THE STATE OF OPEN HUMANITARIAN DATA 2024

MARKING TEN YEARS OF THE HUMANITARIAN DATA EXCHANGE

TABLE OF CONTENTS

1. INTRODUCTION 4

2. KEY MESSAGES 6

3. DATA HIGHLIGHTS 8

4. TEN YEARS OF HDX 10

5. GLOBAL OVERVIEW 14

6. COMPLETENESS BY LOCATION, CATEGORY AND SUB-CATEGORY 16

7. COMPLETENESS BY LOCATION AND CATEGORY 17

8. COMPLETENESS BY LOCATION AND SUB-CATEGORY 21

9. COUNTRY DEEP-DIVE: COLOMBIA 23

10. CONTRIBUTING ORGANIZATIONS 24

11. ORGANIZATION DEEP-DIVE: THE WORLD FOOD PROGRAMME 25

12. CLIMATE IMPACT DATA: TROPICAL CYCLONES 27

13. CONCLUSION 29

ANNEX A: DATA GRID SUB-CATEGORY DEFINITIONS 30

ANNEX B: INCLUSION OF DATA IN THE DATA GRIDS 32

ACKNOWLEDGEMENTS

This report was produced in February 2024 by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data, which manages the Humanitarian Data Exchange (HDX) platform. OCHA thanks all of the organizations that have shared data through HDX, the donors who have supported this work, and the HDX users who are committed to ensuring that humanitarian response is data driven. For additional information, contact the Centre for Humanitarian Data at centrehumdata@un.org.

LIST OF ABBREVIATIONS

ACLED	Armed Conflict Location & Event Data Project
ADAM	Advanced Disaster Analysis & Mapping
CODs	Common Operational Datasets
FTS	Financial Tracking Service
GIS	Geographic Information System
HDX	Humanitarian Data Exchange
HOT	Humanitarian OpenStreetMap Team
HRP	Humanitarian Response Plan
IASC	Inter-Agency Standing Committee
IDP	Internally Displaced Person
IOM	International Organization for Migration
IPC	Integrated Food Security Phase Classification
NGO	Non-Governmental Organization
OCHA	United Nations Office for the Coordination of Humanitarian Affairs
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNHCR	United Nations High Commissioner for Refugees
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund
UNRWA	United Nations Relief and Works Agency for Palestine Refugees in the Near East
WFP	World Food Programme
WHO	World Health Organization

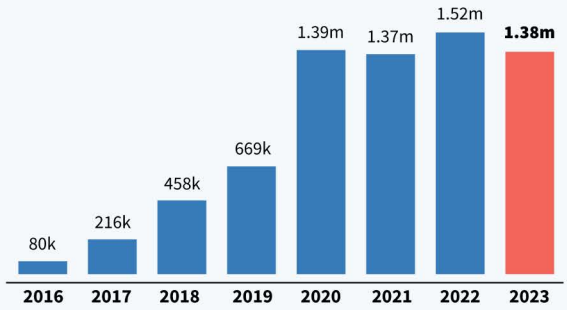
1. INTRODUCTION

In 2023, the demand for humanitarian data reached a record level as the world contended with the effects of war, climate shocks, food insecurity, displacement and disease in crises ranging from Gaza to Sudan. At the same time, data availability across priority humanitarian operations remained steady. Organizations continued to produce and share valuable data about the needs and response to the world's most vulnerable populations, even in the face of restricted access due to violent conflict.

Our insights into data availability and use come from managing the Humanitarian Data Exchange (HDX), an open platform for finding and sharing data across crises and organizations.[1] In 2023, HDX was used by 1.4 million people in 230 countries and territories, maintaining a similar number of users over the past several years.

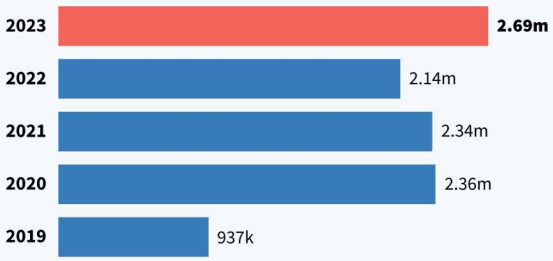
The 237 organizations on HDX contributed 1,800 new datasets, expanding the total catalogue to more than 20,000 datasets. Together, they were downloaded over 2.7 million times, a notable 25 percent increase compared to 2022.

HDX unique users 2016-2023



millions I Creation date: 04 January 2024

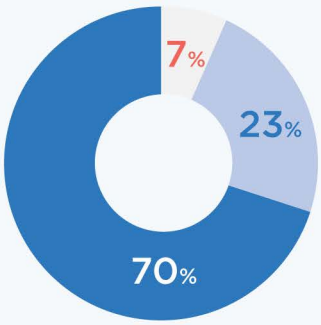
HDX dataset downloads 2019-2023



millions I Creation date: 05 January 2024

Although HDX includes data about all countries in the world, our priority focus is on locations with humanitarian response plans (HRPs).[2] At the start of 2024, we estimate that 70 percent of relevant, complete crisis data is available across 23 humanitarian operations, based on the analysis of the HDX Data Grids (see criteria below). If we add the data that is relevant but incomplete, the total is 93 percent. This leaves 7 percent of categories with data that does not meet the criteria or with no data.

The 23 Data Grids include 397 unique datasets, with a range of 12-20 datasets per location. Datasets included in the Data Grids are downloaded over five times more than the average dataset on HDX.[3]



TOTAL PERCENTAGE DATA COMPLETE

70%

TOTAL PERCENTAGE DATA INCOMPLETE

23%

TOTAL PERCENTAGE NO DATA

7%

Data completeness: Complete Incomplete No Data

[1] https://data.humdata.org/

[2] HRPs are prepared by UN Humanitarian Country Teams in locations where there is an ongoing humanitarian emergency. HRPs are generally prepared annually, and outline an overall strategy and specific activities for each humanitarian cluster or sector.

[3] In 2023, Data Grid datasets were downloaded, on average, 677 times each whereas the average dataset on HDX was downloaded 129 times.

The State of Open Humanitarian Data contains details on the data available for each location, category and sub-category covered in the Data Grids as of 31 December 2023. It includes a country deep-dive for Colombia, which has seen significant gains in data completeness over the past five years, and it showcases the contribution of the World Food Programme (WFP), a longstanding partner of HDX. It also provides an example of how climate data is being used to predict the impact of typhoons in the Philippines and Fiji and trigger anticipatory action.

Although this report covers data activity in 2023, it is published in 2024 which marks the ten year anniversary of HDX. Launched in 2014 with just 800 datasets and a dozen organizations, HDX is now an anchor in the humanitarian data ecosystem. In our piece on 'Ten Years of HDX' we reflect on the big shifts in humanitarian data since we first started our work and consider what the future may bring.

Data Grid Criteria

The HDX Data Grids narrow the focus within each HRP location to a limited set of foundational data needed to understand a humanitarian context. They provide a comparable way to assess data availability across locations and categories and are the basis for the analysis in this report.

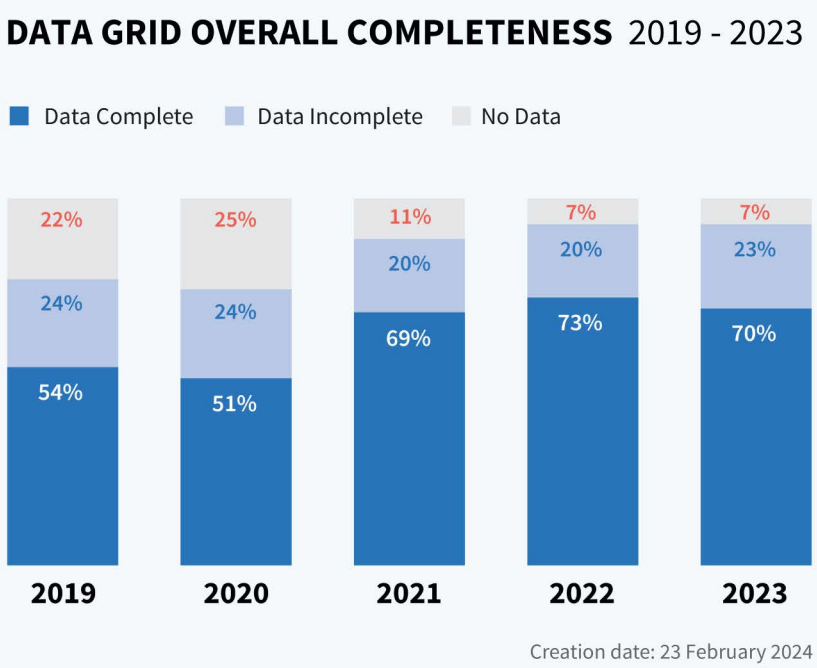
The Data Grids include six categories: affected people; coordination and context; food security and nutrition; geography and infrastructure; health and education; and population and socio-economy. (See Annex A for definitions.)

Data may be included in a Data Grid if it is relevant to the category and sub-national. The data is considered 'complete' if it has broad geographic coverage, is shared in a commonly used format, and is up-to-date. If any of those criteria are not met, then it is considered 'incomplete'. The sub-category is complete if it includes at least one dataset that is complete. If the sub-category contains only incomplete datasets, then that sub-category is considered incomplete. A sub-category is empty if no data meets the above criteria or the data does not exist on HDX. (See Annex B for the Data Grid criteria and curation process.)

In 2023, HDX included Data Grids for 23 locations and 20 sub-categories. The number of locations decreased by two from 2022 given that Iraq and Libya no longer had response plans in 2023. There were no changes to the number of categories or sub-categories.

2. KEY MESSAGES

- Demand for humanitarian data **reached a record level** in 2023 as the world contended with the effects of war, climate shocks, food insecurity, displacement and disease in crises ranging from Gaza to Sudan.
- In 2023, HDX was used by **1.4 million people in 230 countries and territories**. Around **2.7 million datasets were downloaded** throughout the year, a 25 percent increase compared to 2022.
- Twenty-four organizations joined HDX in 2023, bringing the total to 237 organizations.
- The availability of core data for priority humanitarian operations **remained steady**, despite **persistent data gaps for Syria and Venezuela** and with health and education facilities.
- We estimate that 70 percent of relevant, complete crisis data is available across 23 locations with humanitarian operations, based on the analysis of the HDX Data Grids.
- The 23 HDX Data Grids include 397 unique datasets, with a range of 12-20 datasets per location. These datasets are **downloaded over five times** more than the average dataset on HDX.
- Data covering Yemen, Nigeria, Ethiopia and Ukraine was the most downloaded. **Administrative boundaries were the most popular** datasets across all locations.
- The most complete Data Grid sub-categories are conflict events, funding, administrative boundaries, refugees and persons of concern, and baseline population owing to the work of ACLED, OCHA, UNHCR and UNFPA, respectively.
- **Health and education facilities are the most challenging** sub-categories to complete, either because the data is out of date or because of the sensitivity surrounding such information in conflict settings.
- As we have seen in locations and sub-categories that lost ground in 2023, a critical element to data completeness is timeliness – **data needs to stay up-to-date for it to remain useful** and support decision making.
- In 2024, HDX will mark its ten year anniversary. Bringing data together across crises and organizations has undoubtedly created efficiency and **lays the foundation for the AI advances to come**.
- A priority for 2024 will be **increasing access to climate impact data**, which is essential for predicting the future impacts of hazards on vulnerable populations and is a prerequisite for anticipatory action.
- We are also working to make **key figures, such as the number of people in need and the number of people displaced, available programmatically** for use in applications, search engines and business intelligence tools.



