

# ENHANCING THE CAPABILITIES OF UNHCR'S GLOBAL REGISTRATION SYSTEM

TASK TEAM LEAD: ANDREW HOPKINS | JDC FOCAL POINT: PADDY BROCK



UNHCR collects and manages the world's largest and most comprehensive dataset on refugees and asylum seekers. This dataset includes biographical, location and some socioeconomic data, and enables UNHCR and its government partners to plan, deliver and monitor vital protection and assistance services for refugees and asylum-seekers. It can also be used by development and research institutions to construct sampling frames (that is, the source data from which a representative sample is drawn) for household surveys of refugees to enable analysis of the socioeconomic dimensions of refugee crises.

UNHCR's registration data is stored in its Population Registration and Identity Management Eco-System (PRIMES), which brings together UNHCR's registration, identity management and case management applications into one interoperable system. PRIMES includes proGres, UNHCR's registration application, as well as the Rapid Application (RApp), which enables offline collection of data.

## OBJECTIVES AND ACTIVITIES

The objective of this activity was to improve the quality and accessibility of UNHCR registration data and to increase its use. To this end, the JDC provided funding to support the development of PRIMES, specifically:

- Technical work to migrate individual records from proGres v3, an offline, system of separate databases, to proGres v4, a centralized, integrated platform;
- The development of tools and dashboards to enable regional and country operations to monitor and improve data quality;
- Development of the Dataport, a data warehouse and business intelligence tool that automatically pulls data from proGres and allows users to analyze and visualize data, including a range of socioeconomic indicators; and
- The transfer of proGres data to the Raw Internal Data Library (RIDL), where it can be curated, anonymized and disseminated, and considered for subsequent migration to the Microdata Library (MDL).

## RESULTS AND OUTCOMES

Several important results and outcomes have been achieved during the two-year project:

- **The geographic coverage and the volume of the data in proGres v4 has expanded significantly.** The number of country operations in proGres v4 increased from 73 to 117, more than doubling the number of individual records from 10.2 million (5.8 million active individuals) to 21.4 million (12.3 million active individuals). As of February 2022, 96 percent of registered individual's records are in proGres v4. Previously, this data was distributed across multiple stand-alone servers in each country, so that data had to be compiled from multiple servers to calculate aggregate statistics for a country operation, and from different operations to get aggregate regional or global statistics.
- **Migration to proGres v4, combined with the Dataport, maximizes the potential of UNHCR registration data by providing a 'one-stop shop' for UNHCR users to access and analyze data in real time.** Now, users can generate aggregate statistics quickly and easily, using numerous indicators, filters, and breakdowns, and can monitor trends in real time at country, regional and global levels. At headquarters, the Statistical Unit uses the Dataport for analysis and reporting, and senior management

can generate their own reports and data visualizations. In the regional bureaus, Data Identity Management and Analysis units (DIMAs) carry out statistical reporting in real time to monitor trends, conduct oversight, and respond to external requests for information. Operations staff can monitor the number of people in need in each location and plan assistance programs, as well as build custom dashboards using a simple 'drag and drop' functionality.

- **Staff can now access the same individual record at the same time, regardless of where they are, improving efficiency and data quality.** Previously, records had to be transferred back and forth, leading to significant issues with data quality and delays to records becoming available. Now, a caseworker working on a resettlement case in Nairobi can access a refugee's record at the same time as a Protection Officer in the Kakuma refugee camp assesses their special needs while updating the same record.
- **Regional and country operations can monitor and improve their own data quality using tools and dashboards developed by the project.** 31 Key Performance Indicators (KPIs) for data quality have been defined and dashboards developed that enable UNHCR staff to identify data quality issues and drill down to the specific records that need to be corrected or updated. Data quality reports can also be generated during refugee verification exercises, as was done during the verification exercises carried out in Uganda and South Sudan in 2021.
- **The Dataport has been configured to extract, transform, and deliver registration data from proGres v4 to the UNHCR Raw Internal Data Library (RIDL).** The methodology for pre-anonymizing registration data in advance of this process has been established and tested, as has the staging database for implementing it.
- **The Dataport allows users to compare current and historical data.** This new functionality allows case management users to see the evolution of case information over time.
- **proGres v4 data has been used by operations and partners to inform household survey sampling frames,** so that much-needed socioeconomic data can be collected on representative samples of refugee populations. Integration with the KoBo Toolbox means that data from proGres (such as demographic and location data) can be brought together with recent socioeconomic data for combined analysis. This complements the improvements that have been made to proGres socioeconomic data fields; and the further enhancement of proGres to capture more socioeconomic data during registration exercises continues to be explored.

