NYU | CIC | CENTER ON INTERNATIONAL COOPERATION

DATA FOR PEACEBUILDING AND PREVENTION

Towards a Prevention and Peacebuilding Data Hub

Scoping the Future of Data Services and Capacity Building

Branka Panic and Paige Arthur

Table of Contents

Introduction

The complex crises in our world—from rising instability linked to pandemic effects and climate change, to ongoing challenges of civil war, urban violence, violent extremism—require complex analysis and insights. Emerging technologies, data, and data science methods have been recognized as potential tools to help tackle some of these "wicked" problems across the humanitarian-development-peace nexus. Artificial intelligence (AI) and machine learning (ML) potentially open a range of new opportunities for early warning and humanitarian preparedness, assessment and monitoring, service delivery, and operational and organizational efficiency. Advanced data analysis is the core of these efforts, and its success often depends on timely access, as well as the right quality and quantity of data.

Those wishing to develop data-driven insights, however, would benefit from consolidated access to relevant data sets, so that experts can be as creative and effective as possible about modeling and solutions. CIC's Ecosystem Mapping Report, published in October 2020, identified a key **gap around cross-sectoral cooperation** and the lack of a **central resource** for data relevant to peacebuilding and prevention.[1] At the time, we recommended a scoping of options for a **data hub for prevention and peacebuilding** that could consolidate relevant data and provide expert resources for field-level initiatives. This hub would have a potential to centralize information, foster innovation, build capacities for data-driven approaches, and enable partnerships, which would lead to more evidence-based practices and strategies.

A good model for a data-driven hub comes from the humanitarian sector: The Centre for Humanitarian Data in the Hague, managed by the UN OCHA, is a central resource for data services, data literacy, data responsibility, and predictive analytics.[2] With the objective of learning from the Centre's experiences and to envision the future of data-driven peacebuilding work, we facilitated a set of interactive workshops in December 2021 to crowdsource ideas, best practices, and opportunities for a more central hub to support data-driven approaches to prevention and peacebuilding.

This briefing summarizes those discussions and provides suggestions for future steps. It starts by identifying the growing needs for prevention and peacebuilding practitioners for better data services, including, e.g., a consolidated resource of relevant data sets, an entity to maintain the data and identify gaps, and a central resource for guidance on data protection and ethics. It continues by describing how a hub could also build capacity of practitioners on, e.g., data literacy and the safe and effective use of emerging technologies. It concludes by suggesting a way forward, including options for establishing a new hub versus potentially enlarging an existing hub, such as The Centre for Humanitarian Data.

Why learn from the humanitarian sector about data-driven approaches?

On December 1–2, 2021, the **Emerging Technologies in Peacebuilding and Prevention— Lessons Learned from Humanitarian Actors** virtual workshop provided an opportunity to discuss the current and future applications of emerging technologies in peacebuilding and learn from humanitarian actors and their experiences. Humanitarian and peacebuilding work are carried out in similar contexts—in fragile countries, and with vulnerable populations. Similar settings bring similar challenges and responsibilities, which make sharing lessons even more vital to this work. Humanitarian actors are often ahead in innovating both in the use and in the evaluation of the pitfalls of using emerging technologies. Through efforts to work across the humanitarian-development-peace nexus, many actors have already recognized linkages between their efforts and emphasized the need to work together to meet peoples' needs more effectively, mitigate risks and vulnerabilities, and sustain peace.

Although humanitarian and peacebuilding actors often share the same fragile and conflict-affected spaces, working with similarly vulnerable groups, our objectives may not always neatly overlap. **Peacebuilding practitioners must find ways to build trust and social cohesion in fractured societies; humanitarians, while not oblivious to these goals, must focus on alleviating suffering and delivering life-saving goods and services.**

Where do we meet? It is in the space around solutions, "do no harm," and the recognition that we need to get much, much better at anticipation rather than reaction. This is why we exchanged lessons learned—both positive and negative—around the role of very exciting but also potentially harmful technologies for solving practical challenges, such as chatbots, remote sensing, AI, predictive analytics, and many other approaches.

