

UNHCR AS GENERATOR OF DATA AND EVIDENCE

EVALUATION EVIDENCE FOR STRATEGIC DIRECTIONS 2027–2031



UNHCR as Generator of Data and Evidence

About this paper

This paper is part of the UNHCR Evaluation Synthesis Series — 12 papers drawing on 2,421 conclusions, good practices, and lessons learned from 122 independent evaluations commissioned by UNHCR between 2016 and 2026; this paper draws on 19 of those evaluations. To produce the series, an automated Python script extracted these findings verbatim from published evaluation reports, compiling them into a database. This structured foundation enabled the Evaluation Office to develop the papers through iterative dialogue using Claude Sonnet 4.6 (Anthropic). The Evaluation Office is responsible for all analytical judgments and final text.

In brief. UNHCR collects substantial data on populations, needs, activities, and feedback, but collection and use are disconnected. Data is decentralised, siloed across divisions and systems, inconsistently defined, and variable in quality. The result is an organisation that does not consistently assess, at country or global level, what difference its efforts are making. Investment in data systems and analytical capacity has grown in recent years, but uptake and use across operations remain limited. In the end, UNHCR measures more than it learns from what it measures.

1. Registration Is UNHCR's Foundational Data Asset and Its Most Underleveraged

Registration is UNHCR's most developed data infrastructure, and the collection/use gap is already visible within it. PRIMES and the proGres4 case management system represent a substantial institutional investment: age and sex disaggregation is now systematic in most operations, and where proGres4 is functioning well it is the backbone of targeting, referral, and case management.² In Brazil, UNHCR's support for the SISCONARE digital registration system transformed the national asylum application process, enabling more efficient processing, improved outcomes, and better data for advocacy with government and the private sector.⁴ In Mexico, a quality assurance initiative with COMAR explicitly incorporated AGD data collection and capacity building into registration processes, improving the consistency and analytical value of population data.²

The limits are well documented. Disability and gender identity data are collected inconsistently. The reasons cited include limited resources and time during data collection, limited knowledge of the policy and of proGres4's full potential, and cultural context.⁵ In Greece and Kenya, the handover of registration processes to government created data continuity risks: disaggregated data collected under UNHCR management risked being lost when programmes transferred, eliminating the accountability function that registration data provides.² The 2019 data use evaluation found the more fundamental structural problem: UNHCR's approach to data management is heavily decentralised, with information needs defined and tools developed by sectoral and thematic teams independently of each other. Numerous systems exist. But they are not consistently interoperable, making cross-operational learning structurally difficult and the aggregation of data for strategic analysis largely impossible.¹

"There is an internal perception that UNHCR lags behind other UN agencies, partners and others in its ability to produce and use data and evidence, especially quantitative evidence."¹

2. Needs Assessment and Protection Monitoring Are Indispensable to Programming and Funded as If They Are Not

Needs assessment and protection monitoring are under-resourced relative to service delivery, generate data that rarely feeds into programme design, and are treated across operations as periodic exercises rather than continuous systems. The consequence is that UNHCR's evidence base is stronger at the start of an emergency, when data collection is funded and prioritised, than in protracted situations, where continued investment is harder to sustain.^{3,18,19}

In northern Ethiopia, the L3 was not informed by needs assessments; the operation lacked information about rapidly changing needs, and data were poor regarding people with special needs and Eritrean refugees. These gaps directly limited the response's ability to target and adapt.⁶ In Sudan, UNHCR faced persistent difficulty capturing results in ways that promoted learning, compounding the problem of acting without reliable ongoing needs data.⁷ In DRC, despite UNHCR investing the bulk of its L3 resources in protection monitoring, the function did not add as much value as it should have due to a lack of service orientation: data was produced but not converted into action or linked back to programming decisions.⁸ There, the protection cluster reported reaching 8% of targeted people, while WASH and shelter sectors reported 103% and 43% respectively. The gap reflected a reporting system designed to count sectoral outputs, with no mechanism for capturing protection work mainstreamed across other sectors. That work was happening but did not appear in the coverage data.⁸

