OCHA centre for humdata HDX

THE STATE OF OPEN HUMANITARIAN DATA 2023

ASSESSING DATA AVAILABILITY ACROSS HUMANITARIAN CRISES

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This report was produced in February 2023 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 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.

LIST OF ABBREVIATIONS

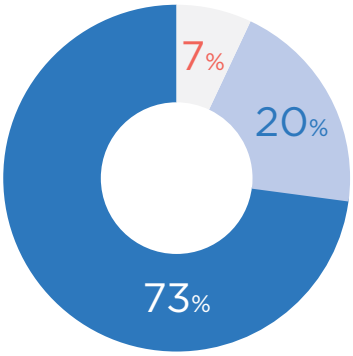
ACLED	Armed Conflict Location & Event Data Project
CH	Cadre Harmonisé
CODs	Common Operational Datasets
FAO	Food and Agriculture Organization of the United Nations
FTS	Financial Tracking Service
GAM	Global Acute Malnutrition
GIS	Geographic Information System
HDX	Humanitarian Data Exchange
HOT	Humanitarian OpenStreetMap Team
HRP	Humanitarian Response Plan
IATI	International Aid Transparency Initiative
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
OSM	OpenStreetMap
SAM	Severe Acute Malnutrition
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
WFP	World Food Programme
WHO	World Health Organization

1. INTRODUCTION

In our fourth year of producing *The State of Open Humanitarian Data*, we can report the highest levels yet for data availability across priority humanitarian operations. These gains can be attributed to the commitment of organizations to sharing and maintaining their data publicly. There was also strong demand for data about the world's largest humanitarian crises, from the war in Ukraine to drought and food insecurity in the Horn of Africa.

Our understanding of data availability and use comes from managing the Humanitarian Data Exchange (HDX), an open platform for finding and sharing data across crises and organizations.[1] In 2022, HDX was used by 1.5 million people in 233 countries and territories, an increase of 8 percent compared with 2021. Organizations added 3,700 new and updated datasets, bringing the total to more than 20,400 datasets, which were downloaded over 1.8 million times. Although HDX includes data about all countries in the world, we concentrate here on 25 locations with humanitarian response plans (HRPs).[2] In 2022, 32 percent of all datasets downloaded from HDX were related to an HRP location.[3]

At the start of 2023, we estimate that 73 percent of relevant, complete crisis data is available across 25 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 25 Data Grids include 459 unique datasets, with a range of 13-22 datasets per location.



Total Percentage Data Complete

73%

Total Percentage Data Incomplete

20%

Total Percentage No Data

7%

Data completeness: Complete Incomplete No Data

This report contains details on the data available for each location, category and sub-category covered in the Data Grids as of 31 December 2022. It includes a country deep-dive for South Sudan and showcases the work of the Humanitarian OpenStreetMap Team (HOT) on community mapping. It also provides examples of how climate impact data is being used to predict the impacts of a hazard on lives and livelihoods and trigger actions to mitigate them.

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.)

[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] We use Mixpanel, a third-party analytics service, to understand the behavior of users on the HDX platform. We do not send identifying information to Mixpanel.

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 2022, HDX maintained Data Grids for 25 locations and 20 sub-categories. The number of locations decreased by two from 2021 given that Zimbabwe and Pakistan no longer had humanitarian response plans in 2022. Two sub-categories were removed – transportation status and affected areas – and one was added for climate impact. (See Annex C for changes to the Data Grids.)

We recognize the valuable contributions of all data-sharing organizations publishing data on HDX, and in particular the 18 organizations that populate the Data Grids, with reference to 175 data sources. Datasets included in the Data Grids are downloaded almost five times more than the average dataset on HDX.[4] Trusted partnerships and focused advocacy efforts have led to many new or updated datasets in the Data Grids this year, including:

- Global malnutrition prevalence rates disaggregated by province in Afghanistan, contributed by OCHA Afghanistan and sourced to UNICEF.[5]
- The number of displaced people by sub-prefecture in the Central African Republic (CAR), validated by the Commission des Mouvements de Population and shared by OCHA CAR.[6]
- The impact of floods with disruption to affected populations, destroyed infrastructure and vegetation in Niger, contributed by OCHA Niger and sourced to the Ministère de l'Action Humanitaire et de la Gestion des Catastrophes du Niger.[7]
- Education facilities by type and location in Somalia, compiled by UNICEF and contributed by OCHA Somalia.[8]
- Sub-national sex and age disaggregated population statistics for Sudan with updated projections for 2022, contributed by UNFPA and sourced to the Central Bureau of Statistics for Sudan.[9]

There were dozens of high-value datasets available on HDX throughout 2022 that go beyond the scope of the Data Grids. A few examples include:

- The list of voyages completed under the Black Sea Grain Initiative since August 2022, compiled and shared by the Joint Coordination Centre.[10]
- Acute food insecurity estimates, cross-border trade and staple food prices for 37 countries, shared by the Famine Early Warning Systems Network.[11]
- Historical typhoon impact data covering the number of people affected and homes damaged from 2014 to 2020 in the Philippines, compiled and shared by the Netherlands Red Cross.[12]
- High-resolution population density maps and demographic estimates and the Facebook Social Connectedness Index, covering dozens of countries and shared by Meta.[13]

[4] In 2022, Data Grid datasets were downloaded, on average, 585 times each whereas the average dataset on HDX was downloaded 116 times.
[5] **https://data.humdata.org/dataset/afghanistan-malnutrition-prevalence**
[6] **https://data.humdata.org/dataset/statistiques-detaillees-des-sites-pdis-en-republique-centrafricaine**
[7] **https://data.humdata.org/dataset/niger-situation-des-inondations**
[8] **https://data.humdata.org/dataset/somalia-education-facilities**
[9] **https://data.humdata.org/dataset/cod-ps-sdn**
[10] **https://data.humdata.org/dataset/black-sea-grain-initiative-vessel-movements**
[11] **https://data.humdata.org/organization/fewsnet**
[12] **https://data.humdata.org/dataset/philippines-typhoon-impact-data-2014-2020**
[13] **https://data.humdata.org/organization/facebook**

2. KEY MESSAGES

- The availability of core data for priority humanitarian operations is at its highest levels in the past four years, but gaps remain with climate impact data, acute malnutrition prevalence and access constraints, among other areas.

- Since 2019, when the Data Grids were first introduced on HDX, we have seen an increase in data completeness from 54 percent to 73 percent in locations with humanitarian response plans. The categories with no data have fallen from 25 percent to 7 percent (see chart below).

- These gains can be attributed to the strong commitment of organizations to sharing and maintaining their data publicly through the HDX platform.

- There was strong demand for data about the world's largest humanitarian crises, from the war in Ukraine to drought and food insecurity in the Horn of Africa. Data covering Ukraine, Ethiopia and Nigeria was the most downloaded in 2022.

- In 2022, HDX was used by 1.5 million people in 233 countries and territories. Around 1.8 million datasets were downloaded throughout the year.

- We estimate that 73 percent of relevant, complete crisis data is available across 25 locations with humanitarian operations, based on the analysis of the HDX Data Grids. If we add the data that is relevant but incomplete, the total is 93 percent. This leaves about 7 percent of the Data Grids with no data.

- Over two dozen new organizations joined HDX in 2022, bringing the total to almost 300 organizations. Eighteen organizations contribute data that is included in the Data Grids.

- The 25 Data Grids included 459 unique datasets, with a range of 13-22 datasets per location. Datasets included in the Data Grids are downloaded almost five times more than the average dataset on HDX.

