

## Support to the development and operation of Microdata Libraries at UNHCR and the World Bank



Task Team Lead: Matthew Welch  
JDC Focal Point: Patrick Brock

The [World Bank Microdata Library](#) is the world's largest repository of development survey and census data, with a catalogue of over 9,500 datasets. From 2017 to 2019, the World Bank Development Data Group (DECDG) supported the initial establishment of the [UNHCR Microdata Library](#), with funding from the DfID-financed "Building the Evidence on Protracted Forced Displacement: A Multi-Stakeholder Partnership".

Both Microdata Libraries (MDLs) use open-source tools developed and/or supported by the World Bank, including systems for microdata documentation (DDI editor), anonymization (sdcmicro), and cataloguing (NADA). The UNHCR MDL is managed in parallel to the UNHCR Raw Internal Data Library (RIDL), which stores and shares operational data (including microdata), accessible to UNHCR staff and partners.

### OBJECTIVES AND ACTIVITIES

The objective was to support the UNHCR and World Bank MDLs in a joint effort to ensure data on forced displacement are better documented and more widely accessible.

Detailed activities included: (i) technical support and development of tools; (ii) knowledge sharing on best practices for preservation, anonymization and dissemination of microdata; (iii) identifying, organizing and tagging microdata from the World Bank's MDL relevant to forced displacement and sharing these with the UNHCR Curation Team; (iv) surfacing new data sets or combinations of existing data sources and sharing these with the UNHCR Microdata Library; (v) curation of new data generated by JDC-sponsored surveys.

The activity was approved on December 14, 2020, JDC released funds in January 2020, with an initial end date in December 2022. JDC granted a no-cost extension to March 1, 2023.

### RESULTS AND OUTCOMES

- **Developed customized tools for the UNHCR MDL.** Rather than transferring exactly the same tools used in the World Bank MDL, the suite of tools was adapted and customized to meet the needs of UNHCR.
- **Enabled UNHCR to publish a growing catalog of data on forcibly displaced populations.** As part of the previous DfID funded project, the World Bank MDL team introduced only the basic software for the UNHCR MDL, but there was not any funding for institutionalizing UNHCR's approach to preservation, anonymization, and dissemination of microdata. With JDC funding, the World Bank MDL team provided an expanded suite of tools and technical support, without which the UNHCR MDL would have remained solely an exploratory demonstration catalog. The UNHCR MDL also provides the platform for other organizations that do not have their own microdata libraries (e.g., DRC and NRC) to publish datasets on forcibly displaced populations.
- **Documented, prepared, and published seven JDC-sponsored surveys in the UNHCR and World Bank MDLs.** JDC has funded several surveys undertaken by the World Bank covering forcibly displaced populations. This activity supported the documentation, preparation, and publication of these datasets in the World Bank and UNHCR MDLs.
- **Built a strong collaborative "friendship" between the technical teams in the World Bank and UNHCR.** The frequent interactions between the two technical teams (including work on standards, tools, and methods) did

not only build the capacity of the UNHCR team, but also built a very strong collaborative working relationship, which helped to facilitate the excellent results achieved. This is a potential model for prospective engagements with other organizations on data sharing (e.g., HDX and OCHA). Moreover, with greater attention to data sharing as part of various UN initiatives, including the High-Level Committee on Management (HLCN) Task Team on Data Sharing, there is an opportunity for the World Bank, UNHCR and the JDC to share best practices with a broader group.

- **Outreach to a broader group on the importance of common standards and tools for metadata on forcibly displaced populations.** JDC facilitated meetings between the World Bank and UNHCR technical teams and a broader group (including WFP, NRC and OCHA). While these meetings did not surface high-quality datasets or precipitate new data sharing agreements, they gave the World Bank and UNHCR teams the opportunity to advocate for common standards and schema for metadata. These messages were well received by the broader group, which augurs well for interoperability and data sharing in the future. Organizations in the broader group were grateful for the opportunity to present their work to the World Bank, which is known to be a leader in this space. Insights from these meetings also led the World Bank team to add functionality to the metadata editor they developed with JDC support so that it can be used by other organizations. (The additional functionality permits the microdata editor to publish to CCAM, which is the technology that both WFP and NRC plan to use in their data sharing platforms.)
- **Contributed to a culture change in UNHCR around the importance of data for improving the lives of forcibly displaced populations.** This activity goes to the heart of the vision that UNHCR set out in its Data Transformation Strategy 2020-2025, for UNHCR to become “a trusted leader on data and information related to refugees and other affected populations, thereby enabling actions that protect, include and empower.” It has provided an important means towards achieving this vision and contributed to the broader appreciation for data taking place in UNHCR.

## CHALLENGES

- **Initial frictions between the World Bank concept note and UNHCR scoping note.** There were some initial frictions and delays due to the World Bank concept note and UNHCR scoping note not aligning. However, the teams ended up with a well described initial concept that satisfied the needs of both organizations, which contributed to the smooth implementation of the activity.
- **Legal issues prevented the World Bank MDL team from providing a new Metadata Editor to UNHCR.** The World Bank MDL team was not able to share its new Metadata Editor with UNHCR due to legal issues relating to the open-source licensing. However, this did not affect the results achieved by this JDC-funded activity, as the World Bank MDL team instead shared the underlying scripts and programmatic methods for data curation, since the UNHCR MDL team has the requisite expertise to use these methods. It is anticipated that the new open-source editor will be launched in October 2023, which will help both the World Bank and UNHCR MDLs engage with non-expert users and curation teams in national statistical offices around the world.
- **Identifying and tagging datasets in the World Bank MDL relevant to forcibly displaced populations was more complicated than anticipated.** It was intended that Natural Language Processing (NLP) approaches would be used to develop advanced semantic search tools that would help researchers locate datasets of interest and enable the World Bank MDL team to create collections relevant to the study of forcibly displaced populations. This was conceptualized as a straightforward procedure, which was conducted but did not immediately deliver the results that were anticipated. Instead, it has evolved into a much bigger piece of work that the World Bank and JDC continue to collaborate on. Moreover, Artificial Intelligence (AI) and GPT-based systems are expanding rapidly, with implications for surfacing data and for how developers/users interact with data that could build upon the Natural Language Processing (NLP) and tagging work that has already been undertaken.

## LESSONS LEARNED AND JDC's ROLE

- **JDC funding ensured dedicated World Bank staff time and a projectized approach.** JDC funding was critical to ensuring that adequate World Bank staff time was dedicated to this activity (insulated from competing demands), while also enabling a projectized approach to the work. Without this support it would not have been possible to get the UNHCR Microdata Library to where it is today.
- **JDC is a “best case” partnership model that should be showcased at the highest levels in the World Bank.** The collaboration between JDC, the World Bank and UNHCR in data and statistics is strong. The technical teams valued JDC’s “partnership attitude”, flexibility, ability to meet collaborators “at a deep technical level”, and open-minded approach to learn from technical exchanges, which they see as critical to JDC’s successful partnership model.
- **Beyond funding, JDC adds value through its in-house expertise, hands-on facilitation and support to teams, outreach to external stakeholders, and leadership skills.** JDC staff facilitated and fostered the strong relationship between UNHCR and Bank technical teams. JDC staff helped address unexpected issues as they arose during implementation, supported communication and collaboration between the technical teams, and ensured that detailed activities were on track to achieving the agreed outcomes. JDC staff also facilitated outreach to external stakeholders, bringing them into meetings and facilitating meaningful discussions.

READ MORE ABOUT OUR ACTIVITIES HERE:

[www.jointdatacenter.org](http://www.jointdatacenter.org)