Digital Humanities and AI in the Social Sciences

Datafication in practice

Religion and Intellectual Culture Unit
12 November 2026

ZMO LEIBNIZ ZENTRUM MODERNER ORIENT

Leibniz Association

Overview

1. **Framing:** two ZMO workshops and a position paper
2. **Global Heshmat:** a sculptor's works on a map
3. **Datafication:** where the work lies, and Kalamu, an app to speed it up
4. **Albani Zaria's sermons:** from DVD to word counts, places and people
5. **The Global South and data:** who digitises what, and who has access
6. **The dialectics of datafication:** what helps, what damages, what endangers
7. **Questions for the discussion:** method, ethics, epistemology
8. **Two more projects:** ZMO Visualisations and Genocide at the Security Council

Framing

Where this comes from, and what datafication means

Two workshops

Both applications prepared at ZMO

Charting New Territory

Scoping workshop · VolkswagenStiftung
Hanover, 18–20 February 2026

- Convened with Vincent Hiribarren (King's College London)
- 26 researchers from 16 countries
- Working groups and World Café rounds instead of keynotes and panels

fmadore.github.io/dh-ai-african-studies-2026

Towards Sustainable and Equitable Practices

DFG Programme Point Sud · STIAS
Stellenbosch, 21–24 September 2026

- Convened with Vincent Hiribarren, Emmanuel Ngue Um and Menno van Zaanen
- 25 papers, 33 participants from 17 countries
- Hybrid, in English and French

fmadore.github.io/stias-dh-ai-workshop-2026

What came out of it

Position paper: “For Whom and For What Purpose?” (*ZMO Programmatic Texts*, 2026)

- **Situated practice:** digitisation agreements as living documents, ontologies open to revision
- **Sovereignty:** design grounded in access and governance; ethics revisited throughout a project
- **Shared problems:** consent, standardisation, ill-fitting metadata, hidden labour, language bias, maintenance, all more visible in African contexts
- **Reading as data:** 217 works annotated, mapped into 145 concepts and 1,219 links

Next

A special issue in preparation

Journal of the Digital Humanities Association of Southern Africa (JDHASA)

Articles drawn from the Stellenbosch papers

Datafication

Rendering social life, objects and records into machine-readable data

van Dijck 2014	Datafication as a paradigm shift in knowledge production, and the belief in data it fosters
Mejias & Couldry 2019	The quantification of social life, whose appropriation they call data colonialism
Drucker 2011	Data are capta: taken, not given
Gitelman (ed.) 2013	“Raw data” is an oxymoron

Question for today: what is decided, and by whom, when a source becomes data?

One object, two forms

A memorial in Hasselt, Belgium, and its record in the Global Heshmat code



Hasselt memorial, 1970 (historical photo, Global Heshmat)

```
// artworks/018-Heilig Kruiskerk.ts
const artwork: Artwork = {
  id: 18,
  slug: "heilig-kruiskerk",
  name: 'Heilig Kruiskerk',
  lat: 50.92058026384286,
  lng: 5.325863864013154,
  country: 'Belgium',
  city: 'Hasselt',
  status: 'located',
  trails: ['east-west', 'faith-remembrance'],
  address: 'Kruisherenlaan 29, 3500 Hasselt',
  desc: 'In memory of 23 Belgian missionaries murdered
    in Buta (Congo) in 1965, Hassan Heshmat was …',
  images: [{src: "Mahnmal_Hasselt_Kopie.jpg",
    caption: "Mural next to the Heilig Kruiskerk"}, … 7 more],
  links: [{label: 'KADOC',
    url: 'https://kadocheritage.be/items/show/11473'}]
};
```