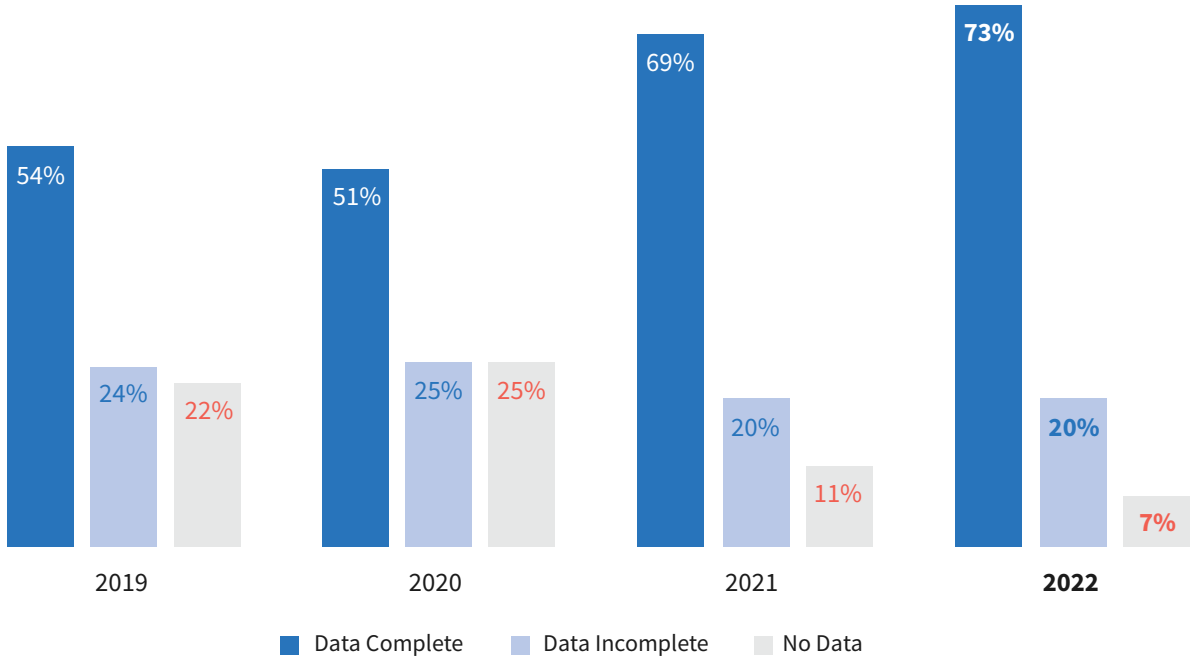
- South Sudan, Afghanistan and Somalia have the most complete Data Grids. Venezuela, Syria and Libya are the countries with the largest data gaps.

- The most complete Data Grid sub-categories are funding, administrative divisions, baseline population, IDPs, refugees and persons of concern, and food prices, owing to the work of OCHA, UNFPA, IOM, UNHCR, and WFP.

- 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.

- A priority for 2023 will be a continued focus on climate impact data, which is essential for understanding and predicting the future impacts of hazards on vulnerable populations. We will also explore implementing data quality grades or scores for some datasets included in the Data Grids.

DATA GRID OVERALL COMPLETENESS 2019 - 2022



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

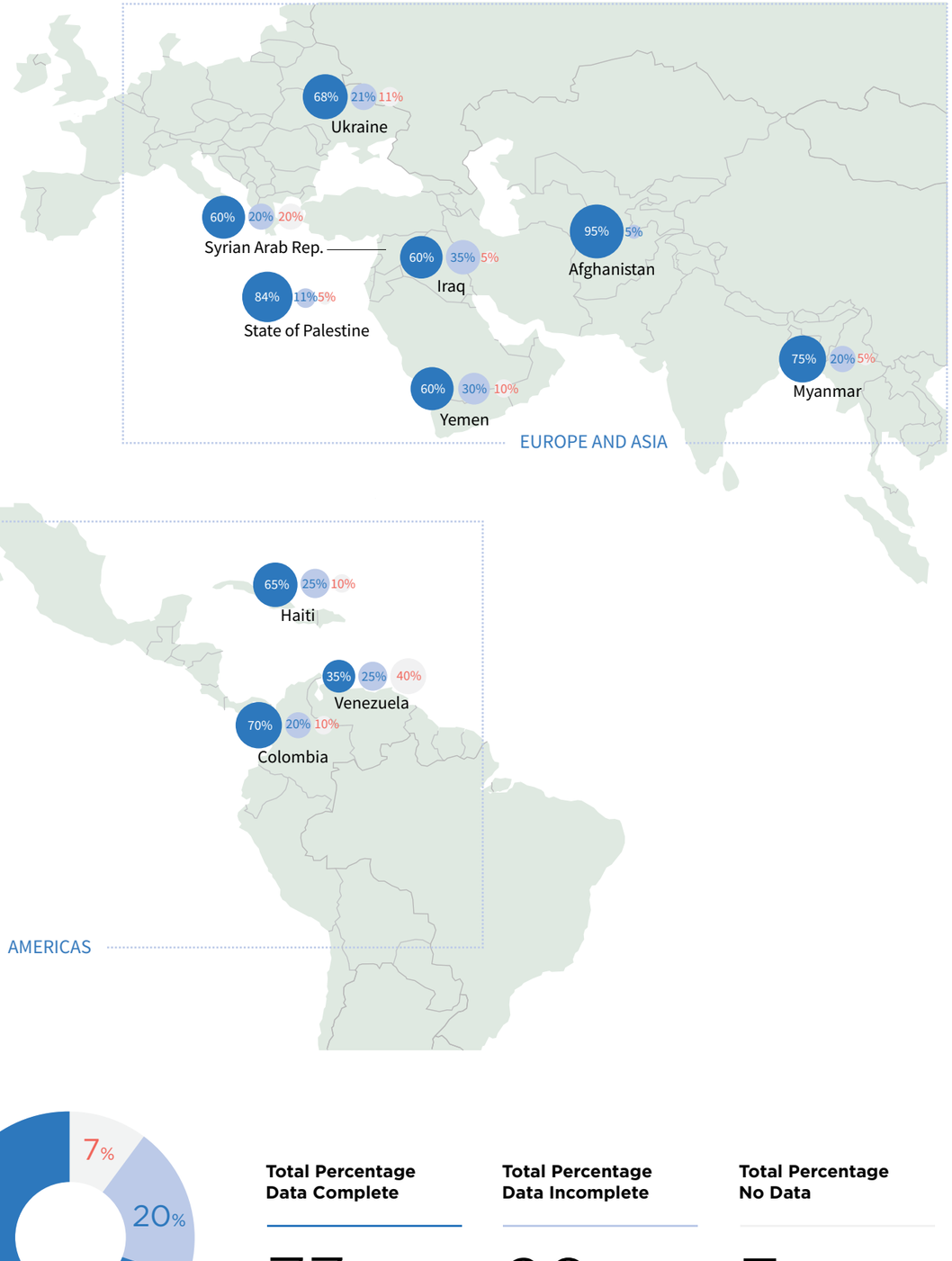
- Access constraints on the delivery of humanitarian assistance (potential sources: national governments, access working groups, OCHA).
- Climate impact data covering the previous 10 years of hazards and their impact on people, infrastructure and vegetation (potential sources: national governments, disaster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

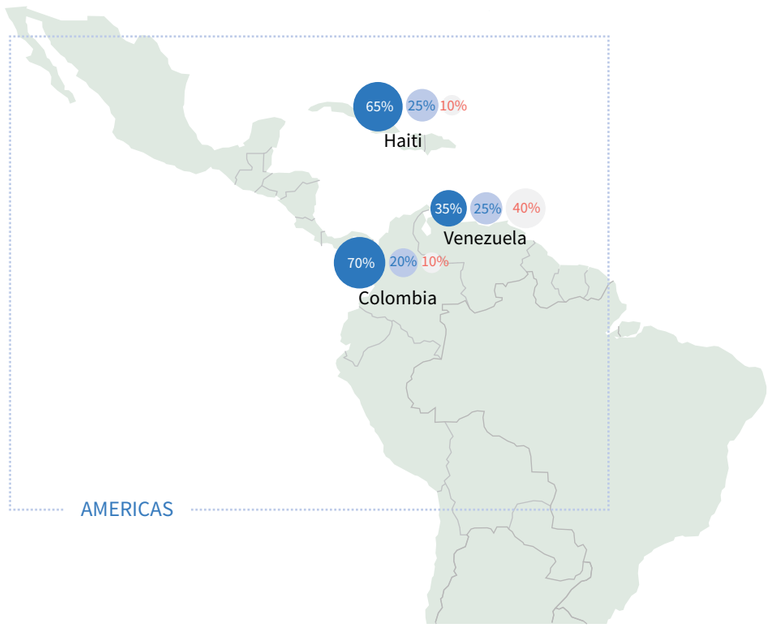
In cases where data is sensitive, organizations can use HDX Connect[14] to share only the metadata and make the underlying data available by request. Alternatively, an incomplete or aggregated version of the dataset that removes sensitive information (such as locations) can be shared publicly.[15]

[14] HDX Connect datasets still contribute to the completeness of a Data Grid. Learn more: **bit.ly/hdx-connect**

