

Democratizing Data: Insights and Lessons on Standardization and Interoperability as a Foundation for Humanitarian Action

Oscar Maria Caccavale, Valerio Giuffrida, Anna Ong and Nynne Warring (World Food Programme, Research, Assessment and Monitoring Division)

Abstract

Dwindling resources and increasing needs call humanitarian organizations to do more with less. Technological innovation can support this effort by accelerating how we create evidence from the data we collect. In doing so, we must ensure that non-sensitive data is openly available and adheres to the Human Rights-Based Approach to Data principles. Currently, humanitarian action is informed by data that is often not collected in a standardized and comparable manner, nor managed, stored and shared transparently and efficiently. To address this challenge, WFP has developed a data ecosystem with built-in interoperability. This entails an end-to-end data pipeline that builds and deploys standardized surveys, uploads data into a centralized database and stores cleaned datasets in a micro data repository. This unlocks the possibility for automated analyses and easy visualization of results while enabling increased comparability and transferability of data. Market data has been used as a first test case of this ecosystem, resulting in a 24-hour turn-around after the data collection to visualization of findings.

Doing more with less

The gap between humanitarian needs and the resources available to address these needs continues to widen. This presents an urgent challenge for humanitarian actors, demanding increased precision and efficiency in humanitarian response. Utilizing tailored new technologies have immense potential in improving the way data is used to build our collective evidence-base, on which timely, appropriate, and effective humanitarian action can be taken.

Only a decade ago, humanitarian operations were still in large part informed by enumerators travelling to the field and conducting pen and paper surveys. Now, technological advances have unlocked novel ways of collecting information and digitized the collection process itself, contributing to an abundance of data and information. However, an expansion of data has not necessarily translated to equally abundant actionable evidence and knowledge. In fact, the potential of the data collected by humanitarian organizations has not yet been fully realized. Often, data is not collected in a standardized and comparable manner across time and space, nor it is managed and stored transparently and efficiently. Furthermore, this data is seldom available as open data for the humanitarian community and beyond, while the way the analysis is conducted often remains undisclosed. All in all, non-sensitive data “available to all - whether public or private, big or small, start-up or giant” (EU, 2020) should be at the foundation of a more democratic approach to the data itself. Additionally, interoperability of data across agencies is critical, which refers to using and re-using existing standards, not

overlooking metadata, using common classifications wherever possible, publishing data in machine-readable formats and ensuring that standards are user-driven (Steele, Orrell, & Anderson, 2017). Taken together, these challenges limit the scope of building consistent evidence over time. Building data streams and systems that can communicate (OCHA, 2022) is crucial for a more efficient and effective creation of actionable evidence.

At WFP, market, household, and monitoring survey data are regularly collected to guide the organization’s response to combat food insecurity. This brief reflects on WFP’s practical experience in tackling these challenges in the context of its food security data. Specifically, it details how WFP has leveraged data standardization and improved governance and illustrates data can be dynamically translated into an evidence base to inform humanitarian response.

How interconnected processes and platforms enable efficiencies in generating evidence

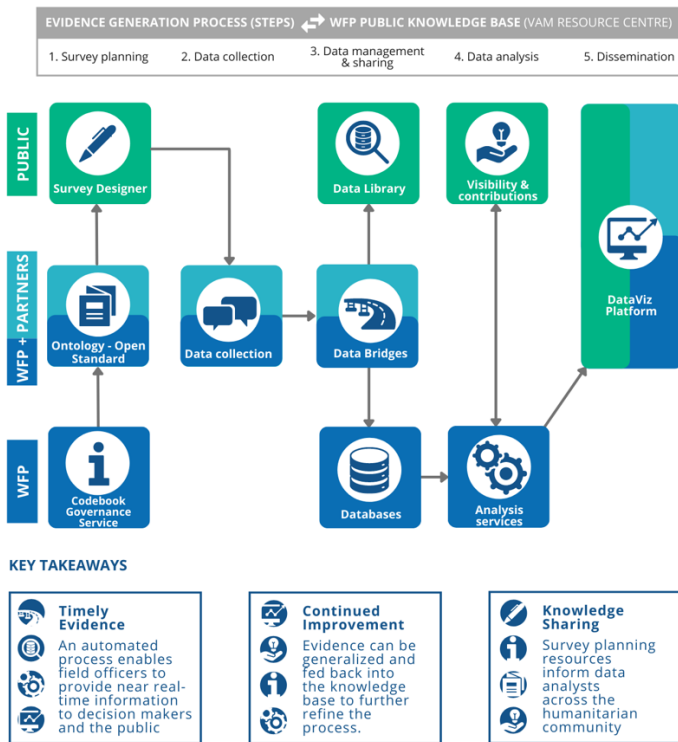
WFP has a vast network of food security analysts across 80 country offices under the technical supervision of the Research, Assessment and Monitoring (RAM) Division. RAM oversees data generation for vulnerability analysis, program outcome monitoring and the development of digital platforms and analytical tools. In implementing the Generic Statistical Business Process Model, RAM identified different steps, from planning of surveys to final dissemination of results (UNECE, Eurostat, & OECD, 2019), each in need of specific technological solutions. For each step of the model, WFP

has designed solutions that optimize the use of the data collected and developed an ecosystem with built-in interoperability.