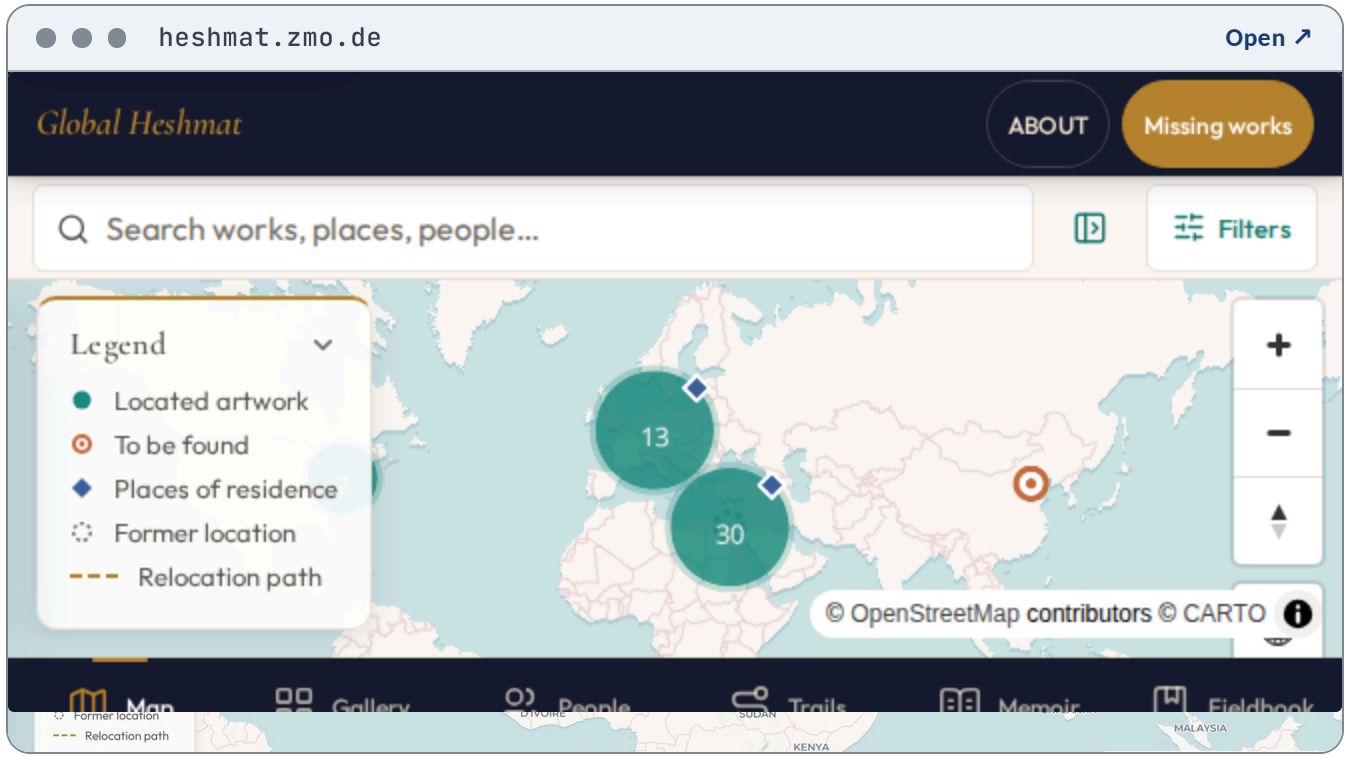
Global Heshmat

Following Hassan Heshmat's works around the world

LEIBNIZ ZENTRUM MODERNER ORIENT

Global Heshmat

Egyptian sculptor Hassan Heshmat (1920–2006) · a project by Sonja Hegasy



29
works located

13
still to be found

160
photographs

31
people profiled

46 entries, each with its own page: story, photographs, sources, a citable address

Missing works keep their entries, with a public call to report them.

LEIBNIZ ZENTRUM MODERNER ORIENT

www.leibniz-zmo.de

What the mapping showed, and what it took

31 profiles from Sonja Hegasy's research, and the decisions behind the site

What the mapping showed

An unexpected cluster: Heshmat's ties to the Armenian community in Cairo.



Back to People

Shant (Chant) Avetisyan

Heshmat and the Armenian-Egyptian (art) community in Cairo Contemporaries and Peers

Shant (Chant) Avetisyan, Egyptian-Armenian artist (1951 – 2018). From 1970 to 1978, he studied in Montreal and Paris. He then returned to Cairo. In the 1970s, for example, he designed costumes for the Armenian folk dance group Sardarapat. Avetisyan bought artworks from Heshmat. “In Egypt I am Armenian, in Europe I am Egyptian, but in China these definitions mean nothing.”

One file per work

Each work and each person is a typed data file: adding one is a one-file edit

Statuses

Located, to be found, relocated: a work's history becomes a field and a line on the map

Names and scripts

Search ignores Arabic vowel marks and accents, but each name still takes one chosen spelling

After the project

Who maintains the record, and who in Egypt can correct it?

Datafication is where the work is

Most of the time goes into turning sources into data; analysis comes after

Datafication

find · scan · transcribe · correct · describe · link

Analysis

visualise · count · compare

AI speeds up the first part by an order of magnitude

Text extraction

Manuscripts, typescripts and print, in several scripts

Transcription

Interviews, lectures and sermons, with speakers told apart

Entities

People, places and organisations, ready to link

Its own problems: errors that read well, bias, cost, and extraction (whose data trains the models, and who profits?)

