

THE STATE OF OPEN HUMANITARIAN DATA:

WHAT DATA IS AVAILABLE AND MISSING ACROSS HUMANITARIAN CRISES

JANUARY 2020



OCHA

centre for humdata



HDX

TABLE OF CONTENTS

1. INTRODUCTION	3
2. KEY MESSAGES	5
3. GLOBAL OVERVIEW	6
4. COMPLETENESS BY LOCATION, BY CATEGORY, BY SUB-CATEGORY	7
5. COMPLETENESS BY LOCATION AND CATEGORY	8
6. COMPLETENESS BY LOCATION AND SUB-CATEGORY	10
7. COUNTRY DEEP DIVE: AFGHANISTAN	11
8. CONTRIBUTING ORGANIZATIONS	13
9. DATA FOR MODELING	14
ANNEX A: DATA GRID SUB-CATEGORY DEFINITIONS	15
ANNEX B: DATA GRID CRITERIA FOR ASSESSING COMPLETENESS	17

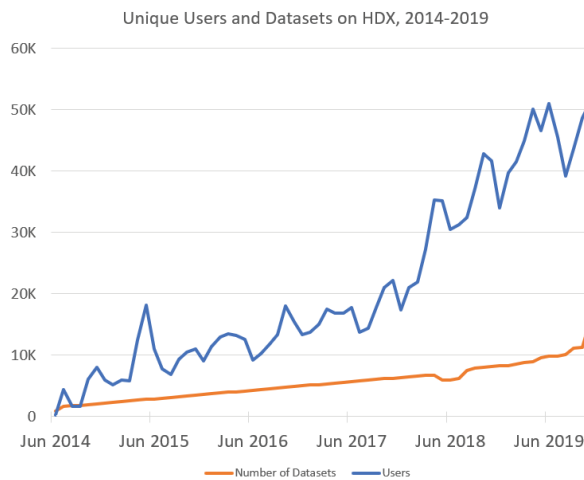
Acknowledgements

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

1. INTRODUCTION

The goal of this report is to increase awareness of the data available for humanitarian response activities and to highlight what is missing, as measured through OCHA’s Humanitarian Data Exchange (HDX) platform. We want to recognize the valuable and long-standing contributions of data-sharing organizations. We also want to be more targeted in our outreach on what data is required to understand crises so that new actors might be compelled to join the platform. Data is not an end in itself but a critical ingredient to the analysis that informs decision making. With nearly 168 million people in need of humanitarian assistance in 2020¹ – the highest figure in decades – there is no time, or data, to lose.

When HDX was launched in 2014, it held around 800 datasets. Over the past five years, that number has skyrocketed to over 17,000 datasets. The data covers every active humanitarian crisis, from Afghanistan to Yemen, and has been shared by dozens of organizations, from ACLED to WFP.² In 2019, HDX was accessed by over 600,000 users.



This is a tremendous achievement for collective action in a sector that relies on cooperation. It also shows the value of an open data platform. OCHA’s work to aggregate data from many sources in one place has undoubtedly created efficiency in the system. Humanitarians, donors, academics, and journalists no longer need to chase contacts to locate data; they can go to HDX and search for it. If the data is not there, the HDX team will help find it.

“Accurate data is the lifeblood of good policy and decision-making. Obtaining it, and sharing it across hundreds of organizations, in the middle of a humanitarian emergency, is complicated and time-consuming - but it is absolutely crucial.”³

- United Nations Secretary-General António Guterres at the opening of the OCHA Centre for Humanitarian Data in The Hague in December 2017

One downside to all of this data sharing is knowing what data is most relevant to understanding a crisis context. In May 2019, HDX added a new feature called the Data Grid to help people in their quest for good and relevant data. Based on extensive user research, the Data Grid places the most important crisis data into six categories: affected people; coordination and context; food security and nutrition; geography and infrastructure; health and education; and population and socio-economy.

¹ 2020 Global Humanitarian Overview: https://www.unocha.org/sites/unocha/files/GHO-2020_v9.1.pdf

² Humanitarian data is defined as: 1) data about the context in which a humanitarian crisis is occurring; 2) data about the people affected by the crisis and their needs; and 3) data about the response by organizations and people seeking to help those who need assistance.

³ Taken from the opening remarks of the UN Secretary-General at the launch of the Centre for Humanitarian Data in The Hague in December 2017: <https://www.un.org/sg/en/content/sg/speeches/2017-12-22/opening-centre-humanitarian-data-remarks>

Within each category, there are several sub-categories. For example, within affected people, there are sub-categories for internally displaced people, refugees, returnees, humanitarian needs, and casualties. Within coordination and context, there are sub-categories for ‘who is doing what where’, affected areas, humanitarian access, and funding, among others. (See Annex A for all sub-category definitions).

There are three main criteria for whether relevant data is included in the Data Grid: 1) disaggregated beyond the national level; 2) commonly-used formats; and 3) timeliness. If at least one dataset meets all criteria, that sub-category is considered ‘complete’. If at least one dataset meets some of these criteria, the sub-category is considered ‘incomplete’. If a dataset does not meet the criteria or does not exist on HDX, the sub-category is considered empty or as having no data. (See Annex B for details on the Data Grid criteria).

