Reference Paper for the 70th Anniversary of the 1951 Refugee Convention

FORCED DISPLACEMENT DATA:
Critical gaps and key opportunities in the context of the Global Compact on Refugees

Natalia Krynsky Baal
World Bank - UNHCR Joint Data Center on Forced Displacement

March 2021

Abstract:

Increased global attention to refugee situations and forced displacement more broadly, brings with it increased scrutiny over the available data. Are the numbers reported accurate? Whose data is trustworthy? How does the situation of refugees or IDPs empirically compare to other vulnerable and impoverished groups? It is no wonder then that the Global Compact on Refugees includes explicit, carefully worded but ambitious recommendations on data. This paper takes a bird's eye view over the current forced displacement data landscape and focuses in to describe a number of persistent and critical gaps. It points out various opportunities through which these gaps may be addressed or at least minimized, and highlights the collective effort required to make headway in this field to improve the policy-relevant data and evidence available on forced displacement.

Acknowledgments: I am grateful to a range of colleagues who provided contributions and inciteful comments on an earlier version of this paper, including Karoline Eberle, Ninette Kelley, Arthur Alik La Grange, Felix Schmieding, Caroline Sergeant and Sigrid Weber. I am also grateful to many other colleagues and mentors who I have learnt from and with over the last decade working on related subject matter: you know who you are.

This paper was originally prepared to serve as background material for the *UNHCR's People Forced to Flee: History, Change and Challenge*. The views expressed in this paper, and any mistakes it may contain, are those of the author's only.

CONTENT

INTRODUCTION

Against a backdrop of increasing numbers of forcibly displaced persons in recent years, the Global Compact on Refugees was agreed upon by Member States. It recognized the increasingly protracted nature of refugee situations and the limited progress in securing any one of the three durable solutions for most of the world's refugees.

The Global Compact on Refugees makes explicit and carefully worded recommendations on data.[1] It calls for "reliable, comparable, and timely data" to "improve socio-economic conditions for refugees and host communities; assess and address the impact of large refugee populations on host countries in emergency and protracted situations; and identify and plan appropriate solutions." It advocates for the application of "relevant data protection and data privacy principles" and calls for "harmonized or interoperable standards for the collection, analysis, and sharing of age, gender, disability, and diversity disaggregated data". It further highlights the importance of strengthening national data collection systems and encourages states to seek support as needed to include refugees and host communities, as well as returnees within them. For practitioners with a decent level of data literacy and some awareness of the current data landscape on forced displacement, this is a tall order to say the least.

Recognizing the critical need to improve the quality and availability of data on forced displacement to better address this complex global phenomenon, should not be exclusively credited to the Global Compact on Refugees and the declarations and processes that led to its development. Improved data was also a recurring topic during the 2016 World Humanitarian Summit culminating in a series of commitments articulated in the Grand Bargain linking impartial and joint needs assessments to more informed funding for humanitarian crises globally.[2] The UN Secretary General's High Level Panel on Internal Displacement has identified data as one of its five priority themes,[3] and improved quantitative and qualitative data on stateless populations is one of the ten points outlined by the Global Action Plan to End Statelessness 2014-2024.[4] From a development perspective, the World Bank has also clearly articulated a strong call for a significant and coordinated investment to address existing forced displacement data challenges in its 2017 Flagship report on the topic. [5]

---

[1] UN, 2018, *Global Compact on Refugees*, paragraph 45-48, https://www.unhcr.org/5c658aed4

[2] The Grand Bargain – A Shared Commitment to Better Serve People in Need, 2016, https://interagencystandingcommittee.org/system/files/grand_bargain_final_22_may_final-2_0.pdf

[3] Terms of Reference HIGH-LEVEL PANEL ON INTERNAL DISPLACEMENT, https://www.un.org/internal-displacement-panel/sites/www.un.org.internal-displacement-panel/files/tor_of_the_panel.pdf

[4] Global Action Plan to End Statelessness 2014-2024, https://www.unhcr.org/ibelong/global-action-plan-2014-2024/

[5] World Bank. 2017, p27-33, *Forcibly Displaced: Toward a Development Approach Supporting Refugees, the Internally Displaced, and Their Hosts*, https://openknowledge.worldbank.org/bitstream/handle/10986/25016/9781464809385.pdf?sequence=11&isAllowed=y

With this increased political attention there is a lot of work underway at national, regional and international levels but more work is needed to coordinate efforts and ensure the positive impact envisioned on policy, operations and public discourse, and, ultimately to improve the lives of forcibly displaced persons. Recognizing that in many of the countries most heavily affected by forced displacement the existing data landscape is generally poor; capacity building will be a big part of making this work. And with momentum from the ongoing COVID-19 pandemic, the importance of integrating innovation into this transformative data agenda has also been thrust to the front of the queue.

This paper provides an overview of key elements that must be integrated into the proposed transformative data agenda on forced displacement. It first provides a brief overview of the current data landscape on forced displacement as observed by the author; it then lays out a selection of persistent and critical gaps, and subsequently responds to each of these by identifying relevant opportunities that can be further enhanced to make a meaningful difference. It concludes by drawing attention to the concrete plans of the newly established WB-UNHCR Joint Data Center set up to significantly contribute to this collective effort that is so urgently needed to improve the situation of people forced to flee.[6]

THE CURRENT DATA LANDSCAPE

According to UNHCR's Global Trends 2019,[7] there were 79.5 million people forcibly displaced at the end of 2019 as a result of persecution, conflict, violence, human rights violations or events seriously disturbing public order. Out of this total, 26 million were refugees, 4.2 million were asylum-seekers and a further 3.6 million Venezuelans were displaced abroad. Significantly, 85% of the world's refugees and displaced Venezuelans were hosted in developing countries. In addition to those who have been forced across international borders, 45.7 million of this total figure were internally displaced people - by far the largest group. Concerning stateless people, UNHCR reports on some 4.2 million people[8], although it is widely recognized that this is a significant underestimate with the unverified global estimate of 10 million people often cited.

UNHCR's Global Trends, alongside other notable publications[9], provides critical analysis of the population figures and demographic breakdown of forcibly displaced populations. A breakdown of refugee and IDP figures is provided, for example, by country of origin, host country, and as far as possible by age and sex – although even here data is incomplete and global estimates include values from statistical modelling[10]. However, very little data is provided on the socio-economic circumstances of the different population groups covered. Although there are some exceptions,

[6] Connections between this background paper and chapters of the main publication are particularly relevant for Part III on 'Solutions', Part IV on 'Improving Life Prospects' and Part V on 'Bridging the Gap'.
[7] UNHCR, 2020, Global Trends: Forced Displacement in 2019, https://www.unhcr.org/5ee200e37.pdf
[8] UNHCR, 2020, Global Trends, p56
[9] IDMC, 2020, Global Report on Internal Displacement, https://www.internal-displacement.org/global-report/grid2020/
[10] UNHCR, 2020, Global Trends, p14-15

an overview of key socioeconomic indicators such as education or poverty level is simply not available with extensive enough coverage and/or access to be included in any comprehensive global level analysis.