We call on partners to generate and **share data that is missing** or incomplete for many humanitarian crises, including:

- **Access constraints** on the delivery of humanitarian assistance (potential sources: national governments, access working groups, OCHA).
- **Climate impact data** covering the previous 10 years of hazards and their impact on people, infrastructure and vegetation (potential sources: national governments, disaster risk management agencies, humanitarian partners).
- Global **acute malnutrition prevalence** rates by administrative division (potential sources: UNICEF, the Nutrition Cluster).
- The location of **education facilities** (potential sources: national governments, UNICEF, UNESCO, the Education Cluster).
- The location of **health facilities** (potential sources: national governments, the Health Cluster, WHO).

In cases where data is sensitive, organizations can use HDX Connect[4] to share only the metadata and make the underlying data available by request. In 2023, the Data Grids included seven HDX Connect datasets: two for Afghanistan, one for Burkina Faso, two for Niger, one for Syria and one for Yemen. Alternatively, an incomplete or aggregated version of a dataset that removes sensitive information (such as locations) can be shared publicly.[5]

[4] HDX Connect datasets still contribute to the completeness of a Data Grid. Learn more: https://centre.humdata.org/a-new-call-to-action-sharing-the-existence-of-data/

[5] Learn more about our process for statistical disclosure control: https://centre.humdata.org/learning-path/disclosure-risk-assessment-overview/

3. DATA HIGHLIGHTS

We acknowledge the invaluable contributions of all organizations that publish data on HDX, with special recognition for the **15 organizations** sharing data included in the Data Grids, which encompass **164 data sources**.

Trusted partnerships and focused advocacy efforts led to many new or updated datasets in the Data Grids in 2023, including:

- Information on the **price of basic commodities in Gaza** before and after hostilities, contributed by the Palestinian Central Bureau of Statistics.[6]
- The number of newly **displaced households in Yemen** disaggregated by district and by their needs, contributed by IOM.[7]
- Who-is-doing-what-where by cluster and the number of **affected people reached by oblast in Ukraine**, contributed by OCHA Ukraine.[8]
- Sub-national sex and age disaggregated **population statistics for Chad** with updated projections for 2024, contributed by UNFPA.[9]
- The number of people in need of **humanitarian assistance in Afghanistan** from 2017 to 2024, produced by humanitarian partners and contributed by OCHA Afghanistan.[10]
- Analysis of **humanitarian access in Ethiopia** at the woreda level, produced in consultation with partners in the Access Working Group and contributed by OCHA Ethiopia.[11]

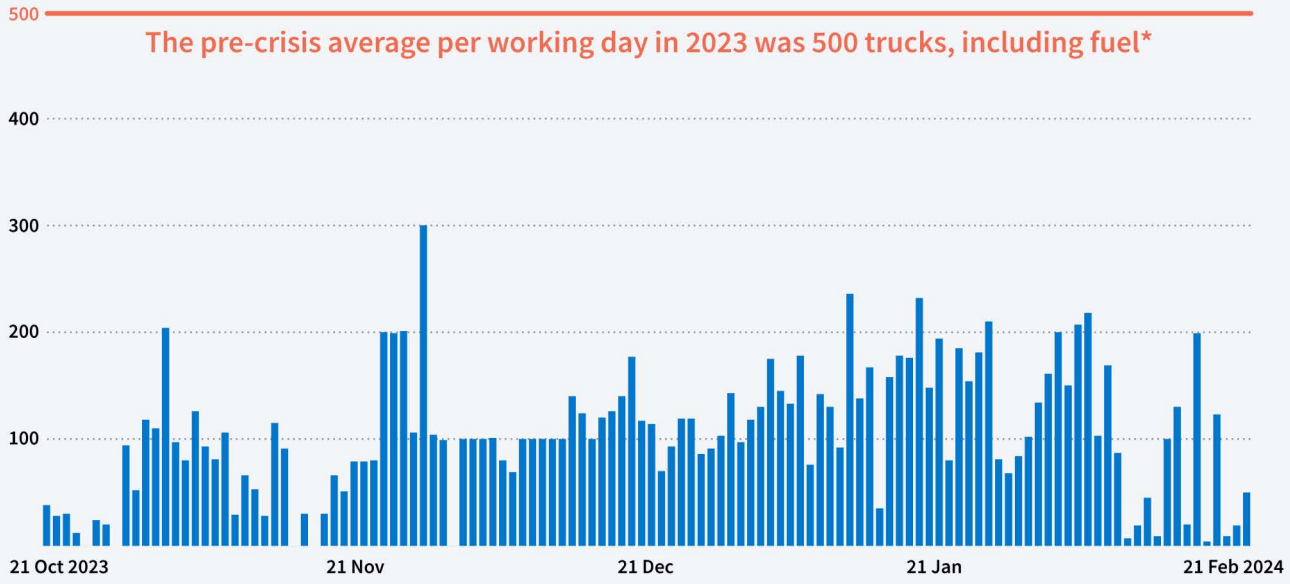
Building footprint imagery from Google Research

[6] https://data.humdata.org/dataset/state-of-palestine-price-of-basic-commodities-in-gaza
[7] https://data.humdata.org/dataset/yemen-displacement-daily-tracking-idps-returnees-iom-dtm
[8] https://data.humdata.org/dataset/ukraine-who-does-what-where-3w
[9] https://data.humdata.org/dataset/cod-ps-tcd
[10] https://data.humdata.org/dataset/afghanistan-humanitarian-needs-overview
[11] https://data.humdata.org/dataset/ethiopia-humanitarian-access

Beyond the Data Grids, HDX featured dozens of **new, high-value datasets** in 2023. Examples include:

- The number and type of **aid trucks and their cargo accessing Gaza** through Rafah and Kerem Shalom crossings, compiled and shared by the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA).[12]
- Alerts following sudden onset humanitarian emergencies with **near real-time risk and impact information**, shared by WFP Advanced Disaster Analysis & Mapping (ADAM).[13]
- A World Risk Index assessing the **vulnerability of 193 countries to humanitarian disasters** resulting from extreme natural events and the adverse effects of climate change, shared by the Institute for International Law of Peace and Armed Conflict.[14]
- **Movement distribution** showing how far people move away from the area where they live on a daily basis with global coverage, created and shared by Meta.[15]
- The **footprint of buildings** derived from high-resolution satellite imagery following sudden onset disasters in Afghanistan, Japan, Libya, Morocco, and Reunion Island, compiled and shared by Google Research.[16]

Supply trucks entering the Gaza Strip



Figures may change as new data becomes available *OCHA UNRWA data via HDX Creation date: 23 February 2024

[12] https://data.humdata.org/dataset/state-of-palestine-gaza-aid-truck-data
[13] https://data.humdata.org/organization/advanced-disaster-analysis-mapping
[14] https://data.humdata.org/dataset/worldriskindex
[15] https://data.humdata.org/dataset/movement-distribution
[16] https://data.humdata.org/organization/google-research

4. TEN YEARS OF HDX

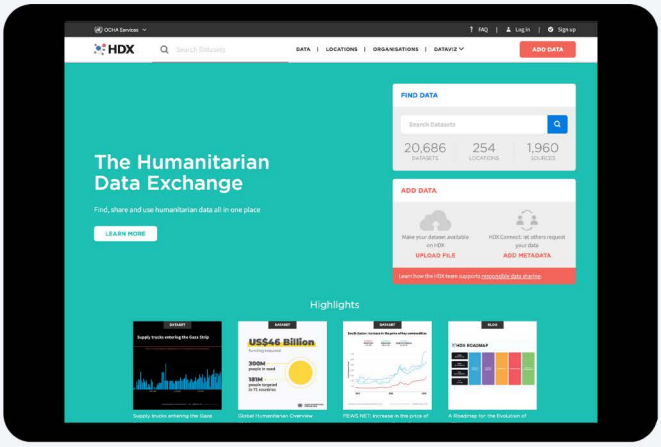
The Humanitarian Data Exchange (HDX) was publicly launched at the Open Knowledge Festival in Berlin in July 2014, featuring 800 datasets shared by over a dozen organizations. Fast forward ten years and HDX is an integral platform for the humanitarian data community: 237 organizations are sharing over 20,000 datasets, covering every active humanitarian crisis, with users in 230 countries and territories.

The concept behind HDX – open data aggregation – was not a new idea. HDX was formed at the peak of the open data movement. Actors spanning multiple sectors had come together to challenge the conventions of information sharing and demanded a departure from the era of closed data systems. They advocated for data to be open: freely available for anyone to use, reuse, and redistribute without restrictions, subject only to attribution.[17] The World Bank's open data initiative was a key inspiration for HDX, providing access to data that had previously been behind a paywall.[18]

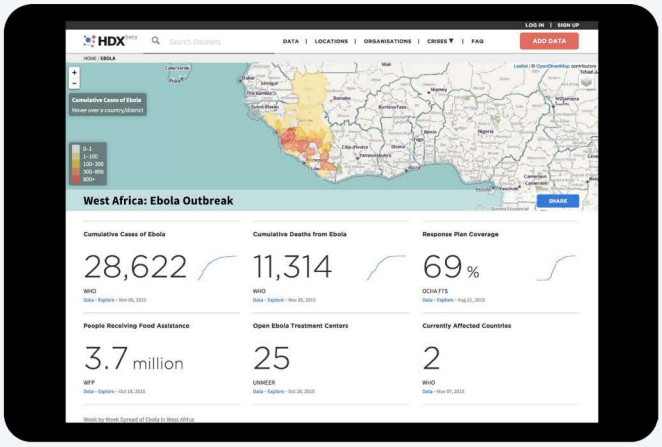
Before HDX, the humanitarian data landscape was fragmented and opaque. Humanitarian data was hard to find, and there was a lack of clarity around whether data could be trusted or reused: Was it the latest version? Was there any metadata about the source and data collection methodology? Was there a license? In the humanitarian sector, getting access to data, and answers to questions, depended to a large extent on personal connections.

Enter HDX

OCHA created HDX with the goal of making data easy to find and use for analysis, in line with its coordination mandate. Our theory of change was that, by providing open and accessible data, HDX could lower the barrier to entry for data use, in turn improving data-informed decision making in humanitarian response.



HDX Homepage, February 2024



West Africa Ebola Outbreak Crisis Page, 2014

While the willingness of humanitarians to share data preceded the open data movement, data was not commonly shared in a manner that maximized opportunities for reuse. To realize our ambition, HDX had to drive behavior change, create incentives and reduce friction for data sharing across the sector.

A clear source of friction, for example, was the dependency on PDF formats for reporting on crises. In 2014, shortly after the launch of HDX, the Ebola outbreak in West Africa was intensifying. Timely information on the location of Ebola cases and deaths was essential to limit the spread of the disease. This data was made available through daily situation reports published in PDFs by WHO. We manually extracted the data from these reports into a machine-readable file that was shared on HDX and sourced to WHO. The Ebola cases and deaths data became the most popular dataset on the platform, proving the value of open data. *The New York Times* used it to create a visual of the outbreak that ran on their homepage, reaching millions of readers.