Kalamu

A desktop app I am developing: the same work, without writing code



- Read scanned and handwritten documents
- Transcribe interviews, in the cloud or offline
- Analyse texts with fields you define: people, places, themes
- Check every result against its source; corrections never overwrite the AI's text
- Export with a record of model, prompt and settings

github.com/fmadore/Kalamu
Free, open source
Windows and macOS

LEIBNIZ ZENTRUM MODERNER ORIENT

www.leibniz-zmo.de

Albani Zaria's sermons

From DVD to word counts: Hausa sermons in the *Islam West Africa Collection* (IWAC)

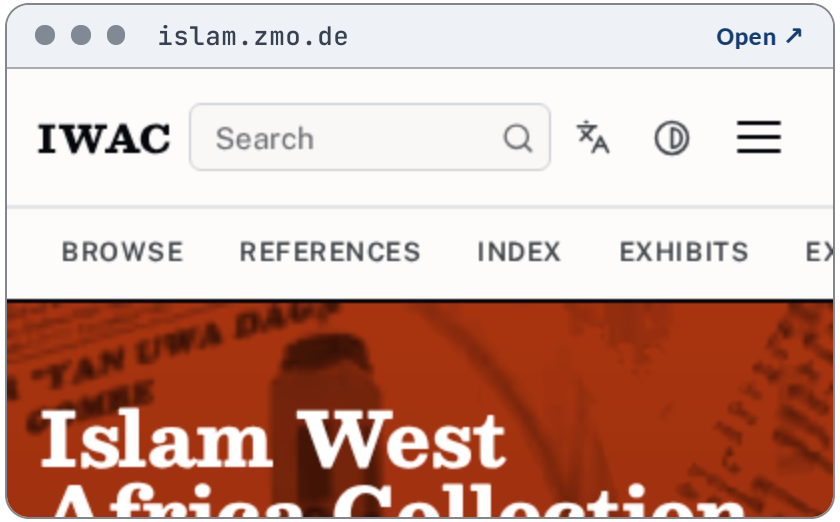
ZMO
LEIBNIZ ZENTRUM MODERNER ORIENT

Who was Albani Zaria?

Muhammad Auwal Adam Albani Zaria (1960–2014)

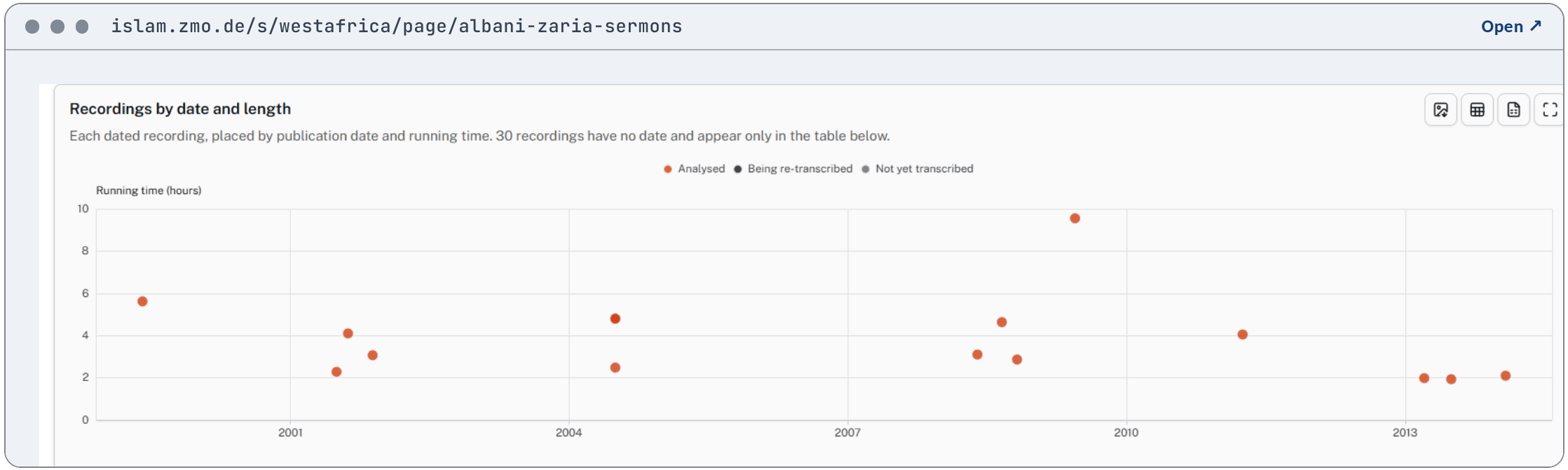
Born	Zaria, northern Nigeria, 27 September 1960
Educated	Barewa College, Zaria
Role	Salafi scholar and Qur'anic exegete; taught at Daarul Hadeethis Salafiyyah, Zaria
Languages	Taught in Hausa, quoting Arabic
Died	Assassinated in Zaria, 1 February 2014

The collection includes the lessons he gave on the day of his death.



