Exploring Web-Available Data for Macro-indicators of Humanitarian Intervention in the aftermath of Disasters

A thesis submitted for the degree of
Doctor of Philosophy (PhD)

Asmat Monaghan

ROYAL HOLLOWAY UNIVERSITY OF LONDON

School of Management
Royal Holloway, University of London

February 2018

ABSTRACT

In 2015 US$28bn was spent in international humanitarian assistance, of this US$10.8bn was raised through United Nations coordinated appeals; 45% short of the US$19.8bn needed by the United Nations. As pressure increases to do more with less, measuring the effectiveness of past humanitarian aid to inform the response to future disasters becomes an imperative. In the humanitarian domain, however, the difficulties and obstacles to obtaining authoritative empirical data to evaluate the effect of humanitarian intervention at the detail level are significant.

In response, this study explores the ability of Web-available curated data in providing macro-level indicators of humanitarian intervention in the aftermath of disasters. In so doing it identifies three macro-level indicators that suggest *mean disaster survival rates* when plotted by *year*, by *humanitarian aid per person* and by *population* growth may signpost the effectiveness of humanitarian intervention.

En-route to finding these macro-indicators, the study clarifies the need for domain-specific key data artefacts, referred to here as *data scaffold*s, to support viable data analysis; develops a Data Veracity framework (DVf) as a toolset to equitably and consistently evaluate the veracity of sourced data; defines a prototype Master Disaster Classification (MDC) model; and creates a baseline amalgamated Master Set of Global Disasters (MSGD). Finally, over and above the knowledge contribution of these created artefacts and the design theory that they support, this work provides a foundation for future research in the humanitarian and data science domain.

Dedications

I dedicate this work to my parents. To my adored late father, who inspired and encouraged me to be ever curious as he believed learning should stop only with your last breath; and to my mother whose incredible love, understanding, patience and support have protected and comforted me all my life. I also want to dedicate this to my husband who has spent days, months and years hovering around me patiently as I worked, making sure I did not need or want for anything. I do not have the words to express my gratitude for the care, support and love he has shown me.

I want my children to know how grateful I am for them and dedicate this to my beautiful and wonderful daughters, their amazing husbands and to my grandchildren, who are my angels and give my life meaning. Finally, I want to thank my incredible parents-in-law for being so unbelievably caring and understanding; even sacrificing family visits, and the opportunity to see their grandchildren and great-grandchildren to allow me the space and time to study.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my supervisor, Professor Mark Lycett, for his invaluable support and guidance without which I would never have completed this work. His knowledge and insights enabled me to see opportunities and possibilities that I would have otherwise missed. Most precious of all is that he helped me ground my thinking when I needed focus and inspired me when I needed motivation. I will be eternally grateful to him for his kindness, empathy and the fact that I was fortunate enough to have him as my supervisor.

DECLARATION

Declaration

I, Asmat Monaghan, hereby declare that this thesis, and the research within it, is my own work carried out under the supervision of Professor Mark Lycett.

This research has also led to the following papers:

> Lycett, M., and Monaghan, A. and 2013, October. Big data and humanitarian supply networks: Can Big Data give voice to the voiceless?. In *Global Humanitarian Technology Conference (GHTC), 2013 IEEE* (pp. 432-437). IEEE.
>
> Lycett, M., and Monaghan, A. (2018). Taming Data Veracity: Developing a Framework for Disaster Data. Submitted to Information Systems Research.

Asmat Monaghan (February 2018)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: INTRODUCTION

1.1 Research Background & Motivation

Over the last decade the world has experienced disasters that have caused the deaths of more than 700 thousand people, injured over 1.4 million people and made homeless around 23 million people (Wahlström, 2015). Population growth, increased urbanisation, climate change, food scarcity and spiralling bloody conflicts are just some of the underlying factors that signal it is unlikely the scale, intensity, effect and frequency of disasters will diminish in the future (HERR, 2011; Purvis, 2015).

Tens of billions of dollars are spent every year in the humanitarian response to disasters (Lattimer et al., 2016), but shortfalls prevail resulting in decisions to stop or cut humanitarian relief programmes (van der Zee, 2015). Resource limitations heighten pressure to achieve more with less humanitarian funding which, in turn, increases focus on measuring the impact of the humanitarian aid given – with a view to using this knowledge to improve the likelihood of survival and address the welfare of victims in future disasters (Purvis, 2015; Moorhead and Sandler Clarke, 2015; Völz, 2005). The push for accountability and transparency by donors expecting both performance (*the proficient execution of actions*) and effectiveness (*the best possible result from those actions*) as differentiators is further spurred by an increasing intolerance of duplication of effort and wasted resources (Thomas and Kopczak, 2005; Riddell, 2008; HERR, 2011; Lancet, 2010). Numerous factors contribute to creating barriers to accountability and transparency, including:

- The ephemeral nature of the network of actors that mobilise in response to disasters, referred to here as *humanitarian supply networks (HSNs)*;
- Data capture and post disaster assessment are a lower priority to saving lives and alleviating suffering;

- Understandable chaos and pandemonium in a disaster zone;
- Silos of activities and limited data-sharing between a multitude of actors seeking ‘competitive advantage’ when securing funds;
- Uncertainty, guesswork and limited rigour when capturing the real needs of the victims.

The movement of humanitarian goods and services is often referred to as humanitarian logistics and is estimated to represents up to 80% of the annual cost of humanitarian aid (Van Wassenhove, 2005; Tatham and Pettit, 2010). This study uses the term *humanitarian supply network (HSN)* to signify a scope beyond humanitarian logistics. HSNs encompass the entirety of the urgent humanitarian response to a disaster – from securing funds, in-kind aid and specialist skills and services, to sourcing and delivering urgently needed goods and services to alleviate the suffering of the victims (Tatham and Hughes, 2011). The argument here is that the whole eco-system that responds to a disaster is a de facto supply network – analogous to an extended commercial supply chain (xSCM) – mobilising swiftly and existing only for the timespan that emergency relief is needed (Wailgum, 2007; Ferrari, 2015).

HSNs should not however be considered direct equivalents to commercial supply chains. The goods and services that flow through commercial supply chains transform as they progress through the various of processes and actors, accruing not only costs but also value in pursuit of profit (APICS SCC, 2012). With HSNs goods and services flowing through the network only reflect accumulating costs; their value derived from their ability to assist in post-disaster survival. Additionally, unlike commercial supply chains, HSNs do not have established processes and flows that have been honed and calibrated over time. Typically humanitarian actors and processes mobilise in the aftermath of a disaster to rapidly form the network connections for the flow the aid. These necessitated network connections are not expected to mature or improve over time but to dissipate once the activities in the disaster zone transform from emergency relief to reconstruction and rehabilitation.

For large-scale disasters the complexity of inter-organisational coordination and cooperation is significant as, added to the ensuing chaos of the disasters, there are typically a vast number of actors involved. A high impact disaster can command a significant national and international response that may include any or all of the following – governments, armed forces, inter-governmental organisations (IGOs), international non-governmental organisations (iNGOs), national non-governmental organisations (NGOs) and a myriad of local charities together with a host of volunteers (Völz, 2005). Efforts to achieve some level of coordination in emergency situations can be unsuccessful and appear counter-productive. For example, after the 2004 tsunami, in Banda Aceh, Indonesia, UN agencies in situ ran 72 co-ordination meetings per week. These meetings are likely to have had little effect in improving coordination, as for most meetings the NGOs did not have the bandwidth to attend and fewer than 10% of the iNGOs in the area were able to take part (Völz, 2005).

The challenge of harmonising the work of innumerable jockeying actors is unlikely to diminish bearing in mind the continued growth of the third sector (Oxford Dictionaries, 2017). At the start of this century 30,000 iNGOS were believed to be in existence (Tatham and Pettit, 2010), now the United Nations Educational, Scientific, and Cultural Organization (UNESCO) alone lists some 67,000 iNGOs with which it has official relationships (UNESCO-UIA, 2017). The terrain becomes even more crowded when you factor in national NGOs. For example, the United States Department of States currently cites some 1.5 million US NGOs (U.S. Dept. of State, 2017); in India, in 2009, it was estimated that there were some 3.3 million NGOs (Shukla, 2010), which at the time equated to one NGO for every 400 people in the country. In this overcrowded domain it is worth noting that many of the major iNGOs are familiar names, such as CARE International, International Rescue Committee (IRC), Médecins sans Frontières, Oxfam International, Save the Children and World Vision, and that 90% of funds mobilised through the NGO community as a whole are controlled by fewer than a dozen of these large players (Ferris, 2007; Care International, 2017; IRC, 2017; MSF, 2017; Oxfam, 2017; Save The Children, 2017; WVI, 2017).

Even a small subset of these abundant organisations, large and small, mobilising in response to a disaster is likely to require a herculean coordination effort. Not-for-profit actors need to justify their existence to their donors to secure continued funding, therefore working as 'team' with other actors who may share, or take, credit is counterintuitive. Unlike the private sector, the 'not for profit' arena is not typically at risk from corporate amalgamation through mergers and acquisitions. The only major threat to the lifespan of an NGO is lack of funding, therefore competition amongst humanitarian delivery agencies for continued financial support is significant and not surprising (Lancet, 2010). This provides a backdrop that is not conducive to information sharing, co-ordination and collaboration between the numerous humanitarian actors.

The impact of humanitarian aid is also difficult to assess because it can emerge over time – well after the lifespan of the relief effort – and is often influenced by other factors such as development programmes and political machinations (Beamon and Balcik, 2008; Riddell, 2008). Additionally, what may intuitively seem an appropriate response to a crisis can actually have no obvious, or an unforeseen, impact. For example, sending food to victims of famine should save lives, but research confirms that the three million sacks of grain delivered as food aid in Darfur saved few, if any, because most of the deaths were from health crisis and not starvation (de Waal, 2005).

Despite these challenges the need for transparency has given rise to endeavours that measure and monitor the contribution and flow of funds (IATI, 2017; OECD/DCD-DAC/HA, 2017; ATI, 2017; FTS, 2017a; EDRIS, 2017; Lattimer et al., 2016). Efforts that allow comparable traceability of the *when*, *what*, *how* and *impact* of funding, nonetheless are so rare that none have been found by this study. There is even a school of thought that, if any such efforts exist, they are in essence illusions of effectiveness hidden behind a fog of bureaucracy (Polman, 2010). There have also been initiatives to capture post-disaster perceptions of victims through surveys (The Fritz Insitute, 2007) or provide *participating* humanitarian actors with tools and processes to evaluate the results of their humanitarian assistance and then voluntarily share their outputs (ALNAP, 2017). Unfortunately, these initiatives are neither consistently nor comprehensively applied across all disasters: Where they are used, their outputs are not designed to facilitate comparative analysis or ongoing and comprehensive evaluations of the impact of humanitarian intervention. Any granular and definitive data that could potentially help measure the impact of humanitarian aid is likely to reside with the major players in the humanitarian sector. If this data exists, it is not likely to be made freely available at the detail level, nor is it likely to be compatible or comparable across actors. Therefore, considering the dearth of relevant and viable data, it is not surprising that little data analysis can be found as to patterns in the outcomes of disasters and the impact of the humanitarian response.

Guided by two consecutive UN endorsed frameworks for disaster risk reduction – the *Hyogo Framework for Action (HFA) 2005 – 2015* (HFA) and its successor, the *Sendai Framework for Disaster Risk Reduction 2015-2030* (SFDRR) – this study takes an aggregate view of the impact, of the humanitarian response to disasters (HYOGO, 2005; Wahlström, 2015). The HFA and SFDRR suggest measures of progress in efforts to mitigate losses from disasters based on indicators (HFA) and targets (SFDRR). Three quantitative measures align across both frameworks – *disaster-related deaths*, *people affected by disasters* and *financial losses caused by disasters* (ibid).

As these factors are already identified within UN endorsed frameworks, this work examines their utility in the aggregate as *macro-indicators of outcome* (M^{i}O) of disasters. Then, building on M^{i}Os, this study goes on to explore available data for macro-level impact of humanitarian aid, *macro-indicators of impact* (M^{i}I). Finally, it attempts to explore factors extrinsic to the humanitarian domain to identify if there exists a relationship between these factors and M^{i}Os and M^{i}Is for *macro-indicators of effect* (M^{i}E) (The European Commission, 2002). The intent is that identified macro-indicators, and the search for these macro-indicators, will yield valuable knowledge that can be used to effect positive changes in the capture, curation and transparency of data in the humanitarian domain. Additional motivation for the work includes a hope that this work will provide a foundation for further research to drive ongoing improvements across the humanitarian sector.

Finally, this research acknowledges that not all disasters are the same, however many of the significant actors (UN, iNGOs etc.) are ubiquitous across most major disasters, as are many of the donors countries who are members of the Development Assistance Committee (DAC) within the Organisation for Economic Co-operation and Development (OECD). Therefore, as there is some consistency in the participation of core delivery agencies across major disasters there is the possibility that the benefit of experience, knowledge and research within these agencies will be reflected in any indication of progressive improvement found in the impact of humanitarian intervention at the aggregate level.

1.2 Research Aim and Objectives

Despite the issues above, there is a growing need to analyse and improve the effectiveness of humanitarian aid (Moorhead and Sandler Clarke, 2015; HERR, 2011; van der Zee, 2015; Purvis, 2015). No universally agreed method or metric of 'goodness' or 'badness' is employed by the various stakeholders evaluating the impact of humanitarian efforts (Wahlström, 2015; HYOGO, 2008). It is also unknown if the currently available quantifiable data are capable of supporting any search for measures of outcome, impact or effect, at any level of aggregation or detail. As a step towards addressing this issue, this research seeks to examine web-available curated data for macro-indictors of the outcome of disaster, impact of humanitarian aid and the effect on disaster outcome from factors extrinsic to the humanitarian sector. Accordingly, the aim of this research is to answer the question:

Can exploration of curated web-available data yield macro-indicators of humanitarian intervention in the aftermath of disasters?

The following objectives are set to help answer this question:

Objective 1:
Review current practices and research relevant to assessing the effect of humanitarian intervention.

Objective 2:
Explore curated web-available data of global disaster losses, humanitarian aid and other factors extrinsic to the humanitarian sector for macro-indicators that may signpost the consequences of humanitarian intervention.

Objective 3:
Evaluate the artefacts, theories and findings from this study in the context of the research domain, identifying knowledge contribution, research limitations and the potential for future research.

1.3 Research Approach

A design science research (DSR) approach is taken for this study. The essence of this approach is "*learning through the act of building*" (Kuechler and Vaishnavi, 2008) This effectively means that knowledge of the *solution and the problem* space is developed through the iterative and/or incremental development of artefacts, with emphasis placed on the utility of the created artefacts. At the heart of the DSR methodology is a design cycle that is repeated until a 'good enough' and 'satisficing' outcome is achieved [*Figure 1-1*] (Simon, 1996; March and Smith, 1995; Vaishnavi, 2008; Hevner et al., 2004).

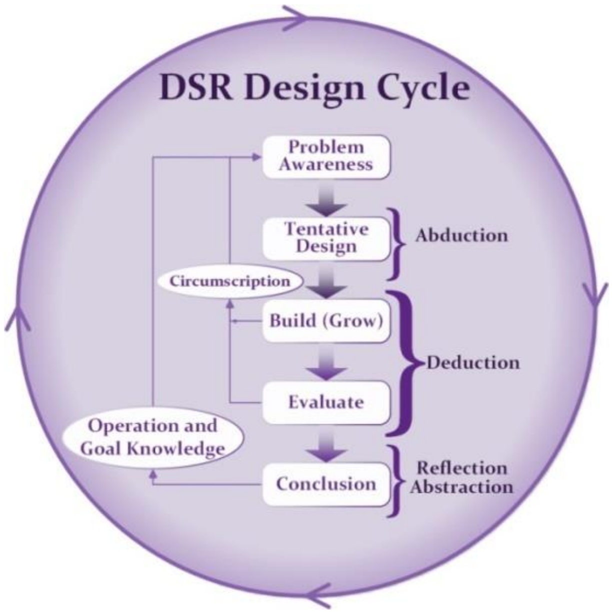


Figure 1-1: DSR Design Cycle
Adapted from (Vaishnavi and Kuechler, 2004b; Hevner, 2007)

Each rotation of the design cycle includes the following steps:

1. **Problem Awareness** explores the problem identified in this study using whatever applicable knowledge is available at the beginning of each loop.
2. **Tentative Design** is the pre-build activity of designing a solution abductively, based on awareness of the problem.
3. **Build (Grow)** is the creation for designed artefacts.
4. **Evaluate** assesses the utility of the artefacts created and harvests the knowledge that emerges during the iteration.

5. **Conclusion**, which is completed at the end of the research to bring the research to a close and articulate the 'satisficing' solution achieved (Simon, 1996).

It is worth noting here that ***Circumscription***, on the flow back to *Problem Awareness* [*Figure 1-1*], is not a DSR step per se, but signifies the process of *learning from what did not 'work'* in order to inform the shape of the next iteration during the *Evaluate* step and assess the 'satisficing' outcome on *Conclusion* of the study (Simon, 1996; Vaishnavi and Kuechler, 2004b; Hevner, 2007).

This constructive research method, sometimes referred as 'improvement research' (Vaishnavi, 2008), is considered well suited to addressing the challenges of seeking macro-indictors of the outcome of disasters and the impact of humanitarian intervention, as well as any effect on the outcome of disasters from extrinsic factors.

1.4 Thesis Structure

This thesis is structured as eight chapters. A brief summary of each of the chapters is included below and schematics of the thesis flow are included at the end of this chapter [*Figure 1-2* & *Figure 1-3*].

Chapter 2 Research Domain

This chapter provides a literature review of the humanitarian space as it pertains to stresses and pressures in the humanitarian sector and the imperative to measure the effect of humanitarian intervention. It describes the definition and recording of disaster, the components of disaster management, the concept and idiosyncrasies of humanitarian supply chains and the paucity of crucial data to measure outcome, impact and effect.

Chapter 3 Research Approach

This chapter describes the concept of 'wicked' problems and characterisation of this research problem of this study as 'wicked'. It goes on to describe the structure and processes of the design science research approach employed by this study, which is particularly suited to 'wicked' problems. Finally, the chapter outlines the research artefacts and design theory outputs of this research and provides an overview of the structure of the study.

Chapter 4 Disasters (Iteration 1)

This chapter describes and explains the foundational iteration of the design cycle. It synthesises knowledge of the problem from the review of the research domain to create the first tentative design of this study. The chapter discusses the acquisition, preparation and examination of the Emergency Events Disasters (EM-DAT), originally considered the single credible comprehensive source of disaster loss data (Guha-Sapir et al., 2017l). The chapter goes on to identify and describe shortcomings in the EM-DAT datasets that inhibit the search for macro-indicators of outcome (M^iOs). Finally, it outlines the consequences of these shortcomings on the remainder of this work and the basis of the change of the structure of the study from three iterations to four iterations.

Chapter 5 Data Veracity (Iteration 2)

The chapter scans the data science domain for a definition of data veracity and toolset that can be used to evaluate the veracity sourced datasets. These are needed to enable a consistent understanding of the veracity of disaster datasets required to address the shortcomings of EM-DAT as a disaster data source for this study (Guha-Sapir et al., 2017l). As no definition or toolset for data veracity is found this chapter outlines the creation of a data veracity framework (DVf) that can be used by this work.

Chapter 6 More Disasters (Iteration 3)

This chapter describes and explains the third iteration of the design cycle which returns to the humanitarian domain. It describes the use of five disaster datasets, in addition to EM-DAT, in order to create a Master Set of Global Disasters (MSGD). This includes the application, and therefore the testing, of the data veracity framework as a prerequisite to constructing the MSGD. The chapter also outlines the creation of Master Disaster Classification (MDC) in support of the MSGD. Finally, it goes on to use the MSGD dataset and MDC model to identify *mean survival rate by year* as a M^iO.

Chapter 7 Aid and Population figures (Iteration 4)

This chapter describes expands on the work of Iteration to search for M^iIs from M^iOs and humanitarian aid and go on to use found macro-indicators to explore other international funding and population figures to identify M^iEs. The chapter concludes the research and as part of this process outlines the core artefacts developed and the utility theory presented and tested.

Chapter 8 Conclusion

This chapter examines the extent to which research objectives, are met. It also articulates salient findings and contribution to knowledge from this work. The chapter goes on to discuss the limitations of the study and posit opportunities for further research.

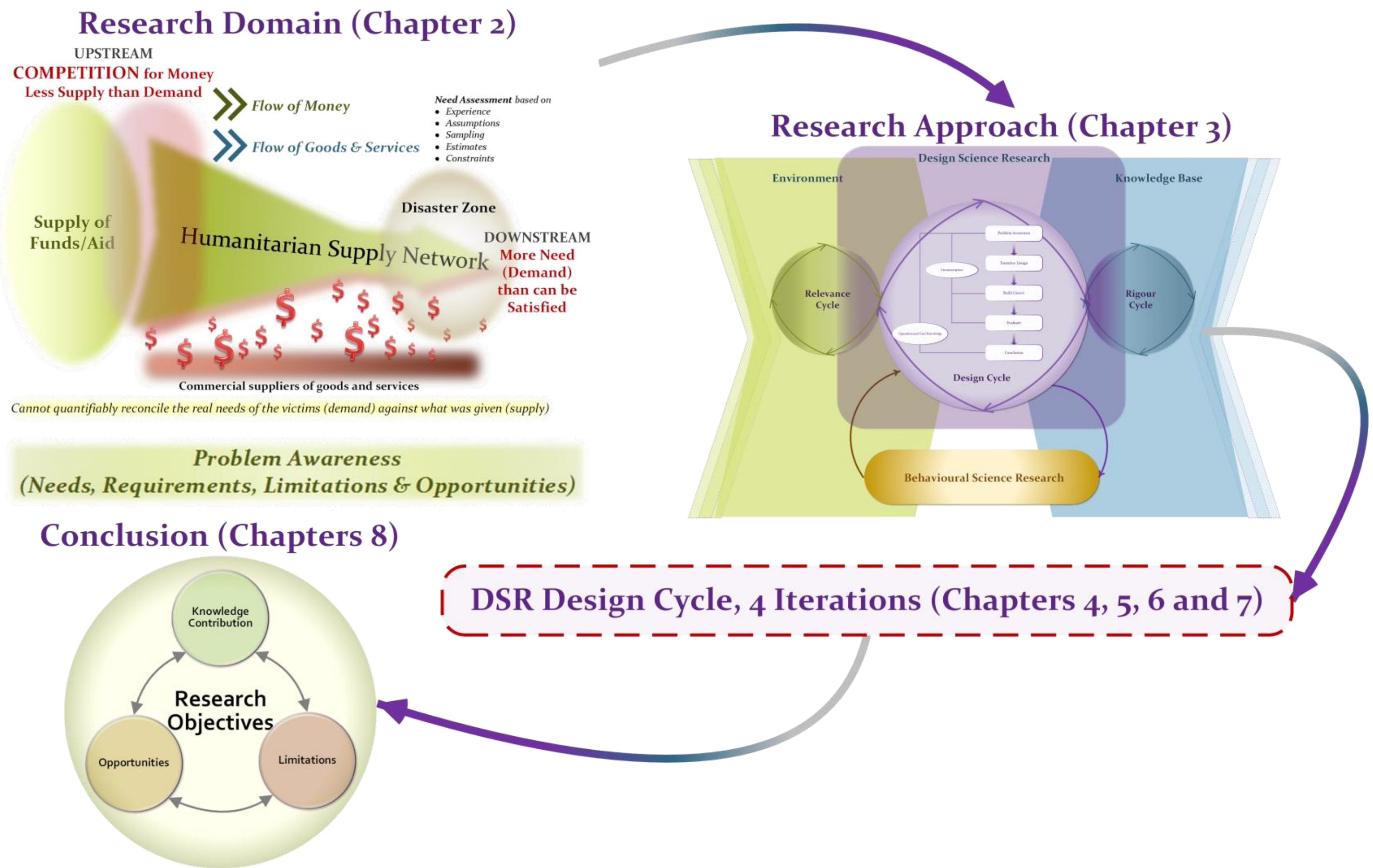


Figure 1-2: Thesis Flow – High-level View, DSR Design Cycle in Context

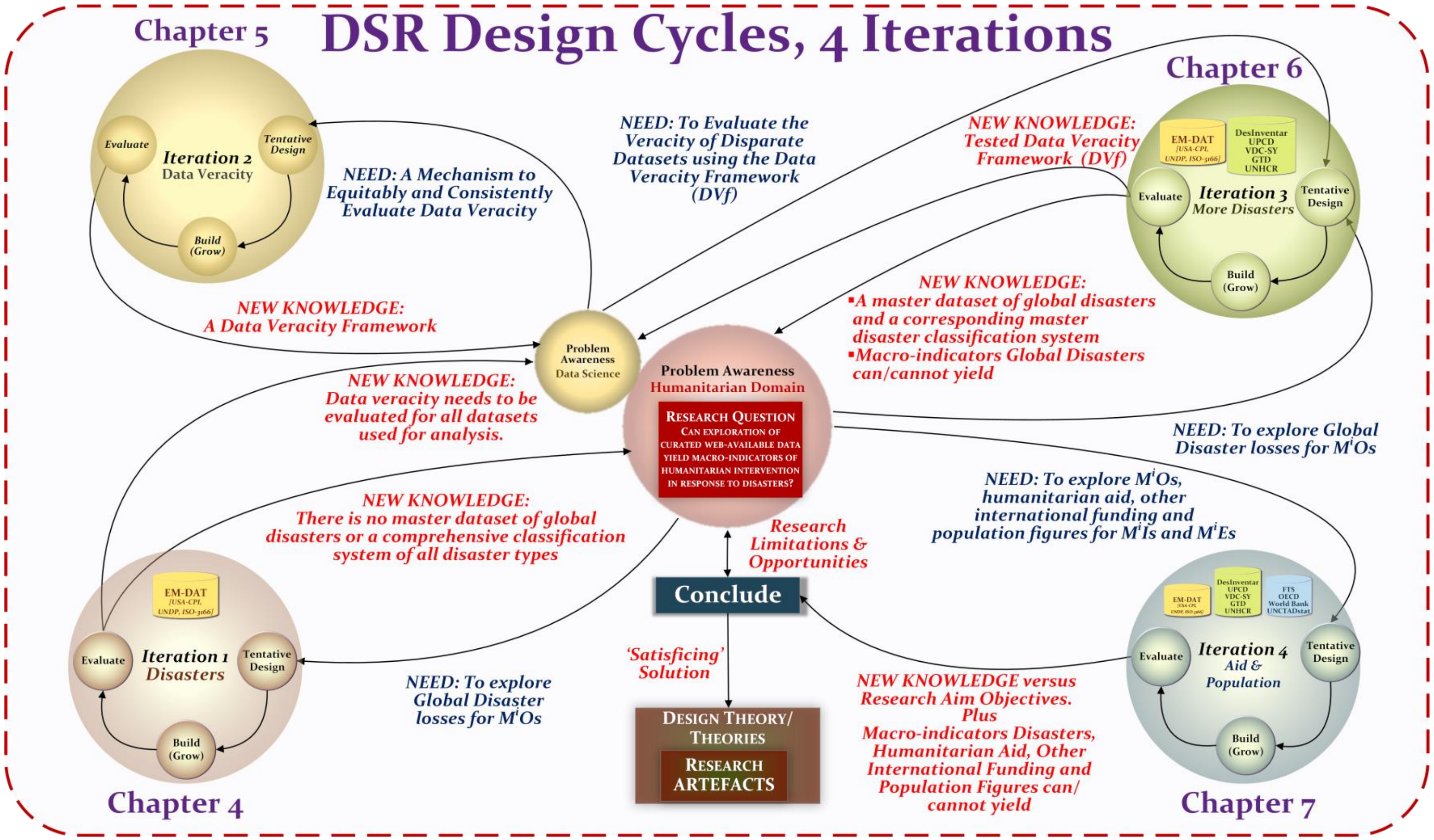


Figure 1-3: Thesis Flow – DSR Design Cycle Iteration Chapters

Chapter 2: Research Domain

2.1 Overview

This chapter primarily describes the stresses and pressures on the humanitarian sector and the imperative to measure the effectiveness of humanitarian efforts. The aim here is to provide a contextual background to the research and lay the groundwork to explain a number of the challenges and obstacles that are addressed or circumvented to complete this study. The chapter also explains concepts and terminology relevant to understanding the research domain.

This chapter is structured as follows. *Section 2.2* provides a brief perspective of the growing need for humanitarian assistance and the imperative to assess the impact of the aid given to those affected by disasters. *Section 2.3* outlines some of the challenges facing attempts to gauge the effectiveness of humanitarian intervention. It also introduces and describes constructs and nomenclature relevant to this study. *Section 2.4* introduces the concept of '*macro-indicators*' of humanitarian intervention in the aftermath of disasters. Additionally, recognising the enabling medium for this work is *data*, *Section 2.5* contextualises this study within the nascent research domain of *data science*. Finally, *Section 2.6* provides an overall summary of the chapter.

2.2 The Problem

The world experiences hundreds of humanitarian crisis each year, and the intensity, frequency and consequences of these disasters are expected to increase (HERR, 2011). It is believed factors contributing to this increase include a rising world population, continued urbanisation, climate change and a growing number of conflicts (HERR, 2011; Gates et al., 2016; Schleussner et al., 2016). Additionally economic interdependencies and an interconnected world in which disease and conflict can cross borders will mean future crisis could differ significantly from past experience (HERR, 2011). It is predicted by 2030, 61% of the world's population, over 5 billion people, will live in urban areas, making it more likely that future disasters will hit urban centres and result in a significant intensification of consequence from disasters (ibid).

The humanitarian sector's purpose is to respond to disasters in order to save lives and alleviate the immediate suffering of the affected (Humanitarian Coalition, 2017). The sector is becoming increasingly underfunded, thus exacerbating the stress on an over-extended system and presaging diminishing resources stretched beyond capacity to meet ever-growing needs (HERR, 2011). Arguably, the tens of billions of dollars spent each year in response to disasters are ominously inadequate in meeting humanitarian needs. In 2015 the UN was able to raise only US$10.9bn of the US$19.8bn needed, a 45% deficit (Lattimer et al., 2016; Purvis, 2015). Such extreme shortfalls in funding inevitably lead to decisions to stop or cut life-saving humanitarian relief programmes (van der Zee, 2015; Belanger et al., 2016). This downward pressure on funds is exacerbated by recent political aversions to funding humanitarian needs by developed countries that have in past been mainstays of the humanitarian effort (Seddon, 2017; Chambre, 2017).

The need to achieve more with less, spurred by an increasing intolerance for the duplication of effort and waste, has created a push for accountability and transparency in the humanitarian sector (Thomas and Kopczak, 2005; Riddell, 2008; HERR, 2011; Lancet, 2010). There is also now an inherent expectation of effectiveness from every penny spent and a requirement to reassure that relief efforts are measurably reducing deaths and alleviating the suffering of those affected by disasters. Such reassurance, however, at a 'per disaster' level is dependent on the availability of accurate and complete data. For example, details of what is truly *needed by* the affected; what is formally *requested for* the affected; and what is ultimately *provided to* the affected; and the *timeliness* of the flow of these provisions and services is crucial to any assessment of humanitarian aid impact (Lattimer et al., 2016). Unfortunately, notwithstanding the considerable efforts to improve transparency by the humanitarian sector, such data is not available (IATI, 2017; HDX, 2017; FTS, 2017a; DesInventar.NET, 2017; Guha-Sapir et al., 2017l).

To date, examples of solutions to empirically assess relief efforts have largely consisted of post-disaster surveys and reports (The Fritz Insitute, 2007; ALNAP, 2017; CHS Alliance, 2017), the outputs of which tend to be qualitative and do not easily lend themselves to comparative analysis. Additionally, these post-disaster assessments are typically voluntary, varying in scope and sporadic therefore do not provide a comprehensive and congruent perspective across the gamut of world-wide disasters that have occurred in recent times.

Other research and guidance on the effectiveness of relief efforts have typically focussed on the performance of processes or have been from the perspective of a specific organisation or theme (Beamon and Balcik, 2008; Sauer, 2016; Ali et al., 2017; Hella et al., 2014; Everett and Friesen, 2010; Christian Aid, 2015; Dabelstein, 1999). Notably, one 2015 paper discusses in detail theoretical methods to evaluate the impact of humanitarian assistance based on the analysis of case studies while bemoaning the lack of sufficient 'high-quality' evidence' and the scarcity of baseline data (Puri et al., 2015). Acknowledging that there are difficulties facing any granular analysis of the effectiveness of humanitarian assistance, this study posits that some directional sense of effectiveness may be visible by way of 'macro-level' aggregated data of disaster losses, humanitarian aid funding and factors external to the humanitarian sector, such as development aid and population data.

2.3 The Challenges

This section describes a number of the challenges confronting any analysis of the effectiveness of humanitarian intervention, at the micro or macro level, such as:

- The lack of common understanding of what constitutes a disaster and the categorisation of disasters;
- The rarity of universal standards, processes and systems to consistently and accurately monitor or capture disaster losses;
- The absence of an innate systemic relationship that reconciles demand (what is needed in the disaster zone), supply (the aid response in goods and services) and the flow of funds.

2.3.1 The Definition & Recording of Disasters

The initial and rudimentary challenge faced by this study is the lack of a single commonly used standard definition of a disaster that clearly and comprehensively describes what constitutes such an event. Numerous definitions of disaster exist (Lerner, 2016; LOG, 2017; DesInventar, 2017c; JHSPH, 2008; John and Thangamani, 2015; IRDR, 2014; Below et al., 2009). Some organisations use quantifiable criteria to consider an event a disaster (Guha-Sapir et al. 2017e), while others have freer, more subjective, descriptions of the events they consider a disaster (JHSPH, 2008; FEMA, 2017; IFRC, 2017; Gortney, 2016). In the absences of a widely accepted standard this study adopts the definition of a disaster used by the United Nations Office for Disaster Risk Reduction (UNISDR) [*Figure 2-1*]:

> **Disaster:** *"A serious disruption of the functioning of a community or a society involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources"*

Figure 2-1: UNISDR definition of a Disaster
(UNISDR, 2009)

Considering the lack of clarity as to what constitutes a disaster, it is not surprising that there are also no known universally accepted standards, processes and systems to record disasters (De Groeve et al., 2014). More often than not web-available disaster data are sourced from a variety of third-parties by organisations or institutions that collate and curate disaster databases (Guha-Sapir et al., 2017k; UCDP, 2017e; GTD, 2017a; UNHCR, 2017a). This obscures the quality and variety of original data collection methodologies (if any exist), which become even more opaque as these data aggregators apply their respective standards and policies as to how they wish to hold and present their datasets (GTD, 2017d; UCDP, 2017e; Guha-Sapir et al., 2017i; UNHCR, 2017a). Even when data is collected and updated in databases more local to the disaster zone, the flexibility and variability of localisation obfuscates any signs of standardised processes and methods that may have been suggested by the open source solution provider (DesInventar, 2017d). Examples of how lack of clarity and consensus in the recording of disasters can be problematic include:

- **Deaths**: At which point is a death counted and attributed to the disaster? Only in the immediate aftermath? What if those injured died a week later? What if thousands died days later because of the lack of resources or aid e.g. medicines, clean water?
- **People Affected**: What is meant by affected, e.g. injured, missing, homeless, forced out of the country, etc.? Are those that move away (self-evacuate) counted? If a secondary issue affects those that survive the disaster, how are these counted, e.g. refugees drowning while escaping conflict?
- **Financial losses**: How are losses, insured and uninsured, calculated and in which currency? If converted to US$, at what exchange rate? How is the loss of critical services, destruction of infrastructure and loss of livelihood costed etc.? Are calculations carried out to quantify any secondary effect if there is a loss of life-critical resources such as healthcare, food and water?

Where even stark disaster loss data, such as numbers of dead and affected, is not provided by sources, the aggregators make do with their own estimates. As there is no known universally accepted standard practice for estimating disaster losses, organisations and institutions apply their own bespoke estimating practices (De Groeve et al., 2014; Guha-Sapir et al., 2017i; DesInventar, 2017b; VDC-SY, 2016b; GTD, 2017e; UCDP, 2017b).

One area of tacit consensus that does seem to exist is in the description of disasters in broad generic terms, it is not uncommon to find disasters referred to as *'sudden-onset'* or *'slow-onset'* and grouped as *'natural'; 'technological'*; or *'complex'* [*Table 2-1*] (Van Wassenhove, 2005).

	Natural	Technological	Complex
Sudden-onset	Earthquake	Chemical Leak	Terrorist Attack
Slow-onset	Drought	Pollution	Refugee Crisis

Table 2-1: Disaster Groupings – Example 1
(Van Wassenhove, 2005)

Although, even these high-level groupings can differ, e.g. '*natural*', '*man-made*' and '*hybrid*' [*Figure 2-2*] (Ibrahim Mohamed, 2007), and apart from '*natural*' disasters, which are typically 'acts of God', the groupings are difficult to scope and bound.

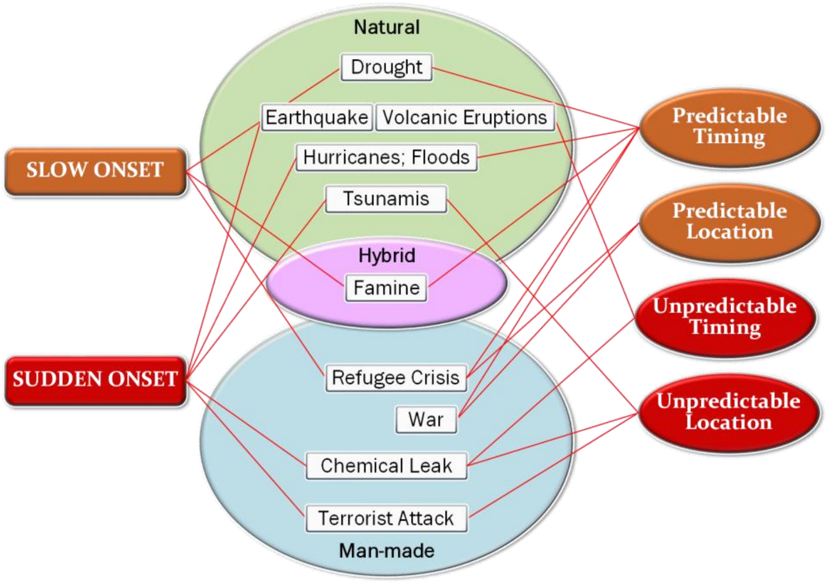


Figure 2-2: Disaster Groupings – Example
(Ibrahim Mohamed, 2007)

Thus the challenge of extracting intelligibility from the various repositories of disaster information across international and administrative boundaries is exacerbated by the variety of definitions, classifications and nomenclatures proposed, adapted or adopted by the numerous humanitarian actors (Guha-Sapir et al., 2017g; DesInventar, 2017c; IRDR, 2014; Coburn et al., 2014; GLIDE, 2017; Lerner, 2016).

2.3.2 The Disaster Management Cycle & Aid

One concept that appears to be generally understood and accepted within the humanitarian sector is the *disaster management cycle*, which consists of four phases of disaster-related activities – *mitigation*; *preparation*; *response* and *recovery* [*Figure 2-3*] (Cozzolino, 2012; Safran, 2004; Vasilescu et al., 2008). It is perhaps simpler to think of these phases as either pre-disaster (*mitigation* and *preparedness*) or post-disaster (*response* and *recovery*).

Figure 2-3: Disaster Management Cycle
(Cozzolino, 2012; Safran, 2004; Vasilescu et al., 2008)

Mitigation includes measures typically put in place by governments to minimise losses from disasters (Cozzolino, 2012; Vasilescu et al., 2008). *Preparedness*, also normally within the national remit, is about contingency planning and pre-positioning of resources, emergency facilities and evacuation mechanisms that can be rapidly deployed when a disaster occurs (Cozzolino, 2012; Safran, 2004; Vasilescu et al., 2008). *Response* equates to the initial actions taken in the immediate aftermath of a disaster to minimise deaths and provide emergency relief to those affected (Vasilescu et al., 2008;

Cozzolino, 2012). *Recovery* incorporates reconstruction and rehabilitation to restore conditions to pre-disaster levels, or better (Vasilescu et al., 2008; Safran, 2004).

Humanitarian aid corresponds to the *response* phase of the disaster management cycle. Humanitarian aid programmes deploy rapidly, often heavily supported through appeals for private donor funding, and from the point of view of these donors, are expected to end when the emergency ends (Scholten et al., 2010; Riddell, 2008). Nonetheless, there are cases when *response* needs subside yet humanitarian aid continues into the post-disaster activities of the *recovery* phase (Riddell, 2014a).

Development aid has no obvious place in the disaster management cycle. Development aid is sourced from governments of developed countries and given to developing countries for economic, political, environment and social development (OECD, 2017a). Development aid programmes take time to plan and implement as their focus is on nurturing and enabling long-term change designed to improve the lives of beneficiaries and their future generations (UNDP, 2017b). For some developing countries, however, pre-disaster phases, i.e. *mitigation* and *preparedness*, are bolstered through development aid initiatives (Riddell, 2014a).

In short, the distinction in the use of humanitarian aid and development aid is not always clear (Riddell, 2014b). Some agencies extend the boundaries of their post-emergency efforts from relief into *recovery* and possibly even through to development initiatives, which may or may not have any relationship to the disaster that initiated the funding (Riddell, 2014a). This means there is ambiguity as to whether humanitarian aid is used to supplement development activities or if it is supplemented by development funding. As a result, cleanly attributing relief effectiveness exclusively to the humanitarian aid given is an unrealisable aspiration (Buchanan-Smith and Fabbri, 2005; Riddell, 2008, 2014b).

2.3.3 The Humanitarian Supply Network (HSN)

The idiosyncrasies of *humanitarian supply networks* (*HSNs*), which are the manifest response to disasters, are argued here to underlie the scarcity of data crucial to measuring the effectiveness of humanitarian intervention. Regardless of how well or poorly the pre-disaster phases of a disaster management cycle are executed, or what type of aid is given and where in the cycle it is applied, the need and ability to assess the effectiveness of disaster management only becomes feasible after a disaster occurs. Additionally, and logically, any measure of effectiveness is contingent on the timeliness and veracity of the data captured during the ensuing humanitarian response, i.e. during the lifespan of the HSN.

It is important to consider the entire system of funding, sourcing and supplying humanitarian assistance in response to a disaster to be a *de facto* supply network (Vasilescu et al., 2008; Safran, 2004). This is a network of supply (*money, goods and services*) that forms horizontally and vertically, typically across organisational, institutional and international boundaries, mobilising for a single purpose – *serving the humanitarian needs of those affected by a disaster* (Tatham and Hughes, 2011). As such HSNs align to the *response* phase of the disaster management cycle and are expected to dissipate when conditions stabilise enough to allow relief activities to transition to *recovery* activities (Safran, 2004; Tatham and Houghton, 2011). An HSN can be considered to be akin to an extended supply chain, *or 'ecosystem' (*Ferrari, 2015), that rapidly forms on the fly to satisfy a very specific set of needs and disperses once the needs (and urgency) diminish. HSNs are perceived here to extend beyond the bounds of humanitarian logistics (the movement of humanitarian goods and services), which in and of itself is considered by some to represent up to 80% of the annual cost of humanitarian aid (Van Wassenhove, 2005; Tatham and Pettit, 2010).

In recent years there has been an increase in the development of thought, theory and practice surrounding HSNs (Ergun et al., 2010). To the extent that there is now a growing body of research relating to their performance (Beamon and Balcik, 2008; Sauer, 2016; Tatham and Hughes, 2011), with scholars proposing that there may be opportunities for the 'not-for-profit' sector to benefit from knowledge of improved supply chain performance gained in the 'for-profit' sector. This is pertinent as humanitarian supply networks are believed to be lagging behind their commercial counterparts by about 15 years (Van Wassenhove, 2005; Tatham and Pettit, 2010; Moore et al., 2011; Tomasini and Van Wassenhove, 2009).

Notwithstanding the above proposition, it is important to take cognizance of some important differences between humanitarian supply networks and their commercial counterparts. Commercial supply chains are *profit-driven* and their flows of goods, services, money and data are inextricably integrated and tightly-coupled [*Figure 2-4*]. Moreover, they are developed over time and designed for longevity and growth.

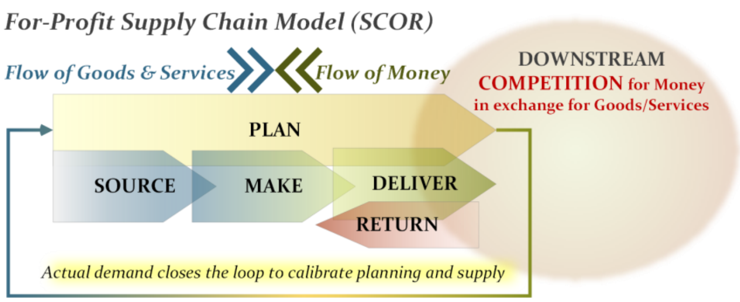


Figure 2-4: 'For Profit' Supply Chain Model (Schematic)
(APICS SCC, 2012)

Humanitarian supply networks, in contrast, are *event-driven* and *temporary*. Their rapidly formed flows of goods, services, money and data are typically disjointed and disconnected. For HSNs speed and cost are key drivers, not profitability and market share (Scholten et al., 2010). Importantly, HSNs do not compete to satisfy demand (i.e. the urgent needs of the affected victims) as demand inevitably exceeds supply and those affected are not in a position to 'shop around' for alternatives (Beamon and Balcik, 2008). Contrarily, heated competition in HSNs is in fact upstream for the 'supply' of funds, as agencies and NGOs vie for money [*Figure 2-5*] (Lancet, 2010; Riddell, 2008; Stephenson, 2005). Additionally, the multitude of actors and lack of cohesion within the humanitarian system make efficiency, agility and transparency difficult goals to attain (Polman, 2010; Völz, 2005).

Figure 2-5: Humanitarian Supply Networks

Emphasis has been placed on defining and describing *humanitarian supply networks (HSNs)* as an explicit term in order to accentuate that after a disaster *supply (aid) flows to meet demand (need)* and that this is notionally equivalent to commercial supply chains. It should however be noted that unlike commercial supply chains, disaster response demand/supply flows are not closed-loop systems matching supply to demand, which is something for-profit organisations continuously strive to optimise. It is argued here that conceptualising the entirety of the post-disaster emergency relief effort as a humanitarian version of an extended commercial supply chain underlines the powerful implications of the lack of an intrinsic feedback-loop; not least of which is a deficit of quantified data to confirm real need is matched by real supply.

2.3.4 The Paucity of Reconcilable Data

It is the case that comprehensive quantifiable data that ties together disaster-specific losses with humanitarian funding for all disasters in recent history is not available. Only disparate databases holding disaster loss <u>or</u> aid funding data, together with an assortment of post-disaster subjective surveys and reports, can be found (The Fritz Insitute, 2007; ALNAP, 2017; CHS Alliance, 2015; Dabelstein, 1999; Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017c; FTS, 2017a; IDS, 2017).

