

Enhancing responsible microdata access to improve policy and response in forced displacement situations



Task Team Lead: Harriet Mugeru and Olivier Dupriez
JDC Focal Point: Patrick Brock and Harriet Mugeru

The initial idea for this activity emerged from a [workshop](#) organized by JDC in May 2020, on access to, and protection of data on forcibly displaced populations. The workshop brought together experts on data protection, anonymization, and de-identification from a range of different industries and sectors, with around 30 participants from UNHCR headquarters, field offices and management. Participants highlighted the imperatives (among others) to define a framework for data risk assessment and management, finalize microdata de-identification standards and protocols to be applied across UNHCR, and expand UNHCR's data protection capacity, including through partnerships.

OBJECTIVES AND ACTIVITIES

The objective of this activity was to develop a framework for making microdata on forcibly displaced populations (FDPs) more accessible to external stakeholders in a responsible manner, with a dual focus on: (1) technical support to UNHCR to develop comprehensive protocols and processes for microdata curation and dissemination; and (2) building consensus around a common framework for responsible microdata dissemination among stakeholders with shared interests and facing similar challenges working with microdata on FDPs (including national statistical offices), that complies with relevant legal, ethical and technical principles.

Specifically, JDC funding supported: (a) a review of existing microdata curation and dissemination practices; (b) development of tools, protocols, and guidelines for microdata protection and dissemination; (c) training, sensitization, and capacity building within UNHCR; (d) building partnerships and coalitions with humanitarian, development, and national statistical offices; and (e) adopting tools, protocols and guidelines for data protection and dissemination for broader stakeholders (beyond UNHCR).

RESULTS AND OUTCOMES

- **Deepened partnership between the World Bank and UNHCR on the UNHCR Microdata Library.** While the World Bank had already begun to work with UNHCR on its microdata library prior to the establishment of JDC, the support provided by JDC through this activity has been pivotal in deepening the World Bank-UNHCR technical partnership on microdata.
- **Preparation of the UNHCR Data Curation Handbook.** JDC funding enabled specialist World Bank staff to advise on the preparation of the UNHCR Data Curation Handbook, which provides technical guidance for data producers within UNHCR, including best practices for data cleaning, risk assessment, anonymization, and the creation of metadata.
- **Contribution towards institutionalizing responsible data sharing within UNHCR and expanding the UNHCR Microdata Library.** UNHCR now has a high-performing team that operates a very good survey catalog and understands all aspects of data protection and documentation; they have the UNHCR Data Curation Handbook and all the appropriate data curation procedures are in place. These improvements have enabled the expansion of the UNHCR Microdata Library. While the UNHCR team should be given credit for building the microdata library and

putting in place all the appropriate data curation procedures, the frequent collaborative exchange with the World Bank team, and the input of their technical expertise, have ensured that this has been a smooth and effective process.

- **Preparation of training materials on responsible data access, used in training events offered to UNHCR staff and external organizations, including national statistical offices.** Data protection and dissemination tools (including on statistical disclosure risk assessment, microdata anonymization, metadata standards, and data curation practice) were packaged into training materials, which have been used in UNHCR-led training events offered to both UNHCR staff and external partners. Training materials created through this activity will be made publicly available, and form the basis for future training events, including World Bank training on data anonymization organized by the Center for Development Data (C4D2) and the DEC Global Institute in Perugia in July 2023, with 30 participants coming from statistical training institutes in the Africa region.
- **Technical exchanges on microdata curation systems and practices.** JDC convened workshops with the OCHA Centre for Humanitarian Data (OCHA-CHD), World Bank and UNHCR on aligning microdata curation and dissemination approaches, guidance, and training. As part of this series of workshops, JDC produced a [Fragility Forum Podcast](#) with the OCHA-CHD and the UK Royal Society Privacy Enhancing Technologies Working Group, exploring the mosaic effect and privacy enhancing technologies. There were also collaborative exchanges on consistency in data curation, cataloguing and data protection with the WFP Microdata Library team, and other partners, including JIPS, ICRC, and NRC.
- **Development of innovative tools and methods for preserving data privacy, including for highly sensitive data.** The World Bank team developed a highly innovative machine learning model that can produce realistic tabular relational synthetic data (i.e., with statistical properties similar to real-world survey data) while at the same time implementing safeguards to prevent the model from reproducing sensitive real-world data. This new synthetic data generation model, which has been developed as an open-source package, offers an approach to fill data gaps, allow for analysis of sensitive microdata while preserving their confidentiality, and generate datasets for modeling and microsimulation. For example, it could be used to “unlock” datasets based on sensitive administrative data collected by UNHCR and OCHA. The model has been described in an academic [paper](#) by Aivin Solatorio and Olivier Dupriez, which will be presented at the *2023 UNECE Expert meeting on Statistical Data Confidentiality* and is expected to garner interest from participating national statistical offices. The approach and some of its applications will be presented at the UNECE Expert Group Meeting on Statistical Data Confidentiality in Wiesbaden in September 2023.
- **Empirical testing of statistical disclosure controls.** In April 2022 when JDC reviewed and rationalized its activities, it was agreed to prioritize work on risk assessment and statistical disclosure control. On risk assessment, the UNHCR team completed the Disclosure Risk Assessment Template as part of the UNHCR Data Curation Handbook. On statistical disclosure control, the World Bank team undertook empirical work making use of the newly developed approach to generating synthetic household survey data, using it to test the efficacy and sensitivity of current anonymization processes, and to explore the mosaic effect. The findings from this work were written up as part of an internal report for UNHCR for use in the future development and optimization of their data curation and dissemination pipeline.
- **Technical input to the [IASC Operational Guidance on Data Responsibility in Humanitarian Action, 2023](#).** The 2023 revision to IASC Operational Guidance was prepared through a collaborative process led by the UN Data Responsibility Working Group and involving stakeholders from across the humanitarian system, and was endorsed by the IASC in April 2023. JDC contributed to new wording on standardization, interoperability, and data documentation in the data quality principle (p18) (nudging the humanitarian data system towards greater accessibility for government and development organizations) as well as the UNHCR Microdata Library listed as an ‘Example of Data Responsibility in Practice’ (p37, and [here](#)).
- **Investment in developing standards and tools for data protection has also been valuable and timely for the World Bank.** Within the Bank, there is growing interest in data protection standards, guidelines and tools that can help teams comply with the World Bank’s Privacy Policy and meet the increasingly high expectations of organizations that share data with the Bank. The JDC-funded activity contributed to the assessment and further development and documentation of these tools and guidelines.

