba

Type	**Person**
Preferred name	**One**
Entity	**One**

The record as the database saw it, before October 2026

Whom the record describes

Researcher Scholarly reference

Author of a 2002 work on Sufi orders in Côte d'Ivoire

Attached by name matching

Politician Côte d'Ivoire

National Assembly vice-president, 1989; foreign minister, 2003

Lieutenant-colonel Burkina Faso

Announced on television the dissolution of the Transition's bodies, September 2015

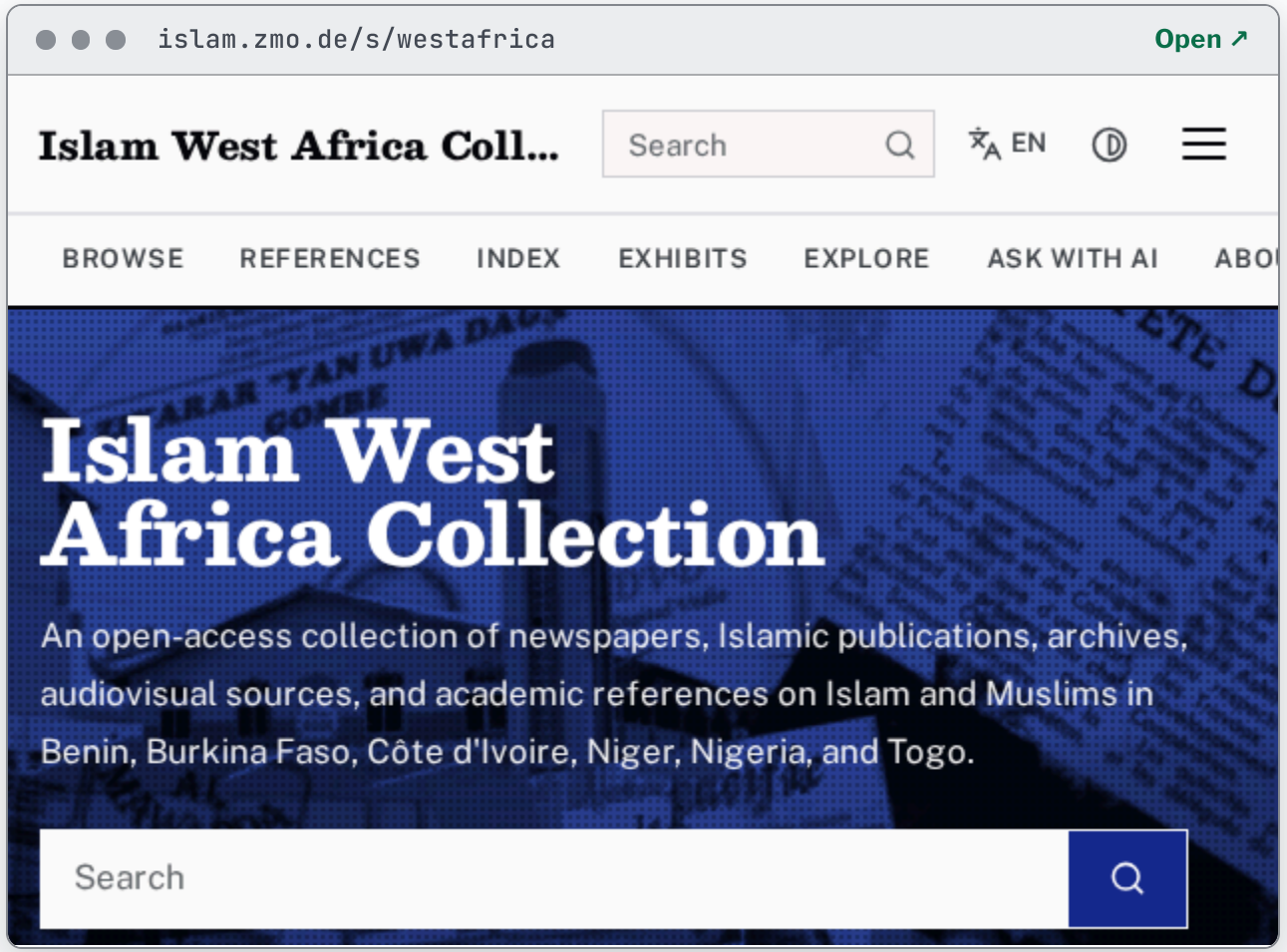
Deputy imam of Yamoussoukro Côte d'Ivoire