More information on the workshop can be found on our website.

The role of a hub in strengthening data services for peacebuilding work

One distinct value of a hub is to centralize a range of "data services" for the prevention and peacebuilding field. This could include, for example, a one-stop-source for relevant datasets; identification and filling of gaps in important data; and making it easier for practitioners to coordinate and share data; among other services.

The benefits of a "go-to location" for prevention and peacebuilding data

Although there are different websites and data resources available, such as The Armed Conflict Location & Event Data Project (ACLED), Uppsala Conflict Data Program at the Department of Peace and Conflict Research, Uppsala University (UCDP), Global Terrorism Database (GTD), International Organization for Migration (IOM) data, World Bank data, etc., there is a challenge in combining and integrating these resources and making them

Key lessons learned from the Humanitarian Data Exchange

The Humanitarian Data Exchange (HDX), created in 2014, is an open platform for sharing data across humanitarian organizations with a goal to make data easy to find and easy to use for analysis. When HDX was launched it had around 800 datasets. Today there are 18,500 different datasets covering every active crisis. While many organizations have a sector specific approach in collecting data (such as IOM collecting data about displacement, or UNHCR collecting data about refugees) OCHA has important role on information management, being tasked to bring and aggregate all these datasets to build common operational picture of a humanitarian crisis. HDX saw a record growth of users, with more that 1.4 million people using the platform in 236 countries and territories and more than 1.8 million datasets downloaded. Prior to HDX finding humanitarian data was very difficult and once the data was found it was very difficult to trust that dataset. With HDX one centralized place was created where users can find the data, see the source, can rely on the information and metadata behind the dataset and reuse it. What is that centralized place for the peacebuilding and prevention community?

You can listen in on our workshop discussion, **Co-creating futures of data services for peacebuilding work**, on our Youtube channel.

useful in the quickest way possible to translate that data to decisions. Developing UN OCHA Humanitarian Data Exchange (HDX) as a centralized place for humanitarian data increased awareness of the data available for humanitarian response activities and drew attention to what was missing. This mobilized efforts to fill in the gaps, and progress has been achieved thanks to advocacy of HDX and its partners and coordinated efforts in the field.

Collecting new and/or better utilization of existing data

Although there is increasingly more data available for humanitarian and peacebuilding actors, its quality and integrity are mixed. The data may also lack of granularity, and be scattered, unstructured or insufficiently protected. Particularly inside the UN, there is a lot of data gathered, but it may be siloed and not interoperable.

HDX's data grids are a step towards creating awareness of what data is available at the country level, with a goal to eventually cover the subnational level.[3] As there is also a need to update the existing datasets regularly, a dataset is removed from the data grids if it is outdated or has been superseded by a relevant dataset. The following is an example of a data grid completeness for Mali and a chart showing an overview of data completeness for all datasets.