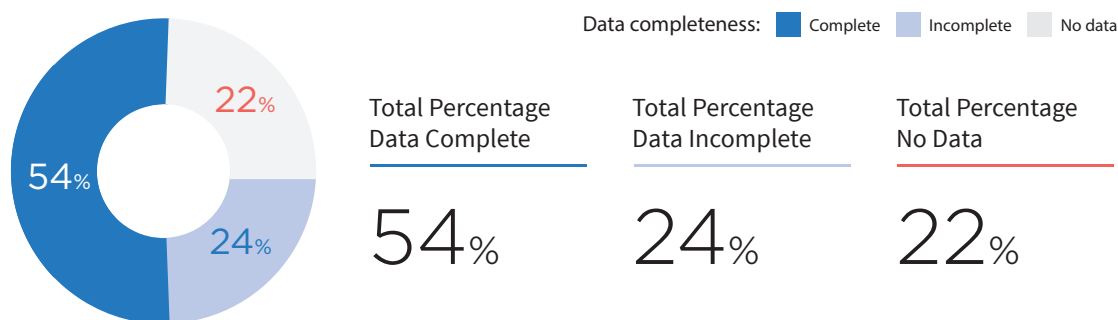
DATA GRID CRITERIA: 1) SUB-NATIONAL; 2) COMMONLY-USED FORMATS; AND 3) TIMELINESS.

Complete	Incomplete	No Data
At least one dataset in the sub-category meets all criteria.	At least one dataset in the sub-category meets some criteria.	Available data in the sub-category does not meet the criteria or does not exist on HDX.

Of course, relevant data will greatly depend on who is looking and what they are looking for. A dataset might have the right data, but not cover the part of the country needed for the analysis. Or it might cover the right geographic area but be in a format that is difficult to work with. For this reason, the HDX team reviews all relevant datasets and assesses them against the criteria. This careful curation process is undertaken daily on all newly shared or updated datasets. So far, some 700 datasets have been taken through this process.

By the end of December 2019, HDX included Data Grids for 14 locations.⁴ These include: Afghanistan, Bangladesh, Central African Republic, Chad, Colombia, Democratic Republic of the Congo, Mozambique, Myanmar, the Philippines, State of Palestine, Somalia, Sudan, Venezuela and Yemen. We will expand the Data Grids to cover all locations with a Humanitarian Response Plan throughout 2020.⁵ We may also expand the categories and sub-categories based on feedback.

As we start 2020, the completeness of all Data Grids combined is 54 percent. That is, 54 percent of relevant, comprehensive data is available across 14 locations. If we add the data that is relevant but incomplete, the total is 78 percent. This leaves 22 percent of categories with data that does not meet the criteria or with no data. The Data Grids include an average of 20-30 datasets per location.



It is important to note that not all humanitarian data can be shared openly. Data about the location of affected people and responders can put people at risk, especially in conflict environments. The HDX Terms of Service⁶ prohibit the sharing of data that includes personally identifiable information. For sensitive, non-personal data

⁴ A location refers to a country or territory.

⁵ See all response plans here <https://www.hpc.tools>. A response plan may be for a country, region, or unique to a specific crisis.

⁶ Organizations sharing data through the HDX platform should ensure that the data was collected in a legal, ethical and responsible manner. For more on the HDX Terms of Service, visit <https://data.humdata.org/about/terms>

that can be shared under certain conditions, HDX offers a feature, HDX Connect, which enables organizations to share only the metadata and make the underlying data available bilaterally upon request. There are two HDX Connect datasets that are included in the Data Grid, both related to Venezuela.⁷

The HDX team will continue to update and expand the Data Grids throughout the year. We look forward to continued collaboration and to closing data gaps.

2. KEY MESSAGES

- Just over 50 percent of relevant crisis data is available across 14 humanitarian operations. Afghanistan and the Central African Republic have the most complete data, while Venezuela has the least.
- The largest data gaps are in the categories for health and education, and food security and nutrition. The categories with the best coverage of data include affected people, and geography and infrastructure.
- Partners have made tremendous progress with data sharing and use over the past several years. The Humanitarian Data Exchange platform hosts over 17,000 datasets shared by hundreds of organizations. The data was accessed by over 600,000 users in 2019.
- Administrative boundaries and place names are among the most critical datasets for ensuring a shared understanding of location in an operating environment but can be difficult to source and maintain.
- Although some data is sensitive and cannot be shared openly, organizations can share the metadata with the underlying data available by request through HDX Connect. This ensures the data is protected from misuse while also increasing transparency and avoiding duplication around what data has been collected.
- Reticence risk — the decision not to share data because of uncertainty — can be addressed through secure data sharing techniques and disclosure control processes.
- Predictive analytics is an exciting area with the potential to transform how response happens but models can only work based on good data. Data cleaning is the most time consuming part of the modeling process. When feasible, data that has been cleaned for use in models, or the modeled data itself, should be shared on HDX.

We call on partners to share or help source the following data that is critical but often missing for many crises:

- The location of affected schools (potential sources: national governments, UNICEF, the Education Cluster).
- The location of health facilities (potential sources: national governments, WHO).
- The authoritative list of populated places with locations (potential sources: national governments, UNFPA).
- The authoritative geographic dataset of sub-national administrative divisions (source: national governments).
- The location of airports (potential sources: national governments, private sector).
- Local transportation routes with an indication of status (potential sources: national governments, WFP, the Logistics Cluster).
- Food security data with population by IPC phase and administrative division in an accessible format (potential sources: Fewsnet, IPC partners).
- Population living under a defined poverty threshold, aggregated by administrative divisions (potential sources: World Bank, Oxford Poverty & Human Development Initiative).
- Global Acute Malnutrition (GAM) and Severe Acute Malnutrition (SAM) rates (potential sources: UNICEF, World Bank, WHO).

⁷ Both datasets are from IOM and can be found at: <https://data.humdata.org/dataset/venezuela-4w-december-ocha> and <https://data.humdata.org/dataset/dtm-round-3-flow-monitoring-of-venezuelan-migration-peru>.