Community-based protection monitoring, when dedicated resources are committed to it, demonstrates the function's potential. In Ethiopia, protection help desks and trained community monitors maintained coverage and generated needs data when humanitarian access was cut, extending UNHCR's analytical reach where centralised data collection was impossible.⁶ In DRC, GBV, CBI, and capacity-building strategies were adapted based on community feedback triangulated with participatory assessment data, a model in which monitoring data functioned as a real-time programming input.⁹ The gap between these cases and the norm is not technical capacity. It is dedicated resourcing and institutional recognition that protection monitoring is a data function, not an operational by-product.

“Protection monitoring was potentially a strategic use of UNHCR resources but did not add as much value to the L3 response due in large part to lack of a service-oriented approach.”⁸

3. UNHCR Monitors What Its Systems Can Count, Not What Its Work Is Meant to Change

The output trap is the most cross-cutting data finding in the evaluation portfolio. Across virtually every emergency, protracted situation, and thematic programme reviewed, monitoring systems track activities and services delivered rather than whether outcomes changed for forcibly displaced and stateless people.^{8,10,21} In DRC, outcome monitoring was absent outside cash-based interventions.⁸ In West and Central Africa, shelter data captured household-level outputs, not community-level outcomes.¹⁰ The protection strategies evaluation found that UNHCR's monitoring system, which gives country offices high discretion in indicator selection, makes it difficult to track performance against strategy objectives over time or to compare across countries.¹¹ COMPASS, introduced in 2021, is an attempt to shift monitoring from outputs to outcomes; evaluations that post-date its introduction consistently note its potential but find effects not yet visible in practice.^{10,20}

Data quality compounds the problem of output-focused monitoring. Even the output data that exists is often unreliable. In Brazil, the uneven quality of quantitative data on planning, budgeting, and monitoring — fragmented across COMPASS, Workday, PROMS, and ActivityInfo — made meaningful analysis of organisational efficiency difficult.⁴ In Peru, a lack of standardised indicator formulation across zones, with new indicators introduced mid-implementation, created confusion and affected programmatic coherence.¹² The 2019 data use evaluation concluded that the core barrier was not measurement difficulty but data governance, interoperability, and analytical capacity. Progress has been made on each since; the gap that remains is in translating available data into analysis that informs decisions.¹

Definitional gaps are the third and least visible problem, and the hardest to fix because they are embedded in the frameworks that generate the data. Urban programming illustrates the structural version of this problem: there is no agreed UN or global definition of what constitutes an urban setting, governments each define it differently, and UNHCR has no urban marker in its RBM, finance, or budget systems. As a result, UNHCR cannot determine how much it spends on urban programming, whom it reaches, or what works across contexts. The limitation lies not in data quality but in the absence of an urban marker at system level. A forthcoming global evaluation on economic inclusion in urban areas addresses this gap.¹

The AGD framework shows the same problem at the conceptual level. In practice, gender is systematically interpreted as a proxy for sex and narrowly understood as a focus on women and GBV, leaving masculinities and the structural causes of gender inequality largely unaddressed.⁵ The D in AGD is frequently operationalised as disability specifically rather than diversity more broadly, and disability data itself remains a recognised gap: there is a dearth of robust data on refugees with disabilities and limited and partial information for SOGIESC, with limited resources, limited knowledge, and cultural context all cited as contributing factors.^{13,5} These three problems (output-focused monitoring, data quality, and definitional gaps) reinforce one another: what is measured cannot be improved if the definitions are contested and the data entering the system is unreliable.