Data on forcibly displaced populations is primarily gathered by national authorities and both national and international humanitarian organizations through systems that aim, explicitly, to be used to inform protection, assistance, advocacy and programs. Existing data can often be used to identify or distinguish a population of concern from other groups, to assess immediate needs of refugees and IDPs or monitor protection risks and incidents that may occur. Most commonly used data systems include registration (especially for refugees), sectoral and multi-sectoral surveys (that collect new data), sectoral and multi-sectoral secondary data reviews (that bring together existing data from diverse sources), and a wide variety of qualitative data capture tools including focus group discussions, key informant interviews and participatory assessments.

In recent years, as demand for more granular data collected at the individual or household level has increased, qualitative assessments have been overtaken by quantitative surveys and an increasing amount of data collection at the household as opposed to community level frequently takes place. As technology advances and becomes more widely available, new approaches are also more readily available in humanitarian settings – such as mobile phone data tracking, social media analysis, satellite imagery etc. – but these still have limited usage.[11]

The data produced through the above mechanisms is of course highly valuable, but many argue it is significantly compromised or limited in important ways.[12] Before elaborating these limitations in the next section of this paper, it is important first to recognize the very challenging contexts in which the vast majority of this data is produced. [13] Firstly, political dynamics have their toll with a complex web of (largely untransparent) incentives driving stakeholders' investments in and analysis of data. These can range from incentives to exaggerate or under-report population figures to increase funding allocations or fit with a current political narrative of the peace/security/governance situation in a given region or country. Funding incentives within the aid sector also have a widespread impact creating competition between agencies and sectors and hindering collaboration. The common (politically motivated) perception of forced displacement as an overwhelmingly humanitarian issue also has a significant impact. Secondly, challenging operational realities also impact data quality and availability. Limited connectivity and institutional capacity as well as fluid population movements are common challenges, as are limited access for reasons of insecurity or simply inadequate infrastructure that can significantly frustrate data production efforts. With most forcibly displaced persons residing in out-of-camp

[11] Some examples can be found here: https://www.flowminder.org/what-we-do/data-science-analysis/ . In the context of the Covid-19 pandemic technological innovations are gathering further momentum, read more here: https://www.jointdatacenter.org/supporting-evidence-driven-responses-to-covid-19/

[12] Sarzin, Zara. 2017. *Stocktaking of Global Forced Displacement Data. Policy Research working paper, no. WPS 7985. World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/26183*

[13] All these issues are elaborated on in this paper: Natalia Baal and Laura Ronkainen, 2017, UNHCR Statistics Technical Series: Obtaining representative data on IDPs: challenges and recommendations, https://www.unhcr.org/598088104.pdf

or urban contexts, the complex nature of many urban environments such as over-crowing, insecurity, anonymity, and other aspects linked to informality, also present significant operational challenges.

CRITICAL GAPS

Standing back and taking stock reveals a series of significant gaps in the current data landscape on forced displacement. Rather than providing an exhaustive analysis, this section aims to outline those gaps – data gaps, analytical gaps, and even system-wide gaps – that are most pertinent in the context of the Global Compact.

Inconsistent definitions & methods

The increased attention on refugee and broader migration flows has been accompanied by an increased scrutiny on the available data. This in turn has quickly resulted in a realization that we are all too often comparing apples and pears. At the core of this problem is the existence of different definitions and methods used across partners and contexts.

As highlighted by the International Recommendations on Refugee Statistics there is a "Lack of comparability between statistics on refugees and asylum seekers produced by different countries, and across displacement situations within countries. This arises due to the lack of consistency of terminology, concepts, definitions and classifications, as well as variation in the methods of data collection, compilation and presentation at national and international levels."[14] This challenge effects even the more advanced data contexts demonstrated by the example of Norway, where Statistics Norway has often pointed out the divergence between their definition and way of counting refugees compared to that used by UNHCR to report global population figures based on an operational definition leading to a disparity of around 100,000 individuals.[15]

Concerning the definition of IDPs, the pictures becomes even murkier. Precisely which events trigger internal displacement? How far or for how long must you be displaced to count? Are children born to IDP parents also to be considered IDPs? Do refugees and forced migrants become IDPs when they return to their country of origin? When should IDPs no longer be defined as IDPs? Answers to these questions and others vary considerably in practice and cumulatively

[14] Expert Group on Refugee and Internally Displaced Persons Statistics, 2018, International Recommendations on Refugee Statistics (IRRS), paragraph 17: https://ec.europa.eu/eurostat/documents/3859598/9315869/KS-GQ-18-004-EN-N.pdf/d331c9cc-1091-43c2-b589-2c250bccc281

[15] The multiple reasons for this disparity are clearly described in this helpful article: Vebjørn Aalandslid, 2017, *Record numbers of displaced persons*, https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/record-numbers-of-displaced-persons

have a significant impact on the ability to compare data from within and across countries.[16] Whilst it has been recognized that "national and regional instruments largely follow the same IDP definition as the UN Guiding Principles on Internal Displacement [it is also clear that] interpretations of this definition and its practical applications for policy implementation and programming often vary, specifically in relation to interpretations on the end of displacement. This results in a range of practices when it comes to capturing internal displacement statistics."[17]

Inconsistent definitions of population categories are also related to the challenge of conflicting results and duplication of efforts that result from limited coordination, which unfortunately still characterizes many data collection efforts in forced displacement contexts (and humanitarian crises more generally). Multiple actors with varying motives or interests invest in duplicative or overlapping data collection efforts, often as part of agency-specific operational and programmatic commitments, that in practice result in unnecessary demands on the affected populations who are asked similar questions repeatedly. This can lead to significant "survey fatigue" and frustrations within the affected population as well as conflicting results – on both population data and vulnerability/needs analysis – that are in turn problematic for senior management and donors alike when undertaking prioritization decisions. Addressing this conundrum gained further high-level recognition and momentum through the Grand Bargain agreement which noted that the "proliferation of uncoordinated needs assessments leads to duplication, wasted resources and putting a burden on affected populations".[18]

Improving familiarity with and use of standard definitions and methods for population data by all stakeholders, as well as enhanced coordination, can go a long way to addressing this problem as explained further below.

Limited coverage of socio-economic data and comparison to hosts

Given the overwhelming focus on informing protection and assistance interventions – i.e. filling humanitarian data needs such as the number of new arrivals in a given location, identifying those with disabilities and other special needs in a certain camp, or assessing emergency shelter or basic food needs of an affected population – available data on forcibly displaced persons is often not well suited to development purposes where contextualized and comparative socio-economic data is critical. However, socio-economic data and analysis in refugee and IDP contexts is often

[16] Expert Group on Refugee and IDP Statistics, 2018, *Technical Report on the Statistics of IDPs*: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-003). Specifically, the challenge of inconsistent definitions on durable solutions, including several concrete examples, was described here: Chaloka Beyani, Natalia Krynsky Baal and Martina Caterina, 2016, *Conceptual challenges and practical solutions in situations of internal displacemen*t, Forced Migration Review: https://www.fmreview.org/solutions/beyani-baal-caterina.