Establishing a hub would allow for better identification of data needs. For example, the prevention and peacebuilding field is still lagging in using unstructured data, such as news reports, articles, imagery, and social media. It would also be able to identify where data needs to be improved (e.g., much of this unstructured data is not available in a usable format).[4]
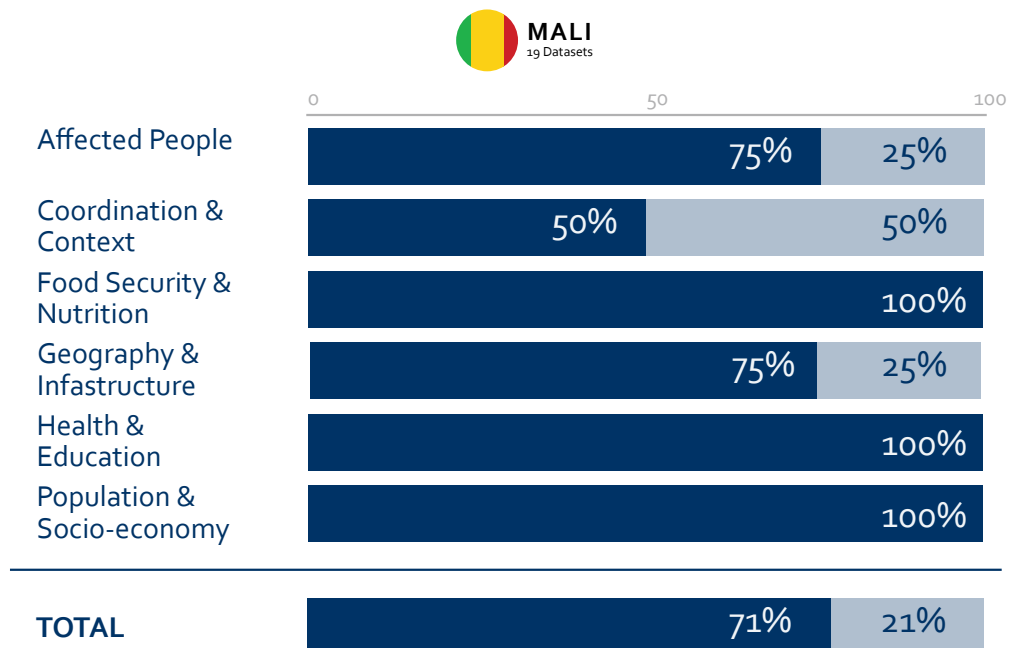

Figure 1: Data completeness for Mali from HDX website. Source: https://data.humdata.org/dashboards/overview-of-data-grids

"Data, data, data...But then what? Our collective energy has been hijacked to a certain extent by this focus on data, and less focused on what we actually end up with data."

— Olivier Cottray
ESRI
Director of Humanitarian Solutions

Tackling definitional and conceptual challenges to effective data use for peacebuilding

Beyond the need for consolidated and better data, another issue relates to methods for assessing the relation between datasets and prevention/peacebuilding concepts. For example, if we are interested in "social cohesion" as a core peacebuilding concept, then which datasets would be most relevant? For this, we need meta-level knowledge graphs that provide maps to connect words that mean similar things. This is the kind of challenge that a data hub could start to bring expertise around.[5]

Simplifying data collection, coordination, and sharing

A hub could be a key driver in developing standards for data collection processes, agreeing on what should and should not be collected, and merging and using data from different sources. There is an importance of agreeing on what the good data is and what the good use of data is. Many actors are slipping into "over-collecting" data without clear goals or use cases—and without a clear sense of when the data would be deleted. Geocoding and temporal coding was mentioned as an important dimension for practitioners in using and sharing datasets. For example, for a practitioner going to the field, to a certain province, city or village getting a quick access to information and data available for this specific area would be extremely beneficial.

Total Percentage Data Complete	Total Percentage Data Incomplete	Total Percentage No Data
59%	**23%**	**17%**
Nov. 24, 2021	Nov. 24, 2021	Nov. 24, 2021

Number of Locations	Number of Categories	Number of Sub-categories
27	**6**	**21**
Nov. 24, 2021	Nov. 24, 2021	Nov. 24, 2021

Global Data Grid Completeness
■ Complete ■ Incomplete ■ No data
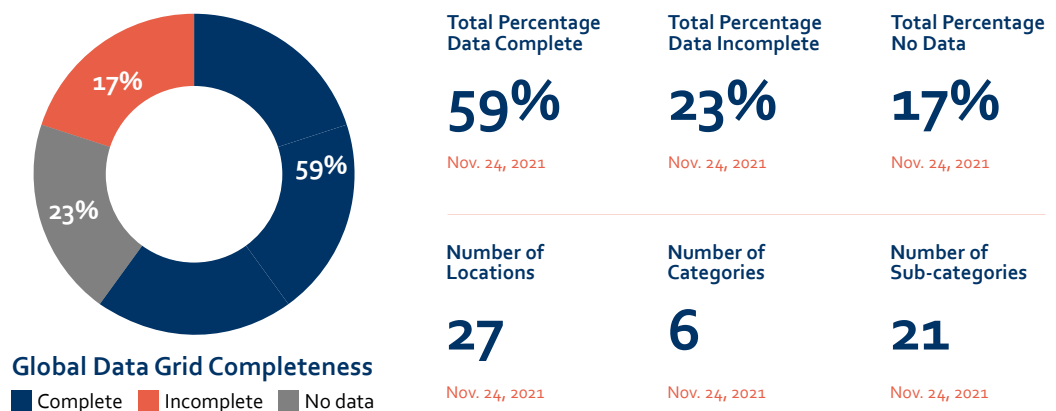
Figure 2: Global overview of data completeness for datasets on the HDX website.
Source: https://data.humdata.org/dashboards/overview-of-data-grids

