PERSPECTIVES FROM IOM-DTM ON IDP DATA

Recommendations for the consideration of the High-Level Panel on Internal Displacement

**Overview**

In recognition of the High-Level Panel on Internal Displacement’s critical mandate, the International Organization for Migration (IOM), the largest global collector of IDP data, would like to provide technical recommendations for consideration on IDP data and evidence. These recommendations are supported by an examination of the current landscape of the IDP data ecosystem, assessing both global and country-specific trends in IDP data as they relate to institutional structures, efficiency, and use. These recommendations summarize perspectives from IOM’s Displacement Tracking Matrix (DTM), a global network of 438 professional and specialized staff who provided the majority of the essential IDP data of 24 Humanitarian Needs Overviews (HNOs) and Humanitarian Response Plans (HRPs) in 2019.[[1]](#footnote-2) With a global footprint in the field of internal displacement data, DTM is unique in scale, volume, and consistency. In 2019, it regularly monitored the locations, needs, and conditions of 27.8 million IDPs and 20.1 million returnees that were formerly displaced, affording the humanitarian community the ability to ground policy and operational planning in up-to-date data. Data generated by DTM significantly contributes to global IDP statistics. These recommendations are thus based upon observations provided by the largest global network directly involved in coordination, management, collection, storage, analysis, and dissemination of IDP data.

1. ***IDP data needs to be valued as one of the common operational foundations for humanitarian response****.* A displacement geo-referenced data baseline that is fully interoperable with other common humanitarian datasets needs to be at the foundation of any IDP or humanitarian data efforts, in-country or globally. At present, country-level operations’ data production is confusingly dispersed across a wide range of actors and processes, often in an uncoordinated fashion. IDP and humanitarian data should instead have consistent and focused strategies and coverage, including planned redundancies where there is risk of humanitarian lapses in data collection due to access constraints.
2. ***Global coordination efforts on IDP data must be rebalanced to prioritize country-level efficiency for collection and use of data alongside transnational coordination***. In contexts where data projects are dispersed across agencies, there is an unbalanced focus on generating evidence to support internationally comparable data according to standards and policies that are not always aligned with country-level operational priorities and affected population needs. While multilateral efforts in recent years have positively contributed to expanded guidance and recommendations for standardization and compatibility in IDP statistics for improved international records, these initiatives are only partially relevant for on-the-ground humanitarian needs, and may in some cases contrast with the priorities of IDPs themselves. With increased access to data collected directly from affected populations, a rebalancing in global coordination efforts on IDP data that prioritizes ground-level humanitarian information needs is required from an accountability perspective, particularly since accountability to affected populations for the use of data collected by humanitarian actors varies in comparison to national statistics which are collected for other reasons.
3. ***IDP data needs to be repositioned in current humanitarian responses through new coordination mechanisms.*** The importance and complexity of humanitarian data has increased exponentially since the cluster system was designed in 2005, the cluster reforms of 2010, and the progress made under the commitments of the Grand Bargain in 2018. Therefore, there needs to be data coordination positioned within the coordination architecture (beyond information management and assessments) to promote a coherent and inclusive country-level data strategy, cooperation, and systems that would allow the humanitarian community to optimize use of capacity, and to fully utilize the information available in collected datasets. Country-level data coordination is also essential to ensure links between humanitarian and development, transition or other IDP data-related processes.
4. ***Approaches to IDP data through official statistics should continue to be complemented with global monitoring of internal displacement that utilizes both formal and informal sources.*** The role of initiatives related to IDP data for national statistical purposes should be reinforced with global independent monitoring mechanisms, such as the yearly estimates currently compiled by the Internal Displacement Monitoring Center (IDMC)*.* Independent global monitoring of internal displacement remains an important tool for overseeing global trends, for complementing national data, and for ensuring impartiality and neutrality in complex humanitarian environments.
5. ***Humanitarian and development stakeholders should commit to make better use of the IDP data already being gathered, which is only possible with investment on data science and data ethics capacities.*** Such an expansion would require support to ensure robust, ethical approaches are adopted and human rights are adhered to in practices such as predictive forecasting, big data displacement monitoring, use of artificial intelligence, deep analysis of prevalence of needs, algorithmic models for prioritization, and beyond.[[2]](#endnote-2) Investing in the responsible use of data science is necessary to ensure accountability for using information already provided by IDPs. Country-level data strategies are essential to guide analysis conducted by specialized data scientists.
6. ***IDP data security should be held to the highest standards, including auditing the security practices of those entrusted with IDP data.*** In proportion to the level of risk posed by different IDP datasets, practitioners should be held to gradated security standards. Inter-agency mechanisms such as the IASC Operational Guidance on Data Responsibility that relate to overall IDP security should be instituted. Responsible collection and sharing of IDP data require the definition of minimum data security capabilities and standards. This recommendation may resonate overly technical. It intends to support advocacy from ICT system experts to address data security alongside other high-level policy priorities. Addressing data security issues among other high-level priorities is essential to mitigate one of the fastest growing risks to humanitarianism.
7. ***The sensitivity of IDP data must be recognized as an area of extraordinary risk, and treated accordingly***. Amid the growth of new data science methods (recommendation 5), greater care and attention must be paid to data responsibility through an adherence to ethical standards, data protection, data security, and data governance practices in increasingly complex environments. Initiatives aimed at improving IDP data and humanitarian data have so far not been complemented with commensurate advocacy and investment in technical areas necessary to ensure responsible growth. At global level, there is the need to establish coordination mechanisms for ethics and data science, to support interaction between ethics experts and data science staff.
8. **IDP data needs alignment and harmonization to support a meaningful country-level Humanitarian Development Peace Nexus (HDPN)**. IDP data provides a mechanism to bring together multiple stakeholders. To enable decision-makers across the HDPN to define strategies and measure impact from a nexus perspective, there needs to be the development of a set of nexus indicators which are simultaneously representative of humanitarian priorities and of peace and development outcomes. Such initiative should consider the different modalities of delivery and intervention from stakeholders involved in the nexus.

