# Creating a Novel Global Dataset to Support UNICEF's Humanitarian Risk Analysis

**Final Report** 

## Workshop in Sustainable Development Practice 2021-22 Columbia SIPA

**Team Members:** 

Aawanti Singh

**Arielle Herman** 

**Carolina Tamayo** 

Jaideep Salil

Jeongmook Lim

Executive Summary	4
Introduction	4
The Missing Link	6
Background	7
Objectives	8
Methodology	8
Research Questions	8
Procedural Questions	8
Analytical Questions	9
Definitions	9
Assumptions	10
Project Phases	11
Phase 1: Initial Research	11
Phase 2: Data Collection	12
Phase 3: Initial Analysis and Presentation of Data	16
Phase 4: Prediction	16
Research Methods	16
Existing Databases' Collection Methodologies	17
Critical Assumptions and Risks	22
Assumptions	22
Risks	22
Outputs and Recommendations	23
Outputs	23
<b>Recommendations for Follow-Up</b>	26
Conclusion	29
References	30
Annexes	32

## Acronyms

ACLED	Armed Conflict Location and Event Data Project		
API	Application Programming Interface		
AWS	Amazon Web Services		
CCCs	Core Commitments to Children in Humanitarian Action		
COs	Country Offices		
EM-DAT	International Disasters Database		
EMOPS	Office of Emergency Programme		
FTS	Financial Tracking Services - OCHA Services		
GHO	Global Humanitarian Overview		
GDP	Gross Domestic Product		
HDI	Human Development Index		
IASC	The Inter-Agency Standing Committee		
IBAN	International Bank Account Number		
IBTRACS	International Best Track Archive for Climate Stewardship		
IDP	Internally Displaced Person		
IFRC	International Federation of Red Cross and Red Crescent Societies		
OECD	Organisation for Economic Co-operation and Development		
OCHA	Office for the Coordination of Humanitarian Affairs		
PTOR	Preliminary Terms of Reference		
PRIO	Peace Research Institute Oslo		
RAPS	Risk Analysis and Preparedness Section in EMOPS		
Sitrep	Situation Report		
SIPA	School of International and Public Affairs		
SDGs	Sustainable Development Goals		
UCDP	Uniform Collateral Data Portal		
UNDP	United Nations Development Programme		
UNICEF	United Nations Children's Fund		
UNHCR	United Nations High Commissioner for Refugees		
Uppsala	Uppsala Conflict Data Program <sup>1</sup>		
USGS	United States Geological Survey		

<sup>&</sup>lt;sup>1</sup> We distinguish Uppsala Conflict Data Program (UCDP) from Uniform Collateral Data Portal (UCDP) by referring to the former as "Uppsala" and the latter as "UCDP".

### **Executive Summary**

The Risk Analysis and Preparedness Section (RAPS) of the Office of Emergency Programme (EMOPS) of UNICEF enlisted a SIPA workshop team to create a composite database, documenting and quantifying the scope and impact of humanitarian events in order to inform risk analysis and emergency preparedness. The SIPA team committed to (1) research existing databases, appropriate variables and proxies to include in the database, and the current RAPS risk analysis process, (2) compile and collect a subset of the relevant data for a "proof-of-concept," outlining a method to continually update the database as new humanitarian events occur, and (3) perform preliminary analysis on the proof-of-concept. The following report results from a semester of research and consistent biweekly meetings with the client. This report may also be useful to the International Federation of Red Cross and Red Crescent Societies (IFRC), which is endeavoring to develop a similar database documenting the impact of natural disasters. Ultimately, the team hopes that this proof-of-concept database will be fully populated to document all humanitarian crises, starting in the year 2000.

#### Introduction

UNICEF has the main objective of promoting the rights of every child, everywhere. Working in over 190 countries, their mission is to deliver the essentials for every child to have an equitable chance in life: health care and immunizations, safe water and sanitation, nutrition, education, emergency relief and more<sup>2</sup>. UNICEF does through advocacy and by implementing programs as well as individual operations, specially focusing on an "equity strategy", which targets the most disadvantaged and excluded children and families. As explained by the client, a child's rights are violated to the degree that the child has an unequal chance in life, socially, politically,

<sup>&</sup>lt;sup>2</sup> UNICEF. "Our Mission," UNICEF USA, accessed December 21, 2021, https://www.unicefusa.org/mission.

economically, or civically. There is growing evidence that addressing inequity not only will give all children the opportunity to fulfill their potential but also will lead to sustained growth and stability of countries<sup>3</sup>.

Within UNICEF, the Office of Emergency Programmes (EMOPS) provides strategic and coordinated support to Country Offices preparing for and responding to humanitarian crises, in collaboration with Regional Offices, to ensure that they have the right capacities to respond effectively and deliver on UNICEF's Core Commitments to Children in Humanitarian Action (CCCs). EMOPS leads efforts to ensure that UNICEF's role in complex emergencies and natural disasters is clearly defined, that the organization is properly equipped to fulfill that role, and that all levels of the organization are prepared to deliver their mandate. Moreover, the Risk Analysis and Preparedness Section in EMOPS (RAPS) supports UNICEF Country Offices in being ready to meet the CCCs, and contribute to UNICEF's general ability to assess and adapt to a multitude of risks. RAPS' aim is to help UNICEF be best prepared and impactful in saving and protecting lives, and in reducing countries' and populations' fragilities and children's vulnerability around the world towards reaching the Sustainable Development Goals (SDG's).