The collection

45 recordings, c. 1999–2014 · 42 transcribed so far: 145 h, 1,089,269 words



Running time (hours) of the 15 dated recordings; the other 30 carry no date. Catalogued for IWAC by Vincent Favier and Aleksei Akseshin; 3 recordings not yet transcribed.

The pipeline

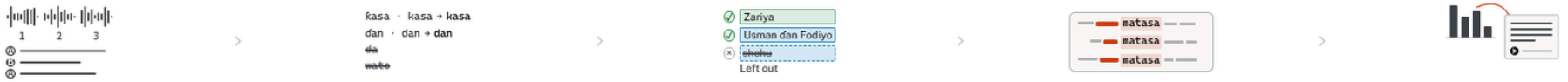
Five stages turn the recordings into figures and passages

islam.zmo.de · Albani Zaria's sermons · how the figures are made — Open ↗

How the Albani Zaria sermons become figures

Replay animation

The Albani Zaria exhibition counts words, terms and names across the transcribed sermons and lessons of Shaykh Muhammad Auwal Albani Zaria. Every figure leads back to a passage, and every passage to its recording.



1 · The recordings

Speech recognition writes each recording out word for word, in 20-minute segments, and estimates who is speaking. Recordings not yet transcribed, and passages in English, are left out.

Both can be wrong, so every passage links back to the recording.

2 · Vocabulary

Spellings with and without the hooked letters (ɓ, ɗ, ƙ, ƴ) count as one word; the commonest words and the speaker's fillers, such as wato, are left out.

The stop words are those published by Bashir, Bichi and Adamu (2024); religious formulae are counted as whole phrases, and a simple rule tells Arabic passages from Hausa.

3 · People and places

A named-entity model proposed names once; only a list of names and spellings curated by hand is counted.

A spelling with another meaning, such as shehu, is left out, so the counts are floors.

4 · Youth and politics

Terms for youth and for politics are found in the text, with the words that occur unusually often near them and the places where the two themes meet.

The word lists are a draft that a Hausa-speaking specialist will check.

5 · Figures and passages

The counts become the page's charts, its map and the words that set each recording apart; passages are quoted only from transcripts that are public.

A count shows where to look; the passage and the recording are the evidence.

The figures describe these recordings only: one preacher, the part of his teaching transcribed so far, and the errors the transcription contains. The word lists and translations are drafts awaiting a specialist's review.

Illustrative example: the drawings show the method with words the exhibition's method section cites; the passages are lines, not quotations.

Every step is a decision

A word list is a classification scheme

Spelling

ɓ/b, ɗ/d, ƙ/k, y/y, q/k, double letters and apostrophes count as one word

Common words

76 published Hausa stop words (Bashir, Bichi & Adamu 2024), plus this speaker's fillers *ainihin*, *wato*

Formulae

sallallahu alaihi wa sallam, *in sha Allah*: counted as whole phrases, outside the word lists

Loanwords

addini, *shari'a*, *ilimi*: Arabic words that count as Hausa

Meaning

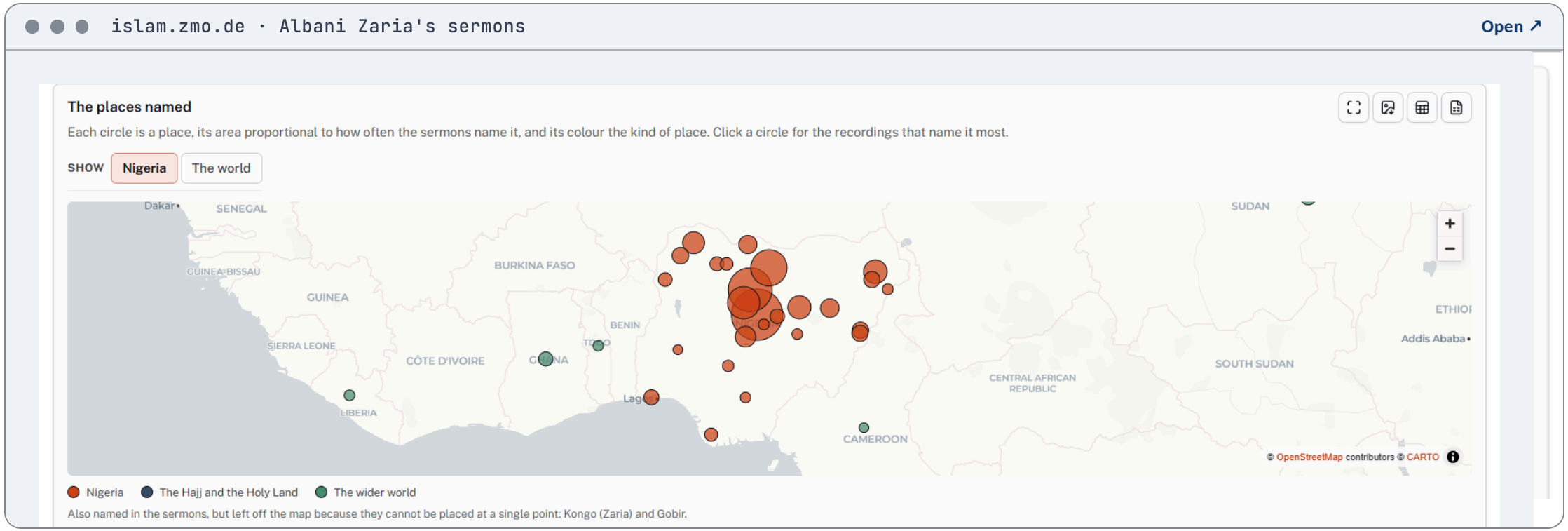
sarki is not counted where it means God (1,184 of 2,094); *kasa*, “country” or “to fail”, stays out of totals