**Introduction**

In recent years ‘data’ has been ascendant in virtually all facets of life. Facilitated by the rapid expansion of internet technologies, this era has been heralded the “Data Age,”[[3]](#endnote-3) in which the amount of data is growing exponentially[[4]](#endnote-4) and touted as “the new oil.”[[5]](#endnote-5) Initially directed towards education policy, “data-driven decision making” (DDD) has become a standard for several forms of policy implementation around the world.[[6]](#endnote-6) Simultaneously, a backlash to this data wave has also arisen, initiating greater consideration towards the risks and limitations of data collection, usage, and impact. The humanitarian sector has seen a parallel trend towards incorporating data for better policy and decision-making, alongside increasing risks and ethical considerations around data initiatives.

It is in this broader context that this paper considers the future of IDP data. For purposes of this report, **IDP data is considered any information that may be useful to respond to IDP needs**. At the moment, the needs of IDPs globally are particularly acute,[[7]](#endnote-7) and the principles of data science are becoming more established for humanitarian actors.[[8]](#endnote-8) Moreover, the recent proliferation of IDP data has transformed the needs of the data ecosystem beyond its traditional scope, previously understood to be within the remit of ‘information management.’ Instead IDP data should be reassessed in the broader context of humanitarian data, development data, protection data, and data in the service of the Sustainable Development Goals (SDGs). These changes necessitate new and innovative ways to not only collect IDP data, but also strategize, analyze, and process it. Indeed, we are currently at a critical juncture with an unprecedented IDP caseload, a rapid expansion of IDP data, and a profusion of methods to analyze that data.

IDP data contributes to an evidence base to understand the nature and extent of IDP needs for humanitarian actors and/or governmental authorities to ensure life-saving assistance and protection to those most vulnerable. There are a variety of approaches to IDP data which tend to have overlapping aims, as do humanitarian, governmental, and development stakeholders, even if their specific goals and emphases may differ. For example, “IDP operational data”[[9]](#footnote-3) is catered towards better efficacy in targeting humanitarian responses, whereas IDP data in the form of “IDP statistics” is particularly useful for advocacy and national planning purposes. Even within IDP statistical data and IDP operational data, each country may take a different approach catered to their specific circumstances, generating significant challenges in data interoperability and comparability across contexts.