Humanitarian crises are widespread around the world. As such, globally, although the total number of war deaths has been declining since 1946:

"Conflict and violence are currently on the rise, with many conflicts today waged between non-state actors such as political militias, criminal, and international terrorist groups. Unresolved regional tensions, a breakdown in the rule of law, absent or co-opted state institutions, illicit economic gain, and the scarcity of resources exacerbated by climate change, have become dominant drivers of conflict. In 2016, more countries experienced violent conflict than at any point in almost 30 years. At the same time, conflicts are becoming more fragmented...Furthermore, the regionalisation of conflict, which interlinks political, socio-economic and military issues across borders, has seen many conflicts become longer, more protracted, and less responsive to traditional forms of resolution."<sup>4</sup>

<sup>&</sup>lt;sup>3</sup> Preliminary Terms of Reference (PTOR)

<sup>&</sup>lt;sup>4</sup> United Nations, "A New Era of Conflict and Violence," United Nations (United Nations), accessed December 21, 2021, https://www.un.org/en/un75/new-era-conflict-and-violence.

Additionally, climate change and industrialization bring increasing numbers of environmental crises,<sup>5</sup> which have international implications for human security.

The increasing number and duration of crises call for strengthened preparedness and strategies to address them. Data is key for this, as it can be used to study risk factors and trends, as well as map the impact of past responses, and consequently inform present and future decision making in humanitarian agencies.

#### The Missing Link

Current data on risk preparedness and humanitarian emergencies is insufficient: existing databases are incomplete and unlinked, which poses challenges to the usefulness and completeness of the information they contain. For example, even UN agencies<sup>6</sup> largely use different databases (either managed by them or external stakeholders), instead of a central, unified database. Current databases being widely used include Desinventar, Peace Research Institute Oslo (PRIO), and International Disasters Database (EM-DAT), but they are decentralized. This is problematic because the details of humanitarian crises and interventions are dispersed and decisions based on them will potentially miss key insights.

Likewise, organizations with a humanitarian and social focus tend to collect information qualitatively, by writing reports that, more often than not, are only shared within the pertaining country office. Even when data is available, its collection is ineffective to identify baselines for future work because they are inconsistently documented across time and across different agencies.

All of the above supports the fact that there is an undeniable gap between how quantitative and qualitative information is collected and used. The project aims to help close said gap by

<sup>&</sup>lt;sup>5</sup> IPCC. "Climate Change Widespread, Rapid, and Intensifying," 2021, https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/.

<sup>&</sup>lt;sup>6</sup> The coordinating role here of OCHA, and to some extent of UNHCR, will be further examined.

combining various sources, helping analyze the needs of humanitarian crises based on certain explanatory variables. Likewise, the project produced a compilation of existing data-bases to further support these efforts.

The quantitative dataset proof-of-concept, once fully populated, will facilitate UNICEF's humanitarian risk analysis process so that the resources are allocated efficiently. It may similarly be useful to other organizations, such as the International Federation of Red Cross and Red Crescent Societies (IFRC), who endeavor to compose similar databases documenting the impact of natural disasters or other humanitarian crises. While the dataset at hand thus far includes 10 country case-studies, it provides a framework for future expansion, applications and collaborations.

#### Background

The project, which will build on the existing databases of humanitarian crises and their corresponding interventions, aims to develop the proof-of-concept for a predictive dataset to enable evidence-based decision-making and improve risk analysis and emergency preparedness. Currently, the process for identifying the resources needed to respond to a humanitarian crisis is often estimated on a qualitative basis. Additional quantitative data reflecting past crises might facilitate a more precise forecasting of future needs at UNICEF, as well as of potential resources. Pooling past crises together, our project is designed to provide baseline estimates that would help calculate important indicators like number of people planned for assistance in contingency plans. Other agencies have attempted to streamline this process but there has yet to be a centralized database that all agencies and various stakeholders can access. Given that UNICEF plays an integral role across a wide range of humanitarian work, a public database by UNICEF would be

extremely useful. We believe this would lead to a more efficient allocation of much needed resources for the people in need after humanitarian crises.

## **Objectives**

The project had the following objectives:

- 1. Research existing database structures and variables of interest.
- 2. Conduct a systematic review of past key humanitarian crises and corresponding databases.
- 3. Analyze EMOPS current risk analysis and emergency-preparedness process.
- 4. Create a comprehensive database framework that will inform UNICEF's future risk analysis and emergency-preparedness for humanitarian events.
- 5. Compile, analyze and prepare a proof-of-concept dataset for validation of the database framework and preliminary comparative analysis.

#### Methodology

#### **Research Questions**

In order to create a useful and comprehensive database of humanitarian events that can be effectively used to analyze risk and allocate resources appropriately, we developed two classes of research questions, procedural and analytical. The procedural questions 1 through 4 are addressed in the first phase of our research. As we limited the scope of the project to research question 5, the remaining analytical questions should be addressed in later stages once the database is fully compiled.

#### **Procedural Questions**

1. How does EMOPS currently respond to humanitarian events and how will the database inform their future responses?

- 2. Can the number of displaced persons accurately approximate humanitarian impact in terms of the number of people in need? If not, which indicators would be relevant?
- 3. Which countries have limited capacities to respond to humanitarian crises, as per United Nations Development Programme's (UNDP) Human Development Index?
- 4. How should protracted and overlapping humanitarian events be recorded in a database?
- 5. What types or aspects of humanitarian events are not being captured in current measures?
- 6. How can the database be efficiently updated?

#### **Analytical Questions**

- 7. How frequently do humanitarian events of varying sizes occur?
- 8. How can we analyze past humanitarian crises and their corresponding interventions in order to better predict and estimate the impact of future humanitarian crises?