One likely, but significant, underlying cause of this dearth of data matching *aid* (supply) to *need* (demand), is the unique nature of HSNs. Disaster victims do not pay for the goods or services they receive. This obvious fact means the receipt of aid is fundamentally disassociated from the funding of aid, which in turn has a profound effect on the systemic necessity for transactional data collection. Unlike commercial supply chains, there is no tightly-coupled '*end-of-chain*' dependency between <u>what is received</u> and <u>payment for what is received</u>. The absence of this crucial dependency effectively means that *what is received* need not match *what is funded*. Consequently *what is received* also need not reconcile to *what is needed* or *what is requested*. Ultimately, and inevitably, any consequences of mismatches between *needed*, *requested*, *received* and *funded* are felt by the victims of disasters. Of particular relevance is that continued funding by donors (for which there is considerable competition) is not typically dependent on accurate and indisputable measurement or effectiveness of the aid provided (Riddell, 2008; Stephenson, 2005; Tatham and Hughes, 2011).

Another obstacle in the path of comprehensive, accurate and compatible data collection is the complexity and confusion caused by the morass of humanitarian actors that can mobilise in the event of a disaster, particularly a large-scale international disaster. The

International Committee of the Red Cross (ICRC) estimate that major disasters can attract an average of a thousand different organisations, which is not surprising considering that earlier this decade the United Nations Development Programme estimated the total number of international non-governmental organizations (iNGOs) to be greater than 37,000 (Polman, 2011). As a concrete example, more than 400 iNGOs, as well as numerous smaller initiatives that bypassed the UN registration process, were operating in Indonesia immediately after the 2004 south Asian tsunami (Völz, 2005). It is therefore understandable that the complexity and scale of the inter-relationships between the myriad humanitarian actors places considerable impediments in the way of capturing accurate coherent data and accrediting positive effect to a specific group or actor (Riddell, 2008, 2014b).

Furthermore, the vagaries of data collection can also prove to be a hindrance to the availability of reliable detailed data that can illuminate any mismatches in *needed-requested-received-funded* aid. Post-disaster data collection can be subjective, biased, incomplete, flawed, intentionally never captured or logistically difficult to obtain (Tatham and Hughes, 2011; Maiers et al., 2005; Riddell, 2008). Take for example victim's needs assessment, which when articulated by aid delivery organisations, are often 'guestimates' that are:

- Based on the perceptions and experience of field workers who are plagued with difficulties of access and coverage; and not based on the real need or input of those affected (Tatham and Hughes, 2011; Riddell, 2008; Darcy and Hofmann, 2003; Riddell, 2014b);
- Potentially inflated due to varying boundaries between relief and development needs and to counteract the limitations caused by donations that are tied by the donor to specific supply sources (Buchanan-Smith and Fabbri, 2005; Riddell, 2014b, 2008; Beamon and Balcik, 2008; Smillie and Minear, 2003);

- Prone to distortion because agencies may ask for what can be funded rather than what needs to be funded (Smillie and Minear, 2003; Riddell, 2014b, 2008).

Considering the difficulties, and possible 'loss of funding' repercussion of less than favourable findings, it is little wonder that gauging the true effect of humanitarian aid is not a high priority for the multitude of actors that operate across HSNs (Riddell, 2014b; Polman, 2010).

2.4 Macro-Indictors of Humanitarian Intervention

The obstacles to objectively, quantifiably and comparatively assessing the effectiveness of humanitarian intervention as it pertains to each *specific* disaster appear to be insurmountable. Hope may lie, however, in free web-available curated data that, in the aggregate, may hold the possibility of signalling directional macro-level changes in disaster outcome, humanitarian aid impact and the effect on or from extrinsic factors such as other international funding and population. Any or all of these macro-indicators, if found, may provide some sense of the overall effectiveness of humanitarian intervention efforts [*Figure 2-6*]:

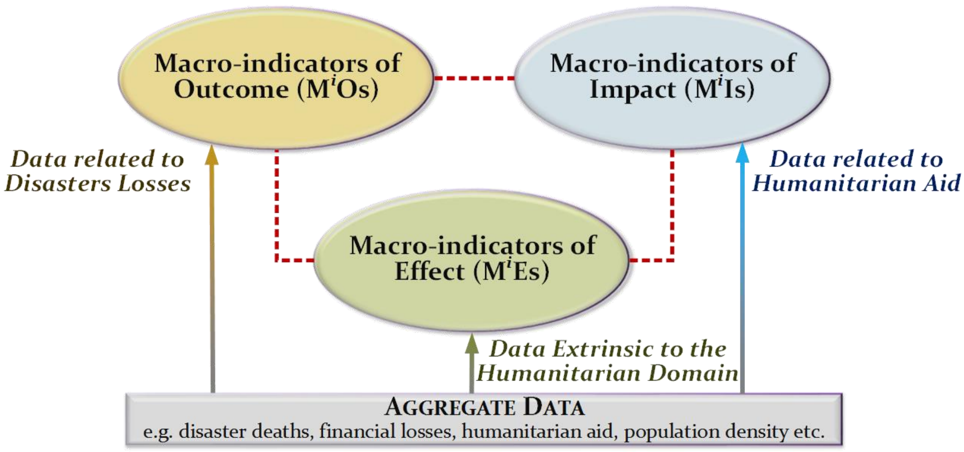


Figure 2-6: Macro-indicators of Outcome, Impact and Effect

This study uses the 3 macro-indicator terms as follows [*Table 2-2*]:

Macro-Indicator of *Outcome* (M^iOs)	An aggregate indicator that shows a pattern of change in the outcome of disasters over a unit of time.
Macro-Indicator of *Impact* (M^iIs)	An aggregate indicator that shows a pattern of change in the outcome of disasters (**M^iOs**) in relation to a measure of humanitarian aid.
Macro-indicator of *Effect* (M^iEs)	An aggregate indicator that shows a pattern of change in the outcome of disasters (**M^iOs**) or the impact of humanitarian aid (**M^iIs**) in relation to (a) the flow of other international funding; and (b) population growth (a) Tests the assertion that the lines between humanitarian aid and other international (development) aid are blurred, which if true is expected will show as a meaningful relationship between this and any **M^iOs** and **M^iIs** (Riddell, 2014a). (b) Tests the assertion that population growth has a negative effect on the outcome of disasters (HERR, 2011).

Table 2-2: Macro-indicators

Taking some direction from a European Commission "Guidelines for the use of Indicators", here *outcome* refers to metrics at the level of disasters losses, e.g. deaths, survivors. financial costs etc. (The European Commission, 2002). Referring back to the problem of questionable transparency and the need to effectively use limited funds to address increasing needs (Thomas and Kopczak, 2005; Riddell, 2008; HERR, 2011; Lancet, 2010); *impact* is used here to gauge the relationship, if any exists, between the flow of billions of US$s of humanitarian aid and the outcome of disasters (Lattimer et al., 2016). These constructs are further expanded by using *effect* as a term used for indicators that may show an effect on the *outcome* of disasters or the *impact* of humanitarian aid. For *effect* this work specifically considers other international funding, to test the assertion that this can be used to bolster disaster management activities (Riddell, 2014a), and population growth, to test the assertion that larger populations equate to worse *outcomes* of disasters (HERR, 2011).

It is worth mentioning here that in identifying possible macro-indicators of *outcome* guidance is taken from two UN endorsed frameworks for *Disaster Risk Reduction (DRR)*, the Hyogo Framework for Action (HFA) 2005–2015 and the Sendai Framework for Disaster Risk Reduction (SFDRR) 2015-2030 (UN, 2017; Wahlström, 2015; HYOGO, 2008). In 2005 the Hyogo Framework for Action (HFA) 2005–2015 offered a blueprint for disaster risk reduction efforts (HYOGO, 2005), and recommended three disaster loss indicators as measures of progress [*Figure 2-7*].

(1) "Number of deaths arising from natural hazard events"

(2) "Total economic loss attributed to natural hazard events"

(3) "Number of people affected by natural hazard events"

Figure 2-7: HFA – Recommended Progress Indicators
(HYOGO, 2008)

Note that the HFA's scope for these indicators was specifically "*natural hazard events*" and the indicators did not suggest targets that would facilitate a measure of progress (HYOGO, 2008).

In 2015 the Sendai Framework for Disaster Risk Reduction (SFDRR) 2015-2030 succeeded the HFA 2005 (Wahlström, 2015), this time recommending seven global targets. The first three of these targets are quantifiable and align to the three progress indicators of the HFA [*Figure 2-8*] (Wahlström, 2015; HYOGO, 2008).

(1) "Substantially reduce global disaster mortality by 2030, aiming to lower the average per 100,000 global mortality rate in the decade 2020–2030 compared to the period 2005–2015;"

(2) "Substantially reduce the number of affected people globally by 2030, aiming to lower the average global figure per 100,000 in the decade 2020–2030 compared to the period 2005–2015;"

(3) "Reduce direct disaster economic loss in relation to global gross domestic product (GDP) by 2030;"

Figure 2-8: SFDRR Targets that align to HFA Progress Indicators
(Wahlström, 2015; HYOGO, 2008)

Conspicuously both disaster risk reduction frameworks use macro-level deaths, people affected and financial losses as indicators or targets of progress. Thus this study's focus on disaster losses as the basis of potential macro-indicators of outcome aligns to the approach taken by the HFA and SFDRR.

There are however two points of distinction between this study and the referenced DRR frameworks worth highlighting (HYOGO, 2008; Wahlström, 2015). First, there is the aforementioned focus of these DRR initiatives on disasters caused by natural hazards (UNISDR DRR, 2017). In contrast this study is interested in *all types of disasters that attract humanitarian funding*. Second, the DRR frameworks are focused on resilience to disasters, i.e. the *mitigation* and *preparedness* phases of the disaster management cycle. This study is interested in exploring the effect of humanitarian intervention, interpreted here as humanitarian funding. The DRR frameworks do not appear to take into consideration that the indicators and targets they are recommending as measures of progress in disaster resilience are as, if not more, likely to show the effects of the response to a disaster than resilience to it. To explain by way of an analogy – *if everyone survived a fire, was it because the first responders were adept enough to save them all or because the building was built to adequate fire-safety standards?* The answer to this could be either, both, or despite deficiencies in one or the other.

Realistically, with data currently available, it is not possible to isolate or delineate cleanly which, or how many, of the disaster management phases (*preparation*, *mitigation*, *response* or *recovery*) had what effect. As evidenced by the HFA and SFDRR recommended indicators/targets, gauging any progress (effect) is only really possible after a disaster occurs – continuing the fire analogy, *there has to be fire to know how many are able to survive it and if the scale of destruction is contained*. Acknowledging the existence of these obscurities, this work takes a simple position of not attempting to distinguish between the four steps of disaster management, but simply view any patterns of disaster losses as a tacit reflection of all disaster management activity.

2.5 The Data Science Context

At its core this research seeks answers to questions within the humanitarian domain. It does however also hold relevance within the data science domain as the *study of data* is central to achieving this work's intended ends. This section therefore discusses *(1)* the nascency of the data science domain; and *(2)* considerations relevant to the realisation of value in data science undertakings.

2.5.1 The Nascency Challenge

The '*data science*' research domain is in its pupal stage of (im)maturity and has yet to benefit from a universally accepted definition (Provost and Fawcett, 2013; Press, 2013a; Bloor, 2013; Ayankoya et al., 2014; Blei and Smyth, 2017). If commentators' arguments as to what constitutes the profile of a data scientist are taken as indicative of current views of what defines data science then it can fairly be considered a broad domain of data-related research and practice with roots in many camps [*Figure 2-9 &* Figure 2-10] (Granville, 2017; Castrounis, 2017; Costa and Santos, 2017).

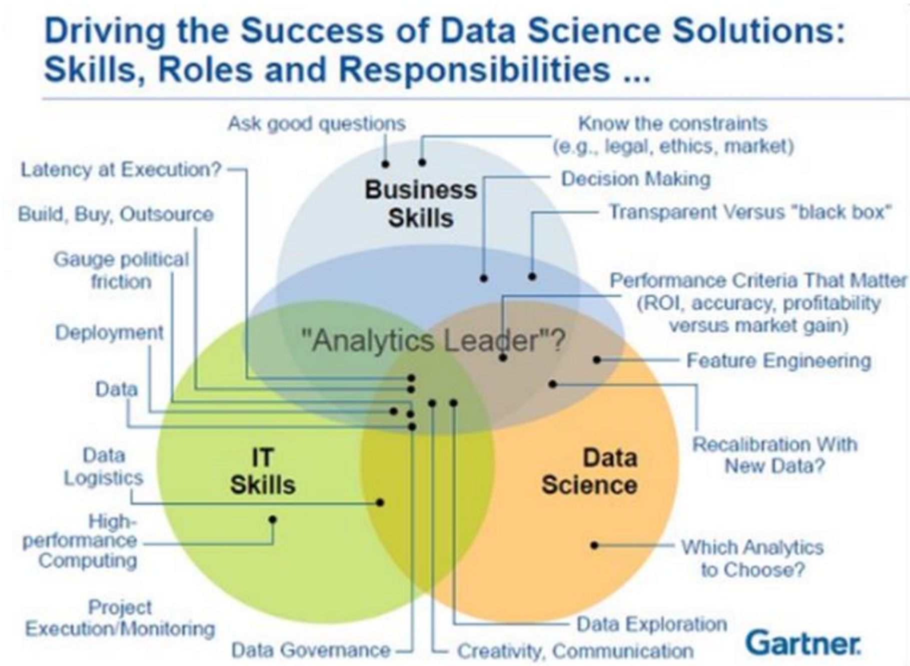


Figure 2-9: Gartner's Depiction of Data Science Skills, Roles and Responsibilities (Granville, 2017)

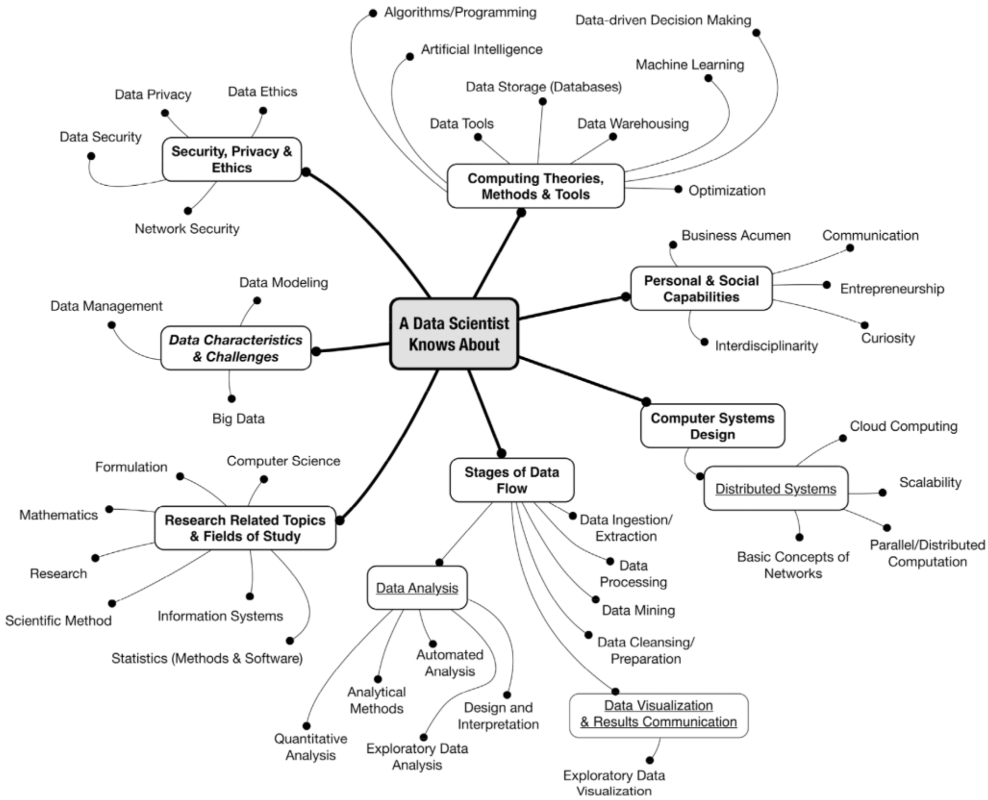


Figure 2-10: Conceptual Model for the Data Scientist profile (knowledge base) (Costa and Santos, 2017)

Hence, for the purposes of this research, data science is simply taken to be a multi-disciplinary research domain encompassing the study of all things '*data*' (Berman et al., 2018; Cao, 2017); similar to biological science being the study of all things living. In other words, data science is perceived here as a contemporary designation for an overarching domain of study that incorporates, and integrates established, as well as emerging, data-related disciplines.

Humanity has nurtured a keen interest in creating, collecting and benefiting from data for millennia. This interest can be tracked back to the earliest examples of tally marking (*some 30,000 years ago*) and census-taking (*over 6000 years ago*) (Bogoshi et al., 1987; ONS, 2013). Historical technological constraints to data generation, transmission and storage now swept away by the near ubiquitous application and use of internet and mobile technologies – rapidly moving towards global coverage –has resulted in previously inconceivable voice and data connectivity (Euromonitor, 2011). This in turn has fuelled an exponential rise in datafication and the proliferation of vast collections of data with seemingly endless possibilities for personal and corporate gain (Lycett, 2013; Cukier and Mayer-Schoenberger, 2013; Manyika et al., 2011).

It is the promise of immense oceans of data tantalisingly offering the prospect of unheard of treasures that has given rise to the concept, tools and technologies of Big Data (analytics) and the convergence and emergence of numerous data-related disciplines under the banner of data science (BBC, 2003; Press, 2013b, 2013c; Berman et al., 2018; Cao, 2017). Thus the materialisation and propagation of the term '*data science*' in the collective consciousness of commerce, research and academia closely coincides with the introduction and prolific use of the term '*Big Data*' (Press, 2013c, 2013b), which has also – to date – defied clarity and common understanding (Mikalef et al., 2017; Lemire and Lefebvre-Naré, 2017; Opentracker; Ward and Barker, 2013; Press, 2013b).

It is this nascency of the data science domain, in and of itself, that is perceived to be a challenge to this study. More specifically, it is conceivable that, as this work progresses, weaknesses and gaps in extant knowledge, tools and practices in the data science research domain will require the exploration of other more developed disciplines for possible solutions *or* the reactive creation of solutions that satisfy the immediate needs of this study. With this in mind, the approach taken by this work is to:

(a) Rely on the extant data science knowledge base first and foremost; *and*

(b) Close any progress-hindering gaps in the data science knowledge base by utilising relevant seminal work in cognate fields; *and finally, if this is insufficient…*

(c) Pragmatically create any necessitated artefacts to enable the study to move forward.

2.5.2 The Realisation of Value

Data science may be in its infancy, but it has certain considerations in common with other data-related disciplines, both mature and emerging, that are worthy of exploring. The most fundamental of these considerations is the realisation of value, or more specifically that the creation of value cannot be assumed simply because of the existence and availability of vast amounts of data (Mithas et al., 2013). A number of further considerations related to the challenges of creating value from data can be found in other related fields of study (Popovič et al., 2018; Huberty, 2015; Sharma et al., 2014; Lycett, 2013; Mithas et al., 2013; LaValle et al., 2011). Three of these *value-centric* considerations are discussed here as examples germane to this work:

(1) Understanding value and how to get it.

With a highly-hyped environmental back-drop of Big Data (analytics) it is not unusual for organisations embarking on data-driven initiatives to be unclear as to the value the effort is actually expected to yield and how to achieve it (LaValle et al., 2011; Mithas et al., 2013). For this research applying the question of value is not moot as, in general terms, the value of this study is ultimately assessed by its knowledge contributions to the humanitarian domain, and possibly the data science domain. In more data-specific terms, its value may be in creating an aerial view of web-available historical data (*the how*) that can offer insights into the aggregate effectiveness of humanitarian intervention, or the lack thereof (*the value*). Note here that insight is considered sufficient to argue value as this is a research endeavour. In organisations insight without effective and attributable action is of questionable value (Sharma et al., 2014).

(2) Remaining ‘as is’ is not conducive to value creation from analytics

Discussed in the context of Big Data and business analytics, there is a (mistaken) underlying assumption in organisations that gaining high-yield value from data can be achieved without changes to existing ways and means of working (Popovič et al., 2018; Huberty, 2015; Sharma et al., 2014). This observation is also relevant to data science and is likely to be illustrated by this study.

For example, it is realistic to accept that the web-available data that can be sourced for this work is unlikely to be perfectly suited to the purposes of the research. A rudimentary reason for this is that for all intents and purposes data collection and curation in the humanitarian domain is not designed to enable analysis of humanitarian intervention. For valuable and reliable data insights, and insight-attributable actions, the life-cycle of the needed data and resultant actions should ideally be designed, implemented and blended to enable a frictionless and auditable flow from data collection to insight-driven action. This is a radical aspiration for the humanitarian domain.

(3) Extracting value from data is interactive

The process of developing insights from data is an interactive engagement between the data and the actors who can elicit and/or act upon emerging insights (Sharma et al., 2014; Lycett, 2013). This is another relevant observation from business/data analytics that is pertinent to data science as a whole.

For this study the actualisation of this consideration is that after the data is prepared, no prescriptive or mechanical approach to exploring and exploiting the data is applied. Instead the process of eliciting knowledge is analogous to engaging in a dialogue with the data, where each exchange is designed improve familiarity and shape and direct the next questions to ask for further enlightenment.

Finally, to remain consistent with the DSR premise of evolving knowledge through the iterative creation of artefacts, further pertinent literature reviews in data science and cognate fields are carried out in situ within iterations that encounter data issues and/or create data science relevant artefacts.

2.6 Summary

This chapter sets the context for this research. It does this by providing a perspective of the humanitarian sector that highlights the progressively increasing demands and diminishing resources it faces in fulfilling its role (HERR, 2011). It also discusses some of the terms and concepts within the humanitarian system that are pertinent to this work. The chapter goes on to describe the need to gauge the effect of humanitarian intervention and identifies some of the challenges of developing such measures. Additionally, the chapter discusses the challenges experienced and approach taken to reviewing influential work within the data science domain that may yield useful and usable knowledge for this work. Finally, taking guidance from UN endorsed disaster risk reduction frameworks it proposes exploring macro-indicators based on aggregate disaster-related losses, aid flows and population factors as a means of obtaining some directional signs of progress of humanitarian intervention efforts (HYOGO, 2008; Wahlström, 2015).

Chapter 3: RESEARCH APPROACH

3.1 Overview

This chapter discusses the Design Science Research (DSR) approach employed for this study. It provides a brief description of 'wicked' problems and why the research problem addressed by this study should be considered wicked; DSR being particularly suited to tackling wicked problems. This work acknowledges that there are a variety of perspectives for this mode of research, which is likely a product of its relative immaturity (March and Smith, 1995; Venable, 2013; Gregor and Jones, 2007; Gregor and Hevner, 2013; Baskerville and Pries-Heje, 2010; Vaishnavi and Kuechler, 2004b). This multiplicity of views is addressed here by making explicit how DSR is applied by this research and the position taken on design theories.

The structure of this chapter is as follows: *Section 3.2* explains the concept of 'wicked' problems and outlines why the research problem motivating this study is 'wicked'. *Section 3.3* outlines the DSR approach as applied to this work taking cognisance of influential DSR papers in the field of Information and Communication Technologies (ICT) . *Section 3.4* describes the shape and flow of the research and sketches out the iterations of the design cycle planned for this study. Finally, *Section 3.5* summarises the salient points of the chapter.

3.2 'Wicked' Problems

'Wicked' problems are complex and highly resistant to resolution (APSC, 2007). These are problems for which knowledge of the problem and/or the solution is insufficient and the problem exists in an environment of many and diverse stakeholders with varying perspectives and agendas; with the 'wickedness' of problems amplifying as the diversity of the environment increases [*Figure 3-1*] (Head and Alford, 2008; Conklin, 2001).

Diversity → Complexity ↓	Single party	Multiple parties, each having only some of the relevant knowledge	Multiple parties, conflicting in values/interests
Both problem and solutions known (Heifetz Type 1)	Tame problem 1	2	3
Problem known, solution not known (relationship between cause and effect unclear) (Heifetz Type 2)	4	5	Wicked problem 6
Neither problem nor solution known (Heifetz Type 3)	7	Wicked problem 8	Very wicked problem 9

Figure 3-1: Typology of Problems
(Head and Alford, 2008)

Wicked problems were originally distinguished from 'tame' problems by way of ten defining characteristics (Rittel and Webber, 1973), but these are later rationalised to a set of six characteristics (Conklin, 2001). It is the six defining characteristics of wicked problems that are considered here, bearing in mind that not all characteristics need to be satisfied for a problem to be considered wicked and that there are degrees of wickedness [*Table 3-1*] (Conklin, 2001).

	Wicked Problems	**Tame Problems**
(1)	You don't understand the problem until you have developed a solution	Have well-defined and stable problem statements
(2)	Wicked problems have no stopping rule	Have has a definite stopping point, i.e. when the solution is reached
(3)	Solutions to wicked problems are not right or wrong	Have solutions that can be objectively evaluated as right or wrong
(4)	Every wicked problem is essentially unique and novel	Belong to a class of similar problems which are all solved in the same similar way
(5)	Every solution to a wicked problem is a 'one-shot operation'	Have solutions which can be easily tried and abandoned
(6)	Wicked problems have no given alternative solutions	Come with a limited set of alternative solutions.

Table 3-1: Six Criteria for Wicked and Tame Problems
(Conklin, 2001)

Consider the backdrop to this study [*Section 2.2*] as summarised and restated below:

As humanitarian funding becomes increasingly constrained and humanitarian needs continue to grow there is a pressing imperative to assess the impact of humanitarian aid so that more can be *achieved with less (Moorhead and Sandler Clarke, 2015; Belanger et al., 2016; HERR, 2011; van der Zee, 2015; Purvis, 2015).*

Now consider the six characteristics of wicked problems [*Table 3-1*], mapping these to what is known of the research problem and domain [*Table 3-2*] (Conklin, 2001):

<table>
<tr><td colspan="2">1. You don't understand the problem until you have developed a solution ✔</td></tr>
<tr><td>There are a myriad of stakeholders in the humanitarian space with equally as many perspectives. This may explain why there is no universal accepted definition of what constitutes a humanitarian disaster. Nor does there appear to be any definitive quantifiable criteria or target to assess what equates to progress and under what conditions. The closest that could be found were the targets in the "Sendai Framework for Disaster Risk Reduction 2015-2030" (Wahlström, 2015); but even this has the 'targets' stated in relatively vague and subjective terms, e.g. "<u>substantially reduce</u> global disaster mortality by 2030" or "<u>substantially enhance</u> international cooperation".</td><td><u>Solution Challenges:</u>
The data is an unknown. The full picture of what data is relevant, available, usable or useful is unknown
The value of the solution is unknown. Whether any solution developed will provide useful and usable insights is unknown.
Measures of progress are unknown. What would constitute an improvement is unknown.</td></tr>
<tr><td colspan="2">2. Wicked problems have no stopping rule ✔</td></tr>
<tr><td>It is not clear what 'good' looks like, therefore it is not clear how to gauge if 'good' is achieved and a solution is complete. The most that can be attempted is a journey of cyclical learning focussed on each cycle improving knowledge of the problem and its evolving solution.</td><td><u>Solution Challenges:</u>
There is no finite suite of data that limits the possibilities of the solution?
The evolutionary limits of the solutions cannot be defined.</td></tr>
<tr><td colspan="2">3. Solutions to wicked problems are not right or wrong ✔</td></tr>
<tr><td>There are no articulated conclusive goals or targets in terms of effectiveness. The process of learning through improvement can only strive for a 'better' solution as more knowledge is gathered, and not a final 'correct' one.</td><td><u>Solution Challenges:</u>
There is no target 'correctness' of a solution.
The design of each evolution of a solution is driven by the need to improve on its predecessors.</td></tr>
</table>

4. ***Every wicked problem is essentially unique and novel*** ✔	
There is no pre-existing solution that can be used to address the problem at the core of this study. As such the solutions will be novel and custom-built specifically for this research.	<u>Solution Challenges:</u> • There are **no industry models or solutions that can be adapted** for this problem.
5. ***Every solution to a wicked problem is a 'one-shot operation'*** ✔	
Each attempted solution will be examined and evaluated then either completely/partially discarded or evolved to deliver a better solution than the last. Each manifestations of the solution will be based on an increased understanding of the problem and solution.	<u>Solution Challenges:</u> • The **utility of the data cannot be assessed until it is used**. • The **value of insights cannot be predicted before they are revealed**.
6. ***Wicked problems have no given alternative solutions*** ✔	
There is no primary solution; therefore there is no alternative solution. Design choices on what to pursue and what to discard will be based on judgement and argued rationale.	<u>Solution Challenges:</u> • The **only way of knowing if any solution will yield meaningful results is to build the solution**.

Table 3-2: Research Problem – Six Criteria for Wicked Problems
(Conklin, 2001)

There is some consensus that attempts to solve wicked problems using top-down linear steps, such as the 'waterfall' – *gather*, *analyse*, *formulate* and *implement* – approach, are destined to fail (Rittel and Webber, 1973; Conklin, 2001). Instead "*opportunity-driven*" approaches are recommended (Conklin, 2001), with caveats against ignoring the "*polarity of design*" – i.e. '*what is needed*' versus '*what can be done*' – warning that the design of any solution must reconcile these two typically opposing forces (ibid). In this respect, Design Science Research (DSR) is appropriate as a research method, as this mode of research epitomises the '*opportunity-driven*' approach and places an emphasis on utility and '*satisficing*' – implicitly addressing the caution against ignoring the "*polarity of design*" (Vaishnavi, 2008; Simon, 1996; Conklin, 2001).

3.3 Design Science Research (DSR)

Design science research is a form of research that seeks to derive knowledge from the design and creation of artefacts (Vaishnavi and Kuechler, 2004b). New knowledge is sourced, not just from the process of designing and building artefacts but also from the analysis, reflection, abstraction, use, utility and value of what is created (ibid). DSR, as referred to earlier, is an ‘opportunity-driven’ approach (Conklin, 2001), of note is that it involves ‘sensing’ the way forward through repeated iterations of a *design-build-evaluate* cycle (Hevner et al., 2004). DSR accommodates the ‘wickedness’ of a problem by accepting that the outputs of each iteration can be less than optimal, as long as the direction of travel is towards improvement (Vaishnavi and Kuechler, 2004b).

The iterative development of artefacts in order to increase knowledge of problems and solutions is at the core of design science research (Hevner et al., 2004). Conclusion of the research is when a ‘good enough’, or ‘*satisficing*’, solution is reached (Vaishnavi, 2008; Simon, 1996). ‘*Satisficing*’ being a portmanteau word of ‘*satisfactory*’ and ‘*suffice*’, or as is more likely ‘*sufficing*’ (Simon, 1996; Reva, 2004). DSR facilitates ‘*learning through the act of building*’, allowing a solution to grow based on the development of new or fuller knowledge of the problem and solution space (Kuechler and Vaishnavi, 2008).

Although the DSR paradigm is well suited to this research problem, perceived weakness in DSR must also be acknowledged and addressed. DSR is a relatively young research method, which has yet to reach its full potential. There remains confusion surrounding its central ideas and contributions as well as continued debate over the need, type and relevance of design theory as an outcome of DSR (March and Smith, 1995; Hooker, 2004; Venable, 2013; Gregor and Jones, 2007; Gregor and Hevner, 2013).

This study addresses these areas of weakness by clearly describing the guidelines and framework of DSR adopted to shape, structure and complete this work and making explicit both the material and abstract artefacts created as output from this research (March and Smith, 1995; Vaishnavi and Kuechler, 2004b; Hevner, 2007; Venable, 2006a). Additionally it takes a utility theory approach to any nascent design theories that emerge from this work (Venable, 2013) .

3.3.1 A THREE Cycle View

DSR should not be viewed as an encapsulated research paradigm. In context it has an inherent relationship with two sets of activities outside the bounds of its *design-build-evaluate* iterations, as well as a symbiotic kinship with behavioural science research [*Figure 3-2*] (Vaishnavi, 2008; Hevner, 2007).

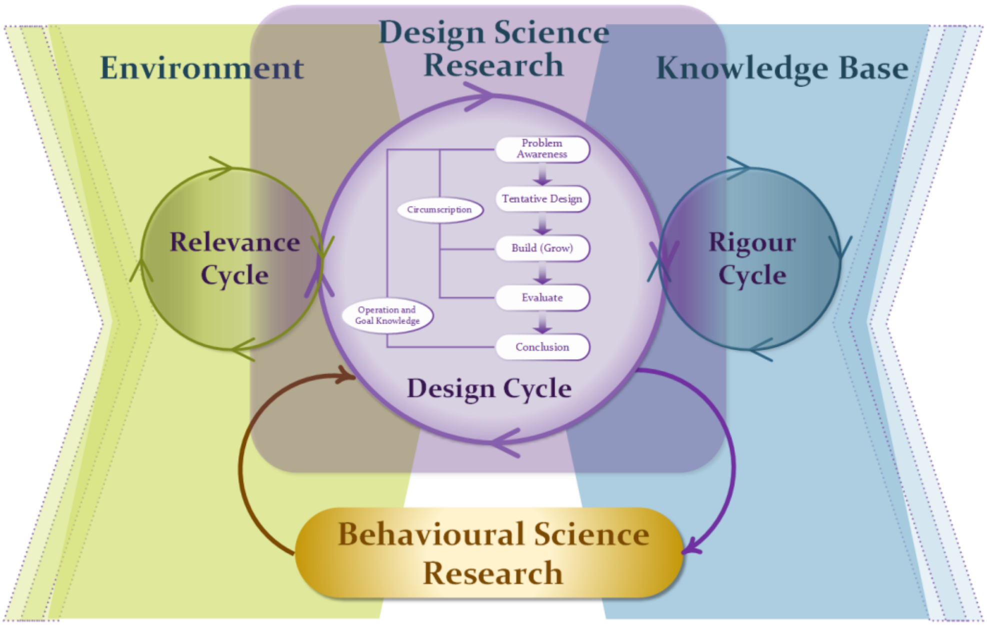


Figure 3-2: DSR in Context
Adapted from (Hevner, 2007; Vaishnavi, 2008; Hevner and Chatterjee, 2010)

A rudimentary description of this diagram is that it shows the design cycle of DSR and places it in context with its environment and knowledge base as well as its relationship with behavioural science research [*Figure 3-2*] (Hevner, 2007):

- The *Design Cycle* is at the core of this research approach. The process entails iterating through the *design-build-evaluate* loop until a satisficing solution is achieved (March and Smith, 1995; Vaishnavi, 2008; Simon, 1996).
- The *Relevance Cycle* provides context and requirements to the study and the environment in which to test the outputs of the research to ensure that the work remains germane to the domain in which it is situated (Hevner, 2007).
- The *Rigour Cycle* grounds the research by drawing from, and adding to, the resources, experience and expertise of existing knowledge (Hevner, 2007).
- The connections to *Behavioural Science Research* recognises the context and knowledge of behavioural (natural) science influences design science and the *utility* of design science on behavioural (natural) science (Hevner and Chatterjee, 2010).

Before discussing the *design cycle* as it is applied to this research it is useful to discuss the other major components of *Figure 3-2,* namely the *relevance cycle, rigour cycle* and *behavioural science research* and their relevance to this work.

The Relevance Cycle

This cycle initiates the research through identified need and ultimately provides the basis by which the research output is evaluated (Hevner, 2007). The 'need' that prompted this research is articulated in *Chapter 2*, but a very simple paraphrasing of this need could be – *to assess the consequences of humanitarian intervention in the aftermath of disasters, in order to identify areas and opportunities to minimise disaster outcome and maximise humanitarian aid impact*.

A key challenge in satisfying this need is that there is no quantitative data, measure or scale for '*minimise*' and '*maximise*'. In other words, there is no gauge for what is good, bad or inconsequential. Based on the premise that you can't improve what you can't measure, the research question underpinning this study (restated here) may offer a first (high-level) step in the direction of addressing this challenge. – *Can exploration of curated web-available data yield macro-indicators of humanitarian intervention in the aftermath of disasters?* – This question establishes the domain (environmental) relevance of this study and sets the scene for the DSR design cycle iterations employed for this research. The relevance cycle is also fundamental to the evaluation of the outputs of this work and assessing the contribution of knowledge to the research domain.

The Rigour Cycle

Three knowledge domains are pertinent to this work:

(1) Humanitarian Sector

(2) Design Science Research

(3) Data Science

(1) Humanitarian Sector

Knowledge of this domain is built from peer-reviewed papers, journal articles, books, news media, magazine articles, blogs, on-line discussion boards, summit/conference submissions, lecture videos and research theses. Sources include academia; iNGOs; NGOs; United Nations organisations; research institutes, multi-media news organisations; on-line blogs and expert discussion boards. The material read and evaluated far exceeds the material ultimately discussed, synthesised and cited in *Chapter 2*.

Notably, this knowledge gathering exercise did not identify a domain-specific methodology or widely accepted theory to inform or shape this study. In terms of knowledge contribution, the material artefacts created through this study hold the potential to increase knowledge in this domain and provide a foundation for future research.

(2) Design Science Research

Exemplar research in DSR within the component disciplines of ICT have helped shape and structure this study. More specifically, published guidelines for design science in information systems research are used as the basis of applying rigour to the research process [*Table 3-3*] (Hevner et al., 2004).

DSR Guidelines	Mapping
(1) Design as an Artefact *Design science research must produce a viable artefact in the form of a construct, a model, a method, or an instantiation*	Artefacts created by this study are cumulatively listed as research progresses through iterations of the Design Cycle. Created artefacts are mapped to the DSR framework (March and Smith, 1995).
(2) Problem Relevance *The objective of design science research is to develop technology-based solutions to important and relevant business problems*	*Chapter 2* provides the context and motivation for this work. *Section 1.2* lists what it intends to achieve and the earlier discussion of the *Relevance Cycle* in this section speaks directly to relevance of the study.
(3) Design Evaluation *The utility, quality, and efficacy of a design artefact must be rigorously demonstrated via well-executed evaluation methods*	Evaluation is by way of descriptive informed arguments based on the utility of created artefacts in answering the research problem (Hevner et al., 2004). Evaluation will take the form of ex ante artificial evaluation. Ex ante because the artefacts created are considered embryonic, created solely for formative purposes. Artificial because while the data is real, but the artefacts are not being evaluated by practitioners within the humanitarian space (Venable et al., 2012).
(4) Research Contribution *Effective design science research must provide clear and verifiable contributions in the areas of the design artefact, design foundations, and/or design methodologies*	The iterations of the design cycle create both material and abstract research artefacts, which through the process of evaluation verify the research contribution of this study.
(5) Research Rigour *Design science research relies upon the application of rigorous methods in both the construction and evaluation of the design artefact*	The rigour applied to this study is evidenced in its adherence to the process steps in the general methodology of DSR and the methods, experience and expertise used to inform and shape this work.

DSR Guidelines	Mapping
(6) Design as a Search Process *The search for an effective artefact requires utilizing available means to reach desired ends while satisfying laws in the problem environment*	This study plans three iterations of the design cycle to grow artefacts and repeatedly search for a solution to the research problem.
(7) Communication of Research *Design science research must be presented effectively to both technology-oriented and management-oriented audiences*	This study makes no assumption of prior knowledge in its audience. The study, its artefacts and the nascent theories are explained from first principles allowing it to be accessible to a more general audience.

Table 3-3: Mapping of DSR Guidelines
(Hevner and Chatterjee, 2010)

(3) Data Science

In the absence of a widely accepted definition of *Data Science,* in the context of this research it has simply been taken to mean the science and study of *all things data*, (Berman et al., 2018; Cao, 2017; Press, 2013a). This is regardless of the perspective, sub-categorisation or overlap with other disciplines – e.g. statistics; modelling; data technologies ('big' and regular), visualisation, analytics, mining, data quality and so on and so forth (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Laney, 2013; Grimes, 2013; Normandeau, 2013; Berti-Equille and Lamine Ba, 2016; Berti-Equille and Borge-Holthoefer, 2015; Lukoianova and Rubin, 2014; Claverie-Berge, 2012; Schroeck, 2012; Few, 2013; Spence, 2006; Cairo, 2012). While the analysis and visualisations employed by this research are relatively basic, this work draws from a broad pool of knowledge when working with data, selectively sourcing insights and guidance as needed. References are made where wisdom, expertise and experience from this knowledge base have specifically influenced or shaped the work. For the craft of data visualisation a more general approach is taken and the overall influence of Few, Spence and Cairo is recognised here (Few, 2013; Spence, 2006; Cairo, 2012).

Behavioural Science Research

Hevner et al argue that design science and behavioural science, a form of natural science, complement each other as technology and behaviour are inextricably intertwined (Hevner and Chatterjee, 2010). This assertion also holds for data and behaviour. Acknowledging DSR's focus on utility, a relationship with natural (behavioural) science, i.e. the activities of discovery and justification, remains (March and Smith, 1995), as neither design nor behavioural science can be considered in isolation [*Figure 3-3*] (Hevner and Chatterjee, 2010).

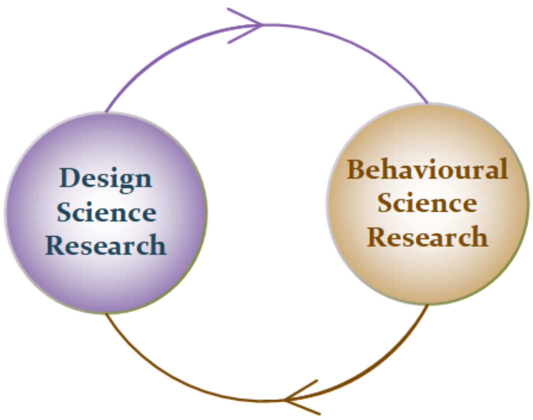


Figure 3-3: DSR & Behavioural Science Complementary Research Cycle
Adapted from (Hevner and Chatterjee, 2010)

The DSR paradigm for the development of design artefacts and theory provides the direction and shape of this research (Venable, 2013; March and Smith, 1995; Vaishnavi and Kuechler, 2004b; Hevner, 2007; Gregor and Jones, 2007). In developing artefacts, and more especially theory, the extrinsic teleology of design science is considered fundamental (Simon, 1996). Equally, the behavioural environment (research domain) informs this study and frames the purpose of this work.

3.3.2 The Design Cycle

The design science research paradigm is inherently pragmatic, it considers the practical consequences and the real effects of the research to be essential to meaning (March and Smith, 1995; Hevner, 2007). This essence of pragmatism also anchors the research to maintain its synergetic relationship with relevance and rigour in order to contribute outputs that offer both utility and research knowledge (Hevner, 2007).

This form of research embraces the view that it is fruitless to attempt to solve a problem we can't fully comprehend, e.g. a wicked problem, with a solution we can't fully imagine, and offers instead an exploratory '*doing to learn*' methodical approach to creating satisficing artefacts and contributing to the knowledge base. DSR's exploratory '*doing to learn*' approach, as adopted here, is portrayed by this design cycle schematic [*Figure 3-4*]. The diagram depicts the progression through one revolution of the design cycle and the cognitive processes associated with key steps.

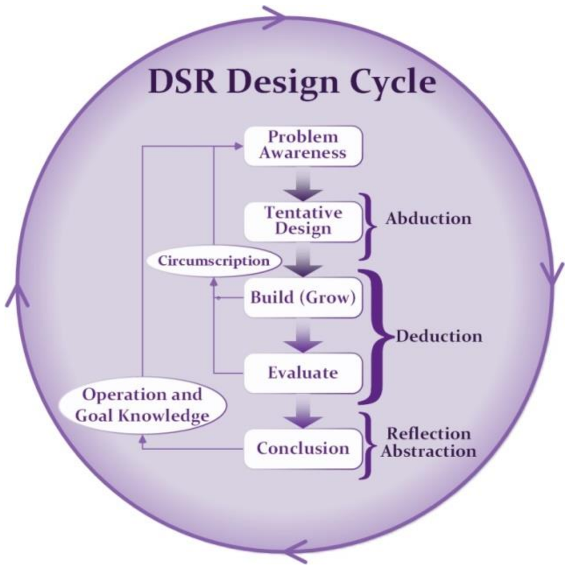


Figure 3-4: DSR Design Cycle
Adapted from (Vaishnavi and Kuechler, 2004b; Hevner, 2007)

NB. the steps in the general methodology of design science research have been relabelled here to echo the terminology used in this study (Vaishnavi and Kuechler, 2004b).

The steps within each design cycle are described here:

Problem Awareness

For the first entry into the Design Cycle (Iteration 1) *problem awareness* reflects knowledge gathered through the literature review [*Chapter 2*]. This knowledge, seeded by a foundational grasp of the research domain and the motivation to work toward addressing the identified need, initiates the research. New knowledge is gained while progressing through the iterations of the design cycle, and by way of evaluation and circumscription when completing the iterations; thereby improving awareness of the problem and the shortcomings of the most recently created solution to inform the next iteration, if there is one. Concluding the final iteration will include reflection and enriched knowledge of the problem and solution space to provide a springboard for future research.

It should be noted that circumscription is not a process step in the DSR framework, but entails the reasoned and common sense assumption that unless stated otherwise things are as expected (McCarthy, 1980). Here this means learning what does not work, *or what is missing*, and using this to improve problem awareness and thus help shape subsequent iterations. Notably reflection and circumscription also facilitate reasoned arguments for opportunities for further research.

Tentative Design

This is a data-centric study, therefore the tentative design is about selecting data to create analytics-ready datasets to satisfy the aim and objectives of the research (Vaishnavi and Kuechler, 2004b). For the first iteration the selection of data is framed primarily by awareness of the problem. For subsequent iterations the tentative design is shaped by an evaluation of progress towards the aim and research objective. Knowledge of the limitations and shortcomings of completed efforts, together with abductive reasoning of what could be used to address these limitations and shortcomings, form the basis of the design for each build of artefacts.

Build (Grow)

During this step the *tentative design* is realised, and proven, or not, as the case may be. For this study *build (grow)* includes all activities that establish the usability, utility and veracity of the data selected in the *tentative design*. This step also includes the analysis and visualisations that are carried out to examine the prepared data and explore it for macro-indicators of humanitarian intervention. It is during this step that the material and abstract artefacts that are outputs of this research are created (March and Smith, 1995; Vaishnavi and Kuechler, 2004b).

Evaluate

The method of evaluation for each revolution of the design cycle is by way of descriptive informed arguments based on knowledge of the research domain (Hevner et al., 2004). Deductive and logical reasoning is used to assess the efficacy and utility of created artefacts and takes cognisance of the aim, objectives and testable propositions of the utility theory of this research (Venable, 2013; Gregor, 2006). Even knowledge gathered through artificial (i.e. not 'real world') evaluation of artefacts created is expected to improve not only understanding of the solution, but also of the problem; thus shaping the next iteration of the design cycle (Venable et al., 2012; Gregor and Jones, 2007; Vaishnavi and Kuechler, 2004b).

Conclusion

This is the last step of the final revolution of the design cycle. This step includes reflection and interpretation of the findings and, in line with the DSR methodology, the outcome from this iteration is expected to be '*good enough*' to complete the study and finalise the knowledge contribution of this research (March and Smith, 1995; Simon, 1996; Gregor and Jones, 2007).