Led the end-of-Ramadan prayer, 1986

One authority file, six countries

What a record asserts

The Islam West Africa Collection



Scope

Open-access archive of Muslim public life in Benin, Burkina Faso, Côte d'Ivoire, Niger, Nigeria and Togo

Holdings

Press articles, Islamic periodicals, documents, audiovisual recordings, photographs, scholarly references

Home

Hosted at the Leibniz-Zentrum Moderner Orient (ZMO), Berlin; built and curated by one person

One curator, one authority file

Curator	**1**
Items	**20,000+**
Words	**32 million+**

Countries	**6**
Authority records	**6,000+**
Record types: persons, organisations, places, events, subjects	**5**

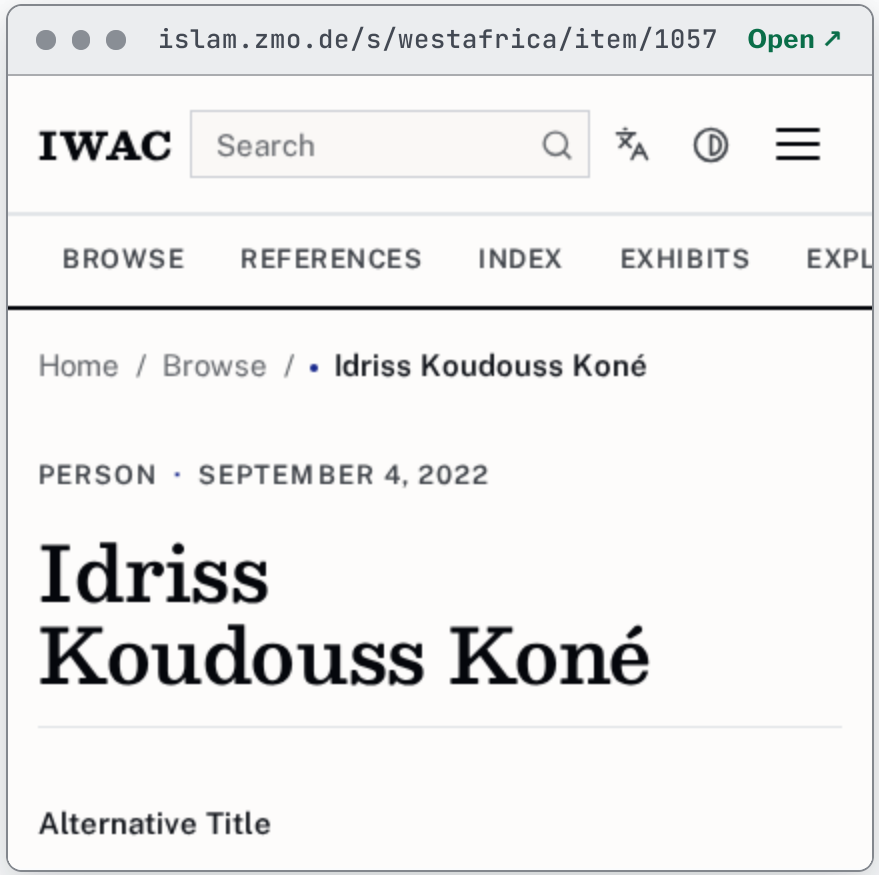
Press articles, Islamic publications and scholarly references all link to the **same** set of records.

The same name in several countries

A script that matches by name turns each row into one entity.

Name	Burkina Faso	Côte d'Ivoire	Benin	Scholarly author
Mamadou Bamba	Lieutenant-colonel, 2015 coup 188934	Politician; deputy imam 188930 188954		Researcher, 2002 1314
Kassoum Coulibaly	Defence minister, 2023–24 127886	Cited 1982–83 67772		
Mamadou Traoré	Army colonel, 2007–09 127891	Ivorian namesake 23631		
Yacouba Traoré	Imam, Muslim community of Poni 127901	Ivorian namesake 64666		
Boukary Savadogo	Ambassador to Saudi Arabia 127895			Author 1295
CNOPM	Hajj commission, 1979–2018 579	Hajj committee, 1993–2007 23540	Hajj committee, 1972–83 127882	

One imam, twenty names



Nineteen alternative names on record 1057

Word order	Koné Idriss Koudouss · Idriss Koné Koudouss · Koudouss Idriss Koné · Koudouss Koné Idriss
Honorific	El Hadj Idriss Koudouss Koné · Imam Koudouss
One s	Idriss Koudous Koné · Idriss Koné Koudous · Idriss Koudous · Koné Idriss Koudous
Given name	Koudouss Drissa Koné · Koné Idrissa Koudouss · Koudouss Idrissa Koné
Shortened	Idriss Koudouss · Koudouss · Koudouss Koné · Idriss Koné · Koudouss Idriss · Koné Koudouss

The reverse also happens: one string for several things, as with *foi* and Danyi (case 3).