#### **Definitions**

- Emergency: UNICEF defines an emergency as a "situation that threatens the lives and wellbeing of large numbers of a population and requires extraordinary action to ensure their survival, care and protection."<sup>7</sup> For our purposes, an emergency is the catalyst for a humanitarian event.
- 2. *Humanitarian event*: UNICEF defines a humanitarian crisis as "any circumstance where humanitarian needs are sufficiently large and complex to require significant external assistance and resources...This may include smaller-scale emergencies; in countries with limited capacities."<sup>8</sup> Notably, this will include man-made and natural events. In order to operationalize this definition, we will need to specify (1) methodology to measure the

<sup>&</sup>lt;sup>7</sup> "CCC | SCOPE," Humanitarian UNICEF, accessed December 21, 2021,

https://www.corecommitments.unicef.org/ccc-1-1.

<sup>&</sup>lt;sup>8</sup> Ibid.

discrete length of an event (see above research questions), and (2) threshold of "sufficiently large" impact, that varies with a nation's resources at the time of the recorded event.

- 3. People in need: In the humanitarian space, people in need is defined as the "gap or discrepancy between the status quo and a different desired state"<sup>9</sup>, according to the guidelines developed by the Inter-Agency Standing Committee (IASC), in the Humanitarian Profile Common Operational Dataset. "Need" is a time sensitive concept, in that it changes over the course of a conflict, with requirements at the immediate onset of a conflict varying/evolving over different phases of an event. Given these characteristics, people in need is considered to be a concept "without widely agreed boundaries"<sup>10</sup>. It is however a subset of the population affected by a crisis.
- 4. *Humanitarian impact*: We approximate overall humanitarian impact based on the number of people in need and the number of displaced persons.

### Assumptions

The key process in this project mostly concerns compiling previously collected data. While we thoroughly researched the data collection methodology of each of the sources, it remains difficult to deconstruct hidden assumptions in the previous data collection, such as potential biases against specific demographics or implicit prioritizations for which measures to record. Further, these implicit assumptions may vary across data collection agencies. For example, different country offices may have different assumptions or biases in recording information that may render the resulting information incomparable across countries. Some resources (such as United Nations High Commissioner for Refugees or UNHCR) additionally compile data themselves from

<sup>&</sup>lt;sup>9</sup> Humanitarian Profile Support. April 2016. IASC Information Management Working Group.

https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/humanitarianpro filesupportguidance\_final\_may2016.pdf

<sup>&</sup>lt;sup>10</sup> Ibid

disparate resources (e.g. governments or country offices), further layering obstructions in understanding assumptions of the data. While it was not possible to examine all relevant situation reports, these reports similarly suffer from a lack of uniformity. Therefore, we supplement our analysis with thorough research of the data collection methodology from each of our sources.

#### **Project Phases**

#### **Phase 1: Initial Research**

For this first phase of our project, we (1) analyzed the EMOPS RAPS' risk analysis procedure, (2) researched the events and structure of existing humanitarian crises databases, and (2) selected countries to research that would facilitate the development of a robust and diverse data structure. In this appraisal, we gleaned relevant indicators to utilize in our proposed database structure that can be used to predict future humanitarian interventions.

- 1. *Risk Analysis Process Review:* We read through the materials provided by EMOPS in order to understand their process, how the database would enhance it, and how the database should be structured in order to best permit incorporation into the RAPS' emergency-preparedness process. We further conducted independent research in order to understand how other agencies analyze risk.
- 2. Existing Databases: In this appraisal, we researched the relevant indicators that best inform preparation for future humanitarian interventions. We additionally determined the necessary categories of indicators to best capture assessment of need during humanitarian crises, which are the following: 1) General Identifiers, 2) Country coping capacity, 3) Population Impact, and 4) Intervention details. We collected data in each of the above-listed categories from the following sources: 1) IBAN, Uppsala, UNHCR, World Bank, UNDP, OCHA.

The challenges we faced in this phase concerned the definitions of our data. These included: (1) inability to disaggregate existing annual indicators by a country's humanitarian crisis (e.g. number dead resulting from a specific event), (2) understanding missing information, and (3) agreeing on a definition of prioritization of which events to include. The ways in which we measure the humanitarian impact was extensively discussed internally, the details of which will be shared later in the report.

#### **Phase 2: Data Collection**

In this phase, we developed a comprehensive database structure that will eventually accommodate the necessary information to analyze risk and inform preparation for humanitarian crises. We examined numerous humanitarian crisis databases, and selected the most relevant variables for UNICEF RAPS strategic planning, in consultation with UNICEF.

Inspired by the "Inform Global Risk Index 2020", which is a public database compiling global humanitarian events and risk-assessments, we chose four broad categories of variables: 1) General identifiers, 2) Immediate impact 3) Coping capacity, and 4) Intervention details.

- 1. **General Identifiers** include information on the country, year, and types of conflicts occuring in that year.
- 2. Immediate Impact reflects variables that approximate the total impact of humanitarian crises in a given year. Importantly, we can not causally attribute this impact to any specific humanitarian crisis because these variables are not disaggregated by event (only by year). Instead, we indicate the annual costs of the conflict-driven humanitarian crises, including the number of people who died due to conflicts, who were internally displaced, and who were externally displaced.

- 3. **Coping Capacity** describes a country's ability to respond to a humanitarian crisis. The indicators chosen illustrate the quality of the existing political, socio-economic, health infrastructure to deal with humanitarian crises of a given country. For instance, HDI, GDP, and child mortality rate were included in this category. This category permits us to evaluate the relative need of two countries experiencing similarly severe crises, especially if one country has more extensive infrastructure or greater access to resources than the other. In this category, we record economic and development indicators.
- **4. Intervention Details** capture the level of international financial response to humanitarian crises. This required looking at the amount of donations that were requested to meet the needs of the people on the ground, as well as the amount that was actually donated.

As we collected data for each of these categories, we continued to identify challenges and adapt the database structure to accommodate information. For example, we initially proposed a database that would list humanitarian crises and associated indicators. However, upon discovering few, if any, consistently disaggregated measures, we transformed the database to describe the state of each country in a given year.