Consolidating standards for data protection and data responsibility

Using data in conflict settings brings additional layers of vulnerability, for example, the misuse of open data disaggregated by ethnicity as a targeting lens for a specific group. This strengthens the requirement that technologists need to have a good context awareness.[6] Many humanitarian actors have developed internal data protection guidelines and policies, such as International Committee of the Red Cross's *Handbook on Data Protection in Humanitarian Action* and the Inter-Agency Standing Committee's *Operational Guidance on Data Responsibility,* offering a set of principles and guidelines for safe, ethical, and effective management of personal and non-personal data for operational response.[7] In cases where data is sensitive, organizations like HDX share only the metadata and make the underlying data available only by request.[8]

Yet there are still gaps. For example, there are no unified standards for data protection for simple tools people use on a daily basis (for example: do Google Docs, Dropbox, email, or other tools entail safe data sharing?). Also, there is currently no regulation on how AI/ML is being used, nor requirements for humanitarian or peacebuilding actors to audit their AI/ML systems for bias or discrimination.[9] The lack of enforcement on protection was also recognized as a challenge that needs to be addressed.[10] Finally, there was a strong call for protection to be part of the design process of any application—including a potential data hub.

Providing ethical guidance on deploying AI/ML in peacebuilding

While data-driven approaches in prevention and peacebuilding work can create benefits, there is always a risk of harm. Assessing the magnitude and depth of the potential harm is not always part of the application decision making process. Making ethical assessments and conducting risk/benefit analysis is often costly and timely process and not in the reach of many peacebuilding organizations. Therefore, having one centralized place that can develop and offer set of standards and decision-making tools could be a valuable part of a hub's mission and activities.

> "*The question with biometrics is not how do we use biometrics well, the question is should we and if so, how do we do that responsibly… This relates to broader discussion around ethics of technology and data in the sector and the fact that we do lack across the board the skills to deliberate effectively and in a meaningful way on real ethical issues that arise with technology and data, and investing in building that capacity is key.*"

—STUART CAMPO
UN OFFICE FOR THE COORDINATION OF HUMANITARIAN AFFAIRS (OCHA), CENTRE FOR HUMANITARIAN DATA

Combating digital neocolonialism and data extractivism

Most of the big tech companies working in humanitarian contexts are located in the United States or the European Union, while most of the work is in the Global South. This raises a question how to reduce historical power imbalances or "digital colonialism," in which valuable data is extracted out of these contexts but does not always return as a benefit to people on the ground.

Localization and the centrality of local actors is key to peacebuilding processes; logically, therefore, a peacebuilding approach should counteract colonial and extractive models. As so many emerging technology tools offer the chance to communicate with ever larger numbers of people, both the humanitarian and peacebuilding communities need to be more creative in thinking how to use these tools for greater inclusion. At the same time, it should not be forgotten that there is still significant proportion of the world's population that is digitally disconnected or digitally illiterate. An estimated 37 percent of the world's population, approximately 2.9 billion people, most of them women, are still offline—out of which an estimated 96 percent live in developing countries.[11] There are still no generally accepted principles of what constitutes good public-private partnership in a conflict context—there is no guidance on what those practices should look like.[12]

Capacity-building for data-driven peacebuilding work

Acknowledging that successful and responsible data-driven work is inextricably linked to building knowledge and skills for that work, there is a parallel need to build capacity for prevention and peacebuilding actors.