“UNHCR’s monitoring and reporting system, which provides a high level of discretion to country offices in the selection of monitoring indicators, makes it difficult to track performance in relation to the strategies over time or to compare performance across countries.”¹¹

4. UNHCR Collects Feedback But Rarely Reports Back to Affected People

Feedback and complaint mechanisms are the data function most directly connected to accountability to affected populations, and the evidence is consistent: feedback is collected, but the loop back to communities is rarely closed. The AAP synthesis, drawing on 17 evaluations, identified four structural barriers: limited dedicated capacity for AAP; weak information management capacity for integrating feedback into programmatic review; lack of available guidance and tools; and weak community engagement.^{9,21} Participatory assessments are widely conducted as an annual compliance exercise. The data they produce rarely feeds into programme adaptation.^{9,2}

The MENA phone-based communication channels evaluation is the portfolio's most detailed analysis of what happens between collection and use. Most country offices collect basic call data such as volume, location, and query type. Few conduct systematic analysis of that feedback to inform programme learning and adaptation. Performance tracking favours

quantitative metrics over user experience and protection outcomes, and efficiency indicators centre on call volume rather than on whether the system is effective.¹⁴ The AGD longitudinal evaluations identify the mechanism: there is a disconnect between the feedback collected through participatory exercises and what is reflected in programming decisions. There are no consistent mechanisms ensuring systematic analysis of feedback data feeds into operations.^{2,13}

The positive models are specific about what closing the loop requires. In Jordan, helpline data is used by senior management in real time: call trends, caller locations, and query subjects inform case prioritisation and programme adaptation; staff working on cash and resettlement use call frequency data to identify priority cases.¹⁴ In Tunisia, operators perform outbound follow-up calls to communicate outcomes of all queries, making the feedback loop a routine practice.¹⁴ In Iraq, all communication and feedback data from multiple channels are combined in a single CFM report shared with management, enabling integrated programme learning.¹⁴ In Egypt, UNHCR included refugee-led organisations in the RCM to collect additional information on needs and priorities, institutionalising community voice within the coordination structure itself.²¹ In Lebanon, SGBV WhatsApp networks reached approximately 60,000 families, generating community-level protection data that would otherwise have been invisible.¹⁵ What distinguishes each case is not technology or channel. It is that the feedback data was routed to someone with the authority and responsibility to act on it.

“Current performance tracking tends to emphasize quantitative metrics over qualitative aspects like user experience or protection outcomes, missing opportunities for meaningful improvement aligned with AAP commitments.”¹⁴

5. UNHCR Is Ceding Its Data Authority to Actors Who Invest in Analysis, Not Just Collection

UNHCR's authority as the definitive source on forced displacement is being actively eroded. Other UN agencies have made deliberate investments in data science, statistics, and econometrics that UNHCR has not matched, and actors who present more compelling and analytically rigorous data are filling the space. The 2019 data use evaluation is explicit: UNHCR is missing opportunities to use data and analysis in advocacy with states, especially when forced displacement discourses are politicised. Beyond missed opportunities, the evaluation finds UNHCR's standing as the authority on displacement data eroding.¹

The cases where UNHCR has used data strategically show what is possible. In Brazil, studies conducted jointly with the World Bank and IFC demonstrated refugees' and migrants' financial and economic contributions to the economy, shifting the advocacy frame from humanitarian obligation to economic case, and providing government and private sector actors with evidence they could act on.⁴ In the 3RP Syria response, a highly effective knowledge and information management function captured the complexity of the crisis in a user-friendly manner that underpinned advocacy credibility and resource mobilisation at the Brussels Conference, generating approximately \$4.84 billion annually; stakeholders identified information quality as a key strength distinguishing the 3RP from bilateral appeals.¹⁶ In Jordan, helpline data aggregated across the CO and used by senior management in real time converted a feedback mechanism into a strategic planning tool, with community communication data shaping programme decisions at the management level.¹⁴

The statelessness evaluation makes the cost of not investing in strategic data concrete: gaps in reliable data on stateless persons' numbers and lived experience made it more difficult to

communicate and prompt action; progress was made more easily where political will already existed, not in the contexts where data was most needed to build the case.¹⁷ The 2019 data use evaluation's conclusion remains valid: the gap between what UNHCR wants from its data and where it currently stands is not insurmountable. Investments in data integration, data capacity, and data norms and standards could materially close it. UNHCR has strong data and information management practices in a number of places. The task is to institutionalise them before the space those they could occupy is permanently filled by others.¹