[17] Expert Group on Refugee and IDP Statistics, 2018, *Technical Report on the Statistics of IDPs,* Chapter 3: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-003)

[18] https://interagencystandingcommittee.org/improve-joint-and-impartial-needs-assessments

either lacking (e.g. poverty rates and similar indicators)[19], geographically limited (e.g. to certain camps or urban areas), or, even when it does exist, is of limited quality as it is rarely aligned to existing statistical standards for key socio-economic indicators (e.g. on employment or poverty). This means that it cannot easily be compared to data on other population groups or national averages and thereby integrated into development planning processes.[20]

Comparative analysis - with other population groups in the same country and across countries – is the first victim of these limitations; our inability to systematically compare forcibly displaced persons with others is a direct consequence of a lack of high-quality, standardized socio-economic data. However, the ability to compare the situation of refugees or IDPs with the national population, host communities or even sub-groups thereof (e.g. economic migrants) is a burning and recurring policy need in so many displacement contexts. Without this it is very difficult to understand if displacement introduces specific hurdles to overcoming poverty in a given context, and thereby both justify and inform the appropriate targeting of development programs. In many countries, survey tools that are aligned to international statistical standards are used to measure the socio-economic profile of the population and ensuring data on refugees were collected in the same – or similar – ways would facilitate this comparison and provide a wealth of policy relevant analysis. This can be done either through including these groups in existing data collection efforts or aligning separately implemented ones to standardized national efforts to ensure results are comparable (see below).

If socio-economic comparison with other groups is a gap, longitudinal analysis – seeing how the situation changes over time to be able to monitor progress and flag any stagnation or worsening of a situation in a timely manner - is a crater. Very few examples exist where the socio-economic situation of a refugee of IDP community can be analyzed over time.[21] Of course it is important to recognize that even in non-fragile or developed contexts, collecting longitudinal data is complicated, costly and organizationally challenging, the benefits are significant for long-term policy outlooks to better understand integration of forcibly displaced persons and other migrant populations.[22]

Although many individual attempts to provide more standardized, more comparative and even more longitudinal data do exist – and the number is growing – the need to ensure the more

[19] World Bank, *Poverty and Shared Prosperity*, 2020, p128, https://openknowledge.worldbank.org/bitstream/handle/10986/34496/9781464816024.pdf . A forthcoming report on persons missing from poverty statistics will also go into more detail on the reasons why refugees, IDPs and others are omitted.

[20] Expert Group on Refugee and Internally Displaced Persons Statistics, 2018, International Recommendations on Refugee Statistics (IRRS), paragraph 17: https://ec.europa.eu/eurostat/documents/3859598/9315869/KS-GQ-18-004-EN-N.pdf/d331c9cc-1091-43c2-b589-2c250bccc281

[21] Here is a rare but interesting example of a longitudinal study: IOM, 2019 Access to Durable Solutions Among IDPs in Iraq: Four Years in Displacement, available in Arabic and English here: https://iraq.iom.int/publications/access-durable-solutions-among-idps-iraq-four-years-displacement

[22] NB the migration literature and practice on measuring integration of economic migrants is diverse. See Prof Russell King and Dr Aija Lulle, Research on Migration: Facing Realities and Maximizing Opportunities, chapter 2: integration, http://ec.europa.eu/research/social-sciences/pdf/policy_reviews/ki-04-15-841_en_n.pdf

sustainable provision of such data will be critical for implementation of the GCR in many regions. The reasons for the above limitations on socio-economic data are multiple and complex, but some central factors include a mixture of systemic and political dynamics. Forced displacement has traditionally been seen a humanitarian problem that will be short-lived/temporary and therefore should be resolved through humanitarian means. This is more politically digestible for many stakeholders involved and, whilst many argue it has never been true in practice, it has resulted in a systemic reality that has kept forced displacement separate from the broader development policy agenda and funding streams. Although perceptions are changing and policy is shifting, overtime this has had a direct impact on the data systems most regularly used that are funded through humanitarian financing mechanisms (often annual) and designed to inform or implemented as part of humanitarian response plans.

Disconnect with national systems

The above gaps and challenges (on population group definitions and socio-economic statistical standards) are intricately linked to this next one: the fact that refugees, IDPs and stateless persons, as well as other related population groups, are largely omitted from official statistics produced through national systems.[23] Given the need to enhance national response and include forcibly displaced persons in national systems and development programmes, this disconnect is particularly pertinent.

National statistical systems comprise a variety of government bodies, ministries, departments and agencies usually coordinated by a national statistical office with overall responsibility for official statistical standards.[24] Data are produced from a range of sources, including the national census, nationally representative thematic surveys, and administrative data systems that are managed by their members. They produce data that are locally owned and provide the backbone for national policy and development planning processes. In many contexts where national statistical systems have limited capacity, including in most fragile and displacement affected countries, support from international partners through financial and technical assistance is provided. The inclusion of forcibly displaced persons – recognized as such – within these systems is more often the exception rather than the rule (see below).

Initially recognized by UNHCR and Statistics Norway when they formally raised the issue to the UN Statistical Commission in 2015[25] the disconnect between forced displacement data and national statistical systems significantly hinders the inclusion of refugees in national policies and systems. Data produced through national statistical systems provide the backbone for national

[23] E/CN.3/2015/9, Report of Statistics Norway and the Office of the United Nations High Commissioner for Refugees on statistics on refugees and internally displaced persons, 2014, https://unstats.un.org/unsd/statcom/doc15/2015-9-RefugeeStats-E.pdf

[24] Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014): https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

[25] E/CN.3/2015/9, Report of Statistics Norway and the Office of the United Nations High Commissioner for Refugees on statistics on refugees and internally displaced persons, 2015, https://unstats.un.org/unsd/statcom/doc15/2015-9-RefugeeStats-E.pdf

development planning and associated budgetary allocations which are closely linked to shaping international development assistance.[26] Reliable statistics, produced through national systems, are also essential for measuring progress in reaching development goals and thereby enhance transparency and accountability of governments. The strategic value of including refugees in national statistical systems to furthering the objectives of the GCR was also highlighted in the final text of the Compact which states: "Upon the request of concerned States, support will be provided for the inclusion of refugees and host communities, as well as returnees and stateless persons as relevant, within national data and statistical collection processes; and to strengthen national data collection systems on the situation of refugees and host communities, as well as returnees."[27]

The political, technical, and operational challenges for integrating forcibly displaced and stateless persons into national data systems are numerous and are exaggerated further by the high demands on already over-burdened and under-resourced national statistical offices and systems. However, significant steps have been taken in recent years under the auspices of the Statistical Commission and there is growing support and momentum to overcome or mitigate the challenges demonstrated through the many countries that are already taking steps in the right direction (see below).

Difficulties of measuring impact

An issue that arises cumulatively from the above listed gaps, is the difficulty of adequately measuring the impact of forced displacement on host countries and host communities in a way that is comparable across contexts. This challenge was explicitly noted during the GCR negotiations as a specific analytical gap that threatened to hinder global progress on burden and responsibility sharing.[28] Without a common approach to take stock of the status quo in terms of where the social and economic impact currently falls – taking into account both hosting refugees but also contributing humanitarian and development assistance internationally – it is difficult to make fair and evidence-informed changes that will help to ease pressure on host countries (one of the GCR's four overarching objectives). Moving towards more equitable sharing of the "burden" of the world's refugee population, more data and evidence is needed to measure the impact arising from hosting, protecting and assisting refugees with methodologies needing to take into consideration the complexity of the issue, differences in level of development and economic growth, as well as local and regional contexts.