## CHALLENGES

During implementation of the project, the team found creative solutions to several challenges, and are still working on others:

- **Ensuring individual records cannot be re-identified from aggregate statistics generated from the Dataport.** To protect the confidentiality of personal data, UNHCR users are assigned one of two levels of access to the Dataport: (1) existing proGres users can generate aggregate statistics only for data in proGres that they already have permission to see; and (2) Excel users with a UNHCR account can generate high-level statistics from proGres data that has been pseudonymized (that is, from which direct identifiers and other sensitive information have been removed).
- **Ensuring the consistent performance of the Dataport as millions of records were migrated into the system.** The Dataport platform was moved to the cloud in January 2021, which increased the performance of the system as millions of records were contained in the system and tens of thousands of queries run each week.
- **Partner government users do not yet have direct access to the Dataport.** Government partners can log in securely and access records in the proGres case management system according to their data sharing agreements with UNHCR. However, the technical architecture currently prohibits direct access to

the Dataport environment by external users. The team found a creative solution to provide government partners with access to the reports generated by the Dataport system by creating a linked aggregate data source that could generate reports from within proGres itself. So, while government users can't access the Dataport directly, they can still benefit from the technology while sensitive data and privacy are protected.

- **The comprehensiveness of data (measured by the ‘Completeness Factor’ KPI) can be low.** Each operation specifies the data fields that must be completed when individuals are recorded in proGres, including seven data fields that are mandatory across all UNHCR operations. The number of required fields reflects a trade-off between the speed and cost of registration and verification exercises and data needs. Over time, individual records expand as UNHCR staff have further interactions with the individual and additional data is added (such as that needed for refugee status determination or resettlement, or to reflect changed circumstances or evolving needs during verification). To improve completeness during verification exercises, quality assurance thresholds were introduced into the document issuance process in proGres to improve data completeness.
- **The current data quality tools provide only current, not historical information.** The team is developing a scoring mechanism that would help to track data quality improvements over time.
- **Some measures of data quality are difficult to improve, for example, the proportion of data that is up-to-date.** There are still subsets of UNHCR data where there hasn't been a physical or virtual interaction with individuals to update information—an issue that has been amplified during the pandemic. Updating information is more important for socioeconomic data fields, like special needs, education, and occupation, which can change over time, whereas biographic information tends to be static.
- **The transfer of data from proGres to RIDL is still a manual process.** While the pipeline has been established and tested, including pre-anonymization and curation procedures, the automation is still outstanding. Before migration can be automated, the team needs to resolve remaining governance and data protection issues.

## LESSONS LEARNED AND JDC'S ROLE

The project team found engagement with JDC ‘incredibly helpful’. Generous funding from JDC has been critical to ensure significant and targeted enhancements to PRIMES, over and above the usual work to keep the system running. The interest of the World Bank and the involvement of JDC has helped the team catalyze and advocate for systems improvements with senior managers in UNHCR, by emphasizing the tangible benefits of these for UNHCR operations and the people that UNHCR protects and assists. Previously this kind of work was perceived as “just a database upgrade [that happens] in the background”. The project has also supported the culture change that is occurring in UNHCR to embrace a wider and deeper data agenda with more emphasis on socioeconomic data and the opportunities it provides. Quarterly reviews with the JDC focal points also helped to ensure that the project achieved its objectives.

READ MORE ABOUT OUR ACTIVITIES HERE:  
<https://www.jointdatacenter.org>