[17] https://opendatahandbook.org/guide/en/what-is-open-data/
[18] https://data.worldbank.org/

The Ebola crisis was a test case for a number of data services that HDX continues to offer today. We created our first crisis page to help non-technical users understand trends across locations and sources. Around this same time, WFP started sharing their food price data for Ebola-affected countries in West Africa on HDX. In exchange, we visualized their data and created the first custom organization page.[19] These services – a crisis page, a branded organization page and data visualization – incentivized data sharing by other organizations and drove adoption of the platform.

> "HDX was the first time datasets were brought together in an easy and consumable way for humanitarian purposes. The data was clean and from a source that could be trusted. By placing data into one central space, users can now connect different sources of information to conduct comprehensive analysis. I believe HDX is one of the core reasons why humanitarians started using data."

Rohini Swaminathan
Head of Geospatial Support Unit, World Food Programme

Ten Years of Growth

HDX has seen significant growth over the last ten years, reaching around 130,000 unique users per month in December 2023. Global crises drive peaks in use, as we have seen with the COVID-19 pandemic, Ukraine war and hostilities in Gaza.

The top user locations have consistently included the United States, United Kingdom, Germany, Kenya, Bangladesh and the Philippines. In locations with humanitarian operations, Ethiopia, Nigeria and Ukraine have the most use. User growth has increased more in HRP locations relative to non-HRP locations over the past ten years.

Our early research in 2014 found that one of the key determinants for how much a user trusted a dataset was based on the organization it came from. We have since worked with hundreds of organizations to help them establish a data sharing presence on HDX.

As of early 2024, there are 237 active organizations sharing data on the platform.[20] Of those, 131 (55 percent) are global, 26 (11 percent) are regional, and 80 (34 percent) are national or local organizations.

Unique Users on HDX 2014-2024

HDX Users Growth in HRP Countries 2016-2023

[19] https://data.humdata.org/organization/wfp
[20] https://data.humdata.org/organization

Big Shifts

Data usage in the humanitarian sector has evolved significantly over the last ten years. For its part, HDX has led an important shift in driving the centrality of data in humanitarian response. We have also had a front row seat to the big shifts with data in the sector. We highlight a few below.

A shift to responsible data

The call for open data has evolved, in some cases, to 'share and protect' data, based on a more complex understanding of the relationship between data and humanitarian response. Over the last decade, as technological advances have made data management easier, humanitarian organizations have faced a new challenge: how to balance the benefits of widespread data sharing with the risks of disclosing potentially sensitive information about people in crisis.

At the time of HDX's launch, there was limited guidance on the management of sensitive data in humanitarian contexts.[21] By 2021, the first system-wide IASC Operational Guidance on Data Responsibility in Humanitarian Action[22] had been endorsed, outlining eight actions for the safe, ethical and effective management of data. This includes the development of Information Sharing Protocols, which govern data and information sharing in an operation.

HDX responded to this shift by introducing HDX Connect, a feature that enables organizations to share only metadata while the underlying data is available by request. We also introduced methods for disclosure control of datasets, such as needs assessments, that may have a high risk of re-identification of people.[23]

A shift to programmatic use

Humanitarian organizations have advanced in their technical capacity in recent years and the use of APIs[24] is growing. As a result, the scale and predictability of humanitarian datasets have greatly improved.

Sixty-eight percent of the 20,000+ datasets on HDX are shared programmatically through APIs, reducing the need to manually add and update individual datasets. This shift to using data pipelines is a sign of increased sophistication among our data partners. The same can be said for a majority of HDX users who want to access data through frictionless services rather than by going to the site to download files. We expect this trend toward automation to continue.

Although positive, this comes with the risk of further fragmentation of the humanitarian data ecosystem, overwhelming data managers and developers with too many systems to learn in order to access the data they need. HDX will leverage its role as a central platform to provide the most essential data in a common system, simplifying programmatic access to data. This new 'humanitarian API' will provide a consistent, standardized and machine-readable interface to query and retrieve data from a set of high-value indicators, starting with those in the HDX Data Grids.

A shift to insight

More humanitarians are using data in their daily workflows than ever before. However, access to data is only as powerful as the insight it provides. As such, conducting analysis is a crucial part of humanitarian data efforts and this work is set to accelerate in the coming years.

[21] Sensitive data is defined as data that, if disclosed or accessed without proper authorization, is likely to cause harm to any person, including the source of the information or other identifiable persons or groups, or a negative impact on an organization's capacity to carry out its activities or on public perceptions of that organization.

[22] https://interagencystandingcommittee.org/operational-response/iasc-operational-guidance-data-responsibility-humanitarian-action/

[23] https://centre.humdata.org/learning-path/disclosure-risk-assessment-overview/

[24] Application programming interfaces (APIs) enable two software components to communicate with each other using a set of definitions and protocols.

When HDX started, we supported data contributors by making a visual of one of their datasets. Then we began to combine dozens of datasets on HDX to create broader situational awareness of a crisis, such as with the data explorers for COVID-19 pandemic and the war in Ukraine.[25] Now we are using data and forecasts to predict the impact of a crisis and trigger action before it unfolds. These impact models rely on climate data, satellite imagery, and novel sources such as building footprint data from Google Research and wealth indices from Meta, all of which are available on HDX.

Advances in AI will enable analysis at a previously unattainable scale, supporting the shift towards proactive approaches in the humanitarian sector. AI has the potential to drive operational gains for humanitarians through enhanced efficiency in routine tasks like report drafting and data formatting. The scale and speed of digestion and analysis generated by AI tools could markedly increase the capacity for large-scale humanitarian analysis.

Search interest in AI over the past five years worldwide, via Google Trends.

The Future of HDX

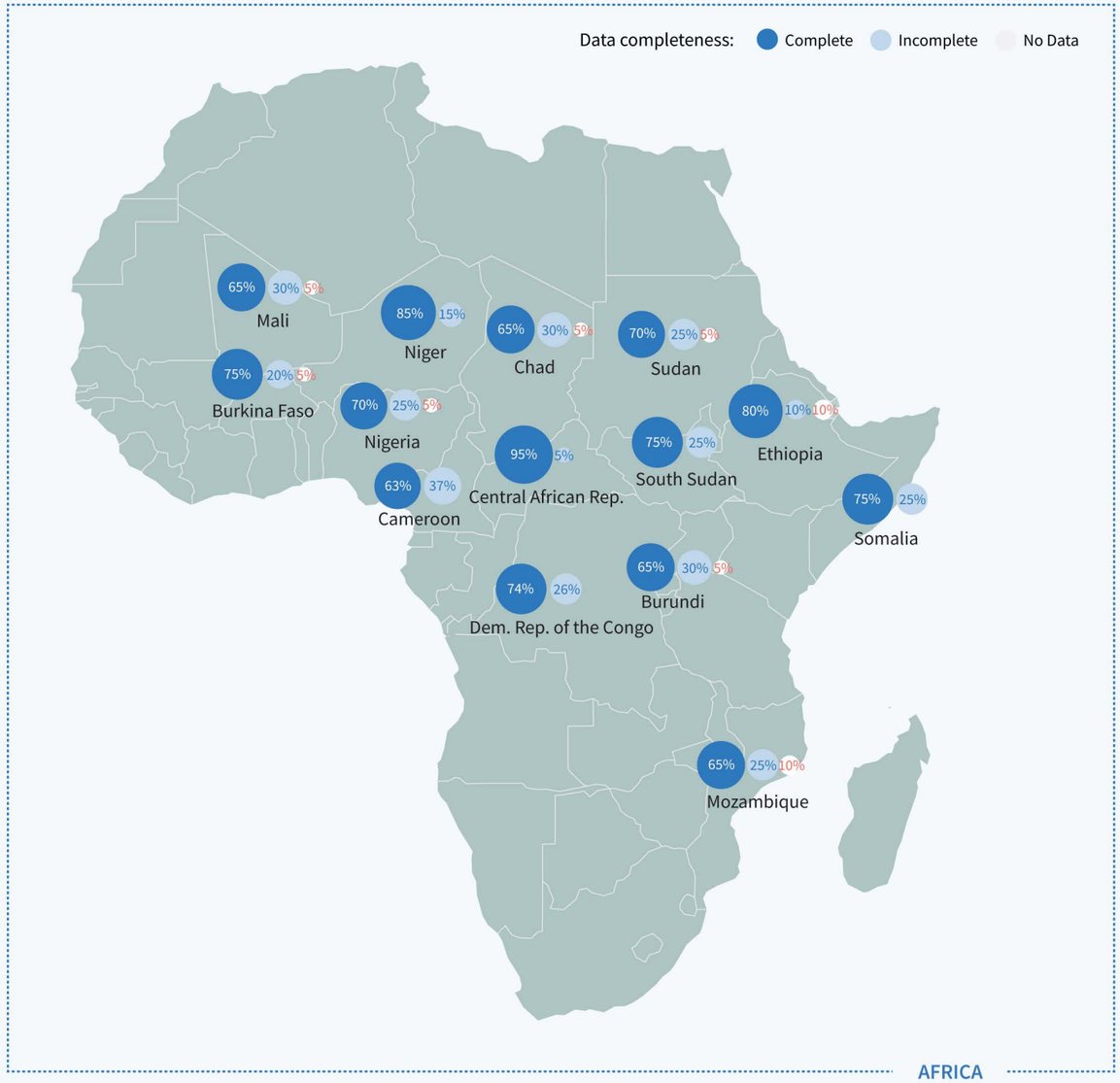
In contrast to ten years ago, data is now firmly on the agenda of the United Nations. The Secretary-General's Data Strategy,[26] created in 2020, is committed to unlocking "the data potential of the UN family to better support people and planet." Demand for better data and analysis has grown in the humanitarian sector, as it contends with unprecedented needs and insufficient funds. Over 300 million people are in need of humanitarian assistance in 2024.[27] In this context, data is more powerful than ever, and the stakes have never been higher.

The future of HDX will meet the changing needs of humanitarians, leveraging advances in emerging technologies to do so. Our product roadmap builds on the progress of the past ten years, and creates new opportunities for faster access to reliable data and insight about humanitarian crises.[28] We will evolve the HDX platform to be more sophisticated in its processing of data at scale with a continued focus on data quality.

Trusted partnerships with individuals and organizations in the humanitarian community will always be at the forefront of our work. Through HDX, and the Centre's broader objectives, we hope to drive another decade of increased use and impact of data in humanitarian response.

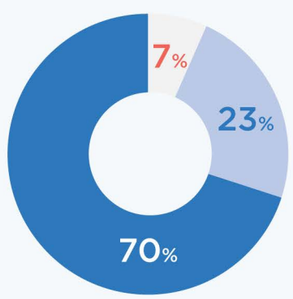
[25] https://data.humdata.org/visualization/ukraine-humanitarian-operations/
[26] https://www.un.org/en/content/datastrategy/index.shtml
[27] https://humanitarianaction.info/document/global-humanitarian-overview-2024
[28] https://centre.humdata.org/a-roadmap-for-the-evolution-of-hdx

5. GLOBAL OVERVIEW



The boundaries and names shown and the designations used on this map do not imply official endorsement or acceptance by the United Nations.
Percentages may not total 100 due to rounding.

