

Extended abstract: Monitoring cross-border refugee flows from Sudan using Wikipedia

Jackson Mason-Mackay

OVERVIEW

Forced displacement represents one of the most rapidly expanding forms of human mobility. The number of forcibly displaced people worldwide has tripled in the last decade, reaching over 120 million according to the United Nations High Commissioner for Refugees (UNHCR, 2025a). Forced migrants now account for one in every six international migrants (United Nations, 2025). Despite this scale, forced migration remains understudied in demographic research.

So far, 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 (Sanlitürk & 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, 2025). The use of digital trace data to measure conflict-induced displacement in the global south remains rare, despite these contexts facing significant information gaps and limitations for field-based data collection.

Sudan's ongoing armed conflict, which began in April 2023, represents the world's largest displacement crisis (UNHCR, 2025a). Over 3 million refugees have crossed into neighbouring countries - Chad, Egypt, Ethiopia, Libya, South Sudan, and the Central African Republic - while 8 million people remain internally displaced (UNHCR, 2025b; IOM 2025). 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 (UNHCRa, 2025).

Sudan presents a challenging context for digital trace analysis. Nationally, the internet penetration rate is just 29 percent (World Bank, 2022). However, refugees often rely on mobile technologies during migration (Kingsley, 2016), and the scale of displacement from Sudan may be sufficiently large that it has impacted trends in online behaviour, even if search patterns are driven by a minority of migrants.

This research examines whether online information-seeking behaviour, as captured through Wikipedia page views for destination locations, can serve as a proxy measure for cross-border displacement flows. The study focuses on the case of Sudanese refugees entering Ethiopia since the outbreak of conflict in April 2023, leveraging the language dimension of Wikipedia to distinguish between different user populations.

Wikipedia is a free online encyclopedia that offers detailed information on geographic locations, which may be useful for migrants who are planning their journey and are unfamiliar with their destination. Research by Vieira, Şanlıtürk and Zagheni (2025) demonstrated that Wikipedia searches for Polish cities were linked to Ukrainian migration patterns, highlighting its potential as an indicator of population movements.

RESEARCH OBJECTIVE

RQ: Is it feasible to use Wikipedia page views as an indicator of refugee flows in contexts with limited digital access?

The research aims to detect signals of migration behaviour rather than develop precise population estimates, recognizing the inherent limitations of digital trace data in representing complete populations. By analysing page views for Ethiopian cities and border towns in Arabic—a primary language in Sudan but not Ethiopia—the study seeks to isolate online activity likely attributable to Sudanese migrants from broader international interest.

DATA

Wikipedia allows for free access to the daily counts of page views¹, which can be accessed through the Wikimedia API or directly from the webpage and includes historical records since 2015. In this analysis, I utilise daily views between January 2022 and May 2025 for Wikipedia pages dedicated to major cities in Ethiopia (250,000+ inhabitants), as well as for border towns that are specific to the migration routes. These page views are analysed separately by language, which serve different functions:

- Arabic (proxy for Sudanese users)
- Amharic (proxy for local/Ethiopian users)
- English (proxy for global interest)
- Russian and Italian (control languages)

As the best available source of “ground truth” data, I use daily cross-border flows from Sudan to Ethiopia reported by IOM since April 2023². IOM monitors key points of entry at the Sudan-Ethiopia border and publishes the number of persons crossed into Ethiopia each day on an online dashboard as part of their Displacement Tracking Matrix (DTM). The number of persons crossed reflects national-level movements and IOM does not provide subnational estimates.

METHODS

In order to reduce noise, both data sources were aggregated from daily to weekly counts. Pearson correlation coefficients are used to measure the strength of association between migration flows and Wikipedia page views for Ethiopian towns, and Vector Autoregressive (VAR) models were used to identify potential lead-lag dynamics.

Building on the VAR results, Granger causality tests were applied to evaluate predictive directionality. Granger causality test asks whether the lagged values of one series improve forecasts of the other (beyond the series’ own lags). It evaluates predictive content, but not structural causation.

In addition, a relative change metric was computed to capture how Wikipedia activity evolved in response to the conflict outbreak. Weekly page views for each location were expressed as a percentage difference from a pre-conflict baseline period (March 2022–March 2023). This provided a measure of how much online activity increased or decreased compared to typical levels before the crisis. Finally, structural-break analysis was applied to identify points in time when the underlying behaviour of the time series shifted. This helped to determine whether major breakpoints coincided with key events in the Sudanese conflict or surges in refugee movements.

Similar methods have been applied to measure displacement, as developed in the pioneering work by Vieira, Şanlıtürk and Zagheni (2025).

¹ Available at pageviews.wmcloud.org

² Displacement Tracking Matrix: dtm.iom.int

RESULTS

Arabic-language Wikipedia page views for towns located along the Sudan–Ethiopia migration corridor showed the strongest association with recorded refugee inflows. The relationship was particularly pronounced for Gondar ($r = 0.68$, $p < 0.001$) and Metemma ($r = 0.60$, $p < 0.001$), both of which lie close to the Sudanese border and serve as key transit points for arriving refugees. Cities further inland demonstrated much weaker correlations, suggesting that online searches in Arabic were concentrated around locations directly linked to migration routes rather than larger, more distant urban centres.