CHALLENGES

- **Areas of overlap with other JDC-supported activities.** The responsible data access activity intersected with the other JDC-supported activities (particularly JDC support to the World Bank and UNHCR Microdata Libraries). However, the World Bank and UNHCR teams working on the various JDC-supported activities adapted to these overlaps and worked very well together. The responsible data access activity “was a glue that stuck [the various JDC-supported activities] all together and allowed them to work together in a constructive way.” In particular, the responsible data access activity clearly supported the work of the UNHCR and World Bank microdata library teams.
- **The broad scope of the initial concept led to some delays.** The delay between approval and the initiation of activities, and the extended time before completion can be attributed, in part, to the broad initial concept for this activity. Not all elements of the original concept had full buy-in from all stakeholders. It took time to reach consensus on and specify the detailed activities and expected outcomes, but this process was helpful for the team and contributed to strong results for this activity.
- **Engagement with data producers beyond UNHCR was challenging.** The original idea was that all organizations producing data on forcibly displaced people would subscribe to a common framework for data protection, documentation, and dissemination. Efforts were made to engage other data producers, beyond UNHCR, which were successful in some instances but not in others. A microdata technical exchange group was convened by JDC (with participants from the World Bank, UNHCR, OCHA, WFP and NRC), which was a useful forum for technical exchanges. NRC have continued to engage separately with UNHCR after being convened by the team. However, other exploratory engagements were less fruitful, and there was less responsive engagement from some data producers (e.g., ICRC) than expected.
- **There were less high-quality socioeconomic microdata on FDPs than anticipated.** At the scoping stage, it was anticipated that socioeconomic data could be gathered from several organizations working with FDPs, including ICRC, UNICEF, and WFP. WFP has an internal microdata library, which is increasingly accessible to external users, and it is hoped that external license requests will soon be possible (although it is not yet clear how much of this data is high quality socioeconomic microdata). NRC is setting up new centralized data systems, and the microdata technical exchange group enabled them to draw lessons from organizations that have existing systems, for example to maximize future interoperability.
- **Potential legal issues associated with accepting socioeconomic data from external data producers.** This activity considered the ethical and technical issues associated with sharing sensitive data on forcibly displaced people, but not the legal dimensions. Potential JDC funders have expressed interest in the possibility of uploading their data to the UNHCR Microdata Library (and including conditionality in grants requiring recipients to conform to UNHCR data curation standards and to upload data to the UNHCR Microdata Library). However, there are potential reputational and legal risks associated with disseminating data collected by organizations that are not official partners of UNHCR. In cases where UNHCR already has a MoU (e.g., with DRC, NRC etc.), these could be amended to include standards or criteria for data uploaded to the UNHCR Microdata Library.
- **Difficulty of measuring the impact of UNHCR Microdata Library.** There is an inclination to measure the success of the UNHCR Microdata Library using the number of views, downloads, or academic citations. However, there may be a small group of data scientists and economists who are viewing and downloading the data, yet they are making impactful use of it. Many researchers (including in the Bank) will use the data for analysis that is never intended to be published in an academic journal but is used for public policy advice or operational purposes. For those researchers that do intend to publish in a peer-reviewed journal, this usually takes around two years so there will be a lag in citations. A better metric would be the number of people who make use of the analysis or read the academic papers based on data from the UNHCR Microdata Library, but this would be difficult to measure.

LESSONS LEARNED AND JDC'S ROLE

- **“A+ collaboration” between UNHCR, World Bank and JDC.** The UNHCR, World Bank and JDC teams worked exceptionally well together. JDC staff were very helpful in clarifying and adjusting the scope and expectations for this activity.
- **Flexibility is appreciated.** The teams made some adjustments during implementation to avoid duplication with other JDC-supported activities. The team was also grateful for the flexibility permitted by JDC to request a no-cost extension to finalize the report.
- **Spotlights on how available microdata is being used.** Given the challenges of measuring the impact of the Microdata Libraries, it may be useful to prepare some short pieces that spotlight the use of microdata for various purposes, including operational purposes.
- **Reinforces JDC’s strategic focus on data science,** given its importance for knowledge building and knowledge sharing.

READ MORE ABOUT OUR ACTIVITIES HERE:

www.jointdatacenter.org