NUMBER OF LOCATIONS	NUMBER OF CATEGORIES	NUMBER OF SUB-CATEGORIES	NUMBER OF CONTRIBUTING ORGANIZATIONS
23	6	20	15

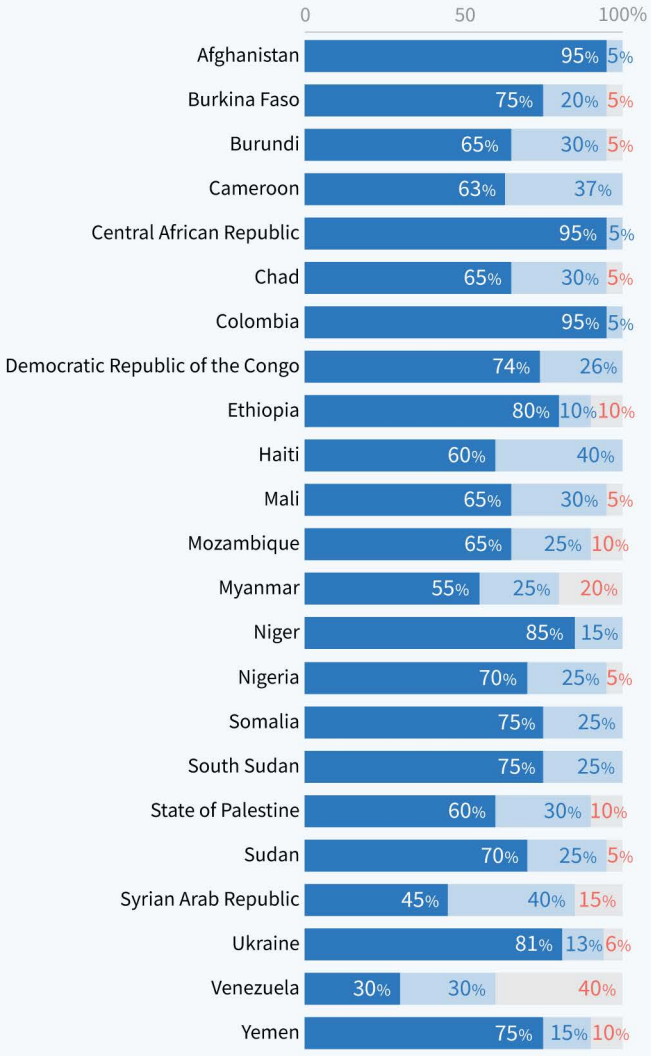


TOTAL PERCENTAGE DATA COMPLETE	TOTAL PERCENTAGE DATA INCOMPLETE	TOTAL PERCENTAGE NO DATA
70%	23%	7%

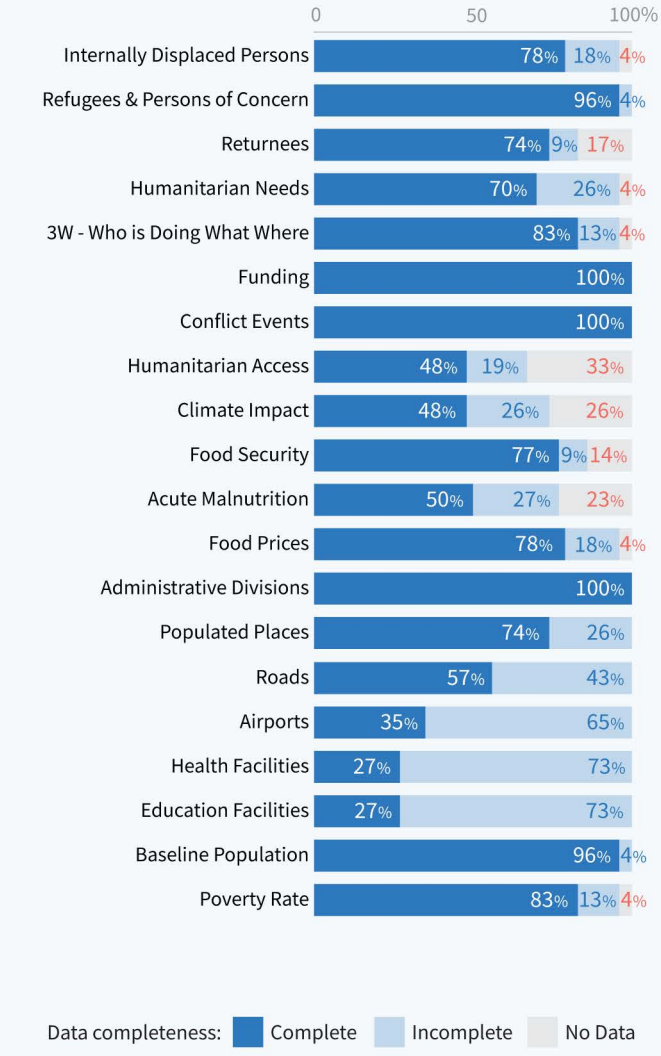
6. COMPLETENESS BY LOCATION, CATEGORY AND SUB-CATEGORY

Although data availability across crises remained stable, a number of locations, categories and sub-categories lost some of their recent gains in 2023. This was the case for South Sudan, which went from 100 percent to 75 percent complete, and Somalia, which went from 90 percent to 75 percent complete. There was a large drop in completeness in health facilities as a dataset that was covering all Sub-Saharan countries became outdated and is therefore considered incomplete. Overall, the sub-categories with no data remained 7 percent compared with 2022.