3.3.3 DSR Artefacts

This section brings the focus of the chapter to the DSR artefacts of *constructs*, *models*, *methods* and *instantiations* (March and Smith, 1995; Vaishnavi and Kuechler, 2004a, 2004b; Venable, 2015; Gregor and Jones, 2007), which are the four types of artefacts most commonly and consistently cited as outputs of DSR [*Table 3-4*] :

Constructs	The conceptual vocabulary that forms the specialised language of the problem/solution space.
Models	Propositions or statements that express relationships between constructs; proposals for how things are or should be.
Methods	Steps (algorithms or guidelines) used to perform tasks.
Instantiations	Working artefact made real, sometimes even before the other types of artefacts, to provide a proof of concept.

Table 3-4: The 4 Types of DSR Output
(March and Smith, 1995; Vaishnavi and Kuechler, 2004b):

Additionally *instantiations* can be described as *material* artefacts, i.e. 'physical' manifestations that emerge from the design, while *constructs*, *models* and *methods* can be described as *abstract* artefacts, i.e. concepts, representations and processes that emerge from the design (Gregor and Jones, 2007).

Furthermore, the types of artefacts, as well as the premise of *material* and *abstract* artefacts, are interpreted and adjusted to accommodate this work, which is a *study of data* and not technology (IT/IS). With this in mind *Table 3-5* maps the planned DSR outputs from this study to March and Smith's Research Framework (March and Smith, 1995) and relates these to the *build (grow) – evaluate* steps of the (Vaishnavi and Kuechler, 2004b; Hevner, 2007).

Research Framework			Research Activities			
			Design Science		**Natural Science**	
			Build (grow)	Evaluate	Theorise	Justify
Research Outputs	Artefacts	**Constructs**	[a] ⟶		[c] - - - ->	
		Models				
		Methods				
		Instantiations	[b] ⟶			
[a]	Macro-indicators of disaster outcome and the impact and effectiveness of humanitarian intervention (MiOs, MiIs and MiEs).					
[b]	Data analysis outputs and visualisations.					
[c]	A (behavioural science) hypothesis relating the availability, or lack thereof, of humanitarian data and the flow of humanitarian aid that emerges from the domain knowledge and may be worthy of future research.					

Table 3-5: DSR Output to Research Framework Mapping v.0
(Vaishnavi and Kuechler, 2004b; Hevner, 2007; March and Smith, 1995)

The constructs and instantiations (ref: **[a] & [b]**) identified here are the bare minimum research artefacts expected from this work based on the aim and objectives of the research. Also included here, in order to ensure it is recorded, is a behavioural (natural) science hypothesis (ref: **[c]**). Hypothesis [c] posits that the practice of data collection in humanitarian supply networks (HSNs) is an '*unnatural act*' and as such is left undone or badly done. Unlike commercial supply chains, where integrated transactional activities collects vast amount of reconcilable data as a by-product of doing business, in HSNs data collection is disconnected from the flow of funds, goods and services. Securing donor funds to fuel the HSNs is not contingent on measures of matching real supply to real demand. It is believed here that **[c]**, which falls under the scope of behavioural (natural) science research, may lead to studies of the characteristics of *humanitarian supply networks (HSNs)* at the root of some of the issues that have necessitated and shaped this research (March and Smith, 1995; Hevner and Chatterjee, 2010).

3.3.4 Design Theory

Seminal DSR authors distinguish natural (behavioural) science and design science as the former being concerned with the way things are and the latter being concerned with the way things ought to be (Simon, 1996). DSR commentators consider theories in natural (behavioural) sciences to be "*deep*" and *"principled explanations"* of "*phenomena*", i.e. '*what is*'. Whereas in design science, which is about the design of artefacts "*to attain goals*", i.e. '*what has yet to be*', a design theory is argued unnecessary by some and of a variety of types by others (Simon, 1996; March and Smith, 1995; Hevner et al., 2004; Hooker, 2004; Venable, 2015, 2013, 2006a, 2006b; Gregor and Jones, 2007; Vaishnavi and Kuechler, 2004b; Pries-Heje and Baskerville, 2008; Baskerville and Pries-Heje, 2010; Walls et al., 1992; Nunamaker et al., 1990).

This work does not subscribe to the view that design theories are unnecessary (March and Smith, 1995; Simon, 1996; Hooker, 2004); instead the position taken here is that design theory in DSR is an overarching *seeding inspiration*, *evaluatory context* and *intrinsic contribution* (Venable, 2006b). Design theories provide cohesion and relevance to the design effort, mooring created artefacts to the research domain, providing a context for evaluation, and ultimately offering a means of summarily articulating the knowledge contribution of the work (Venable, 2006b). As a design theory is to be a part of this work, a clear stance is also taken as to the nature of a design theory – *explanatory*, *prescriptive* or *utility* (March and Smith, 1995; Hooker, 2004; Hevner et al., 2004; Vaishnavi and Kuechler, 2004b; Baskerville and Pries-Heje, 2010; Walls et al., 1992; Gregor and Jones, 2007; Venable, 2006b; Nunamaker et al., 1990). All three perspectives of design theories are discussed here, but the argument that design theories are ***utility theories*** is considered to be the most cogent.

Design Theories are Explanatory

Design theories of the explanatory type are said to explain "*why a component is being constructed into an artefact*" (Baskerville and Pries-Heje, 2010). This perspective of explanatory design theories also describes the relationship between the general requirements and the general components of the design research as follows: *"The definitions of general requirements and general components must be circular. Requirements specify (and explain) the reasons for components. Components are justified by requirements"* (ibid). This assertion, however, about circular definitions does not tally with the description of circular definitions provided as the quoted specification can more straightforwardly be stated as "*Requirements specify (and explain) the reasons for components"* <u>and justify them</u> – without losing meaning. Rewritten it simply describes a one-way relationship of what *requirement*s do for *components*. A didactic explanation of what requirements do for components is considered inadequate to the task of reflecting the '*improvement*' offered by an "*improvement research*" paradigm, aka DSR (Venable, 2006b, 2013; Gregor, 2006).

Moreover, explanatory theories are described as offering ex poste explanations that make testable propositions infeasible (Venable, 2013; Gregor, 2006). This is in contrast to the needs of this study as testing the propositions of the design theory is viewed to be part of the evaluation step in the design cycle. Consequently, this form of theory is not adopted for this study.

Design Theories are Prescriptive

Design theories of a prescriptive nature are about rules and guidelines, expressed in terms of '*can*' and '*will*' that, if followed, will result in the desired outcome (Walls et al., 1992; Gregor and Jones, 2007; Venable, 2013). While the use of '*can*' dilutes the implied mandatory nature of something that is referred to as *prescriptive*, a prescriptive design theory nevertheless suggests an enforced imposition of a definitive solution (Venable, 2013). The feasibility of this is, at the very least, doubtful for research problems that best suit the DSR paradigm (ibid). There is cognitive dissonance in considering a definitive set of rules or guidelines, i.e. prescriptive theories, to be the result of a methodology that advocates the opportunistic search for '*good enough*' and '*satisficing*' outcomes (Simon, 1996; Vaishnavi, 2008; Venable, 2013). To explain, a theory that is definitive enough to be imposed as a result of a DSR study implies the theory is the *best* found, i.e. others were evaluated and found lacking. Apart from the infeasibility of finding the *best* in a solution space where neither the problem space or solution space are fully understood, there is the jarring incongruity of the absolute of an implied 'best' emerging from a paradigm in which merit is measured relatively in terms of 'better' or 'worse' (Venable, 2013; Simon, 1996; Conklin, 2001 ; Vaishnavi, 2008). Therefore the notion that design theories should be prescriptive is also set aside.

Design Theories are Utility Theories

Utility theories are described as theories that define the relationship between the solution space and problem space in terms of utility, asserting that a particular type or class of solution improves a particular type or class of problematic situation (Venable, 2006b). This study takes the stance that design theories should be in the form of utility theories because this form of theory:

- is an adaptive match to the peculiarities of artificial phenomena as opposed to natural phenomena (Simon, 1996; Venable, 2013);
- incorporates pertinent definition of the problem and solution space to provide contextual relevance (Venable, 2006b);
- embeds improvement towards goals and in so doing emphasises the teleological aspect of the work (Venable, 2006b).

Furthermore, it is argued here that there are two areas of distinction between natural science research and design science research that are important when considering the nature of design theory. First, in natural science the phenomena made intelligible through theory are not man-made, whereas in design science the phenomena made intelligible through theory are anchored to purposefully man-made artefacts (Venable, 2013; March and Smith, 1995). It therefore stands to reason that the intelligibility of design science phenomena through theory is incomplete without an articulation of the utilitarian purpose of the artefacts created through the research (Venable, 2006b, 2013; Simon, 1996).

Second, natural science theories are expected to be timeless, replaced only through better research, whereas design science theories are expected to become obsolete as the man-made reality for which they are created progresses beyond their utility (Venable, 2013). Hence, utility theories frame utility in the context of the type of problem and solution addressed, thus admitting their relevance may not outlive changes in the problem space and advancement in the solution space.

The Utility Theory of this study

Utility theories assert that a particular type or class of solution will improve a particular type or class of problematic situation (Venable, 2006b). They link concepts in the solution space to the issues they address in the problem space and form the basis of evaluating the artefacts created [*Figure 3-5*] (*ibid*).

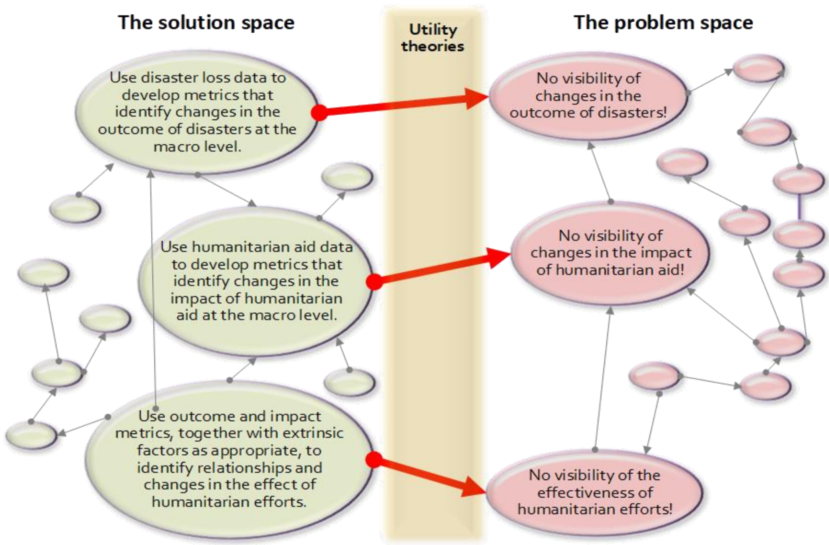


Figure 3-5: Utility Theories – Solution & Problem Space (Venable, 2006b)

The form of utility theory best suited to this work can be constructed to argue *'solution **x** (when applied properly) will be efficacious in improving problem **y**'* (Venable, 2006b). Thus the statement of *utility* can be written as [*Figure 3-6*]:

Statement of Utility

Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector, when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level, the inability to gauge the consequences of monies spent, and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.

Figure 3-6: Statement of the Utility Theory

Note that the structure of this statement reflects the prototypical structure of utility theories suggested by Venable [*Table 3-6*] (Venable, 2006b):

	Solution Space	Utility	Problem Space
	Form	Function	Purpose
	Artefact [*What*]	Efficacy [*How*]	to Address [*Why*]
STATEMENT	*Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector…*	*…when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level…*	*…the inability to gauge the consequences of monies spent and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.*

Table 3-6: Structure of the Utility Theory Statement

3.3.5 DSR and the Study of Data

As outlined in *Section 3.2* the research problem addressed by this work can reasonably be considered wicked (Rittel and Webber, 1973). DSR's modus operandi of iteratively creating progressively improving artefacts – i.e. *constructs*, *models*, *methods* and/or *instantiations* – embodies the '*opportunity-driven*' approach recommended to solve wicked problems (Conklin, 2001 ; March and Smith, 1995; Kuechler and Vaishnavi, 2008; Hevner et al., 2004). However, in choosing DSR as the research methodology for this study it is important to address three potential misconceptions:

1. *DSR is exclusively used for IT and IS research*

Conspicuously, the richest seams of influential design science paradigmatic research can be found in the disciplines of information system (IS) (Baskerville and Pries-Heje, 2010) and information technology (IT) (March and Smith, 1995; Venable, 2013, 2006a; Venable et al., 2012, 2016; Venable, 2015; Gregor and Jones, 2007; Vaishnavi and Kuechler, 2004b; Kuechler and Vaishnavi, 2008; Walls et al., 2004; Hevner et al., 2004; Hevner, 2007). This does not however mean that the DSR mode of research is exclusively applicable to the creation of knowledge and artefacts in the fields of IS and IT – only that researchers and practitioners in IS and IT are its most prominent adopters and commentators.

It is argued here that the DSR approach of 'sensing' the way forward through repeated iterations of a *design-build-evaluate* cycle is equally suited to addressing wicked problems in other fields of study (Hevner et al., 2004; Rittel and Webber, 1973). For example, research into reforestation as a countermeasure to the wicked problem of climate change could feasibly be completed using the DSR approach. In that the overall methodology remains structurally applicable and only the outputs (*constructs*, *models*, *methods* and *instantiations*) need reflect the field and focus of the research (Rayner, 2006; Boucher, 2012). Recognising this, as yet largely untapped, broader applicability of DSR this research applies this methodology to the study of data relevant to humanitarian intervention.

2. *DSR will need to be adapted to create 'data' artefacts*

The interests of this work lie in the capacity of curated web-accessible data to provide new perspectives and knowledge by way of a '*datafied*' window into the humanitarian domain (Lycett, 2013; Normann, 2001). The information systems and technologies used to capture, hold, manage, transmit, or even analyse and visualise the data, are simply a means to an end within the bounds of this research. It does not hold direct relevance if the data used by this study is primarily captured and managed on paper or via state-of-the-art ICT as long as the data is ultimately accessible via the web, is from a credible source, has some veracity and holds value through utility. Similarly, the systems and technology used in the data preparation and analysis carried out for this study are not for the most part pertinent, unless they affect the data's utility, nor are they in and of themselves a research output. That said, recommendations for future work or changes to process or policy from this work may have IT or IS implications insofar as their potential to yield datasets with greater utility in the future.

The DSR approach has not been applied here by rote to create the type of artefacts that are typically exampled in IS/IT research papers (Peffers et al., 2007; Hevner et al., 2004). Instead, without modification to the principles of DSR, all research outputs (material and abstract) from this work are data or data-related. The general terms of *abstract* artefact and *material* artefact have been applied here is as follows (Gregor and Jones, 2007): abstract artefacts include constructs, models, methods and theories; material artefacts include datasets and metadata instantiations that can have application beyond this study, or data visualisations/analytical models that can inform others for different purposes.

3. *DSR is used in place of Data Science Methodologies*

There are two aspects of the relationship between the use of DSR for this research and data science methodologies worth clarifying:

(a) DSR is an overarching research approach

DSR provides structure, process, coherence and focus to the overall research and is distinct from any tools, techniques or methods utilised during the course of the work. As an example: the first iteration of a data-driven DSR study may employ statistical methods to improve knowledge, the next iteration may go on to use data-mining techniques to build on that knowledge – neither the statistical methods nor the data-mining techniques violate or complicate the use of DSR as a research methodology.

(b) What data science methodologies?

There are numerous techniques for extracting knowledge from data relevant to the data science domain (Granville, 2016), however the domain, and associated disciplines, are not awash with broadly applicable frameworks and methodologies. The three most commonly referred to are:

- Knowledge Discovery in Databases (KDD) process for data mining (Fayyad et al., 1996).
- Cross-Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al., 2000).
- Sample, Explore, Modify, Model, Assess (SEMMA) data mining process (SAS, 2014).

Interestingly, these methodologies are primarily for data mining and are also not universally used (Piatetsky, 2014).

In conclusion, DSR is an approach that easily lends itself to data-centric research. The fundamental methodology and outputs can remain true to their fundamental form and framework, only the specifics of the outputs change from the IS/IT examples typically detailed in seminal literature (March and Smith, 1995; Vaishnavi, 2008). Instead when using DSR as an approach for data-centric research the *constructs*, *models*, *methods* or *instantiations* created are data or data-related artefacts. Furthermore, the use of DSR as an overarching approach to research does not preclude the use of more specific data manipulation and/or analytical methods and techniques.

3.4 The Research Structure

This section briefly outlines the shape and structure of this research, providing an overview of the four iterations of the DSR design cycle ultimately completed [*Figure 3-7* & Figure 3-8].

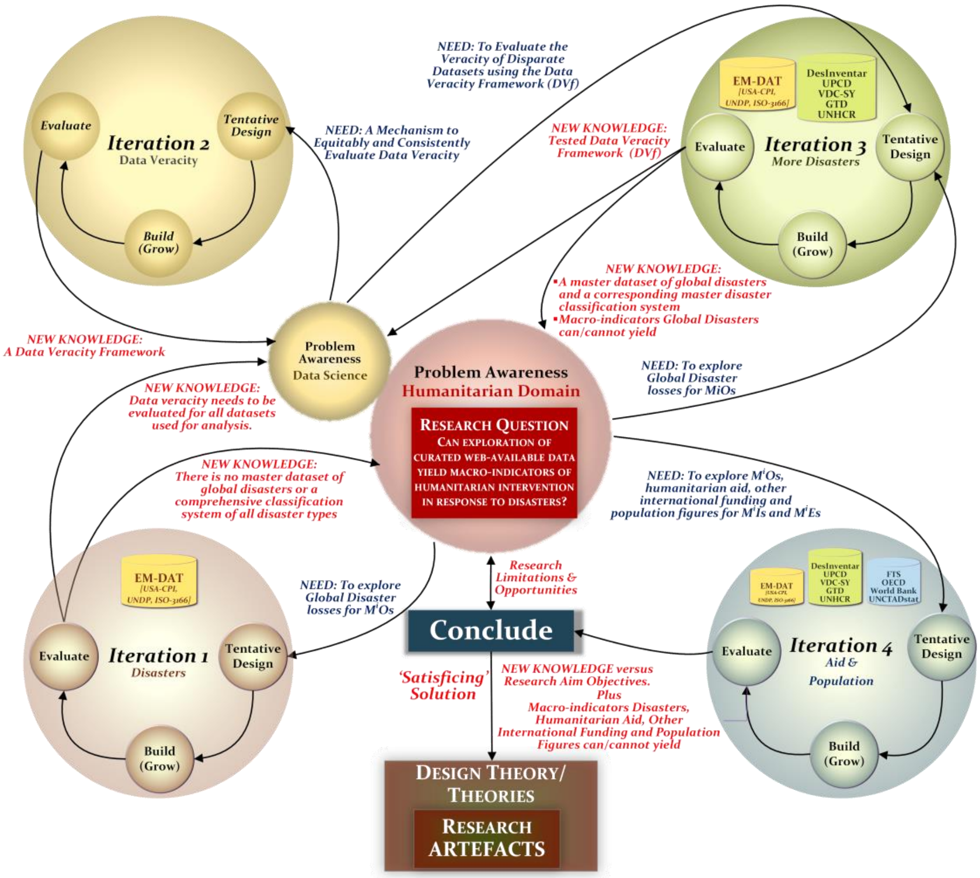


Figure 3-7: The 4 iterations of the DSR design cycle as revised after Iteration 1

Iterations

1 **Explore Disaster Losses for Macro-indicators of Outcome (M^iOs)**

Acquire, prepare examine and explore EM-DAT datasets for MiOs.

EM-DAT disaster data and classification are found to be incomplete and exhibit issues of veracity. Reshape the remainder of the study to acquire and amalgamate other disaster datasets and create a classification system to support the amalgamated dataset. Address the issues of veracity that may also be present in other datasets to be used, by obtaining or building a toolset to evaluate data veracity as a prerequisite to utilising the additional datasets.

↓

2 **Develop a Mechanism to Equitably and Consistently Evaluate the Veracity of all utilised datasets**

As no data veracity toolset is found, one is built in the form of a Data Veracity framework (DVf)

↓

3 **Create an amalgamated Master Set of Global Disasters, a Master Disaster Classification system, then explore the amalgamated Disaster Losses for Macro-indicators of Outcome (M^iOs)**

In addition to EM-DAT, five disaster loss datasets are acquired, prepared, examined, evaluated for veracity and amalgamated to create a Master Set of Global Disasters (MSGD). A Master Disaster Classification (MDC) model and reference dataset are created to support the MSGD. Finally, the **M^iO** of *mean survival rate by year* is identified

↓

4 **Explore M^iOs, humanitarian aid, other international funding and population figures for Macro-indicators of Impact and Effect (M^iIs & M^iEs).**

Humanitarian aid and Other International Funding flow datasets are acquired, examined and evaluated for veracity. Population growth values are also obtained. The **M^iI** of *mean survival rate by humanitarian aid per person* and the **M^iE** of *mean survival rate by population* are identified.

Figure 3-8: The Narrative Flow of the DSR Iterations

3.5 Summary

This chapter provides an overview of the design science research (DSR) approach employed for this study. It describes the concept of ‘wicked’ problems, explaining why the problem addressed by this study is considered ‘wicked’. The chapter goes on to assert the suitability of the design science research paradigm to tackle ‘wicked’ problems in general, and this research problem in particular.

The chapter then details the context and steps of the DSR design cycle and discusses the research artefacts created. It also explains the utility theory approach taken for the design theories presented by this study. Additionally, clarification is provided that the academic papers referenced in support of the DSR methodology are from the field of IT/IS, where software or software-related artefacts are more likely outputs of the research, however this research has adopted the methodology for the study of data to create data and data-related artefacts as outputs. Finally, it provides an outline of the flow of the DSR design cycle iterations used to complete this research.

Chapter 4: DISASTERS (ITERATION 1)

4.1 Overview

This chapter describes and discusses Iteration 1 of the DSR design cycle for this research. The focus of Iteration 1 is to use a credible disaster loss dataset, identified as the Emergency Events Database (EM-DAT), to search for macro-indicators of the outcome (M^iOs) of humanitarian disasters (Guha-Sapir et al., 2017l). In other words, explore disaster losses for indicators that may signal patterns of change in disaster outcomes. *Figure 4-1* is a basic schematic of the flow of Iteration 1.

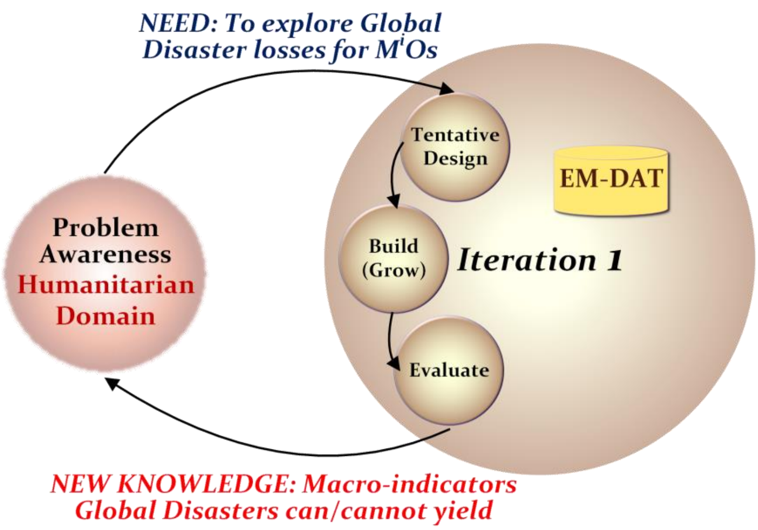


Figure 4-1: Iteration 1 of the Design Cycle

The structure of the chapter is as follows: *Section 4.2* provides an overview of the need (problem), summarising and recapping the points discussed in *Chapter 2*. *Section 4.3* discusses the *Tentative Design*, which essentially centres on the selection of EM-DAT as a single source of disaster loss data. *Section 4.4* describes the *Build (Grow)* step, which includes preparing, examining and working with the EM-DAT dataset. *Section 4.5 Evaluates* the work and outputs of Iteration 2 before *Section 4.6* closes with a summary of the chapter.

4.2 Problem Awareness

For this first iteration of the design cycle, knowledge of the problem is primarily based on the literature review as described in *Chapter 2*. Salient aspects of the problem are restated here as a reminder.

As humanitarian funding becomes increasingly constrained and humanitarian needs continue to grow there is a growing imperative to assess the effectiveness of humanitarian intervention in order to achieve more with less (Moorhead and Sandler Clarke, 2015; Belanger et al., 2016; HERR, 2011; van der Zee, 2015; Purvis, 2015; Lattimer et al., 2016). There are, however, a number of challenges to assessing the effectiveness of humanitarian intervention including:

- Lack of detailed data that can be used to trace disaster zone (real) needs to the subsequent fulfilment of those needs (Tatham and Hughes, 2011; Maiers et al., 2005; Riddell, 2008).
- Disaster response evaluations, where they exist, are typically self-certifying reports by aid agencies or sporadic victim perception surveys (The Fritz Insitute, 2007; ALNAP, 2017; CHS Alliance, 2017). Unfortunately, even these initiatives are not applied comprehensively or consistently enough to facilitate the creation of quantifiable comparable metrics of the effectiveness of humanitarian intervention efforts.
- The lack of visibility as to how, where and during which phase of the disaster management cycle, humanitarian aid is actually spent (Buchanan-Smith and Fabbri, 2005; Riddell, 2008, 2014b; Scholten et al., 2010).
- Any measure of post-disaster response effectiveness will also implicitly reflect other steps in the disaster management cycle, such as preparedness and mitigation (Safran, 2004; Cozzolino, 2012; Riddell, 2014a).

4.3 Tentative Design

The tentative design of the first iteration of the design cycle focusses on exploring web-available curated disaster loss data to identify indicators that may exhibit changes in the aggregate outcome of disasters. At the core of this iteration is a single disaster loss dataset, the Emergency Events Database (EM-DAT) managed by the Centre for Research on the Epidemiology of Disasters (CRED) (Guha-Sapir et al., 2017l; CRED, 2017).

EM-DAT is the only dataset selected because of a reasoned expectation that it is comprehensive enough to provide sufficient view of all types of disasters and the losses caused by these disasters. This expectation is based on explicit and implicit information from the EM-DAT site that suggests the database contains records of all disaster types and disaster-related losses from 1900 to present time (Guha-Sapir et al., 2017i, 2017a). Furthermore, the veracity of the dataset is assumed from the level of credence placed in its data by research studies and institutions (Voigt et al., 2016; Guha-Sapir et al., 2017j; Toya and Skidmore, 2007; Kourosh and Richard, 2008; Strmberg, 2007; Alcántara-Ayala, 2002; Pears-Piggott and Muir-Wood, 2016; Blaikie et al., 2014; Sodhi, 2016; Corey et al., 2016; Raschky and Schwindt, 2016).

This iteration therefore explores for macro-indicators of outcome (M^{i}Os) by concentrating on EM-DAT as the sole source of disaster data, initially considering it to be comprehensive, credible, meticulously curated and accessible account of disaster losses (Guha-Sapir et al., 2017l). The search for M^{i}Os is guided by the type of disaster loss indicators and targets suggested by the two UN endorsed disaster risk reduction frameworks of [*Figure 4-2*] (HYOGO, 2008; Wahlström, 2015).

(1) Number of deaths arising from natural hazard events
(2) Total economic losses attributed to natural hazard events
(3) Number of people affected by natural hazard events

Hyogo Framework for Action (HFA) 2005–2015 (HYOGO, 2008)

(a) Substantially reduce global disaster mortality by 2030, aiming to lower the average per 100,000 global mortality rate in the decade 2020–2030 compared to the period 2005– 2015;
(b) Substantially reduce the number of affected people globally by 2030, aiming to lower the average global figure per 100,000 in the decade 2020–2030 compared to the period 2005–2015;
(c) Reduce direct disaster economic loss in relation to global gross domestic product (GDP) by 2030;

Sendai Framework for Disaster Risk Reduction (SFDRR) 2015-2030 (Wahlström, 2015)

Figure 4-2: Indicators/Targets from UN endorsed DRRs

Note that no prescriptive definition of what may or may not constitute M^iOs is provided in the design step as M^iOs are expected to emerge from improved knowledge of the data.

4.4 Build (Grow)

This section describes the acquisition, preparation, examination and exploration of the EM-DAT disaster loss dataset (Guha-Sapir et al., 2017g). It outlines the pre-acquisition understanding and assumptions about the data as well as the knowledge developed of the data through acquiring, preparing, examining and exploring it. Also described is the use of auxiliary datasets to support, verify or supplement EM-DAT.

4.4.1 EM-DAT

A brief overview of EM-DAT from the site is as follows (Guha-Sapir et al., 2017a): "*In addition to providing information on the* ***human impact*** *of disasters - such as the number of people killed, injured or affected – EM-DAT provides disaster-related* ***economic damage estimates*** *and disaster-specific* ***international aid contributions***" (Guha-Sapir et al., 2017a). Foundational, pre-acquisition, information about the dataset include:

- EM-DAT is a long-standing (established in 1988) source of disaster loss information; it holds data for disasters going as far back as 1900 (Guha-Sapir et al., 2017l).

- EM-DAT data is sourced from some 27 named international entities including governments, UN agencies, the World Bank, major re-insurance companies, international NGOs, research institutes and press agencies (Guha-Sapir et al., 2017k).
- The type of data held includes: country of occurrence; temporal information; and human and financial effects (Guha-Sapir et al., 2017e, 2017g; Below et al., 2009; IRDR, 2014).
- To be included in EM-DAT database an event must meet one of four criteria [*Table 4-1*] (Guha-Sapir et al., 2017e):

Ten (10) or more people reported killed	Declaration of a state of emergency
Hundred (100) or more people reported affected	Call for international assistance

Table 4-1: EM-DAT Inclusion Criteria
(Guha-Sapir et al., 2017e)

- Disasters are classified according to EM-DAT's classification systems, which is based on UNISDR's Peril Classification and Hazard Glossary (Guha-Sapir et al., 2017g; IRDR, 2014; IRDR, 2017; UNISDR, 2017).
- Disaster entries are uniquely identified by a disaster number and country of occurrence (Guha-Sapir et al., 2017i).

Assumptions of data completeness, accessibility and veracity include:

- EM-DAT holds data for all disaster types as its inclusion criteria does not exclude disasters by *type* (Guha-Sapir et al., 2017e).
- EM-DAT is easily accessed and acquired because it is curated for the primary purpose of aiding analysis and decision-making (Guha-Sapir et al., 2017a).
- EM-DAT is veracious as its meticulously maintained analytical data is cited by credible institutions and publications – reasoning:

    (1) EM-DAT's data update processes indicate diligence – "*entries are constantly reviewed for inconsistencies, redundancy, and incompleteness*" (Guha-Sapir et al., 2017f). Additionally, data is said to be updated and consolidated by CRED daily, a process that is supplemented by monthly checks and annual

revisions (Guha-Sapir and Hoyois, 2012; Guha-Sapir et al., 2017f; CRED, 2017).

(2) EM-DAT is centrally maintained for analytical purposes (Guha-Sapir et al., 2017a), it is therefore reasonable to assume it is less likely to suffer from unresolved issues that are more typical when data is a *by-product* of transactions (Guha-Sapir et al., 2017e, 2017g, 2017i; Below et al., 2009; IRDR, 2014).

(3) EM-DAT is frequently used in disaster research and it is also suggested as a data source in the disaster risk reduction Hyogo Framework for Action (Voigt et al., 2016; Guha-Sapir et al., 2017f; Toya and Skidmore, 2007; Kourosh and Richard, 2008; Strmberg, 2007; Alcántara-Ayala, 2002; Pears-Piggott and Muir-Wood, 2016; Blaikie et al., 2014; Sodhi, 2016; Corey et al., 2016; Raschky and Schwindt, 2016; HYOGO, 2008).

This baseline view of EM-DAT is now tested by the experience of acquiring and working with its data.

Acquiring the Data

Attempts to acquire EM-DAT data very quickly dispels any illusion of straightforward data access, which was based on the database's stated objective to facilitate decision-making. No method, automated or 'on request', is found to obtain a full set of EM-DAT data. There is even a message on the site stating – "*Requests of the entire set will not be treated*" (Guha-Sapir et al., 2017c). Alternative methods are therefore needed to obtain as large a set of EM-DAT data as possible.

(c) Three EM-DAT Extracts

Two historic extracts of EM-DAT were obtained from an R archive (Goteti, 2016). One extract is from 2014 and contains data spanning 1900–2013; the other is from 2015 and contains data spanning 1900–2014. Both datasets contained only 14 of the 53 variables documented as held on the EM-DAT database (Guha-Sapir et al., 2017i).

As 2015 disasters are missing from the EM-DAT R archives, a third dataset, spanning 1900–2015, is constructed in 2016 via screen scrapes from the Disaster List page of the EM-DAT site (Guha-Sapir et al., 2017c). The limitations of the screen display means that once again only 14 variables are captured, but not the same 14 as the previously acquired EM-DAT R archives (Guha-Sapir et al., 2017c). These three versions of EM-DAT once merged de-duplicated and rationalised result in a final consolidated EM-DAT dataset of **22,011** entries from 1900 to 2015.

(d) Verifying the Data Extracts

As acquiring and consolidating EM-DAT extracts is not a straightforward process, it is useful to carry out a simple visual verification to ensure there are no obvious signs of unintentional omissions or inclusions (duplications). To do this, two charts are created using the consolidated EM-DAT dataset to mimic two disaster trend charts from the EM-DAT site [*Figure 4-3: EM-DAT*] (Guha-Sapir et al., 2017d). Reassuringly the overall shapes of the trends appear to match.

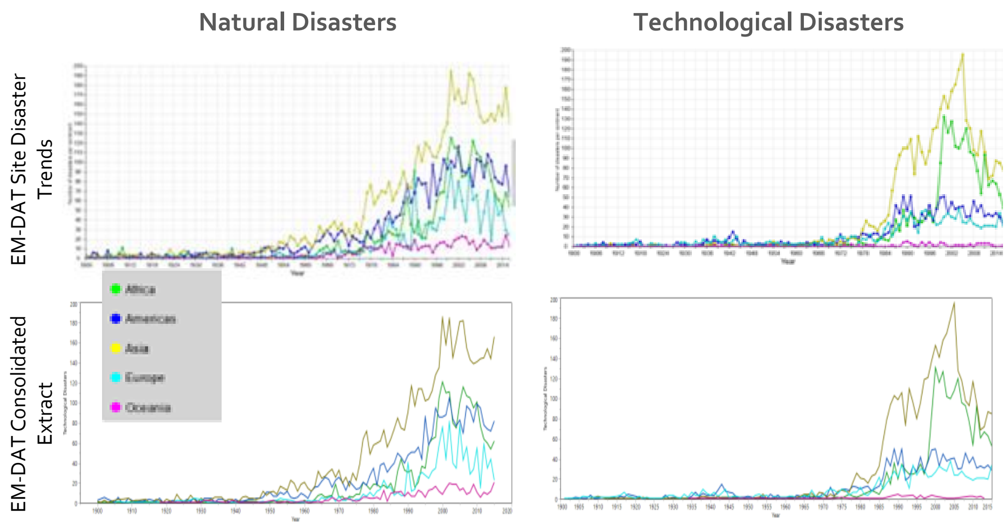


Figure 4-3: EM-DAT Site vs. Consolidated Extract

Note the global increase in disasters from the late twentieth century. This may be a product of the 1988 launch of EM-DAT and not an actual increase in disasters per se.

Preparing the Data

A number of actions are taken to clean and prepare the data for analysis. The more significant of these actions are listed here:

Financial losses are made comparable over the years

Financial losses in EM-DAT are recorded in 1000s of US$s based on the year in which they are incurred; therefore these must be adjusted to enable comparison over time. This leads to the need for the *auxiliary dataset* of **US Consumer Price Index** (USA CPI) to calculate all financial loss entries to the base year of 2015, the last year of entries in the consolidated dataset (BLS, 2016).

Outdated/incorrect country names or codes are rectified

The *auxiliary dataset* of **ISO 3166 Codes** – an international standard for countries, dependent territories and special areas of geographical interest (ISO, 2017) – is used here to verify, and if necessary correct, geographic information.

- Country names or assignments in **4,985** entries are corrected to address errors or reflect changes to sovereignty.
- **305** ISO country codes in the consolidated EM-DAT dataset are incorrect and need to be manually rectified.

Identifying UNDP development status

One of the methods used to estimate the human effect of disasters is dependent on the UNDP development status of the disaster-affected country. i.e. *developing* or *developed* (UNDP, 2017b). Entries that contain estimates using this method are not flagged in EM-DAT. Hence a third *auxiliary dataset*, **Developing Regions classifications** from the United Nations Development Programme (UNDP, 2017b) is needed to distinguish developing from developed countries and thereby identify EM-DAT entries that may contain estimated human effects of disasters.

Missing/incomplete Start and End dates

3,947 entries have missing or incomplete Start or End dates. To circumvent issues with the date fields, the possibility of using specific start or end dates is abandoned and a Year field is created and populated from the first 4 characters of the disaster number (DisNo), which is fully populated and structured Start Year-Sequence Number (YYYY-nnnn).

Duplicate entries removed

Over and above the removal of duplicates created by merging three overlapping EM-DAT data extracts, a further **5** entries are removed as these are 'source' duplicates. That is, these are original entries from EM-DAT that have valid 'unique' identifiers but are in fact for the same disaster recorded more than once in EM-DAT. These may be the result of corrections made to EM-DAT meant to replace previous entries, but the previous entries are left in place.

A Unique Identifier created for future look-ups

A unique identifier is created for each entry by concatenating the disaster number (DisNo) and the ISO 3166 3 character country code. This creates a key that can be used for future look-ups and referencing, if needed.

Examining the Data

Before examining the data in more detail it should be noted that, contrary to information provided on the site, EM-DAT does not hold "*disaster-specific* ***international aid contributions***" (Guha-Sapir et al., 2017a). Other findings include:

Criteria and Classification Inconsistencies

- **1,891** single country disaster entries do not match EM-DAT's human effect inclusion criteria, i.e. at least 10 deaths or 100 people affected (Guha-Sapir et al., 2017e).
- For **391** multi-country disasters entries, equating to **125** disasters (of which 10 disasters affected from 6 to 11 countries), the 'per disaster' human effect does not comply with the EM-DAT

inclusion criteria (Guha-Sapir et al., 2017e). **249** of the 391 multi-country disasters entries contain no human effect whatsoever.

- Only **21** EM-DAT disaster classifications are used as defined (Guha-Sapir et al., 2017g). **38** unique combinations of disaster classification found in the EM-DAT dataset are variations of the standard specified on the site (*these are corrected to ensure consistency*). *Appendix B.1* maps numbers of EM-DAT entries to EM-DAT Disaster Classifications and also provides a mapping of EM-DAT's '*natural*' disaster classification structure to the IRDR classifications that are used as its the basis (IRDR, 2014).

Data Skew and Gaps

- Natural disasters outnumber all other groups of disasters. Disaster entries in EM-DAT are predominantly for EM-DAT's grouping of *natural* disasters (62.7%) with the remaining entries mostly for EM-DAT *technological* grouping (37.2%). Only 14 entries are categorised as *complex* in the database [*Figure 4-4*].
- **26%** of disaster entries are for the first 90 years covered by EM-DAT, **1900–1989**. The remaining **74%** of disaster entries are for the subsequent 26 years, **1990–2015**. [*Figure 4-4*].

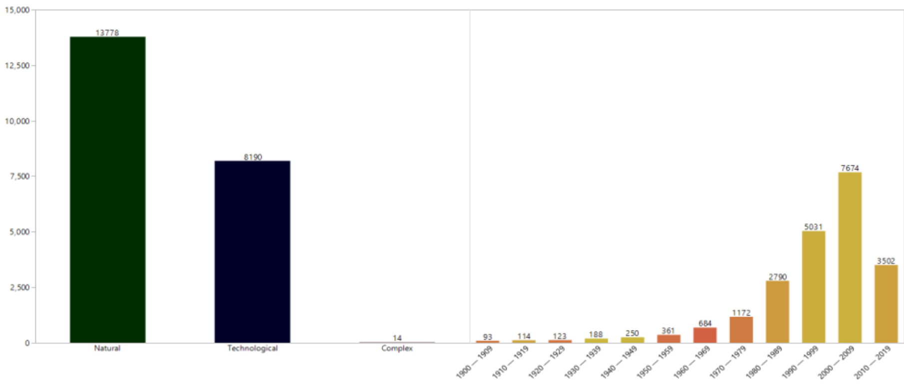


Figure 4-4: EM-DAT - Distribution (Disaster Groups & Decades)

- **1,485** individual entries, almost **7%** of the datasets, are recorded to have no human effect whatsoever and most of these, **893** entries (just over **4%** of the dataset), do not show any financial losses either. Notably, most of the occurrences of ‘no human effect’ entries are in the most recent date ranges [*Figure 4-5*].

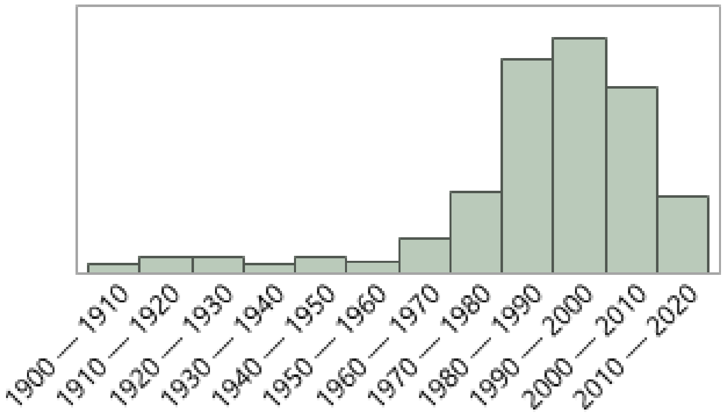


Figure 4-5: Distribution of 'no human effect' EM-DAT entries

Missing or Anomalous Financial Losses

In the absence of a definitive explanation, the relationship between estimated damage and insured damage is interpreted as *X damage (EstDamage) of which Y (InsDamage) is insured*; therefore, to avoid the risk of double-counting, these number are not added together (Guha-Sapir et al., 2017h).

- **17,337** entries (**~ 79%**) do not have any US$ estimated damage.
- **934** have US$ insured losses, but **181** entries (1985 – 2011) of these show no US$ estimated damage.
- Of the **753** entries that contain values in both variables, **21** entries have values for InsDamage greater than EstDamage.
- **17,156** entries have no values for estimated damaged or insured loss, which is anomalous considering **5,385** of these are for transport accidents.

Estimated Human Effect Values

Only two human effect fields are consistently made available across the EM-DAT extracts – TotDeaths and TotAffected. These are composite fields [*Table 4-2*]:

Visible	Not Visible at Detail Level
TotDeaths	= Deaths + Missing
TotAffected	= Affected + Injured + Homeless

Table 4-2: Components of Composite EM-DAT Human Effects (Guha-Sapir et al., 2017i)

When actual human effect of disasters are not available, one of two estimating methods are used for subfields of TotDeaths and TotAffected (Guha-Sapir et al., 2017i):

(1) 2 with or without zeros, i.e. word description of losses becomes a numbers beginning with 2. e.g.: *'hundreds' missing → 200 missing; 'thousands' injured → 2000 injured*.

(2) Human effect is a multiple of damaged houses, i.e. *houses damaged* ***x3*** (developed countries) *or* ***x5*** (developing countries).

Examining the extent to which these methods are likely to have been used in EM-DAT [*Table 4-3*]:

Method 1:	**1,552** entries (over **7%** of the dataset); **671,243,400** people
Method 2:	**6,165** entries (over **28%** of the dataset) **5,261,925,385** people

Table 4-3: EM-DAT Likely Estimated Human Effect

Volatility of the Data

For **20%** of the entries in the consolidated EM-DAT dataset values of human and financial effects changed between the 3 EM-DAT extracts, from the time of the first extract to the time of the last.

4.4.2 The Search for M^{i}Os

Preparation and examination of EM-DAT makes clear that EM-DAT is inadequate as a single source for finding M^{i}Os, primarily because the database does not contain all disaster types. Disaster entries in EM-DAT are predominantly for *natural* disasters, the remaining entries are almost entirely for *technological* disasters. A number of other types of disasters, e.g. those related to conflict, are not represented at all. Nevertheless, it is useful to explore EM-DAT to understand the type of knowledge that can be gained from it.

It is important to note here that exploration of EM-DAT focusses on the occurrence and human effect of disasters. This is because of the insufficiency of financial loss data – around 79% of the entries hold no estimated damage; around 96% hold no insured damage; overall almost 78% of the entries in EM-DAT do not contain any information of financial losses. For completeness a number of visualisations of financial losses have been created and can be found in *Appendix F*.

Disasters and Human Effect (ALL Years)

In addition to the original data from EM-DAT, such as *deaths* and *People Affected*, a new field is calculated, *Survival Rate*. This field is calculated as *People Affected* (surviving victims) expressed as a percentage of the total human effect (*Deaths* plus *People Affected)*. *Figure 4-6* depicts charts for *Disasters*, *Deaths*, *Affected* and *Survival Rate*. Observations worthy of note from these charts include:

- *Deaths* before the mid-1900s are such that they significantly extend the chart scale rendering subsequent bars in the chart barely discernible. In contrast, for *Affected* higher volumes appear from the 1960s onwards.
- *Survival Rate* shows no relationship to the patterns of movement of *Deaths* and *Affected*.

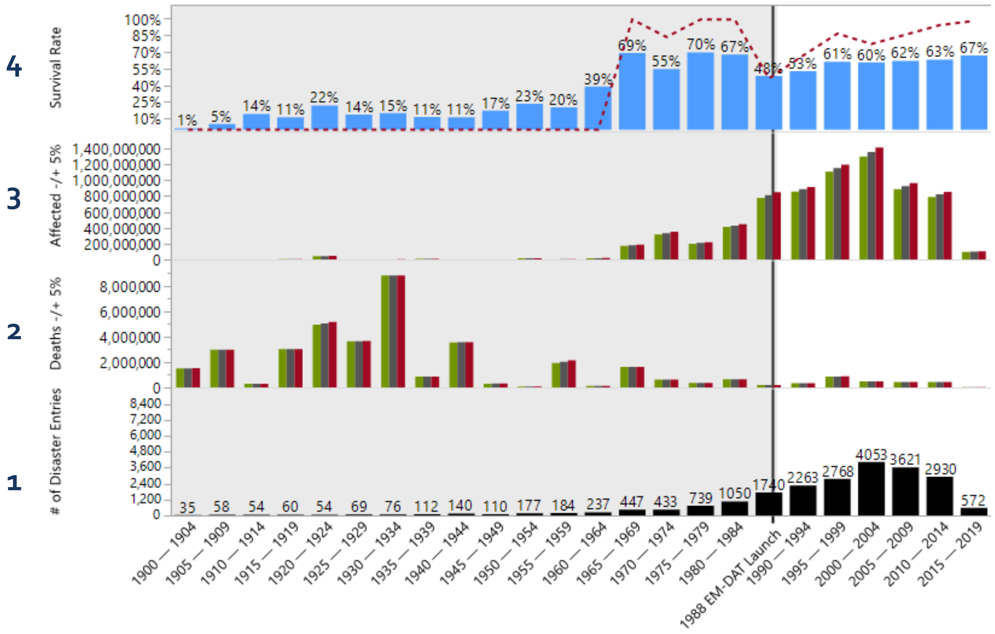


CHART			
1	Number of **Disaster Entries**		
	Left Bars	Centre Bars	Right Bars
2	MINUS 5% variation applied to estimates	Sum of **Deaths**	PLUS 5% variation applied to estimates
3		**Sum of People Affected**	
4	Bars = **Mean of Survival Rate** Red dotted line = **Median of Survival Rate**		

Figure 4-6: EM-DAT Disasters and Human Effect by 5-Year timeslots

- In the early 1900s *Survival Rate* appears to be poor, improving dramatically in the mid-1900s. It is believed this significant shift is likely to be a product of the availability of information rather than improved survival. That is, decades' old data of disaster-related deaths is more readily found than equivalent data of people affected by disasters – more data about deaths than people affected will falsely depress *Survival Rates*.
- From the 1960s the *mean survival rate* range jumps to 50–70%; more remarkably from ~1990 the *Survival Rate* range tightens to 60–67%, remaining within this 7% spread over a period 25 year.
- The implication of visibly staggered bars, particularly in the *Affected* chart, is of significance. In this chart the centre bar is the sum of people affected values and the bars on either side reflect

+/- 5% change in the sum of all guesstimated people affected values. The staggered bars indicate that estimating errors have a noticeable impact on aggregate numbers in that there is a perceptible weakness in the veracity of the data which is caused by the scale of guestimated, as opposed to factual, values.