Mention, link and record

In the source

Mention

A string found in a source, at a date, in a context.

`“Imam Koudouss”`

→

On the article

Link

Asserts that this mention is that entity.

`article → item 1057`

→

In the authority file

Authority record

Asserts that one entity exists, under one preferred name, with a type and a country.

```
Idriss Koudouss Koné ·
person · Côte d'Ivoire
```

What IWAC stores

The link and the record. The mention is not kept as an object of its own.

The workflow

Which steps are model calls, which are code, which are agents

Five steps, and who acts at each

Press of Burkina Faso, 8,045 articles, 20–22 October 2026

1



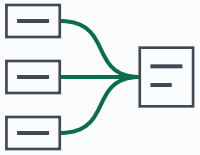
Extract

Model

one call per article

Names of persons, organisations and places; one to six subjects.

2



Match

Code

no model

Exact keys first. Ambiguous or fuzzy candidates are flagged, never linked.

3



Check in context

Model

one call per article

Keep, choose or drop, only among the ids it was offered.

4



Write

Scripts

no model

Dry run first. Each link annotated with its model: `iwac:nerModel`.

5



Leftovers

Agents

then the curator

Reviewer agents propose new records. The curator approves each plan.

How the country is decided

Rules in the prompt

- “Use the text first, the newspaper’s country second.”
- “A body of another country is not this one.”
- A contradiction is only “a different country, an incompatible calling … or dates that cannot fit”.
- “Choose only among the candidates of that entry. Never invent an id.”

What the model is given for one name

Name as found	**Kassoum Coulibaly**
Newspaper’s country	**Burkina Faso**

67772	**Coulibaly Kassoum** · person · Côte d'Ivoire *No description*
127886	**Kassoum Coulibaly** · person · Burkina Faso *Minister of state, minister of defence*

Keep | **Choose** | **Drop: leave unlinked**

Illustration of the input’s shape, not the transcript of a real call

The first full run, Burkina Faso

Articles	**8,045**
Matches dropped, after the re-check	**1,776**
Links written: 10,036 place, 40,763 subject	**50,799**
Matches kept in context	**88,250**
Fuzzy or ambiguous names linked by the check	**522**
New authority records from the leftovers	**674**

Claude Haiku 5.5. Extraction 20 October (1 h 38, about $6.50); leftovers 21 October; context check and write 22 October. From the pipeline changelog and run files.

The other countries

Togo: 1,570 articles extracted and matched on 23 October; no context check, nothing written. Côte d'Ivoire, Benin, Niger and Nigeria: not run.