However, developing such an ecosystem poses two challenges: first, the technical challenge of implementation and second, the design of an evidence generation ecosystem that adheres to a Human Rights-Based Approach to Data as part of the 2030 Agenda in terms of the principles of participation, data disaggregation, self-identification, transparency, privacy, and accountability (OHCHR, 2018; Pontificia Academia, 2020). This approach extends the principle of “do no harm” to the data itself, its systems and algorithms (UN Global Pulse, 2016; IASC, 2021).

The figure below illustrates how WFP RAM has been shaping this data ecosystem in line with the steps of 1) planning, 2) data collection, 3) data management and sharing, 4) analysis, and 5) dissemination.

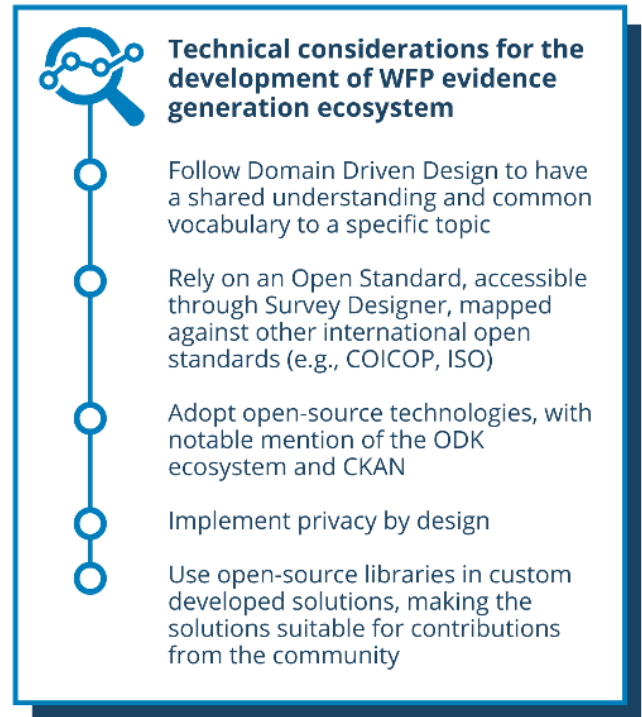
Figure 1. Steps of WFP RAM shaping the data ecosystem



The planning phase (1) pins down the specific questions, methodologies and indicators that will eventually address the research question. Years of experience in surveys across different countries have been condensed in the Survey Designer, a web-based application developed with a participatory process that maps all the questionnaires in use through a Codebook Governance Service, so that users can drag and drop from a range of pre-built standardized components to create their own survey instrument in a matter of

minutes. This means that the survey created is tailored to the needs of the user, while also standardized and avoiding risk of programming errors or miscoding indicators.

Figure 2. Technical considerations for the development of WFP evidence generation ecosystem



Once a standardized survey instrument has been produced using the WFP Survey Designer, it is ready for deployment in the field. At this stage, data collection (2) is typically conducted through open-source mobile data collection platforms. After data collection is completed, the data is uploaded from field offices and stored in a centralized database using a web-based WFP application called Data Bridges. This app instantly performs automatic quality checks and data validation (e.g. the detection and potential correction of outliers) while providing structured storage for the data. This storage is then accessible in two ways: (a) programmatically between applications and (b) through a user’s interface. The cleaned data and the meta-data are eventually stored in a web-based repository called WFP Data Library, from where data can be explored without the ‘traditional’ lengthy processes of recoding, cleaning, and reorganizing the dataset (3).

A substantial advantage of this end-to-end data pipeline (Survey Designer > Data Bridges > Data Library) is that it unlocks the possibility for more complex and automated analyses (4), which can be further customized depending on the context. All the analysis

services developed throughout this pipeline are transferrable and re-usable, they can be shared in a public repository to help inform the work of others who can provide contributions to improve the existing scripts. Finally, analytical findings and reports are visualized in WFP's DataViz platform, making the data publicly available - within the limits of data protection and privacy (OCHA, 2021) - and the results replicable (5). Simultaneously, the knowledge generated throughout this process - whether methodological guidance on indicators, training materials or templates in support of any of these activities - is stored in WFP's public knowledge base, known as the VAM Resource Centre.

Market data: an example of end-to-end data pipelines

An example of the model described above is demonstrated by the world's largest publicly available food price database, which WFP has managed since 2009. Prices from local markets are collected on a weekly or monthly basis by WFP and its partners. The data is uploaded by WFP country offices through Data Bridges, which allows users to create their own schema that matches the country's existing data structure, building on a standardized ontology so that the data can be stored in the centralized database. Once data is uploaded, it is automatically processed and analysed to develop price trends, issue alerts for spiking prices and provide a forecast for the next six months. Data and results are available in the DataViz platform and can be consumed through an application programming interface (API).