3. GLOBAL OVERVIEW

Number of Locations

14

Number of Categories

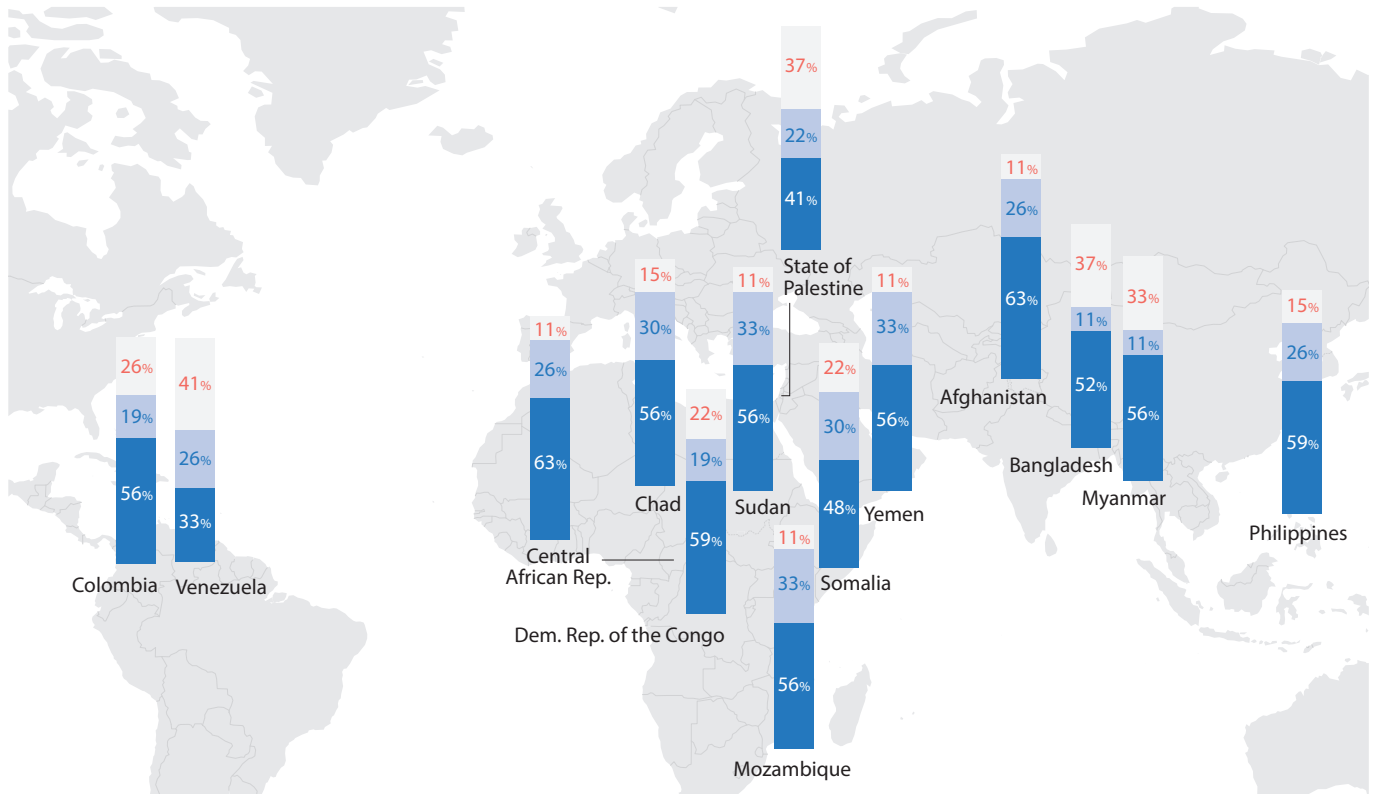
6

Number of Sub-Categories

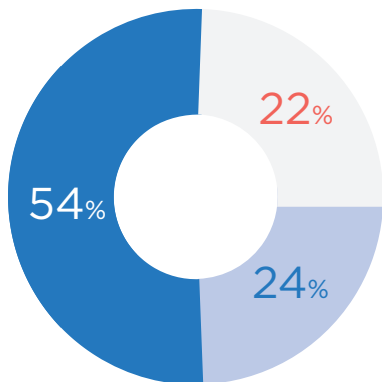
27

Number of Contributing Organizations

28



Data completeness: ■ Complete ■ Incomplete ■ No data



Total Percentage Data Complete

54%

Total Percentage Data Incomplete

24%

Total Percentage No Data

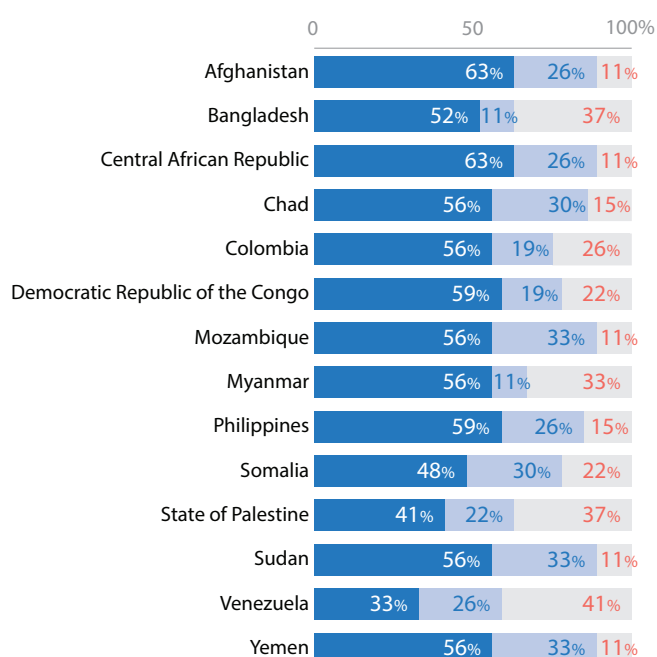
22%

4. COMPLETENESS BY LOCATION, BY CATEGORY, BY SUB-CATEGORY

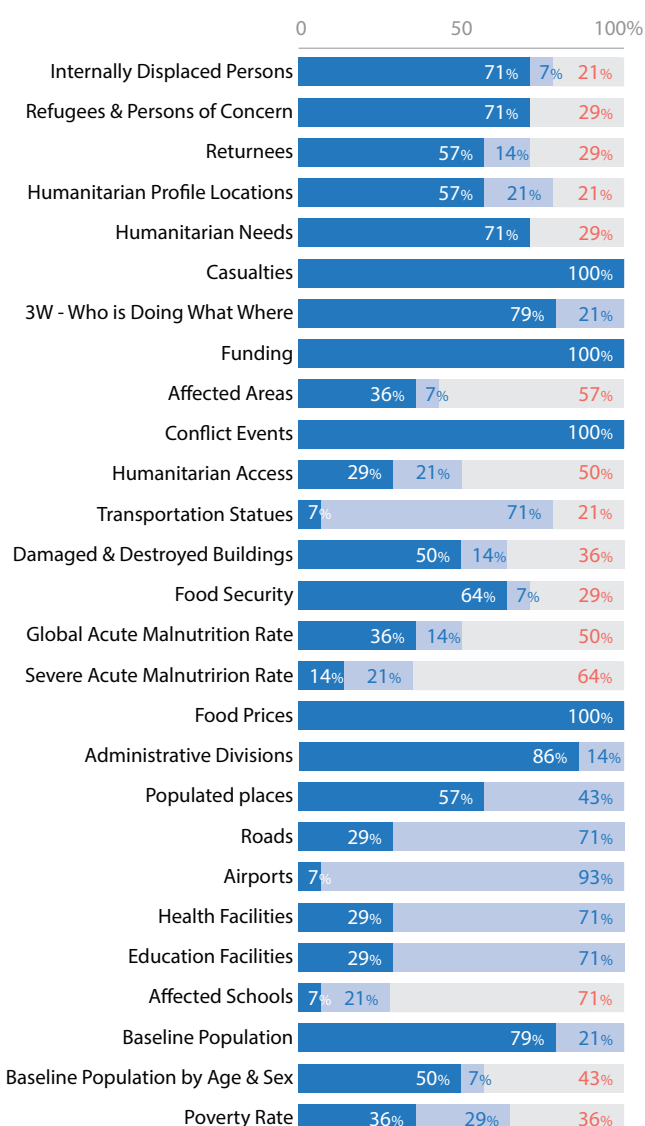
The overall figure of 54 percent complete across 14 locations masks a broad range of what is available and missing in each location, category and sub-category. While Afghanistan and the Central African Republic have the most relevant and comprehensive data at 63 percent complete, Venezuela has the least at 33 percent complete. Venezuela also