Finally, *Disaster Entries* in *Figure 4-6* increase in the 1980s, which may be related to the launch of EM-DAT in 1988 resulting in more diligent data collection. As first identified in *Figure 4-4,* the distribution of EM-DAT data is **26%** in time period **1900–1989** and **74%** in time period **1990–2015**. This indicates that post EM-DAT launch data in EM-DAT may be more complete. Therefore in order to avoid significant gaps in historic data this study narrows its focus to the **1990-2015** timespan.

Disasters and Human Effect (1990-2015)

Figure 4-7 depicts *Disasters*, *Deaths*, *Affected* and *Survival Rate* in the 1990-2015 timespan. Notably, the *mean survival rate* [*Chart 4*] remains remarkably steady from 1993 to 2015, exhibiting only very shallow movements of no more than 5% in either direction of 62%. Even the increase in the occurrence of disasters in 2000, 2002 and 2005, does not translate to a commensurate change in *mean survival rate*. With limited data available it is only possible to surmise possible explanations for this phenomenon, e.g.:

- The gap in disaster types, e.g. no conflict data, is creating a partial view that depicts a steadier *mean survival rate* than may be the case in a truer 'big picture' view.
- The guestimated human effect numbers, particularly for people affected, are resulting in near steady *mean survival rate*.
- Humanitarian intervention over these years is having the effect of steadying the *mean survival rate*.

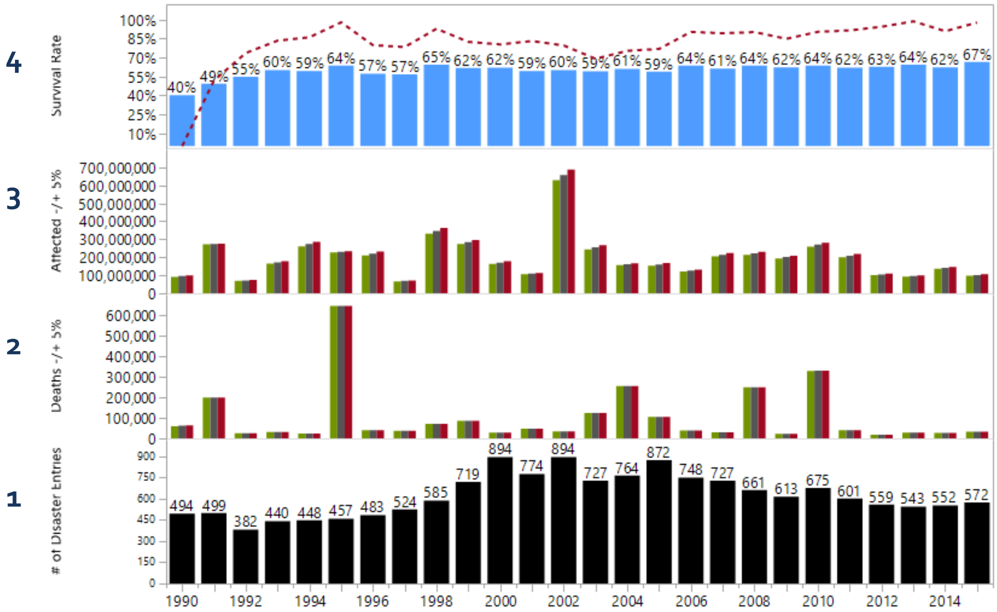


CHART			
1	Number of **Disaster Entries**		
	Left Bars	Centre Bars	Right Bars
2	MINUS 5% variation applied to estimates	Sum of **Deaths**	PLUS 5% variation applied to estimates
3		**Sum of People Affected**	
4	Bars = **Mean of Survival Rate** Red dotted line = **Median of Survival Rate**		

Figure 4-7: EM-DAT Disasters and Human Effect by Year, 1990–2015

As the gaps in disaster types and the existence of weakness in veracity are now known, it can be taken as a given that the means of *Survival Rate* shown is for a subset of disaster data that is less than reliable. This in turn suggests no suppositions can be made from this data as to a relationship between humanitarian intervention and *Survival Rate*.

Disasters and Human Effect by Group (1990-2015)

Breaking down the 1990-2015 *Disasters*, *Deaths*, *Affected* and *Survival Rate* charts into disaster groups, as in *Figure 4-8,* exposes some key differences between EM-DAT's *natural* versus *technological* disasters. Note, there is no point in charting the 14 entries identified as complex disasters as the volume of this data is too small to exhibit any useful information in the aggregate.

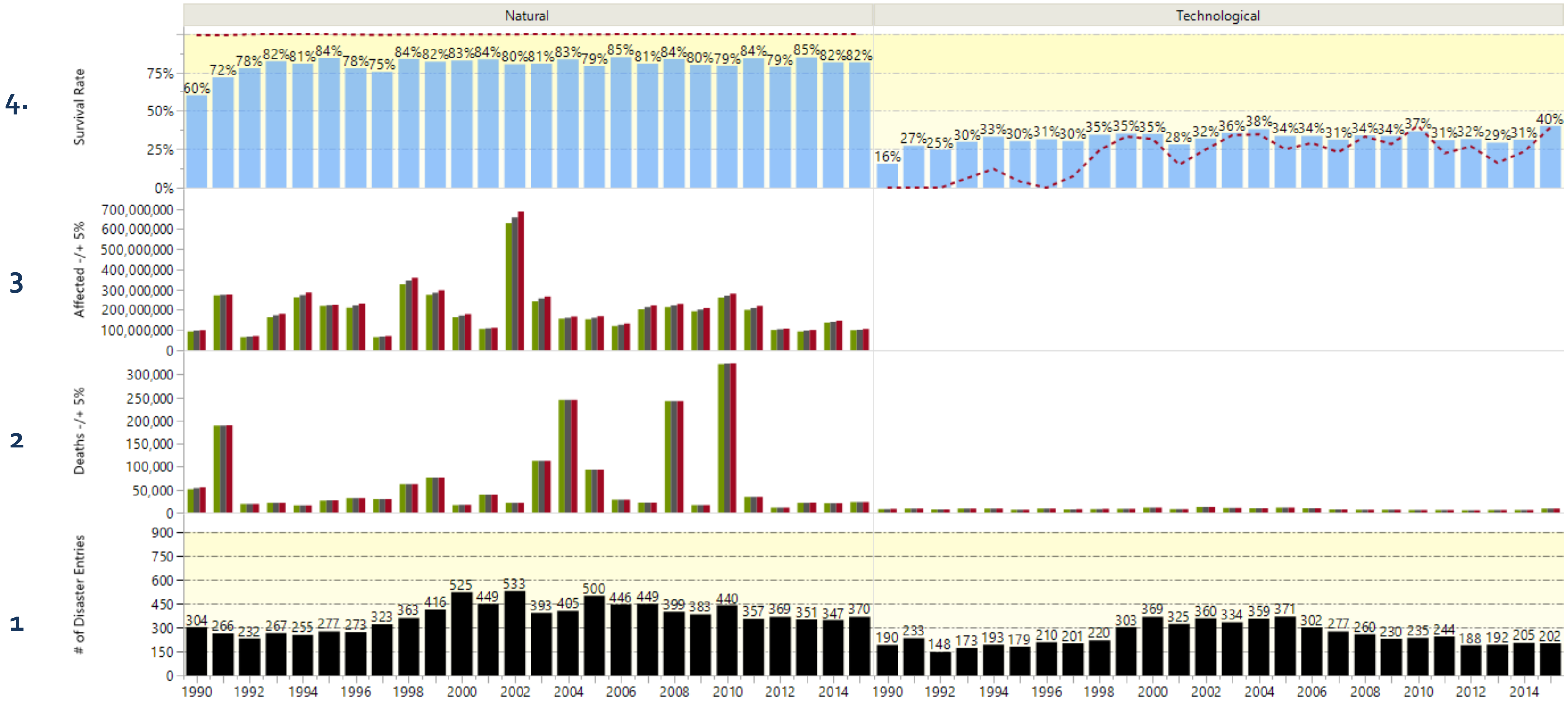


CHART	
1	Number of **Disaster Entries**
4	Bars = **Mean of Survival Rate** Red dotted line = **Median of Survival Rate**

CHART	Left Bars	Centre Bars	Right Bars
2	MINUS 5% variation applied to estimates	Sum of **Deaths**	PLUS 5% variation applied to estimates
3		**Sum of People Affected**	

Figure 4-8: EM-DAT Disasters and Human Effect by Year split by Group, 1990–2015

Some noteworthy observations from these *natural* and *technological* sets of charts include:

- The *mean survival rate* for both disaster groups appears to remain relatively steady from 1993 to 2015. The range of the *mean survival rate* during this period for *natural* disasters is 80% ± 5% and 32% ± 6% for *technological* disasters.
- The *mean survival rates* are far higher for *natural* disasters than for *technological* disasters.
- The scale of *affected* in *technological* disasters is significantly smaller to those *affected* in *natural* disasters; to the extent the bars for *affected* in *technological* disasters disappear if viewed alongside *natural* disasters on the same scale [*Figure 4-8*]. The only way to gain visibility of the bars for *affected* in *technological* disasters is to view these disasters in isolation [*Figure 4-9*].

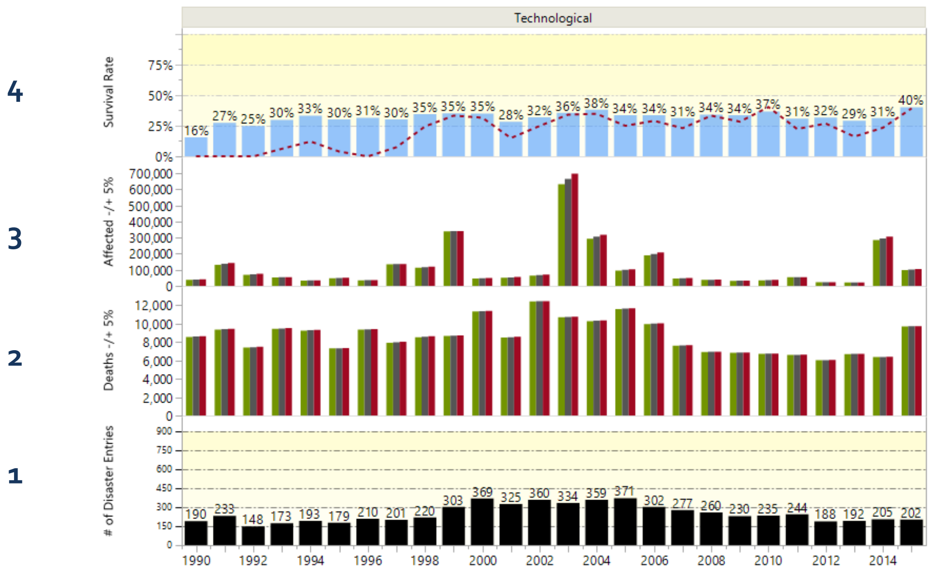


CHART			
1	Number of **Disaster Entries**		
	Left Bars	**Centre Bars**	**Right Bars**
2	MINUS 5% variation applied to estimates	Sum of **Deaths**	PLUS 5% variation applied to estimates
3		**Sum of People Affected**	
4	Bars = **Mean of Survival Rate** Red dotted line = **Median of Survival Rate**		

Figure 4-9: EM-DAT *Technological* Disasters and Human Effect by Year, 1990–2015

- For *Affected* in *natural* disasters most sets of bars are noticeably staggered. This is indicative of a significant number of guestimated values influencing the *Survival Rate*.
- While the volumes of EM-DAT entries for *natural* versus *technological* disasters are different, the occurrence of each type of disaster appears to rise and fall in near tandem. This is checked using a simple line chart [*Figure 4-10*]. No valid reason is obvious as to why two very different types of seemingly unrelated events could occur in near matching patterns over a 26-year period. The suspicion is that this may be more a product of record-keeping rather than disaster occurrence.

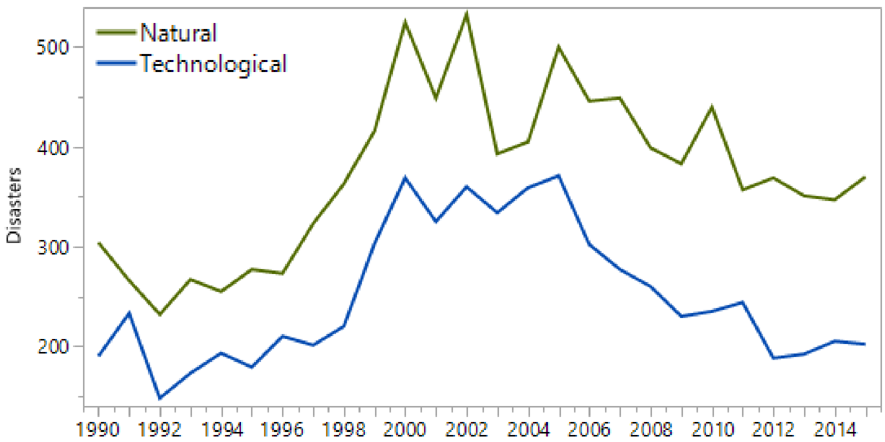


Figure 4-10: EM-DAT *Natural* vs *Technological* Disasters, 1990–2015

In summary, there is little here to provide insights of a relationship between humanitarian intervention and disaster outcome. Even though the data indicates victims are less likely to survive a *technological* disaster than a *natural* one, questions of veracity make this conclusion less reliable. In particular, the inexplicable similarity of pattern in annual occurrences of *natural* and *technological* disasters increases doubt about the veracity of the data.

Disasters & Survival by Group/Region (1990-2015)

Figure 4-11 breaks down the 1990-2015 disaster group charts of *Disasters* and *Survival Rate* by region in the hope that this can provide a better understanding of the data.

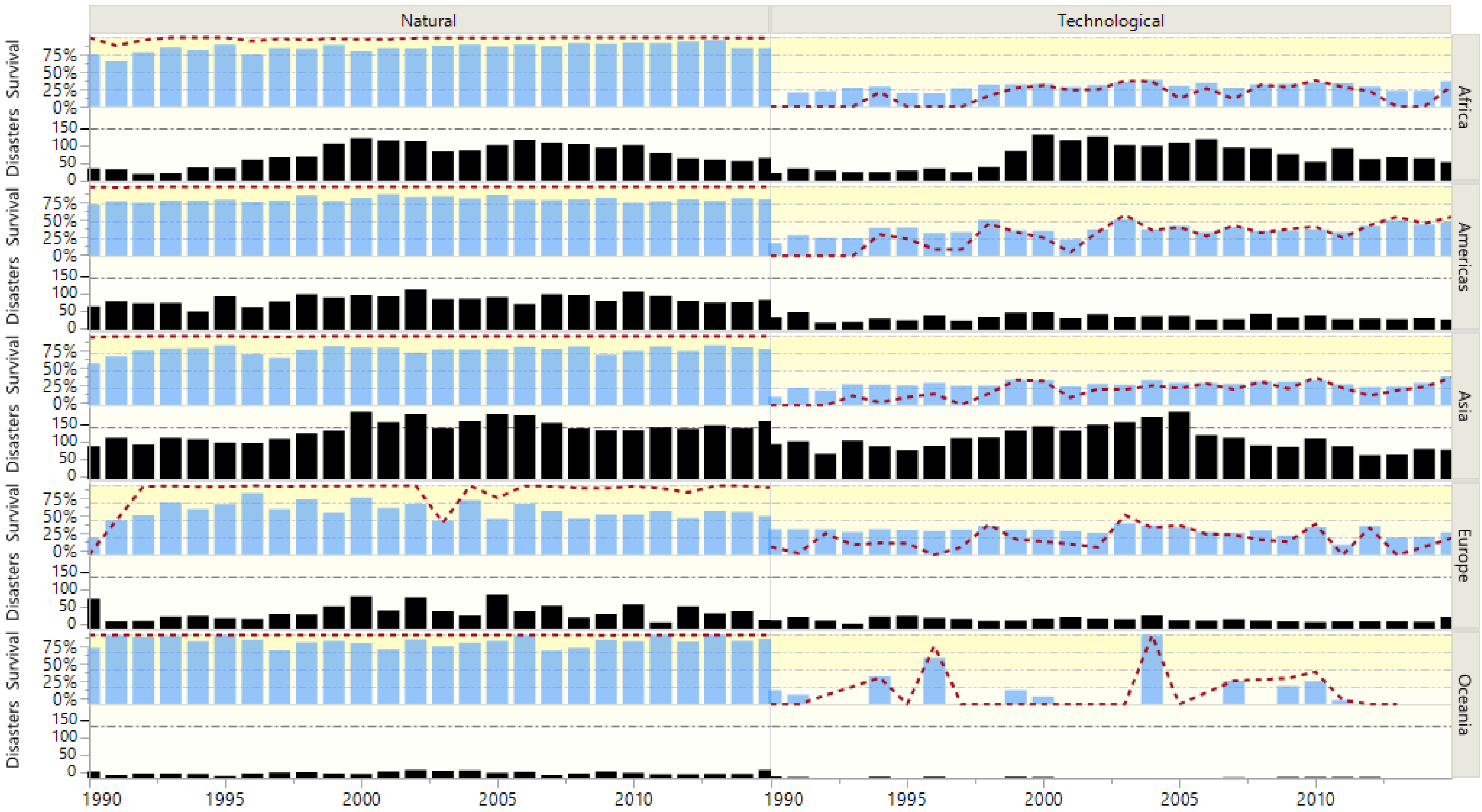


Figure 4-11: EM-DAT Disasters & Survival by Year split by Group/Region, 1990–2015

Observations from *Figure 4-11* include:

- *Technological* disasters, which in EM-DAT are subdivided into types of accidents (industrial, transport and miscellaneous), appear to be far more lethal than *natural* disasters, *in all regions*.
- Oceania experienced the least number of disasters of all regions and most of the disasters that did affect the region are *natural*, with sporadic occurrences of *technological* disasters.
- *Natural* disasters in Europe have some of the lowest, and most erratic, *mean survival rate*. This is unexpected considering Europe constitutes mostly developed countries that can reasonably be assumed to have better disaster mitigation initiatives, and more adept first response services, than lesser developed regions.
- From the distribution of disaster entries it can be seen that the bulk of disasters from 1990 to 2015 occurred in three regions: Africa, Asia and the Americas. The mean and medians of the average *Survival Rates* of *natural* disasters in these regions is almost level for the most recent twenty-five years, with the medians hovering just above the bars of the means. This appears to be the case even if the frequency of disasters fluctuates. Whereas, when we look at the charts for *technological* disasters in Africa, Asia and the Americas not only is the average *Survival Rate* poor, the medians tracking below the means suggests that most of the *Survival Rates* are worse than those depicted by the bars.

4.5 Evaluate

To recap from *Section 3.3.2*, the method of evaluation applied in this study is *descriptive informed arguments based on knowledge of the research domain* (Hevner et al., 2004). Using this method, artefacts are evaluated for utility against the aim of the research, thus acknowledging the teleological nature of DSR (Simon, 1996; Hooker, 2004). Deductive and logical reasoning provides a means of harvesting and synthesising the knowledge gained to test the utility theory that inspired the work; or even allow new design theories to emerge (Venable, 2013; Gregor, 2006).

The evaluation step also facilitates a better understanding of both the solution and problem space, which in turn enables the shaping of the next iteration and the seeding of future research opportunities (Venable et al., 2012; Gregor and Jones, 2007; Vaishnavi and Kuechler, 2004b). With these goals in mind the evaluatory reasoned arguments of this section centre on the following:

- **Tentative Design ⇔ Build (Grow) Alignment**: evaluating the alignment, or misalignment, of the design and the realisation of that design may provide insights that affect the study.
- **DSR Artefacts**: evaluating the output of the *Build (Grow)* step against the aim and objectives of the research is fundamental to assessing progress and continued validity.
- **Knowledge ⇔ Consequence**: evaluating the knowledge gained from the iteration and discusses the consequences of this improved awareness.
- **The Utility Theory**: testing the study's utility theory based on the knowledge gained and artefacts created.

Distilled to its essence the evaluation step of each iteration addresses the questions – *what did it achieve* and *what did it fail to achieve* – the answers to which are used to shape the remainder of the work.

4.5.1 Tentative Design⇔Build (Grow) Alignment

Here the *tentative design* of the iteration is evaluated against the *build (grow)* step to identify how closely it aligns to the needs of developing the artefacts. Even though '*tentative*' in DSR allows elaboration of the design during the *build (grow)* step to facilitate the construction of artefacts (Vaishnavi, 2008), this is useful information that may affect or influence the next iteration of the study, or future research opportunities.

The tentative design for Iteration 1, in summary form, simply states that disaster occurrences and losses be explored to search for macro-indicators of outcome, M^{i}Os. The Emergency Events Database (EM-DAT), as a single source, is presumed to be complete and veracious enough to support this search (Guha-Sapir et al., 2017l). These presumptions of data completeness and veracity are based on EM-DAT documentation and the use of its data by credible publications (Guha-Sapir et al., 2017i, 2017f, 2017e, 2017j).

Use of EM-DAT in the build (grow) step identified the following:

- The data cannot be used for analysis without correcting numerous country codes and references, standardising its financial losses to a base year and identifying the loss figures that are likely estimates. This entails the use of three additional datasets:

  (1) **International Organization for Standardization, Country Codes** (IS0-3166, 2017) for country naming conventions and codes to rectify incorrect or outdated country information in almost 23% of the entries.

  (2) **USA Consumer Price Index** (USA-CPI) from the Bureau of Labor Statistics (BLS, 2016) to bring all financial losses to parity, in this case 2015, to enable comparisons of losses over time.

  (3) **Developing Regions** classifications from the United Nations Development Programme (UNDP, 2017b) are needed to distinguish developing from developed countries as one of the estimation methods used in EM-DAT hinges on the development status of the country.

- The dataset does not represent all disaster types. Almost two-thirds of the entries are for *natural* disasters, all but 14 of the remaining entries are for *technological* disasters and the 14 entries are all that represent complex disasters. As it is known that other types of humanitarian crisis exist, such as those caused by conflict and deracination, it is safe to say EM-DAT does not represent all disaster types (HERR, 2011; Belanger et al., 2016; Lattimer et al., 2016; van der Zee, 2015).

- Over 8,200 entries are likely estimated values for human effects.
- Almost 1,000 entries are included even though they do not meet the key quantifiable inclusions criteria for EM-DAT.
- Other limiting issues include
  - Temporal detail is limited to years as this is the only date information that is complete (and obtainable) from the disaster identifier, both start date and end date variables have entries that are missing or incomplete.
  - Less than 25% of entries hold any information of financial loss, therefore financial loss data is considered insufficient for meaningful analysis.
  - While entries in the dataset date back to 1900, the first 90 years are poorly represented, therefore the focus of any analysis is best restricted to years beyond 1990.

In summary, the design aspirations of this iteration are not met during the build (grow) step because of the limitations of the identified disaster loss dataset.

4.5.2 DSR Artefacts

The only DSR artefacts originally envisaged in *Section 3.3.3* realised to any extent in this iteration are those that fall into the category of *[b] Data analysis outputs and visualisations* [*Table 4-4*] (March and Smith, 1995; Vaishnavi and Kuechler, 2004b; Hevner, 2007).

<table>
<tr><th colspan="3" rowspan="3">Research Framework</th><th colspan="4">Research Activities</th></tr>
<tr><th colspan="2">Design Science</th><th colspan="2">Natural Science</th></tr>
<tr><th>Build (grow)</th><th>Evaluate</th><th>Theorise</th><th>Justify</th></tr>
<tr><td rowspan="4">Research Outputs</td><td rowspan="4">Artefacts</td><td>Constructs</td><td>[a] →</td><td></td><td>[c] --→</td><td></td></tr>
<tr><td>Models</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Methods</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Instantiations</td><td>[b] →</td><td></td><td></td><td></td></tr>
<tr><td>[a]</td><td colspan="6">Macro-indicators of disaster outcome and the impact and effectiveness of humanitarian intervention (MiOs, MiIs and MiEs).</td></tr>
<tr><td>[b]</td><td colspan="6">Data analysis outputs and visualisations.</td></tr>
<tr><td>[c]</td><td colspan="6">A (behavioural science) hypothesis relating the availability, or lack thereof, of humanitarian data and the flow of humanitarian aid that emerges from the domain knowledge and may be worthy of future research.</td></tr>
</table>

Table 4-4: DSR Output to Research Framework Mapping v.1
(Vaishnavi and Kuechler, 2004b; Hevner, 2007; March and Smith, 1995)

The insufficiency of EM-DAT as a single source of disaster loss data limits the usefulness of these outputs for this research. That said, the visualisations of *mean survival rate* identify this calculated human effect metric as a potential M^{i}O meriting further exploration.

4.5.3 Knowledge ⇔ Consequence

This section assesses the knowledge gained from this design cycle iteration and argues the consequences of this improved awareness. Key nuggets of knowledge that emerge from this iteration include:

(a) EM-DAT does not represent all disaster types

From examining the dataset it is clear that it predominantly represents *natural* disasters and man-made accidents since 1990 rather than, as *implied* by the site's documentation, all disaster types from 1900.

(b) Not all humanitarian crises are called disasters

Examination of EM-DAT confirms that a number of humanitarian crises, such as conflict and deracination, are not recorded in this disaster database (Guha-Sapir et al., 2017g), but these other types of crises also receive humanitarian aid (van der Zee, 2015; Belanger et al., 2016). This mismatch between what is considered a disaster by EM-DAT and what needs funding as a disaster can be seen when comparing the EM-DAT bar charts with a chart of UN Coordinated Appeals [*Figure 4-12*] (Lattimer et al., 2016).

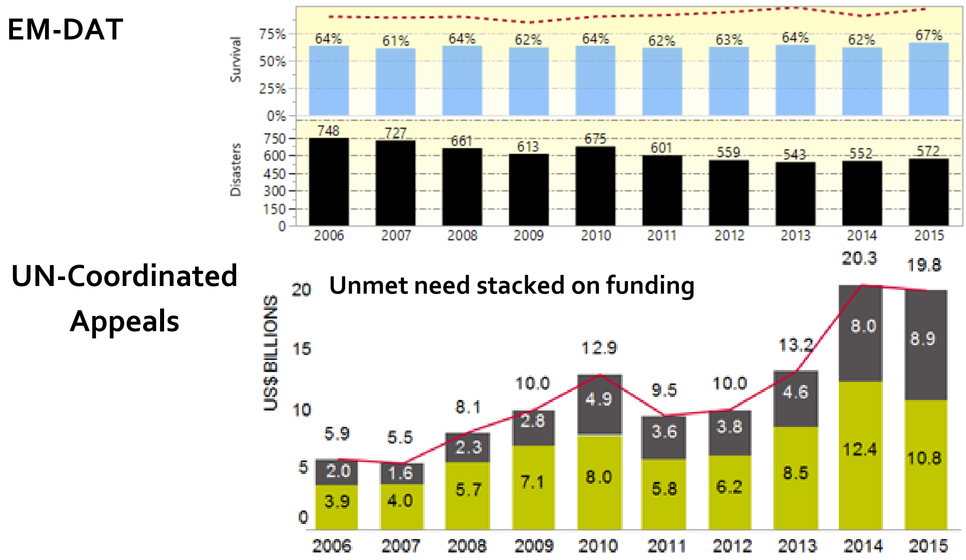


Figure 4-12: EM-DAT vs UN-Coordinated Appeals (2006 – 2015)
(Lattimer et al., 2016)

The differing chart patterns affirm the explanation that, unlike the EM-DAT dataset, UN funding is not restricted to *natural* disasters or man-made accidents (UNISDR, 2009; van der Zee, 2015; Lerner, 2016; Guha-Sapir et al., 2017g; Belanger et al., 2016).

(c) There is no single definitive source of disaster data

The selection of EM-DAT as a potential single source of disaster data is the result of an extensive search for the *most* veracious dataset that offered the *most* comprehensive view of disaster data. The selection is not based on an idealistic expectation of absolute completeness. Even the more realistic assumption of *sufficient* completeness is proven false. That said, EM-DAT is still believed to be the best option available therefore it is highly unlikely that a data set exists that holds a complete view of most humanitarian crises.

(d) There is no classification model for all disaster types

EM-DAT's classification system is extensive for *natural* disasters based on IRDR's Peril Classification and Hazard Glossary (IRDR, 2014); fairly rudimentary for *technological* disasters; and non-existent for any other group of disasters (Guha-Sapir et al., 2017g). This together with the fact that not all humanitarian crises are called disasters, it is understandable that a search for a comprehensive taxonomy of all disaster types does not yield any results.

(e) EM-DAT data for disasters before 1990 is sparse

EM-DAT has entries for disasters that date back to 1900, but the first 90 years are sparsely populated and the richest seam of data is from 1990, two years after EM-DAT's launch in 1988.

(f) EM-DAT temporal granularity is limited to annual

Numerous disaster Start and End dates in EM-DAT are incorrect or incomplete, only the *year* component in the DisNo (disaster identifier) is available for all entries (Guha-Sapir et al., 2017i). Therefore, while this study is primarily interested in macro-level data, it should be noted that even if monthly geographical or categorical events in the aggregate are considered worth exploring, this is not feasible using EM-DAT data.

(g) EM-DAT financial losses are insufficient for study

Only around 22% of EM-DAT entries hold any financial loss at all. There is no obvious logic as to which entries may or may not reflect US$ losses. As EM-DAT cites major international re-insurance firms in its list of data sources it is surprisingly that this information is not better populated (Guha-Sapir et al., 2017k).

(h) Survival Rate is worth investigating

Although EM-DAT is only a subset of global disasters, the charts of *mean survival rate* may foreshadow this created variable as a possible M^{i}O if applied to a more complete set of disaster data

[*Figure 4-13*]. The other human effect variables, when viewed in isolation, show no hope of a potential pattern or relationship. In the aggregate, there is no obvious relationship between the bars of *Deaths* and *Affected*; their numbers appear to move independently of each other. Whereas the near level bar chart of *mean survival rate*, suggests that it may be possible to find some significance in its behaviour if it is based on more complete data.

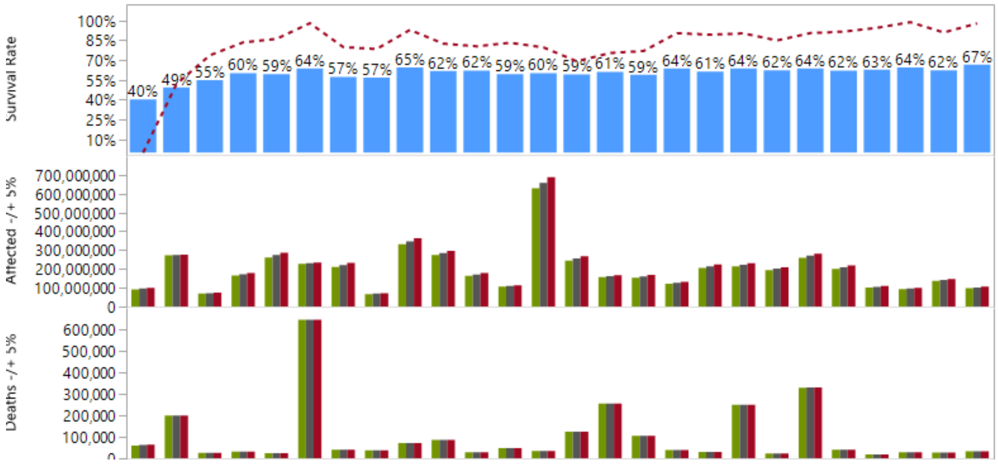


Figure 4-13: EM-DAT Possible M^{i}Os, 1990–2015

(i) Data veracity must be evaluated and not assumed

EM-DAT is collated and maintained by a highly credible source and used by reputable publications (CRED, 2017; Guha-Sapir et al., 2017j). Nonetheless, preparing and examining EM-DAT data reveals numerous issues of veracity that undermine the reliability of any analysis or visualisation that may be based on this data. Therefore, while the credibility of the source or the users of the data may aid in the selection of a dataset, these are clearly inadequate criteria by which to assess the veracity of its data. Data veracity must therefore be evaluated based on the data not on assumptions of source and usage.

Table 4-5 maps the knowledge garnered from this iteration to the consequence of that knowledge on this study and, where applicable, on future opportunities for research.

Knowledge	Evaluate	Consequence
(a) EM-DAT does not represent all disaster types	Humanitarian funding is a key aspect of the response step of the disaster management cycle. It equates to aid across a breadth of humanitarian crises that may or may not be referred to as disasters. A balanced search for the effectiveness of humanitarian intervention based on relevant factors (e.g. disaster losses, humanitarian aid etc.) even in the aggregate must refer to the same, or almost the same, population of disasters. Therefore, a master set of global disasters, and a corresponding disaster classification system, are '*data scaffolds*' without which data analysis of the humanitarian sector is significantly restricted.	To continue this study a master dataset of global disasters must first be created by amalgamating a number of disaster loss datasets. A disaster classification model to categorise all disasters in the amalgamated master dataset of global disasters is also needed.
(b) Not all humanitarian crises are called disasters		
(c) There is no single definitive source of disaster data		
(d) There is no classification model for all disaster types		
(e) EM-DAT data for disasters before 1990 is sparse	Analysis of this EM-DAT disaster loss dataset is best limited to the period 1990 to 2015 and the human effect of disasters. It is also not possible to use this data to carry out analysis of daily or monthly effects of disasters.	Any analysis of amalgamated disaster losses data that includes EM-DAT is also as a result constrained to 1990 to 2015, human effect and annual.
(f) EM-DAT temporal granularity is limited to annual		
(g) EM-DAT financial losses are insufficient for study		
(h) Survival Rate is worth investigating	The annual *Survival Rate* remains almost level over most of the years explored, even when the *death* and *affected* data on which *Survival Rate* is based varies considerably. This may indicate that it has relevance as a potential indicator.	*Survival rate* needs to be calculable in any disaster loss dataset created.
(i) Data veracity must be evaluated and not assumed	The type and extent of any weaknesses in the veracity of sourced datasets must be understood in order to gauge the reliability of analysis carried out using these datasets.	A means of equitably evaluating and comparing the veracity of sourced data is a prerequisite to creating a master dataset of global disasters.

Table 4-5: Iteration 1 Knowledge ⇔Consequence Mapping

4.5.4 The Utility Theory

Restating the utility theory statement of this study [*Table 4-6*]:

	Solution Space	Utility	Problem Space
	Form	Function	Purpose
	Artefact [*What*]	Efficacy [*How*]	to Address [*Why*]
STATEMENT	*Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector…*	*…when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level…*	*…the inability to gauge the consequences of monies spent and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.*

Table 4-6: Structure of the Utility Theory Statement

As DSR artefacts in the solution space are not realised in Iteration 1, testing the propositions of this utility theory is moot. Nevertheless, two important realisations emerge as a consequence of the knowledge gained from this iteration:

(1) That it is infeasible to attempt to explore and analyse data in a research domain without essential domain-specific data support structures. For the humanitarian sector, such support structures – from this point these are referred to as '*data scaffolds*' – include a *master set of global disasters* and a *master disaster classification* system.

(2) That the credibility of the data source and the use of the data by reputable publications cannot be taken as de facto credentials of the veracity of that data. Therefore, a means of evaluating, measuring and comparing the contextual (relevant to purpose) veracity of datasets is a prerequisite to understanding the reliability of analytical output.

As a result, testing of the study's utility theory is now dependent on the creation of *data scaffolds* (1) and a means of equitably *evaluating the veracity of the data* used (2).

4.5.5 Repercussions of New Knowledge

The repercussion of improved awareness of the problem space from this iteration is a revision of the structure of the study. The focus at the start of this iteration is to explore a single global disaster data for macro-indicators [*Figure 4-14*]. As a result of this iteration it becomes clear that two '*problems*' need to be addressed before focus can return to searching for macro-indicators in global disaster losses [*Figure 4-15*].

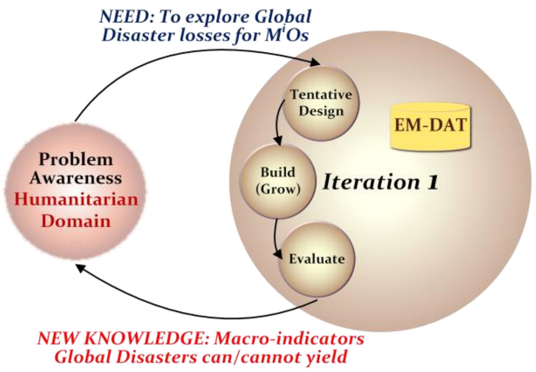


Figure 4-14: Initial view of Iteration 1

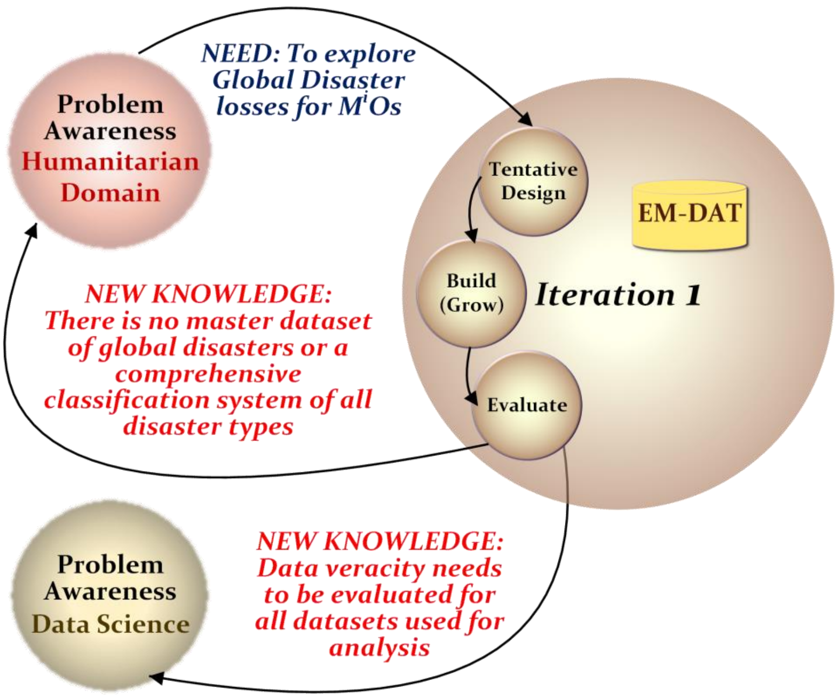


Figure 4-15: Revised view of Iteration 1

The absence of a master dataset of global disasters and the lack of a corresponding disaster classification system needs to be addressed before macro-indicators can be sought. Therefore, an amalgamated disaster dataset and supporting disaster classification system need to be created. Using any dataset, let alone and an amalgamated one, for analysis while blind to the veracity of its data will result in outputs of unknown reliability. Therefore, inability to consistently and equitably evaluate the veracity of datasets must be addressed first to ensure the veracity of each sourced dataset is known when amalgamating disaster data. *Note: the data veracity 'problem' is outside the humanitarian domain and is applicable to all types of data*.

4.6 Summary

This chapter discusses Iteration 1 and the progressive realisation that the assumptions and expectations at the beginning of the iteration are less than accurate. The chapter starts by summarising the originating problem, highlighting once again the imperative to gauge the effectiveness of humanitarian intervention efforts. It goes on to specify the *Tentative Design* for this iteration, which is essentially the use of data from CRED's Emergency Events Database (EM-DAT) to search for macro-indictors of disaster outcome (CRED, 2017; Guha-Sapir et al., 2017l). The *Build (Grow)* step details the acquisition, preparation, examination of the data, highlighting any issues found with the data, before going on to describe the exploration of the data.

Finally, the discussion moves to the *Evaluate* step. Here an explanation is provided as to why this iteration cannot fulfil its original intent, arguing that the reason for this also seeds two further utility theories and reshapes the remainder of the study. The contention being that the research must now incorporate the provision of a means to evaluate data veracity as well as the creation of a master dataset of global disasters and associated disaster classification system.

Chapter 5: DATA VERACITY (ITERATION 2)

5.1 Overview

This chapter describes and discusses Iteration 2 of the DSR design cycle for this research. The focus of this iteration is to develop a means of evaluating the veracity of datasets such that the extent to which their data can be relied upon can be equitably and consistently assessed and compared. The chapter therefore describes the design and development of a framework to evaluate the veracity of different datasets. *Figure 5-1* is a basic schematic of the flow of Iteration 2.

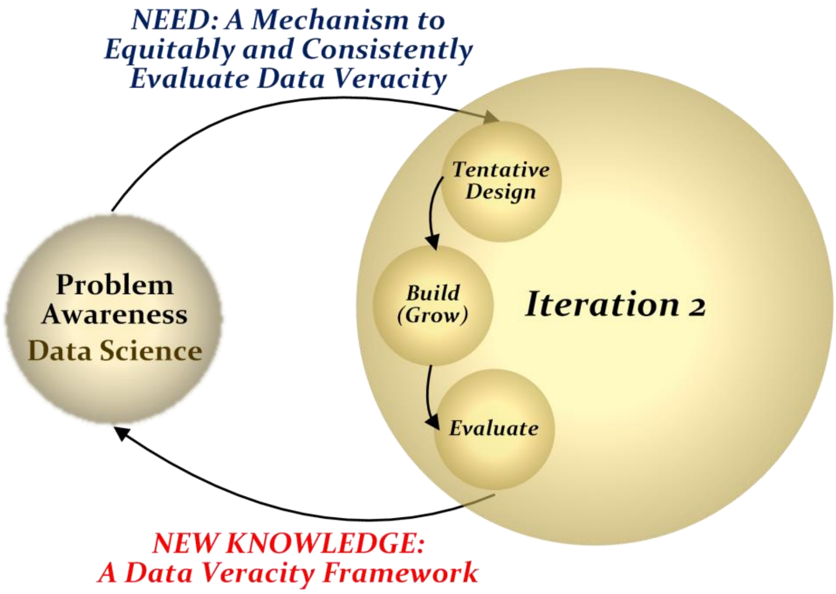


Figure 5-1: Iteration 2 of the Design Cycle

The structure of the chapter is as follows: *Section 5.2* discusses the need to evaluate data veracity and the lack of existing mechanisms that enable such an evaluation, even in the growing domain of data science. *Section 5.3* discusses the *Tentative Design* of the data veracity framework to be built. *Section 5.4* describes the *Build (Grow)* step to create the data veracity framework. *Section 5.5 Evaluates* the work and outputs of Iteration 2 before *Section 5.6* closes with a summary of the chapter.

5.2 Problem Awareness

This section discusses the importance of assessing the veracity of data used for analysis in general and the need to create a data veracity framework in support of this study in particular. Analysis built on data of unknown or weak veracity is of equally unknown or weak veracity (Clarke, 2016; D'Mello, 2016). Creating algorithms, models and visualisations without understanding the veracity of the data on which these artefacts are based is to risk analyses that form false perceptions, or potentially worse, trigger erroneous actions (D'Mello, 2016). Logic dictates that data that are not veracious are likely to lead to insights that are misleading or incorrect (Clarke, 2016). Hence, regardless of the scale or magnitude of the data being analysed there is a need to gauge the extent to which the data can be trusted and identify any weakness that must be addressed or accepted, or that may in fact render the data unfit for purpose.

Having stressed the importance of knowing the veracity of the data being used, it is acknowledged that in current literature no standardised and universally accepted definition of data veracity can be identified. Instead it is often simply described in the negative, i.e. what data should not be in order to be considered veracious – e.g. ‘data in doubt’, ‘uncertain’, ‘abnormal’, ‘noisy’ and ‘biased’ (Normandeau, 2013; Berti-Equille and Lamine Ba, 2016; Berti-Equille and Borge-Holthoefer, 2015; Lukoianova and Rubin, 2014; Claverie-Berge, 2012; Schroeck, 2012). Without a standard definition of data veracity it is not surprising that an accepted model or framework for the evaluation of data veracity also cannot be found. The conclusion drawn here is that in order for this research to progress, a definition of data veracity must first be specified and then used as the basis of designing and creating a framework to evaluate data veracity.

Another finding considered worthy of note before proceeding to the design and creation of a data veracity framework is the significant gaps in standard definitions for other data characteristics relevant to the study of data; many of these gaps creating space in which a confusion of opinions can thrive. This is particularly true of the burgeoning 'bandwagon' of Big Data, which has many qualifying designations beginning with 'V' e.g. volume, velocity, variety, visibility, viability, variability, veracity etc. (Grimes, 2013; Laney, 2001, 2012, 2013; IBM, 2013). Though primarily interested in *veracity*, the knowledge that these other gaps exist provides better context of the vagueness surrounding data veracity in the data science domain, providing food for thought as to its relationship with other Vs of data.

This study takes the view that veracity is one of ***six*** Vs of data, namely *volume*, *velocity*, *variety*, *veracity*, *virtue*, and *value* deserving of further study, more standardised definitions, and agreed upon evaluatory models. The first three Vs – *volume*, *velocity* and *variety* – are **definitional** qualities that are measures of magnitude used to identify the data as *big*, or not as the case may be (Laney, 2001, 2013); whereas, the remaining three Vs –*veracity, virtue*, and *value* – are **aspirational** qualities that can be applied to data of all magnitudes. Here *virtue* is taken as the ethical and moral provenance and use of data (Floridi and Taddeo, 2016; Boyd and Crawford, 2012); and *value* as alluding to the worth of the data (Glikman and Glad, 2015; Baldwin, 2015).

5.3 Tentative Design

In the absence of a clear and widely-accepted definition of ***data veracity***, it is specified here to be ***a determinant of the extent to which the data is a credible and trustworthy basis for contextually cogent and tenable insights***. In designing a data veracity framework guidance is taken from seminal work in data quality (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996), as well as referencing concepts from the Big

Data multiverse, where veracity has some traction as an aspirational *V* of Big Data (Laney, 2013; Grimes, 2013; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014; Powers Dirette, 2016; Normandeau, 2013).