We include variables that approximate the total impact of humanitarian crises in a given country in a given year, and do not attempt to disaggregate these statistics. This iterative process yielded the following variables for each category:

## Table 1. Data Dictionary

Category	Variable	Description	Source
General Identifiers	ID	Numeric value describing the order in which the country was entered into the database	created
	Code	Three letter alpha code	<u>IBAN</u>
	Year	The year of a given country being documented	created
	Country	Full name of country	created
	Major Event	Major event impacting a given country in a given year, as identified by individual research	created
	Tags	Tags describing the type of major events impacting a country in a given year	<u>created from</u> <u>Uppsala</u>
	Context	List of the major events impacting a country in a given year	<u>created from</u> <u>Uppsala</u>
Population	Deaths	Number of deaths yearly during the course of the event	<u>Uppsala</u>
Impact	Internally Displaced	Number of internally displaced persons yearly during the course of the event	<u>UNHCR</u>
	Externally Displaced	Named list of the 5 countries that received the most refugees, e.g. CountryCode: Number Displaced	UNHCR
	Total Externally Displaced	Number of total externally displaced persons yearly during the course of the event	<u>UNHRC</u> (example for <u>Afgh)</u>
	Total Population	Total population in a country at a given year	World Bank
	People in Need	People affected by armed conflict conflict	<u>Global</u> <u>Humanitarian</u> <u>Overview</u>
Coping Capacity	Child Mortality rate	Child mortality or the under-five mortality rate refers to the probability of a child dying between birth and exactly 5 years of age, expressed per 1,000 live birth. (annually)	<u>UNICEF</u>
	GDP	Value of production of goods and services in a country (annually)	World Bank
	HDI	A composite metric which tracks life expectancy, education development and per capita income of a country	<u>UNDP</u>
Intervention	Request	Response Plan Requested (USD million)	<u>OCHA</u>
	Funded	Response Plan Funded (million)	<u>OCHA</u>
	Target Met	Percent of target met (Funded / Request)	created

In this process, we limited the scope of our project to 10 countries between the years of 2000 and 2020. In order to develop the most robust data structure possible, we selected a diverse range of countries that experienced humanitarian crises with a varied typology. We explored the following categories: 1) protracted crises, 2) electoral violence / Coup D'etat, 3) intrastate wars, 4) interstate wars, and 5) large-scale crises.

## Table 2. Selected Countries

Humanitarian Crises	Research Qs	Challenge	Countries
Protracted Crises	2	Other indicators may be necessary to identify the population impact.	<ol> <li>Afghanistan</li> <li>Colombia</li> </ol>
	4	Temporal spillover: protracted timelines obscure baseline impact.	
Coup D'etat / Electoral Violence	2	The impacts of events that span less than a year may be confounded with other events.	<ol> <li>Mali</li> <li>Kenya</li> </ol>
	3	The baseline information years before and after the event are relevant in understanding impact and coping capacity.	
Intrastate war	2	Other indicators may be necessary to identify the population impact.	<ol> <li>Myanmar</li> <li>South Sudan</li> </ol>
	4	Spatial spillover: coping capacity of responding countries (e.g. Bangladesh) may be relevant in comparing crises.	
Interstate war	2	Record externally displaced individuals across borders	<ol> <li>7) Ethiopia</li> <li>8) Eritrea</li> </ol>
	4	Spatial spillover: dynamic impact between the two countries.	
Large Scale Crises	3	Protracted crises may demonstrate relationships with coping capacity indicators.	9) Yemen 10) Iraq
		Represent the Middle East.	

#### **Phase 3: Initial Analysis and Presentation of Data**

In the preliminary analysis phase, we (1) determined the validity of our collected data in comparing impacts of events and (2) performed basic preliminary analysis.

- 1. *Validity of measures*: We confirmed our hypothesis concerning the validity of the chosen indicators in the final dataset. We further measured the consistency of missing data across entries to determine if sufficient indicators exist to enable comparison of events.
- 2. *Preliminary Analysis*: We conducted basic analysis and visualizations. This included descriptive statistics, such as frequency and distribution of the impact variable, and a timeline visualization of the events.

#### **Phase 4: Prediction**

This phase was planned only if time would permit. Our hope was to specify the appropriate methodology after understanding the structure of the composite database and further research performed in our initial phase. Yet, due to time constraints and higher prioritization of compiling and preparing our database for future usage, we did not reach this stage of phase 4 at the end of our workshop time period.

#### **Research Methods**

The sources of our data came from pre-existing databases and other reports that identified the size of humanitarian crises, and if possible, the resources that were needed. Hence, we researched the World Bank, International Disasters Database (EM-DAT), Desinventar, Global Food Observatory (GFO), World Food Programme (WFP), Flood Observatory, Uppsala Conflict Data Program (Uppsala), The Armed Conflict Location and Event Data Project (ACLED), Uniform Collateral Data Portal (UCDP), United States Geological Survey (USGS), International Best Track Archive for Climate Stewardship (IBTRACS), WFP and Situation Reports from UNICEF, OCHA and

other UN agencies that work in the humanitarian field. Some databases had critical information that other databases did not include. Therefore, our approach relied on a multifaceted data collection technique, which ensured that our project's deliverables were reliable and comprehensive.

#### Existing Databases' Collection Methodologies

#### Uppsala Conflict Data Program: Deaths, Context

Total deaths per year were collected from Uppsala Conflict Data Program, Department of Peace and Conflict Research. This database had an excellent and relatively complete reservoir of data illustrating the intensity of man-made conflicts per country from 1989 to 2020. For a given year and country, the database had the recorded number of deaths due to a conflict disaggregated into the types of conflict such as "state-based violence", "non-state violence". It must be noted that although these numbers of deaths are not always consistently uniform across major databases to the decimal, we chose Uppsala as our common source for the variable, "Deaths", at the recommendation of our client and for the sake of uniformity across years and countries in our database.