has the highest percentage of no data at 41 percent, illustrating the challenge of data collection and sharing in that operation.

Health and education is the least complete category at 21 percent, with incomplete data for the location of health and education facilities, and missing data for affected schools. While HDX does include significant data related to health and education statistics, this data is generally at the national level, e.g. one number for the level of primary school enrollment in a country. We do plan to expand the Data Grid in the future to include a sub-category for out-of-school children⁸. We welcome feedback on other relevant data and sources to include.

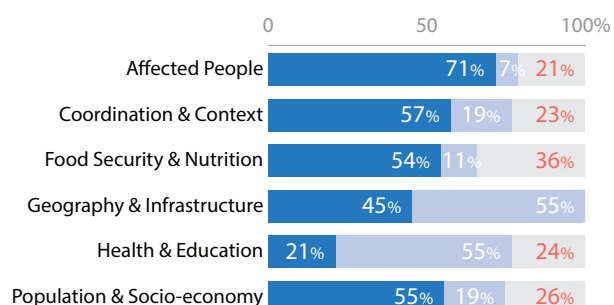
BY LOCATION



BY SUB-CATEGORY



BY CATEGORY



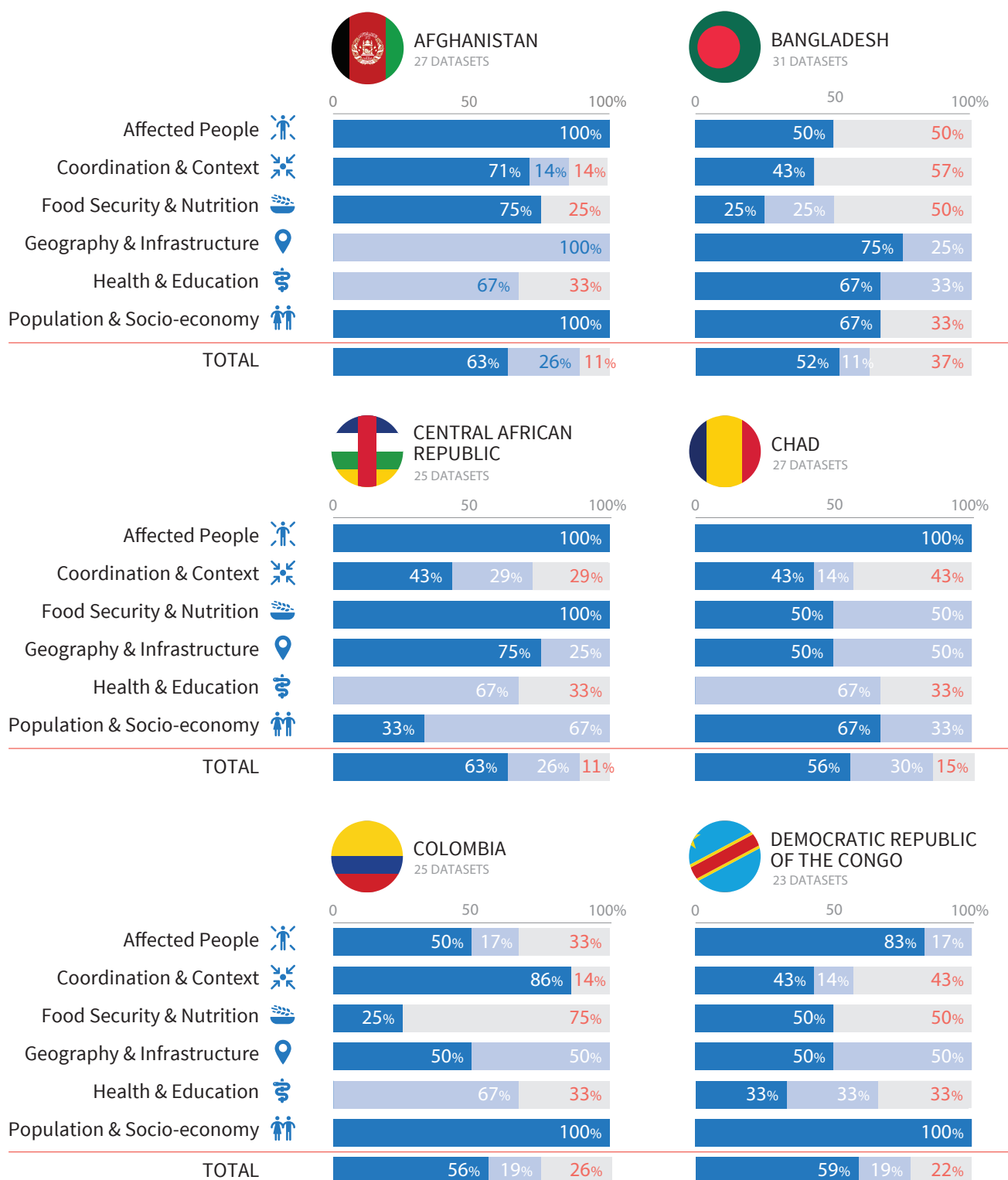
Data completeness: ■ Complete ■ Incomplete ■ No data