By analysing, rationalising and grouping the various dimensions and criteria discussed by data quality and Big Data commentators, together with some abductive reasoning, a tentative design for a **Data Veracity framework** (**DVf**) is conceived (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014; Powers Dirette, 2016; Normandeau, 2013). The design of the DVf must provide a *model* that specifies a set of characteristics against which data can be assessed. Based on this model a *profile* must be created to capture findings when the data is assessed and a mechanism is needed to calculate a veracity *index* to enable comparison across datasets [*Figure 5-2*].

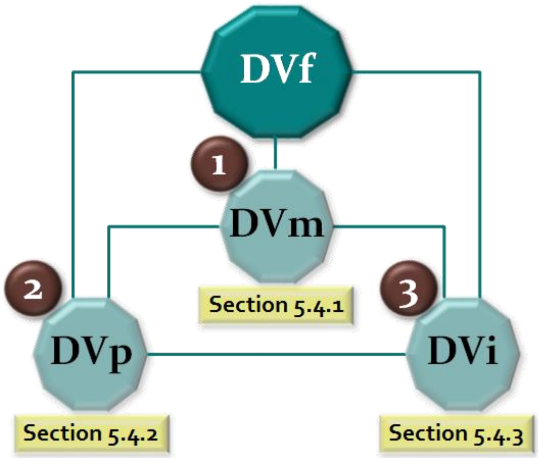


Figure 5-2: Data Veracity framework (DVf) – 3 Interrelated Constituent Parts

(1) A **Data Veracity model** (**DVm**): a hierarchical structure of characteristics that collectively determine the veracity of data.

(2) A **Data Veracity profile** (**DVp**): a means of capturing descriptive veracity information about a dataset.

(3) A **Data Veracity index** (**DVi**): a means of calculating a numeric indicator of the overall veracity of a dataset that can be used for comparison against the DVis of other datasets.

5.4 Build (Grow)

The tentative design of the *Data Veracity framework* (DVf), which consists of three constituent parts, includes inherent dependencies. The *Data Veracity model* (DVm) needs to be defined first to provide a foundational structure to the way the *Data Veracity profile* (DVp) must be constructed. Finally, the creation of the *Data Veracity index* (DVi) is dependent on the structure of the *Data Veracity model* (DVm) and the ability to use it is dependent on a populated DVp. Therefore, this build (grow) step constructs the DVf in the following order ***DVm*** → ***DVp*** → ***DVi*** [*Figure 5-2*].

It should be noted that in constructing the DVf research in the areas of data quality and Big Data is used as a knowledge base in general and individual papers cited where specifically applicable (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014; Powers Dirette, 2016; Normandeau, 2013).

5.4.1 Data Veracity model (DVm)

As stated earlier, veracity is considered here to be a determinant of the extent to which the data is a credible and trustworthy basis for contextually cogent and tenable insights. With this in mind, sifting and sorting through the various commentaries on data quality and Big Data, veracity has been structured as a hierarchy of dimensions that are either *Elucidatory* or *Expository*. The distinction being that *Elucidatory* dimensions are related to the '*hygiene*' of the data, whereas *Expository* dimensions are related to the '*credibility*'.

Dimensions that are ***Elucidatory*** address the question of how *clean* and *uncluttered* the data is (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014; Normandeau, 2013). Dimensions that are ***Expository*** address the question of how *precise* and *accurate* the data is (Strong et al., 1997; Lee et al., 2002;

Wang and Strong, 1996; Wand and Wang, 1996; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014; Powers Dirette, 2016; Normandeau, 2013).. As a precursor to explanation, *Figure 5-3* provides a summary of the full hierarchy of dimensions that are defined for the veracity of data, shown as level 1 (L1), level 2 (L2) and level 3 (L3) dimensions (L3 dimensions being those that are assessed in practice).

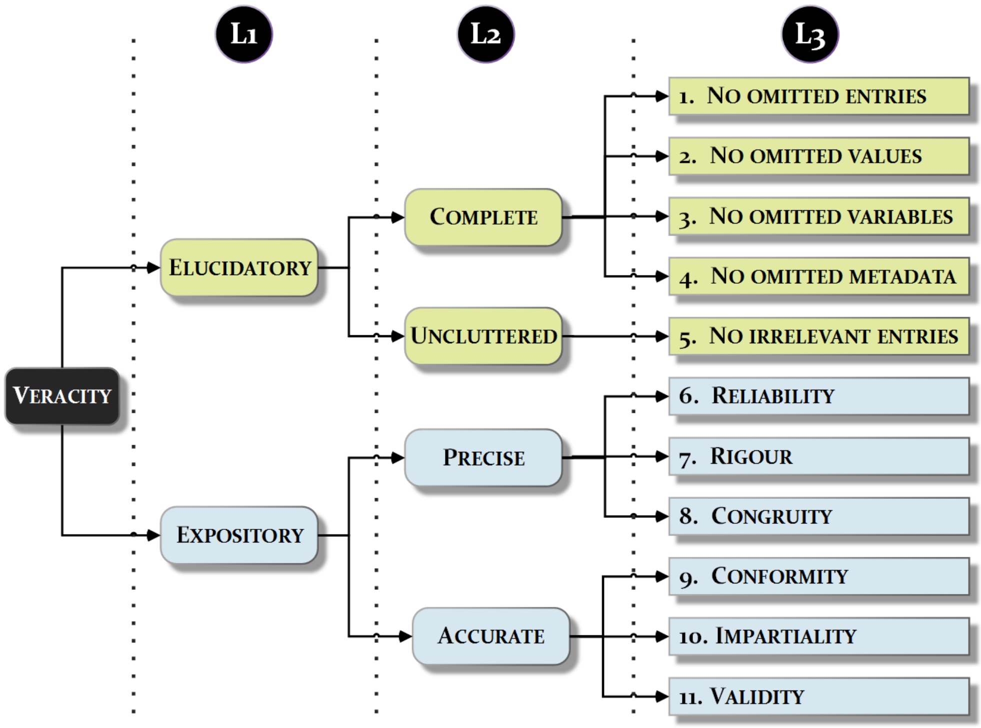


Figure 5-3: Veracity - Definitional Hierarchy

These dimensions are further explained here:

- **Elucidatory** dimensions facilitate an assessment of the clarity of the dataset as a whole, examining the *extent to which the data is comprehensive, narratively coherent and unadulterated*. These dimensions fall into two subgroups, *complete* and *uncluttered*:
    - **Complete** – *How* complete *is the dataset with respect to its purpose?* (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014).

1.	**No omitted entries**	Are there missing observations?
2.	**No omitted values**	Are there empty/incomplete fields?
3.	**No omitted variables**	Are the observations incomplete?
4.	**No omitted metadata**	Is there missing information about the data?

- **Uncluttered** – Is the dataset free from 'noise' and clutter? Does it include inappropriate, spurious or misleading entries? (Lukoianova and Rubin, 2014; Normandeau, 2013).

5.	**No irrelevant entries**	Does it include inappropriate, spurious or misleading entries?

- **Expository** dimensions help expose the proximity of the data values to facts. – To what extent is the data a reflection of reality rather than a perception of reality? – These fall into two subgroups, precise and accurate:

  - **Precise –** *How exact and unambiguous is the data?* (Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Berti-Equille and Lamine Ba, 2016; Powers Dirette, 2016).

6.	**Reliability**	The data is not **volatile** or **uncertain.** *Does the data deviate from the correct, intended or original values? Are there abnormalities in the data? Is it vague or confusing? .*
7.	**Rigour**	The data has been meticulously collected or measured as opposed to estimated or assumed. *Is the data a product of scrupulous data gathering or assumptions/guesswork?*
8.	**Congruity**	The data has values that are consistent and congruous.

  - **Accurate –** *How close is the data to real values?* (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Lukoianova and Rubin, 2014; Powers Dirette, 2016; Normandeau, 2013).

9.	**Conformity**	The data conforms to facts. *To what extent is the data a reflection of reality?*
10.	**Impartiality**	The data is unbiased. *Is the data skewed in any way?*
11.	**Validity**	The data is applicable to the problem and up-to-date. *Is the data relevant? Is it obsolete?*

5.4.2 Data Veracity profile (DVp)

Figure 5-4 illustrates the template created using the eleven L3 dimensions of the DVm, which enables the capture of findings when the dataset being evaluated is assessed against each L3 dimension.

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		EM-DAT	Action(s)
L2	L3		
Complete	1. No omitted entries		
			•
Complete	2. No omitted values		
			•
Complete	3. No omitted variable		
			•
Complete	4. No omitted metadata		
			•
Uncluttered	5. No irrelevant entries		

Expository Dimensions			
Data Veracity Sub-Dimensions		EM-DAT	Action(s)
Precise	6. Reliability		
			•
Precise	7. Rigour		
			•
Precise	8. Congruity		
			•
Accurate	9. Conformity		
			•
Accurate	10. Impartiality		
			•
Accurate	11. Validity		
			•

Figure 5-4: Data Veracity Profile – Template

The DVp allows reasonably detailed notes to be made for the dataset of interest as it is assessed against the L3 dimensions of the DVm. This information provides a base of reference when attempting to interpret any analysis of the data and provides veracity 'meta-data' if the dataset is used by others. A usable DVp template with notes can be found in *Appendix A.1*.

5.4.3 Data Veracity index (DVi)

The Data Veracity index represents the strength of a dataset's veracity in numeric terms and is calculated bottom-up from the lowest level dimensions of the DVm. Each L3 dimension is assigned a weighting that can be tailored to the relative importance placed on each dimension per dataset or across a suite of datasets. The L3 dimensions can then be scored to reflect how closely the dataset fits the dimension. *Figure 5-5* illustrates the DVm alongside the scale of weightings (i.e. 1 – 3) and scores (i.e. 0 – 5) assignable to the L3 dimensions.

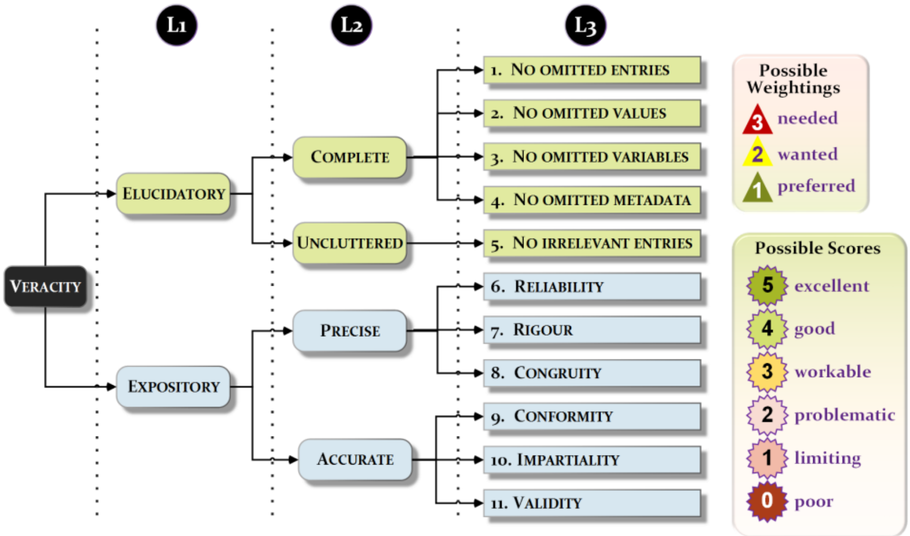


Figure 5-5: Data Veracity index (DVi)

The calculation of an L3 DVi is described in *Figure 5-6*:

(1) Calculate a ***weighted score*** by multiplying the L3 dimension's score by its weighting: $L3\ Dimension\ Score\ \times L3\ Dimension\ Weighting$

(2) Calculate a ***weighting context*** by dividing the L3 dimension's weighting by the sum of all L3 weightings: $\frac{L3\ Dimension\ Weighting}{\sum_1^{11} L3\ Dimension\ Weightings}$

(3) Calculate the ***index*** by multiplying the ***weighted score*** by the ***weighting context***:

$$[L3\ Dimension\ Score\ \times L3\ Dimension\ Weighting\ ] \times \left[\frac{L3\ Dimension\ Weighting}{\sum_1^{11} L3\ Dimension\ Weightings}\right]$$

Figure 5-6: Calculating an L3 DVi

The rationale applied in these relatively simple calculations being:

- The ***weighted score*** reflects the relative importance placed on each dimension. For example:

  Dataset A is a small dataset of deaths during surgery in one hospital.

  Dataset B is a dataset holding millions of Tweets about the same hospital.

  The L3 dimension of '*reliability*' is assigned a weighting of ***3*** (*needed*) in *Dataset A* and ***1*** (*preferred*) in *Dataset B*, and scored as ***2*** (*problematic*) as a result of an evaluation of both datasets. Therefore the ***weighted score*** is:

  Dataset A: $\boldsymbol{L3\ Score\ \times L\ Weighting} = 2\ \times 3 = \mathbf{6}$

  Dataset B: $\boldsymbol{L3\ Score\ \times L\ Weighting} = 2\ \times 1 = \mathbf{2}$

- The ***weighting context*** places the weighting for each dimension in context with the overall importance of data veracity for the dataset. Extending the example above:

  If the <u>sum of all of the weightings</u> for:

  Dataset A is **31** (**10** L3 dimension weighted **3** and **1** L3 dimension weighted **1**)

  Dataset B is **13** (**10** L3 dimension weighted **1** and **1** L3 dimension weighted **3**)

Dataset A: $$\left[\frac{\boldsymbol{L3\ Weighting}}{\boldsymbol{\sum_1^{11} L3\ Weightings}}\right] = \frac{3}{31} = \mathbf{0.1}$$

Dataset B: $$\left[\frac{L3\ Weighting}{\sum_1^{11} L3\ Weightings}\right] = \frac{1}{13} = 0.08$$

  That is, '*reliability*' may be equally as important as '*rigour*' in data involving deaths, and equally less of an expectation as '*rigour*' in a social media feed.

- The ***index*** then places the relative importance of the L3 dimension in context with overall expectations from the dataset (i.e. the total of weightings). Returning to the example:

Dataset A: $$[\boldsymbol{L3\ Score\ \times L3\ Weighting}] \times \left[\frac{\boldsymbol{L3\ Weighting}}{\boldsymbol{\sum_1^{11} L3\ Weightings}}\right] = \mathbf{6\ \times 0.1 = 0.6}$$

Dataset B: $$[\boldsymbol{L3\ Score\ \times L3\ Weighting}] \times \left[\frac{\boldsymbol{L3\ Weighting}}{\boldsymbol{\sum_1^{11} L3\ Weightings}}\right] = \mathbf{2\ \times 0.08 = 0.16}$$

Each higher level DVi, i.e. *L2*, *L1* and *dataset*, is calculated as the mean of its subordinate L3 dimensions. The range of the DVi varies depending on the level as shown below [*Table 5-1* & *Figure 5-7*]. The permutations of weightings and scores that result in these maximums can be found in *Appendix H.1*.

DVi	MIN	MID	MAX
L3	0	1.73	3.46
L2	0	1.185	2.37
L1	0	1.07	2.14
Dataset	0	0.68	1.36

Table 5-1: DVi Levels and their Min-Max ranges

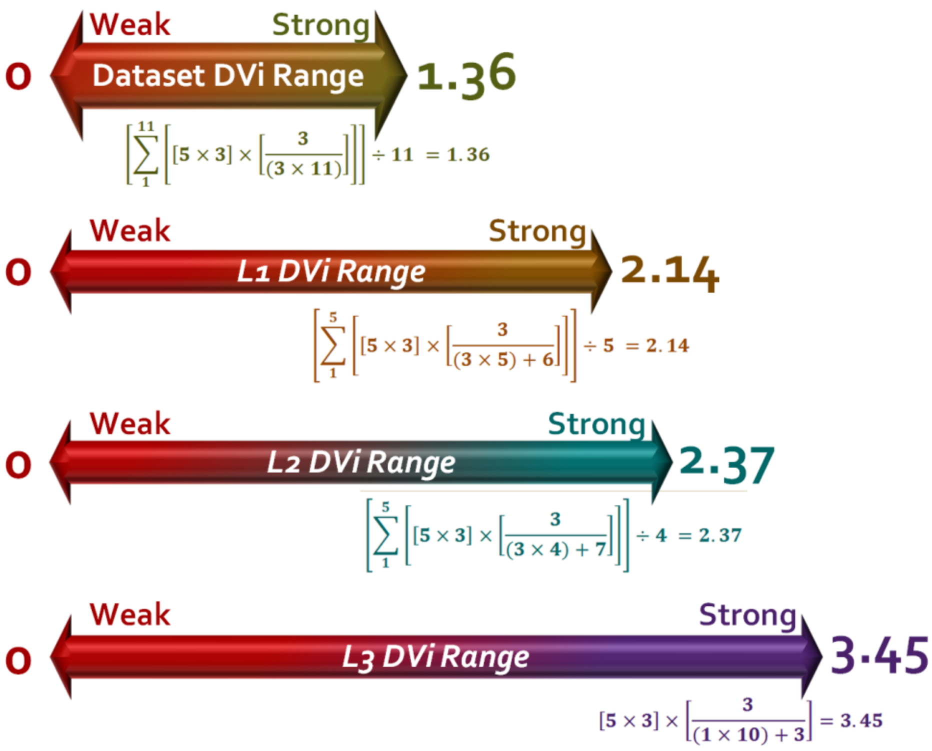
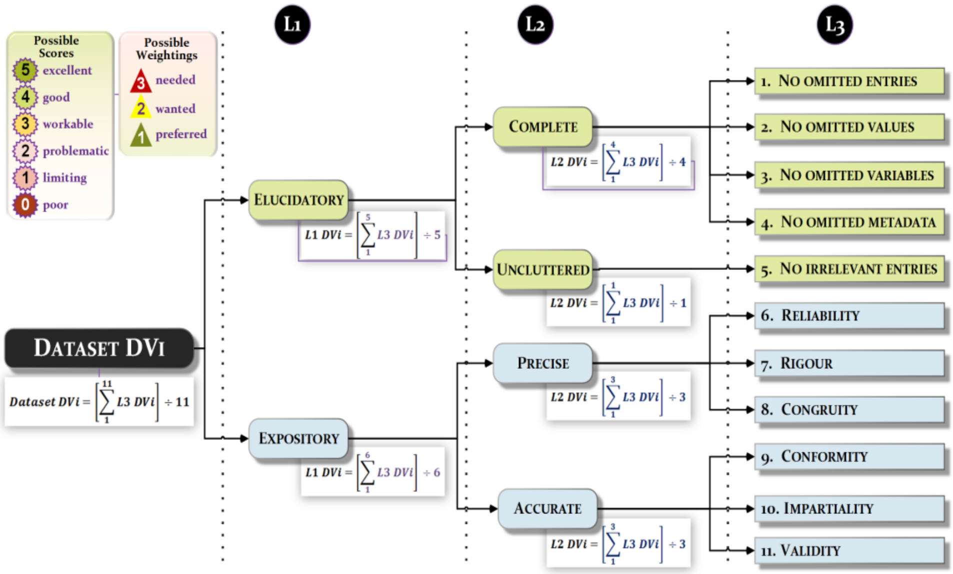


Figure 5-7: DVi Scale per DVm Level

A final note on DVi calculations, when a dataset is created by amalgamating multiple DVf evaluated datasets, the approach taken here is to err on the side of caution and base the DVi values of the amalgamated dataset on the weakest DVi of each of the eleven L3 dimensions of the contributing datasets.

The diagram below summarises the salient calculations used for the different DVi levels [*Figure 5-8*]:



$$L3\ DVi = [L3\ Dimension\ Score \times L3\ Dimension\ Weighting\,] \times \left[\frac{L3\ Dimension\ Weighting}{\sum_1^{11} L3\ Dimension\ Weightings}\right]$$

$$Amalgamated\ Dataset\ DVi = \left[\sum_{1}^{11} \bigwedge_{1}^{n} L3\ DVi\right] \div 11$$

Where:

n = the number of contributing datasets

L3 DVi = the DVi value of equivalent L3s in each dataset

Figure 5-8: DVi Key Calculations

In building the DVi indexing mechanism it is acknowledged that data veracity has to be viewed in context. For example, if 'noise' and clutter is a less crucial dimension in one dataset (e.g. a dataset of tweets, where clutter is a given) than another (e.g. records of deaths) then this dimension can have different weightings within the same study. The indices are relative calculations, therefore will adjust accordingly. Finally, to allow automated calculation of indices when L3 dimension scores are assigned, the DVi is made manifest using a spreadsheet template [*Figure 5-9*]. *Note that even though the minimum (weakest) scores of L3 dimension are used as the basis of the overall DVi for the amalgamated dataset, columns are included to show the results if maximum, median and mean scores are used.*

Veracity Dimensions				Weighting	Dataset 1			Dataset 2			Dataset *n*			OVERALL				
L1	L2		L3		Reasoning for Score	Score	Index	Reasoning for Score	Score	Index	Reasoning for Score	Score	Index	Score	Min Index	Max Index	Mid Index	Mean Index
Elucidatory	Complete	1	No omitted entries	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		2	No omitted values	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		3	No omitted variables	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		4	No omitted metadata	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
			Complete				0.0			0.0			0.0		0.0	0.0	0.0	0.0
	Uncluttered	5	No irrelevant entries	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
			Uncluttered				0.0			0.0			0.0		0.0	0.0	0.0	0.0
	Elucidatory Index						0.0			0.0			0.0		0.0	0.0	0.0	0.0
Expository	Precise	6	Reliability	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		7	Rigour	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		8	Congruity	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
			Precise				0.0			0.0			0.0		0.0	0.0	0.0	0.0
	Accurate	9	Conformity	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		10	Impartiality	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
		11	Validity	1			0.0			0.0			0.0	0	0.0	0.0	0.0	0.0
			Accurate				0.0			0.0			0.0		0.0	0.0	0.0	0.0
	Expository Index						0.0			0.0			0.0		0.0	0.0	0.0	0.0
Veracity Index							0.0			0.0			0.0		0.0	0.0	0.0	0.0

Figure 5-9: DVi Spreadsheet

5.5 Evaluate

Evaluation of this iteration is cast in the same mould as the equivalent design cycle step of the previous iteration, with the focus now on the data veracity toolset needed to enable this study to progress (Hevner et al., 2004; Simon, 1996; Hooker, 2004; Venable et al., 2012; Venable, 2013; Gregor, 2006; Vaishnavi and Kuechler, 2004b). As such this section will discuss the alignment of the *build (grow)* step with the *tentative design*; the *DSR artefacts* created in this iteration; the *knowledge* gained and the *consequences* of this knowledge; and finally test the propositions of the *design theory* or *theories* relevant to this iteration.

5.5.1 Tentative Design⇔Build (Grow) Alignment

The *build (grow)* step closely aligns to the *tentative design* outlined for this iteration. In summary, the build (grow) step follows the design brief of creating a DVf of three interrelated parts – a DVm to describe the anatomy of data veracity; a DVp to capture descriptive knowledge of a dataset's veracity; and a DVi to provide a numeric measure of a dataset's veracity that can be compared across datasets [*Figure 5-10*].

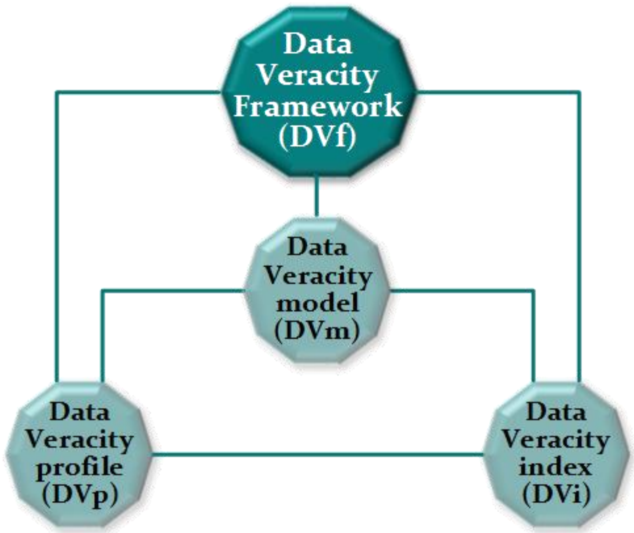


Figure 5-10: Constituent Parts of the DVf – Designed and Built

5.5.2 DSR Artefacts

The artefacts created in this iteration are in addition to those anticipated at the start of this study [*Table 3-5*]. *Table 5-2* maps these additional DSR artefacts to the research framework discussed in *Section 3.3.3* (March and Smith, 1995; Vaishnavi and Kuechler, 2004b; Hevner, 2007).

Research Framework			Research Activities			
			Design Science		Natural Science	
			Build (grow)	Evaluate	Theorise	Justify
Research Outputs	Artefacts	Constructs	[a] →		[c] - - ->	
		Models	[d] →			
		Methods	[e] →			
		Instantiations	[b] →			
[a]	Macro-indicators of disaster outcome and the impact and effectiveness of humanitarian intervention (M^iOs, M^iIs and M^iEs).					
[b]	Data analysis outputs and visualisations.					
[c]	A (behavioural science) hypothesis relating the availability, or lack thereof, of humanitarian data and the flow of humanitarian aid that emerges from the domain knowledge and may be worthy of future research.					
[d]	**Data Veracity framework (DVf) and Data Veracity model (DVm)**					
[e]	**Data Veracity profile (DVp) and Data Veracity index (DVi)**					

Table 5-2: DSR Output to Research Framework Mapping v.2
(Vaishnavi and Kuechler, 2004b; Hevner, 2007; March and Smith, 1995)

The research artefacts created in this iteration have only been tested using manufactured sample data for the DVi, as described in the examples used to explain the calculations in *Section 5.4.3*. A fuller evaluation can only really be carried out when the DVf is used in earnest to evaluate real datasets, which is planned for the next iteration of this research. That said, two areas of weakness can be identified here:

(1) There is no formal method to define the context of the evaluation.

(2) Nuanced differences in the L3 dimensions may lead to varying interpretations.

Regarding the first of these, it is believed that the DVf would benefit from a method of placing any DVps and DVis created in context with the scope of the study to be completed. For example, the veracity evaluation of a disaster dataset may be different depending on whether the scope of a study is deaths caused by *natural* disasters 2010 – 2015, or losses from all disasters since the beginning of the last century. This is not an issue for this work as any DVps and DVis created are within the scope of this research. That said, the absence of a context-setting artefact may prove problematic and should be remedied if these artefacts are shared as 'metadata' beyond this work.

As for the second issue, some of the more nuanced L3 dimensions, e.g. conformity, congruity impartiality etc., unless fully and uniformly understood may result in varying interpretations resulting in incompatibility across shared DVps and DVis. Once again, this is not a problem here as all DVps and DVis are to be completed by the same researcher. If the DVf is to be used by other works, more detailed definitions, examples and training materials are needed.

5.5.3 Knowledge ⇔ Consequence

Nuggets of knowledge that emerge from this iteration include:

(a) There is no standard definition for data veracity

Data veracity is typically expressed in terms that describe '*what it is not*' rather than '*what it is'* e.g. ***not*** uncertain, biased, noisy etc. and even these terms are vague and can vary amongst commentators (Normandeau, 2013; Berti-Equille and Lamine Ba, 2016; Berti-Equille and Borge-Holthoefer, 2015; Lukoianova and Rubin, 2014; Claverie-Berge, 2012; Schroeck, 2012).

(b) There is no standard method to evaluate data veracity

An extensive search for an existing accepted method of evaluating data veracity yielded no result. This is unsurprising as it is not clear how a characteristic can be assessed if it is not defined. This raises the question of how do other initiatives that are focussed on the study of data assure themselves and others of the veracity of the data they use. A declaration that outputs can be relied upon simply because of the absence of data that is '*in doubt*', '*noisy*' or '*biased*' lacks the substance to be reassuring.

(c) THREE aspirational Vs for a 3-dimensional view of data

The search for an accepted definition of data veracity involved scouring numerous publications relevant to the study of data for answers. While none of these sources offered complete answers, a subset from data quality and Big Data did contribute usable snippets of knowledge that assisted in the construction of the bespoke DVf of this work (Strong et al., 1997; Lee et al., 2002; Wang and Strong, 1996; Wand and Wang, 1996; Laney, 2013; Grimes, 2013; Normandeau, 2013; IBM, 2013).

An interesting additional realisation emerged from material in the subject area of Big Data is that there may be potential in developing frameworks for three **aspirational qualities** of all data, namely ***veracity***, ***virtue*** and ***value*** (Laney, 2001, 2013; Normandeau, 2013; Grimes, 2013; Floridi and Taddeo, 2016; Boyd and Crawford, 2012; Glikman and Glad, 2015; Baldwin, 2015). It is believed a 3-dimensional aspirational view of data (of any magnitude) used for analysis is an invaluable perspective that can aid fully realising the worth of that data [*Figure 5-11*]. Even achieving this 3-D view is aspirational because currently there are no universally accepted definitions for these characteristics, let alone mechanisms that can be used to evaluate datasets along these three dimensions.

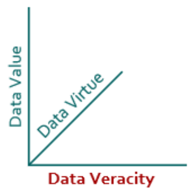


Figure 5-11: The Three Aspirational V-Dimensions of Data

Table 5-3 maps the nuggets of knowledge from this iteration to the consequence of that knowledge on this study and, where applicable, on future opportunities for research.

Knowledge	Evaluate	Consequence
(a) There is no standard definition for data veracity	Preparation and examination of EM-DAT has already raised significant questions of data veracity. This work needs to amalgamated disaster loss datasets to obtain a more complete view of humanitarian crises. A lack of visibility of the comparative veracity of each dataset contributing to an amalgamation of disaster loss would result in a blind spot as to the reliability of any outputs created using this amalgamate dataset.	A means of evaluating data veracity is considered a mandatory prerequisite to creating an amalgamated disaster dataset. Therefore a bespoke solution is built to evaluate data veracity in the form of the DVf. As the DVf is untested it carries with it the inherent risk of results that may be unexpected, inadequate, or even incorrect.
(b) There is no standard method to evaluate data veracity		
(c) THREE aspirational Vs for a 3-dimensional view of data	Data ***veracity*** is one of <u>**three aspirational qualities of data**</u>, the other two being ***virtue*** and ***value.*** These qualities are referred to in literature, but as yet have escaped definition or a means of evaluation. While ***veracity*** is of particular interest to this work, it is believed that assessment of ***value*** and ***virtue*** would provide a 3- dimensional understanding of the data.	While there may be benefit in evaluating the ***value*** and ***virtue*** of the datasets the lack of standard definitions and evaluation mechanisms is problematic, thus not pursued any further here. Developing a means of evaluating data veracity is on the critical path to completing this study, Developing complementary solutions for the other two aspirational qualities is beyond the immediate needs, resources and scope of this research. The development and application of complementary frameworks for all three aspirational qualities of data is considered an opportunity for future research.

Table 5-3: Iteration 2 Knowledge ⇔Consequence Mapping

5.5.4 The Utility Theory

It is not possible to test the propositions of the utility theory that underpins this study as the DSR artefacts created in this iteration are prerequisites to other research artefacts that contribute to the theory. This aspect is thus deferred to the chapters that follow.

5.6 Summary

This chapter discusses iteration 2 and the realisation and addressing of the absence of a means to systematically evaluate data veracity. The chapter discusses the importance of understanding the veracity of data used for analysis, arguing that if the trustworthiness of data is unknown the reliability of any output created using that data is also unknown. The chapter goes on to describe the *Tentative Design* of the Data Veracity framework (DVf) needed for this study. Explaining the three interrelated parts required for the DVf to come into existence, namely: the Data Veracity model (DVm) to explain the *anatomy of veracity*; the Data Veracity profile to descriptively capture a dataset's fit to the DVm; and the Data Veracity index (DVi) to quantify the scale of a dataset's fit to the DVm.

The *Build (Grow)* step outlines each of the three constituent parts of the DVf. Describing the hierarchy and dimensions of the DVm. Providing the DVp as a template to capture descriptive knowledge from the evaluation of dataset against the dimensions of the DVm. Explaining in some detail the rationale and calculations of the DVi, the creation of which is made available for use in the form of a spreadsheet template.

Finally, the discussion moves to the *Evaluate* step. Here it is confirmed that the creation of the DVf aligns to the intent and expectations of the tentative design. This DSR design cycle step then goes on to evaluate the created artefacts, but before doing so highlights that as these artefacts were not originally envisaged or discussed. It therefore revisits the research framework originally presented in *Section 3.3.3* to place these ‘new’ artefacts in context with the expected outputs of this study (March and Smith, 1995; Vaishnavi and Kuechler, 2004b; Hevner, 2007). The evaluation of the research artefacts from this iteration determines that the DVf suite created here may be adequate for the purposes of this study but will need further development to allow the capture of context and minimise risk of misinterpretation if it is to be used by others. Evaluation of this iteration moves on to summarise the key nuggets of knowledge gathered from the data science research domain and the consequences of this knowledge on this study and, if applicable, future opportunities for research.

Chapter 6: MORE DISASTERS (ITERATION 3)

6.1 Overview

This chapter describes and discusses Iteration 3 of the DSR design cycle for this research. The chapter returns the research to its original path of exploring credible disaster loss data to identify macro-indicators of outcome (M^iOs) by first creating then using the *data scaffolds* of a *master set of global disasters* and a corresponding *master disaster classification model*. Notably, the *master set of global disasters* is created from datasets that are first evaluated using the *DVf* developed for this purpose in Iteration 5. *Figure 6-1* is a basic schematic of the flow of Iteration 3.

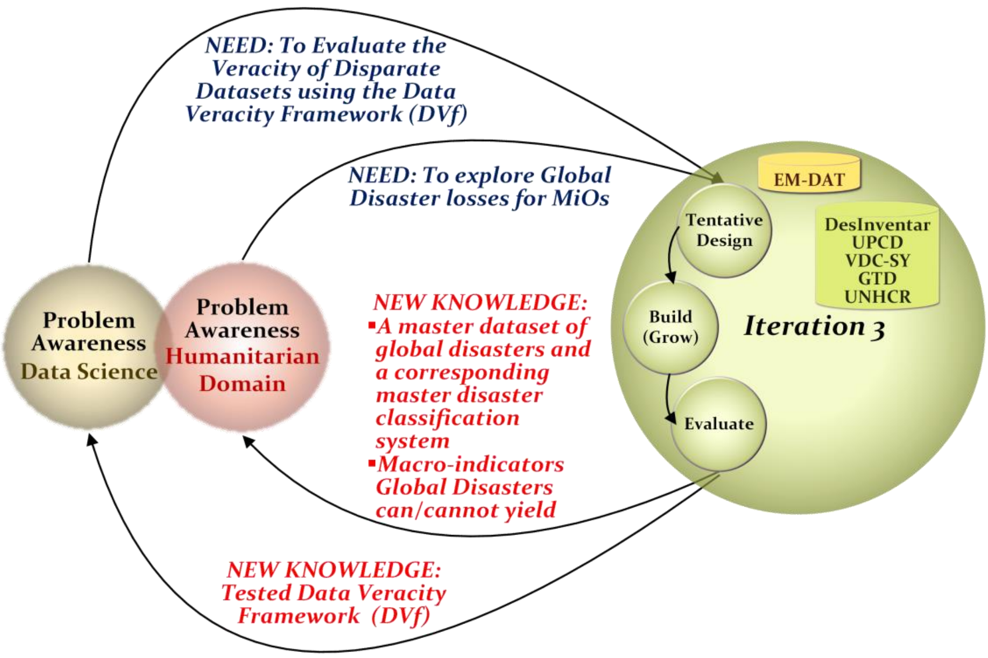


Figure 6-1: Iteration 3 of the Design Cycle

The structure of the chapter is as follows: *Section 6.2* discusses the problem addressed by this iteration which is a continuation and expansion of the problem space of Iteration 1 [*Chapter 4*] and reaffirms the need for key data scaffolds. *Section 6.3* discusses the *Tentative Design* of the master set of global disasters and the master disaster classification system, and follows up on the potential identified in Iteration 1 of Survival Rate as a possible M^{i}O. *Section 6.4* describes the *Build (Grow)* step to create the master set of global disasters, the master disaster classification system, and explore the newly amalgamated disaster dataset for M^{i}Os. *Section 6.5 Evaluates* the work and outputs of Iteration 3 before *Section 6.6* closes with a summary of the chapter.

6.2 Problem Awareness

The problem space for this iteration is a continuation and expansion of the Iteration 1 [*Chapter 4*] problem space, which could not be addressed because of the absence of the key data scaffolds of a *master set of global disasters* and a corresponding *master disaster classification* model. The term *'data scaffold'* is used here to encompass any artefact that provides context, frame, form or support to exploit data for actionable insights, enlightening discoveries or pre-emptive predictions relevant to a domain. Examples of such artefacts would include data frameworks, data governance models, data models, master datasets, taxonomies, ontologies etc. that allow a common understanding and a shared perspective of the domain.

Inherent in most sectors are a proliferation of enterprise-wide transactional systems that force some congruence in the creation, coverage, management and quality of key datasets, e.g. product masters, supplier masters, patient profiles, customer/client masters, general ledgers, inventory levels etc.. The lists of such datasets may vary depending on the sector, but their existence plays a significant role in facilitating the analysis of, and insights from, the burgeoning oceans of data being generated (Cheung and Chung, 2014; McGilvray and Thomas, 2008). Where challenges of integration, compatibility and interactivity exist many more mature sectors have responded with data standardisation, data architectures and data governance to facilitate consistency and interoperability between islands of automation and datasets (Data Foundation, 2016; Otto, 2011; Hoven, 2003), together with taxonomies and the occasional use of ontologies (Smith, 2003; Hlava, 2012).

For the humanitarian sector a master dataset of global disasters and a comprehensive disaster classification model as '*data scaffolds*' is an elemental expectation. Unfortunately, while there are ongoing initiatives to create and populate datasets of various disasters, there is no unified master for disaster data (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; UNHCR, 2017c; GTD, 2017e; VDC-SY, 2016b; ADRC, 2017; GLIDE, 2017; ReliefWeb, 2017; Guha-Sapir and Below, 2002; Velasquez et al., 2002). Similarly, a unifying disaster taxonomy has yet to be found (LOG, 2017; DesInventar, 2017c; JHSPH, 2008; Duffield, 1994; John and Thangamani, 2015; IRDR, 2014; Below et al., 2009; Lerner, 2016).

The lack of such fundamental artefacts may be an idiosyncrasy of the humanitarian domain and linked to the unusual characteristics of humanitarian supply networks. As the recipients of aid do not pay for what they receive there is no irrefutable need to reconcile demand satisfaction and monies. This appears to have the effect of removing the requirement for integrated transactional activity across the entire network that would typically allow data to be captured as a part of a mandatory process. Thus in the humanitarian domain, collecting data related to the effect and response to disasters is an additional task exclusively for reporting and justification purposes, which can be considered an 'unnatural act' rather than a naturally occurring by-product of transactional activity.

As a result, the humanitarian sector, which exists solely to address humanitarian crises, lacks the '*data scaffolds*' necessary to provide an integrated source of **<u>most</u>** disasters and a classification system for **<u>most</u>** disaster types. Such a suite of artefacts does not appear to be available even from the international bodies monitoring humanitarian funding and for whom these artefacts would provide better context to the flow of aid. As a result, for the work here to continue, these two basic structures need to be created. Therefore to search for M^{i}Os in disaster loss data this iteration must first:

(1) Evaluate the veracity of a select set of disparate disaster loss datasets before amalgamating them to create a master set of global disasters.
(2) Construct a master disaster classification system to correspond with the disasters in the master set of global disasters.

6.3 Tentative Design

The *Tentative Design* for this iteration includes:

(1) The humanitarian crises datasets selected to be included in the Master Set of Global Disasters (MSGD).
(2) The structure of the Master Set of Global Disasters (MSGD).
(3) The design principles and overall structure to be applied in constructing a Master Disaster Classification model (MDC).
(4) The design principle applied in searching for M^{i}Os.

6.3.1 MSGD Sources

Six disaster loss datasets are identified as appropriate candidates for the MSGD, each to be incorporated into a predefined structure for the new dataset:

(1) Emergency Events Database (EM-DAT) for humanitarian crises most of which are categorised as *natural* or *technological* disasters (Guha-Sapir et al., 2017l);
(2) Disaster Inventory System (DesInventar), which holds disaster types that are believed to be of a similar ilk to those in EM-DAT, but unlike EM-DAT, DesInventar does not restrict the data collected based on an inclusion criteria. Additionally, DesInventar is an open source solution that is deployed in multiple countries or districts, therefore it is a collection of

databases and not centrally curated (DesInventar.NET, 2017; DesInventar.ORG, 2017);

(3) Uppsala Conflict Data Program (UCDP), which provides a perspective of conflict-type global humanitarian crises, excluding those in Syria, and the losses experienced because of these crises (UCDP, 2017b; UCDP, 2017a);

(4) Violations Documentation Center in Syria, which plugs the gap in UCDP by providing data of the human effect of the conflict in Syria **(VDC-SY)** (VDC-SY, 2016b);

(5) Global Terrorism Database (GDT), which provides a perspective of global terrorism-type humanitarian crises and the losses caused by these crises (GTD, 2017e);

(6) United Nations High Commissioner for Refugees (UNHCR), which provides a view of people uprooted from their homes and/or their countries (UNHCR, 2017c).

The reasons for selecting EM-DAT for this research are described in *Chapter 4*. Even if EM-DAT on its own is not sufficient for the purposes of this research it is still a significant and well-perceived source of disaster loss data (Voigt et al., 2016; Guha-Sapir et al., 2017j; Toya and Skidmore, 2007; Kourosh and Richard, 2008; Strmberg, 2007; Alcántara-Ayala, 2002; Pears-Piggott and Muir-Wood, 2016; Blaikie et al., 2014; Sodhi, 2016; Corey et al., 2016; Raschky and Schwindt, 2016). The rationale for selecting each of the other data sources – **(2)** to **(6)** in the list above – can be found in *Appendix C.1*. It should be noted that in building the MSGD there is no expectation of a complete and definitive view of all disasters losses caused by all disaster types, the best that is hoped for is a *more complete* view of *most disaster losses* associated with *most humanitarian crises*.

6.3.2 MSGD DVi Weightings

The L3 dimension weightings for all datasets to be evaluated for the MSGD are shown below [*Figure 6-2*]:

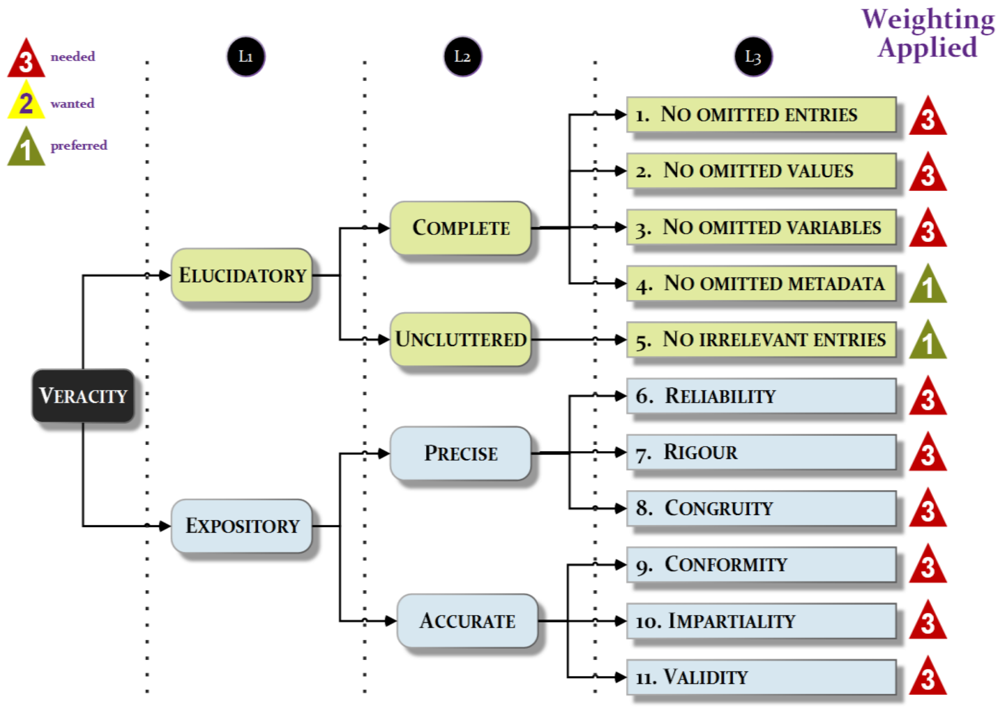


Figure 6-2: Veracity Profile DVi – Applied Weightings

All bar two L3 dimensions are weighted ***3*** (*needed*). The two that are assigned a weighting of ***1*** are: '***4.*** *no omitted metadata*' and '***5.*** *no irrelevant entries*'. These are considered manageable weaknesses that have workarounds e.g. deduction or third party literature for ***L3 dimension 4*** and exclusion of spurious entries for ***L3 dimension 5***. They are also part of the ***L1 Elucidatory Dimension***, which is by and large the aspect of disaster loss data weakness addressed in this study by amalgamating multiple datasets, whereas there is currently no known workaround for data that is imprecise or inaccurate.

6.3.3 MSGD Structure

Each of the datasets selected as contributors to the MSGD varies in structure and level of detail. Regardless of the originating database structure, the disparate source data needs to be made compatible and comparable. The variables in the MSGD are as follows [*Table 6-1*]:

MSGD_ID	A unique sequential identifier. Created as a key to uniquely identify each entry in the MSGD.
SOURCE_ID	Identifies the name of the original source **dataset** for the entry e.g. EMDAT3, EMDAT4, DesInventar Laos, GTD etc.
SOURCE_DisNo	The unique identifier used in the original dataset. If the original dataset does not have a unique identifier this value is created as an extra variable in the source data. This variable is used to link the MSGD to each source dataset.
YEAR	The year relevant to the event. (Copied from source.)
ISO_alpha3	The three character ISO code used to identify the country in which the event occurred or, in the case of refugee movement, country of origin (ISo-3166, 2017).
Location	A more specific location of the event. (Copied from source.)
MDC_ID	The link to the **Master Disaster Classification** (**MDC**). (Created in parallel with the MSGD)
Total Deaths	Deaths cause by the event. (Copied/calculated from source.)
Total Affected	People affected by the event. (Copied/calculated from source.)
Total Human Effect	$Total\ Deaths\ + Total\ Affected$
Survival Rate	$\left[\frac{Total\ Affected}{Total\ Human\ Effect}\right] \times 100$ Acknowledging the knowledge consequence from *Section 4.5.3*.
US$Loss_2015 (000s)	Highest financial losses recorded adjusted to 2015 levels based of USA CPI (BLS, 2016).
Soft Total Deaths	$= Total\ Deaths$ if DVp evaluates weak veracity
Soft Total Affected	$= Total\ Affected$ if DVp evaluates weak veracity
Soft US$Loss_2015 (000s)	$= US\$Loss\_2015$ if DVp evaluates weak veracity
ISO_alpha3 (Destination)	The three character ISO code of the country where refugees settled or seek asylum (ISo-3166, 2017).