Three cases

Where the workflow went wrong or could not decide

Case 1: one acronym, three countries' committees

Record 654 · Burkina Faso

CNOPM

32

articles, most from Ivorian and Beninese papers

12 To the Ivorian committee · 23540

7 To a new record for Benin's state committee · 127882
Daho-Express and Ehuzu, 1972–1983

2 To a Burkinabè record · 665

1 To the AMOP · 46321

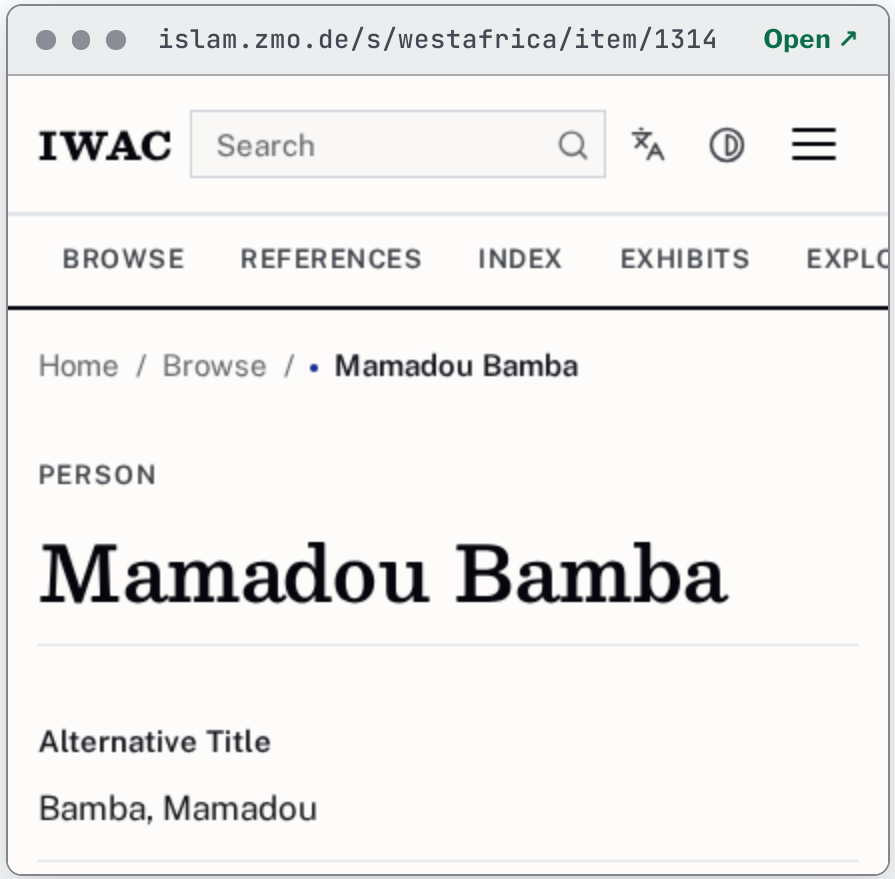
10 Left on 654 for the curator
settled one by one on 23 October

The rule that followed

Every generic national body carries its country in its title: "Ministère de la Santé (Burkina Faso)".



Case 2: the researcher and his namesakes



1314	Scholarly reference	Researcher, 2002 work on Sufi communities in Côte d'Ivoire	Kept
188930	Côte d'Ivoire	Politician, Assembly vice-president 1989, foreign minister 2003	New
188934	Burkina Faso	Lieutenant-colonel, spokesman of the failed coup of September 2015	New
188954	Côte d'Ivoire	Deputy imam of Yamoussoukro, end-of-Ramadan prayer 1986	New
No record	*"El Hadji Mamadou Bamba", 1979: cannot be placed*		Unlinked

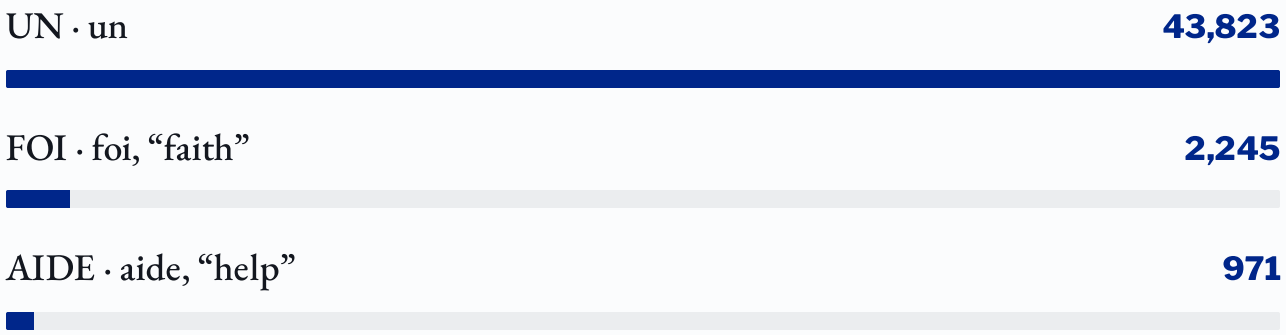
"Unknown" has to be representable

Authors and actors share one person set

Case 3: acronyms that are also words

The authority records hold 408 acronyms. Twelve are also common words in the articles.

Times written in lower case in the articles



Links on the FOI record (23699) that were the word *foi* — **317 of 410**

Matches of “Foi” and “Aide” in the Burkinabè run, all dropped — **199 + 27**

The rule now: capitals only

`FOI`	✓ key
`F.O.I`	✓ same key
`Foi`	× no key
`foi`	× no key

Lesson

An alternative label is also a matching key.