Following commitments made within the GCR and consolidated in the General Assembly resolution on the Office of the United Nations High Commissioner for Refugees in December

[26] Eurostat, 2013, *Guide to Statistics in European Commission Development Co-operation*, Part B 'Statistics in Development', https://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_in_development_cooperation_-_national_statistical_systems
[27] UN, 2018, Global Compact on Refugees, paragraph 46, https://www.unhcr.org/5c658aed4
[28] UN, 2018, *Global Compact on Refugees*, paragraph 48, https://www.unhcr.org/5c658aed4

2018,[29] UNHCR is coordinating a process amongst interested Member States, to take this work forward. It aims to reach consensus on a common methodology or methodologies to measure the impact of hosting, protecting and assisting refugees, and is benefitting from technical support of the World Bank and the Joint Data Center. Progress so far includes the successful completion of three workshops during 2019 with representatives from a range of Member States (both refugee-hosting and donor countries) who are actively participating in the process.[30]

Critical gaps or data related challenges that have arisen in this process concern capacity development, coordination and technical issues linked to well-recognised impact evaluation standards and the onerous data needs for a comprehensive methodology. Results from the process so far have flagged the need for capacity development of relevant national actors to be able to identify and aggregate the required data, as well as fully understand the objectives of the task. Given the multi-sectoral nature of the impact of hosting, protecting and assisting refugees and therefore the need for data from various line ministries and national institutions, the challenge of coordination and the delineation of clear lines of responsibility, have also been identified as pertinent challenges in several contexts.

The technical challenges were summarized neatly in the Progress Report issued in July 2020 which highlighted firstly, the need for a counterfactual comparison to adequately measure impact. "In other words, the difference in outcomes needs to be assessed relative to a scenario without refugee presence, all other elements remaining the same. This is challenging because refugee influxes are dissimilar to the subjects typically addressed by randomized control settings and do not usually allow for a valid comparison scenario." Secondly, it recognized the complexities of delineating relevant data given that refugee influxes often accompanied by "economy-wide shocks and spillovers from neighbouring conflicts. The effects of these are difficult to disentangle from those associated with refugee emergencies." And lastly, it highlights the complexities specific to certain sectors or topics that should be part of the comprehensive approach to measure impact.

These factors have been central to shaping the way forward; support to which presents a key opportunity for implementation of the GCR (see below).

Lack of disaggregated data

Understanding the experiences and realities of forcibly displaced persons in a differentiated manner, is critical for offering adequate protection and facilitating durable solutions. Disaggregated data – by age, gender, disability and other diversity criteria such as ethnicity or religion – is widely recognized as important to inform more targeted interventions but is also often missing. As mentioned above, even a breakdown of refugee and IDP population data by

[29] https://undocs.org/en/A/RES/73/151

[30] UNHCR, July 2020, Progress Report: Measuring the Impact of Hosting, Protecting and Assisting Refugees, https://www.unhcr.org/5f0570754.pdf

age and sex at the global level currently makes use of statistical modelling, as explained in UNHCR's *Global Trends*.[31]

Due to the prevalence of refugee registration systems which for the most part provide data on sex, age and nationality of refugees, the picture is far bleaker for IDPs about whom the quality of data sources varies considerably. IDMC have estimated the proportion of IDPs in 2019 by specified age groups, however limited explanation of the data sources or methods used to produce this estimate are given.[32] In a separate dedicated briefing paper on the topic of age disaggregated data for IDPs, IDMC explain these estimates are based on applying UNFPA's estimated national population percentage breakdown to IDP population data per country, thereby confirming these estimates do not advance our understanding on whether different age groups are disproportionally impacted by displacement.[33] For stateless populations, further data gaps are apparent. UNHCR confirms that reported data disaggregation is scarce, with sex-disaggregated data "available for 28 of the 76 countries reporting on stateless populations, covering 73 per cent of the reported stateless population."[34]

For data on disabilities, significant gaps exist for all population groups. Many reports apply the World Health Organization's (WHO) estimates that state roughly 15% of the world's population have a disability[35] and apply this estimate to refugee and IDP populations. This corresponds, for example, to the estimated 6.8 million of the 45.7 million persons internally displaced worldwide by conflict and violence at the end of 2019[36] and 3.9 million of the 16 million refugees. Through an analysis of four case studies – the Nepal earthquake (2015), the protracted crises in Somalia (1991-ongoing), the Rohingya refugee crisis (2017 – ongoing) and the European refugee and migrant crisis (2015) – a report from UNICEF and partners helpfully summarizes the challenges that contribute to this particular data gap. These include the pre-crisis situation in terms of data capacity and availability, the methodologies used by humanitarian responders, data sharing and coordination challenges and the capacities of states to meet their obligations under the Convention on the Rights of Persons with Disabilities.[37]

---

[31] UNHCR, 2020, *Global Trends*, p14-15

[32] IDMC, 2020, *Global Report on Internal Displacement*, p68, https://www.internal-displacement.org/global-report/grid2020/

[33] IDMC, *Briefing Paper: Number of IDPs by age at the end of 2019*, p2, see https://www.internal-displacement.org/sites/default/files/publications/documents/202004-age-disaggregated-IDP-data-paper.pdf

[34] UNHCR, *Global Trends*, p60: https://www.unhcr.org/5ee200e37.pdf

[35] WHO, *Disability and Health*, 16 January 2018, https://www.who.int/news-room/fact-sheets/detail/disability-and-health

[36] A/HRC/44/41, Report of the Special Rapporteur on the human rights of internally displaced persons, *Persons with disabilities in the context of internal displacement*, p5, https://undocs.org/en/A/HRC/44/41

[37] UNICEF, 2019, *INCLUDING EVERYONE: Strengthening the collection and use of data about persons with disabilities in humanitarian situations*, p17, https://reliefweb.int/sites/reliefweb.int/files/resources/Including%20everyone%20-%20Strengthening%20the%20collection%20and%20use%20of%20data%20about%20persons%20with%20disabilities%20in%20humanitarian%20situations.pdf

Oftentimes gaps in disaggregated data are due to the unit of analysis used in data collection systems. Broadly speaking, when data is collected at the individual level (or at household level when household rosters are used), data on the age and gender and special needs of affected populations can be easily collected. However, when it is collected at the group or community level it can be much more difficult and more unreliable. Social and political sensitivities also come into play concerning the collection of data on many of the criteria that fall into the "diversity" category, such as religious beliefs, ethnicity or sexual orientation.[38] These characteristics are often widely recognized as an important indicator for protection programming and to help secure solutions for affected populations but are often, simply, too sensitive to ask.

Challenges to making use of data

The best motivation for building the necessary momentum and focus to address data gaps and challenges is derived from the potential impact that results can have, which immediately raises the question of data usage: to what extent is the data produced actually used? Unfortunately, however, the connection between data production and data use in humanitarian and forced displacement settings, is not always linear and has much room for improvement.

Data usage can of course imply a broad range of activities undertaken by a variety of different stakeholders: from program design of civil society and international organizations, to funding allocations of international donors across and within crises, from effective coordination of humanitarian response within and across clusters or similar platforms, to data-rich research of the international research community, from targeting assistance to the most in need to evidence-based advocacy to impact policy dialogue and protection outcomes. In short, data on forced displacement has many potential uses and whilst it is difficult to summarize all the challenges currently experienced by those disparate stakeholders, there are a few issues that can be highlighted as particularly crippling: data literacy; data quality and coherence; and data accessibility.