Context is drawn from Uppsala, but discretion of what is entered into the database is up to the researcher feeding the data.

#### UNHCR: Internally / Externally Displaced Persons

In order to access the data frames utilized for this variable, we searched the country name and "Data on forcibly displaced populations and stateless persons" at <u>https://data.humdata.org/</u>. After following the link, we selected the spreadsheet entitled "End-year Stock Population Figures for Forcibly Displaced Persons originating from <country>" (not to be confused with the spreadsheet of almost the same name except on persons residing in the given country). UNHCR data structure,<sup>11</sup> definition,<sup>12</sup> and methodologies<sup>13</sup> are all available on the website. UNHCR gathers data from a wide variety of sources, including country offices, governments, NGOs, and "unspecified combinations of the aforementioned,"<sup>14</sup> yielding a high degree of uncertainty in their collected data. Notably, governments tend to report narrower figures compared to UNHCR estimates.<sup>15</sup> In countries that lack official registers, UNHCR estimates the refugee population.<sup>16</sup> Reported figures are generally rounded based on magnitude to reflect varied degrees of certainty (e.g. figures below 1000 are generally rounded to the nearest 10, whereas figures between 100,000 and 1 million are rounded to the nearest 10,000).<sup>17</sup> While data disaggregated by age and gender are starting to become available,<sup>18</sup> this information was not yet reliably available in the majority of the years of our research. This may be useful to include in the future when fully populating the proposed database.

While UNHCR is a prime example of methods for consistently updating large datasets with huge numbers of sources, they lack a "comprehensive handbook on reporting standards and best practices aimed for data collection and coordination."<sup>19</sup> This became clear when handling the data, as many figures demonstrate varied numbers of significant digits. For example, some 7-digit

<sup>&</sup>lt;sup>11</sup>UNHCR, "Data Content," UNHCR, accessed May 11, 2022, <u>https://www.unhcr.org/refugee-statistics/methodology/data-content/</u>.

<sup>&</sup>lt;sup>12</sup>UNHCR, "Methodology," UNHCR, accessed May 11, 2022, <u>https://www.unhcr.org/refugee-statistics/methodology/</u>.

<sup>&</sup>lt;sup>13</sup>"Chapter 1: Sources, Methods and Data Considerations," UNHCR Statistical Yearbook, 2005, <u>https://www.unhcr.org/464049e32.pdf</u>.

 <sup>&</sup>lt;sup>14</sup>Lily Chen, Petra Nahmis, and Sebastian Steinmueller, "UNHCR Statistical Reporting on Statelessness," UNHCR, 2019, <u>https://www.unhcr.org/statistics/unhcrstats/5d9e182e7/unhcr-statistical-reporting-statelessness.html</u>.
 <sup>15</sup>Ibid.

<sup>&</sup>lt;sup>16</sup>Ibid.

<sup>&</sup>lt;sup>17</sup>UNHCR, "Chapter 1: Sources, Methods and Data Considerations," UNHCR Statistical Yearbook, 2005, <u>https://www.unhcr.org/464049e32.pdf</u>.

<sup>&</sup>lt;sup>18</sup>Ibid.

<sup>&</sup>lt;sup>19</sup>Lily Chen, Petra Nahmis, and Sebastian Steinmueller, "UNHCR Statistical Reporting on Statelessness," UNHCR, 2019, <u>https://www.unhcr.org/statistics/unhcrstats/5d9e182e7/unhcr-statistical-reporting-statelessness.html</u>.

figures were recorded to the last digit (e.g. 1,244,072 internally displaced in Colombia 2003), while others were rounded (e.g. 3,000,000 internally displaced in Colombia in 2006, 2007, and 2008). This clearly demonstrates that figures are reported under varying degrees of certainty, however, there are no values in the country displacement datasets that are reported as "unavailable." This is likely related to the common practice of assuming that unreported values (countries or populations not recorded in a given database) are zero<sup>20</sup> rather than simply "unavailable." While some attempt to correct for unreported (or unavailable data) by imputing unreported values according to trendlines,<sup>21</sup> in our proof-of-concept database, we transfer the UNHCR data exactly as originally reported.

**Bank:** World Total Population, Child Mortality GDP. Rate. HDI We gathered the total population, under five mortality rate, GDP, and HDI from the World Bank. Total population was disaggregated by country per year. According to the World Bank, this data was compiled from various sources such as "census reports and other statistical publications from national statistical offices" and "United Nations Statistical Division".<sup>22</sup> GDP was aggregated from sources such as World Bank national accounts and OECD National Accounts.<sup>23</sup> Under-five mortality rate is an indicator expressing the number of children who die by the age of 5 years, per 1000 live births. It serves as a baseline indicator for how a country is coping on issues

relating to children's rights.<sup>24</sup>

<sup>&</sup>lt;sup>20</sup>Moritz Marbach, "On Imputing UNHCR Data," *Research & Politics* 5, no. 4 (October 2018): 205316801880323, https://doi.org/10.1177/2053168018803239.

<sup>&</sup>lt;sup>21</sup>Ibid.

<sup>&</sup>lt;sup>22</sup>World Bank, "Population, Total | Data," data.worldbank.org, accessed May 11, 2022, <u>https://data.worldbank.org/indicator/SP.POP.TOTL</u>.

<sup>&</sup>lt;sup>23</sup>World Bank, "GDP (Current US\$) | Data," data.worldbank.org, accessed May 11, 2022, <u>https://data.worldbank.org/indicator/NY.GDP.MKTP.CD</u>.