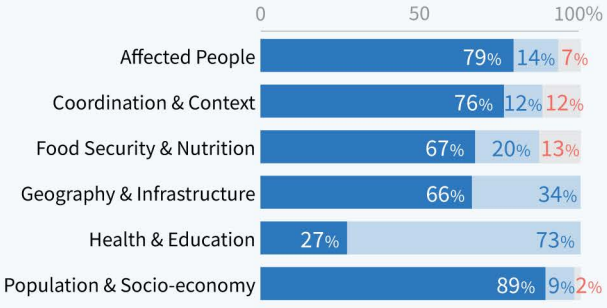
BY LOCATION



BY SUB-CATEGORY



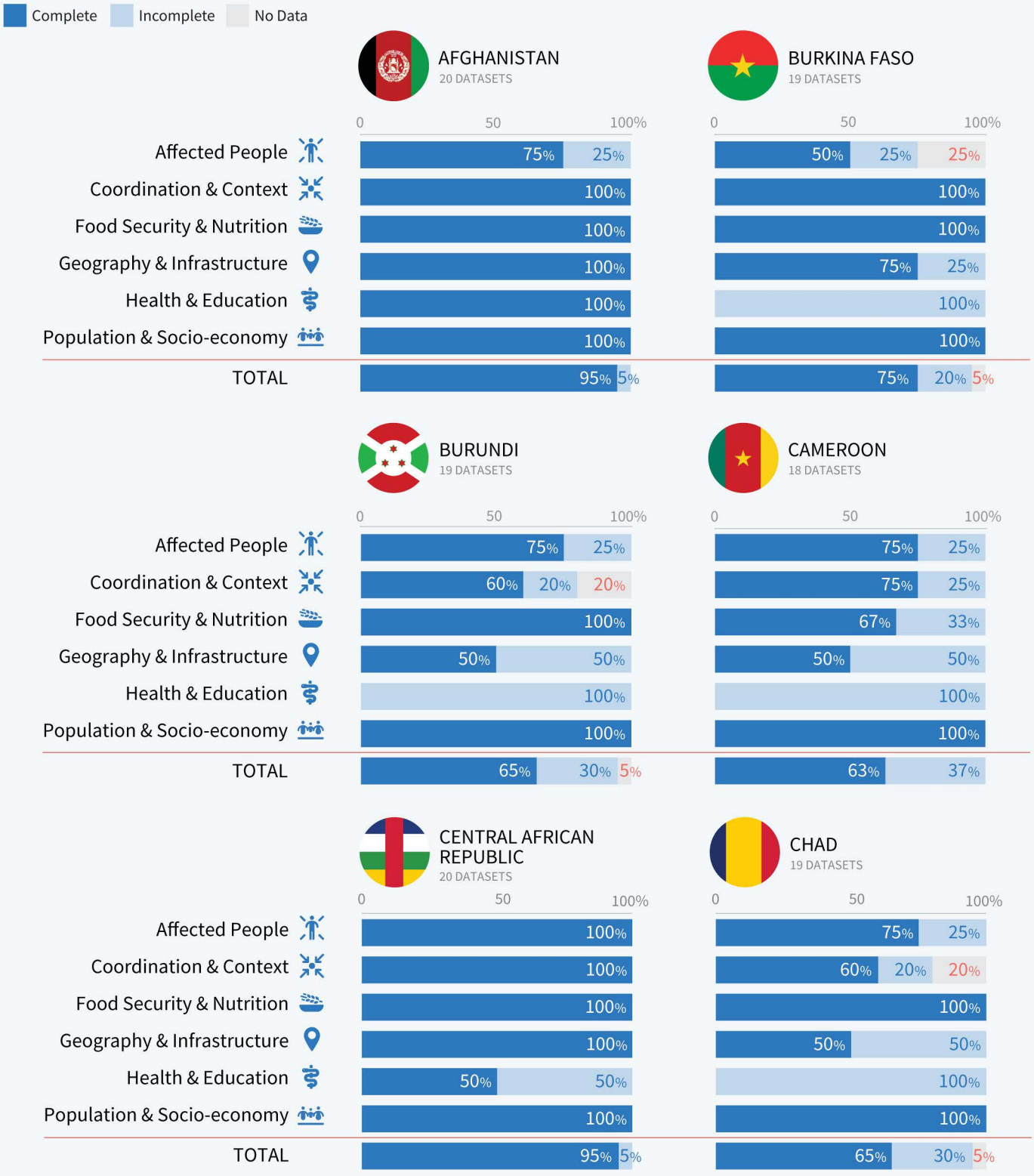
BY CATEGORY

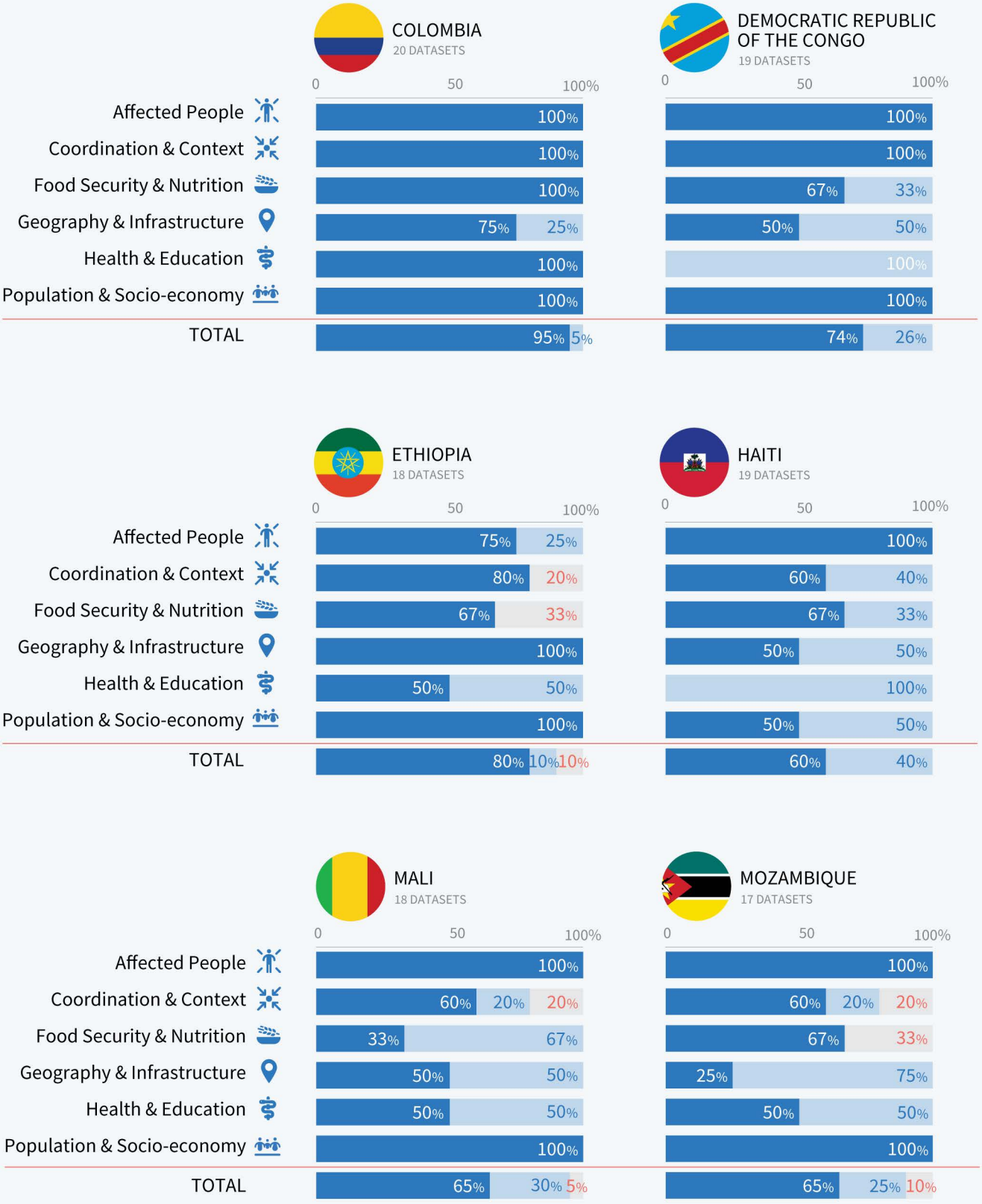


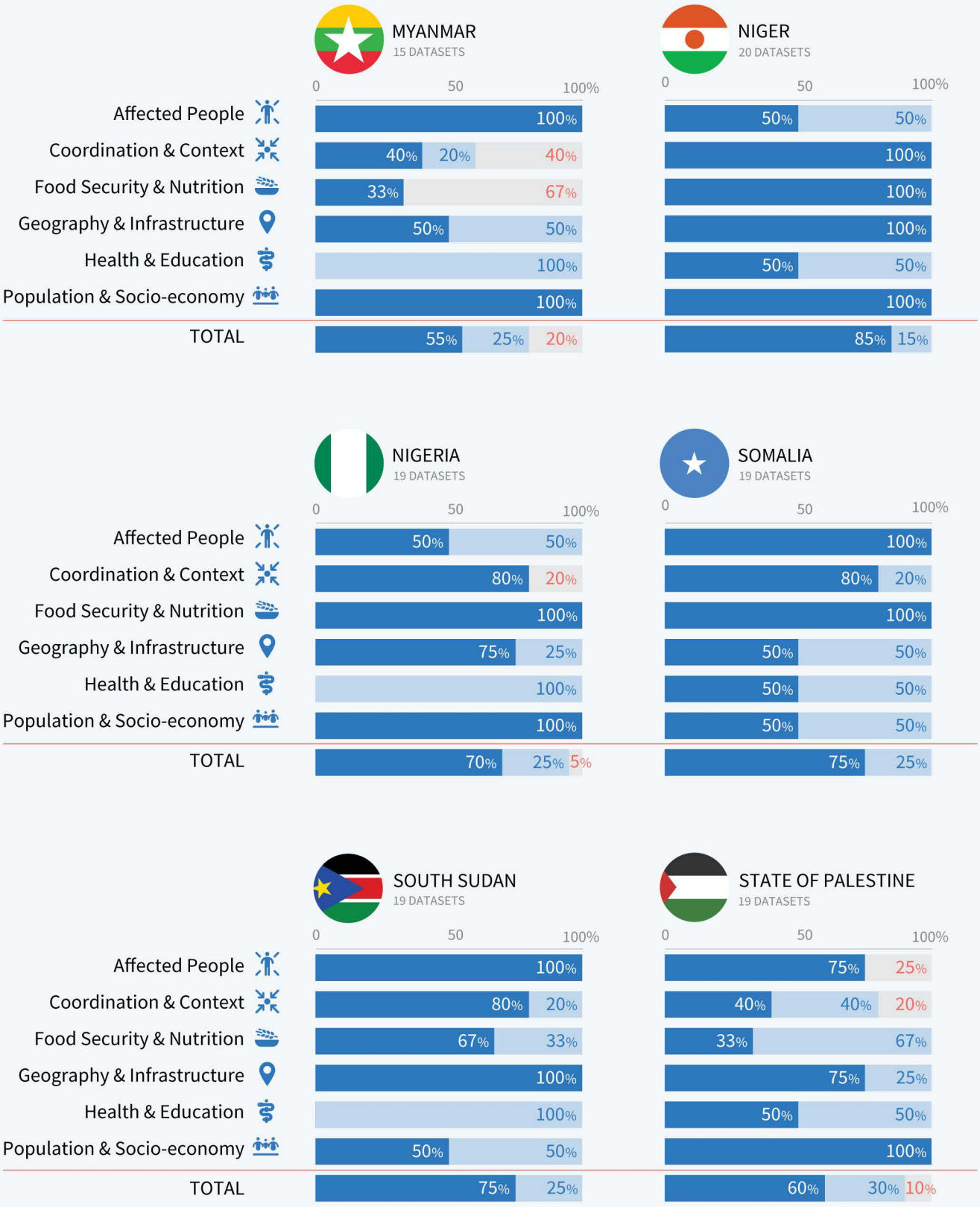
7. COMPLETENESS BY LOCATION AND CATEGORY

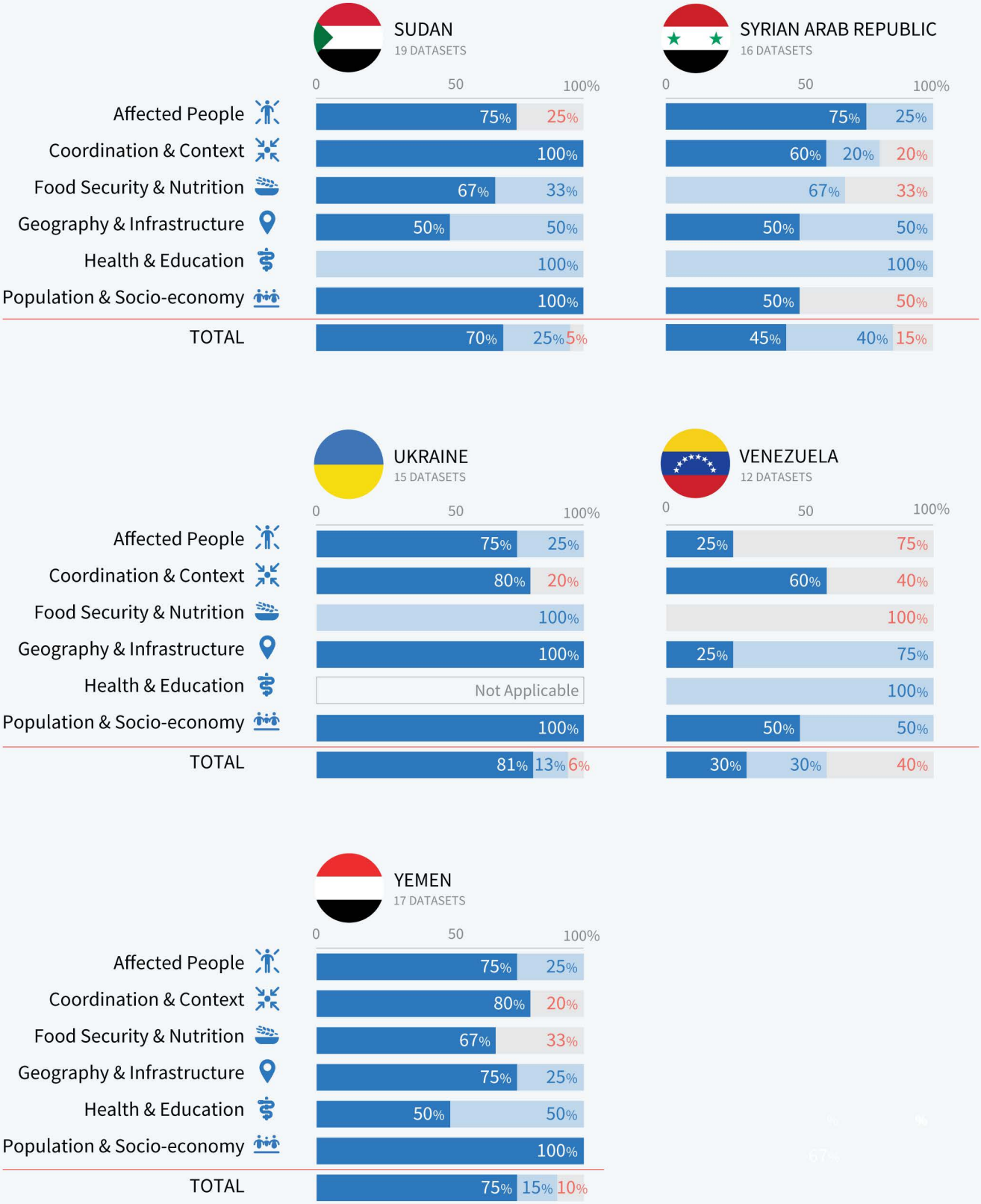
Afghanistan (95 percent), Central African Republic (95 percent) and Colombia (95 percent) are the countries with the highest levels of data completeness. The next most-complete locations for data are Niger (85 percent), Ukraine (81 percent) and Ethiopia (80 percent). Venezuela and Syria are the countries with the largest data gaps.

The most complete categories are population and socio-economy (89 percent), affected people (79 percent) and coordination and context (76 percent). Health and education is the least complete category at 27 percent, owing to incomplete data for the location of health and education facilities. The most popular category, in terms of total dataset downloads, is geography and infrastructure. A total of 397 unique datasets are included in the 23 Data Grids. Each Data Grid includes between 12 to 20 datasets.