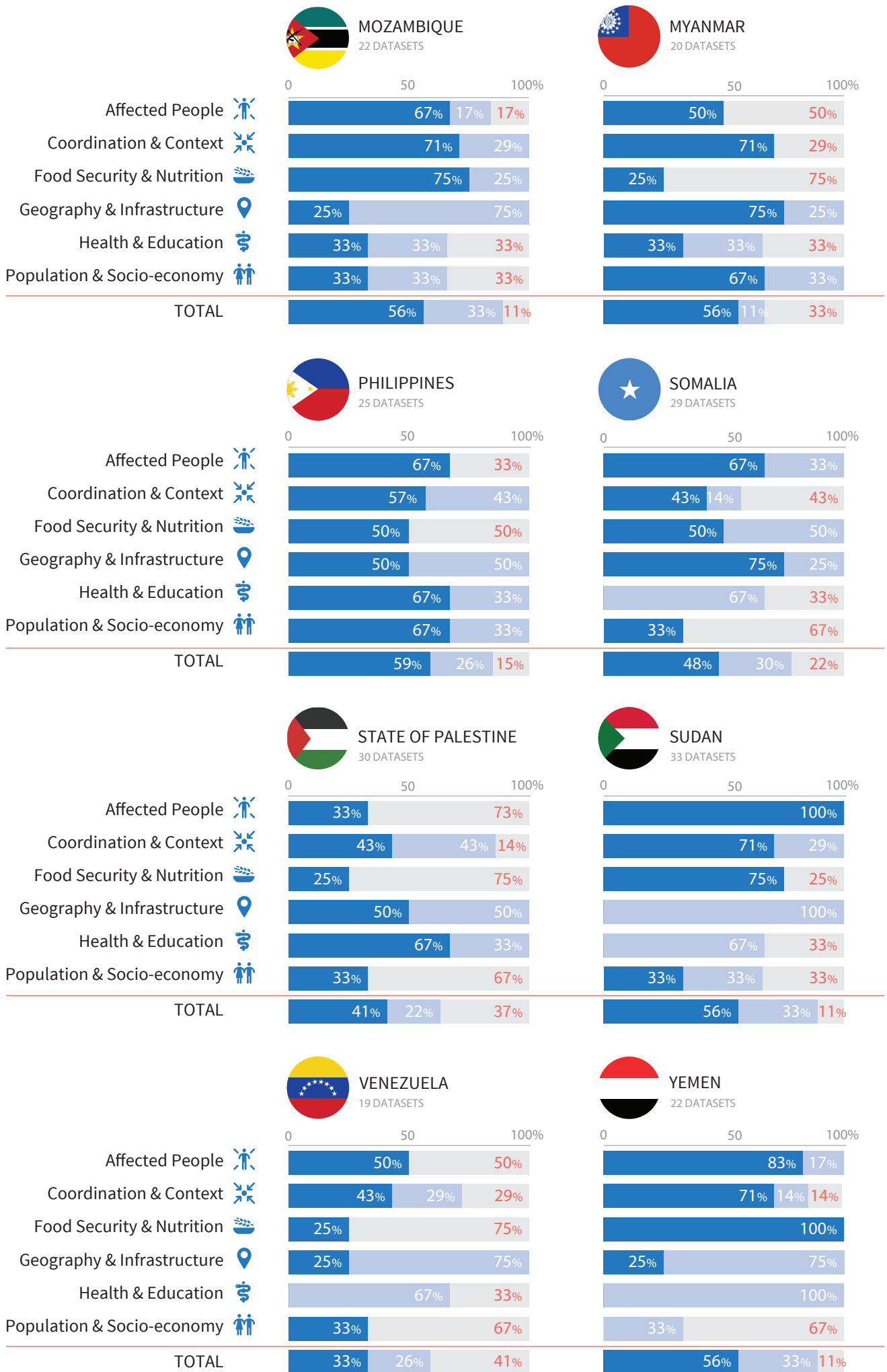
⁸ Data on out-of-school rates by country is available on HDX from UNESCO and UNICEF. The Centre has been working to source education in emergency datasets through a partnership with the Education Above All Foundation. Current data can be found here: <https://data.humdata.org/dashboards/education-in-emergencies>

5. COMPLETENESS BY LOCATION AND CATEGORY

This section provides an overview by location and category so that responders in specific humanitarian operations can see where the largest gaps are. In some contexts the data may be available but cannot be shared openly. In these cases, we encourage partners to use HDX Connect to share the metadata with the underlying data available by request.