Speakers

Voices are told apart only within each 20-minute segment; the 35 dialogue segments can be excluded

The places named

Each circle a place, its area proportional to how often the sermons name it

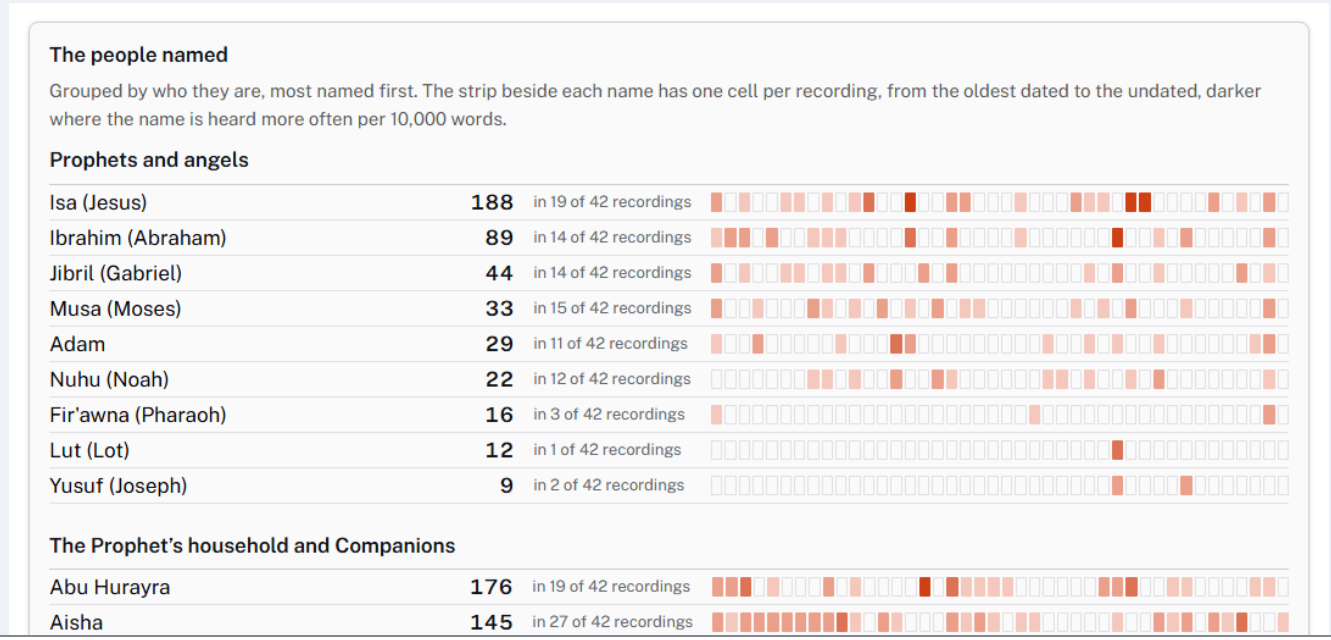


LEIBNIZ ZENTRUM MODERNER ORIENT

The people named

Seven groups of people, the most named first in each

islam.zmo.de · Albani Zaria's sermons — Open ↗



The people named

Grouped by who they are, most named first. The strip beside each name has one cell per recording, from the oldest dated to the undated, darker where the name is heard more often per 10,000 words.

