*Extended abstract: Monitoring cross-border refugee flows from Sudan using digital trace data*

*Author: Jackson Mason-Mackay, PhD student*
*Affiliations: Max Planck Institute for Demographic Research (MPIDR), University of Manchester*

## OVERVIEW

Forced migration is one of the fastest growing forms of human mobility. The number of forcibly displaced people worldwide has tripled in the last decade (UNHCR, 2025) and now accounts for one in every six international migrants (United Nations, 2025). Despite this scale, forced migration remains understudied in demographic research.

The application of digital trace data to study displacement has primarily focused on upper-middle and high-income receiving countries where digital access is strong – such as Syrians in Turkey (Sanliturk & Billari, 2024), Venezuelans in Colombia (Palotti et al., 2020), Puerto Ricans in the United States (Alexander, Polimis, & Zagheni, 2019), and Ukrainians in Poland (Vieira, Şanlıtürk and Zagheni, 2026). The use of digital trace data to measure conflict-induced displacement in the global south remains rare, despite these contexts facing significant information gaps.

Sudan's ongoing armed conflict, which began in April 2023, represents the world's largest displacement crisis (UNHCR, 2026). Over 4 million refugees have crossed into neighbouring countries, while over 8 million people remain internally displaced (UNHCR, 2026; IOM 2026). Sudan's displacement patterns align with global trends, whereby most displacement is contained regionally and does not involve large distances. Globally, 73 percent of refugees live in low- and middle-income countries and 63 percent live in neighbouring countries (UNHCR, 2025).

Sudan presents a challenging context for digital trace analysis. Nationally, the internet penetration rate is just 29 percent (World Bank, 2022). However, refugees across many contexts often rely on mobile technologies during migration (Kingsley, 2016), and the scale of displacement from Sudan may be sufficiently large that it can be detected through trends in online behaviour, even if search patterns are driven by a minority of migrants. Most movement out of Sudan involves overland travel, and platforms like Wikipedia and Google may offer useful information about locations that are relevant to specific travel routes.

This study examines whether records of online information-seeking are linked to patterns of cross-border migration from Sudan into neighbouring Egypt and Ethiopia. I compare cross-border migration records (from IOM and UNHCR, April 2023 to May 2025) with patterns of online interest for locations along migration corridors, as captured by page views on Wikipedia (filtered by language) and search volume on Google Trends (filtered by country). Sudanese migrants are assumed to be more likely to search in Arabic rather than other languages.

## RESEARCH OBJECTIVE

RQ: Can online search activity serve as an indicator of refugee flows in low-income and conflict-affected settings?

This research aims to detect signals of migration behaviour rather than develop precise population estimates, recognising the inherent limitations of digital trace data in representing complete populations. It is hoped that findings from the study could help inform the development of monitoring systems for future displacement crises.

DATA

The data sources used in this study include:

1. View counts for Wikipedia pages about different transit towns along the Sudan-Ethiopia corridor (pageviews.wmcloud.org), disaggregated by language.
2. Relative search volume (0-100 scale) on Google Trends for the names of transit towns along the Sudan-Egypt corridor (trends.google.com), filtered for searches originating from within Sudan.
3. Border crossings from Sudan into neighbouring countries since April 2023, as reported by IOM.
4. Refugee stocks of Sudanese nationals in neighbouring countries since April 2023, as reported by UNHCR.

METHODS

- Google search volume is analysed monthly, using place names as "Topics" in order to capture broad activity.
- Wikipedia page views are grouped to weekly counts and analysed by language: Arabic (proxy for Sudanese users), Amharic (proxy for Ethiopian users), English (proxy for global interest), Russian (control language).
- Association between migration and online activity is measured using Spearman's Rank and Pearson correlation coefficients. Lead-lag relationships are explored using Vector Autoregressive (VAR) models and granger-causality tests.
- A set of controls are measured for both Wikipedia (control pages) and Google (control Topics).

RESULTS

Preliminary analysis of the Sudan-Egypt corridor found a sharp rise in information-seeking for transit towns in northern Sudan at the onset of the conflict - both for Wikipedia views (in Arabic) and Google searches (made in Sudan) - with online activity rising and falling with the first wave of Sudanese refugees into Egypt (April-June 2023).

Weekly crossings of Sudanese nationals into Ethiopia are found to be positively correlated with Wikipedia views for transit towns along the Sudan-Ethiopia corridor ($r = 0.47\text{-}0.55$, $p < 0.001$), specifically in the Arabic language, with weaker or insignificant associations for English, Amharic, and Russian. Granger causality tests found that Wikipedia views for locations on the Sudanese side of the border are predictive of border crossings (1 week lead, $p < 0.05$), while border crossings are predictive of page views for locations on the Ethiopian side (1–2 week lag, $p < 0.05$). No significant relationship was found for "control" Wikipedia pages of Ethiopian cities that are further inland and not directly linked to migration routes.

This study is ongoing, but initial findings suggest that online activity may capture meaningful signals of refugee movements in conflict settings, where timely migration information is often limited.