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

3. 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

25

Number of Categories

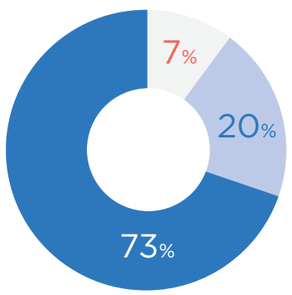
6

Number of Sub-Categories

20

Number of Contributing Organizations

18



Total Percentage Data Complete

73%

Total Percentage Data Incomplete

20%

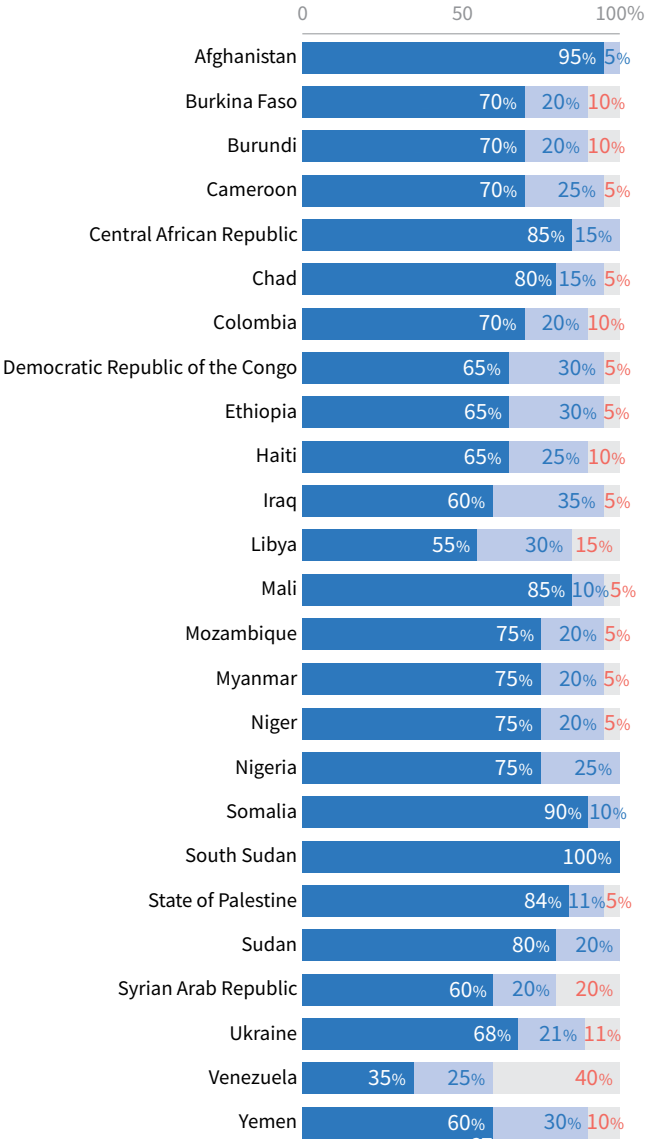
Total Percentage No Data

7%

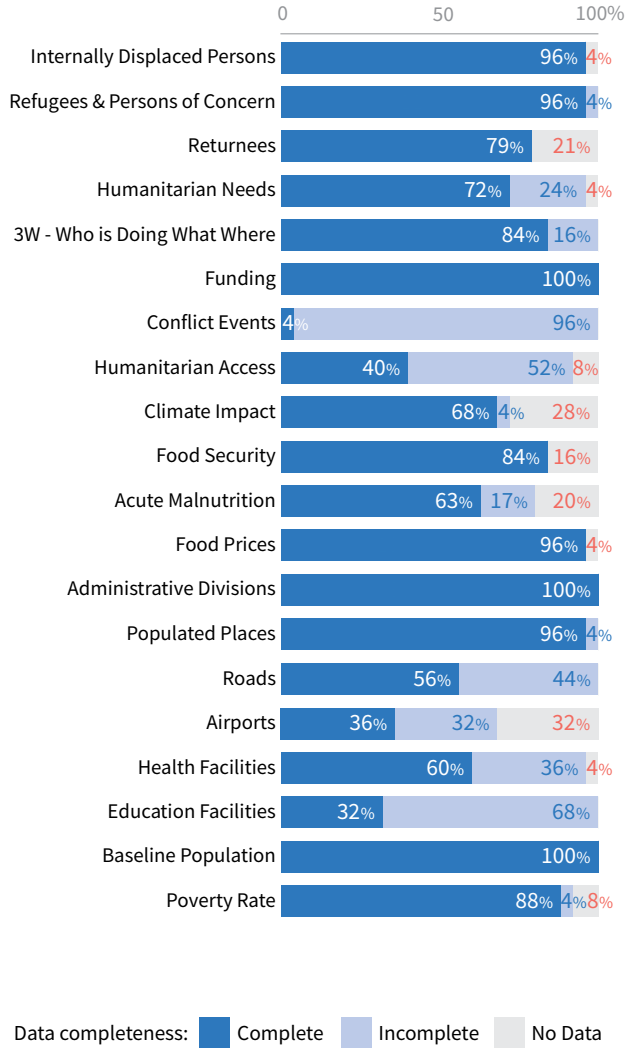
4. COMPLETENESS BY LOCATION, CATEGORY AND SUB-CATEGORY

By the end of 2022, there were improvements in data availability across location, category and sub-category. Six countries – Afghanistan, CAR, Nigeria, Somalia, South Sudan and Sudan – no longer have any sub-categories with missing data, although some data remains incomplete in geographic coverage or frequency. The geography and infrastructure category become more complete with data added on populated places for a number of countries. Overall, the sub-categories with no data decreased from 11 percent to 7 percent year on year.

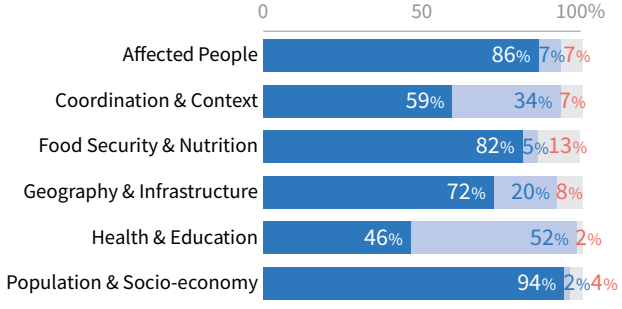
BY LOCATION



BY SUB-CATEGORY



BY CATEGORY



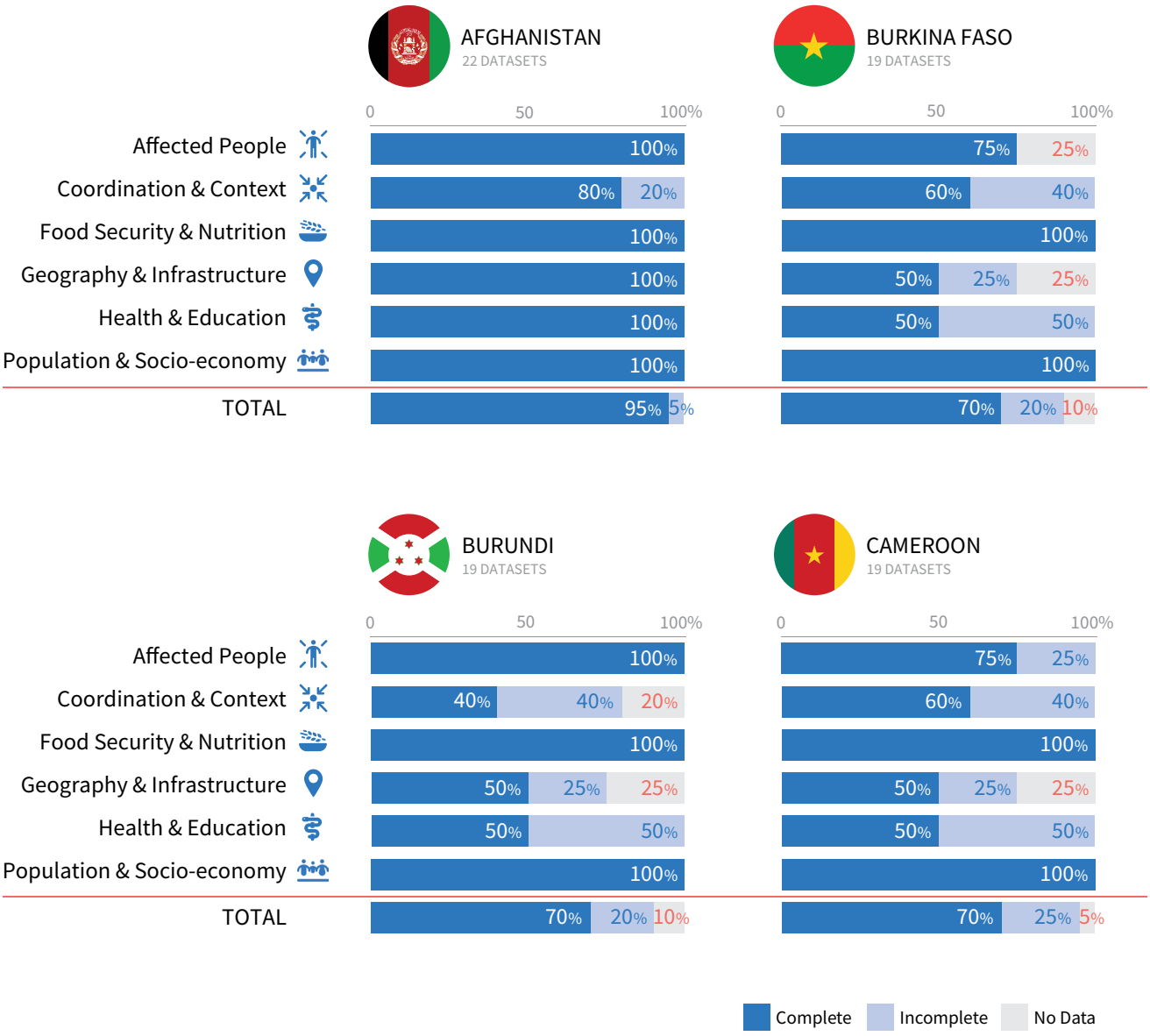
5. COMPLETENESS BY LOCATION AND CATEGORY