Danyi and Dankpen, read as Denmark

93 of 124

place links on Danemark (428) were Togolese agency reports, 2020–2024, naming the prefectures of Danyi and Dankpen

Wrong links, removed **75%**



Who made the error?

A pass older than the `iwac:nerModel` annotation. Nothing records which model or script made these links.

Lesson

Provenance belongs on each link.

What stays human

One curator, some 50,000 links from one country run

Where the curator sits

1	**Rules before runs**	Written by the curator, rewritten after reading what went wrong
2	**Ambiguity is never settled by code**	The model chooses among the records offered, or leaves the name unlinked
3	**Every new record is approved**	731 proposed in the first pass: 18 held back as possibly existing, 39 flagged
4	**Merges and splits are the curator's**	59 merge pairs drawn from the reviewers' remarks, withdrawn unapplied
5	**Agents never write**	Scripts do: dry run, backup of each item, dump of every payload

Who validates?

The same model rereads

Not independent

Extractor and checker are both Claude Haiku 5.5.

A deterministic test

Independent, but narrow

A full-text count overturned wrong drops: 4 on the FOI, 10 on other acronyms.

A human sample

None yet

No hand-checked error rate for any country.

What escapes the loop is what no rule anticipated.

Four modelling decisions the cases forced

Country in the title

For generic bodies, identity includes jurisdiction.

Unlinked as an outcome

A mention may stay a mention when the source cannot settle it.

Provenance on each link

`iwac:nerModel`, `iwac:descriptionModel`: which model proposed this value.

Undated descriptions

Written today about articles decades old. Patched by a prompt rule; IWAC lacks a temporal marker.

Questions for the Alevi Archive

1. Where does the identity of a person or a *cemaat* live, and who decides that two factoids are about the same entity?

2. Can a probabilistic identification stay marked as such once it is in the graph?

3. Is a check by the same model a verification at all? What would count as an independent one?

4. When the source cannot settle an identity: an unlinked mention, a provisional record, or a “possibly the same as” relation?

Thank you

The curator approves records, merges and splits. The links still await a hand check.

Scan for these slides

slides.frederickmadore.com

Frédérick Madore

Researcher and Data Curator

Africa Multiple Cluster of Excellence, University of Bayreuth

UNIVERSITÄT BAYREUTH

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

Appendix

Held back for questions

How step 2 matches names

Code, no model. Looser keys are tried only once every exact key has missed.

- Exact keys: lower case, no accents, spaces or dashes
- A key shared by two records is ambiguous, never resolved
- Word order: “Coulibaly Kassoum” = “Kassoum Coulibaly”; never for initials
- Plural fold for non-person subjects; accents kept, “Jeunes” ≠ “Jeûne”
- Acronyms of 2–6 capitals get a case-sensitive key only
- “(Country)” titles are also keyed bare within that country’s run

Twelve acronyms that are also words

UN 43,823 in lower case	FOI 2,245	AIDE 971	AU
CAR	CET	AIT	UNIR
UL	SEMI	AD	UP

The other ten acronym records had 722 item links; 579 name the record in full or in capitals. The model read the other 143, and a full-text check overturned ten of its drops (seven “Nations-Unies”, two Tidjania, one bishops’ conference). 108 links were removed, 99 of them “aide”.

When reviewers propose merges

Withdrawn, 59 pairs

- The Yirgou clashes into the village of Yirgou
- The arrest of Inoussa Kanazoé into the man
- The FAIB into the AIMB
- Nas-Air into Air Afrique

Event, place, person and organisation are types a language model blurs.

Merged, 21 October 2026

- BID: 14587 → 636
- CCI-BF: 126600 → 126067
- FAMa: 126779 → 126000
- Kaaba: 15008 → 76313

Further questions

1. Should provenance sit on every link, and what does that cost in a factoid model where links are the data?

2. How do you tell apart two people, or two *ocak*s, of one name in different regions? Is that decision a factoid, with a source and an author?

3. Does every factoid pass a person, and would it still if the data grew tenfold?

What the first extraction prompt picked up

Generic themes extracted as subjects	**13,582**	of them “Islam”	**4,983**
of them “Religion”	**1,115**	Mentions of the article’s own source outlet	**1,700**

The revised prompt, tested on 18 articles: subject entries fell from 185 to 127, and “Islam” from 12 articles to none. The Burkinabè run used the first prompt; Togo, the revised one.