Prophets and angels

Isa (Jesus)	188	in 19 of 42 recordings
Ibrahim (Abraham)	89	in 14 of 42 recordings
Jibril (Gabriel)	44	in 14 of 42 recordings
Musa (Moses)	33	in 15 of 42 recordings
Adam	29	in 11 of 42 recordings
Nuhu (Noah)	22	in 12 of 42 recordings
Fir'awna (Pharaoh)	16	in 3 of 42 recordings
Lut (Lot)	12	in 1 of 42 recordings
Yusuf (Joseph)	9	in 2 of 42 recordings

The Prophet's household and Companions

Abu Hurayra	176	in 19 of 42 recordings
Aisha	145	in 27 of 42 recordings

Most named, all groups

1. Al-Bukhari **502**
2. Usman dan Fodio **397**
3. Ja'afar Mahmud Adam **279**
4. Abubakar Gumi **252**
5. Isa (Jesus) **188**
6. Abu Hurayra **176**

Al-Bukhari in 36 of 42 recordings; Usman dan Fodio in 25.

Youth: who is being talked about?

Occurrences of the youth word list across the 42 analysed transcripts

islam.zmo.de · Albani Zaria's sermons — Open ↗

The word list and its counts

How often each term occurs in the analysed transcripts, and in how many segments and recordings. Words with more than one meaning are counted and quoted but kept out of the combined figures until a specialist has checked them. The glosses are a guide for the reader; the analysis works on the Hausa.

TERM	GLOSS	CATEGORY	OCCURRENCES	SEGMENTS	RECORDINGS	FORMS FOUND
yara	children; a boy, a girl	Young people	999	264	41	yara, yaro, yaran, yaron, yarinya, yarinyar
ɗalibai	students; a student	Young people	595	170	37	daliban, dalibai, dalibi, dalibin, daliba, dalibar
matasa	young people; a young man, a young woman	Young people	449	144	33	matasa, matasan, matashi, matashiya, matashin, matashiyar
'ya'ya	offspring (as in tarbiyyar 'ya'ya)	Young people	287	98	36	'ya'yan, 'ya'ya, ya'yan, ya'ya
almajirai	Qur'anic pupils	Young people	164	80	30	almajiran, almajirai, almajirin, almajiri
'yan mata	girls, young women	Young people	146	63	27	yan mata, 'yan mata, yan matan, 'yan matan
samari	young men; a young man	Young people	129	76	31	samari, saurayi, samarin, saurayin, samartaka

Yara (children) is the largest youth term: much of the theme concerns children and families.

Makaranta (school, education theme) occurs 996 times, in 40 of 42 recordings.

Almajirai, Qur'anic pupils: 164 times in 30 recordings.

Politics: the vocabulary of government

Occurrences of the political word list across the 42 analysed transcripts

islam.zmo.de · Albani Zaria's sermons — Open ↗

The word list and its counts

How often each term occurs in the analysed transcripts, and in how many segments and recordings. Words with more than one meaning are counted and quoted but kept out of the combined figures until a specialist has checked them. The glosses are a guide for the reader; the analysis works on the Hausa.

TERM	GLOSS	CATEGORY	OCCURRENCES	SEGMENTS	RECORDINGS	FORMS FOUND
demokuraɗiyya	democracy	Democracy	263	45	18	demokradiyya, dimokuradiyya, demokradiya, demokradiyyar, demokradiyar, dimokuradiyyar, demokra'iyya, demokradiyance
siyasa	politics	Politics and politicians	439	133	35	siyasa, siyasar, siyasan, siyasance
'yan siyasa	politicians	Politics and politicians	108	71	24	'yan siyasa, yan siyasa, yan siyasar, 'yan siyasar
gwamnati	government	Government	690	198	38	gwamnati, gwamnatin, gwamnatocin, gwamnatoci
gwamna	governor	Government	262	105	32	gwamna, gwamnan, gwamnoni, gwamnonin
hukuma	authority, public body	Government	260	91	30	hukumar, hukuma, hukumomi, hukumomin
	leader, head:					

sarki (emir, king): 2,094 occurrences, but 1,184 left out because they refer to God, as in *Allah Maɗaukakin Sarki*