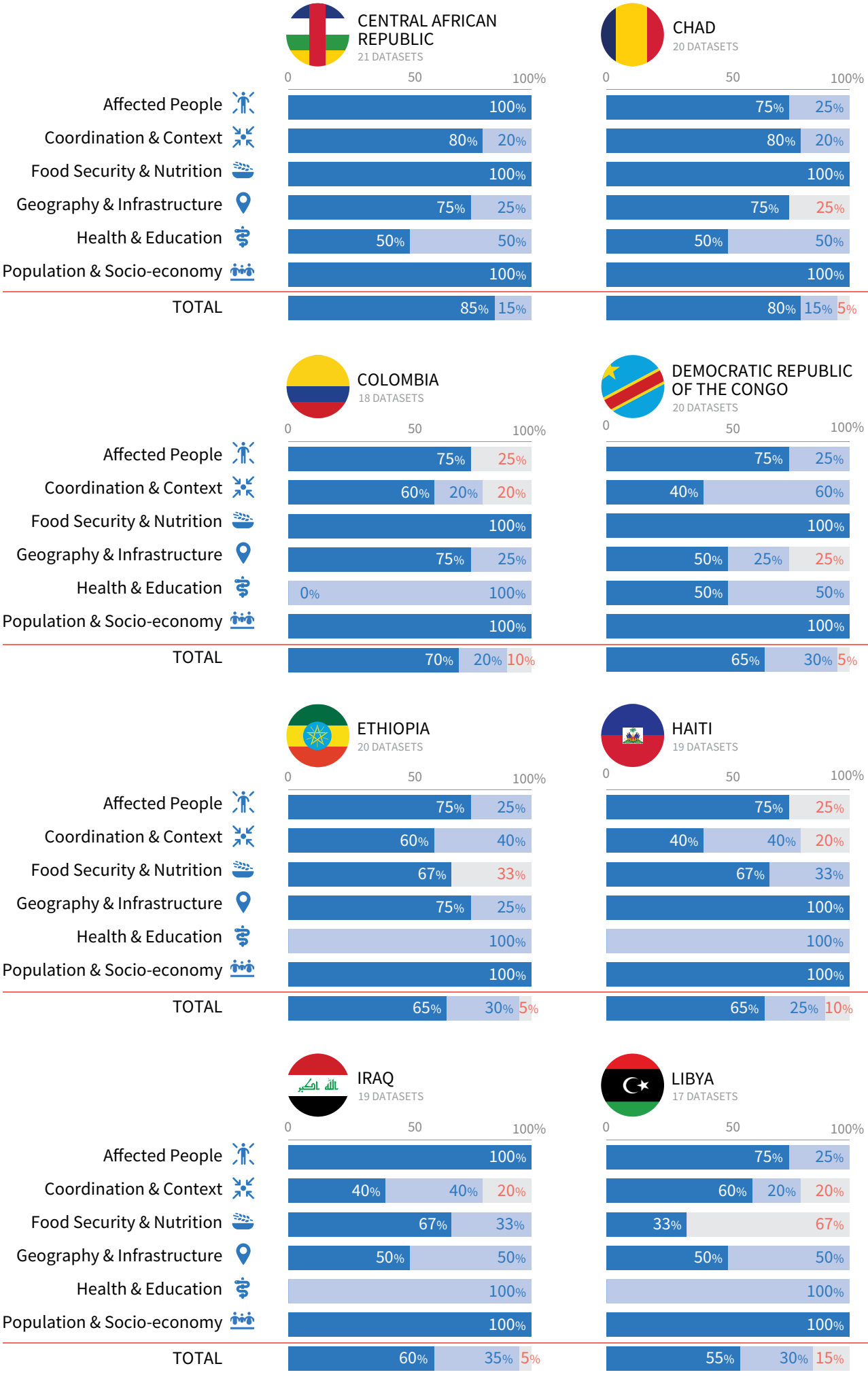
South Sudan (100 percent), Afghanistan (95 percent) and Somalia (90 percent) are the countries with the highest levels of data completeness. The next most-complete locations for data are the Central African Republic (85 percent) and Mali (85 percent).

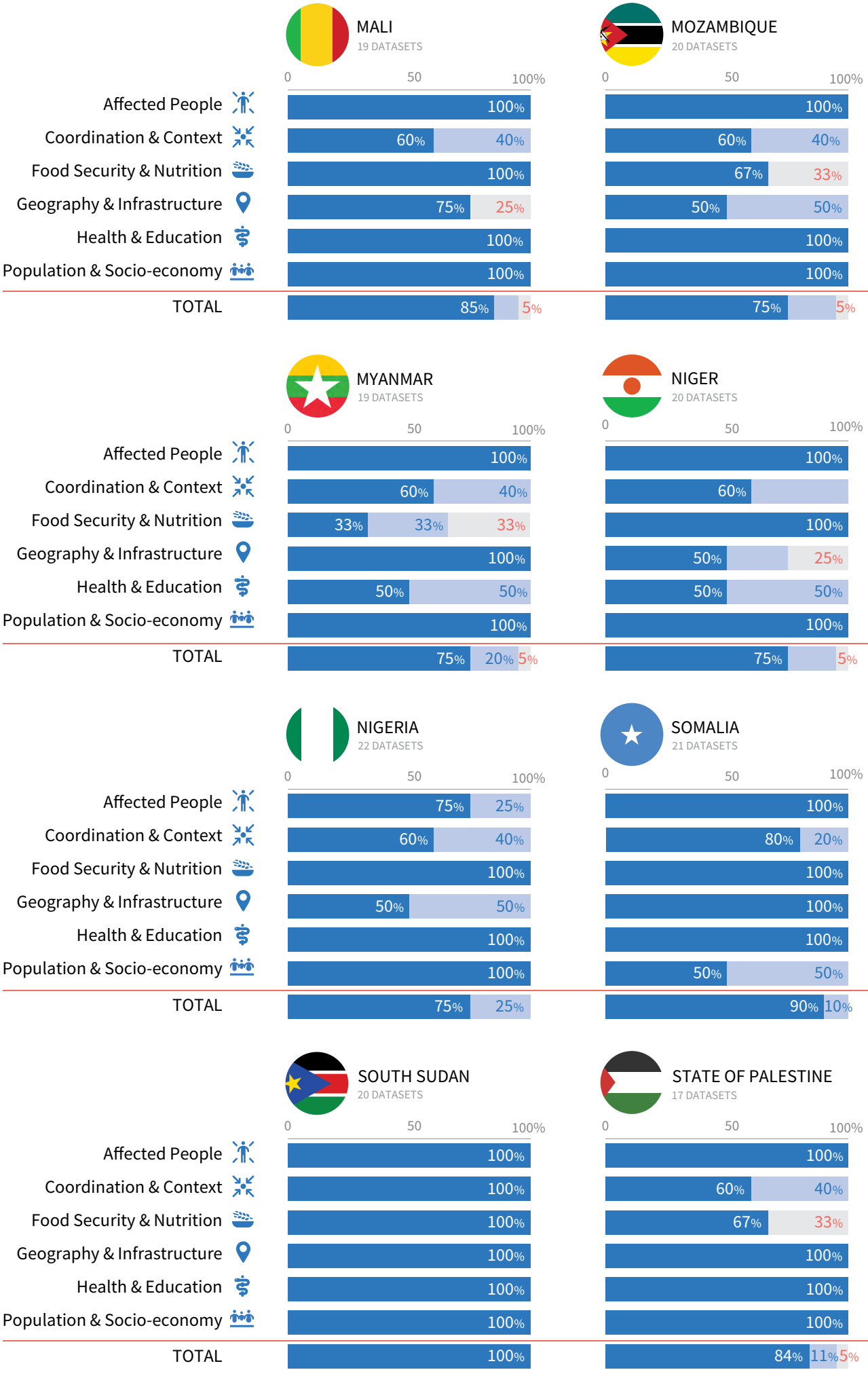
Venezuela (40 percent of the sub-categories with no data), Syria (20 percent of the sub-categories with no data) and Libya (15 percent of the sub-categories with no data) are the countries with the largest data gaps.