*"As the UN Refugee Agency, UNHCR ought to be an authority on what is happening with refugees and other persons of concern, but this authority is being eroded by other players who are presenting and sharing data and information that is sometimes more compelling and interesting than UNHCR's data."*¹

6. Conclusion

Investment in data systems has grown, but the gap between collection and use persists across every layer of UNHCR's data architecture. Registration captures population data comprehensively but without the analytical infrastructure to convert it into strategic decisions. Needs assessment and protection monitoring are funded as periodic exercises rather than continuous systems, so the evidence base is strongest at the start of an emergency and thinner when programming decisions depend on it. Results frameworks compound the problem by measuring what systems can count rather than what programmes are meant to change, and COMPASS has not yet closed that gap in practice. Feedback mechanisms are established across operations, but data routed to no one with authority to act produces accountability to no one. The positive models documented here show that the problem is one of design, not capacity. UNHCR's next Strategic Directions depend on evidence-based advocacy and demonstrable results. Building that evidence base is an institutional choice UNHCR has not yet operationalised.

Endnotes

Page numbers refer to the physical page position in the PDF version of each evaluation report and may differ from the printed page numbers shown in the document.

1. [Evaluation of UNHCR's Data Use and Information Management Approaches](#) (2019), p. 8.
2. [Longitudinal Evaluation of UNHCR's Age, Gender and Diversity \(AGD\) Policy: Final Report](#) (2023), p. 69.
3. [Evaluation of UNHCR's Response to the L3 Ethiopia](#) (2023), p. 56.
4. [Country Strategy Evaluation Brazil](#) (2025), p. 78.
5. [Longitudinal Evaluation of the Implementation of UNHCR's Age, Gender and Diversity \(AGD\) Policy: Year 1 Report](#) (2022), p. 9.
6. [Evaluation of UNHCR's Response to the L3 Ethiopia](#) (2023), p. 14.
7. [Independent Evaluation of the UNHCR South Sudanese Refugee Response in White Nile, Sudan](#) (2018), p. 14.
8. [Independent Evaluation of UNHCR's Response to the L3 Emergency in the Democratic Republic of the Congo](#) (2018), pp. 28, 48.
9. [UNHCR's Approach to Accountability to Affected People \(AAP\): Synthesis of Evaluations](#) (2022), p. 9.
10. [West and Central Africa Regional Shelter and Settlement Evaluation](#) (2022), p. 11.
11. [Evaluation of UNHCR's Implementation of Three of its Protection Strategies: the Global Education Strategy, the Updated SGBV Strategy, and the Child Protection Framework](#) (2017), p. 58.
12. [Country Strategy Evaluation Peru](#) (2024), p. 15.
13. [Longitudinal Evaluation of UNHCR's Age, Gender and Diversity \(AGD\) Policy: Synthesis Report Year 2](#) (2023), p. 26.
14. [Independent Evaluation of Data Systems in MENA: Multi-country Evaluation of Phone-Based Communication Channels](#) (2025), p. 80.
15. [Evaluation of UNHCR Prevention and Response to SGBV in the Refugee Population in Lebanon](#) (2018), p. 25.
16. [Regional Refugee and Resilience Plan for Syria \(3RP\) Evaluation](#) (2022), p. 53.
17. [Evaluation of UNHCR-led Initiatives to End Statelessness](#) (2021), p. 9.
18. [Country Strategy Evaluation Somalia](#) (2025), p. 71.
19. [Evaluation of UNHCR's Multi-country L3 Emergency Response to the Sudan Situation](#) (2025), p. 82.

20. [Evaluation of UNHCR Approach to Gender-Based Violence \(GBV\) Prevention, Risk Mitigation and Response \(Phase II\)](#) (2025), pp. 42-43.
21. [Evaluation of UNHCR's Multi-country L3 Emergency Response to the Sudan Situation](#) (2025), pp. 32, 77-78.