The number of datasets included in each Data Grid depends on how the data is organized but in general, each sub-category includes one to two datasets. With 27 sub-categories, we would expect a complete Data Grid to include between 50-60 datasets. For the current 14 locations, the number ranges from 19 datasets (Venezuela) to 33 datasets (Sudan).

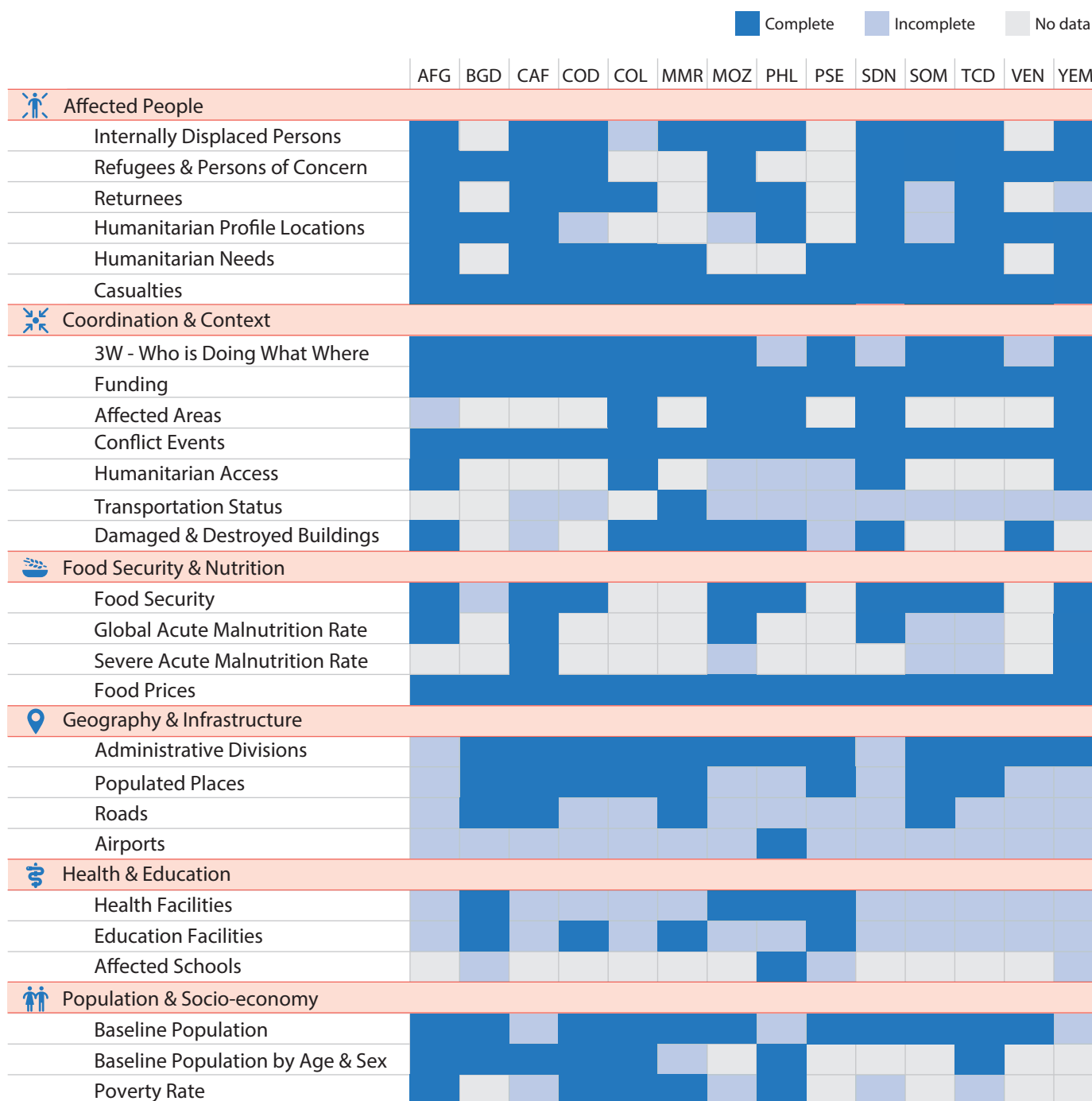




6. COMPLETENESS BY LOCATION AND SUB-CATEGORY

This section provides a closer look at the sub-categories by location. Here we can see that the geography and infrastructure category has no missing data, even if some of the sub-categories are incomplete. This is largely due to the work of OCHA and governments to provide data on administrative boundaries. In cases where this data is incomplete, it may be that the geographic coverage is somehow limited or that the boundaries have not been endorsed as a ‘common operational dataset’.⁹ Within this same category, Humanitarian OpenStreetMap also provides extensive data on points of interest (roads, buildings etc) through its crowd-sourced platform. All other categories have some missing data.

AFG Afghanistan COD Dem. Rep. of the Congo MOZ Mozambique SDN Sudan VEN Venezuela
 BGD Bangladesh COL Colombia PHL Philippines SOM Somalia YEM Yemen
 CAF Central African Republic MMR Myanmar PSE State of Palestine TCD Chad



⁹ <https://data.humdata.org/dashboards/cod>



7. COUNTRY DEEP DIVE: AFGHANISTAN

The Data Grid for Afghanistan is one of the most complete at 63 percent. We have listed the individual datasets included in each category and sub-category to showcase the types of data and the contributing organizations. Unique to the Afghanistan Data Grid is the dataset on the 'impact of the use of explosive weapons on health facilities' shared by the Explosive Weapons in Populated Places (EWIPA) Community.