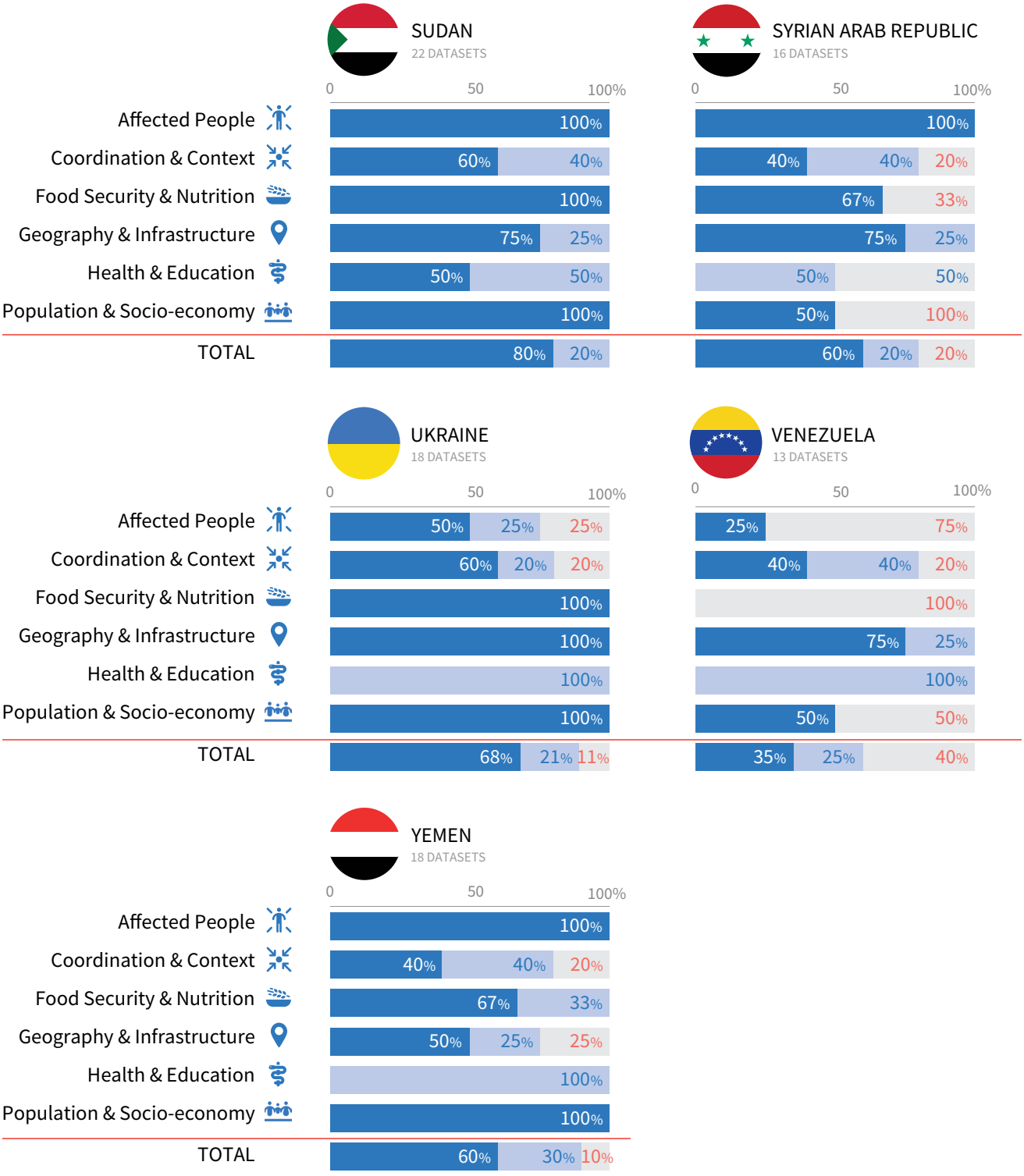
The most complete categories are population and socio-economy (94 percent), affected people (86 percent) and food security and nutrition (81 percent). Health and education is the least complete category at 46 percent, owing to incomplete data for the location of health and education facilities. The most popular category, in terms of total dataset downloads, is geography and infrastructure.

A total of 459 unique datasets are included in the 25 Data Grids. Each Data Grid includes between 13 to 22 datasets.




CENTRAL AFRICAN REPUBLIC
21 DATASETS
CHAD
20 DATASETS
COLOMBIA
18 DATASETS
DEMOCRATIC REPUBLIC OF THE CONGO
20 DATASETS
ETHIOPIA
20 DATASETS
HAITI
19 DATASETS
IRAQ
19 DATASETS
LIBYA
17 DATASETS
0
50
100%
Affected People
Coordination & Context
Food Security & Nutrition
Geography & Infrastructure
Health & Education
Population & Socio-economy
TOTAL
100% 80% 20% 100% 75% 25% 50% 50% 100% 85% 15%
75% 25% 80% 20% 100% 75% 25% 50% 50% 100% 80% 15% 5%
75% 25% 60% 20% 20% 100% 75% 25% 0% 100% 100% 70% 20% 10%
75% 25% 40% 60% 100% 50% 25% 25% 50% 50% 100% 65% 30% 5%
75% 25% 60% 40% 67% 33% 75% 25% 100% 100% 65% 30% 5%
75% 25% 40% 40% 20% 67% 33% 100% 100% 100% 65% 25% 10%
100% 40% 40% 20% 67% 33% 50% 50% 100% 100% 60% 35% 5%
75% 25% 60% 20% 20% 33% 67% 50% 50% 100% 100% 55% 30% 15%