COLOMBIA
20 DATASETS
DEMOCRATIC REPUBLIC OF THE CONGO
19 DATASETS
ETHIOPIA
18 DATASETS
HAITI
19 DATASETS
MALI
18 DATASETS
MOZAMBIQUE
17 DATASETS
0
50
100%
Affected People
Coordination & Context
Food Security & Nutrition
Geography & Infrastructure
Health & Education
Population & Socio-economy
TOTAL
100%
75%
25%
67%
33%
50%
80%
20%
60%
40%
10%
65%
30%
5%
74%
26%

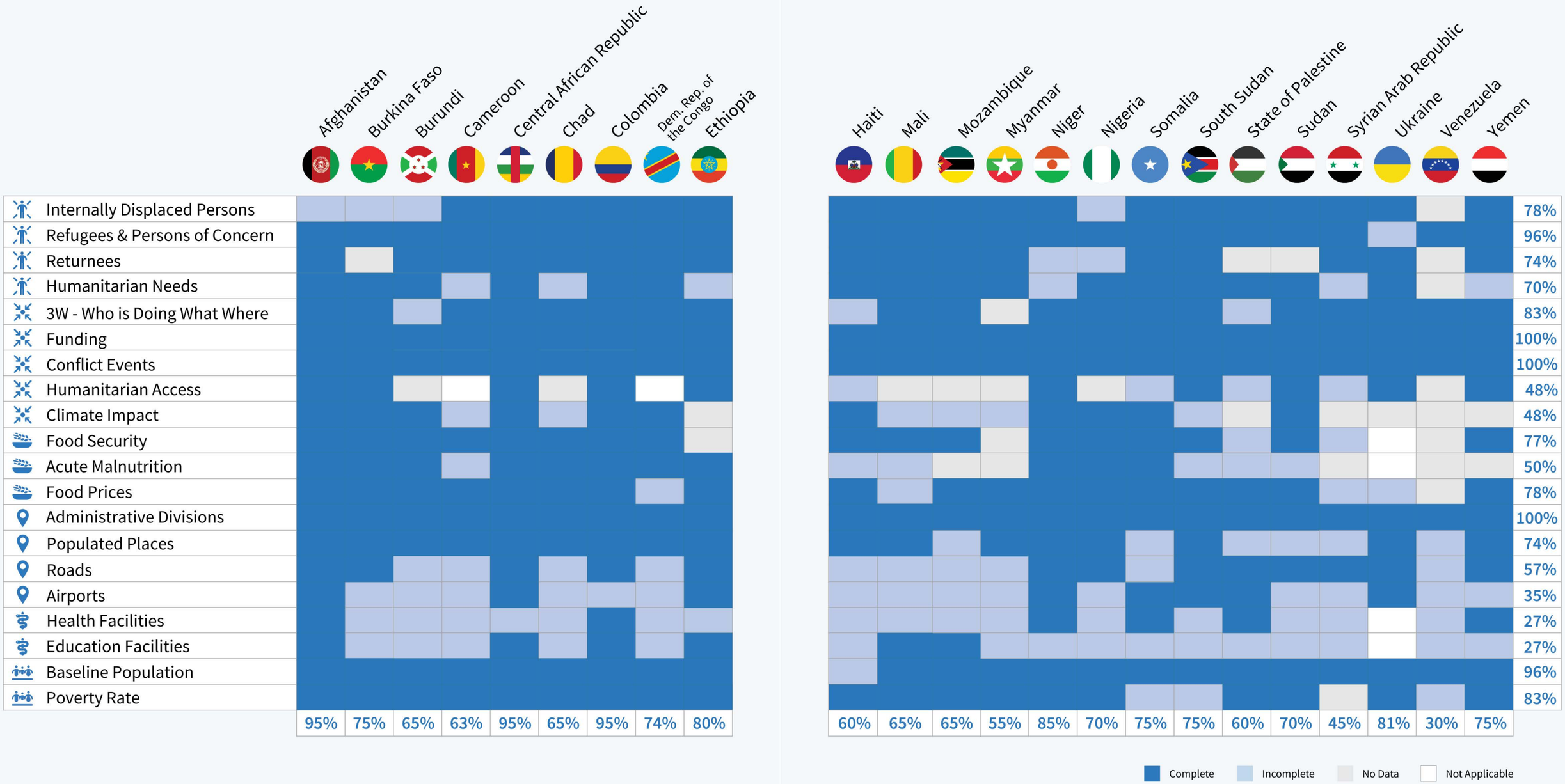

MYANMAR
15 DATASETS
NIGER
20 DATASETS
NIGERIA
19 DATASETS
SOMALIA
19 DATASETS
SOUTH SUDAN
19 DATASETS
STATE OF PALESTINE
19 DATASETS
0
50
100%
Affected People
Coordination & Context
Food Security & Nutrition
Geography & Infrastructure
Health & Education
Population & Socio-economy
TOTAL
100%
40% 20% 40%
33% 67%
50% 50%
100%
100%
55% 25% 20%
50% 50%
100%
100%
100%
100%
50% 50%
100%
85% 15%
50% 50%
80% 20%
100%
75% 25%
100%
50% 50%
70% 25% 5%
100%
100%
100%
80% 20%
100%
100%
50% 50%
50% 50%
50% 50%
75% 25%
75% 25%
40% 40% 20%
33% 67%
75% 25%
50% 50%
100%
60% 30% 10%


SUDAN
19 DATASETS
SYRIAN ARAB REPUBLIC
16 DATASETS
UKRAINE
15 DATASETS
VENEZUELA
12 DATASETS
YEMEN
17 DATASETS
0
50
100%
Affected People
Coordination & Context
Food Security & Nutrition
Geography & Infrastructure
Health & Education
Population & Socio-economy
TOTAL
75% 25%
100%
67% 33%
50% 50%
100%
100%
70% 25% 5%
75% 25%
60% 20% 20%
67% 33%
50% 50%
100%
50% 50%
45% 40% 15%
75% 25%
80% 20%
100%
100%
Not Applicable
100%
81% 13% 6%
25% 75%
60% 40%
100%
25% 75%
100%
50% 50%
30% 30% 40%
75% 25%
80% 20%
67% 33%
75% 25%
50% 50%
100%
75% 15% 10%

8. COMPLETENESS BY LOCATION AND SUB-CATEGORY

The most complete sub-categories are conflict events, administrative boundaries and funding (all at 100 percent), followed by refugees and persons of concern, and baseline population (both at 96 percent), owing to the work of ACLED, OCHA, UNHCR and UNFPA, respectively.

Health and education facilities (both at 27 percent) are the most challenging sub-categories to complete, often because the data is out of date or is crowd sourced and may not be comprehensive. The sub-categories with the highest amount of missing data are humanitarian access, climate impact and acute malnutrition.



9. COUNTRY DEEP-DIVE: COLOMBIA

There has been steady progress with the Colombia Data Grid over the past five years. Overall data completeness has improved from a low of 48 percent in 2020 to a high of 95 percent in 2023. This encouraging trend can be credited to the proactive involvement of the OCHA Office and humanitarian partners, as well as enhanced collaboration between the UN and the Government, notably with the Departamento Administrativo Nacional de Estadística.

The Colombia Data Grid includes 20 datasets shared by nine organizations with reference to 12 sources. Together, these datasets were downloaded 6,709 times in 2023. The most popular dataset is the administrative boundaries, followed closely by population statistics.

A few other data highlights include:

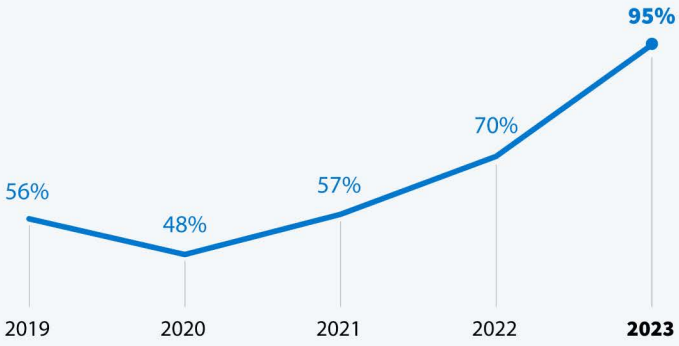
- **Incidents affecting humanitarian access** at the municipal level, collected from several sources including civil society and Government, and contributed by OCHA Colombia.[29]
- Data about **forcibly displaced populations** and stateless persons within Colombia and across the border, collected and shared by UNHCR.[30]
- Information about roads and their importance in the **transportation network**, collected by the National Road Institute in Colombia and contributed by OCHA Colombia.[31]

"With the invaluable support of the HDX team and the assistance of our colleagues from the Humanitarian Country Team, we were able to enhance our completion of Colombia's HDX Data Grid. This effort helped us unveil the diverse and rich landscape of data from our country, demonstrating that through teamwork and dedication, we can unlock the immense potential of data to drive impactful humanitarian efforts."

Santiago Felipe Rengifo Rubio
Data Analyst, Bogotá, OCHA Colombia

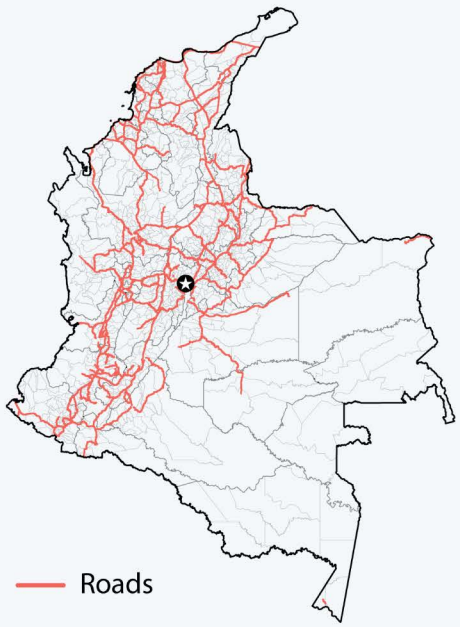
COLOMBIA DATA GRID

95%
Dataset completeness



Creation date: 20 February 2024

COLOMBIA ROADS



Creation date: 21 February 2024

[29] https://data.humdata.org/dataset/acceso-humanitario-colombiaegory
[30] https://data.humdata.org/dataset/unhcr-population-data-for-col
[31] https://data.humdata.org/dataset/colombia-roads

10. CONTRIBUTING ORGANIZATIONS

Out of the 237 active organizations sharing data on HDX, 15 contribute data that is included in the Data Grids, with reference to 164 data sources. Seven of these 15 organizations have been part of HDX since it was established in 2014. This includes the Humanitarian OpenStreetMap Team, IOM, OCHA, OurAirports, UNHCR, WHO and WFP.

OCHA is counted as one organization even though 29 entities[32] share data. Taken together, OCHA added 218 datasets or 55 percent of the data in the Data Grids. This substantial contribution underscores OCHA's important role in data coordination within the humanitarian sector.

Armed Conflict Location & Event Data Project

Food Security and Nutrition Working Group, West and Central Africa

Humanitarian OpenStreetMap Team

Integrated Food Security Phase Classification

International Organization for Migration

Myanmar Information Management Unit

OurAirports

Oxford Poverty & Human Development Initiative

Palestinian Central Bureau of Statistics

United Nations Office for the Coordination of Humanitarian Affairs

United Nations Population Fund

United Nations High Commissioner for Refugees

United Nations Children's Fund[33]

World Food Programme

World Health Organization

[32] This includes 27 OCHA country and regional offices as well as OCHA Field Information Services and OCHA Financial Tracking Service.
[33] This covers UNICEF offices in two countries: DRC and South Sudan.

11. ORGANIZATION DEEP-DIVE: THE WORLD FOOD PROGRAMME

The World Food Programme (WFP) is the world's largest humanitarian agency fighting hunger worldwide, delivering food assistance in emergencies and working with communities to improve nutrition and build resilience. It has been a valuable partner of HDX since 2014 when it became the first UN agency to join the platform.

The data sharing that began during the West Africa Ebola Outbreak has grown over the last ten years to include over 500 datasets covering 145 countries and territories. In 2023 alone, these datasets were downloaded over 140,000 times on HDX.