Table 6-1: MSGD variables

Note: The values contained in these variables are copied, created or calculated where applicable, and as needed, depending on what is available in the source dataset and the outcome of evaluating its veracity.

6.3.4 Master Disaster Classification (MDC)

There is no readily available disaster classification system that encompasses all crises events that are in receipt of humanitarian aid. The classification model used for EM-DAT falls short of the needs of this work as it does not include crises situations such as conflicts and deracination [*Appendix B.1.1 & B.1.2*] (Guha-Sapir et al., 2017g). Therefore, a suitable *Master Disaster Classification* (MDC) model must be created here to suit the needs of this research. A pragmatic design approach is taken to the creation of the MDC, which here only needs to be sufficiently inclusive to classify all events held in the source datasets selected for the MSGD.

For humanitarian crises that are 'natural' in origin, the Integrated Research on Disaster Risk (IRDR) Peril Classification and Hazard Glossary (IRDR, 2014) is used. IRDR is a decade-long research programme sponsored by a number of international organisations including the UNISDR (IRDR, 2017; UNISDR, 2017). A benefit of using IRDR is that both EM-DAT and DesInventar 'natural' disaster classifications purport to be based on it (IRDR, 2017; Guha-Sapir et al., 2017i; DesInventar.NET, 2017; DesInventar.ORG, 2017), albeit examination of EM-DAT in *Section 4.4.1* revealed EM-DAT does not strictly adhere to IRDR classifications [*Appendix B.1.3*] (IRDR, 2014; Guha-Sapir et al., 2017g; DesInventar, 2017c). For humanitarian crises that are not '*natural*', such as conflict, terrorism and deracination, a UN study of "*alternative classification schemes for man-made hazards*" (Lerner, 2016), and a "*taxonomy of threats*" developed by the Centre of Risk Studies, University of Cambridge (Coburn et al., 2014) are used as guides.

The structure of the created MDC falls into three groups *Naturogenic*, *Anthropogenic* and *Deviant*, below which disaster classifications are structured hierarchically into levels akin to IRDR levels: *Family* → *Main Event* → *Peril* (IRDR, 2014).

- ***Naturogenic***, where the root of the event lies in nature. These are '*natural*' disasters in EM-DAT parlance, but rather than considering disasters to be '*natural*', this recognises the source of the crisis is '*natural*'. Where possible IRDR classifications are used for these, but if a '***naturogenic***' event does not map cleanly to IRDR, the name in the source dataset is used as a guide to creating a new entry in the '***naturogenic***' group (IRDR, 2014).
- ***Anthropogenic***, where the root of the crises is the result of an unintended human action e.g. accident, technical error, carelessness etc.
- ***Deviant***, these are humanitarian crises caused by deviant – *departing from usual or accepted standards or expectations* – human acts, where intent or apathy, cause unacceptable outcomes for people and society, e.g. war, terror, migration, famine etc.

For ***'anthropogenic'*** and '***deviant***' events, inspiration for the MDC model comes from the source disaster dataset and, where appropriate, is modified or enhanced using the UN study of "*alternative classification schemes for man-made hazards*" (Lerner, 2016) or the "*taxonomy of threats*" developed by the Centre of Risk Studies, University of Cambridge (Coburn et al., 2014).

6.3.5 Macro-Indicators

The metric of interest for this iteration of the DSR design cycle is the macro-indicator of outcome (M^iO). As with Iteration 1, no prescriptive definition of what may or may not constitute M^iOs is provided in the design step, as M^iOs are expected to emerge from improved knowledge of the data. That said, the interesting behaviour of EM-DAT *Survival Rate* in Iteration 1 is acknowledged in the creation of the MSGD, where it is a calculated variable for all entries. Moreover, observation of the behaviour of *mean survival rate* in Iteration 1, also make it a key candidate for exploration in this iteration once the MSGD is created.

6.4 Build (Grow)

This step of the iteration creates the Master Set of Global Disasters (MSGD), testing the Data Veracity framework (DVf) with each of the disaster datasets that contribute to the MSGD (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017c). The Master Disaster Classification (MDC) model and associated reference dataset are constructed in parallel to the MSGD. On completion of the MSGD and MDC, the MSGD is explored for Master Indicators of Outcome (M^iOs). In summary, this build step will create the following artefacts:

- The MSGD, an amalgamation of six humanitarian crises datasets.
- Data Veracity profiles (DVps) and Data Veracity indices (DVis) for each of the six datasets amalgamated to create the MSGD, and a composite DVi of the MSGD (DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017c).
- An MDC model that at a minimum classifies events and perils held in the MSGD.
- An MDC reference dataset as a physical manifestation of the created MDC model, with a unique identifier for each entry that allows the MSGD to link to it.
- Summaries, models and visualisations using the newly built MSGD to search for Macro-indictors of Outcome (M^iOs).

6.4.1 Master Set of Global Disasters (MSGD)

The sequence of populating the MSGD is as follows:

(1) **EM-DAT**, building on the work from Iteration 1 [*Chapter 4*]. This dataset of mostly naturogenic and anthropogenic disasters is the first to be added to the MSGD (Guha-Sapir et al., 2017l).

(2) **DesInventar** for ‘locally’ collected granular disaster loss data (DesInventar.NET, 2017; DesInventar.ORG, 2017).

(3) **Uppsala Conflict Data Program** (UCDP) for conflict data (UCDP, 2017b; UCDP, 2017a).
(4) **Violations Documentation Center in Syria** (VDC-SY) for data relevant to the Syrian conflict currently missing from the UCDP (VDC-SY, 2016b; Croicu and Sundberg, 2015).
(5) **Global Terrorism Database** (GTD) for disaggregated terrorist acts (GTD, 2017e).
(6) **United Nations High Commissioner for Refugees** (UNHCR) Persons of Concerns for communities uprooted from their homes (UNHCR, 2017c).

A DVp and DVi is completed for each dataset before it is added to the MSGD, thus allowing the entries with weaker veracity to be flagged in the MSGD.

EM-DAT ➔ MSGD

Details of the acquisition, preparation and examination of EM-DAT are in Iteration 1 [*Chapter 4*] (Guha-Sapir et al., 2017l). Here reference is made to a structured evaluation of its veracity using the DVp and DVi templates of the DVf created in Iteration 2 [*Chapter 5*]. A completed DVp of EM-DAT can be found in *Appendix A.2*. This EM-DAT DVp provides the evaluated findings, and the response to these findings, against the dimensions defined in the DVm hierarchy. EM-DAT's DVi scores and indices are shown here [*Figure 6-3*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions			Weighting	EM-DAT		
L1	L2	L3		EM-DAT	Score	Index
Elucidatory	Complete	1. No omitted entries	3	Not all years and disasters types held	2	0.62
		2. No omitted values	3	Very little financial loss data. 1485 missing human effect numbers	2	0.62
		3. No omitted variables	3	Less than half of the stored variables were available	2	0.62
		4. No omitted metadata	1	Information on the handling of zero or missing not available	2	0.07
			Complete			0.48
	Uncluttered	5. No irrelevant entries	1	893 entries for disasters that had no effect	2	0.07
			Uncluttered			0.07
	Elucidatory Index					0.4
Expository	Precise	6. Reliability	3	Entries are subject to change without a reason or audit trail	3	0.93
		7. Rigour	3	A high proportion of human effect is not collected but guessed	1	0.31
		8. Congruity	3	Imbalance in disaster classifications across disaster groups	2	0.62
			Precise			0.62
	Accurate	9. Conformity	3	~36% of human effect not anchored in facts	1	0.31
		10.Impartiality	3	Emphasis more on natural disaster and the most recent decades	2	0.62
		11.Validity	3	Entries are for disasters, even though a number are pointless (no effect)	3	0.93
			Accurate			0.62
	Expository Index					0.62
Data Veracity index (DVi)			29			0.52

Figure 6-3: EM-DAT DVi

EM-DAT's DVi of 0.52 is below the mid-point of 0.68 on the Dataset DVi Scale [*Figure 6-4*].

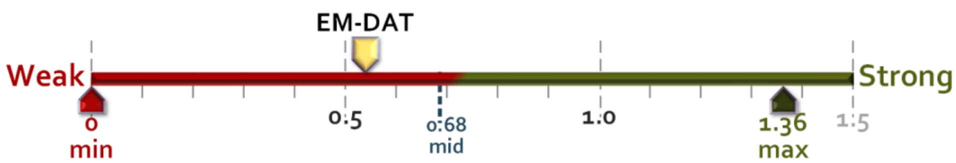


Figure 6-4: EM-DAT on the Dataset DVi Scale

Unsurprisingly, EM-DAT's indices echo its insufficiency for this research and the significant proportion of suspected guestimates for human effect. Of particular note is that EM-DAT's *Elucidatory* index ('*complete*' and '*uncluttered*') is only 0.4, which underlines the need to supplement it using other data sources for the purposes of this work. Even its *Expository* index (‘*precise*’ and ‘*accurate*’) is a relatively weak 0.62, reflecting the reliance on estimates, the lack of absolute adherence to its own inclusion criteria and the imbalance of focus and emphasis [*Figure 6-5*].

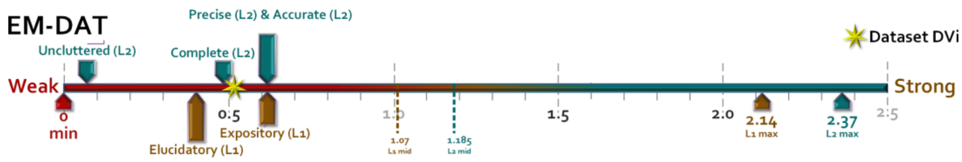


Figure 6-5: EM-DAT on the L2 & L1 DVi Scale

Having evaluated the veracity of EM-DAT, its data is added to the MSGD. The *EM-DAT → MSGD* variable mappings can be found in *Appendix D.1,* but some key logic decisions are described here:

- Financial losses are taken to be either estimated damage (tot_dam/EstDamage) or insured losses (insur_dam), whichever is the greater.
- Region is not taken from EM-DAT to the MSGD, but as EM-DAT entries can be referenced using the SOURCE_DisNo from the MSGD this information remains accessible.
- Soft Total Deaths = Total Deaths; Soft Total Affected = Total Affected; and Soft USLoss_2015 (000s) = USLoss_2015 (000s) if human effect values match EM-DAT’s estimating methods or EM-DAT’s inclusion criteria is not met.

Finally, as per the design principle discussed in *Section 6.3.4* a base MDC model for *Naturogenic* disasters is built using IRDR classifications and extended to include any incremental disaster classifications needed to support EM-DAT [*Appendix B.2*] (IRDR, 2014; Guha-Sapir et al., 2017g). The MDC is then made real as a reference dataset, with MDC_ID as its key. This allows the MDC_ID variable for each entry in the MSGD, now populated with EM-DAT data, to be updated to link to the MDC reference dataset.

DesInventar ➔ MSGD

Details of the acquisition, preparation and examination of DesInventar data can be found in *Appendix C.2* (DesInventar.NET, 2017). A detailed DVp for DesInventar is located in *Appendix A.3*, while DesInventar's DVi scores and indices are included here [*Figure 6-6*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions			Weighting	DesInventar		
L1	L2	L3		Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	High volume of data but still not complete for country, year and/or disaster groups	1	0.31
		2. No omitted values	3	Numerous missing disaster types; 'No effect' entries = 5,404; No **human** effect entries >33,000	1	0.31
		3. No omitted variables	3	Numerous databases are missing > 1 variable for human or financial effect or definition of event	2	0.62
		4. No omitted metadata	1	Reasoning for entries vary. No documentation as to how decision are made locally	1	0.03
		Complete				0.32
	Uncluttered	5. No irrelevant entries	1	1000s of spurious/duplicate entries cleared, but many remain	1	0.03
		Uncluttered				0.03
	Elucidatory Index					0.26
Expository	Precise	6. Reliability	3	Sporadic/varied updates of single databases, but GAR2015 database does not appear to be volatile.	1	0.31
		7. Rigour	3	No validation appears to be applied.	0	0.0
		8. Congruity	3	IRDR compliance < 33% of entries. Varied 'free range', often local language disaster types < 66% entries.	1	0.31
		Precise				0.21
	Accurate	9. Conformity	3	Most values suspect - lax/no validation and inconsistent usage of fields.	1	0.31
		10. Impartiality	3	Naturigenic > 78% entries and 99.7% of entries are for the period 1970 to 2017.	1	0.31
		11. Validity	3	Mostly disasters, but the databases may also have be used for other purposes.	2	0.62
		Accurate				0.41
	Expository Index					0.31
Data Veracity index (DVi)			29			0.29

Figure 6-6: DesInventar DVi

DesInventar DVi of 0.29 is almost half that of EM-DAT, and therefore very low, well below the mid-point of 0.68 on the Dataset DVi Scale [*Figure 6-7*].

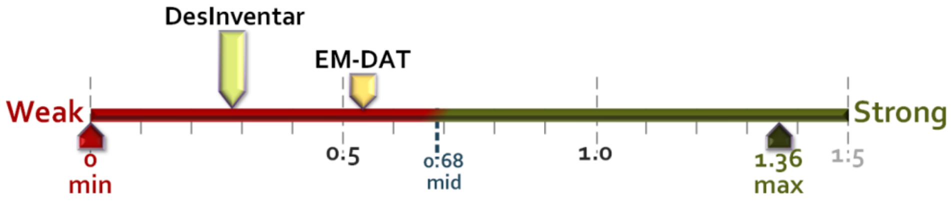


Figure 6-7: DesInventar on the Dataset DVi Scale

The lack of validation and the voluntary, sporadic and inconsistent usage of the DesInventar solution by varying countries over varying time periods also translate to very low DVi values at the L2 and L1 levels of the DVm [*Figure 6-8*].

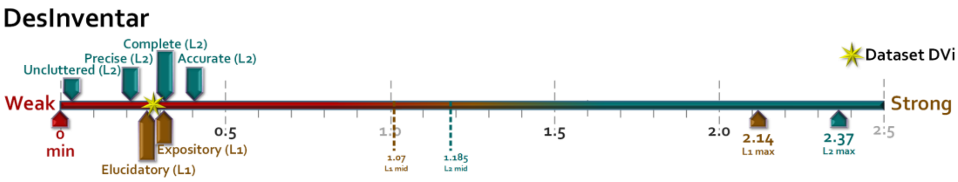


Figure 6-8: DesInventar on the L2 & L1 DVi Scale

Having evaluated the veracity of DesInventar, its data is added to the MSGD and the *DesInventar* → *MSGD* variable mappings can be found in *Appendix D.2*. DesInventar's DVi, however, is so low that it is deemed that none of its data can be considered 'firm' (veracious), therefore the following flags are set in the MSGD for all DesInventar entries:

- Soft Total Deaths = Total Deaths;
- Soft Total Affected = Total Affected; and
- Soft USLoss_2015 (000s) = USLoss_2015 (000s).

UCDP GED50 → MSGD

Details of the acquisition, preparation and examination of data, specifically UCDP GED50, from the Uppsala Conflict Data Program (UCDP) can be found in *Appendix C.3* (UCDP, 2017d; UCDP, 2017a). A detailed DVp for UCDP GED50 is located in *Appendix A.4*, while its DVi scores and indices are included here [*Figure 6-9*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions			Weighting	UCDP GED50		
L1	L2	L3		Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	No conspicuous missing entries. Syria is intentionally omitted.	4	1.24
		2. No omitted values	3	Only two entries are missing values.	5	1.55
		3. No omitted variables	3	All variables are available in the download.	5	1.55
		4. No omitted metadata	1	Full and detailed documentation is available and accessible.	5	0.17
		Complete				1.13
	Uncluttered	5. No irrelevant entries	1	Most entries are relevant; only 2 'empty' entries flagged for exclusion.	5	0.17
		Uncluttered				0.17
	Elucidatory Index					0.94
Expository	Precise	6. Reliability	3	Each version of the dataset is static and clearly identifiable. This version is up to 2015 and will not change.	5	1.55
		7. Rigour	3	Meticulously maintained,but outdated countries Yugoslavia and Soviet Union needed correction.	4	1.24
		8. Congruity	3	Only 3 classifications defined and described. Appear to be used equitably and without ambiguity.	5	1.55
		Precise				1.45
	Accurate	9. Conformity	3	best_est is firmest total fatality variables. For ~10% of entries high_est is used, therefore 'softer' data.	4	1.24
		10.Impartiality	3	No evidence of bias found	5	1.55
		11.Validity	3	Relevant data for conflict disasters.	5	1.55
		Accurate				1.45
	Expository Index					1.45
Data Veracity index (DVi)			29			1.21

Figure 6-9: UCDP GED50 DVi

The UCDP's GED50 dataset attained the highest DVi so far with 1.21 [*Figure 6-10*].

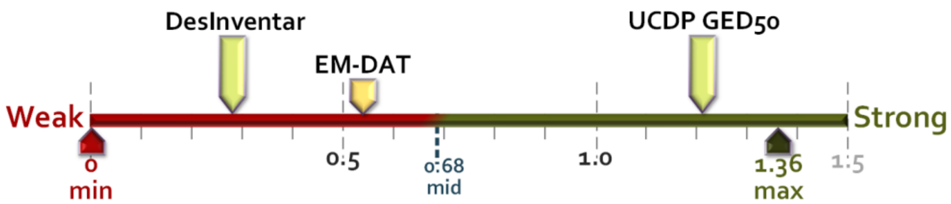


Figure 6-10: UCDP GED50 on the Dataset DVi Scale

The L1 and L2 DVi values of this dataset also reflect the relative importance placed by this study on the different dimensions of veracity. The weightings are such that the *expository* dimensions (i.e. *accurate* and *precise*) will always have a more significant influence on the overall dataset DVi than the elucidatory dimension (i.e. *complete* and *uncluttered*).

For this work, *elucidatory* weaknesses are considered to be somewhat addressable via the amalgamation of datasets and the exclusion of ‘noisy’ entries. Notably, weakness in the *expository* dimensions can only be addressed by flagging suspect entries as ‘soft’ and acknowledging that their existence reduces the reliability of results based on them. The UCDP GED50 L1 Elucidatory DVi of 0.94 versus its L1 Expository DVi of 1.45 echoes this slant [*Figure 6-11*]. The dataset’s overall score is pushed up to 1.21, well above the midpoint of 0.68, because of this emphasis as the dataset appears to be fastidiously maintained.

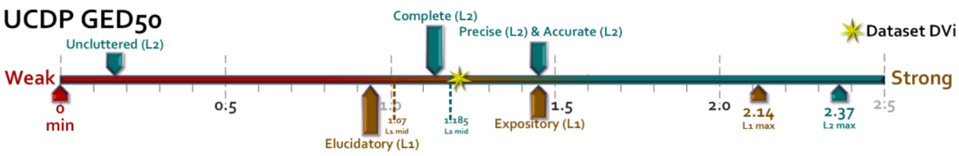


Figure 6-11: UCDP GRD50 on the L2 & L1 DVi Scale

Having evaluated the veracity of UCDP GED50, its data is added to the MSGD and the *UCDP GED50 → MSGD* variable mappings can be found in *Appendix D.3*. For the **13,090** entries in UCDP GED50 have no value for best_est, high_est is used to populate the MSGD fields of Total Deaths and Soft Total Deaths. The latter to indicate these entries are of lesser veracity.

VDC-SY ➔ MSGD

Details of the acquisition, preparation and examination of VDC-SY data can be found in *Appendix C.4* (VDC-SY, 2016b; VDC-SY, 2016a). A detailed DVp for VDC-SY can be found in *Appendix A.5*, while its DVi scores and indices are included here [*Figure 6-12*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions L1	L2	L3	Weighting	VDC-SY Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	2011 to 2015 Deaths < 152K, 100K less than 250K reported by the BBC. Numerous entries likely to be missing.	2	0.62
		2. No omitted values	3	Some values are missing for date, location (province) and victim.	3	0.93
		3. No omitted variables	3	No known variables are missing.	5	1.55
		4. No omitted metadata	1	Very limited information is provided about the data.	2	0.07
		Complete				0.79
	Uncluttered	5. No irrelevant entries	1	758 entries (0.34% of the dataset) have no date.	2	0.07
		Uncluttered				0.07
	Elucidatory Index					0.65
Expository	Precise	6. Reliability	3	Does not appear to be volatile, but unable to confirm if this is the case.	3	0.93
		7. Rigour	3	Notwithstanding the missing values , the data is meticulously collected.	4	1.24
		8. Congruity	3	One entry = one person, therefore little room for a incongruity.	5	1.55
		Precise				1.24
	Accurate	9. Conformity	3	Some vagueness/blanks in qualifiers may be an indicator of 'softness'.	3	0.93
		10.Impartiality	3	Ratio non-regime vs regime fatalities may indicate recording bias.	2	0.62
		11.Validity	3	A valid fit for conflict related humanitarian crises.	5	1.55
		Accurate				1.03
	Expository Index					1.14
Data Veracity index (DVi)			29			0.91

Figure 6-12: VDC-SY DVi

The VDC-SY dataset's DVi of 0.91 is less than the UCDP GED50 dataset, for which it is filling the gap of Syria data, but still high as it is greater than the *Dataset DVi* midpoint of 0.68 [*Figure 6-13*].

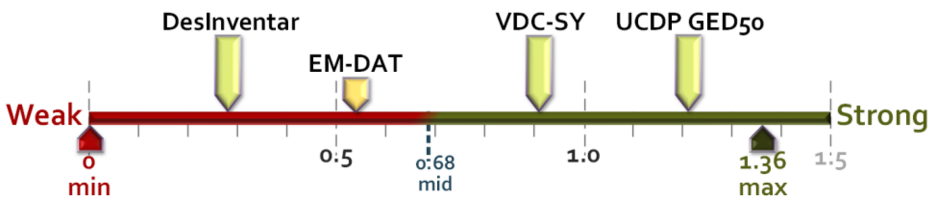


Figure 6-13: VDC-SY on the Dataset DVi Scale

The L1 and L2 DVi values of this dataset are also visually represented in the diagram below [*Figure 6-14*].

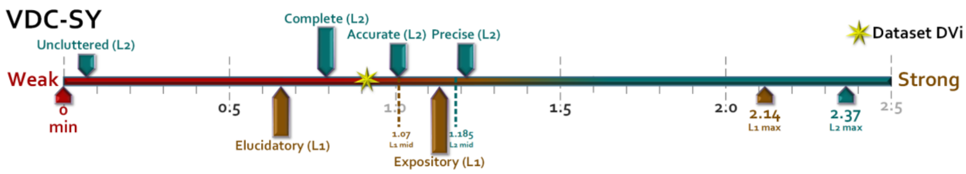


Figure 6-14: VDC-SY on the L2 & L1 DVi Scale

From the VDC-SY site it is understood that the collators of this database risk, and in some cases lose, their lives obtaining this data, therefore the importance of this information to them drives the integrity with which it is maintained. The difficulty in collecting the data and the attention to detail in maintaining it may be echoed in the relative weakness of its *elucidatory* dimension (i.e. *complete* and *uncluttered*) compared to its *expository* dimensions (i.e. *accurate* and *precise*).

Having evaluated the veracity of VDC-SY, its data is added to the MSGD and the *VDC-SY → MSGD* variable mappings can be found in *Appendix D.4*. For *deaths*, if the victim or location is unknown, which is the case for **4,465** entries, then Soft Total Deaths is set to Total Deaths to identify these entries has having lesser veracity than other VDC-SY entries. Similarly, for *affected*, if the victim or location is unknown, which is the case for **2,761** entries, then Soft Total Deaths is set to Total Deaths to identify these entries has having lesser veracity than other VDC-SY entries.

GTD ➔ MSGD

Details of the acquisition, preparation and examination of GTD data can be found in *Appendix C.5* (GTD, 2017e; START, 2017). A detailed DVp for GTD can be found in Appendix A.6, while its DVi scores and indices are included here [*Figure 6-15*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions			Weighting	GTD		
L1	L2	L3		Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	No 1993 data; Method changes may mean loss of data. **Deaths:Affected** ratio **~8:1**, may be missing Affected?	1	0.31
		2. No omitted values	3	>100K 'empty' entries (broader use). >5K entries have unknown type/number kidnapped = flagged 'soft'.	3	0.93
		3. No omitted variables	3	No known variables are missing.	5	1.55
		4. No omitted metadata	1	Very well documented.	5	0.17
		Complete				0.74
	Uncluttered	5. No irrelevant entries	1	103,219 entries 'empty' (because of broader us). Flagged for exclusion.	2	0.07
		Uncluttered				0.07
	Elucidatory Index					0.61
Expository	Precise	6. Reliability	3	Three data collection process changes acknowledged on site.	3	0.93
		7. Rigour	3	Current data collection processes (>1998) meticulously applied and sources cited.	4	1.24
		8. Congruity	3	'Statistical accuracy' across events maintained. Collection methodology changed - 1998, 2008, 2012.	4	1.24
		Precise				1.14
	Accurate	9. Conformity	3	Unknown values flagged as -99 flagged as 'soft' entries.	4	1.24
		10.Impartiality	3	Not sure if dominance of deaths, rather than affected, is bias or fact.	4	1.24
		11.Validity	3	Where entries are not 'empty' they are relevant.	4	1.24
		Accurate				1.24
	Expository Index					1.19
Data Veracity index (DVi)			29			0.92

Figure 6-15: GTD DVi

The GTD dataset's DVi of 0.92 indicates the veracity of this dataset is reasonably high as it is greater than the *Dataset DVi* midpoint of 0.68 [*Figure 6-16*].

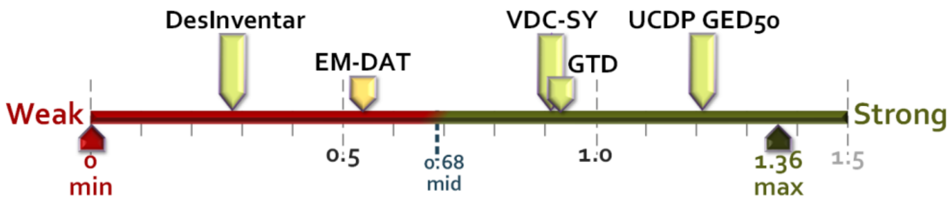


Figure 6-16: GTD on the Dataset DVi Scale

The L1 and L2 DVi values of this dataset are also visually represented in the diagram below [*Figure 6-17*].

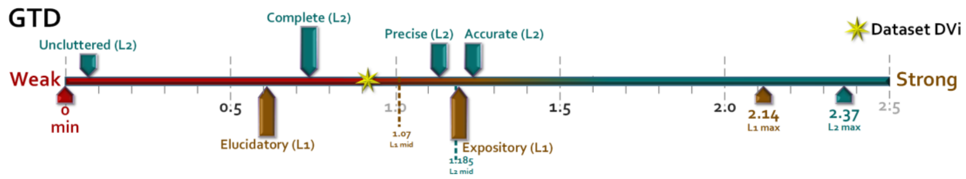


Figure 6-17: GTD on the L2 & L1 DVi Scale

GTD DVi is relatively strong when it comes to *expository* dimensions (i.e. *accurate* and *precise*), which is not surprising considering the level of detail that is maintained. As the dataset stores more than losses for each terrorist incident, there are missing values and a lot of 'clutter' from entries that are 'empty' of human or financial effect. This is reflected in the *elucidatory* dimensions (i.e. *complete* and *uncluttered*).

Having evaluated the veracity of GTD, its data is added to the MSGD and the *GTD → MSGD* variable mappings can be found in *Appendix D.5*. If the type of attack or number of hostages is unknown, entries are taken to be of lesser veracity, and flagged by assigning the following values:

- Soft Total Deaths = Total Deaths (2,283 entries);
- Soft Total Affected = Total Affected (154 entries); and
- Soft USLoss_2015 (000s) = USLoss_2015 (000s) (253 entries).

UNHCR ➔ MSGD

Details of the acquisition, preparation and examination of UNHCR data can be found in *Appendix C.6* (UNHCR, 2017a). UNHCR's on-line documentation states that there may be an underrepresentation of affected people (ibid), citing two reasons: (1) the inability to register all individuals who unlawfully settle outside official camps; (2) the lack of refugee registers in industrialised countries (ibid). This caveat regarding missing information is taken into consideration in the detailed DVp for the UNHCR dataset, which can be found in *Appendix. A.7*; the DVi table for UNHCR PoC data is included here [*Figure 6-18*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions L1	L2	L3	Weighting	UNHCR Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	Disaggregate entries below Year, country of Origin and Country of destination are not available.	3	0.93
		2. No omitted values	3	Over 43% of entries are 'empty'. 4,687 entries for unidentified countries.	2	0.62
		3. No omitted variables	3	Not all variables held are made available for download.	1	0.31
		4. No omitted metadata	1	Only high level explanations are available. No explanation of zero-value entries provided.	2	0.07
		Complete				0.48
	Uncluttered	5. No irrelevant entries	1	141,160 'empty' or spurious entries excluded.	2	0.07
		Uncluttered				0.07
	Elucidatory Index					0.4
Expository	Precise	6. Reliability	3	No change management information provided.	3	0.93
		7. Rigour	3	Site documentation confirms estimates where definitive information cannot be obtained.	2	0.62
		8. Congruity	3	The data is at too high a level to identify incongruities that may/may not be an underlying issue.	3	0.93
		Precise				0.83
	Accurate	9. Conformity	3	14,710 entries (~ 5%) were identified as potentially guestimated numbers.	3	0.93
		10.Impartiality	3	The data does not provide any indication of bias.	4	1.24
		11.Validity	3	49,197 entries are for returnees; therefore about 16.5% of the dataset had no relevance.	3	0.93
		Accurate				1.03
	Expository Index					0.93
Data Veracity index (DVi)			29			0.69

Figure 6-18: UNHCR DVi

UNHCR is the last of the humanitarian crises datasets to be evaluated; therefore the DVi dataset scale [*Figure 6-19*] now displays the relative DVi position of all six datasets of the MSGD.

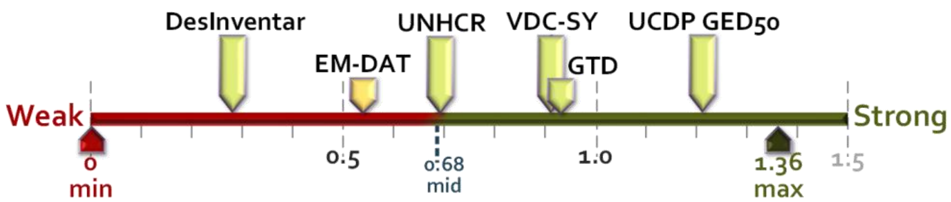


Figure 6-19: UNHCR on the Dataset DVi Scale

The UNHCR dataset's DVi is 0.69 and, although this places it above the *Dataset DVi* midpoint of 0.68, it has one of the three lowest DVi values. The L1 and L2 DVi values of this dataset are visually represented in the diagram below [Figure 6-20].

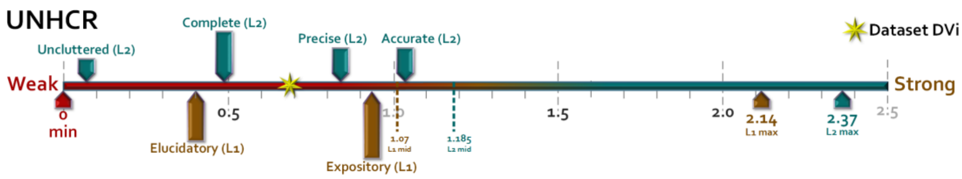


Figure 6-20: UNHCR on the L2 & L1 DVi Scale

UNHCR's DVi is weaker than expected from a UN agency; it contains a significant proportion of unusable or highly suspect entries that are excluded from analysis [*Table 6-2*]. This has the effect of dragging down the DVi's *elucidatory* dimensions (i.e. *complete* and *uncluttered*). There remain 14,701 entries of weaker veracity which influences the dataset's *expository* dimensions.

UNHCR Entries EXCLUDED	
Empty Entries	129,177
ONE person, ONE origin, ONE year	952
Spurious negative one	1
Values 1-4 Marked as *	4,208
Total	**134,338**

Table 6-2: UNHCR Excluded entries

Having evaluated the veracity of UNHCR, its data is added to the MSGD and the *UNHCR → MSGD* variable mappings can be found in *Appendix D.6*. The remaining 14,701 entries that are suspect are flagged as 'soft' (i.e. Soft Total Affected = Total Affected), these are:

- 3,475 entries for ~90 million people of unknown origin;
- 11,226 entries (equating to 15,855 people) where *fewer than TEN people a year from the same country* are counted as PoCs. These are flagged because it is considered unlikely that very small numbers of people leaving their home or homeland in a year is the result of a humanitarian crisis.

Master Set of Global Disasters (MSGD)

Having created a baseline MSGD by acquiring, preparing, evaluating and amalgamating **EM-DAT**, **DesInventar**, **UCDP**, **VDC-SY**, **GTD** and **UNHCR** datasets (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017c), it is best to inspect this consolidated dataset – as this newly created terrain of data may influence any further analysis. Therefore, before searching for M^{i}Os, the proportional composition, temporal distribution, constituencies of disaster effects, overall veracity and finally the 'soft' versus 'firm' entries of the MSGD are examined and, where useful, summarised.

(a) Proportional Composition by Source Dataset

The figure below shows two tables that provide a breakdown of the years and entries in the MSGD from each source dataset [*Figure 6-21*]. The table above represents the entire dataset; the table below is the MSGD once the 277,349 entries flagged as 'empty', spurious or irrelevant are excluded.

Master Set of Global Data (MSGD)				
COMPOSITION				
All Entries				
Data Source	First Year	Last Year	Number of Entries	% of Total
EM-DAT	1900	2015	22,011	2.3%
DesInventar	1200	2017	322,489	33.9%
UCDP	1989	2015	128,264	13.5%
VDC-SY	0	2016	22,049	2.3%
GTD	1970	2015	156,772	16.5%
UNHCR	1951	2016	298,441	31.4%
All	**0**	**2017**	**950,026**	**100.0%**

Master Set of Global Data (MSGD)				
COMPOSITION				
After Excluding Spurious/Irrelevant Entries				
Data Source	First Year	Last Year	Number of Entries	% of Total
EM-DAT	1900	2015	21,091	3.1%
DesInventar	1200	2017	290,457	43.2%
UCDP	1989	2015	128,262	19.1%
VDC-SY	1970	2016	22,033	3.3%
GTD	1970	2015	53,553	8.0%
UNHCR	1951	2016	157,281	23.4%
All	**1200**	**2017**	**672,677**	**100.0%**

Figure 6-21: MSGD Proportional Composition, by Source Dataset

Note that 'empty', spurious and irrelevant entries are not deleted outright as they may yet have a story to tell in future investigations. Appendix E contains additional charts showing empty entries.

Interestingly the proportional contribution from each dataset changes once flagged entries are excluded. For example, DesInventar's representation in MSGD jumps almost 10% from 33.9% to 43.2%; whereas GTD's representation reduces by more than half as it falls from 16.5% to 8%. Another point worth noting is the lack of concurrence of the date ranges that underpins the need to work with a subset of the data once all selected datasets are incorporated.

(b) Temporal Distribution of Entries

Exploring EM-DAT it became clear that its richest seam of data is from 1990 to 2015, events for years before 1990 are less diligently recorded and there is no data in EM-DAT for years after 2015 [*Section 4.4.2*]. This in and of itself imposes 1990 to 2015 as the maximum feasible date range of focus for the MSGD and also aligns to the constraint identified in Iteration 1 [*'knowledge ⬄ consequence' (g) of Section 4.5.3*].

All but one of the source datasets covers the date range 1990–2015 (the exception being VDC-SY) but, as this mostly represents the current conflict in Syria, this is to be expected. Sub-setting the dataset to 1990–2015 excludes a further 89,794 entries from analyses and the data source composition of the remaining 582,883 entries can be seen in *Figure 6-22*.

Master Set of Global Data (MSGD)				
COMPOSITION				
1990 - 2015 (SUBSET)				
Data Source	First Year	Last Year	Number of Entries	% of Total
EM-DAT	1990	2015	15,787	2.7%
DesInventar	1990	2015	242,244	41.6%
UCDP	1990	2015	125,899	21.6%
VDC-SY	2002	2015	20,372	3.5%
GTD	1990	2015	32,463	5.6%
UNHCR	1990	2015	146,118	25.1%
All	**1990**	**2015**	**582,883**	**100.0%**

Figure 6-22: MSGD Source Dataset Entries 1990-2015

(c) Constituencies of Disaster Effects

Table 6-3 details the number of entries in the 1990–2015 subset of MSGD by disaster effect, or not, as the case may be.

Data Source	Entries (1990-2015)	No Human Effect	No Financial Effect	Deaths	Affected	Human Effect	Financial Effect
EM-DAT	15,787	409	12,324	13,038	11,364	15,378	3,463
DesInventar	242,244	1,287	239,491	52,438	212,960	240,957	2,753
UCDP	125,899		125,899	125,899	-	125,899	-
VDC-SY	20,372		20,372	18,122	9,557	20,372	-
GTD	32,463	2,339	29,902	29,460	978	30,124	2,561
UNHCR	146,118		146,118	-	146,118	146,118	-
All	**582,883**	**4,035**	**574,106**	**238,957**	**380,977**	**578,848**	**8,777**
Percentage of Total		**0.7%**	**98.5%**	**41.0%**	**65.4%**	**99.3%**	**1.5%**

Table 6-3: MSGD (1990–2015), Entries with/without Disaster Effect

This helps suggest where further analysis of the data may, or may not, prove useful and what constraints may apply. For example, only 8,777 of 582,883 entries provide any financial loss data. This is only 1.5% of MSGD 1990–2015 total and is sourced from 3 datasets: EM-DAT, DesInventar and GTD. This meagre representation of financial effects is probably because of inadequate recording and not because financial losses are relatively rare when compared to human effects. As a result, exploring financial effects in the MSGD for M^iOs is considered infeasible.

Another realisation from the spread of disaster-effect containing entries is that exploring UCDP or UNHCR in isolation is likely to provide a partial perspective. UCDP conflict effects are all about fatalities, whereas UNHCR effects are all about people who are uprooted from their homes. Therefore, analysing each dataset individually does not factor in possible cause and effect relationships. For example, a conflict recorded in the UCDP dataset may be the cause of a mass movement of people recorded in the UNHCR dataset.

Figure 6-23 depicts area charts of MSGD entries, human effects and financial effects highlighting the selected and, most richly recorded, period of 1990–2015.

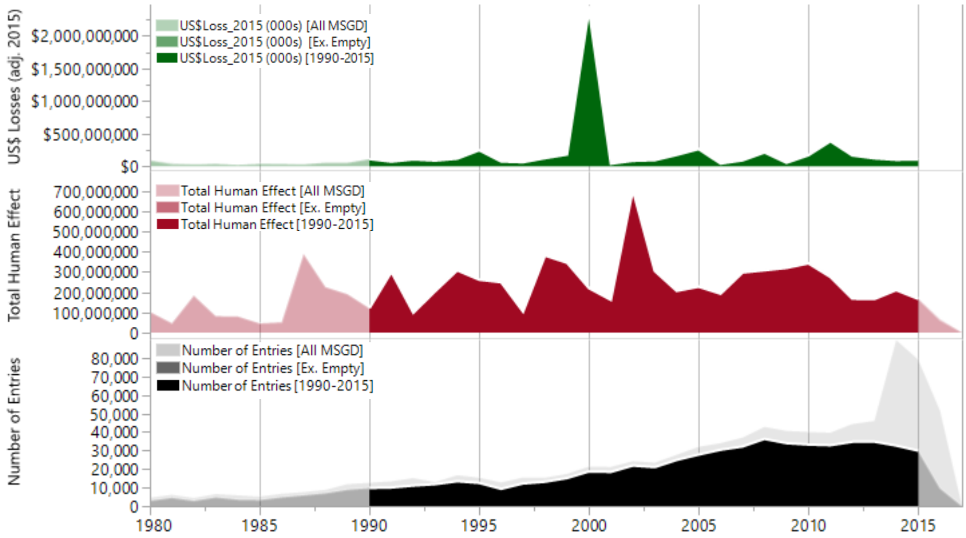


Figure 6-23: MSGD Entries, Human Effect & US$ Losses

Table 6-4 details the numbers of deaths, people affected, total human impact (the sum of deaths and affected) and the financial losses recorded by each data source for 1990–2015.

Data Source	Number of Deaths	Number of People Affected	Total Human Effect	Total Financial Effect ($000s)
EM-DAT	2,637,786	5,253,623,314	**5,256,261,100**	$3,237,071,951.000
DesInventar	3,949,329	722,911,570	**726,860,899**	$2,444,451,505.103
UCDP	2,181,292		**2,181,292**	
VDC-SY	151,739	63,583	**215,322**	
GTD	151,758	20,887	**172,645**	$11,892,160.300
UNHCR		671,729,987	**671,729,987**	
All	**9,071,904**	**6,648,349,341**	**6,657,421,245**	**$5,693,415,616.403**

Table 6-4: MSGD (1990–2015), Disaster Effects

Basic observations from *Table 6-4* include:

- The 1.5% of entries that record financial loss equate to almost US$5.7 trillion of losses adjusted to 2015 levels. Therefore it may still be useful to identify if there is any pattern as to when, and for which countries, financial losses are recorded, even though this is not expected to yield any M^{i}Os [*Appendix F.2*].
- For EM-DAT and DesInventar, both datasets of predominantly *naturogenic* and *anthropogenic* disaster types, the number of deaths are a near miniscule proportion of the human effect of disaster – 0.05% for EM-DAT; 0.54% for DesInventar. For

disaster datasets that contain only *deviant* disaster types, i.e. UCDP, VDC-SY and GTD, the proportion of entries of deaths increases dramatically to 100%, 70.47% and 87.9% respectively.

- A significantly higher number of deaths are recorded in DesInventar than in EM-DAT.
- The number of deaths, 151,739, recorded in the VDC-SY for the Syrian conflict over 5 years (2011–2015) *Table 6-5*] is almost equivalent to the number of terrorism deaths, 151,758, in the GTD for the 26 years examined [*Table 6-4*].

Year	Number of Deaths	Number of People Affected	Total Human Effect
2002		1	1
2008		1	1
2010		1	1
2011	6,422	19,539	25,961
2012	46,639	21,380	68,019
2013	45,769	17,233	63,002
2014	31,452	4,313	35,765
2015	21,457	1,115	22,572
All	151,739	63,583	215,322

Table 6-5: VDC-SY The Human Effect of the Syrian Conflict

(d) MSGD Data Veracity

Higher level DVi dimensions for each of the MSGD's source datasets are depicted here in relation to each other [*Figure 6-24]*.

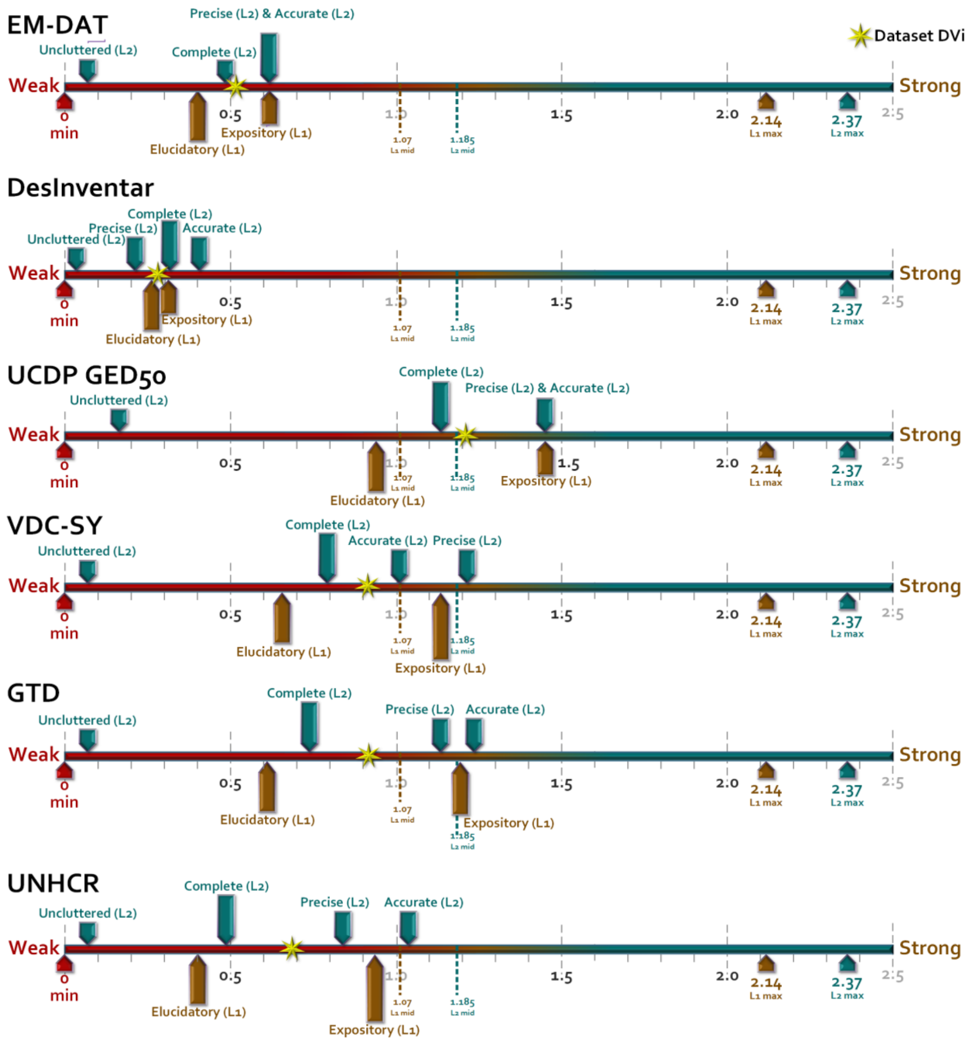


Figure 6-24: MSGD Source Datasets - DVi Dimension Scale

The individual scores and indices of the component datasets can be viewed in comparison on the next page [*Table 6-6*].