Ensuring organizations have the fundamental building blocks to use data effectively

Many AI enthusiasts claim that data-driven technologies can improve humanitarian action, enable organizations to do more with less, and ultimately help more people and save more lives. Many would also claim, however, that not all humanitarian actors are fully ready and fit for big data and AI applications. Similar conclusions must be drawn, perhaps even more strongly, for the prevention and peacebuilding sector.

Key Lessons from The Centre for Humanitarian Data

The Centre for Humanitarian Data works to increase digital literacy of non-technical humanitarians, bridge the gap between technical staff and decision-makers, and increase demand for, and collection of, better data. There are several lessons learned from the Centre for Humanitarian Data on data literacy capacity building that can serve as a good sharing lesson for peacebuilders:

- **One-off data trainings are not enough to develop data literacy**. Such trainings often concentrate on building technical skills. To be effective, this needs to be complemented with development of strategic data skills (the knowledge and capacities to manage data like an asset) and "sense making" data skills (the knowledge and capacities to translate strategic priorities into data exercises and use data products to make data-informed decisions).

- **Integrating data literacy into existing trainings** and programs is a good approach to maximize efficiency.

- Another lesson shared is the need to **go back to basics,** to build an understanding of how data does and does not represent the world. As there are so many different things that training can focus on, **organizations need to decide what their strategic priorities are and design** for those priorities.

You can listen to our workshop, **Co-creating futures of capacity building for data driven peacebuilding work,** on our YouTube channel.

Many humanitarians and peacebuilders, both in the field and at the decision-making positions, are still lacking digital and data literacy, and their organizations—especially but not exclusively smaller ones—are not always equipped to understand and interact with data-driven technologies, systems, and tools. Hence, more education is needed in the sector about the fundamental building blocks required for the safe, effective, and sustainable use of emerging technologies.

Avoiding "solutionism"

At the same time, what some notice is that many in the sector have an "AI fascination bias," seeing the technical solution for every problem. More in-depth knowledge is needed, and capacities built to recognize this trap and make appropriate decisions when and how to use emerging technologies. Before using these tools, it is important to have a business case analysis of what is the value for bringing in the data science and which tools could realistically be adopted and sustained in the working culture of the organization.

Building capacities is also not only about the tools and using new technologies or about building data skills—we also need to build capacities of data scientists for recognizing the political relevance of data interpretation—the data we select and the terms we use may have a significant and sometimes sensitive political implications,

Stronger data literacy

There is an acknowledgment today that using data can be a core competency for organizations, and that we need to build data literacy for peacebuilding and prevention practitioners, which was not the case a decade ago. There is also a growing awareness that data literacy is more than a narrow set of technical abilities (coding, visualization, etc.), and therefore it is not a challenge that can be met by hiring purely technical staff as a supplement. Data literacy includes, for example, the skill of recognizing data that is relevant for the question being asked.

"Data is always about people. Even machine data or aggregate data that we may think has no relationship to a human being...it will always come back to human beings...Protection is more complicated in a conflict context—it's more than the typical enterprise safety and security concerns around technology."

—Rakesh Bharania
Salesforce.org
Director, Humanitarian Impact Data

While building data literacy, we also need to be aware of variety of organizations in peacebuilding field. Many of them are small, community-driven actors that are already overwhelmed with the daily challenge of their work. Finding a way to tailor trainings that are easily integrated into their workflow and do not shift time away from immediate peacebuilding priorities will be key to getting these organizations to adopt and use data better.

Capacity for protection of data and people

Collecting and sharing data always comes with a risk of data being used by "threat actors"—meaning those who would use data to threaten, diminish, or harm peacebuilders and their work—including governments.[13] Building capacities to collect and use data must be coupled with building capacities to protect data and people—starting with the capacity to understand different layers of vulnerabilities and risk in conflict settings. Physical harm is a real concern in humanitarian and peacebuilding settings, and they go beyond General Data Protection Regulation (GDPR) type of privacy concerns or compliance lens. Even aggregated data, which would usually not raise privacy concerns, can be used to harm people in a humanitarian conflict lens.[14] Recognizing these vulnerabilities and nuances of data and people protection in conflict settings must be part of capacity building.