AFFECTED PEOPLE

 8 Datasets



Internally Displaced Persons

-  [Afghanistan Displacement Data - Baseline Assessment \[IOM DTM\]](#) International Organization for Migration
-  [Afghanistan - Conflict Induced Displacements in 2019](#) OCHA Afghanistan


Refugees & Persons of Concern

-  [Afghanistan Voluntary Repatriation 2019](#) UNHCR Afghanistan
-  [Afghanistan Displacement Data - Baseline Assessment \[IOM DTM\]](#) International Organization for Migration
-  [Pakistani Refugee in Afghanistan 2017](#) UNHCR Afghanistan


Returnees

-  [Afghanistan Voluntary Repatriation 2019](#) UNHCR Afghanistan
-  [Afghanistan Displacement Data - Baseline Assessment \[IOM DTM\]](#) International Organization for Migration




Humanitarian Profile Locations

-  [Afghanistan Displacement Data - Baseline Assessment \[IOM DTM\]](#) International Organization for Migration

Humanitarian Needs

-  [Afghanistan: Humanitarian Needs Overview](#) OCHA Afghanistan



Casualties

-  [Afghanistan - Natural Disaster Incidents in 2019](#) OCHA Afghanistan
-  [Afghanistan - Conflict Data](#) Armed Conflict Location & Event Data Project
-  [Afghanistan - Casualties](#) OCHA Afghanistan


COORDINATION & CONTEXT

 6 Datasets

3W - Who Is Doing What Where

-  [Afghanistan - Operational Presence](#) OCHA Afghanistan
-  [Current IATI aid activities in Afghanistan](#) International Aid Transparency Initiative

Funding

-  [Afghanistan - Requirements and Funding Data](#) OCHA FTS

Affected Areas

-  [Whole of Afghanistan Assessment - Hard to Reach Dataset - August 2018](#) REACH Initiative

Conflict Events

-  [Afghanistan - Conflict Data](#) Armed Conflict Location & Event Data Project

Humanitarian Access

-  [Afghanistan - Natural Disaster Incidents in 2019](#) OCHA Afghanistan

Transportation Status

No Data

Damaged & Destroyed Buildings

-  [Afghanistan - Natural Disaster Incidents in 2019](#) OCHA Afghanistan



FOOD SECURITY & NUTRITION

4 Datasets

Food Security

- [Afghanistan - Acute Food Insecurity](#) HDX

Global Acute Malnutrition Rate

- [Afghanistan - Prevalence of Global Acute Malnutrition \(GAM\)](#) OCHA Afghanistan
- [ICA Afghanistan, 2019 - Prevalence of Global Acute Malnutrition \(GAM\), 2018](#) International Organization for Migration

Severe Acute Malnutrition Rate

No Data

Food Prices

- [Afghanistan - Food Prices](#) World Food Programme



GEOGRAPHY & INFRASTRUCTURE

5 Datasets

Administrative Divisions

- [Afghanistan Administrative Level 0-2 and UNAMA Region Boundary Polygons, Lines, and Points](#) OCHA Afghanistan

Populated Places

- [HOTOSM Afghanistan Populated Places \(OpenStreetMap Export\)](#) Humanitarian OpenStreetMap Team

Roads

- [HOTOSM Afghanistan Roads \(OpenStreetMap Export\)](#) Humanitarian OpenStreetMap Team
- [Afghanistan - Roads](#) OCHA Afghanistan

Airports

- [Airports in Afghanistan](#) OurAirports



HEALTH & EDUCATION

4 Datasets

Health Facilities

- [HOTOSM Afghanistan Health Facilities \(OpenStreetMap Export\)](#) Humanitarian OpenStreetMap Team
- [Impact of The Use of Explosive Weapons On Health Facilities](#) EWIPA Community
- [Afghanistan - Healthsites](#) Global Healthsites Mapping Project

Education Facilities

- [HOTOSM Afghanistan Education Facilities \(OpenStreetMap Export\)](#) Humanitarian OpenStreetMap Team

Affected Schools

No Data



POPULATION & SOCIO-ECONOMY

2 Datasets

Baseline Population

- [Afghanistan - Estimated Population 2016/2017 \(Archived\)](#) OCHA Afghanistan

Baseline Population by Age & Sex

- [Afghanistan - Estimated Population 2016/2017 \(Archived\)](#) OCHA Afghanistan

Affected Areas

- [Afghanistan - Poverty Rate Dataset - August 2018](#) HDX

8. CONTRIBUTING ORGANIZATIONS

The following 28 organizations share data that is included in the Data Grid. An organization on HDX can be a legal entity or an informal group.¹⁰ Organizations such as WFP and IOM share data (food prices and displacement tracking) across all Data Grid locations. This is also the case for OCHA's Financial Tracking Service (funding) and ACLED (casualties and conflict events). Other organizations may share a single dataset related to a specific crisis, such as the dataset on the location of schools in Yemen that has been shared by the Education Cluster for that country.

Organizations on HDX can be listed as the source or the contributor of the data. The entities listed below have created organizations on HDX and manage their data directly. Although most organizations are both the source and contributor for the data, there are cases where this varies. For instance, as part of its coordination role, OCHA aggregates data on humanitarian needs but the data is collected by multiple partners.¹¹

With modeled data, the dataset may include multiple sources but the organization that has done the modeling contributes the data. This is often the case with datasets on population estimates, food insecurity and poverty rates.