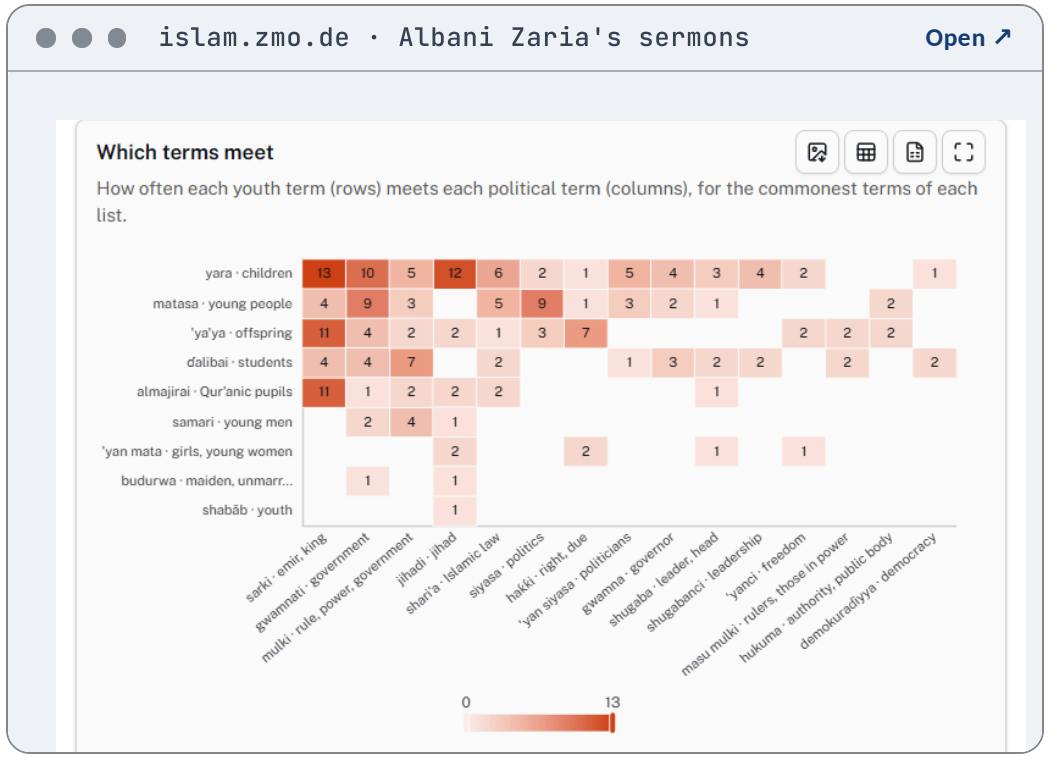
Three in four uses of *demokuraɗiyya* (201 of 263) come from a single lecture, **Demokuradiyyah** (9 h 33).

LEIBNIZ ZENTRUM MODERNER ORIENT

www.leibniz-zmo.de

Where youth meets politics

A meeting: a youth term and a political term within ten words, same speaker



Most frequent pairs: yara × sarki (13), yara × jihadi (12), 'ya'ya × sarki (11), almajirai × sarki (11)

89 segments contain at least one meeting; the exhibit ranks them and links each to its 20 minutes of audio.

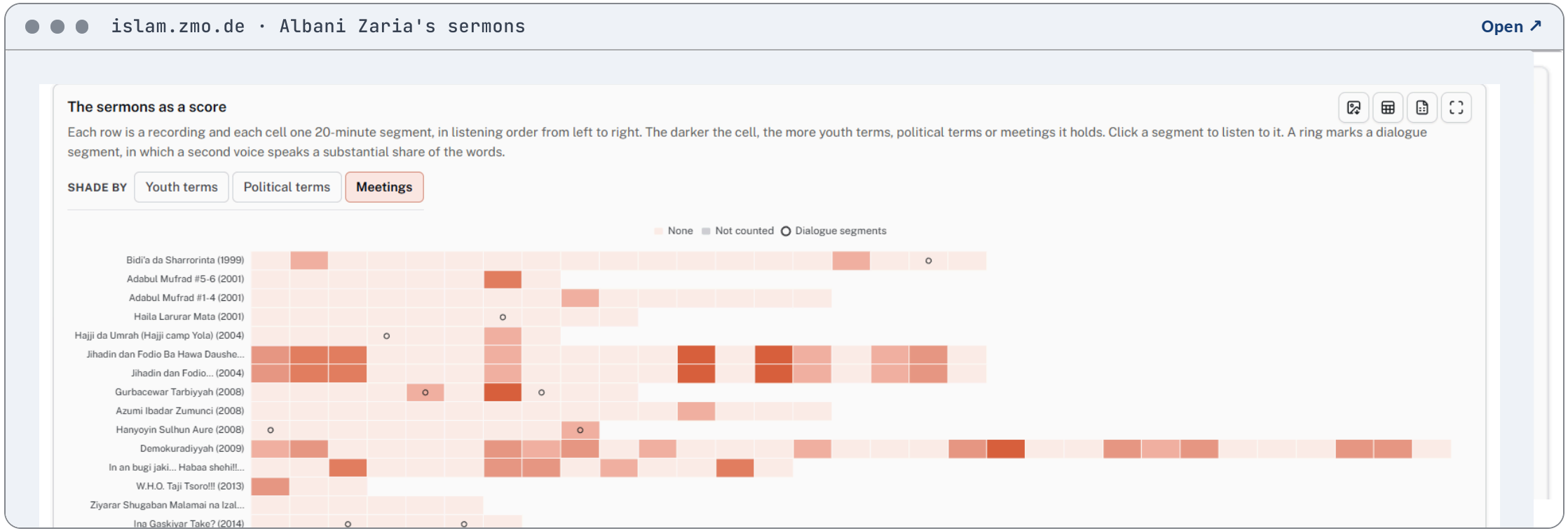
The count points to the 20 minutes worth hearing; the analysis comes from listening to them.