WFP datasets on HDX include:

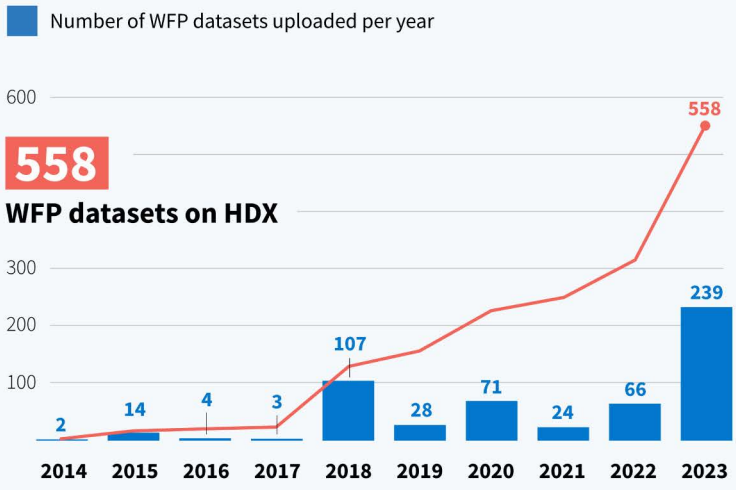
- **Food prices** covering countries, commodities and markets, updated monthly. The data goes back as far as 1992 for a few countries, although many countries started reporting from 2003 or thereafter.

- **Climate data** covering rainfall and vegetation health with between 20 and 40 years of history for some indicators, updated every two weeks.

- **Alerts following sudden onset humanitarian emergencies** with near real-time risk and impact information.

- **Road and street networks**, extracted from the OpenStreetMap database for use in geographic information system (GIS) applications, updated monthly.

Number of WFP datasets uploaded on HDX

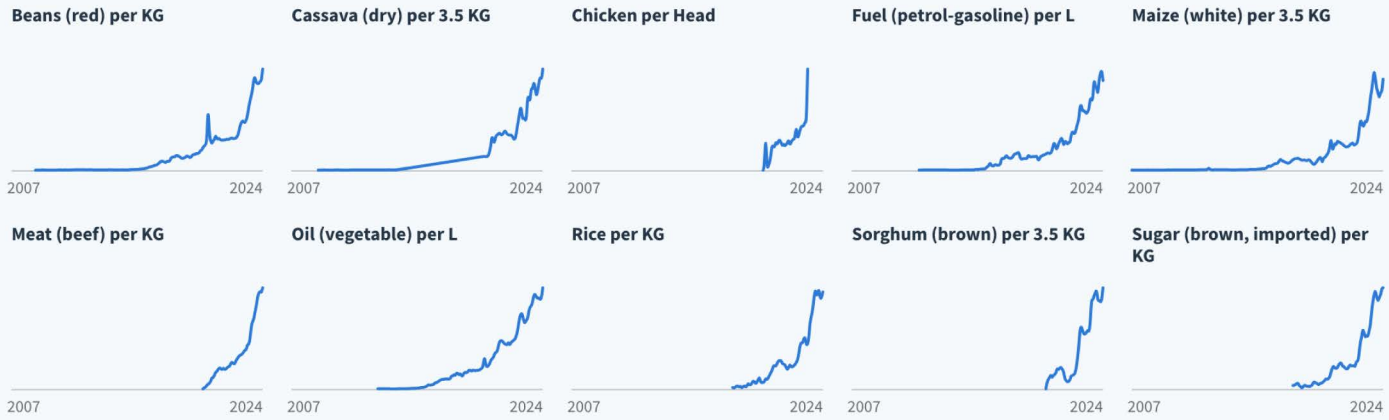


Creation date: 22 February 2024

- **Global airports**, produced from sources such as OpenStreetMap or OurAirports with additional WFP logistics information, updated as needed.

WFP's food prices data stand out as one of the most sought-after resources on HDX. These datasets fall within the food security and nutrition category of the HDX Data Grids, providing coverage for all HRP locations except Venezuela. The data includes basic commodities such as maize, rice, beans, fish and sugar for 98 countries and some 3,000 markets. The ability to afford and access food is a critical component of food security. Regular real-time monitoring of food prices enables humanitarian organizations to quickly identify unusually high prices that could drive up the number of people who are hungry.

South Sudan Food Prices 2007-2024



A sample of the commodities included in WFP food prices for South Sudan.[34]

In 2023, WFP shared first-of-a-kind climate datasets that contain indicators for rainfall and vegetation.[35] The data covers 129 countries, including all HRP locations, and can be used to analyze historical rainfall patterns and compare current conditions to previous years. It can help organizations understand areas currently affected by drought and floods and anticipate needs stemming from climate shocks in humanitarian contexts. For example, the Normalized Difference Vegetation Index (NDVI) datasets can be used to assess vegetation conditions throughout the rainy season and identify the locations most affected by drought to inform better targeting of assistance.

Before this data was made available on HDX, individuals or organizations had to access the large-volume geospatial data directly from the data providers. Processing this data also requires specific expertise and very high computational power. WFP has used its own technical capacity to repackage the data in a consumable format. In doing so, WFP hopes to broaden the number of humanitarian organizations that can use this data to better anticipate and prepare for climate hazards. In 2023, the rainfall and vegetation datasets were downloaded over 26,000 times.

The most recent addition to HDX has been WFP's Advanced Disaster Analysis and Mapping (ADAM) data. To further support early action and humanitarian response, WFP performs a 24/7 automated data harvesting, analysis and mapping of earthquakes, tropical storms and floods. Within minutes of the hazard, alert data is made available on HDX. In addition, WFP prepares a technical analysis that evaluates and identifies programme strategies such as resilience building, disaster risk reduction or capacity building at a national level and makes a report available on request.

Although the popularity of WFP data is clear from the number of downloads, it can be challenging to know exactly how the data is being used for insight and impact. In late 2023, a brief survey was added to WFP data on HDX at the point of download. Over 200 people from academia, nonprofits, government and the private sector have provided feedback so far, demonstrating broad use of WFP's data. For example, academics in Senegal are using WFP's rainfall climate data to train AI models to predict crop yield production. A food policy institute in the United States is monitoring food price data to understand the impact of the war in Ukraine on food prices in Nepal. And a local NGO in Ethiopia is using the food price data to better understand commodity trends across the country.

We are grateful to WFP for their enduring partnership with HDX. We hope that other organizations will be inspired by their commitment to making data openly available and easy to use for analysis.

[34] **https://data.humdata.org/dataset/wfp-food-prices-for-south-sudan**

[35] The primary data is acquired from international centers, processed at WFP's Humanitarian Data Cube cloud platform and shared to HDX as analysis-ready tabular formats designed for humanitarian use. Learn more in this blog: **https://centre.humdata.org/wfp-climate-data-on-hdx/**

12. CLIMATE IMPACT DATA: TROPICAL CYCLONES

Anticipatory action can reduce the impact of climate-related disasters on vulnerable communities by triggering pre-agreed finance and actions based on reliable forecasts and early warning systems. Hazard forecasts are continually improved, thanks in large part to climate datasets with global coverage. However, the relationship between the hazard and the actual humanitarian impact can be harder to establish. Understanding the predicted humanitarian impact, such as homes damaged or agricultural crops destroyed, is essential for determining where assistance might be needed. Unfortunately, the lack of current and historical climate impact data can often be the missing link in predicting humanitarian impact from forecasts.

In 2022, a new sub-category for climate impact was added to the HDX Data Grids to increase awareness and track availability of this data across countries. Over the course of 2023, the percentage of complete datasets dropped from 68 percent to 48 percent. This is largely due to datasets not being updated year on year. This was the case for flood impact data for Cameroon, Chad, Mali and South Sudan as well as cyclone impact data for Mozambique. Additionally, the data that is shared may be incomplete because it only covers one hazard rather than the multiple hazards affecting a country.

Over the past several years, OCHA and partners have been increasing the number of anticipatory action frameworks globally. The frameworks cover a range of locations and hazards, such as floods in Yemen or drought in Niger, and aim to prevent or reduce acute humanitarian impacts before they fully unfold.

One area of progress has been predicting the impact of tropical cyclones on vulnerable communities. These storms, also known as hurricanes or typhoons, affect one in 10 people globally each year,[36] a figure that has doubled in recent decades and is expected to increase with climate change.[37]

Number of major[38] tropical cyclones impacting HRP and non-HRP countries since 2000

Country (HRP only)	Number of major tropical cyclones since 2000	Country (non-HRP only)	Number of major tropical cyclones since 2000
Haiti	10	Japan	76
Mozambique	9	China	52
Honduras	8	Philippines	49
Yemen	5	Mexico	33
Colombia	4	Australia	29
Venezuela	4	United States	25
Myanmar	4	Madagascar	23

Ten of the current HRP locations have faced at least one major tropical cyclone since 2000.[39] Haiti in particular has been impacted by a major cyclone roughly every two years. Tropical cyclones can also cause extensive damage in non-HRP locations and lead to requests for humanitarian assistance. In February 2023, Cyclone Freddy hit Madagascar, Malawi and Mozambique causing nearly 1,000 deaths and a catastrophic loss of livestock in Malawi.[40] By using reliable impact models, some of these impacts can be reduced.

The Philippines is the world's most disaster-prone country, experiencing nearly 20 tropical cyclones a year.[41] In 2021, OCHA facilitated an anticipatory action framework with the potential to reduce the impact of cyclones on over 310,000 vulnerable people. The trigger is based on a typhoon impact model that predicts the level of housing damage based on the forecasted track and intensity.[42]

[36] **Global Population Profile of Tropical Cyclone Exposure from 2002 to 2019 | RAND**

[37] **Climate change is probably increasing the intensity of tropical cyclones**

[38] Major cyclones are at or above a category 3, according to the Saffir-Simpson scale, while within 250km of the country.

[39] A cyclone passing within 250km of the country while it has wind speeds at least 95 knots (roughly equivalent to a Category 3 on the Saffir-Simpson scale).

[40] **World Risk Report 2023**

[41] **Towards a global impact-based forecasting model for tropical cyclones**

[42] The model is available at: **https://github.com/rodekruis/Typhoon-Impact-based-forecasting-model**. The work was done in collaboration with the The Netherlands Red Cross and OCHA Philippines. The model and its application as a trigger for anticipatory action have been peer reviewed: **https://centre.humdata.org/peer-review-of-510s-typhoon-model-and-its-use-in-the-philippines/**

The machine-learning model uses historical impact data to 'learn' how to make impact predictions at the municipal level, so that impact can be anticipated and assistance can be directed more precisely. The visual below shows the damage that the model is predicting for an historical event (Typhoon Melor in 2015), which is then compared with the actual reported damage to estimate model performance.[43]

Currently, the model runs every time there is a forecasted event in the region. It estimates the expected level of damage to houses, which is the main indicator used to trigger the framework. Since the endorsement of the framework in September 2021, the threshold has not been reached.