Data literacy - the ability to 'speak data', to communicate with data and master a basic understanding of data sources and analytical methods - has not been a prevalent skillset amongst the humanitarian community. As with all other sectors, both public and private "data literacy has become important, for almost everyone. [Organizations] need more people with the ability to interpret data, to draw insights, and to ask the right questions in the first place"[39] and it is critical for data literacy to be present at all levels of decision-making if data is going to be used effectively. OCHA's Center for Humanitarian Data has recognized this challenge and has been working to address it through small but concrete steps in recent years.[40] UNHCR's new Data Transformation Strategy 2020-2025 has also put data literacy at its center "by making sure that our people have

[38] See here for fuller list of diversity categories: https://www.globalprotectioncluster.org/themes/age-gender-diversity/
[39] https://hbr.org/2020/02/boost-your-teams-data-literacy
[40] https://centre.humdata.org/wp-content/uploads/2020/02/data_literacy_factsheet.pdf

the knowledge and skills to use data responsibly and effectively" and investing in "general capacity-building among all managers on data interpretation and use."[41]

Data quality and coherence, or its notable absence in many contexts, also deserves a specific mention as a hindrance to effective data use. In recent years this has particularly been highlighted by the donor community where limitations in data quality and incoherent, conflicting results have hampered decision-making. This challenge is epitomized in the Grand Bargain commitments and the work that has resulted as part of its workstream to "Improve joint and impartial needs assessments" (for example, the Joint Intersectoral Analysis Framework (JIAF) that was recently endorsed by the IASC's Operational Policy and Advocacy Group (OPAG) and was rolled out in 2020 for all Humanitarian Needs Overview processes globally)[42]. Challenges resulting from incoherent data sources have also been regularly highlighted by UNHCR and other operational agencies when conflicting data and untransparent methods have complicated action. As one previous UNHCR Representative and Head of their Global Data Service put it: "When multiple actors bring multiple contradictory data or when there is no consensus, decision-makers will not act. Often the next step is to start all over again, with a collaborative approach."[43]

Data accessibility and data sharing is also a prevalent challenge, with both operational and policy-related negative impacts. Whilst there is consensus that streamlining safe data sharing procedures between operational partners and responsibly increasing the accessibility of data on forced displacement for researchers brings value in strengthening both policy and practice, the avenues for currently doing so are troubled. At the center of this challenge is a real protection concern about data that identifies vulnerable individuals or groups as there is a significant risk if it falls into the wrong hands that the subjects of that data can subsequently be targeted and harmed. Misuse of data by others also poses a clear reputational risk to agencies and organizations that collected and disseminated the data, especially when, as in the case of UNHCR, they have a mandate to protect the population who are the subject of that data. Beyond protection, unfortunately interagency rivalry and competition over funding creates further disincentives to accelerate data sharing. Due to the above, a lot of data currently collected by UNHCR and other stakeholders, through surveys, assessments and administrative systems or registries, often remains under-utilized in the design of policies and programs (for example when it cannot even be efficiently shared with operational and trusted partners) and for conducting research on forced displacement within the academic sector and beyond (for example when limited capacity/resources hinders the anonymization/de-identification of datasets). With a growing recognition of the potential use of existing data, many key stakeholders are taking action to develop tools and systems to share it more systematically and responsibly (see below).

[41] UNHCR, 2019, *DATA TRANSFORMATION STRATEGY 2020-2025: Supporting protection and solutions*, https://www.unhcr.org/5dc2e4734.pdf

[42] https://interagencystandingcommittee.org/improve-joint-and-impartial-needs-assessments

[43] Karl Steinacker, then Head of UNHCR's Global Data Service, speaking at a conference on Making Data Useful, organized by the Joint IDP Profiling Service in 2017, see report here: https://www.jips.org/uploads/2018/10/JIPS-Conference-2017-report-vf.pdf

These system-wide gaps together - data literacy, data quality and coherence, and data accessibility – do not fully explain the limitations in data usage, but they go some way to elaborating some fundamental issues that many stakeholders working in this space would quickly identify with.

KEY OPPORTUNITIES

Many of these gaps and challenges are significant and altogether can appear overwhelming, however in mid-2020, despite the surrounding global public health pandemic, the forced displacement data picture should be one of optimism. There is significant momentum and several important opportunities that taken together signal a seismic shift in improving data on forced displacement in the GCR era. Although space to act may have been limited by the pandemic, the crisis cannot undo the recent milestones and collective understanding of the importance of this agenda. Several of these opportunities are listed below.

Standardized definitions and national systems

Over the last few years, significant momentum has grown around the development of international statistical recommendations on both refugee and IDP statistics that focus on clarifying definitions and enabling countries to include forced displacement within their national statistical systems. The recommendations are endorsed by UN Statistics Commission[44] and thereby are primarily directed at member states and UN agencies, but of course have broader implications and potential use.

The International Recommendations on Refugee Statistics (IRRS) and the International Recommendations on IDP Statistics (IRIS),[45] endorsed in 2018 and 2020 respectively, provide for the first time a common international framework for statistics on refugees and IDPs, including clear definitions. For refugees, they outline a classification framework that clearly defines three categories: persons in need of international protection (including refugees), persons with a refugee background (e.g. naturalized/former refugees) and persons returned from abroad after seeking international protection (e.g. repatriating refugees and asylum seekers).[46] For IDPs, the recommendations outline the first internationally recognized internal displacement statistical

[44] "UN Statistical Commission, established in 1947, is the highest body of the global statistical system bringing together the Chief Statisticians from member states from around the world. It is the highest decision-making body for international statistical activities, responsible for setting of statistical standards and the development of concepts and methods, including their implementation at the national and international level." Read more here: https://unstats.un.org/unsd/statcom/

[45] https://ec.europa.eu/eurostat/web/expert-group-on-refugee-statistics

[46] Expert Group on Refugee and Internally Displaced Persons Statistics, 2018, *International Recommendations on Refugee Statistics (IRRS)*, p30: https://ec.europa.eu/eurostat/documents/3859598/9315869/KS-GQ-18-004-EN-N.pdf/d331c9cc-1091-43c2-b589-2c250bccc281

framework that builds on the UN Guiding Principles to define clear criteria for becoming an IDP (entering the national "stock"), delineates three sub-groups of IDP (those in locations of displacement, those in locations of return and those in other settlement locations) and outlines the criteria for exiting the national "stock" including elements of the statistical measure for determining the end of displacement.[47] This last point is of particular significance due to the policy and operational implications of a common definition of durable solutions that can be reflected in national and international data efforts which are often criticized for over-inflation and lack of coherence. Work is ongoing to support the further development/refinement of these recommendations.

These recommendations represent a major step forward not only because they have been officially endorsed by the Commission, but because they were developed through a collaborative process by the Expert Group on Refugee and IDP Statistics (EGRIS) that includes membership of around 45 countries - both refugee hosting states and those affected by internal displacement. Beyond definitions, they outline recommendations that aim to facilitate inclusion of forcibly displaced persons into national statistical systems, for example by focusing on data sources, basic/essential indicators for reporting, and coordination at national and international levels. Given the momentum built through the successes of this group a growing number of countries are already taking steps to include either refugees or IDPs into their national systems, including in places as diverse as Morocco, Kenya, Uganda, Colombia, Ethiopia, Mexico, Ukraine, Central African Republic and the Kurdistan Region of Iraq.