LEIBNIZ ZENTRUM MODERNER ORIENT

www.leibniz-zmo.de

The sermons as a score

Each row a recording, each cell a 20-minute segment: click one to listen to it

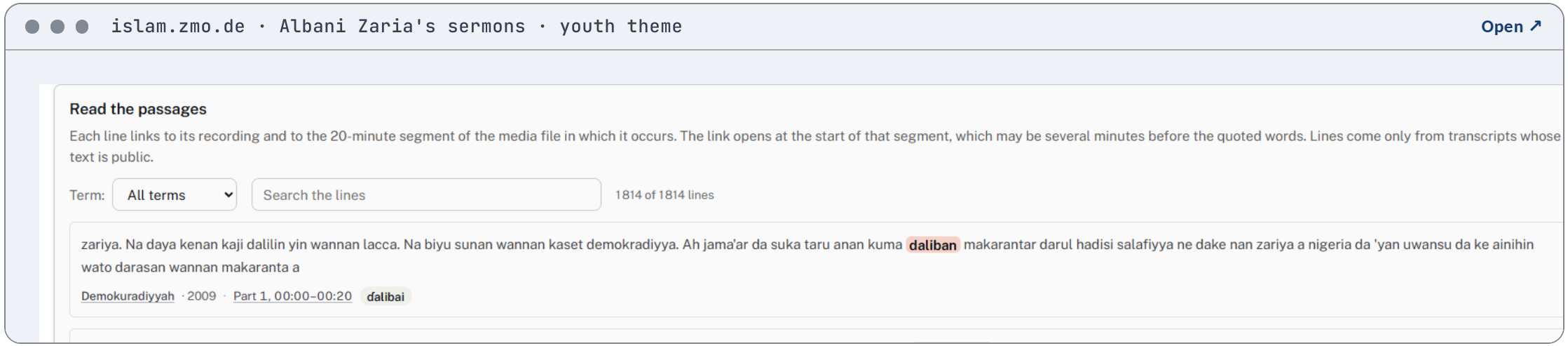


LEIBNIZ ZENTRUM MODERNER ORIENT

www.leibniz-zmo.de

Read the passages

On the exhibit page, every count leads back to the words it comes from



1 Pick a term

or search the lines: 1,814 for the youth list

2 Read it in context

the term highlighted in its sentence

3 Go to the recording

each line links to its 20-minute segment

4 Public text only

drawn from transcripts already public on IWAC

Discussion

Methodological, ethical and epistemological stakes of AI and digital humanities tools

The Global South and data

Digitisation and AI can reproduce older asymmetries

Who digitises

Mostly Western-funded: “digital imperialism” (Breckenridge), a “digital saviour complex” (Shringarpure)

What gets digitised

Selection deepens archival silences; keyword search sidelines what is not online

Who has access

The digital divide: unequal internet access remains a major obstacle (Barringer & Wallace 2014)

Data colonialism

Life turned into raw material for profit, as land once was (Couldry & Mejias 2019, 2024)

Data extractivism

African texts, voices and annotation labour feed models whose value flows north

Data flaring

The reverse: African-language data never collected or kept (Adebara 2025)

One answer, in IWAC: FAIR and CARE principles; catalogue where custodians prefer; credit rights holders.

The dialectics of datafication

Which datafication helps, which damages, and which endangers?

The counter-position

Some projects try to make interlocutors impossible to datafy, so that they do not end up in prison: dispersing data points to prevent mapping.

Datafication varies by region (authoritarian or not), by discipline and by goal.

- Historians, already surrounded by texts, tend to see DH as a treasure
- Anthropologists are often hostile: everyday experience is not data
- Fieldwork is already extraction; AI changes speed and scale
- What would consensual, qualitative practice look like?

Questions for the discussion

1 Method

Who checks the word lists, and in which language?

What does a 20-minute segment cut through?

Can error rates be measured for Hausa–Arabic speech?

2 Ethics

Whose voice becomes data, and who decides?

Should such recordings go through commercial AI services?

When should material stay out of the data altogether?

3 Epistemology

Is a count evidence, or only a pointer?

What does building one's own categories teach that automatic labelling skips?

Fieldwork, archives, AI: is it all extraction?

Two more projects to explore

ZMO Visualisations

ZMO as data: 1,985 publications by 671 authors, 1994–2026



zmo-berlin.github.io/visualisations

Genocide at the Security Council

How is the word used in 167,600 UN Security Council speeches, 1946–2024?



fmadore.github.io/genocide-at-the-security-council

LEIBNIZ ZENTRUM MODERNER ORIENT

www.leibniz-zmo.de