<sup>&</sup>lt;sup>24</sup>World Bank, "Under-Five Mortality Rate," data.worldbank.org, accessed May 11, 2022, <u>https://www.who.int/data/nutrition/nlis/info/under-five-mortality-rate</u>.

The human development index is a proxy for measuring a country's level of achievement in some key aspects of human development. This index was initially created to ensure that important aspects of human development (i.e. longevity and education) were not masked by merely income per capita. More precisely, the HDI values are calculated by taking the geometric mean of indices in three dimensions: health, education, and standard of living. The health index is proxied by the life expectancy at birth. The education index is calculated by considering the average number of years of schooling completed for adults aged 25 years and more and the expected years of schooling for children starting school. The standard of living index is taken from the gross national income per capita. Although there is an HDI ranking for all countries every year, we decided to provide the raw score of HDI to more accurately reflect a country's level of human development irrespective of its standing in the world. We believe that a trend in a country's HDI values over the years can be a helpful first clue to a country's relative response capacity during a humanitarian crisis.<sup>25</sup>

#### Financial Tracking Services (FTS) of UN OCHA: Requested, Funded

These two variables are a snapshot of the requested and funded quantities for humanitarian response plans come from the Financial Tracking Services (FTS) of UN OCHA. FTS makes public all reports of humanitarian funding to affected countries, including those of internationally coordinated response plans, bilateral fundings to the governments, as well as financial contributions that were not part of the coordinated response plans. These two variables, therefore, refer to the total USD amount that was requested or funded by all types of humanitarian funding calls. The USD is not adjusted for inflation and is thus at nominal prices of that given year.<sup>26</sup> When

<sup>&</sup>lt;sup>25</sup>UNDP, "Human Development Index (HDI)," Human Development Reports, accessed May 11, 2022, <u>https://hdr.undp.org/en/content/human-development-index-hdi</u>.

<sup>&</sup>lt;sup>26</sup>UNOCHA, Financial Tracking Service, "Home," accessed May 11, 2022, <u>https://fts.unocha.org/</u>.

the database is fully populated, it may be helpful to provide an inflation-adjusted measure, in order to permit more comprehensive comparisons throughout time.

#### Inter-Agency Standing Committee: People in Need

"People in need" "refers to a gap or discrepancy between the status quo and a different desired state," according to the guidelines developed by the Inter-Agency Standing Committee (IASC), in the Humanitarian Profile Common Operational Dataset. "Need" is a time sensitive concept, in that it changes over the course of a conflict, with requirements at the immediate onset of a conflict varying/evolving over different phases of an event. Given these characteristics, people in need is considered to be a concept "without widely agreed boundaries". It is however a subset of the population affected by a crisis. Working off the IASC's Humanitarian Profile Common Operational Dataset, People in Need is understood to be a subset of the population affected in a crisis. The figures are created by summing the number of internally displaced, refugees, asylum seekers, and others of concern. It does not include casualties, such as the number injured. However, the literature does acknowledge that there is ambiguity on the methods and best practices employed to estimate the number of people in need.

The data for people in need was accessed from the Global Humanitarian Overview (GHO), which is an annual report published by OCHA. It was started in the year 2015, so data is scare for years prior to 2015. Colombia is the only country for which people in need data was drawn from an individual country humanitarian needs overview report, as it is not covered in the GHO.

## **Critical Assumptions and Risks**

This section covers various risks and assumptions which provide nuances which impacted the implementation of the project.

## **Assumptions**

Subject Area	Assumption
Output Structure	Internal agreement about the data structure/format of the data output.
Data Availability	Data across various platforms to be available publicly or could be shared with the SIPA team at all stages.
Distinction between events	Clear secondary research/academic theory available for categorization/delineation of man-made humanitarian crises.
Academically rigorous predictions	Since the data sources are varied and lack uniformity, the prediction models would not stand the test of academic rigor.

## <u>Risks</u>

Phase 1: Initial Research			
Subject Area	Probability	Impact	Mitigation Strategy
<b>Impact Indicators</b> : While displacement indicators could be standardized across various events, other indicators of impact would vary across events. For example, public services would be disrupted during a humanitarian crisis borne of natural causes, however violence would also have an additional impact in the case of a man-made humanitarian crises.	High	High	Frequent communication with the UNICEF EMOPS Risk Analysis team for a dynamic understanding of all types of events.
<b>Timeline of events:</b> Data is required from the early 2000s for events. The data reporting on events would differ across this	High	High	Critical analysis of past data repositories (UNICEF, OCHA and others) to understand past procedures in handling timelines.

time period which would make uniform reporting a hurdle.				
Phase 2: Data Collection	Phase 2: Data Collection			
Subject Area	Probability	Impact	Mitigation Strategy	
<b>Country Reports:</b> While UNICEF is active in all developing countries the format of reporting in all countries might not be the same.	High	Low	Internal team coordination about handling of various components of the reports.	
<b>Country Coverage:</b> As UNICEF is active in over 150 countries, data collection for all the countries over a 20 year period would be unwieldy and be out of reach within the time frame of the capstone of the project period.	High	Medium	Consultation with the UNICEF team to narrow down the list of countries to cover.	
<b>Coordination between Situation Reports</b> <b>and other Data Sources:</b> The reporting format in Situation Reports would not be similar with the other data sources	High	Low	Following Sitrep reporting guidelines/formats to extract information from other data sources.	
Phase 3: Analysis and Presentation of Data				
Validity of measure indicators: the consistency in the availability of information contributing to indicators can impact their validity.	High	High	Identifying patterns in available information for analysis.	
Phase 4: Prediction Model				
<b>Incomparable data:</b> the diversity of data sources and differences in information availability may result in incomparable data in some cases.	Medium	High	Increasing available data with Sit reps and additional data sources beyond existing databases.	