Vector Autoregressive (VAR) models and granger causality testing indicated that fluctuations in border crossings preceded changes in Wikipedia page views, rather than the reverse. For example, border crossings improved the prediction of Arabic page views for Gondar ($p < 0.05$) and Metemma ($p < 0.01$), whereas Wikipedia activity did not significantly predict future border movements ($p > 0.05$). The estimated lag length of one to two weeks suggests that increases in migration flows were typically followed by heightened online information-seeking shortly after arrival. This temporal ordering implies that refugees were more likely to use Wikipedia to navigate or familiarise themselves with their destinations post-arrival rather than for pre-departure planning.

Arabic page views showed the strongest correlations with border crossings, while English ($r = 0.23$, $p = 0.015$) and Amharic ($r = 0.27$, $p = 0.005$) exhibited weaker associations. Italian and Russian displayed no significant relationships ($p > 0.1$), indicating that the signal is primarily driven by Arabic-language users rather than general international interest.

The relative change analysis revealed a sharp increase in Arabic views for Gondar during the third week of April 2023 - immediately following the conflict outbreak - with page views more than doubling from pre-conflict levels and peaking at roughly 650 percent above baseline by August 2023. Structural-break tests confirmed that these shifts were statistically significant and aligned closely with key conflict and displacement events.

CONCLUSION

Together, these findings indicate that language-specific Wikipedia activity can capture meaningful signals of cross-border displacement, when considering locations that are specific to the migration route (such as border towns), and using a language that is associated with migrants but not residents (Arabic).

This research remains exploratory, and future work will seek to refine and extend the approach by incorporating additional digital sources - particularly Google Trends - to capture pre-departure information-seeking that may not appear in Wikipedia activity. Integrating multiple digital indicators with official displacement data may help to improve the temporal sensitivity and reliability of monitoring systems, contributing to a more comprehensive understanding of forced migration in data-scarce contexts.

REFERENCES

- Alexander, M., Polimis, K., & Zagheni, E. (2019). The Impact of Hurricane Maria on Out migration from Puerto Rico: Evidence from Facebook Data. *Population and Development Review*, 45(3), 617–630.
<https://doi.org/10.1111/padr.12289>

- International Organization for Migration (IOM). (2025). *Displacement tracking matrix (DTM)*. <https://dtm.iom.int/sudan>
- Kingsley, P. (2016). *The New Odyssey: The Story of Europe's Refugee Crisis*. Guardian Faber Publishing.
- Leasure, D. R., Kashyap, R., Rampazzo, F., Dooley, C. A., Elbers, B., Bondarenko, M., Verhagen, M., Frey, A., Yan, J., Akimova, E. T., Fatehkia, M., Trigwell, R., Tatem, A. J., Weber, I., & Mills, M. C. (2023). Nowcasting Daily Population Displacement in Ukraine through Social Media Advertising Data. *Population and Development Review*, 49(2), 231–254. <https://doi.org/10.1111/padr.12558>
- Merisalo, M., & Jauhiainen, J. S. (2020). Digital Divides among Asylum-Related Migrants: Comparing Internet Use and Smartphone Ownership. *Tijdschrift Voor Economische En Sociale Geografie*, 111(5), 689–704. <https://doi.org/10.1111/tesg.12397>
- Palotti, J., Adler, N., Morales-Guzman, A., Villaveces, J., Sekara, V., Herranz, M. G., Al Asad, M., & Weber, I. (2020). Monitoring of the Venezuelan exodus through Facebook's advertising platform. *PLOS ONE*, 15(2), e0229175. <https://doi.org/10.1371/journal.pone.0229175>
- Sanliturk, E., & Billari, F. C. (2025). Search for a New Home: Refugee Stock and Google Search. *International Migration Review*, 59(1), 61–89. <https://doi.org/10.1177/01979183241275452>
- United Nations. (2025). *International migrant stock 2024: Key facts and figures*. *UN DESA/POP/2024/DC/NO. 13*.
- United Nations High Commissioner for Refugees. (2025a). Global trends: Forced displacement in 2024. UNHCR. <https://www.unhcr.org/refugee-statistics>
- UNHCR (2025b) Sudan Crisis: Socioeconomic Profile Overview. UNHCR Operational Data Portal (ODP): <https://data.unhcr.org/en/documents/details/114175>
- Vieira, C. C., Sanliturk, E., & Zagheni, E. (2025). Wikipedia as a Tool for Tracking Mass Migration Flows: Insights from the Russian Invasion of Ukraine. *[Unpublished manuscript]*.
- World Bank (2022). Individuals using the Internet (% of population). <https://data.worldbank.org/indicator/IT.NET.USER.ZS>