Capacity for interpretation, policy, and legal expertise

Before using AI-related technologies in sensitive places we need to pull in our policy and legal experts before we approach the technology side. In the peacebuilding sector, many still do not have access to this kind of expertise. Also, we need to recognize the need for building the "translation skills" for complicated concepts, and through this translation to facilitate better conversations between technical and non-technical staff. This will also help technology experts, who often do not have enough insight into the complexities and challenges of humanitarian, prevention, and peacebuilding work to understand the full policy dimensions of their technical work.

Capacity building with an open dialogue mechanism

Capacity building should not be a one-way session that provides information to practitioners in a form of a course or a training; rather, it should be developed as a meaningful dialogue mechanism and interactive feedback process. Dialogue is important for understanding which products would be most relevant for practitioners (visualizations, dashboards, static maps, or prediction models can have a completely different "buy-in" for different audiences and stakeholders). Without context and subject matter expertise, we will not know which questions are important to tackle. Lastly, data driven approaches are often too segmented from the people who own the problems and solutions on the ground and who bear the risks of using the data. We need dialogue with and data literacy for those actors as well.

Recommendation for next steps for a Prevention and Peacebuilding Data Hub

While there is a consensus that the peacebuilding field would benefit from having a **centralized data hub**, there are different views on whether the hub should be a separate entity, or possibly integrated into an existing data hub, such as HDX. One interesting idea would be to **explore whether HDX could open a "nexus" window that would start to integrate data across the humanitarian-development-peacebuilding nexus**. This could even be done on a pilot basis.

The biggest challenges for creating a separate hub are the high start-up costs as well as inefficiencies in maintaining data that might already be maintained elsewhere. For example, 80 percent of HDX effort goes to maintenance—making sure the links to datasets are not broken, that the data can be downloaded, that the metadata is complete, etc.[15]

Hence, there is good reason to explore whether peacebuilding and prevention data could be integrated into an existing hub, like HDX, which already includes relevant datasets. But there are some challenges with this approach as well—mainly because existing hubs have different objectives and use cases. For HDX, one issue is that it is currently concentrated on 27 humanitarian crises, which may not line up with the needs of the peacebuilding research community. The UN's Peace and Security Data Hub is another resource, but it focuses exclusively on organizing data and information generated by the UN.[16]

What many of the workshop participants shared is the concern that their organizations do not have internal capacities and resources to dedicate to building data literacy, data protection, and capacity for policy and legal expertise. Out of these concerns the proposal was raised to build a centralized, open resource for capacity building—yet again, following the example of the Centre for Humanitarian Data.

As next steps, we suggest conducting an in-depth analysis of benefits and technical possibilities of:

- Growing HDX to include prevention and peacebuilding data

- Growing HDX to become an integrated place for a "nexus" data

- The potential of the UN Peace and Security Hub to include external data

- Having a centralized resource for capacity building

- Capacity-building efforts being hosted by the same data hub

- Utilizing existing resources from other fields, such as the "data literacy curriculum for humanitarians" already developed by the Centre for Humanitarian Data

- Sharing resources and learning from the private sector (adapting tools such as Microsoft Responsible AI resources, or Salesforce Ethics by Design)

In addition, and in case the decision has been made for establishing a separate hug, we suggest:

- Conducting a feasibility study to look at a scope and model of a potential separate hub, including hub's mission and vision, audience and primary users, operational structure, focus areas, activities, and membership.

- Defining the role of a hub as a provider of data services, data policy and ethics, data literacy capacity building, and community building

- Exploring potential locations and the added value of having multiple locations/hubs across the world

- Exploring potential funding streams and sustainability models for such a hub

With the current UN secretary-general's commitment to data-driven transformation, and following many efforts, commitments, and energy from member states, academia, and civil society to advance their data-driven peacebuilding work, there is now a momentum to bring the idea of Prevention and Peacebuilding Data Hub to reality.