The main datasets used in the model include:

- **Major typhoon events since 2014** and their impact on affected people and damaged houses, provided by the Philippines Disaster Response Operations Monitoring and Information Center.[44]

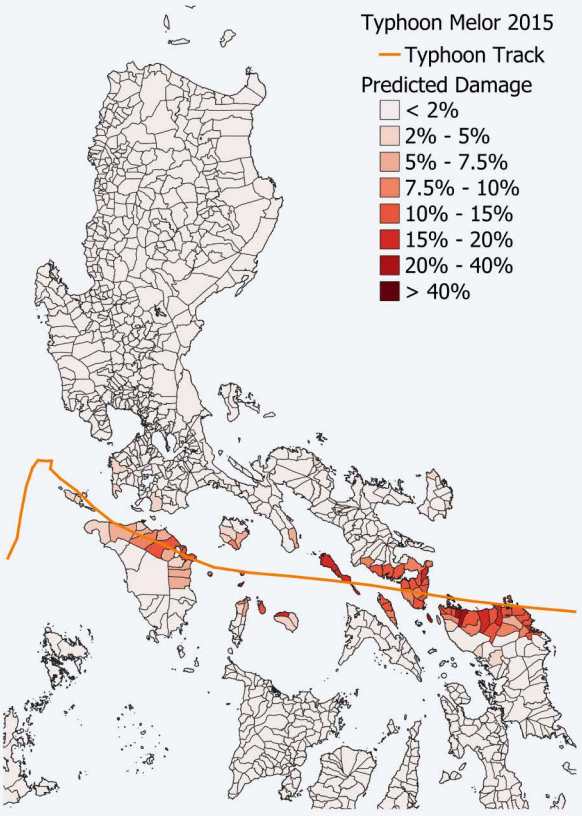
- The type of **roof and wall material of houses**, from the 2015 national census data.[45]

- **Historical cyclone path and intensity**, including cyclone track, maximum wind speed and total rainfall, from multiple sources.[46]

- Topographic data such as **slope, elevation and length of coast**, from the US Geological Survey.[47]

- High resolution **settlement data** to represent the distribution of population within a municipality.[48]

The Philippines' typhoon impact model is being adapted for use in other locations, such as Fiji.[49] The extended model uses frequently available datasets for inputs, such as wind speed, rainfall forecasts, topography, building footprints, population, and wealth indices. However, sub-national impact data remains an essential ingredient. The more sub-national impact data available to train the model, the greater the accuracy in predicting the humanitarian impact.



The percent of housing damage by municipality as predicted by the model in the Philippines for Typhoon Melor in 2015.

National authorities and other actors involved in disaster response can support the development of cyclone impact models by openly sharing current and historical data on the impact of climate hazards. All stakeholders should invest in data collection to enable more proactive disaster response, protect vulnerable communities, and build impact prediction tools to manage the escalating risks of climate change.

[43] A major limitation of model performance is the phenomenon of rapid intensification of tropical cyclones, which remains a major scientific challenge mainly due to the incomplete understanding of its physical mechanisms.

[44] **https://data.humdata.org/dataset/philippines-typhoon-impact-data-2014-2020**

[45] **https://psada.psa.gov.ph/index.php/home**

[46] The Japan Meteorological Agency **https://www.jma.go.jp/jma/indexe.html** and the International Best Track Archive for Climate Stewardship **https://www.ncei.noaa.gov/products/international-best-track-archive**

[47] **https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1**

[48] High Resolution Settlement Layer **https://www.ciesin.columbia.edu/data/hrsl/**

[49] For more information on the development of the model see: **Towards a global impact-based forecasting model for tropical cyclones**

13. CONCLUSION

We will continue to update the Data Grids throughout the year as organizations share new, relevant data. The current status for each location is always available on HDX, both on the relevant location page and on the Overview of Data Grids page.[50] As part of our internal Data Grid Governance Group, we will regularly review the categories and sub-categories to see if they should be removed or expanded.

Humanitarian crises are dynamic and so is the data needed to understand them. The availability and completeness of data will fluctuate year-to-year as new data requirements come into focus. As we have seen with some of the locations and categories that lost ground in 2023, a critical element to data completeness is timeliness – data needs to stay up-to-date for it to remain useful and support decision making.

A priority for 2024 will be increasing access to climate impact data, which is essential for predicting the future impacts of hazards on vulnerable populations and is a prerequisite for anticipatory action. We are also working to make key figures within the Data Grids, such as the number of people in need and the number of people displaced, available programmatically for use in applications, search engines and business intelligence tools.

We welcome feedback on possible improvements. Please be in touch with questions or comments at centrehumdata@un.org.

[50] https://data.humdata.org/dashboards/overview-of-data-grids

ANNEX A: DATA GRID SUB-CATEGORY DEFINITIONS

CATEGORY	SUB-CATEGORY/DEFINITION
Affected People	**Internally Displaced Persons** Tabular data of the number of displaced people by location. Locations can be administrative divisions or other locations (such as camps) if an additional dataset defining those locations is also available.
	Refugees and Persons of Concern Tabular data of the number of refugees and persons of concern either in the country or originating from the country disaggregated by their current location. Locations can be administrative divisions or other locations (such as camps) if an additional dataset defining those locations is also available or if the locations' coordinates are defined in the tabular data.
	Returnees Tabular data of the number of displaced people who have returned.
	Humanitarian Needs Tabular data of the number of people in need of humanitarian assistance by location and humanitarian cluster/sector.
Coordination & Context	**3W - Who is doing What Where** List of organizations working on humanitarian issues, by humanitarian cluster/sector and disaggregated by administrative division.
	Funding Tabular data listing the amount of funding provided by humanitarian cluster/sector.
	Conflict Events Vector data or tabular data with coordinates describing the location, date and type of conflict event.
	Humanitarian Access Tabular or vector data describing the location of natural hazards, permissions, active fighting or other access constraints that impact the delivery of humanitarian interventions.
	Climate Impact Tabular or vector data containing current and historical impacts of climate events relating to floods, droughts and storms. The data should specify the location of the event, date of the event, and contain at least one indicator of impact such as spatial extent of event, disruption to affected populations, destroyed infrastructure and/or affected vegetation.
Food Security & Nutrition	**Food Security** Vector data representing the IPC/CH acute food insecurity phase classification or tabular data representing population or percentage of population by IPC/CH phase and administrative division.
	Acute Malnutrition Tabular data specifying the global acute malnutrition (GAM) or severe acute malnutrition (SAM) rate by administrative division.
	Food Prices Time series prices for common food commodities at a set of locations.

CATEGORY	SUB-CATEGORY/DEFINITION
Geography & Infrastructure	**Administrative Divisions** Vector geographic data describing the sub-national administrative divisions of a location, usually a country, including the names and unique identifiers, usually p-codes, of each administrative division. To be considered 'complete', and included here, the humanitarian community working in the location has to have endorsed a preferred set of administrative boundaries as the Common Operational Dataset (COD).
	Populated Places Vector data or tabular data with coordinates representing the location of populated places (cities, towns, villages).
	Roads Geographic data describing the location of roads with some indication of the importance of each road segment in the transportation network. The data should exclude or indicate roads that are not usable by typical four-wheel-drive vehicles (footpaths, etc.).
	Airports Geographic data representing all operational airports including a name or other unique identifier and an indication of what types of aircraft can use each.
Health & Education	**Health Facilities** Vector data or tabular data with coordinates representing health facilities with some indication of the type of facility (clinic, hospital, etc.).
	Education Facilities Vector data or tabular data with coordinates representing education facilities with some indication of the type of facility (school, university, etc.).
Population & Socio-economy	**Baseline Population** Total population disaggregated age and sex categories, aggregated by administrative division.
	Poverty Rate Population living under a defined poverty threshold, aggregated by administrative division and represented as a percentage of total population or as an absolute number.

ANNEX B: INCLUSION OF DATA IN THE DATA GRIDS

Data Grid curation involves the careful evaluation of datasets on HDX for inclusion in one or more of the sub-categories in accordance with our standard criteria. The HDX team conducts this evaluation when a dataset is newly added or updated on HDX.

The HDX team supports the Data Grids in three ways:

1. Reviewing whether any dataset on HDX could be added to a Data Grid;
2. Conducting targeted outreach with humanitarian organizations to find relevant datasets; and
3. Identifying potential datasets with partners in the research community, academia and/or government agencies.

The first step in determining whether a dataset should be included in a Data Grid is to check whether the dataset meets the thematic requirement defined in Annex A. Datasets that are not considered relevant are automatically excluded.

The second step is to determine if the dataset is sub-national. To fulfill this requirement the data must be disaggregated to at least the first administrative division. If the data is only available at the national level, it is typically excluded. The sub-national requirement has been partially or completely waived for two datasets: UNHCR's data on forcibly displaced populations and stateless persons and OCHA FTS's requirements and funding data.

The third step is to assess whether a thematically relevant, sub-national data set is complete or incomplete. A dataset is considered complete if it satisfies all of the following criteria:

1. Broad geographic coverage;
2. Available in commonly used formats; and
3. Timely (full definitions are given below).

If the dataset does not satisfy one or more of these criteria, it is considered incomplete.

Finally, the dataset is compared against existing datasets for that location to determine if it should be added to the Data Grid. If the sub-category is empty, or if the data would complement other datasets in a sub-category, the HDX team will add it to the Data Grid. A complementary dataset is one that adds unique and useful information to the sub-category.

The sub-category is considered complete if it has at least one complete dataset available for the sub-category. If the sub-category contains only incomplete datasets, then that sub-category is considered incomplete. Overall category completeness refers to the proportion of sub-categories in the category that are complete. Similarly, completeness for a location refers to the proportion of sub-categories that are complete in the location.

Sub-categories are considered empty if no datasets on HDX meet the above-mentioned criteria. In general, data can be missing for three reasons:

1. It is not collected (e.g., because nobody is present to do so, because it is unsafe to access areas to collect it, because it requires investment and resources that are not available, or because nobody prioritizes it as a gap to fill).

2. It is collected but not publicly shared (e.g., because the collecting organization does not have an open data policy, because the data is sensitive and should not be shared, because the collecting entity fears sharing the data with actors they do not know and trust, or because of a lack of resources to clean and share it).

3. It is collected and shared but is not shared on HDX, or known about by the HDX team.

DETAILED COMPLETENESS CRITERIA

The criteria for evaluating completeness for relevant, sub-national data are detailed below:

Does the data have broad geographic coverage with explicit location information?

- Is the dataset geographically comprehensive, or as comprehensive as possible? If the dataset is disaggregated by administrative divisions, does it cover all of them? If it does not, is the meaning of a missing administrative division defined in the metadata? If there is no comprehensive list to compare against (for example, with spontaneous displacement locations), does the dataset make it clear if it attempts to be comprehensive or not?

- Are location references defined? The dataset should contain explicit geographic data (i.e., Geographic Information System data or tabular data with latitude and longitude fields). If not, the dataset should be joinable to an available dataset that defines those locations.

Is the data in commonly-used formats?

- Is it stored in a common file format? We include CSV, XLS, XLSX, SHP, GEOJSON, etc. Formats like GPKG and others that are more difficult for the typical humanitarian data specialist would be marked 'incomplete.'

- Is the data tidy? Field names and data rows should be easy to determine. There should not be sub-total rows interspersed with data rows. The required data for the category should be in a single table on the same tab. For tabular data with coordinates, the x and y columns (usually longitude and latitude) should be in decimal degree format and separated into two columns.

Is the data timely?

- Has the dataset become out of date? Depending on how frequently the dataset is expected to be updated, the HDX team considers the age of the data and whether the dataset should have been superseded.