What all these diverse approaches have in common is a shared reality of **the centrality of IDP data as a foundation for humanitarian assistance**. IDP data is often central to decision-making in humanitarian interventions and must be understood as such to allow for humanitarian responses to operate at their most optimal. To this end, it is of paramount importance that humanitarian crises in particular have a baseline of IDP data to provide situational awareness at the outset of a crisis when it is most needed, in a form that is interoperable with other common humanitarian datasets.

Much of the growth in new methods of collecting and analyzing data is predicated on indirect or proxy collection of data, such as through cellphone tower data, social media, and other indirect means. While there certainly are advantages for such indirect data collection, it is no replacement to the intrinsic value of collecting data in proximity with IDPs themselves. Moreover, data-related solutions tend to be most effective when they emerge from on-the-ground actors. Additionally, big data approaches can disadvantage those who lack access to digital technologies and thus become ‘illegible’ to indirect big data collection approaches.[[10]](#endnote-9)

1. **Traditional IDP Data Structures**

Given the rapidity with which the data ecosystem has been developing, institutional structures and processes have often been slow to adapt to the increasing volume and importance of data. Volume aside, the speed of IDP data consumption is incongruent with the speed of data production, and many types of data analysis are still lacking, hampered by a lack of analysis capacity. Thus, broadly speaking, in the case of IDP data, institutional structures are still catching up in terms of efficiency, new organizational design, and interoperability. The work of coordinative groups like JIAG, IMWG, and EGRIS[[11]](#footnote-4) represent significant steps in the right direction, but much remains to be done, particularly in the humanitarian field and its application of IDP data, or “operational data.” **Indeed, in comparison to national statistical IDP data, operational data has received relatively little support—a misalignment that must be recalibrated** (see elaboration of statistical IDP data and operational data below).

Country-specific differences in IDP data collection and analysis can make it difficult to harness the full potential of the data collected. On the one hand, country-by-country differences simply make transnational analysis difficult to execute. On the other hand, siloed country approaches, varied coordination structures, and ad-hoc crisis management approaches make it difficult to share best practices and to utilize data in the most effective ways possible. Additionally, the current humanitarian cluster system (discussed shortly) was designed in 2005, before data became so integral to humanitarian processes.

1. *IDP Operational Data*

IDP operational data for humanitarian purposes is directed towards discerning immediate, actionable needs of displaced populations for humanitarian actors and sometimes governments from the onset of a crisis, and is generally coordinated through disaster-response agencies within governments or multilateral institutions within the Inter-Agency Standing Committee (IASC) framework. Moreover, IDP data is ever evolving, in contexts were mobility is suddenly impacted by a multiplicity of events. As an important element of a humanitarian response, the mechanisms that inform IDP data collection and analysis are usually organized according to the IASC cluster system. Unlike IDP statistical data, however, there exist relatively fewer standardization and coordination bodies concerned with the collection and analysis of operational data. Additionally, because operational data are needed so quickly in dynamic circumstances, there exist a wider range of approaches to generating and using IDP operational data than statistical data, particularly with the expansion of ‘big data’ techniques and increasing smartphone adoption.

Current structures in place are predominantly governed by the cluster system established in 2005, before data had the central role that it does today in humanitarian responses. Within the cluster system, IDP data is collected and analyzed by different clusters and NGOs according to their needs. Owing to the work of OCHA, information management working groups (IMWGs) are established within the cluster system (where funding permits) to coordinate, share, and harmonize the data between clusters. However, pending a variety of factors, such Information Management Working Groups may or may not be established, and may operate with varying degrees of efficiency and tend to be primarily focused on reporting of services of cluster members rather than undertaking data collection that is aimed at focused on inter-sectoral analysis of needs. Otherwise, the primary relevant entity currently within the cluster system are information management officers (IMOs) who are chiefly focused on tasks associated with a cluster and broader coordination and programme information needs. In part because of these relatively simple structures, the amount of collected data that is being used remains minimal.