Armed Conflict Location & Event Data Project

Demographic and Health Surveys Program

Drew University

Education Cluster, Central African Republic

Education Cluster, Yemen

Education Sector, Cox's Bazar

Explosive Weapons in Populated Areas Community

Food and Agricultural Organization of the United Nations,
Somalia Water and Land Information Management

Food Security and Nutrition Working Group,
West and Central Africa

Global Healthsites Mapping Project

Humanitarian OpenStreetMap Team

Inter Sector Coordination Group, Bangladesh

InterAction

International Federation of Red Cross and
Red Crescent Societies

International Organization for Migration

Laboratorio de Ciencias Sociales, Colombia

Myanmar Information Management Unit

Nutrition Cluster, Somalia

OpenStreetMap, Democratic Republic of the Congo

OurAirports

REACH Initiative

United Nations Operational Satellite Applications
Programme

United Nation High Commissioner for Refugees

United Nations Office for the Coordination of
Humanitarian Affairs

United Nations Office for Disaster Risk Reduction

World Food Programme

World Health Organization

WorldPop

¹⁰ HDX includes over 250 active organizations. Some organizations on HDX have one entry for the entire organization, such as IOM. Others have multiple entries for individual field offices, such as OCHA Sudan and OCHA Afghanistan. See more here: <https://data.humdata.org/organization>

¹¹ The HDX team also manages an HDX organization on the platform for data that it pulls from open data platforms or in cases where the partner would like the data to be shared but does not want to manage the process directly. In these cases, HDX is listed as a contributor but we have not included HDX in the list of contributing organizations.

9. DATA FOR MODELING

As more data is collected and shared by partners, and as more data scientists enter the humanitarian sector, advanced applications of data to gain insight into operations are becoming commonplace. The push for anticipatory action (responding before a crisis escalates) has brought a growing interest in using data to create predictive models. This might include models to predict food insecurity, cross-border displacement or the impact of a weather event.

Models can only work based on reliable, timely and comprehensive data. We encourage data scientists to use the Data Grid as a tool to access the latest data to keep their projections up to date. We also encourage partners to share time series data to help understand the past and to be able to extrapolate future trends.

To improve transparency, the Centre is creating a catalogue of predictive models with basic information on ‘who is doing what, where and when’ and whether the model has been peer reviewed. We hope that the catalogue will become the main reference point for information on models in use in the humanitarian sector. We also hope that as more models are developed, the underlying data will be shared through HDX. As we see progress in this area, we may create an additional category in the Data Grid for modelled data. The information in the table below is not comprehensive. If your organization is working on a model, please be in touch with the Centre.

Who is doing what where with predictive models

PARTNERS	MODELS
World Food Programme Food and Agriculture Organization World Bank African Risk Capacity Famine Early Warning System Network	Analysis of food security indicators and projections for anticipatory humanitarian action
UN Country Teams	Projection of humanitarian needs for Humanitarian Response Plans
United Nations Children’s Fund National Aeronautics and Space Administration Met Office University of Florida	Global Cholera Risk Model to forecast risk of cholera outbreaks
United Nations High Commissioner for Refugees	JETSON model for cross-border displacement in Somalia
International Organization for Migration	Prediction of arrival in camps in Nigeria
The Netherlands Red Cross, 510 Global	Typhoon impact model for forecast-based financing in the Philippines
Internal Displacement Monitoring Centre Potsdam Institute for Climate Research ETH Zurich	Climate change projections of disaster-induced displacement risk
Danish Refugee Council Save the Children International	Mixed migration and forced displacement models
European Commission Joint Research Centre and partners	INFORM risk index for humanitarian crises
World Food Programme Pacific Disaster Center	Sudden-onset hazard impact model for rapid assessments

Annex A

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 Profile Locations

Vector or tabular data with coordinates representing the locations at which displaced people are gathered.

Casualties

Number of deaths and/or persons injured, disaggregated by location. Values can be cumulative totals or a time series of new deaths and/or injured persons.

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.

Affected Areas

Vector data or tabular data by administrative division which describe the type and/or severity of impacts geographically.

Damaged and Destroyed Buildings

Vector data with locations of damaged/destroyed buildings and an indication of damage level or tabular data indicating percentage or total number of buildings in each damage category by administrative divisions.

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.

Transportation Status

Vector or tabular data representing local transportation routes with an indication of status or current practicability.

Conflict Events

Vector data or tabular data with coordinates describing the location, date, and type of conflict event.

Funding

Tabular data listing the amount of funding provided by humanitarian cluster/sector.

Food Security & Nutrition



Food Security

Vector data representing the IPC phase classification or tabular data representing population or percentage of population by IPC phase and administrative division.

Global Acute Malnutrition Rate

Tabular data specifying the global acute malnutrition (GAM) rate by administrative division.

Severe Acute Malnutrition Rate

Tabular data specifying the severe acute malnutrition (SAM) rate by administrative division.

Food Prices

Time series prices for common food commodities at a set of location.

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.).

Affected Schools

Vector data or tabular data with coordinates representing education facilities that have been affected by a crisis with some indication of the nature of the effect and the operational status of each facility.

Population and Socio-economic Indicators



Population

Total population aggregated by administrative division.

Population by Age and Sex

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

All datasets on HDX are evaluated against the following criteria for inclusion in the Data Grid:

- Is the data in a common format? We include CSV, XLS, XLSX, SHP, etc. Formats like JSON, geopackage 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 subtotal 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.
- Are location references defined? The dataset should contain explicit geographic data (as in GIS data or tabular data with lat/long fields)? If not, the dataset should be joinable to an available dataset that defines those locations.
- Is the dataset comprehensive? 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? This comprehensiveness requirement means that crowd-sourced datasets, like those derived from OpenStreetMap, cannot always be considered complete, even though they may be the most complete dataset available.
- Is the dataset up-to-date? The dataset should be the latest available.

When a dataset is added to HDX, it may be automatically added to the Data Grid based on its metadata tags. However, within one day an HDX team member will evaluate the dataset to determine if it sufficiently meets the criteria to remain in the Data Grid.