## **Outputs and Recommendations**

## **Outputs**

The database and other visualizations are some of the key outputs of this project. The possible

outcomes of these products are analyzed below. We have the following outputs: 1) Database, 2)

Database Dictionary, 3) Visualization Dashboard, 4) Pre-existing Database Review, and 5) a datareshaping script.

- 1. Database: It provides a tool for humanitarian organizations to track conflicts. The three different categories (i) General Identifiers (ii) Population Impact (iii) Coping Capacity and (iv) Intervention capture different aspects of a conflict, all of which provide data that are necessary for humanitarian agencies to plan for future events. The General Identifiers category is self-explanatory as it gives the base details of the country and context. The Event Details gives an overview of the Main Event and other indicators like number of both internal and externally displaced, which is especially relevant for humanitarian agencies after a crisis. The Coping Capacity category provides an overview of the country's present status with regard to certain development indicators like Child Mortality, HDI and GDP. Lastly, the Intervention category gives some information about the past funding requirement vs. fund received for conflict-affected countries. These categories along with the indicators have been carefully designed to enable humanitarian agencies to effectively plan for countries which are affected by conflicts or are unstable and veering towards one. This dashboard aggregates data from other sources and it has been especially designed for personnel in the humanitarian space to understand and plan for emergencies arising out of conflicts.
- Database Dictionary: The data dictionary is a breakdown of each of the indicators in our database. Brief descriptions on what the indicator captures, the source from where it was accessed, and general comments are noted here.

3. **Visualization Dashboard**: The <u>UNICEF Humanitarian Risk Analysis and Preparedness</u> <u>Dashboard</u><sup>27</sup> allows organizations in the humanitarian space to assess the magnitude and track the evolution of conflict. It is intended to serve as a focal point, the basis upon which strategic planning and response monitoring can be anchored. Currently, it visualizes data for three indicators: people in need, internal displacement and external displacement. The map provides a snapshot of key information like people in need, displacement, conflict related death and HDI in a text box, which is displayed when the user hovers over a country of interest.

The internally displaced racing bar chart visualizes the conflicts of greatest concern in the world, ranked according to highest values for displacement over a period of twenty years. This allows humanitarian actors to track trends in conflicts around the world. Similarly, the externally displaced line chart allows for tracking of the refugee situation for countries in conflict.

The dashboard is also useful as an exploratory tool. For example, analysis of the visualizations revealed the discrepancy between the relatively stable external displacement situation for Yemen, and the worsening internal displacement and people in need values over the same period. It serves as a launch board to pursue interesting lines of inquiry like: Are people fleeing conflict in Yemen facing border blockades that hinder movement, which in turn drives down external displacement values?

4. **Pre-existing Database Review:** A significant portion of the team's research time was spent on reviewing other databases, which led to a mapping exercise of the current databases which are available in the humanitarian space. These include Dedinventar, Peace

<sup>&</sup>lt;sup>27</sup> Screenshot of interactive dashboard in Annexes.

Research Institute Oslo (PRIO), EM-DAT, Uppsala and others. A database review document was created which maps the key indicators captured in existing databases along with the regions covered and the unit of measurement. This enables humanitarian agencies to navigate the often complicated data landscape which exists when it comes to assessing natural disasters and man made conflict. While there is considerable data which exists on both these types of events, it is often located across different sources and it is hard to comprehend for policy personnel who are not researchers. This mapping exercise will facilitate personnel from all types of agencies to navigate and utilize the data landscape more effectively.

5. **Data Reshaping Script**: This script is provided in R and reshapes the provided data frame to be in a long format that is more compatible with Tableau.

## **Recommendations for Follow-Up**

- 1. **Expert Interviews**: We reached out to UNICEF data specialists in order to conduct expert interviews and thus expand our research of relevant indicators and variables in the field of risk analysis and prediction. However, we were unable to complete this process due to scheduling and time constraints. This could be pursued.
- 2. Literature Review: We surveyed relevant literature in order to determine which variables and indicators were necessary to include in the database, especially for prediction purposes. We further attempted to identify (1) appropriate proxies for humanitarian impact, (2) appropriate measures to use for prediction of humanitarian events, and (3) appropriate handling of missing data. While the outcome of this review is reflected in the design of the global database, experience gained with its use might justify a further assessment of relevant indicators.

- 3. Validation of the Selected Indicators: A method to periodically update and validate the data of chosen indicators still needs to be developed.
- 4. Updating the Database: Using Amazon Web Services (AWS) in combination with website Application Programming Interfaces (API) is a potential solution to keep the database updated. An API is the medium through which one can access through a coding language the same information that is available on the website. This allows one to entirely script the interaction, leaving no room for human error. While not all websites have APIs, the databases we access for this project do. Furthermore, AWS would allow the script to be applied automatically given a certain catalyzing event, such as a date (e.g. January 1st) or the addition of new information to a given database. The combination of these two resources would allow automatic updating of portions of the database, except for the context column, which must be updated manually. Coding packages have already been written for many of these databases that facilitate access to the APIs,<sup>28</sup> such that the scripts need not be written entirely from scratch. One can access the database's API in the table 3 below.

Database	API Link	Packages and Resources
<u>Uppsala</u>	UCDP API	R: <u>UCDPtools</u> (archived package, coding structure useful)
<u>UNHCR</u>	<u>UNHCR APIs</u> <u>RSQ API</u> <u>UNHCR Popstats Data</u>	R: <u>UnhcrDataPackage</u> Github Account: <u>https://github.com/UppsalaConflictDataProgram</u>
World Bank	The World Bank API	R: <u>wbstats</u> R: <u>WDI</u>

Table 3. API and Package Source Links

<sup>&</sup>lt;sup>28</sup> Felix Haass, "List of useful R packages for quantitative conflict research," *Bretterblog* (blog), September 3, 2014, <u>https://bretterblog.wordpress.com/2014/09/03/list-of-useful-r-packages-for-quantitative-conflict-research-4/</u>.