As such, a question emerges as to whether there is sufficient space in the current humanitarian architecture to account for the increased importance and complexity of IDP and humanitarian data. Current data structures are not adequately efficient in addressing needs or prioritizing data-driven decision-making. At present, data management seems to be driven by multiple streams of humanitarian programming through the cluster system, development programming, or triggered by globally led efforts of harmonization rather than built as a country-level data ecosystem that is strategically conceptualized. In this way, IDP data is siloed, underutilized in response planning, and typically not deployed in a systematic manner to garner a holistic view of a displacement crisis. At times, these limitations can result in an inaccurate or distorted understanding of displacement.

Moreover, IDP is central to longer-term HDPN processes and concerns. Addressing country-level strategies to align IDP data relevant for humanitarian, transitional and development programming is paramount to measuring contributions to the nexus. Use of initiatives led by humanitarians, such as to support building socio-economic profiles that are useful for longer-term planning, or Stability Indexes[[12]](#footnote-5) should be evaluated by HDPN actors to support defining best practice. Drawing on the IASC Light Guidance for Collective Outcomes, there is potential for using existing IDP data to conceptualise a joint framework for the analysis of collective outcomes and answer the question of “who are those that are left behind.”[[13]](#endnote-10) Existing IDP baseline data (Recommendation 1) on gender, age, disability, social status can not only help contribute to providing a holistic overview of displacement, but also better identify the location, causes, risks and consequences of ‘those left behind.’[[14]](#endnote-11) In this vein, IDP data can act as a link between humanitarian, development, and peace actors to solidify and institutionalize greater synergy between operational data and to address HDPN priorities, while also enabling the measurement of collective outcomes and SDGs.

**As such, to account for the increased importance and complexity of IDP data in humanitarian responses, this paper proposes the need to revaluate the importance of IDP data, with the establishment of global and country-level data strategies, through structures similar to those of a humanitarian cluster. Issues such as data ethics, solutions, baselines, data science, accountability for use of data, data security, monitoring, and international standards are only a few examples of the complex array of the currently uncoordinated issues affecting IDP data. Failing to address this gap may result in ineffective humanitarian responses, and insufficient resource mobilization.**

*IDP Statistical Data*

IDP statistical data is used for the primary purpose of gaining insight into the overall number and demographics of IDPs within a specific displacement context, as well as related humanitarian needs and vulnerabilities. Official IDP statistics are typically published by national statistics offices (NSOs) to better inform longer-term policy, to more rigorously define IDP populations, and to assess more long-term, detailed vulnerabilities for IDPs and measure progress. The United Nations Statistical Committee (UNSC) ordinarily plays a guidance and collaboration role in IDP statistics practices between countries, as do the Inter-Agency Standing Committee (IASC)’s frameworks and norms.

That said, to realize more effective, comparable IDP statistical data, the data ecosystem is also reliant on operational data and the successful bridging of the gap between operational and statistical data. Indeed, the ambitions of EGRIS cannot be achieved without the continued foundational data provided by operational actors. In cases where governments lack the capacity or willingness to collect IDP statistical data, the gap in data is currently largely filled by humanitarian and development agencies, whose operational planning and responses require different types of information beyond statistics.  For these reasons, supporting mechanisms to better transition operational data into statistical data is of key importance in the IDP data ecosystem.

While efforts should be strengthened in building national capacities, including through National Statistical Offices and National Disaster Management Agencies, stakeholders such as IDMC also play a crucial role in monitoring global IDP trends, adding a neutral and complementary perspective that is politically independent. **It is of paramount importance for the humanitarian community to ensure that objective and neutral monitoring capacity is maintained and strengthened.**

1. **Emerging trends in IDP data**

The most significant emerging trend in the last 8 years of IDP data has been the increasing importance and necessity for data-driven coordination and evidence-driven programming by humanitarian agencies, donors, and governments. There has been a similar proliferation of data actors over the same period, further contributing to an expansion in the overall volume of IDP data.