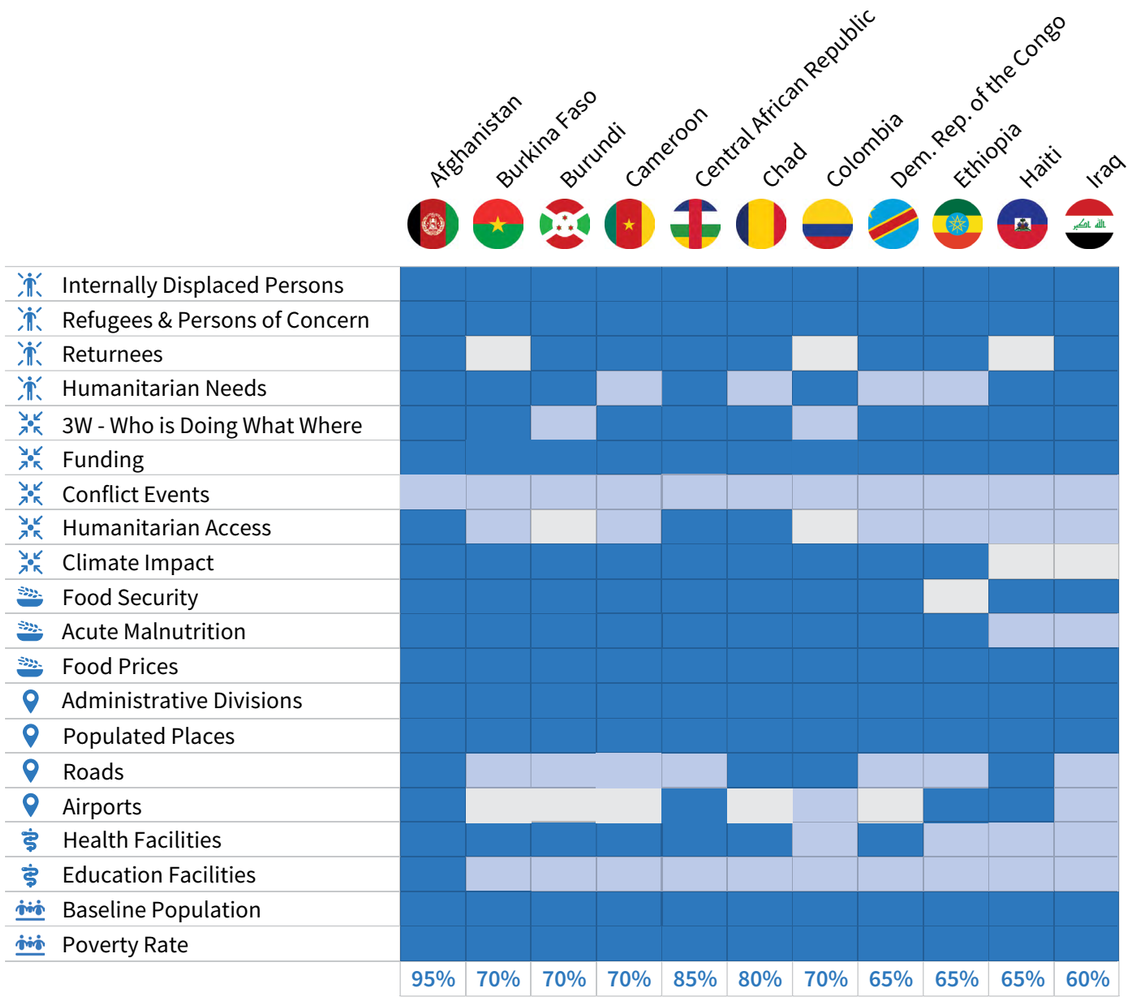

MALI
19 DATASETS
MOZAMBIQUE
20 DATASETS
MYANMAR
19 DATASETS
NIGER
20 DATASETS
NIGERIA
22 DATASETS
SOMALIA
21 DATASETS
SOUTH SUDAN
20 DATASETS
STATE OF PALESTINE
17 DATASETS
0
50
100%
Affected People
Coordination & Context
Food Security & Nutrition
Geography & Infrastructure
Health & Education
Population & Socio-economy
TOTAL
100%
60% 40%
67% 33%
75% 25%
50% 50%
85% 5%
75% 5%
33% 33% 33%
75% 20% 5%
80% 20%
90% 10%
84% 11% 5%

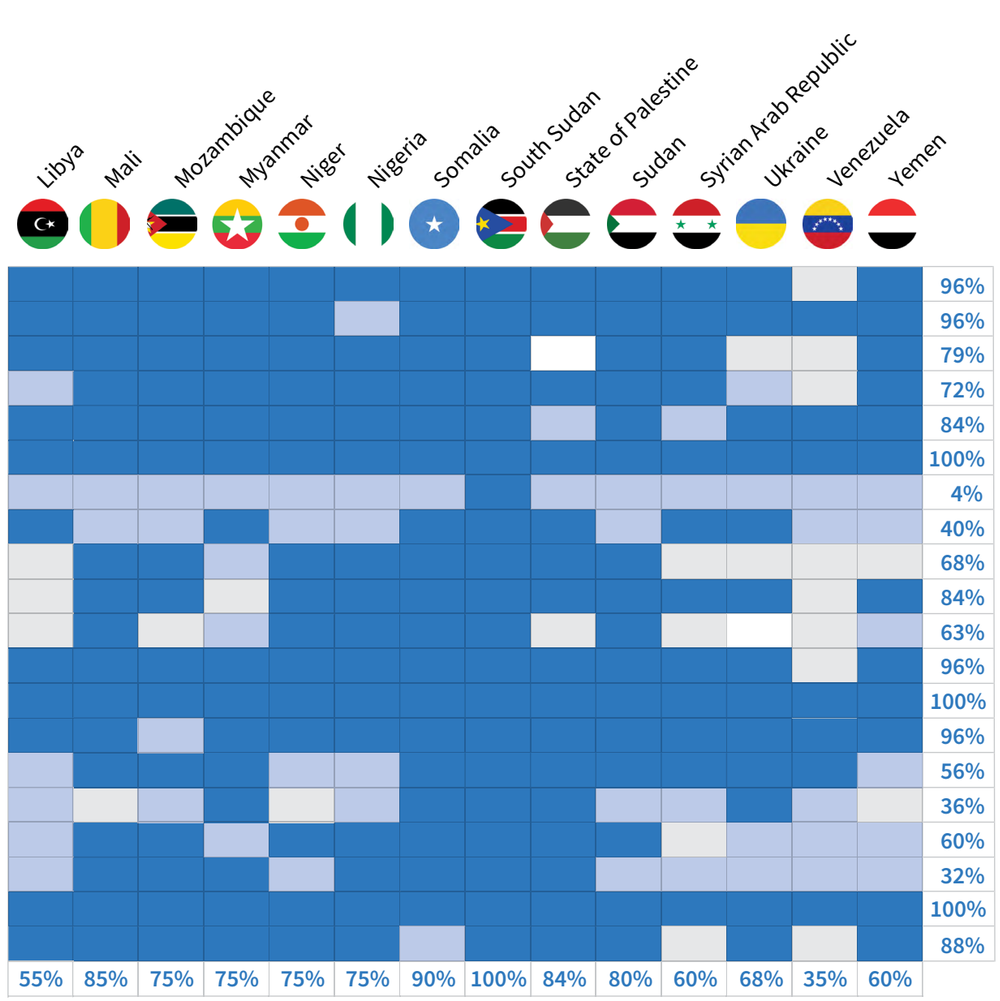

SUDAN
22 DATASETS
SYRIAN ARAB REPUBLIC
16 DATASETS
UKRAINE
18 DATASETS
VENEZUELA
13 DATASETS
YEMEN
18 DATASETS
0
50
100%
Affected People
Coordination & Context
Food Security & Nutrition
Geography & Infrastructure
Health & Education
Population & Socio-economy
TOTAL

6. COMPLETENESS BY LOCATION AND SUB-CATEGORY

The most complete sub-categories are funding, administrative divisions and baseline population (all at 100 percent) followed by IDPs, refugees and persons of concern, and food prices (all at 96 percent), owing to the contributions of OCHA, UNFPA, IOM, UNHCR and WFP. Education facilities (32 percent), airports (36 percent) and humanitarian access (40 percent) are the most challenging sub-categories to complete.

This view of the Data Grids also shows where data is consistently shared across all locations but is considered incomplete. This is the case for conflict events data (political violence, civilian-targeting and demonstrations) which is shared on a weekly basis for each country. ACLED makes more granular datasets available through a registration process on their website so the data on HDX is therefore considered incomplete.





7. COUNTRY DEEP-DIVE: SOUTH SUDAN

There has been steady progress with the South Sudan Data Grid over the past year owing to the engagement of the OCHA Country Office and humanitarian partners. South Sudan became the first location to reach 100 percent completeness, up from 76 percent complete last year.

The Data Grid includes 20 unique datasets shared by nine organizations. Together, these datasets were downloaded 8,348 times in 2022. The most popular dataset was the administrative boundaries for the country. Another sought-after dataset was the number of people affected by flooding, by state and county, over the previous two years, which is included in the climate impact sub-category.