Creation of a centralized hub, either as a separate entity or integrated with an existing hub, will be a crucial step in strengthening effective use of data-driven approaches and increasing effectiveness of prevention and peacebuilding practices. A hub will also respond to needs for a one-stop source for developing analytical expertise in the field and ensuring responsible and ethical adoption and implementation of data-driven approaches to conflict prevention and peacebuilding. It will strengthen national international and ground-level work, by opening its resources to the diversity of actors in the field, and with this it will enable and strengthen highly needed cooperation across sectors.

Endnotes

1   Branka Panic, "Data for Peacebuilding and Prevention Ecosystem Mapping: The State of Play and the Path to Creating a Community of Practice," New York: NYU Center on International Cooperation, (2020); https://cic.nyu.edu/sites/default/files/data_for_peacebuilding_and_prevention_-_ecosystem_mapping_-_october_2020.pdf

2   "The Centre for Humanitarian Data", accessed March 3, 2022, https://centre.humdata.org/

3   CJ Hendrix, "To Find and Fill data Gaps," The Centre for Humanitarian Data, May 21, 2019, https://centre.humdata.org/introducing-the-hdx-data-grid-a-way-to-find-and-fill-data-gaps/

4   For more details, see the workshop session, "Can the machine be a peacebuilder's best friend?," January 2022; available at https://www.youtube.com/watch?v=z51X4TbdAeE&list=PLh9Xp7DYZRYGwJiAtSQnWbM0AsOvFViZw&index=12.

5   We thank Gary Milante for this particular point.

6   For more details, see the workshop session, "Best Practices for Providing and Securing Data in Humanitarian Crisis Situations," January 2022; available at https://www.youtube.com/watch?v=I9X6y4phhUM&list=PLh9Xp7DYZRYGwJiAtSQnWbM0AsOvFViZw&index=19.

7   See Christopher Kuner and Massimo Marelli, "Handbook on data protection in humanitarian action," ICRC; available at https://www.icrc.org/en/data-protection-humanitarian-action-handbook; and "Operational Guidelines Data Responsibility in Humanitarian Action," Interagency Standing Committee, February 2021, https://interagencystandingcommittee.org/system/files/2021-02/IASC%20Operational%20Guidance%20on%20Data%20Responsibility%20in%20Humanitarian%20Action%20-%20February%202021.pdf.

8   At HDX, every dataset goes through a system where an algorithm looks for pieces of personal information (such as email, last name, location). Once such data is located, the dataset contributor is contacted to flag and address the sensitive information. Another approach used is to anonymize the datasets by estimating the likelihood of a person being reidentified in the dataset.

9   Sarah W. Spencer, "Humanitarian AI: The hype, the hope and the future," Humanitarian Practice Network, November 2021; https://odihpn.org/publication/humanitarian-artificial-intelligence-the-hype-the-hope-and-the-future/

10   No penalties or fines have been taken for any of the recent UN data hacks or breaches or linked to the alleged unethical sharing of data by UNHCR.

11   UN International Telecommunication Union, "Facts and Figures 2021: 2.9 Billion People Still Offline," June 10, 2022; available at www.itu.int/hub/2021/11/facts-and-figures-2021-2-9-billion-people-still-offline/.

12   Sarah W. Spencer, "Humanitarian AI: The hype, the hope and the future," Humanitarian Practice Network, November 2021; https://odihpn.org/publication/humanitarian-artificial-intelligence-the-hype-the-hope-and-the-future/

13   As shown in the recent examples of Afghanistan and Myanmar, where governments gained access to biometrics and other data that can be misused against the vulnerable populations.

14   For example, aggregate data that reveals that 80 percent of people in a particular village are members of a certain ethnic minority could be used for targeting purposes.

15   HDX was initially hosting some of the Peace and Security Pillar data of the UN, now integrated into Peace and Security Data Hub, and many of the HDX datasets are already used by peacebuilders.

16   See the UN Peace and Security Data Hub at psdata.un.org.

NYU | CENTER ON
—— | INTERNATIONAL
CIC | COOPERATION

726 Broadway, Suite 543

New York, NY 10003

cic.nyu.edu