These trends are likely only to accelerate as the rise of mobile technology, data science, and artificial intelligence (AI) lead to a profusion of approaches and tools in dealing with IDP data. Several organizations, corporations, and agencies are pioneering new methods of gathering and processing IDP data for a wide array of initiatives, from automated web trawling to better collate global media related to IDPs,[[15]](#endnote-12) to using cellular triangulation and satellite imagery to discern IDP mobility flows, and beyond.

The COVID-19 pandemic has accelerated the push towards alternative methods to collect and analyze IDP data through the widespread development of digital contact tracing systems. Though digital contact tracing systems are not IDP-specific, they represent a form of digital mobility tracking that, in some forms, approximates technologies that have been used (or proposed for use) to generate data on IDP mobility. Given the extraordinary circumstances of the pandemic, there is an increased openness to more intrusive digital approaches to data collection, with the potential to change norms around data collection of populations around the world.[[16]](#endnote-13)

More specific to IDPs, travel restriction measures and internal mobility restrictions have adversely affected access to camps and camp-like settings, and the impact of the COVID-19 pandemic on pre-existent mechanisms of IDP data collection and monitoring modalities. Proxy methods of remote information gathering through phone calls and online systems has in some cases become essential to the continued administration of IDP camps and settlements.

The most pressing issue in IDP data efficiency is not a lack of data but the absence of a coherent data strategy, shared vision, analysis, sub-sectors, and interoperability. This poses primary obstacles to the optimum and efficient use of IDP data. In many cases, the data for a wide range of measurements and data science activities already exist. In this sense, the priority for IDP data effectiveness and humanitarian accountability may be the development of a set of common norms to drive the better use of data for humanitarian delivery. **Data can be better used by the strategic application of data science, as well as by increasing data literacy among decision makers. Indeed, the use of data and analysis in decision-making is proportional to the level of data literacy.**

*IDP Data Ethics*

Commensurate with the increasing power of data science methods is a parallel increase in ethical concerns. IOM has taken a leading role in the Data Science and Ethics Group’s (DSEG’s) development of a “Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector,” which lays out a practical guide to developing data science activities in an ethically responsible manner.[[17]](#endnote-14) As in all matters of humanitarian data science, activities related to IDP data involve issues of privacy, fairness, accountability, data weaponization, etc., whilst abiding by the humanitarian principles of humanity, neutrality, impartiality, and independence. As these concerns and principles relate to the emerging transformations in IDP data, two central issues should underpin the continuing expansion of AI data science tools in the humanitarian sector:

* **Maintaining a needs-based approach**: With such a profusion of data-related tools and practices, there is a noted tendency for humanitarian actors to lose sight of a needs-based approach to IDP data. While the point may seem obvious, humanitarian actors are often approached with new, alluring technologies (or funding to use new technologies), even if said technologies are not the most appropriate way to address an issue on the ground. Practitioners in the field must be cognizant that neither technological capability nor budgetary incentives should drive new IDP data initiatives. Instead, IDP data should be collected and assessed only insofar as it clearly addresses IDP needs.[[18]](#endnote-15)
* **Data Minimization**: Even whilst the importance and needs of IDP data are increasing, collecting more data is not always the best solution, and can often prove counterproductive. Those in the IDP data field are reminded to practice the principle of data minimization, that is, only collecting a minimum amount of data necessary for specific purposes, to avoid unintended harm or privacy violations.[[19]](#endnote-16)

Given the expanding applications of data science to the humanitarian sector, new challenges are likely to emerge even if these concerns are carefully considered. As such, continued support to inter agency led humanitarian data science ethics groups like Data Science and Ethics Group (DSEG) is of paramount importance to adapt to the changing terrain of humanitarian data science.