AFFECTED PEOPLE

3 Datasets

Internally Displaced Persons

- South Sudan Displacement - [IDPs, Returnees] - Baseline Assessment [IOM DTM]
  International Organization for Migration

Refugees & Persons of Concern

- South Sudan - Data on forcibly displaced populations and stateless persons UNHCR - The UN Refugee Agency

Returnees

- South Sudan Displacement - [IDPs, Returnees] - Baseline Assessment [IOM DTM]
  International Organization for Migration

Humanitarian Needs

- South Sudan: Humanitarian Needs Overview
  OCHA South Sudan

COORDINATION & CONTEXT

6 Datasets

3W - Who Is Doing What Where

- South Sudan : Operational Presence OCHA South Sudan
- Current IATI aid activities in South Sudan
  International Aid Transparency Initiative

Funding

- South Sudan - Requirements and Funding Data
  OCHA FTS

Conflict Events

- South Sudan: Access and Security Incidents
  OCHA South Sudan

Humanitarian Access

- South Sudan: Humanitarian Access Incidents
  OCHA South Sudan

Climate Impact

- South Sudan: Flood Data OCHA South Sudan

FOOD SECURITY & NUTRITION

3 Datasets

Food Security

- South Sudan: Acute Food Insecurity Country Data
  Integrated Food Security Phase Classification

Acute Malnutrition

- South Sudan: Nutrition GAM rates UNICEF South Sudan

Food Prices

- South Sudan - Food Prices World Food Programme

GEOGRAPHY & INFRASTRUCTURE

4 Datasets

Administrative Divisions

- South Sudan - Subnational Administrative Boundaries
  OCHA South Sudan

Populated Places

- South Sudan: Settlements OCHA South Sudan

Roads

- South Sudan: Road Networks (IHDP) HDX

Airports

- South Sudan: Airstrip OCHA South Sudan

HEALTH & EDUCATION

2 Datasets

Health Facilities

- Health Facilities in Sub-Saharan Africa
  OCHA South Sudan

Education Facilities

- South Sudan - Village Assessment - Education Facilities [IOM DTM] International Aid DTM

POPULATION & SOCIO-ECONOMY

2 Datasets

Baseline Population

- South Sudan - Subnational Population Statistics
  OCHA South Sudan

Poverty Rate

- South Sudan: Resilience Capacity Index
  OCHA South Sudan

8. ORGANIZATION DEEP-DIVE: HUMANITARIAN OPENSTREETMAP TEAM

The Humanitarian OpenStreetMap Team, known by its acronym HOT, supports humanitarian action and community development through open mapping. When disaster strikes anywhere in the world, HOT is able to mobilize thousands of volunteers who come together online and on the ground to create open data that enables disaster responders to reach those in need.

The data is derived from a crowdsourcing model that leverages OpenStreetMap (OSM), a free, open source, editable map of the world. The data is contributed to OSM in one of three ways:

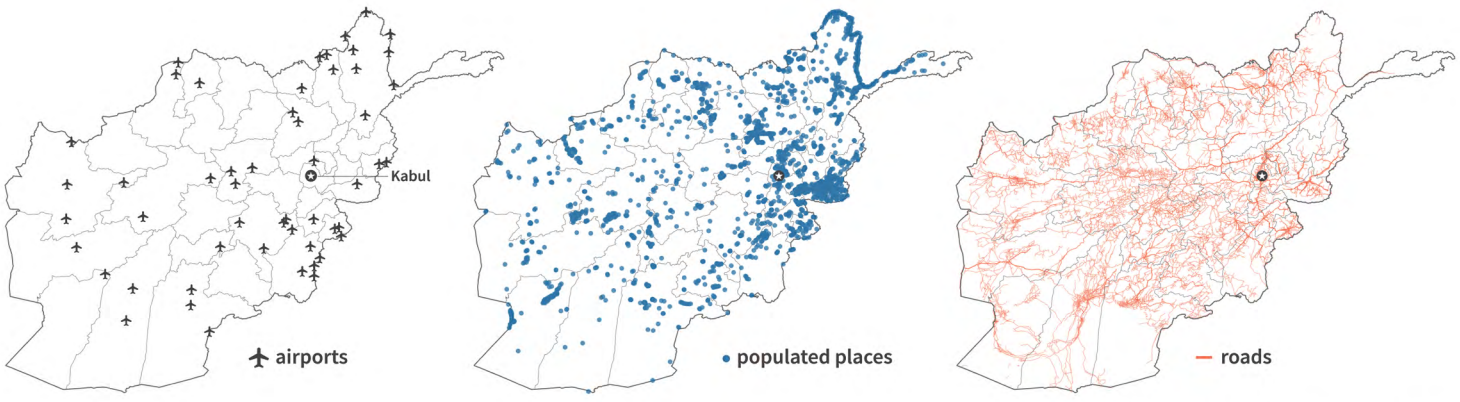
- In-person, mobile data collection that is focused on the location of specific infrastructure, such as education and health facilities.
- Remote sensing and digitization from satellite imagery or unmanned aerial vehicles to create building and road datasets.
- Importing existing datasets from governments and organizations.

Given their commitment to open source data and tools, HOT became one of the first organizations to join HDX when it launched in 2014. Since then, they have added over 4,400 datasets covering 249 countries and territories, which currently accounts for about 20 percent of all data on HDX. The data includes the essential ingredients for map making: road networks; populated places; airport, railways and port locations; education and health facilities; and rivers and waterways. These details are critical for humanitarians to understand the physical geography and infrastructure of a country, and this data is consistently among the most downloaded on HDX.

The humanitarian open mapping community, which is currently estimated to be around 423,000 people from around the world,[16] is continuously updating the data on OpenStreetMap. The HOT and HDX teams realized early on that it would be impossible to keep the datasets on HDX up-to-date without an automated process. In 2016, we worked together to build an extension to the HOT Export Tool – called the HOT Bot[17]– which sends the data to HDX on a regular cycle. This eliminates the complex task of configuring a custom export in the HOT Export Tool and pre-packages the data in common GIS data formats broken down by country and region.

The datasets exported by HOT are included in the Data Grids if they are deemed to be the best available source for a sub-category. This determination is made by the HDX team in collaboration with the Information Management Working Group in each country. In 2022, data attributed to HOT was included in the Data Grids for nine countries covering three sub-categories – airports, populated places and roads.

AFGHANISTAN AIRPORTS, POPULATED PLACES AND ROADS



[16] Source: **https://tasks.hotosm.org/**

[17] **https://data.humdata.org/user/osm2hdx**

This includes:

- Afghanistan (airports, populated places, roads)
- Burkina Faso (roads)
- Burundi (populated places, roads)
- Cameroon (roads)
- Central African Republic (roads)
- Democratic Republic of the Congo (roads)
- Libya (roads)
- Niger (roads)
- Ukraine (roads)

Open mapping data is not included in the Data Grids in cases when it is too sensitive to share (e.g., the location of health facilities in an active conflict), or when there is a more complete version of the data available from an official source.

For HOT, data quality is a primary concern. Additions and improvements are contributed by a large number of partners in the humanitarian and development space, as well as commercial partners using OSM data in their workflows and products. To complement existing data validation, HOT and HDX are working to provide additional insights into factors such as data accuracy and completeness.

In 2022, the data exported by HOT to HDX was downloaded almost 67,000 times by users in 194 countries and territories. Road network data for Afghanistan, Ethiopia and Pakistan were among the most popular. Datasets shared by HOT will continue to be in high demand at the onset of a new crisis, providing first responders with the geographical awareness that is essential for humanitarian action.

We are grateful to Humanitarian OpenStreetMap for being a steadfast data sharing partner. Visit their site[18] to learn more about becoming part of the humanitarian open mapping community and explore how to request, create and use geospatial data for development and humanitarian purposes.

[18] **https://www.hotosm.org/community/**

9. CONTRIBUTING ORGANIZATIONS

Out of the 293 organizations[19] sharing data on HDX, 18 contribute data that is included in the Data Grids, with reference to 175 data sources. These 18 organizations are consistently among the top organizations on HDX with the most datasets downloaded. OCHA is counted as a single organization even though over 30 field offices[20] and two entities at Headquarters contribute data to HDX. In total, OCHA field offices contributed 214 datasets or 43 percent of all data in the Data Grids.

Armed Conflict Location & Event Data Project

Food Security and Nutrition Working Group, West and Central Africa

Global Healthsites Mapping Project

Geo-Referenced Infrastructure and Demographic Data for Development

Humanitarian OpenStreetMap Team

Insecurity Insight

Integrated Food Security Phase Classification

International Aid Transparency Initiative

International Organization for Migration

Myanmar Information Management Unit

United Nations Office for the Coordination of Humanitarian Affairs

OurAirports

Oxford Poverty & Human Development Initiative

REACH Initiative

United Nations Population Fund

United Nations High Commissioner for Refugees

United Nations Children's Fund[21]

World Food Programme

[19] An organization on HDX can be a legal entity or an informal group and may be listed as the source or the contributor of the dataset. The entities listed in this section 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 and shares data on humanitarian needs but the data is collected by multiple partners.

[20] **https://www.unocha.org/where-we-work/ocha-presence**

[21] This covers UNICEF offices in two countries: DRC and South Sudan.

10. CLIMATE IMPACT DATA

The most significant change to the Data Grids in 2022 was the inclusion of a new sub-category for climate impact data. This data should contain the current and historical impacts of climate events relating to floods, droughts and storms, as well as their location, date and impact on affected populations, infrastructure and/or vegetation. By the end of the year, this sub-category was 68 percent complete with data missing for seven countries with humanitarian operations.