Another example is the Market Functionality Index (MFI), an initiative launched in 2020 (WFP, 2020) to assess the transparency and predictability of the features ruling supplies and prices in local markets through a composite index that summarizes nine key market dimensions. The MFI was designed from the outset to follow the standardized process while offering an end-to-end data pipeline, together facilitating its rapid uptake. The MFI has been used in 52 countries to date, gathering information from 2,421 markets and 45,141 traders, becoming an approach used also by several Governments working with WFP. In practice, the MFI involves a standardized survey deployed by enumerators that interview traders and make observations in local markets. Then, the data collected is uploaded and verified through Data Bridges, after which automated analysis scripts are executed to create the final indices and allow visualizations. Overall, after data is collected, less than 24 hours is required to complete the process. In contrast, 'traditional' market assessments use non-standard survey instruments,

which poses challenges in tracing changes over time and leading to lengthy reporting processes. The end-to-end data pipeline enables a much stronger degree of comparability, replicability, and transparency of market analyses. While the MFI data is a public good, the platform itself is available to interested organizations and governments should they want to collect the same market data.

What comes next?

This business process model and its implementation in the market data streams serve as a useful illustration of the power of standardization and interoperability in democratizing data. Moving forward, WFP intends to expand this ecosystem to support the needs of different business domains and increase public access to non-sensitive data through APIs. However, changing the way large organisations manage data does not happen instantaneously, and the process is not rolled out at scale. The experience to date shows that this not only requires the right technological solutions, but it also must start from a foundation of shared understanding, joint business objectives and a common vocabulary, within and across organisations.

The platforms and processes described in this brief offer promising potential for partnerships. Partners can build on the same open-sourced technology, use existing ontologies and thus achieve interoperability across analytical platforms and organizations (UN Secretary General, 2020). Building on this model would allow users the flexibility to adopt the system in parts or fully, without losing the opportunity to curate data in a coherent workflow relevant to their existing infrastructure.

Lastly, these advances need to happen responsibly (OCHA, 2020). On one hand, for organizations like WFP, supported by private and public actors and empowered with a life-saving mandate, data democratization is no longer optional. Large organisations cannot act as gatekeepers of the essential information needed to protect the lives of the most vulnerable and advance the Sustainable Development Goals. Open source and open data systems can dramatically reduce the time needed to act, in particular when emergencies strike. At the same time, automated data processes should never dictate or bias decision making. If the Human Rights-Based Approach to data is ignored, we risk unwittingly hiding human shortcomings and biases in complex processes. The more complex the solutions, platforms and analyses, the more transparency and replicability is needed.

References

- EU. (2020). *A European strategy for data*. Brussels: European Commission.
- IASC. (2021). *Data Responsibility in Humanitarian Action*. Inter-Agency Standing Committee. United Nations. Retrieved from <https://interagencystandingcommittee.org/system/files/2021-02/IASC%20Operational%20Guidance%20on%20Data%20Responsibility%20in%20Humanitarian%20Action-%20February%202021.pdf>
- IMF. (2022). *World Economic Outlook – Update. Rising Caseloads, a Disrupted Recovery, and Higher Inflation*. International Monetary Fund.
- OCHA. (2020). *Humanitarian Data Ethics*. Centre for Humanitarian Data. UN OCHA. Retrieved from https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/a39c4674-d976-42ae-af31-e037c1ffb9fe/download/guidance_note_ethics.pdf
- OCHA. (2021). *Data Responsibility Guidelines*. Centre for Humanitarian Data. UN OCHA.
- OCHA. (2022). *State of Humanitarian Data 2022*. Centre for Humanitarian Data. UN OCHA.
- OHCHR. (2018). *A Human Rights-Based Approach to Data. Leaving No one behind in the 2030 Agenda for Sustainable Development*. Office of the High Commissioner for Human Rights. United Nations. Retrieved from <https://www.ohchr.org/documents/issues/hrindicators/guidancenoteonapproachtodata.pdf>
- Pontificia Academia. (2020). *Rome Call for AI Ethics*. Rome: Pontificia Academia.
- Steele, L., Orrell, T., & Anderson, B. (2017). *The frontiers of data interoperability for sustainable development*. Echelon Creative .
- UN Global Pulse. (2016). *Risks, Harms and Benefits Assessment, Level 1*. UN Global Pulse.
- UN Secretary General. (2020). *Data Strategy of the Secretary General for Action by Everyone, Everywhere*. New York: United Nations.
- UNECE, Eurostat, & OECD. (2019). *Generic Statistical Business Process Model*. Retrieved from UNECE Wikistat web site: <https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model>
- WFP. (2020). *Market Functionality Index (MFI)*. Retrieved from <https://www.wfp.org/publications/market-functionality-index-mfi>