However, much work still needs to be done to advocate for and support more countries (and institutions) to incorporate the recommendations, with a particular need for capacity development in certain regions and countries. The EGRIS is currently working to keep up momentum and coordinate efforts in this direction through various advocacy and dissemination activities and facilitating peer-to-peer exchange of experience. [48] It is also working to develop standardized training materials and tools to support the work at country level and to further refine the IDP statistical recommendations concerning the measure for durable solutions and related concepts. Support to the Expert Group's continued efforts, whether financial, political, or operational, represents a key opportunity in advancing this agenda.

Inspired by the success of EGRIS, a group of statistical experts from national statistical institutions and international organizations (led by UNHCR and UNFPA), is currently working to develop statistical recommendations on stateless persons that will complement the existing ones on refugees and IDPs.[49] Like the others, these will provide clear definitions and recommendations

[47] Expert Group on Refugee and IDP Statistics (EGRIS), 2020, International Recommendations on Internally Displaced Persons Statistics, https://ec.europa.eu/eurostat/documents/3859598/12257846/KS-GQ-20-005-EN-N.pdf/714a7ba0-7ae6-1707-fef4-984a760e0034?t=1610984164036

[48] Terms of Reference for the Expert Group on Refugee and IDP Statistics (EGRIS): Third Phase/Implementation of Recommendations (2020-2024), https://unstats.un.org/unsd/statcom/51st-session/documents/BG-item-3n-terms-of-reference-for-EGRIS-E.pdf

[49] Petra Nahmias, *Better statistics to help end statelessness,* Jan 2020, https://www.unhcr.org/blogs/better-statistics-to-help-end-statelessness/

to enhance inclusion of these groups in national systems. Although politically statelessness is a much more sensitive topic, the goal is to have these technical recommendations adopted by the UN Statistical Commission in 2023 which will be an important milestone for significantly improving data collection at the national level over time and therefore a key achievement for the Global Action Plan to End Statelessness. Enabling the success of this process and, hopefully, supporting implementation of the recommendations once finalized will require careful coordination, advocacy and high-quality technical collaboration.

Enhancing socio-economic data and comparative analysis

There is a recognizable growth in the quality and quantity of socio-economic data on forcibly displaced persons and this trend needs sustainable and strategic support to ensure it brings value in the long run and does not remain a series of individual but ad hoc successes.

Results from recent efforts have produced evidence to inform development interventions. One recent example comes from the *Uganda Refugee and Host Communities 2018 Household Survey*[50], conducted by the Ugandan Bureau of Statistics in collaboration with the World Bank. This household survey was both nationally and regionally (West Nile and Southwest regions) representative and efforts were made to align its questionnaire to the national survey in order to facilitate comparison to official statistics. Its analysis revealed that around half of all refugees in Uganda (46%) are living in poverty with high levels of aid dependence (54% reporting aid as their main source of income). This poverty rate is much higher than for the host population which is 17%. It further revealed that 3 out of 4 refugees are unemployed nationally, compared to 36% of the host population. However, there were no major differences in access to basic services for refugees and hosts particularly in the West Nile and Southwest regions. It found that in some instances, refugees reported even more favorable access rates. For example, access to water through piping systems is higher for refugees at 94% than to 66% for hosts. The analysis of access to services demonstrated that investments benefitting host communities are needed to contribute to overall development and coexistence of both populations. Given the poverty and unemployment rates, there is a clear need to tap into refugees as a source of labor that could also help diversifying the economy and increase resilience to shocks.

The *2018 Kalobeyei Socioeconomic Profiling Survey*[51], conducted by the World Bank and UNHCR, is another example of high-quality socio-economic data on refugees that allows comparisons to host communities. Through alignment with the 2015/16 Kenya Integrated Household Budget

[50] World Bank, *Informing the Refugee Policy Response in Uganda: Results from the Uganda Refugee and Host Communities 2018 Household Survey*, http://documents1.worldbank.org/curated/en/571081569598919068/pdf/Informing-the-Refugee-Policy-Response-in-Uganda-Results-from-the-Uganda-Refugee-and-Host-Communities-2018-Household-Survey.pdf

[51] UNHCR & World Bank, *Understanding the Socioeconomic Conditions of Refugees in Kalobeyei, Kenya: Results from the 2018 Kalobeyei Socioeconomic Profiling Survey*, https://www.unhcr.org/ke/wp-content/uploads/sites/2/2020/04/Kalobeyei_Socioeconomic-Report-1.pdf

Survey (KIHBS), findings are also comparable to national official statistics. Findings similarly show high levels of poverty amongst refugees in Kalobeyei with 58% who are poor which is much higher than the national rate of 37% but lower than the rate in Turkana County that hosts the refugees studied. It also revealed low levels of employment for refugees with only 37% of working-age refugees in Kalobeyei as opposed to 72% of Kenyans nationally in employment. The survey also adds value to policy and programming decisions by showing that hosts and refugees of Turkana County were performing worse according to a range of socio-economic indicators compared to national averages. Insights from these types of surveys that cover both refugees and their host communities can therefore also help identifying where development and humanitarian action is needed the most and how they need to work in tandem to complement each other, even though there are often delays between the production of such analytical results and their impact on action taken.

There has similarly been a significant growth in socio-economic surveys in IDP contexts. A cross-country study by the World Bank on durable solutions for IDPs and refugees in Nigeria, Somalia, South Sudan, and Sudan[52], highlights that enhanced socio-economic microdata[53] on forced displacement can reveal important differences that would be over-looked if only aggregated data is used. The collaborative profiling exercise included in this work that took place in Darfur, Sudan for example demonstrates that poverty levels and unemployment rates are high amongst both IDPs and non-IDPs in the urban and peri-urban areas of El-Fashir, but that both rates are significantly higher for displaced communities.[54] A series of profiling studies in three regions of Kurdistan Region of Iraq (Erbil, Duhok and Sulaymaniyah) conducted under the leadership of regional Governorates and statistical authorities was able to provide a comparative analysis between IDPs, Syrian refugees and local populations in various geographic locations. Findings from these exercises covered critical issues such as housing and employment and helped shift humanitarian programming to include host communities and urban development plans to be inclusive of refugees and IDPs.[55]

---

[52] World Bank, *Informing Durable Solutions for Internal Displacement in Nigeria, Somalia, South Sudan, and Sudan*, https://openknowledge.worldbank.org/bitstream/handle/10986/32627/136740-A-overview.pdf?sequence=4&isAllowed=y

[53] Microdata is defined as unit-level information on a given population including individuals, households, or establishments (for example business enterprises or farms). See *OECD Glossary of Statistical Terms*: https://stats.oecd.org/glossary/detail.asp?ID=1656. In this context it refers to data at the level of individuals or households.