		Python: <u>WBGAPI</u> Python: <u>world-bank-data</u>
<u>UNICEF</u>	SDMX API	Github Account: <u>SDMX</u>
<u>OCHA</u>	FTS public API	R: <u>fts</u> (archived package, coding structure useful) Python: <u>fts3-rest-API</u>

- 5. **Published Data Collection Methodologies**: Refining the data collection method employed in developing this database, to be more consistent and methodic, would improve replicability. Adopting measures utilized by data collection methodologies of existing organizations will be helpful.
- 6. **Build Predictive Capacity Based on This Data (QMSS)**: The ultimate goal of developing this tool is to establish broad ideas about what percentage of populations are affected in a conflict, with differing frequencies in different types of contexts. With this information, baseline probabilities can be developed to define the number of people affected and plan appropriate responses. To make progress towards this goal, we suggest partnering up with consultants with coding skills, for example, students at QMSS. Elena Krumova heads their consultancy program that runs through the fall, spring, and summer semesters.
- 7. **Partner with Research Institutions**: Partnering with institutions like Uppsala and ACLED, as well as other organizations working in the in humanitarian conflict data collection space, will allow for the refinement of this proof-of-concept, especially when it comes to developing a context typology to complement the existing conflict typology.

## Conclusion

The main deliverable of this project is a proof of concept database that consolidates information from multiple existing sources, such as UNICEF situation reports, Humanitarian Data Exchange (a platform by OCHA) EM-DAT etc. It has been specifically designed to enable analysis along a long time span and in a varied typology of humanitarian crises. We have additionally identified intermediary deliverables that will facilitate the compilation of the data, such as a list of indicators for comparison of humanitarian emergency events. In this final report we have provided a detailed description of the methodology we developed for the data compilation. In addition, within the temporal and material constraints of our project, we have made some suggestions on how to carry the project forward. It is the team's hope that the results of this project will contribute to the quality of UNICEF's work in the humanitarian arena: for the sake of all the children under its care.

#### References

Chen, Lily, Petra Nahmis, and Sebastian Steinmueller. "UNHCR Statistical Reporting on Statelessness." UNHCR, 2019.

https://www.unhcr.org/statistics/unhcrstats/5d9e182e7/unhcr-statistical-reporting-statelessness.html.

- Haass, Felix. "List of useful R packages for quantitative conflict research." *Bretterblog* (blog), September 3, 2014. <u>https://bretterblog.wordpress.com/2014/09/03/list-of-useful-</u>r-packages-for-quantitative-conflict-research-4/.
- IASC Information Management Working Group. "Humanitarian Profile Support," April 2016.

https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/doc uments/files/humanitarianprofilesupportguidance\_final\_may2016.pdf.

- IPCC. "Climate Change Widespread, Rapid, and Intensifying IPCC IPCC," 2021. https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/.
- Marbach, Moritz. "On Imputing UNHCR Data." *Research & Politics* 5, no. 4 (October 2018): 205316801880323. <u>https://doi.org/10.1177/2053168018803239</u>.
- UNDP Human Development Report Office. "Human Development Index (HDI)." Accessed May 11, 2022. <u>https://hdr.undp.org/en/content/human-development-index-hdi</u>.
- UNHCR. "Chapter 1: Sources, Methods and Data Considerations." UNHCR Statistical Yearbook, 2005. <u>https://www.unhcr.org/464049e32.pdf</u>.
- \_\_\_\_\_. "Data Content." UNHCR. Accessed May 11, 2022.

https://www.unhcr.org/refugee-statistics/methodology/data-content/.

\_\_\_\_\_. "Methodology." Accessed May 11, 2022. <u>https://www.unhcr.org/refugee-</u> <u>statistics/methodology/</u>.

UNICEF. "CCC | SCOPE." Accessed December 21, 2021.

https://www.corecommitments.unicef.org/ccc-1-1.

- . "Our Mission." UNICEF USA. Accessed December 21, 2021. https://www.unicefusa.org/mission.
- UNOCHA Financial Tracking Service. "Home." Accessed May 11, 2022. https://fts.unocha.org/.
- United Nations. "A New Era of Conflict and Violence." United Nations. United Nations. Accessed May 11, 2022. <u>https://www.un.org/en/un75/new-era-conflict-and-violence</u>.

World Bank. "GDP (Current US\$) | Data." Accessed May 11, 2022. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD. \_. "Population, Total | Data." Accessed May 11, 2022.

https://data.worldbank.org/indicator/SP.POP.TOTL.

https://www.who.int/data/nutrition/nlis/info/under-five-mortality-rate.

## Annexes

## Annex 1. Original Preliminary Terms of Reference

The original description of the project.

## Annex 2. Database (Googlesheet, RDS data)

The database, entitled Database20220513, is provided in the following two formats: Googlesheet, RDS data.

## Annex 3. Database Dictionary

A spreadsheet describing the terms and structure of the database.

## Annex 4. R Reshaping Script

The script, entitled manipulate\_data.R, to reshape the database into a format more suitable for Tableau. It performs two transformations: 1) transforms the external displacement column into long format, and 2) transforms the tags into wide format. Finally, it uploads the transformed data to the existing googlesheet.

### Annex 5. Typology of Existing Databases

Spreadsheet listing out the unique indicators of existing humanitarian crisis databases.

### Annex 6. Tableau Dashboard Visualization

Accessible via the following <u>link</u>.