Veracity Dimensions			Weighting	EM-DAT		DesInventar		UCDP GED50		VDC-SY		GTD		UNHCR		MSGD DVi				
L1	L2	L3		Score	Index	Score	Index	Score	Index	Score	Index	Score	Index	Score	Index	Score	MIN	MAX	MID	MEAN
Elucidatory	Complete	1. No omitted entries	3	2	0.62	1	0.31	4	1.24	2	0.62	1	0.31	3	0.93	1	0.31	1.24	0.62	0.67
		2. No omitted values	3	2	0.62	1	0.31	5	1.55	3	0.93	3	0.93	2	0.62	1	0.31	1.55	0.78	0.83
		3. No omitted variables	3	2	0.62	2	0.62	5	1.55	5	1.55	5	1.55	1	0.31	1	0.31	1.55	1.09	1.03
		4. No omitted metadata	1	2	0.07	1	0.03	5	0.17	2	0.07	5	0.17	2	0.07	1	0.03	0.17	0.07	0.1
		Complete			0.48		0.32		1.13		0.79		0.74		0.48		0.32	1.13	0.61	0.66
	Uncluttered	5. No irrelevant entries	1	2	0.07	1	0.03	5	0.17	2	0.07	2	0.07	2	0.07	1	0.03	0.17	0.07	0.08
		Uncluttered			0.07		0.03		0.17		0.07		0.07		0.07		0.03	0.17	0.07	0.08
	Elucidatory Index				0.4		0.26		0.94		0.65		0.61		0.4		0.26	0.94	0.51	0.54
Expository	Precise	6. Reliability	3	3	0.93	1	0.31	5	1.55	3	0.93	3	0.93	3	0.93	1	0.31	1.55	0.93	0.93
		7. Rigour	3	1	0.31	0	0.0	4	1.24	4	1.24	4	1.24	2	0.62	0	0.0	1.24	0.93	0.78
		8. Congruity	3	2	0.62	1	0.31	5	1.55	5	1.55	4	1.24	3	0.93	1	0.31	1.55	1.09	1.03
		Precise			0.62		0.21		1.45		1.24		1.14		0.83		0.21	1.45	0.99	0.92
	Accurate	9. Conformity	3	1	0.31	1	0.31	4	1.24	3	0.93	4	1.24	3	0.93	1	0.31	1.24	0.93	0.83
		10. Impartiality	3	2	0.62	1	0.31	5	1.55	2	0.62	4	1.24	4	1.24	1	0.31	1.55	0.93	0.93
		11. Validity	3	3	0.93	2	0.62	5	1.55	5	1.55	4	1.24	3	0.93	2	0.62	1.55	1.09	1.14
		Accurate			0.62		0.41		1.45		1.03		1.24		1.03		0.41	1.45	1.03	0.96
	Expository Index				0.62		0.31		1.45		1.14		1.19		0.93		0.31	1.45	1.04	0.94
Data Veracity index (DVi)			29		0.52		0.29		1.21		0.91		0.92		0.69		0.29	1.21	0.8	0.76

Table 6-6: MSGD Data Veracity Index (DVi)

Table 6-6 also shows the MSGD DVi, based on the weakest of each L3 dimension. The principle applied is that the veracity of a combined dataset is set by its weakest contributors; i.e. '*a chain is only as strong as its weakest link*'. To provide perspective MSGD's DVi based on mean, median and maximum L3 results are also calculated [*Table 6-6* & *Figure 6-25*]

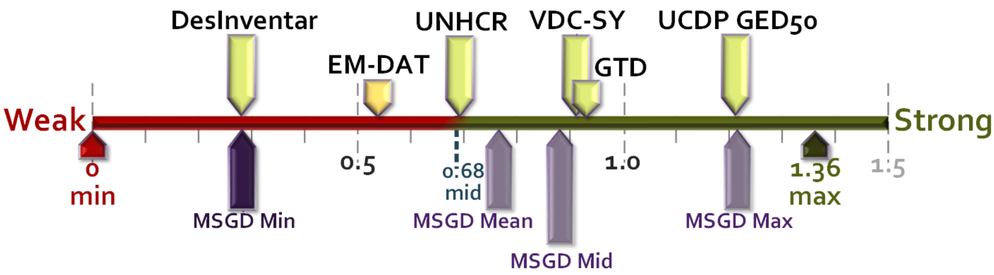


Figure 6-25: MSGD Data Veracity Index (DVi)

A simple visual scan shows that basing MSGD's DVi on the minimums of its contributing datasets results in DesInventar dragging the overall MSGD's DVi to its level. That said, using any of the other values – *median*, *mean* or *maximum* – would result in a perception of higher veracity than justified considering 41.6% of the entries in MSGD 1990–2015 are from DesInventar [Figure 6-22].

Comparing the datasets, the weakest veracity indices are for DesInventar and EM-DAT, both of which hold predominantly naturogenic disasters. DesInventar's poor DVi is understandable as the solution is designed to provide flexibility and freedom for localised tailored implementations. This also has the effect of making it easy to introduce inconsistencies, errors and gaps in the data. The poor DVis score for EM-DAT is harder to reconcile as it is centrally updated and curated in a controlled research environment, and other 'university curated' disaster datasets selected for this study, e.g. UCDP and GTD, have noticeably higher veracity indices.

(e) Summary of 'soft' and 'firm' data

The table below provides a summary of the types of entries flagged as 'soft', i.e. of weaker veracity, in the MSGD [*Table 6-7*].

	'SOFT' Human Losses	'SOFT' Financial Losses
EM-DAT	Estimating logic appears to be applied; OR the inclusion criteria are not met.	Anomalies between Estimated Damage and Insured Damage
DesInventar	Veracity is so poor (DVi=0.29) that all losses are flagged as 'soft'.	
UCDP	best_est=0 therefore high_est used.	
VDC-SY	When locations or victims are unknown or missing.	
GTD	When type of attack or victim details are ambiguous.	
UNHCR	Origin unknown or there is suspiciously consistent single (or very low) number of annual values.	

Table 6-7: MSGD Flagging of 'Soft' Numbers

'Soft' and 'firm' disaster effects are summarised here [*Table 6-8* & *Table 6-9*]:

Data Source	Total Human Effect	Total Human Effect [Firm]	Total Human Effect [Soft]
EM-DAT	**5,256,261,100**	1,201,159,929	4,055,101,171
DesInventar	**726,860,899**		726,860,899
UCDP	**2,181,292**	1,831,959	349,333
VDC-SY	**215,322**	116,870	98,452
GTD	**172,645**	158,268	14,377
UNHCR	**671,729,987**	642,273,584	29,456,403
All	**6,657,421,245**	**1,845,540,610**	**4,811,880,635**

Table 6-8: MSGD 1990-2015 'Soft' and 'Firm' Human Effect by Data Source

Data Source	Total Financial Effect ($000s)	Total Financial Effect ($000s) [Firm]	Total Financial Effect ($000s) [Soft])
EM-DAT	$3,237,071,951.000	$1,735,774,933.000	$1,501,297,018.000
DesInventar	$2,444,451,505.103		$2,444,451,505.103
UCDP			
VDC-SY			
GTD	$11,892,160.300	$11,772,363.799	$119,796.501
UNHCR			
All	**$5,693,415,616.403**	**$1,747,547,296.799**	**$3,945,868,319.604**

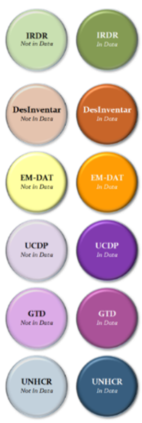
Table 6-9: MSGD 1990-2015 'Soft' and 'Firm' Financial Effect by Data Source

6.4.2 Master Disaster Classification (MDC)

The MDC is created here by applying the pragmatic design principle of using the datasets contributing to the MSGD as guides and establish the scope of the MDC. Thus the first building block of the MDC is the IRDR, the Integrated Research on Disaster Risk Peril Classification and Hazard Glossary (IRDR, 2014) as this is referred to by EM-DAT and DesInventar as the basis of their classifications (Guha-Sapir et al., 2017g; DesInventar, 2017c). The process maps as many of the naturogenic disasters identified in the MSGD to the IRDR base, where this is not possible, the creation of new classifications is guided by the source dataset. Other potential sources of inspiration include the UN study of "*alternative classification schemes for man-made hazards*" (Lerner, 2016), and the "*taxonomy of threats*" developed by the Centre of Risk Studies, University of Cambridge (Coburn et al., 2014).

The higher level structure of the MDC and the number of '*perils*' at the bottom of each hierarchy are included here [*Figure 6-26*]; the full hierarchical model of the MDC can be found in *Appendix B.2* which also includes tables mapping MDC entries the number of MSGD entries. While building the MDC model (in conjunction with populating the MSGD), the MDC reference dataset is also constructed with a reference id (MDC_ID) linking each MSGD entry to its classification in the MDC reference dataset.

MDC Colour Key



IRDR, Integrated Research on Disaster Risk Peril Classification and Hazard Glossary (IRDR, 2014)

DesInventar, Disaster Inventory System (DesInventar.NET, 2017)

EM-DAT, Emergency Events Database (Guha-Sapir et al., 2017l)

UCDP, Uppsala Conflict Data Program (UCDP, 2017b)

GTD, Global Terrorism Database (GTD, 2017e)

UNHCR, United Nations High Commissioner for Refugees (UNHCR, 2017c)

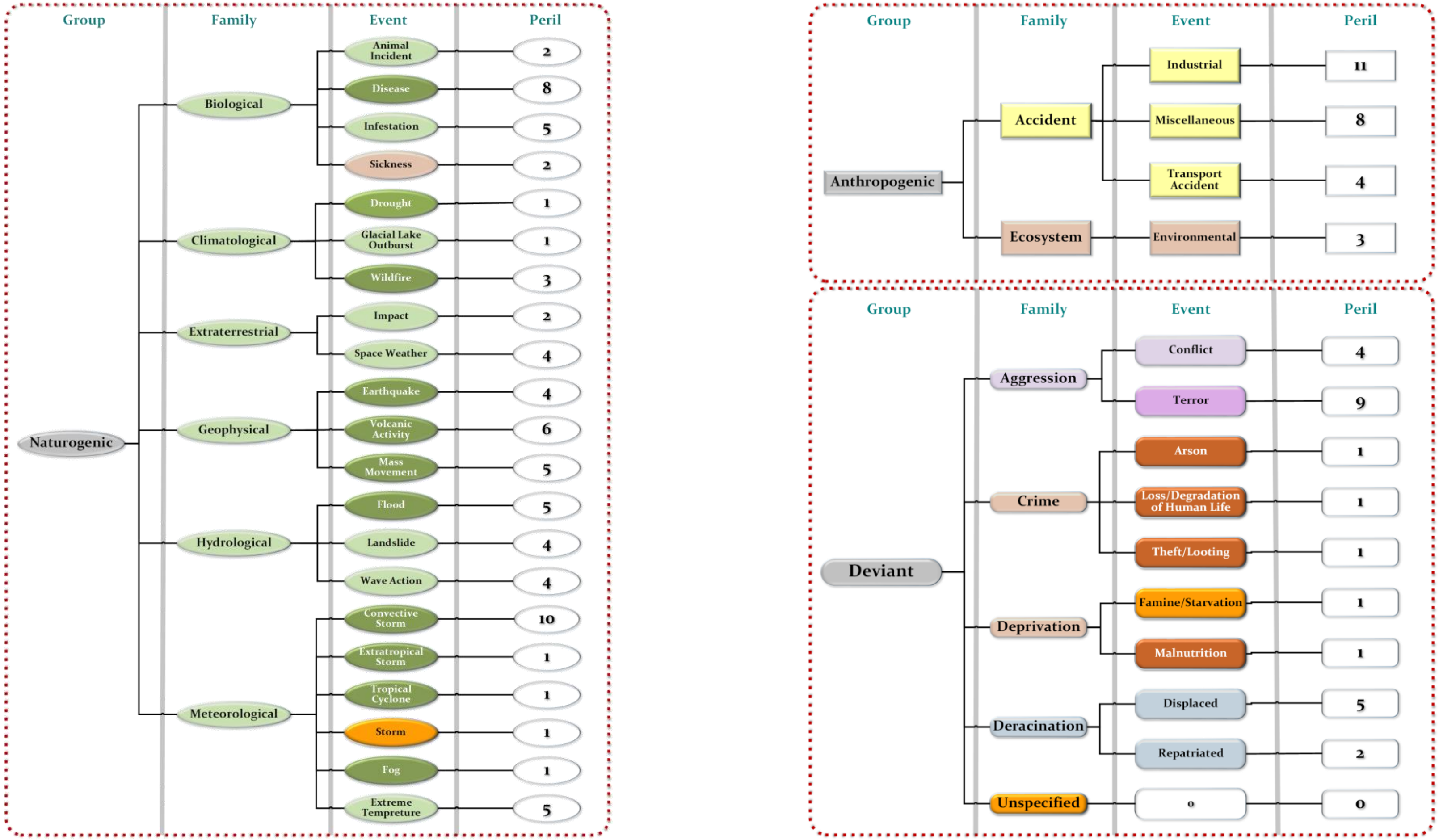


Figure 6-26: Master Disaster Classification (Higher levels)

6.4.3 The Search for M^{i}Os

The MSGD created here provides a more comprehensive set of disasters to explore for M^{i}Os. It is a key '*data scaffold*' in the humanitarian system and a fulcrum around which most all other humanitarian data pivots. Returning to the guidance of the HFA and SFDRR, two fundamental questions are asked to help with this exploration (HYOGO, 2008; Wahlström, 2015):

(1) *Have the financial effects of disasters diminished over time?* If so, this may be an indicator of improving disaster preparation and mitigation, and is less likely to be an indicator of emergency response. This is because response activity is not typically about recovering from financial losses but about saving lives.

(2) *Have the human effects of disasters diminished over time?* If so, this may be an indicator of improving humanitarian intervention as a whole including, if not in particular, the humanitarian response to disasters.

Of these, questions attempting to answer question **(1)** has already been deemed infeasible as only 1.5% (8,777 entries) of the MSGD 1990–2015 hold any financial loss information. Nevertheless for the sake of completeness a number of visualisations are created for financial losses and can be found in *Appendix F*. The remainder of this section focusses on exploring the MSGD for the human effects of disasters based on question (***2)***:

Have the human effects of disasters diminished over time?

The charts in *Figure 6-27* starts the search for answers to this question by looking at trends at the highest level of aggregation over the most fully populated year range 1990–2015 held in the MSGD.

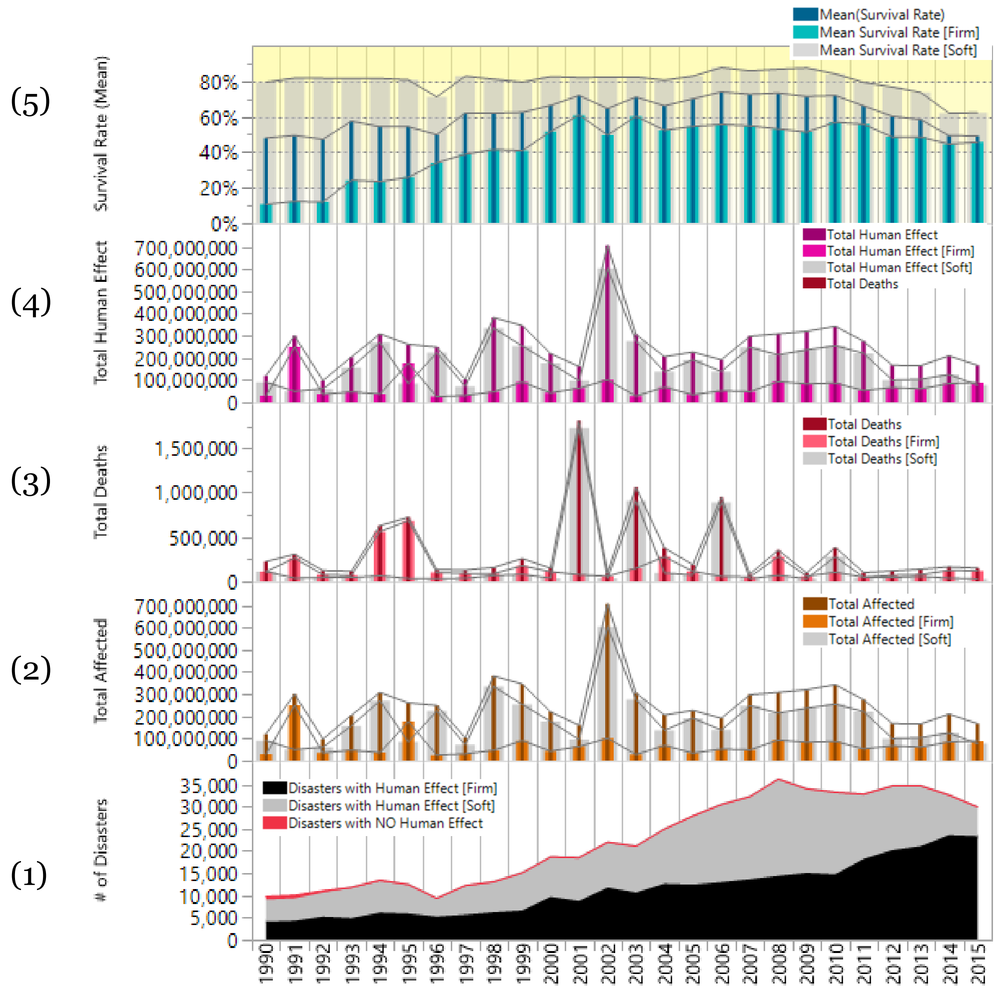


Chart	
(1)	The total number of disaster entries per year – *lowermost chart*.
(2)	The total number of people affected by disasters per year.
(3)	The total number of deaths caused by disaster per year.
(4)	The total human effect of disasters per year – $Total\ Deaths + Total\ Affected$
(5)	The *mean survival rate* per year – *topmost chart* – $\left[\frac{Total\ Affected}{Total\ Human\ Effect}\right]\%$

Figure 6-27: MSGD Human Effect of Disasters 1990 – 2015 (**ALL**)

*Note that the **# of Disasters** notation is a simplification of **# of Disaster <u>ENTRIES</u>**. It is not a literal 'number of discrete disasters', as there may be multiple entries per disaster. Multi-entry disasters are infrequent but possible for EM-DAT entries, not obvious in UCDP, GTD and UNHCR, and constitute almost all of VDC-SY in which entries represent unique date/location based 'incidents', most of which are part of the multi-year Syrian conflict.*

Cursory observations from the charts in *Figure 6-27* are:

- The charts for deaths, people affected and human effect appear to be erratic with no discernible pattern.
- The dominance of soft data in all five charts is considerable. This suggests that any analysis based on undifferentiated disaster effects should be viewed with caution because of the large proportion of data with lesser veracity.
- More disaster entries do not equate to more human effect.
- The greatest spikes in deaths and people affected are in the soft less reliable numbers.
- Survival rates for weaker veracity data are high, remaining in the vicinity of 80% for most of the first twenty years examined. Whereas, firm *Survival Rates* start at a low 10% – roughly translating to an average of only 1 in 10 people surviving a disaster in 1990 – rising to a peak of 61% in 2001 – i.e. an average 6 in 10 people surviving a disaster.
- Both firm and soft aggregate *Survival Rate* means appear to be on a decline, with numbers decreasing to 45.7% for firm data (from a high of 56.9% in 2010) and 62.2% for soft data (from a high of 87.5% in 2009) by 2015.
- Interestingly the relative lack of volatility in the *mean survival rate* is regardless of the behaviour of the number of disaster entries, deaths or people affected.

Disaster Entries and Effects by Disaster Group

In the hope of obtaining more clarity, these charts are broken down to the disaster group level [*Figure 6-28*]. *Note that the Y-scales can vary for each set of disaster group charts. The Y-scale is only changed where bars are likely to become invisible if the larger scale is retained.*

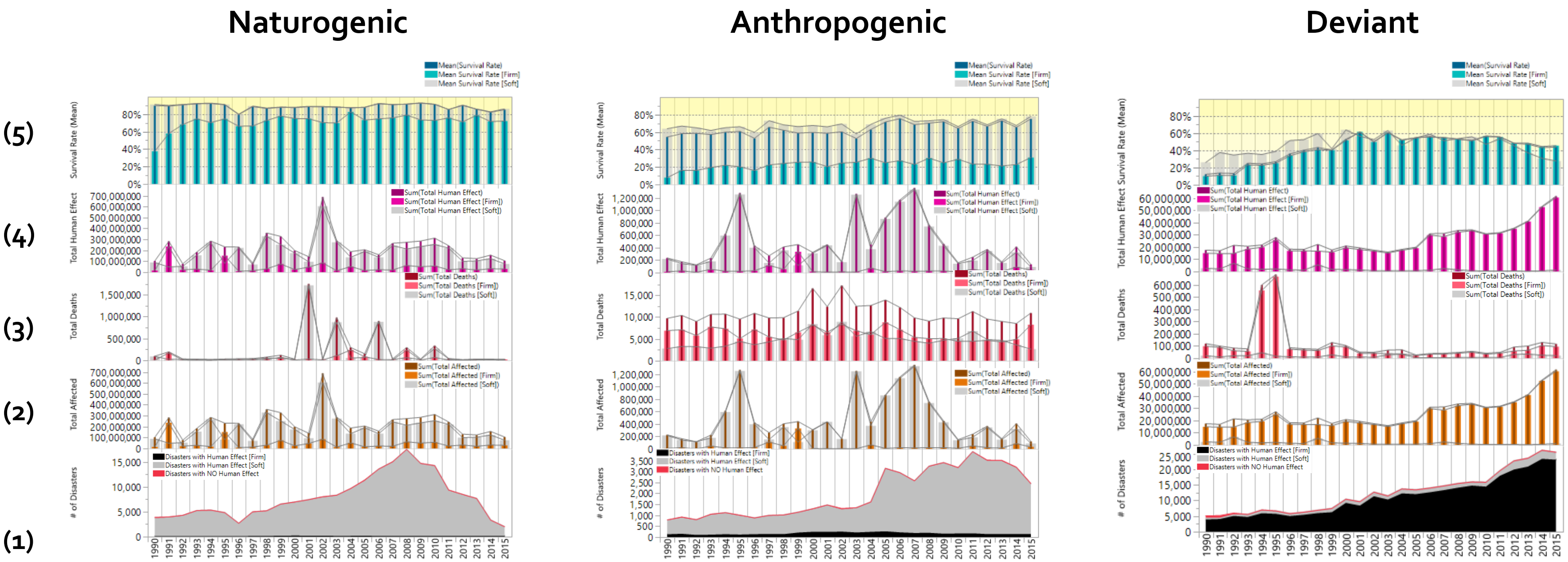


Chart	Y-axis	Scale
(5)	Survival Rate	> 80%
(4)	Total Human Effect	> 700,000,000
(3)	Total Deaths	> 1,500,000
(2)	Total Affected	> 700,000,000
(1)	# of Disaster Entries	> 15,000

Chart	Y-axis	Scale
(5)	Survival Rate	> 80%
(4)	Total Human Effect	> 1,200,000
(3)	Total Deaths	> 15,000
(2)	Total Affected	>1,200,000
(1)	# of Disaster Entries	> 3,500

Chart	Y-axis	Scale
(5)	Survival Rate	> 80%
(4)	Total Human Effect	> 60,000,000
(3)	Total Deaths	> 600,000
(2)	Total Affected	> 60,000,000
(1)	# of Disaster Entries	~ 25,000

Figure 6-28: MSGD Human Effect of Disasters 1990 – 2015 by Disaster Group

Chart (1) Number of Disaster Entries

The area chart of each disaster group tells a different story:

- Naturogenic disasters, having peaked in 2008, are in decline., but as most entries for naturogenic data are flagged as soft therefore this chart cannot be viewed as definitive.
- Anthropogenic disasters, which span a narrower scale than naturogenic disasters, doubled in occurrence over a 15-year period, 1990–2004, then almost doubled again in one year, 2005. After peaking in 2011 they have been in gradual decline but, even allowing for the differing Y-scales in their respective charts, still has more entries (*2,460 undifferentiated, 147 firm, 2,313 soft*) than naturogenic disasters (*2,000 undifferentiated, 141 firm, 1,859 soft*) in 2015.
- Deviant disasters, in contrast to the other disaster groups, have increased almost 5-fold over the 26-year period examined. Admittedly this includes more granular entries from VDC-SY for the multi-year conflict in Syria. That said, even if the granular VDC-SY entries are excluded the scale of increase still remains significantly high at over 425%. *Note that charts of annual disaster entries stacked by disaster group, created both with and without VDC-SY numbers, can be found in Appendix H.2.*

Chart (2) Total Affected & Chart (3) Total Deaths

These charts are discussed as relevant with chart (4) observations.

Chart (4) Total Human Effect

Comparing chart (4) for all three disaster groups:

- The charts for naturogenic and anthropogenic human effect totals are based on data of predominantly weak veracity, i.e. there are significant grey bars. For both disaster groups the primary source of weak data is from Total Affected (C*hart (2)*). For anthropogenic disasters some relative veracity is gained from the underlying Total Deaths (*Chart (3)*). Exploring this in more detail, naturogenic disasters and anthropogenic disasters rely primarily on data from EM-DAT and DesInventar [*Figure 6-29*], both of which have the lowest DVi scores, ***0.52*** and ***0.29*** respectively, of all the datasets contributing to the MSGD. It is primarily DesInventar's contribution creating the underlying weakness in the veracity of naturogenic and anthropogenic disaster losses.

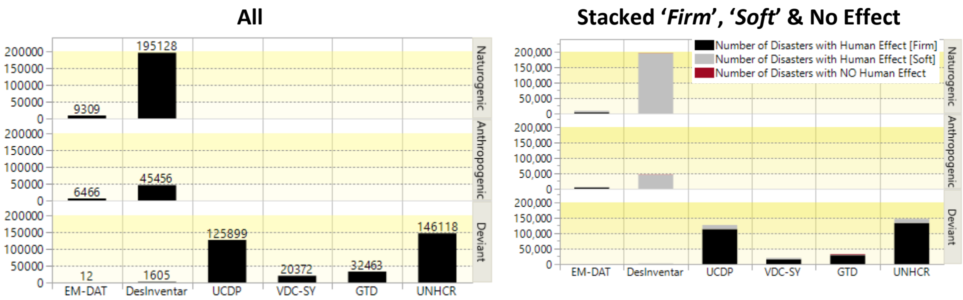


Figure 6-29 MSGD 1990–2015 Entries by Source & Disaster Group

Figure 6-30 depicts only EM-DAT anthropogenic disasters, confirming EM-DAT's 'firm' numbers provide the underlying strength to the annual death toll of anthropogenic disasters.

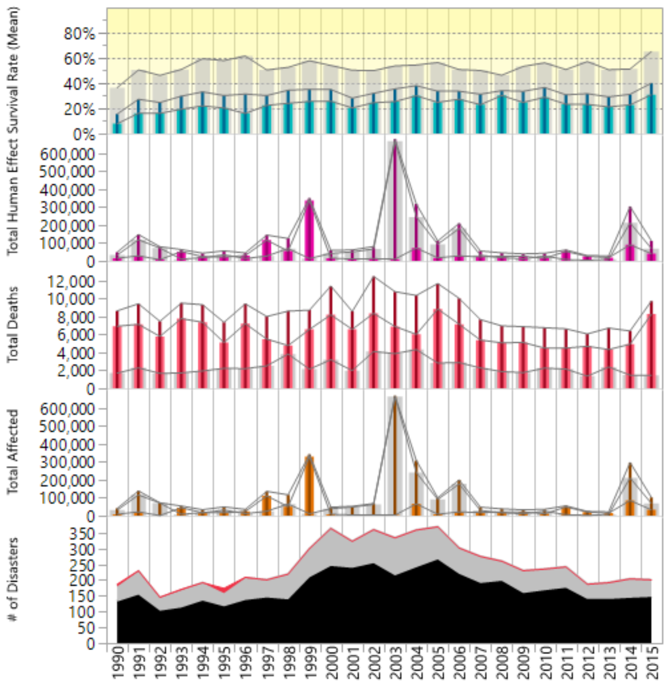


Figure 6-30: EM-DAT Anthropogenic Disasters

Appendix H.3 offers additional stacked charts of disaster data source and human effect by disaster group.

- Neither naturogenic nor anthropogenic disasters exhibit a trend in human impact based on either the undifferentiated numbers or their underlying 'firm' and 'soft' numbers. While, anthropogenic disasters fluctuate between peaks and toughs in human effect, naturogenic disasters exhibit distinct single year spikes in people affected (2002) and deaths (2001, 2003 and 2006). *Figure 6-31* shows the events these spikes represent for these four years.

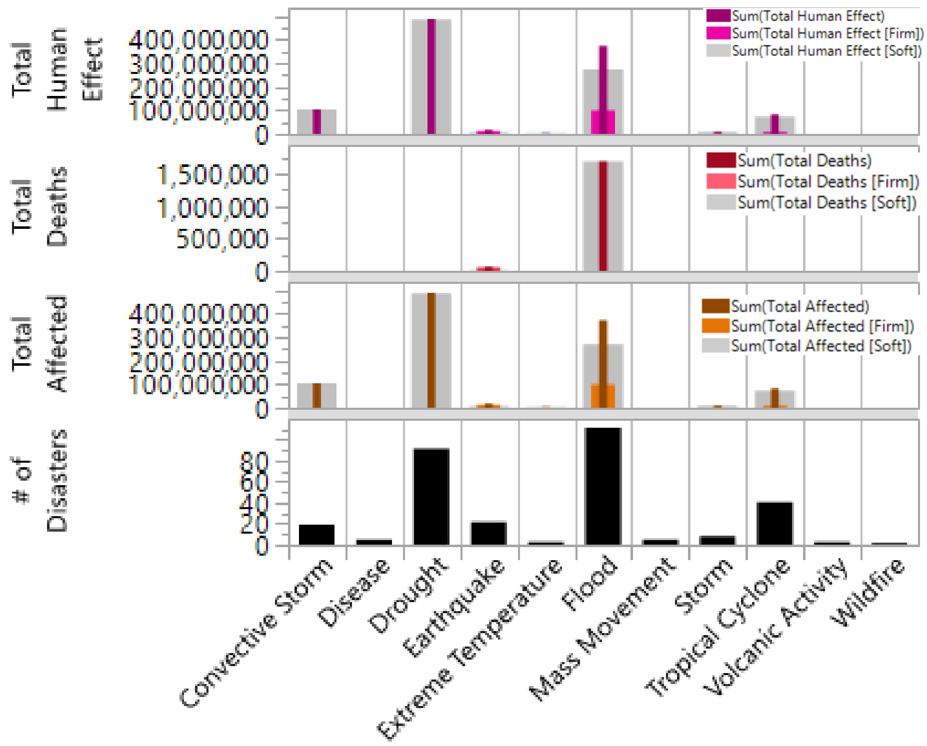


Figure 6-31: MSGD Naturogenic – Human Effect Spikes (2001–2003 & 2006)

The bar charts show that most extreme human effects for these years are from droughts and floods. *Figure 6-32* is a map depicting the 22 disaster-affected countries with total human effects of one million or more during these years. Notice India and China have total human effect in the hundreds of millions. A detail breakdown of high human effect entries for the spike years of 2001, 2002, 2003 and 2006 can be found in *Appendix H.4*.

Country	Total Human Effect
China	592,901,379
India	365,145,213
Ethiopia	21,910,897
Thailand	10,949,868
Philippines	9,877,592
Mexico	9,736,475
Indonesia	7,396,581
Pakistan	6,735,150
Vietnam	6,440,544
Cambodia	6,244,387
Tanzania	5,600,000
Mozambique	4,917,180
Malawi	3,226,284
Afghanistan	2,400,013
Bangladesh	2,236,281
Peru	1,840,227
Argentina	1,529,073
Sri Lanka	1,528,264
Madagascar	1,509,182
Russian Federation	1,330,704
Uganda	1,205,124
Rwanda	1,000,000

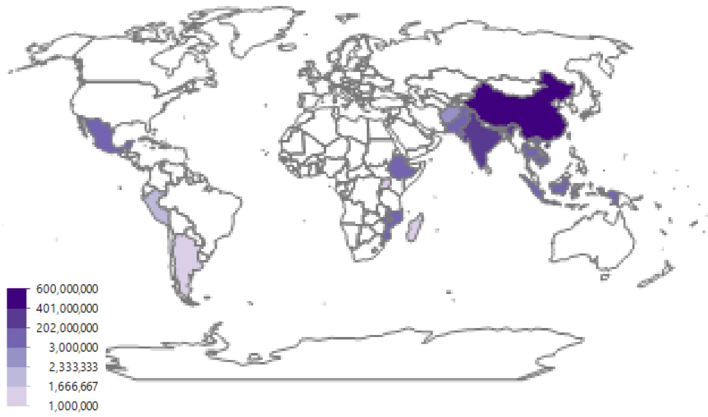


Figure 6-32: 22 Countries with Total Human Effect ≥ ONE million
(Combined 'spike' years 2001, 20032003 & 2006)

- Deviant disasters and their effects charts are noticeably based on data of firm veracity. Two of the three human effect charts show an upward trend in recent years in line with the upward trend in the volume of deviant disaster entries in MSGD. Chart (3) Total Deaths is the exception, fluctuating between lows of around 20,000 and highs of around 120,000 in all years other than 1994 and 1995, when the death toll rises radically to highs of 400,000 to 600,000. The data confirms these deaths are the result of two extreme events, the genocide in Rwanda (459,119 deaths in 1994) and the famine in North Korea (610,000 deaths in 1995).

Chart (5) Survival Rate

Reviewing this chart for each of the three disaster groups is interesting, but still does not yield any indication that progress is being made in improving the outcome of disasters.

- For both naturogenic and anthropogenic disasters the *mean survival rate* is consistently higher for data with weaker veracity. This is more marked for anthropogenic disasters, where soft *Survival Rates* barely fall below 60%, but firm *Survival Rates* struggle to hit highs of 30%. Whereas for naturogenic disasters the difference between soft and firm *Survival Rates* is much smaller with soft *Survival Rates* reaching highs above 90%.

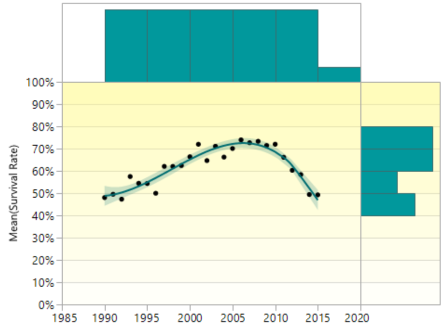
- Another curiosity about the survival patterns of naturogenic and anthropogenic disasters is the limited volatility they exhibit for both soft and firm data. For example:
  - For naturogenic data the human effect of disasters spikes in 2001, 2002, 2003 and 2006, but these spikes do not translate to greater fluctuations in *mean survival rate*. From 1997 to 2010 the soft *Survival Rate* remains in the area of 90%, and the firm *Survival Rate* remains in the vicinity of 70%.
  - For anthropogenic data a number of significant spikes can be seen in people affected numbers from 1995–2007, yet the *Survival Rate* for this time period for soft data ranges between 66% and 72% and for firm data remains with reach of 20%.
- For deviant disasters the shape of the chart is different from the other two disaster groups. Both soft and firm survival values start low (~10% for firm and ~26% for soft) in 1990 and both data veracity levels gradually increase to their highest levels of ~60% by 2000/2001. For the next decade both sets of *Survival Rates* remain more or less steady in the 50%–60% range before entering a decline with 2015 *Survival Rates* closing at around 27% for soft data and 45% for firm data.

M^{i}O: Mean *Survival Rate by Year*

The descriptive analyses and visualisations so far may have improved familiarity with the disaster data held in the MSGD, but do not answer the question posed earlier i.e. *'have the human effects of disasters diminished over time?'* Charts of the basic human effects – *Total Deaths*, *Total Affected*, and *Total Human Effect* – show no pattern, *but Mean Survival Rate by Year* is worth investigating as it has already been highlighted as a potential M^{i}O in Iteration 1 [*'Knowledge ⇔ Consequence' (j) of Section 4.5.3*]. Consider the following null hypothesis, H_0, and the analysis in *Figure 6-33*.

H_0 There is no relationship between the *year (time)* and *mean survival rate*.

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -28.49964 + 0.0145792*YEAR - 0.0013838*(YEAR-2002.5)^2 - 9.8159e-5*(YEAR-2002.5)^3

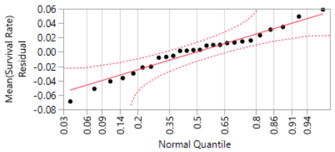
Summary of Fit

RSquare	0.89311
RSquare Adj	0.878535
Root Mean Square Error	0.031664
Mean of Response	0.61737
Observations (or Sum Wgts)	26

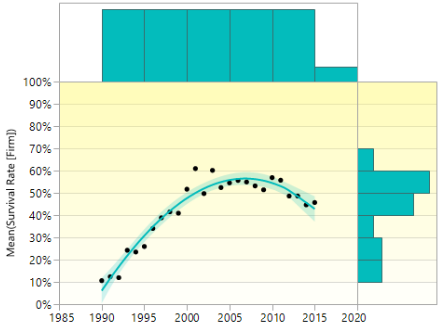
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	3	0.18430037	0.061433	61.2733
Error	22	0.02205749	0.001003	Prob > F
C. Total	25	0.20635786		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -32.29235 + 0.0163901*YEAR - 0.0018181*(YEAR-2002.5)^2 - 1.0932e-5*(YEAR-2002.5)^3

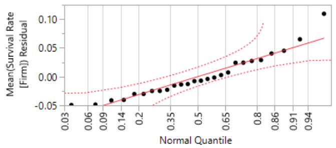
Summary of Fit

RSquare	0.940824
RSquare Adj	0.932755
Root Mean Square Error	0.039963
Mean of Response	0.426508
Observations (or Sum Wgts)	26

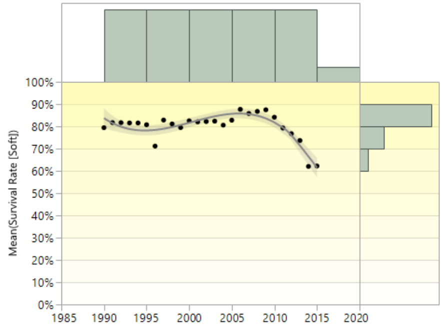
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	3	0.55861735	0.186206	116.5916
Error	22	0.03513568	0.001597	Prob > F
C. Total	25	0.59375303		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -16.63845 + 0.008729*YEAR - 0.000736*(YEAR-2002.5)^2 - 0.0001127*(YEAR-2002.5)^3

Summary of Fit

RSquare	0.791777
RSquare Adj	0.763383
Root Mean Square Error	0.031222
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	3	0.08154866	0.027183	27.8853
Error	22	0.02144582	0.000975	Prob > F
C. Total	25	0.10299448		<.0001*

Residual Normal Quantile Plot

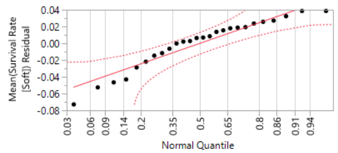


Figure 6-33: Mean Survival Rate x Year, Polynomial Line of Fit (Degree=6)

Figure 6-33 shows scatter plots of the mean survival rate for *undifferentiated*, *firm* and *soft* MSGD 1990–2015 data. Looking at the trend of *mean survival rate* of undifferentiated (all) data, it appears that global *Survival Rates*, having gradually increased from 1990 (~50%) to 2005 (~70%), but dropped back to below 50% by 2015. The firm data plot, while still a curve, starts lower and does not fall back to its 1990 levels. The means of *Survival Rate* for firm data also span a much greater range than those based on undifferentiated data; starting in the vicinity of 10% in 1990 and rising to a peak of around 60% by 2005. The trend for soft (or lesser veracity) data is oddly different from the other two plots. For soft data the *mean survival rate* starts and remains in the vicinity of 80% from 1990 to around 2010, it then dips to circa 60% over the space of five years.

To assess the statistical significance of these trends a polynomial line of fit of degree 6, which offers the best fit, is added to each plot. *(Note the measures of statistical significance applied are the proximity of R^2 to 1, the closer the better; the size of the F value, the larger the better; and Prob>F<0.05.)* The R^2 values of all three models indicate a good fit. While the F Value for firm data (=116.5916) is considerably higher than that of soft data (=27.8853), the probability of either F value is very small (Prob>F = <0.0001). Additionally, the diagnostic residual normal quantile plots confirm fit. Based on this the null hypothesis can be rejected as there is statistical significance between the year of disaster occurrence and the *mean survival rate*. With some assurance of statistical significance, ***Mean Survival Rate by Year* is considered a M^{i}O**.

The premise underpinning the search for M^{i}Os is that changes in a M^{i}O may signal the effectiveness of humanitarian intervention and the humanitarian response to disasters. This being the case it is unfortunate that at all levels of veracity the disaster data in the MSGD shows a downward trend in *mean survival rate* since 2010, indicating that fewer people survived disasters each consecutive year.

Revisiting *Figure 6-33*, the polynomial lines of fit also reveal an interesting drop in R^2 value from firm data (R^2=0.940824) to soft data (R^2 = 0.791777). This is considered worthy of further investigation from different perspectives. In support of this *Appendix G.4* contains plots and reports equivalent to *Figure 6-33* for each disaster group (*Naturogenic*, *Anthropogenic* and *Deviant*) and for each major geographic region (*Africa*, *Americas*, *Asia*, *Europe* and *Oceania*). Salient information from *Appendix G.4* is included and discussed here by disaster groups [*Figure 6-34* & *Table 6-10*] and by region [*Figure 6-35* & *Table 6-11*].

Survival Rate by Disaster Group

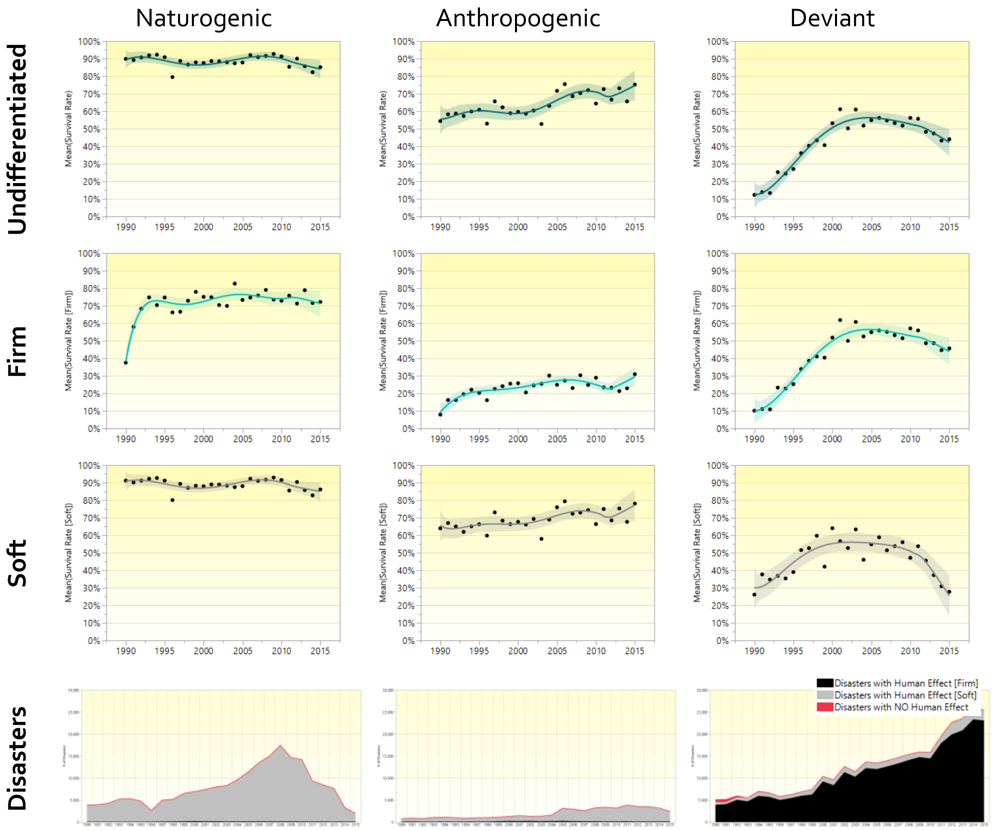


Figure 6-34:Disaster Groups x Mean Survival Rate
Polynomial Line of Fit (Degree=6)

Disaster Groups		Naturogenic	Anthropogenic	Deviant
Total Human Effect	Undiff.	5,947,074,115	13,050,322	697,295,668
	Firm	1,189,923,784	1,188,931	654,427,895
	Soft	4,757,150,331	11,861,391	42,867,773
Percentage	Firm	20%	9%	94%
	Soft	80%	91%	6%
R^2	Undiff.	0.454107	0.681799	0.948717
	Firm	0.840663	0.772979	0.953389
	Soft	0.443829	0.450127	0.775329
F Value	Undiff.	2.6342	6.7851	58.5816
	Firm	16.7073	10.7821	64.772
	Soft	2.527	2.5922	10.928
Prob > F	Undiff.	0.0496*	0.0006*	<.0001*
	Firm	<.0001*	<.0001*	<.0001*
	Soft	0.0571	0.0524	<.0001*

Table 6-10: MSGD 1990-2015 Disaster Groups by Veracity

Naturogenic Disasters

These disasters exhibit the highest *Survival Rate* of the disaster groups at all levels of veracity. The *Survival Rate* fluctuates at the 90% level for soft data. For firm data it starts below 40% in 1990, climbing steeply in the next three years to over 70%, remaining in this higher region for the remainder of the timeline.

Most of the naturogenic disaster entries are soft (80%) and the polynomial line of fit indicates that the null hypothesis cannot be rejected for soft naturogenic disaster data. Therefore, the level and trend of naturogenic *Survival Rates* based on soft data cannot be relied upon as an indicator. Whereas, based on the polynomial line of fit (and resulting R^2, F value, and Prob>F) for firm naturogenic disaster data the null hypothesis can be rejected.

Interestingly, the plot for undifferentiated naturogenic disasters is almost identical to that of soft naturogenic disasters. This suggests the 20% of firm human effects from naturogenic disasters do not have a discernible effect on the trend once absorbed in the undifferentiated numbers. In fact, the strength of influence of the weaker veracity data is also visible in the undifferentiated disasters' line of fit as here the R^2 and F value are weak and the Prob>F is borderline at 0.0496.

Ultimately, taking *Mean Survival Rate by Year* to be a M^{i}O, the circa 1.2 billion firm human effect numbers for naturogenic disasters may be enough to tell a reliable story. The plot indicates that it has remained in the vicinity of 70%, and relatively stable, for over 20 years. As a result conjectures can be drawn, e.g. *humanitarian intervention has ensured the outcome of naturogenic disasters has not deteriorated for over 20 years*. In any case, victims of naturogenic disasters fare better than victims of anthropogenic or deviant disasters.

Anthropogenic Disasters

The *Survival Rate* plots of both soft and firm anthropogenic disasters exhibit a slight upward slope, but their range of movement is dramatically different. The firm data's *mean survival rate* range is around 10%–30%, while the soft data's range is around 60%–80%. It appears that less reliable data provides a much rosier view of the outcome of this group of disasters. As with naturogenic disasters, the polynomial line of fit results for firm anthropogenic disaster data are such that the null hypothesis can be rejected, while for soft anthropogenic disaster data disaster data the null hypothesis cannot be rejected. This being the case, only the plot based on firm anthropogenic human effect numbers offers a statistically significant view of *Survival Rate*. As the best achievement of firm anthropogenic disasters is a *mean survival rate* of 30%. A conjecture from this could be that *humanitarian intervention is relatively ineffectual in affecting the outcome of anthropogenic disasters*.

Deviant Disasters

The *Survival Rate* plots for both firm and soft data show differing trends and over different ranges for deviant disasters. The firm data plot rises from around 10% in 1990 to around 60% 2005, and then slowly slopes down to around 45% by 2015. The soft data plot starts in the region of 25%, rising to around 65% by 2000, plateauing from 2000 to 2010, before dropping back to near 25% by 2015. Notably the results of the polynomial line of fit for both firm and soft deviant disaster data veracity levels confirm that the null hypothesis can be rejected, therefore both indicate statistical significance. As soft data represents only 6% of deviant data, the firm data plot is taken here to be a more substantive data story. This plot rises steadily from 10% to 60% over the first fifteen years examined then gradually falls to 50% over the next ten years. This could be construed to mean *humanitarian intervention having made inroads in improving outcomes of deviant disasters, has been falling short for the most recent ten years*.