[54] Sudanese Government's Joint Mechanism for Durable Solutions and United Nations Country Team Sudan, 2019, *Progress Towards Durable Solutions in Abu Shouk and El Salam IDP Camps*, https://www.jips.org/uploads/2019/12/JIPS-Sudan-profilingreport-2019.pdf. See interactive Story map of results here: http://dswgsudan.org/en/2019-progress-towards-durable-solutions-abushouk-elsalam-idp-camps/#section=0&page=0&subpage=0

[55] Erbil Governorate, Kurdistan Region of Iraq, April 2016, *Displacement as challenge and opportunity Urban profile: Refugees, internally displaced persons and host community*, https://www.jips.org/uploads/2018/09/original_ErbilUrbanProfilingApril2016English.pdf

From these few examples the added value of high-quality socioeconomic microdata is quickly apparent, particularly to support the achievement of durable solutions for both refugees and IDPs.[56] The ability to compare between population groups – forcibly displaced and host communities – as well as with nationally representative official statistics (through alignment with well-established statistical standards for socio-economic indicators in some cases), also bring significant value to the policy relevance of these analyses from the perspective of national and regional policy development and program design.

Increasing availability of this type of socio-economic data on refugees and IDPs will also contribute to enhanced visibility of forced displacement within the framework of the Sustainable Development Goals and related dialogue at the national, regional and global levels as more countries will be able to include data disaggregated by forced displacement in their voluntary national reviews.[57] Although the inclusion of forced displacement in the global SDG indicator framework has been a politically sensitive and difficult one in recent years, the inclusion in March 2020 of a new indicator on refugees in the framework has been an important milestone.[58] It cements refugees clearly in the key international framework for development, which is also providing a guiding framework for investments in development data and capacity building globally – thereby opening up further opportunities. The indicator itself – "*The number of refugees by country of origin as a proportion of the national population of that country of origin"*[59] is only part of the story, this success opens the door for discussion with more countries on the inclusion of refugees and other forcibly displaced persons within the framework more broadly which links up to conversations around inclusion in national systems and enhanced availability of data on refugees and IDPs for relevant indicators across the framework. Combined with the proposal for a short list of prioritized SDG indicators that are most relevant for refugee and IDP inclusion,[60] it could also bring some momentum to dialogue between international custodian agencies of relevant SDG indicators to ensure indicator dissemination portals and mechanisms (such as the World Bank's SDG and World Development Indicator portals, or UNICEF's Data

---

Duhok Governorate, Kurdistan Region of Iraq, August 2016, *Displacement as challenge and opportunity Urban profile: Refugees, internally displaced persons and host community*, https://www.jips.org/uploads/2018/10/Profiling-report-KRI-Duhok-2016.pdf
Sulaymaniyah Governorate and Garmian Administration, Kurdistan Region of Iraq August 2016, *Displacement as challenge and opportunity Urban profile: Refugees, internally displaced persons and host community*, https://www.jips.org/uploads/2018/10/Profiling-report-KRI-Sulaymaniyah-2016-lr.pdf

[56] A particularly useful toolkit for this purpose was developed by the Joint IDP Profiling Service in collaboration with various partners. It presents a series of indicators and guidance for enabling durable solutions analysis in IDP contexts. See here for more information and to access the toolkit: https://inform-durablesolutions-idp.org/

[57] https://sustainabledevelopment.un.org/vnrs/

[58] Petra Nahmias and Natalia Krynsky Baal, Dec 2019, *Including forced displacement in the SDGs: a new refugee indicator,* https://sustainabledevelopment.un.org/vnrs/

[59] UNHCR, 2020, Global Trends, p26

[60] A proposal was developed by the Expert Group on Refugee and IDP Statistics and accepted by the Inter-Agency and Expert Group on SDG Indicators to be included in their recommendations to the UN Statistical Commission. The list of 12 priority indicators can be accessed here "Data Disaggregation and SDG Indicators: Policy Priorities and Current and Future Disaggregation Plans" https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Data-Disaggregation-E.pdf

Warehouse) facilitate visibility and accessibility of the indicator data disaggregated by forced displacement status as it becomes available.

Beyond the policy relevance of such data there are additional benefits that this work can contribute to for more efficient program implementation through targeting and the inclusion of forcibly displaced persons in social registries and administrative databases. Making use of representative survey data on refugees to inform the development of econometric targeting tools that aim to predict consumption based on a small selection of indicators, is an approach that can be used to pursue the integration of refugees in national social protection systems. In contexts where these systems are in the process of being built – which is for example the case in many Sub-Saharan African countries such as DRC and Chad – there is an opportunity to explore this additional use of recently produced survey data for this purpose. Although it is a methodology not without questions due to quality concerns with the original data, exploratory work on this has been completed in Chad[61] and is currently being explored in several other countries.

Investing in more socio-economic microdata on forced displacement is of critical importance, to advance the implementation of the GCR through the inclusion of refugees in development assistance and national systems. However, to maximize the added value of it, simply producing more data is not enough. Investments should focus on aligning new data with well-established international standards for socio-economic indicators and statistics which will facilitate comparison with official statistics within and across countries; it should as far as possible capitalize on opportunities to integrate forced displacement into national statistical systems and be conscious of capacity development needs for more sustainable impact; and it should additionally forge linkages with the SDG framework and associated reporting mechanisms to increase visibility of forced displacement in development policy dialogue and create opportunities for peer-to-peer learning between affected countries.

Supporting a common approach to measuring impact

As mentioned above, a process for developing a common methodology or methodologies for measuring the impact of hosting, protecting and assisting refugees is well underway. Although some delays have incurred due to political attention diverted by the COVID-19 pandemic during 2020, interest amongst Member States continues and the commitment for UNHCR, who is coordinating this effort, to report on progress at regular intervals is still in place.

Due to the data-related challenges summarized above, participants in the process agreed to focus on simple methodologies to quantify the costs incurred in specific sectors from hosting refugee populations; furthermore they prioritized looking into fiscal costs in the education sector as an

[61] Theresa Beltramo, Hai-Anh H. Dang, Ibrahima Sarr and Paolo Verme, April 2020, *Estimating Poverty among Refugee Populations A Cross-Survey Imputation Exercise for Chad*, World Bank Group Policy Research Working Paper no. 9222, http://documents1.worldbank.org/curated/en/511711588016782589/pdf/Estimating-Poverty-among-Refugee-Populations-A-Cross-Survey-Imputation-Exercise-for-Chad.pdf

initial step and, given the impact of the COVID-19 pandemic interest in tackling the health sector next is growing. At this juncture, it is important for stakeholders involved to take steps to keep up momentum despite the challenges and continue to pursue a transparent and participatory process that builds trust around the more technical parts of the work. It will also be important to focus on regional and country-level dialogue to complement the global engagement pursued so far, and make technical support available for countries interested in taking the work forward within their own contexts and - ideally – showcasing results so that others can follow. Further down the road, similar efforts should be pursued for internal displacement contexts which would raise some similar and divergent political and technical challenges.

More disaggregated data

Through investment in microdata (as elaborated in the above section on socioeconomic data) more disaggregated data on sub-groups of those affected by forced displacement can be made available. This will help to fill critical data gaps, outlined above, on age and gender disaggregation for refugee, IDP and stateless populations but should also focus on context-relevant priority diversity criteria as much as possible.