With the climate crisis intensifying emergencies around the world, humanitarians are increasingly having to incorporate climate data and forecasts into their analysis and planning. This is especially the case for anticipatory action, which relies on models to predict the impacts of a hazard on lives and livelihoods and trigger actions to mitigate them.

Since 2019, the Centre has supported OCHA's anticipatory action pilots for various hazards in twelve countries.[22] This has ranged from drought in Chad and Somalia to floods in Nepal and Bangladesh. Our technical support involves conducting historical analysis of the hazard and its impact, assessing available models and forecasts, and designing trigger mechanisms for the release of funds. By the end of 2022, the UN Central Emergency Response Fund had released $89 million[23] in funding for anticipatory action based on trigger mechanisms developed by the Centre and OCHA offices.

Three types of climate data are used to build, evaluate and validate trigger mechanisms:

- data on what happened (the hazard)
- data on the damage (the impact)
- data on what is likely to happen next (the forecast)

The data that is most often missing for this analysis is the impact data. Without it, we are not able to model the expected impacts of future shocks.

As an example, in Niger we used household survey data to understand the impact of seasonal rainfall levels on livelihoods. Once we understood this, we were able to use seasonal rainfall forecasts to inform humanitarian partners about the likely impact of a below average season. Similarly, in the Philippines[24] we worked with the Netherlands Red Cross to use historical typhoon track and rainfall data to train a model that estimates housing damage caused by typhoons. This model is now being used to estimate the expected impact based on typhoon forecasts provided by meteorological institutions. A historical typhoon impact dataset,[25] covering 2014 to 2020, is now available on HDX.

By the end of 2022, eighteen climate impact datasets had been added to HDX and the Data Grids, in collaboration with OCHA offices and partners. A few examples include:

- Mozambique: Data on the number of people affected by cyclones at the district level from 2016 to 2022, shared by OCHA Mozambique and produced by Instituto Nacional de Gestão e Redução do Risco de Desastres.[26]
- Mali: Data on the number of people affected, houses destroyed, water and sanitation infrastructure damaged and livestock lost as the results of seasonal floods, shared and maintained by OCHA Mali.[27]
- Burundi: Data on the impact from floods and heavy rains on populations, infrastructure and/or vegetation, shared by IOM via OCHA Burundi.[28]

[22] **https://centre.humdata.org/anticipatory-action/**

[23] The $89 million has been allocated for drought in Niger, Somalia and Ethiopia; for flooding in Bangladesh, South Sudan and Nepal; and for cholera in the Democratic Republic of the Congo.

[24] **https://centre.humdata.org/peer-review-of-510s-typhoon-model-and-its-use-in-the-philippines/**

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

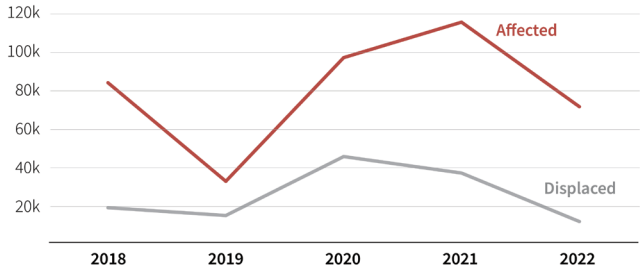
[26] **https://data.humdata.org/dataset/impact-of-cyclones-2016-to-2022**

[27] **https://data.humdata.org/dataset/mali-suivi-des-inondations**

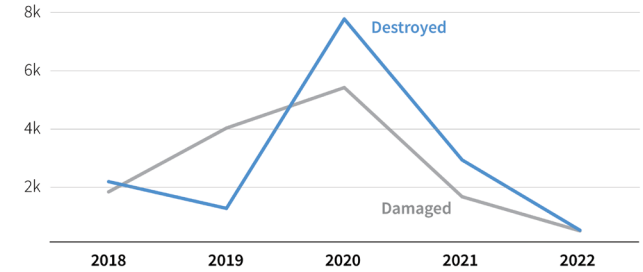
[28] **https://data.humdata.org/dataset/burundi-emergencies-tracking-tool-ett**

BURUNDI: CLIMATE IMPACT TRENDS 2018-2022

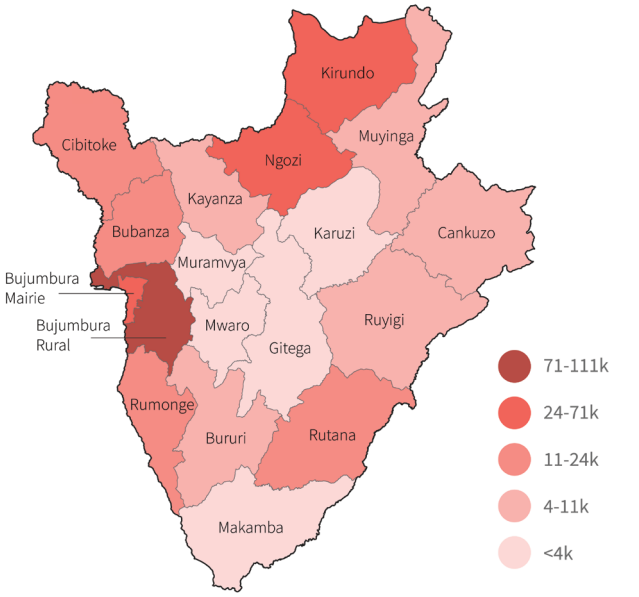
People affected and displaced



Houses destroyed and damaged



Total people affected by province 2018-2022



IOM ETT via HDX | Creation date: 02 February 2023

Beyond the Data Grids, we are also looking to make a wider set of climate data more accessible to the HDX community. For example, WFP recently shared earth observation datasets[29] on rainfall and vegetation health covering 33 countries with either a humanitarian response plan or an anticipatory action framework. With this data, humanitarians can assess current drought conditions and also analyze historical trends to anticipate the impact of rainy seasons.

Despite this progress, climate impact data remains scarce for certain types of climate shocks and locations and the historical time period is often limited. In the year ahead, we plan to share practical guidance on how to interpret and use common forecasts, such as seasonal rainfall forecasts or riverine flood projections. And we are developing tools[30] to make it easier for humanitarians to use climate data.

11. 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.[31] 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.

A priority for 2023 will be a continued focus on climate impact data, which is essential for understanding and predicting the future impacts of hazards on vulnerable populations. We will also explore implementing data quality grades or scores for some datasets included in the Data Grids.

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

[29] https://data.humdata.org/organization/wfp?q=climate

[30] https://github.com/OCHA-DAP/ocha-anticipy

[31] 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 three datasets: UNHCR's data on forcibly displaced populations and stateless persons; OCHA FTS's requirements and funding data; and IATI activities data.[32]

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 IATI datasets are a good example of a complementary dataset as they add unique and useful information to the 3W sub-category. Another example is a dataset that provides data for a geographic area not covered by the existing dataset. If there is already a complete dataset available for the sub-category, the HDX team will reach out to in-country partners to determine which dataset is being used operationally and that dataset will be included in the Data Grid.

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:

[32] See Annex C of The State of Open Humanitarian Data 2022 for the information on the decision-making process for sub-national dataset waivers: **https://bit.ly/3YbKeDM**

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.

ANNEX C: CHANGES TO THE DATA GRIDS

Over the last year, we have continued to solicit feedback from OCHA field offices and our partners to ensure that the sub-category definitions and curation criteria reflect data needs in humanitarian operations. The Data Grid Governance Group discusses the feedback that we receive and makes decisions around changes to the Data Grids.

In 2022, the Data Grid Governance Group met three times and approved the following changes:

1) Adding a sub-category for climate impact

We expanded the Data Grids to include a new sub-category for climate impact which covers current and historical hazards and their impact on people, infrastructure and/or vegetation. Read more about this decision in our blog post from November 2022.[33]

2) Removing sub-categories for transportation status and affected areas

a. *Transportation status*: The difference between the transportation status and roads sub-categories was often unclear. In more than 35 percent of locations, the same dataset was used for both sub-categories.

b. *Affected areas*: The definition for the affected areas sub-category focused on the type and/or severity of geographical impacts. This was interpreted differently across locations, making it an inconsistent measure for data availability. When the sub-category was complete, it was a mix of data on the impact of climate events and the severity estimates included in a Humanitarian Needs Overview dataset.

[33] **https://centre.humdata.org/prioritizing-climate-impact-data/**