**In the case of IDP data, the priority of strong ethical norms in data governance, protection, and usage are of crucial importance**. Given the unique vulnerabilities of IDPs, IDP data is intrinsically sensitive, and could pose a host of risks to IDPs’ wellbeing if mismanaged. As malicious actors become more adept at data science activities, handling IDP data with extreme care is even more important. Additionally, achieving a high degree of data security for IDP data is an additional priority in the age of big data and data security breaches. **To ensure the safety of IDP data, datasets should be categorized according to their level of sensitivity and granted stringent security measures in proportion to that sensitivity.**

1. DTM Annual Summary Report 2019: Total population tracked by DTM globally 75.4 million; Total IDPs 27.8 million; Total returnees 20.1 million; Total other categories 27.5 million (includes residents and migrants); Countries with DTM operations 78, of which 31 focus on internal displacement. Staff structure includes 438 technical experts, 166,379 key informants, and 6,170 data collectors. There were 622,432 downloads of 1,620 reports and 200,738 downloads of 687 datasets produced and shared publicly by DTM during 2019. [↑](#footnote-ref-2)
2. See <https://displacement.iom.int/reports/workshop-report-forecasting-human-mobility-contexts-crises?close=true>. [↑](#endnote-ref-2)
3. John Paulsen, “Enormous Growth in Data Is Coming — How to Prepare for It, and Prosper From It,” *Seagate Blog* (blog), April 4, 2017, https://blog.seagate.com/business/enormous-growth-in-data-is-coming-how-to-prepare-for-it-and-prosper-from-it/. [↑](#endnote-ref-3)
4. Paulsen. [↑](#endnote-ref-4)
5. Joris Toonders, “Data Is the New Oil of the Digital Economy,” *Wired*, July 23, 2014, https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/. [↑](#endnote-ref-5)
6. Erik Brynjolfsson and Kristina McElheran, “The Rapid Adoption of Data-Driven Decision-Making,” *American Economic Review* 106, no. 5 (May 2016): 133–39, https://doi.org/10.1257/aer.p20161016. [↑](#endnote-ref-6)
7. Federica Mogherini and Donald Kaberuka, “COVID-19 Makes Addressing Internal Displacement Even More Urgent,” *Al Jazeera*, May 26, 2020, https://www.aljazeera.com/indepth/opinion/covid-19-addressing-internal-displacement-urgent-200525080109674.html. [↑](#endnote-ref-7)
8. Data Science and Ethics Group, “A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector,” Humanitarian Data Sc, accessed April 28, 2020, https://www.hum-dseg.org/ai-applied-ethics-toolkit. [↑](#endnote-ref-8)
9. See below for definition [↑](#footnote-ref-3)
10. Data Science and Ethics Group, 21–22. [↑](#endnote-ref-9)
11. The Joint Inter-sectoral Analysis Group, Information Management Working Group, and the Expert Group on Refugee and IDP Statistics. [↑](#footnote-ref-4)
12. For an example of the Stability Index in Cameroon, See [DTM Cameroon’s Stability Index Report](https://dtm.iom.int/reports/cameroon-%E2%80%93-stability-index-dashboard-1-12%E2%80%9425-august-2019) [↑](#footnote-ref-5)
13. IASC Light Guidance on Collective Outcomes, May 2020, 7. [↑](#endnote-ref-10)
14. IASC Light Guidance on Collective Outcomes, May 2020, 7. [↑](#endnote-ref-11)
15. See, for instance, dataminr, at dataminr.com. [↑](#endnote-ref-12)
16. Liza Lin and Timothy Martin, “How Coronavirus Is Eroding Privacy,” *Wall Street Journal*, April 15, 2020, sec. World, https://www.wsj.com/articles/coronavirus-paves-way-for-new-age-of-digital-surveillance-11586963028. [↑](#endnote-ref-13)
17. Data Science and Ethics Group, “A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector.” [↑](#endnote-ref-14)
18. Data Science and Ethics Group, 20. [↑](#endnote-ref-15)
19. Data Science and Ethics Group, 27. [↑](#endnote-ref-16)