Calls for more disaggregated data are of course not new. The Inter-agency Standing Committee's (IASC) Gender Handbook from 2006 called for it: "Data on the population affected by the crisis should always be broken down by age and sex and other relevant factors such as ethnicity or religion."[62] And the often cited "Sex and Age Matter" report from 2011 presented solid analysis of why disaggregated data is not collected more systematically (citing 12 distinct reasons including those related to capacity, willingness/interest and available, harmonized tools) and outlined a series of key recommendations for strengthening its collection at different stages of a crisis.[63] Many of the key messages of this report, despite being almost a decade old, still hold true. A further nuance, however, to modernize and complicate the call for more disaggregated data comes through the lens of data ethics. Using this lens, what data is collected, how it is managed, analyzed used or shared should be guided by an ethical compass including questions on the potential harmfulness of the activity and its overall cost-effectiveness.[64] In other words, as the Protection Information Management principles – a multi-stakeholder initiative that has developed a framework for enhancing the ability of humanitarian information management systems to advance protection outcomes for crisis affected populations – has put it, actions should be "proportional to both the identified risk and costs vis-à-vis the expected response"[65]

---

[62] Inter-agency Standing Committee (IASC), December 2006, *Gender Handbook in Humanitarian Action: Women, Girls, Boys and Men Different Needs – Equal Opportunities*, p8 https://www.globalprotectioncluster.org/_assets/files/tools_and_guidance/IASC_Gender_Handbook_EN.pdf

[63] OCHA, Feinstein, Tufts, Care International, 2011, Dyan Mazurana, Prisca Benelli, Huma Gupta and Peter Walker, *Sex and Age Matter*, see page 80-81 for the 12 reasons and https://www.care.at/wp-content/uploads/2016/05/Sex_and_Age_Disag_Data.Feinstein.Final_Report_1_.pdf

[64] OCHA Centre for Humanitarian Data, Jan 2020, *Guidance Note Series – Data Responsibility in Humanitarian Action Note #4: Humanitarian Data Ethics*, https://centre.humdata.org/wp-content/uploads/2020/02/guidance_note_ethics.pdf

[65] http://pim.guide/guidance-and-products/product/principles-protection-information-management-may-2015/

Another opportunity to generate more disaggregated data in a responsible manner could be to develop careful sampling strategies to generate overall statistics from microdata of representative refugee/IDP samples.[66] The advantage of careful sampling is that a smaller but meaningful subpopulation of forcibly displaced persons could generate more precise knowledge about the overall population without needing to conduct a complete census (therefore increasing cost-effectiveness) or rely on estimates based on applying national averages (which themselves might be significantly different). Adequate sampling is certainly a challenge due to the quality and coverage of population data on forcibly displaced persons in many contexts, but the increasing availability of other data sources – such as geospatial data or administrative data – could help to shape appropriate sampling strategies by identifying displaced persons and/or where they are likely to reside. Government or UNHCR registration data, IOM Displacement Tracking Matrix data or sampling frames from national censuses/household surveys – depending on what is available and most up to date in a given context, could facilitate this process even further.

Increasing data use

Avenues to improve the quantity and quality of data on forced displacement only gets us over the half-way line; there is still important work that needs to be done to ensure the data gets into the hands of the people who need it. As described above key issues include ensuring safe and responsible accessibility of data, minimizing the damaging impacts of conflicting data sources, and making sure that relevant decision-makers have good enough data literacy skills that they know what to do with the data when they get it.

Institutional transformations with respect to data management in many organizations working on forced displacement is a critical piece of this puzzle. Many relevant organizations have initiated such processes that include organizational restructuring, recruiting new talent, standardizing systems and other significant investments that aim to enhance institutional data capacity and – in turn – enable more impactful data use. UNHCR's Data Transformation Strategy for 2020-2025 is one such example, but similar efforts in IOM, OCHA and various international NGOs[67] should also be recognized as should those whose main focus is not migration/forced displacement but who are making efforts to increase their focus on these populations.[68]

[66] Some recently released sampling guidance for use in internal displacement contexts: https://www.jips.org/uploads/2020/05/JIPS-SamplingGuideForDisplacementSituations-June2020.pdf

[67] One example comes from the Norwegian Refugee Council that lists "Become a leader in using data and technology to deliver better" as one its four main ambitions: https://www.nrc.no/globalassets/pdf/policy-documents/programme-policy/nrc-global-strategy-2018-2020_web.pdf

[68] On example comes from UNFPA's Strategy for the 2020 Round of Population & Housing Censuses (2015-2024) which includes a commitment (p27) to "enhance the use of census data to generate estimates on refugees and displaced persons, consistent with UN International Recommendations on Refugee Statistics and the UN/Eurostat Technical Report on Statistics of Internally Displaced Persons." https://www.unfpa.org/sites/default/files/pub-pdf/Census_Strategy_Final_July.pdf

Data literacy to enhance data use often is – and should be - a central feature of these institutional reform processes. This can include basic data literacy skills of staff members at all levels; the ability to read bar charts and line graphs, an understanding of main data collection methods or sources used and their limitations, or the ability to identify key data gaps would all fall into this category. Also significant for enhancing the use and therefore potential impact of data, is core data literacy of senior managers so that when data is made available it can directly inform decision-making processes and in turn data becomes valued and adequate resources are allocated. This directly relates to addressing challenges around the quality and coherence of data, as without senior management support they are unlikely to be addressed in an effective and systematic way.

Enabling the use of data beyond immediate institutional needs by investing in responsible and safe data sharing and accessibility so that others – operational and research partners[69] – can also effectively make use of the data produced/managed internally can further increase its impact. Safely increasing data accessibility can foster innovation and inquiry that can in turn strengthen policy and improve protection and operational outcomes for affected populations. It can foster coordination between governments and organizations working to improve the lives of affected people through trust building, reduced duplication of efforts and evidence-based dialogue. Expanding our horizon further, it can also lead overtime to an improved and evidence-based public understanding of forced displacement.

Increasing availability and access to microdata on forcibly displaced persons, as described above, must be accompanied by clear data protection standards that proactively minimize the risk of potential harm to the data subjects. Enabling progress in this direction will require the development and implementation of an agreed upon framework recognized by various key stakeholders that will provide adequate guidance and tools for the responsible and systematic dissemination of microdata. Given the wealth of data that UNHCR holds on forced displacement and its solid protection mandate, leadership from UNHCR in this regard will be critical. Institutional transformations by other data holders will also be necessary to better embrace and adopt new tools and systems.

CONCLUSION

This paper has summarised a series of critical gaps that the author has identified as relevant for a better understanding of the current data landscape on forced displacement. It has also elaborated on a series of key opportunities, in response to each of these challenges, that will help to inform future efforts to improve the ability of data to be used for a positive impact on the lives of affected populations. The Global Compact on Refugees has been used as an initial starting point for the author's analysis and reflections.

[69] See SOWFD background paper on Forced Displacement Academic Research Trends

Many of the points raised in the paper are incorporated into the plans of the recently established World Bank-UNHCR Joint Data Center on Forced Displacement.[70] The JDC aims to work with partners to make significant headway over the next few years to transform the forced displacement data landscape with a primary focus on enhancing socio-economic data and analysis on refugees, IDPs, stateless persons and their hosts. Through supporting the development and implementation of statistical standards, methods and tools, producing high-quality data and analysis integrated as far as possible within national statistical systems, enhancing safe and responsible access to microdata and injecting momentum into the forced displacement data-driven research community, the JDC will bring together the capacities of the World Bank and UNHCR to make a significant difference. These four priorities, each one integrated into the paper's analysis, form the basis of the JDC's strategy for 2021-2023. As interlinked and interdependent opportunities, they will be pursued simultaneously and in collaboration with affected Governments, communities, and other key stakeholders.

[70] https://www.jointdatacenter.org/