Comparing Disaster Groups

For all three disaster groups the focus here has been on firm and soft data, with the undifferentiated data plots and reports included to provide perspective of the effect of blending the veracity levels. Interestingly, examining *Survival Rates* as the macro-level outcome of humanitarian intervention provides varying perspective of the relative success of these efforts.

For naturogenic disaster there seems to be a near status quo, in that the possibility of surviving a naturogenic disaster has barely changed in twenty-six years. The implication being that increased funding and the burgeoning number of actors now operating in the humanitarian domain have stopped the survival rate from deteriorating, but have yet to improve it (Lattimer et al., 2016; Purvis, 2015; UNESCO-UIA, 2017; U.S. Dept. of State, 2017; Shukla, 2010).

For anthropogenic disasters the trend in *mean survival rate* is upwards, but if relying on only veracious data, it can be seen that the chances of survival remains woefully low even at their best of 30%. This raises the question of whether disaster preparation, mitigation and response efforts are geared to man-made disasters.

For deviant disasters, which include conflict, terrorism, famine and deracination, the data exhibits a recent downward trend, with an approximate 50:50 being the most recent chance of survival. As these disasters are typically the result of human intent or apathy, this even chance of surviving a deviant disaster may be mirroring the tug-of-war between humanitarian efforts and the aggression or disinterest that underpin such disasters.

An impression formed in comparing these three groups of disasters is that while humanitarian funds may be spent across all three groups (Lattimer et al., 2016; Purvis, 2015), disaster management and humanitarian response efforts appear to be honed to, and therefore more successful in, the naturogenic group. Of note is that, of the three groups the only group with a significantly high proportion of veracious human effect numbers, is deviant disasters; a group in which data acquisition may arguably be the most challenging (and hazardous) for data collectors. Therefore, the view of naturogenic and anthropogenic disasters has to be based on the relatively meagre quantities of veracious data that is found for these groups of disasters.

Survival Rate by Region

The regional models of *mean survival rate*s all show statistical significance [*Figure 6-35* & *Table 6-11*]. The R^2 values are consistently weaker for soft data than for firm data, a difference that is most marked for Europe. The proportion of firm to soft data is less extreme for each of the regions than when the data is split by disaster group, yet there appears to be no coincidence between the model statistics – R^2 and F Value – and the split of firm to soft data. As this view of the data spreads the source datasets across regions – something that does not happen when the data is viewed by disaster group – this effectively absorbs dataset weaknesses rendering them indiscernible.

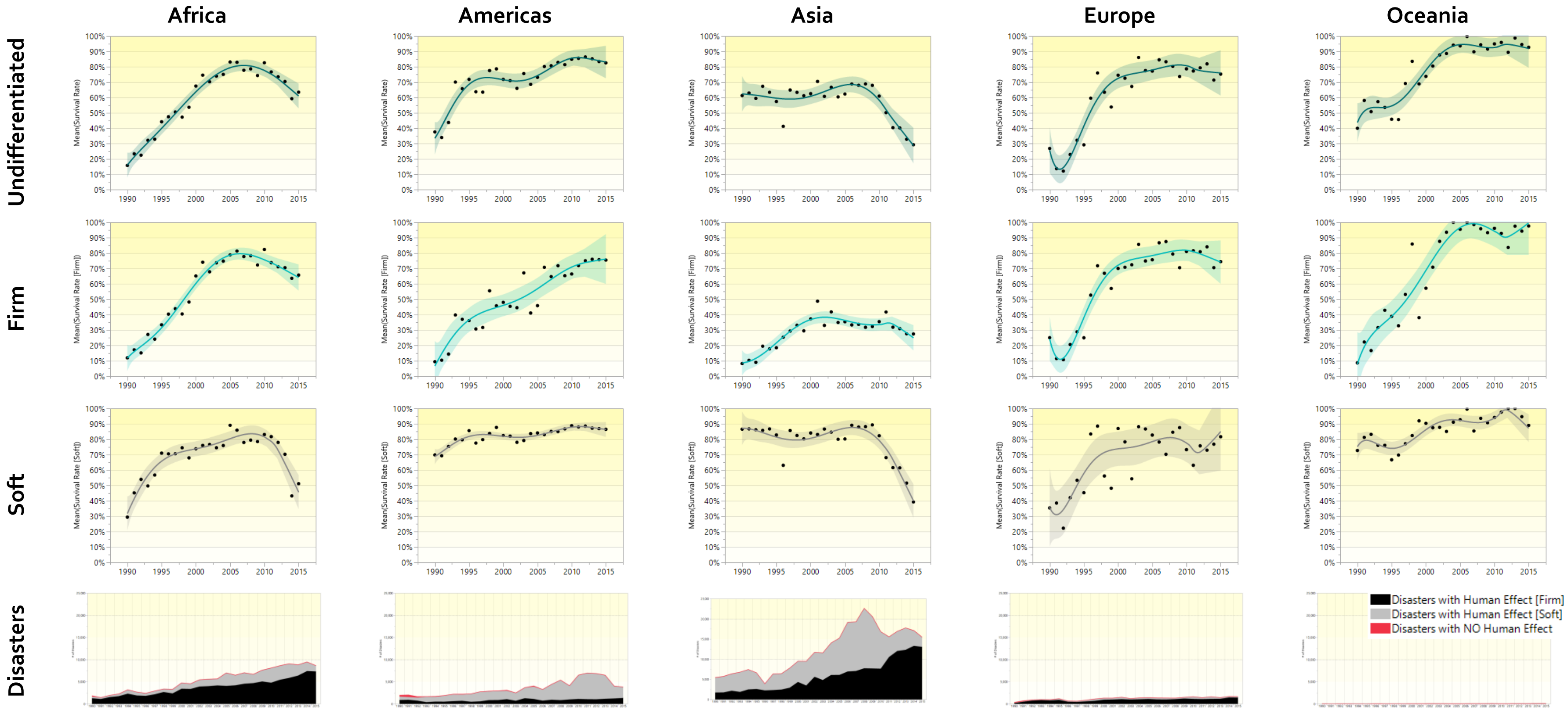


Figure 6-35: Fit Line – Regions x Mean Survival Rate, Polynomial Line of Fit (Degree=6)

Regions		Africa	Americas	Asia	Europe	Oceania
Total Human Effect	Undiff.	836,249,484	459,663,110	5,201,658,053	90,464,100	30,091,494
	Firm	256,523,419	127,456,975	1,370,517,100	66,196,939	14,992,850
	Soft	579,726,066	332,206,135	3,831,140,953	24,267,161	15,098,644
Percentage	Firm	31%	28%	26%	73%	50%
	Soft	69%	72%	74%	27%	50%
R^2	Undiff.	0.969052	0.888774	0.800276	0.921465	0.91432
	Firm	0.972667	0.878112	0.865476	0.935447	0.908414
	Soft	0.887984	0.832161	0.856173	0.653295	0.813713
F Value	Undiff.	99.1565	25.304	12.6886	37.1551	33.7927
	Firm	112.6873	22.8135	20.3731	45.8889	31.4093
	Soft	25.103	15.7006	18.8505	5.9669	13.8322
Prob > F	Undiff.	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*
	Firm	<.0001*	<.0001*	<.0001*	<.0001*	<.0001*
	Soft	<.0001*	<.0001*	<.0001*	0.0012*	<.0001*

Table 6-11: MSGD 1990-2015 Regions by Veracity

As the null hypothesis is rejected for all five regional models, the distinctive downward trends for Africa and Asia are considered credible, troubling therefore worthy of further investigation.

(a) Africa

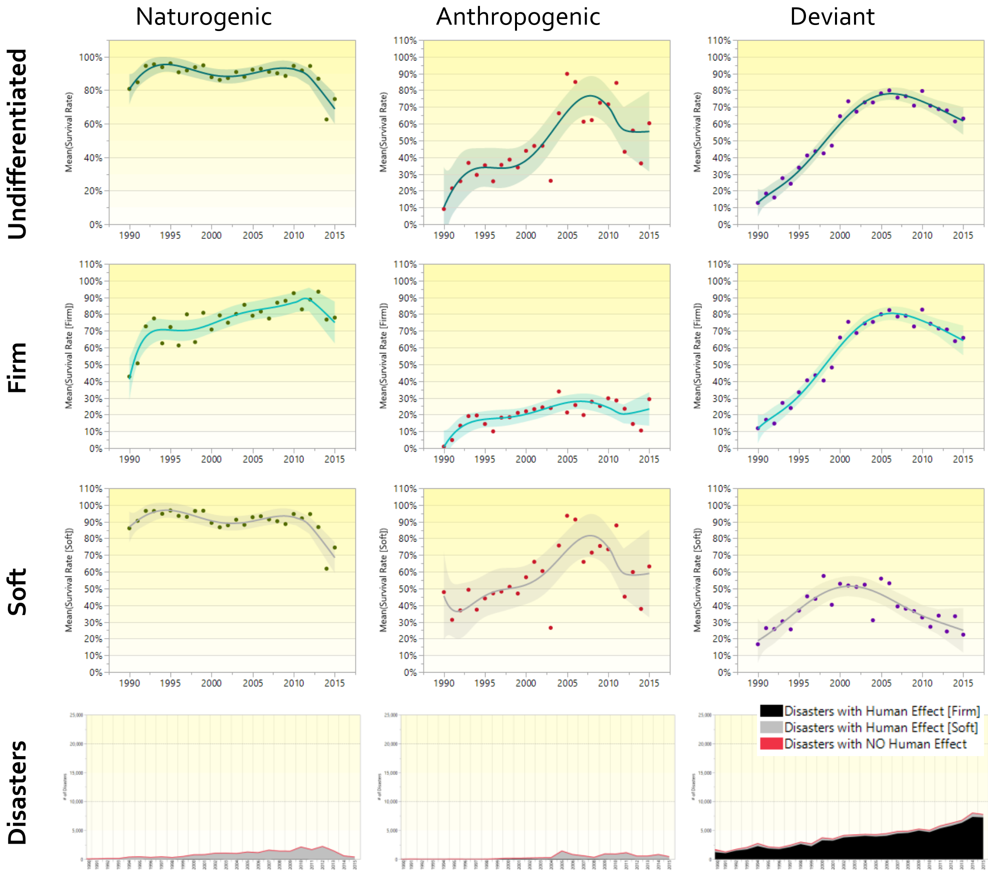


Figure 6-36: Africa Disaster Groups x Mean Survival Rate
Polynomial Line of Fit (Degree=6)

AFRICA by Disaster		Naturogenic	Anthropogenic	Deviant
Total Human Effect	Undiff.	620,501,741	928,235	214,819,508
	Firm	47,963,506	158,786	208,401,127
	Soft	572,538,235	769,449	6,418,382
Percentage	Firm	8%	17%	97%
	Soft	92%	83%	3%
R^2	Undiff.	0.685236	0.746267	0.971717
	Firm	0.781614	0.668321	0.971421
	Soft	0.692735	0.582549	0.740564
F Value	Undiff.	6.8938	9.3136	108.7973
	Firm	11.3337	6.3807	107.6368
	Soft	7.1393	4.419	9.0393
Prob > F	Undiff.	0.0005*	<.0001*	<.0001*
	Firm	<.0001*	0.0008*	<.0001*
	Soft	0.0004*	0.0058*	<.0001*

Table 6-12: Africa MSGD 1990-2015 Disaster Groups by Veracity

Figure 6-36 and *Table 6-12* are Africa-specific version of the disaster group plots and polynomial line of fit models created earlier. A simple visual scan identifies that the overall Africa *mean survival rate* plot is almost identical to the deviant disaster Africa *mean survival rate* plot [*Figure 6-37*]. In both cases the underlying firm data shapes the overall plots, which is not unexpected for deviant disasters in Africa as the human effect numbers are 97% firm [*Table 6-12*]. It is however surprising to see that this is also the case for the overall Africa data, where the human effect numbers are only 31% firm [*Table 6-11*].

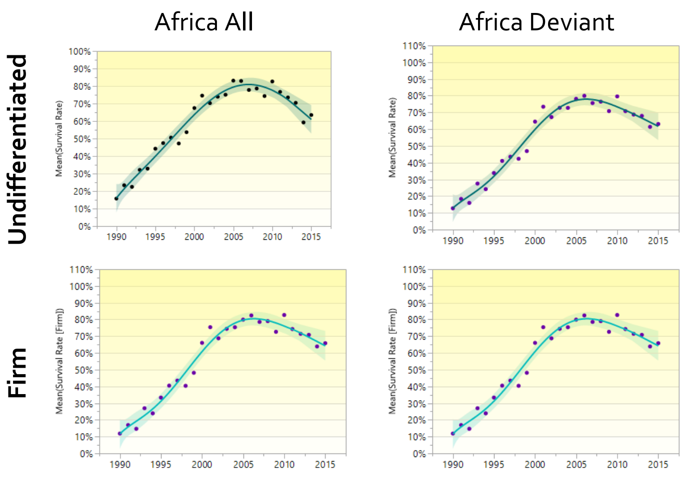


Figure 6-37: Africa All vs. Africa Deviant

Even more unexpected is the lack of influence the soft naturogenic numbers have on the overall shape of the Africa's regional plot, considering the ***soft*** naturogenic human effect numbers (>572.5 million) dwarf all other disaster group and levels of veracity of human effect numbers in Africa, even when combined [*Table 6-12*]. Yet these soft naturogenic values do not appear to hold sway in changing the plot of the overall *mean survival rate* Africa plot.

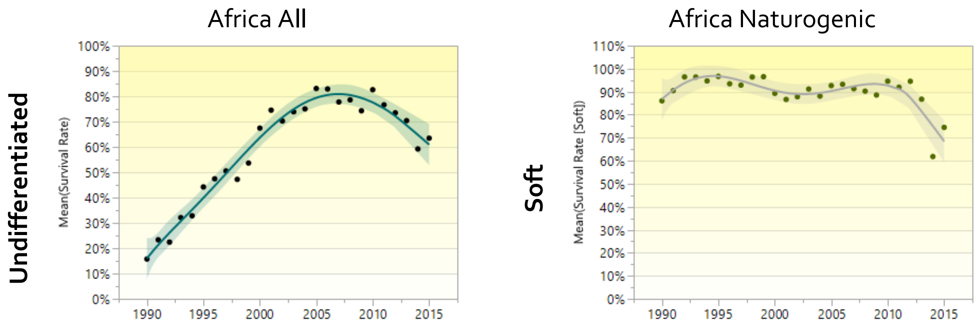


Figure 6-38: Africa All (undifferentiated) vs. Africa Naturogenic (Soft)

That said, this phenomenon is believed to be the result of the volume of disaster entries per disaster group, as can be seen in the '*Disaster*' area charts [*Figure 6-36*]. Deviant disaster entire volume far exceeds that of the other two disaster groups in the region. Therefore –

- *if most of the disasters that occur in Africa are deviant, and;*
- *in the aggregate, disaster management and humanitarian aid is such that mean survival rate is declining for deviant disasters, therefore it is understandable that the mean survival rate in Africa is declining.*

(b) Asia

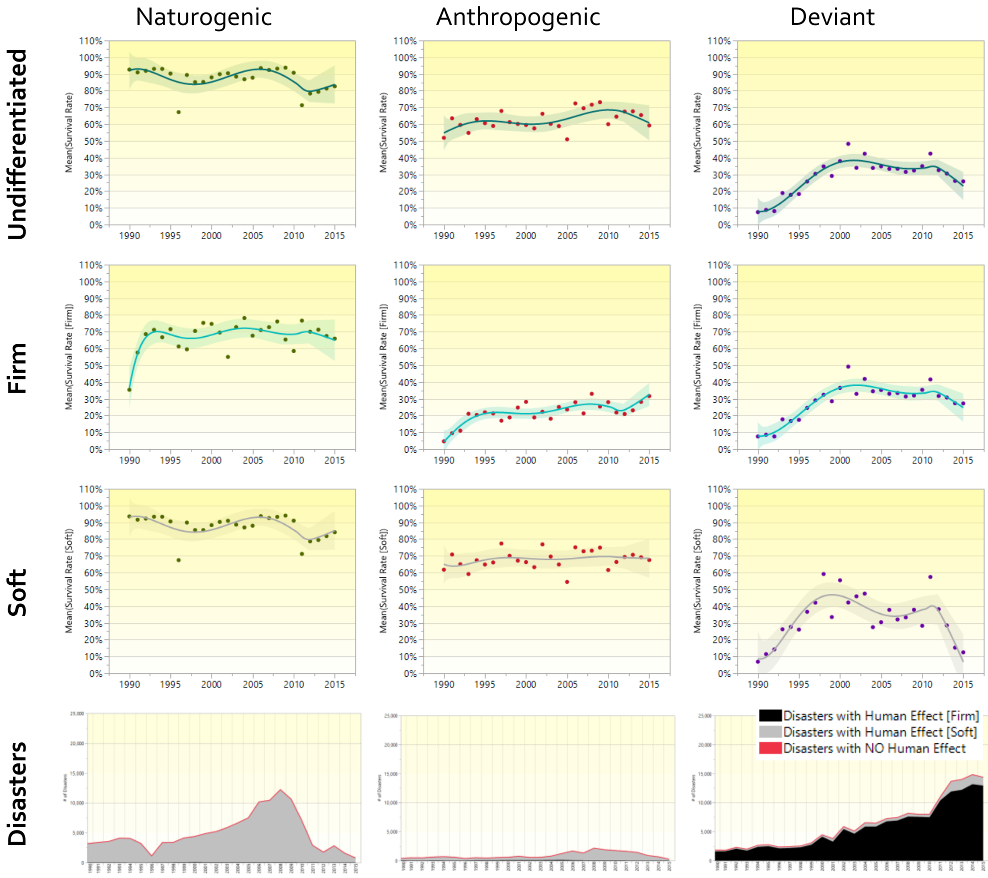


Figure 6-39: Asia Disaster Groups x Mean Survival Rate
Polynomial Line of Fit (Degree=6)

ASIA by Disaster Group		Naturogenic	Anthropogenic	Deviant
Total Human Effect	Undiff.	4,874,539,882	3,122,248	323,994,803
	Firm	1,052,658,128	742,468	317,116,504
	Soft	3,821,881,754	2,379,780	6,878,299
Percentage	Firm	22%	24%	98%
	Soft	78%	76%	2%
R^2	Undiff.	0.445048	0.356142	0.869455
	Firm	0.613099	0.766578	0.86884
	Soft	0.44045	0.085767	0.722532
F Value	Undiff.	2.5395	1.7516	21.0907
	Firm	5.018	10.3996	20.9768
	Soft	2.4926	0.2971	8.2461
Prob > F	Undiff.	0.0562	0.1634	<.0001*
	Firm	0.0031*	<.0001*	<.0001*
	Soft	0.0598	0.2971	0.0002*

Table 6-13: Asia MSGD 1990-2015 Disaster Group by Veracity

Figure 6-39 and *Table 6-13* are Asia-specific versions of the disaster group plots and polynomial line of fit models created earlier. A simple visual scan confirms the disaster group plots bear little resemblance to the higher level regional plot for Asia; this likely to be a product of the mix of soft versus firm disaster entries and human effect in the regional data.

Interestingly, when taken to this level of detail, in Asia the only models that remain statistically significant across all three disaster groups are for firm data. These provide a different perspective of *mean survival rates* in the region than the regional plot of Asia which shows a sharp decline from around 70% *mean survival rate* in 2008 to 30% by 2015. Taken at face value the implication of this could be taken to mean the effect of disaster management and humanitarian aid has been diminishing rapidly in this region. That said, when examining the firm data plots of each of the three disaster groups in the region, the interpretation could be very different:

- For naturogenic disasters *mean survival rate* having risen steeply from 1990 (~35%) to 1993 (~70%) has remained in the vicinity of 70% for the remaining 23 years studied. – The implications of this could be, *humanitarian intervention in the region is capable of maintaining a near stable outcome for naturogenic disasters*.
- For anthropogenic disasters *mean survival rate* rose from 5% in 1990 to around 20% in 1995, but has only very gradually improved to 30% by 2015. – The implications of this could be that *humanitarian intervention is ill-equipped to deal with man-made disasters in the region and the best that has been achieved is still only an overall three out of ten people surviving a man-made disaster in Asia*.
- For deviant disasters *mean survival rate* rose from below 10% in 1990 to just above 40% by 2000, but appears to be meandering gradually downwards since then, reaching below 30% by 2015. – The implications of this could be that *since the turn of the century*

> *disaster management and emergency response efforts to address crises such as conflict, terrorism, famine etc. are inadequate to the task, therefore in the aggregate, by 2015 only three in ten people are likely to survive a deviant disaster in Asia.*

To review, it is taken here to be a given that disaster management and humanitarian aid act (or should act) to address all humanitarian crises in order to minimising the effects of disasters to the best of their ability. As *Survival Rate* is taken in this study to be a measure of an effect of disasters, changes in *mean survival rate* is interpreted to reflect the ability of disaster management and humanitarian aid efforts to minimise the human devastation caused by the crises.

The analyses of the MSGD and the various lines of fit models have provided an interesting perspective of survival rate changes by disaster groups and by geographic regions. While, even the most statistically significant of the line of fit models cannot be used to identify a definitive *cause* of better or worse chances of survival, they do provide an indicator (M^iO) of which regions and groups of disasters, merit further investigation of disaster management efforts and humanitarian funding.

6.5 Evaluate

The evaluation step in this iteration follows the same blueprint for evaluation as the previous two iterations by discussing the alignment of the *build (grow)* step with the *tentative design*; the *DSR artefacts* created in this iteration; the *knowledge* gained and the *consequences* of this knowledge; and finally testing the propositions of the *design theory* or *theories* relevant to this iteration. (Hevner et al., 2004; Simon, 1996; Hooker, 2004; Venable et al., 2012; Venable, 2013; Gregor, 2006; Vaishnavi and Kuechler, 2004b).

6.5.1 Tentative Design⇔Build (Grow) Alignment

The *tentative design* for this iteration is essentially about the construction of three sets of artefacts: (1) the Master Set of Global Disasters (MSGD); (2) the Master Disaster Classification Model (MDC); (3) the Macro-Indicators of Outcome (MiO). Of these, the first two are considered *data scaffolds* for the humanitarian domain without which the contextual relevance of data analyses carried for the domain in general and this study in particular are severely constrained, if not impeded. The last aligns to the aim and research question of this study. Taking each artefact set in turn:

(1) the Master Set of Global Disasters (MSGD)

This includes: (a) data sources; (b) the weightings applied to the DVm L3 dimensions for data veracity evaluation; (c) the structure of the MSGD. Evaluating the alignment of the *tentative design* to what is built:

(a) MSGD data sources

The five disaster loss dataset identified to augment EM-DAT for the MSGD are acquired, prepared and examined as per the *tentative design* (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017b). Notably, the scale of effort and time needed for this exercise belies the simplicity of the design, which principally selects, and justifies the selection of, disaster loss data to provide a more panoramic view of humanitarian crises. As an example, the acquisition and preparation of DesInventar alone was a complex and inordinately labour-intensive exercise involving two source sites, each with numerous datasets – for the same and different countries – holding mostly unvalidated data in varying data structures and in a variety of languages.

Additionally, while these datasets broaden the view of humanitarian crises they do not fill all the gaps in key information. Financial losses, where found, are sparingly or confusingly populated. For example, local currency losses can be found with USD losses, where the USD losses may be conversions of the local currency losses or in addition to the local currency losses. In essence, while the *tentative design* of the MSGD aligns to the MSGD instantiation, compromises are made and gaps remain because of the limitations of the data available.

(b) DVm L3 weightings

The weightings as specified in the tentative design are used without issue and incident and appear to fulfil the objective of creating meaningful and comparable Data Veracity indices (DVis) for each of the source datasets and the MSGD as a whole.

(c) MSGD structure

The data structure specified in the tentative design is used without modification to hold data from all six of the MSGD source datasets.

(2) the Master Disaster Classification Model (MDC)

The pragmatic design principle underlying the creation of the MDC is that it only needs to be *sufficiently inclusive* to classify all events held in the source datasets selected for the MSGD. This principle of *sufficiency*, however, comes into conflict in the adoption of IRDR Peril and Hazard Glossary as the first building block of the MDC (IRDR, 2014). In keeping true to the structure of the IRDR's glossary (for naturogenic disasters), sixteen classifications in the MDC are created for disasters that have never occurred. Additionally, the disaster loss sources used to build the MSGD are found to be more pertinent to the structure of the MDC than specific classes found in the alternative classification schemes identified in the *tentative design* (Lerner, 2016; Coburn et al., 2014).

(3) the Macro-Indicators of Outcome (M^iO)

The *tentative design* of this iteration suggests that Survival Rate may offer some potential as a M^iO. The *build (grow)* step of this iteration confirms that using the MSGD 1990–2015 dataset *mean survival* *rate* does exhibit statistically significant trends at the macro-level. Therefore it may signpost the effects, or inadequacies, of preventative and corrective disaster management efforts and humanitarian funding. Notably, this is the only M^{i}O identified from this data as the base data for deaths and people affected show no discernible pattern of behaviour. Moreover, financial losses are missing from 98.5% of the MSGD entries and there is little prospect of identifying a trend relevant to disaster management or humanitarian aid from the remaining 1.5% of entries.

6.5.2 DSR Artefacts

This iteration adds to the pool of research artefacts [*Table 6-14*]:

<table>
<tr><th colspan="3" rowspan="3">Research Framework</th><th colspan="4">Research Activities</th></tr>
<tr><th colspan="2">Design Science</th><th colspan="2">Natural Science</th></tr>
<tr><th>Build (grow)</th><th>Evaluate</th><th>Theorise</th><th>Justify</th></tr>
<tr><td rowspan="4">Research Outputs</td><td rowspan="4">Artefacts</td><td>Constructs</td><td>[a][f]</td><td>→</td><td>[c]</td><td>→</td></tr>
<tr><td>Models</td><td>[d][g][h]</td><td>→</td><td></td><td></td></tr>
<tr><td>Methods</td><td>[e]</td><td>→</td><td></td><td></td></tr>
<tr><td>Instantiations</td><td>[b][i][j][k][l]</td><td>→</td><td></td><td></td></tr>
<tr><td>[a]</td><td colspan="6">Macro-indicators of disaster outcome and the impact and effectiveness of humanitarian intervention (MiOs, MiIs and MiEs).</td></tr>
<tr><td>[b]</td><td colspan="6">Data analysis outputs and visualisations</td></tr>
<tr><td>[c]</td><td colspan="6">A (behavioural science) hypothesis relating the availability, or lack thereof, of humanitarian data and the flow of humanitarian aid that emerges from the domain knowledge and may be worthy of future research.</td></tr>
<tr><td>[d]</td><td colspan="6">Data Veracity framework (DVf) and Data Veracity model (DVm)</td></tr>
<tr><td>[e]</td><td colspan="6">Data Veracity profile (DVp) and Data Veracity index (DVi)</td></tr>
<tr><td>[f]</td><td colspan="6">Expansion of the construct of 'data scaffolds' for the humanitarian domain</td></tr>
<tr><td>[g]</td><td colspan="6">Data structure of the Master Set of Global Disasters (MSGD)</td></tr>
<tr><td>[h]</td><td colspan="6">Classification structure of Master Disaster Classification Model (MDC)</td></tr>
<tr><td>[i]</td><td colspan="6">Master Set of Global Disasters dataset</td></tr>
<tr><td>[j]</td><td colspan="6">Master Disaster Classification reference dataset</td></tr>
<tr><td>[k]</td><td colspan="6">Data Veracity profile (DVp) and Data Veracity index (DVi) instances for each of the six datasets amalgamated for the MSGD</td></tr>
<tr><td>[l]</td><td colspan="6">Mean Survival Rate by Year as an actualised MiO</td></tr>
</table>

Table 6-14: DSR Output to Research Framework Mapping v.3
(Vaishnavi and Kuechler, 2004b; Hevner, 2007; March and Smith, 1995)

This table [*Table 6-14*] maps these additional DSR artefacts to the research framework of *Section 3.3.3* (March and Smith, 1995; Vaishnavi and Kuechler, 2004b; Hevner, 2007). The constructs of M^{i}O and '*data scaffold*', introduced earlier in this study are explored further in this chapter and argued by way of example as being viable as new constructs in the domains of humanitarian intervention and data science respectively. The MSGD and MDC are developed as examples and manifestations of the humanitarian domain's '*data scaffold*' and *mean survival rate* is identified as a M^{i}O. It is not argued here that the MSGD and MDC are the only two data scaffolds needed in the humanitarian domain, but that these are two structural supports without which this study would not be possible. Similarly *mean survival rate by year* is presented here as an example of a M^{i}O that can be created from the disaster data available.

There is a potential weakness in both of the *model-type* artefacts created here. The MSGD structure and the MDC model are pragmatic and bespoke solutions that balance the needs of this study and the availability of data. These artefacts do not represent an in-depth study of what should be held in a master dataset of disaster losses or a classification system that should encompass all possible disaster types. This is outside the scope and resources of this research. This is why the instantiations of these models as datasets are considered prototypes that serve the purposes of this work.

The data veracity instantiations created here test the DVf developed in *Chapter 5*. It is found that using the DVp to evaluate the veracity of each dataset ensured disciplined and equitable assessment, protecting against tacit and undocumented knowledge of the weakness of the data. Similarly, the process of weighing-up the relative importance of each of the veracity dimensions and then scoring each dataset's fit to these dimensions forces a more considered evaluation of data veracity. It also provides a metric that allows comparison between datasets and an assessment of any innate weaknesses of amalgamated datasets.

The instantiations of exploratory visualisation and summaries are created *for each dataset* added to the MSGD and for the *amalgamated* MSGD. Created here as a means to an end they are reusable products that can have utility beyond this research. They provide insights as to the contribution and data gaps of the various data sources and help explain the compromises needed to make disparate data structures compatible enough to build the MSGD. Furthermore, they help clarify the scope and limitations of each dataset, illustrating why key avenues of analysis are not pursued, e.g. patterns in financial losses or seasonal changes in survival rate.

Finally, ***mean survival rate by year*** emerges as a **M^{i}O** as it exhibits statistically significant patterns of change that may be indicative of the influence of humanitarian funding on the outcome of disasters. Mean *survival rate* plots also offer some validation of the data veracity evaluations, as statistically meaningful *mean survival rate* trends appear to be closely related to the strength of data veracity.

6.5.3 Knowledge ⇔ Consequence

Nuggets of knowledge that emerge from this iteration include:

(a) All disaster datasets used have gaps

The six disaster datasets used for the MSGD all have gaps in potentially useful information (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017b). UCDP, and VDC-SY hold no financial data, therefore the financial effect of conflict e.g. the cost of lost homes, hospitals, roads, schools etc. is not available (UCDP, 2017b; VDC-SY, 2016b). EM-DAT, DesInventar and GTD are designed to hold financial data but have so little of it that it cannot be used (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; GTD, 2017e). UCDP records fatalities but not how many are wounded or made homeless as a result of each incident (UCDP, 2017b). UNHCR provides annualised movement of people which forces all other datasets to yearly views to maintain compatibility (UNHCR, 2017b). Also as the UNHCR PoC download is annualised, even if the cause of each deracination is held in the underlying database, it could not be included in the download.

(b) If it's not deviant it is likely to be *weak*

The six disasters datasets fall into two broad categories: those for events conventionally considered to be '*disasters'*, e.g. hurricanes, earthquakes etc.; and those for events that are humanitarian crises, but not typically labelled as disasters, e.g. genocide, displaced people etc. The veracity evaluation of the six datasets revealed that datasets that hold predominantly non-deviant *disasters*, i.e. conventional disasters, are the least veracious – EM-DAT and DesInventar (Guha-Sapir et al., 2017l; DesInventar.NET, 2017). Whereas those related to conflict appear to be much more diligently collected and curated (UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e).

(c) Most human effect in disaster data is 'soft'

Most human effect values in the MSGD 1990–2015 are from soft, less veracious, entries [*Table 6-15*]. Therefore, an important caveat in interpreting any metric calculated from these values is that the results cannot be considered definitive, at best they are indicative. Notably, if the data is segregated by datasets that exclusively hold the deviant group of disasters and those that can hold a mix of disaster groups, the proportions of firm to soft data change radically. The 'deviant only' datasets are more veracious than not with only 4% of entries identified as soft [*Table 6-15*]. In contrast, the remaining datasets comprise mostly of weak veracity data, with almost 80% of the data identified as soft. Additionally, with deviant disasters equalling only just 10% of the MSGD 1990–2015, firm data is overwhelmed by soft data when all entries are viewed collectively.

MSGD 1990-2015 Human Effect		Veracity Split	
		firm	soft
90%	EM-DAT & Desinventar	20%	80%
10%	*Deviant Disaster Datasets*	96%	4%

Table 6-15: MSGD 1990–2015 Deviant vs Other (Firm/Soft Data)

(d) The best *Survival Rates* are from the weakest data

Soft data appears to exhibit the highest survival rate *Figure 6-40* (*extract of Figure 6-33*) This is reinforced in *Figure 6-41,* which shows that the deviant disaster group (with the lowest proportion of soft data) also has the poorest survival rate for disaster groups:

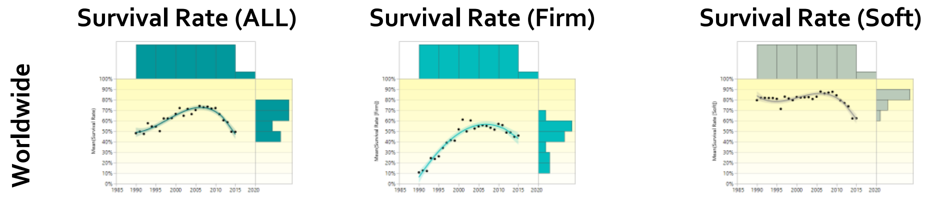


Figure 6-40: MSGD 1990–2015 Survival Rate x Year

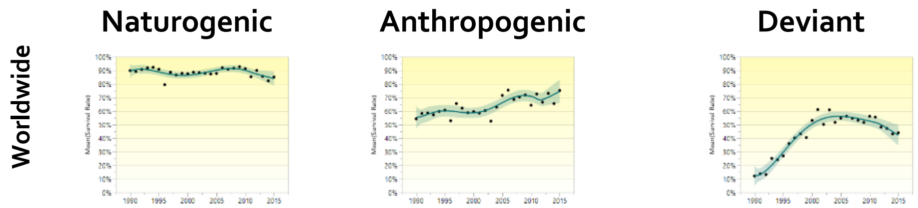


Figure 6-41: MSGD 1990–2015 Worldwide *Mean Survival Rate by* Disaster Group

Naturogenic and anthropogenic disaster loss data are exclusively sourced from EM-DAT and DesInventar [*Figure 6-42*] (Guha-Sapir et al., 2017l; DesInventar.NET, 2017).

Disaster Group	Dataset	Total Human Effect					
		FIRM		SOFT		ALL	
		People	%age of ALL	People	%age of ALL	People	% of Total
Naturogenic	EM-DAT	1,189,923,784	23%	4,042,996,793	77%	5,232,920,577	78.60%
	DesInventar		0%	714,153,538	100%	714,153,538	10.73%
Anthropogenic	EM-DAT	1,188,931	36%	2,155,478	64%	3,344,409	0.05%
	DesInventar		0%	9,705,913	100%	9,705,913	0.15%
Deviant	EM-DAT	10,047,214	50%	9,948,900	50%	19,996,114	0.30%
	DesInventar		0%	3,000,308	100%	3,000,308	0.05%
	UCDP	1,831,959	84%	349,333	16%	2,181,292	0.03%
	VDC-SY	116,870	54%	98,452	46%	215,322	0.00%
	GTD	158,268	92%	14,377	8%	172,645	0.00%
	UNHCR	642,273,584	96%	29,456,403	4%	671,729,987	10.09%
Unspecified	DesInventar		0%	1,140	100%	1,140	0.00%
	Total	1,845,540,610	27.7%	4,811,880,635	72.3%	6,657,421,245	100%

Figure 6-42: MSGD 1990 – 2015 Disaster Group/Data Source Veracity split

The human effect values on which these plots are based have very high proportions of soft data. All DesInventar values are flagged as soft and EM-DAT human effect numbers for naturogenic disasters are 77% soft and for anthropogenic disasters are 64% soft.

(e) Weak data veracity weakens statistical significance

Comparing R^2, F Value and Pro>F of disaster groups [*Figure 6-43*]:

Disaster Groups	Total Human Effect Firm		R^2	F Value	Prob > F
Naturogenic	1,189,923,784	20.0%	0.84066	16.7073	<.0001*
Anthropogenic	1,188,931	9.1%	0.77298	10.7821	<.0001*
Deviant	654,427,895	93.9%	0.95339	64.772	<.0001*

Disaster Groups	Total Human Effect Soft		R^2	F Value	Prob > F
Naturogenic	4,757,150,331	80.0%	0.44383	2.527	0.0571
Anthropogenic	11,861,391	90.9%	0.45013	2.5922	0.0524
Deviant	42,867,773	6.1%	0.77533	10.928	<.0001*

Figure 6-43: MSGD 1990–2015 Disaster Group Veracity & Statistical Significance

For naturogenic disasters human effect values are only 20% firm, the R^2 of the *mean survival rate* line of fit drops from 0.84066 for data flagged as veracious to 0.44383 for data of weaker veracity. Similarly, the F Value, drops from 16.7073 to 2.527. Most interesting is that even the small proportion of data that is firm for naturogenic disasters results in a Prob>F that allows the null hypothesis to be rejected; whereas this is not the case for the 80% of soft human effect values.

This pattern repeats for anthropogenic disasters. Firm data R^2 is 0.77298, while soft data R^2 is 0.45013. F Value drops from 10.7821 for firms data to 2.5922 for soft data. Once again the very small proportions of firm data (9.1%) results in a statistically significant plot, while the large proportion (90.9%) of soft data does not. Only deviant disasters have statistically significant *mean survival rate* plots for both firm and soft data, but even then there is a notable downward shift of R^2 and F Value between firm and soft data. This coincidence of the polynomial line of fit of *Mean Survival Rate by Year* and the veracity of human effect data appears to validate the data veracity evaluations carried out for the data. In that, meaningful results are less likely when the veracity of the data is weak. *Table 6-16* maps knowledge gleaned from this iteration to its consequence.

<table>
<tr><th>Knowledge</th><th>Evaluate</th><th>Consequence</th></tr>
<tr><td>(a) All disaster datasets used have gaps</td><td>These are believed to be best datasets available, therefore the gaps are because the needed information is never collected or is not made available to the public.</td><td>This study has no choice but to use the master dataset created as no other data solution is available. Compromises are made in the creation of the dataset, but the interpretation of any analysis from this data carries the caveat that even the most comprehensive view of disaster data cannot be considered to be a complete view.</td></tr>
<tr><td>(b) If it's not deviant it is likely to be weak</td><td rowspan="2">Unfortunately, estimations play a significant part in recording disaster losses. Deviant disasters are more typically conflict, terror or refugee related. Of these data types, the collection for highly volatile 'aggression' related humanitarian crises is likely to be hazardous, therefore more challenging to obtain. Yet, the UCDP, VDC-SY and GTD datasets are the most veracious of the datasets used. The persons of concern data maintained by UNHCR are not as veracious, yet still more reliable than EM-DAT and DesInventar which hold the bulk of the human effect numbers.</td><td rowspan="2">While it is outside the gift of this research to resolve the veracity weaknesses of the sourced data. Flagging the weakness is important to ensure the analytical results from this data carry a health warning when the data used is of weak veracity.</td></tr>
<tr><td>(c) Most human effect in disaster data is 'soft'</td></tr>
<tr><td>(d) The best Survival Rates are from the weakest data</td><td>Interestingly there appears to be a relationship between weaker veracity and better mean survival rates, which give the impression that guestimated figures may in fact be optimistic.</td><td>When viewing mean survival rate as a M^{I}O, weak veracity data is best considered as overestimated.</td></tr>
<tr><td>(e) Weak data veracity weakens statistical significance</td><td>The inclusion or use of weaker veracity data coincides with weaker or no statistical significance of plots of mean survival rate over years. This is an unexpected but useful finding that helps validates the use of the data veracity toolset.</td><td>This raises questions to how best to use Survival Rate as a M^{I}O. Should mean survival rate only be considered a viable indicator if it is based on firm data? Does statistical significance mean the data, though soft, is not far off 'reality'? In which case, should statistically significant undifferentiated or soft mean survival rate plots also be considered?</td></tr>
</table>

Table 6-16: Iteration 3 Knowledge ⇔Consequence Mapping

6.5.4 The Utility Theory

Restating the utility theory statement of this study [*Table 6-17*]:

	Solution Space	Utility	Problem Space
	Form	Function	Purpose
	Artefact [*What*]	Efficacy [*How*]	to Address [*Why*]
STATEMENT	*Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector…*	*…when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level…*	*…the inability to gauge the consequences of monies spent and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.*

Table 6-17: Structure of the Utility Theory Statement

The behaviour of the *mean survival rate* as it is plotted against each elapsed year provides support to this utility theory. In that, the statistically significant trend in a *mean survival rate* plot may be signalling humanitarian intervention. For example, declining *mean survival rate* may be signposting an inadequacy in life-saving efforts. Also, the difference in range and fluctuation of *mean survival rate* across the three disaster groups (naturogenic, anthropogenic, deviant) could be interpreted as the ability of the humanitarian sector to influence the outcome of disasters varying by disaster group.

Figure 6-44 illustrates the dependency flow and relationship between the constructs relevant to this iteration within the utility theory.

- The macro-indicator construct of M^iO relies on the construct of *Data Scaffolds*, manifested as the artefacts of the MSGD and MDC (*from this iteration*).
- The construct of *Data Scaffolds* (the MSGD and MDC) is in turn dependent on the construct of the *Data Veracity framework* (*from Iteration 2*).

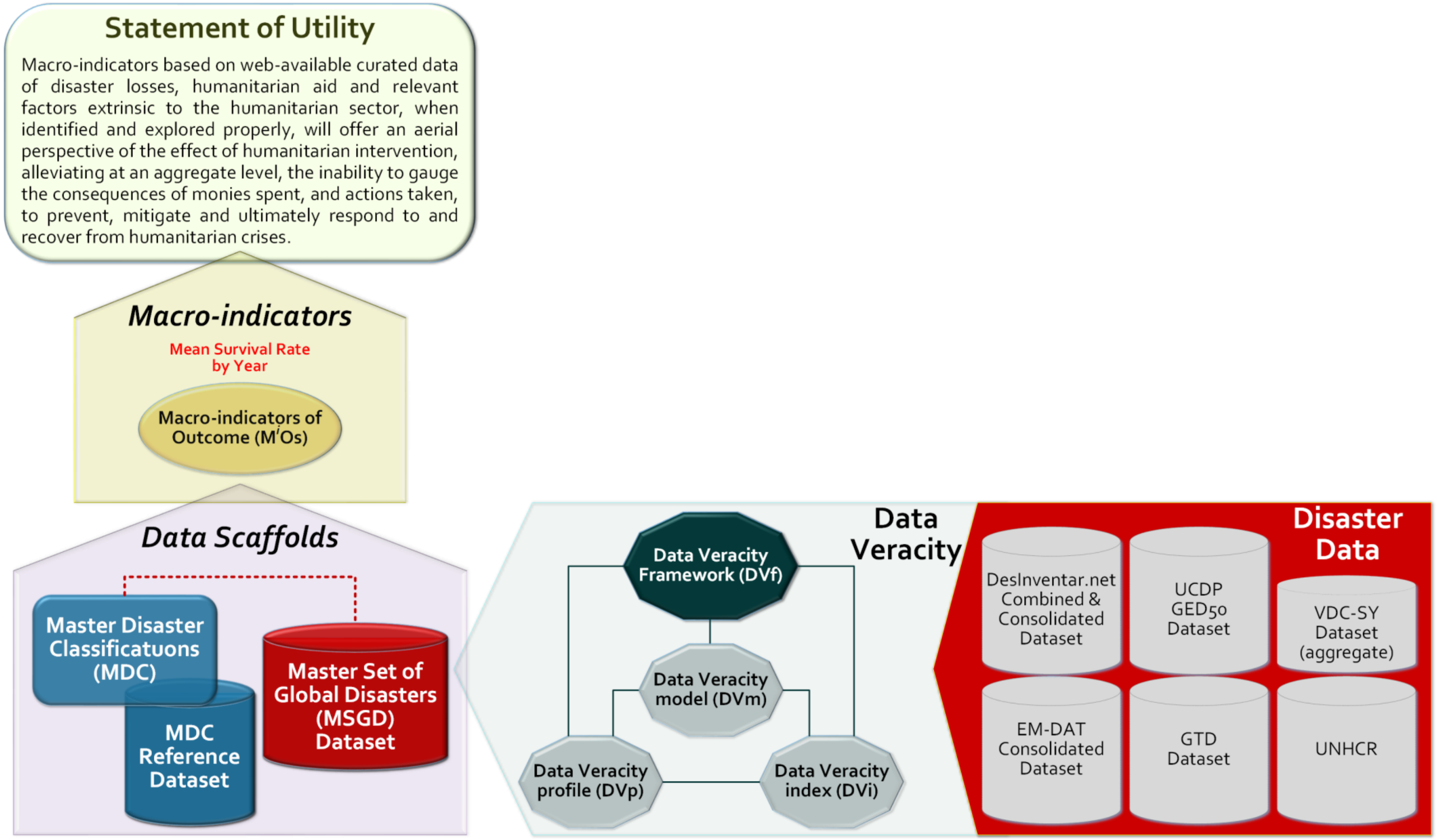


Figure 6-44: Constructs Contributing to the Utility Theory (Iteration 3)

6.6 Summary

This chapter discusses Iteration 3 and constitutes the most substantial and labour-intensive iteration of the DSR design cycle. It starts by setting out the problem that the absence of key '*data scaffolds*', such as a master disaster dataset supported by a disaster classification system, needs to be addressed before any search for indicators for the outcome of humanitarian intervention can begin.

The chapter goes on to design and build these '*data scaffolds*' using data from six sources and partial disaster classification systems within, or proposed for, the humanitarian sector (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b; VDC-SY, 2016b; GTD, 2017e; UNHCR, 2017b; IRDR, 2014; Lerner, 2016; Coburn et al., 2014). In creating the master disaster dataset, the toolsets of the DVf developed in Iteration 2 are utilised for each source dataset; thus not only evaluating the veracity of each dataset, but also testing the DVf toolset. ***Mean Survival Rate by Year*** is identified as a **M^{i}O** and a relationship between weak data veracity and weak, if any, statistical significance is also revealed. The M^{i}O displays distinctively different plots depending on data veracity, disaster group and geographic region. Soft data typically yields better M^{i}O levels but weaker statistical significance.

Finally, the *Evaluate* step discusses the alignment of the design to the artefacts built in this iteration, before mapping the created artefacts to the DSR research framework. The weaknesses and gaps in the disaster dataset are discussed together with the behaviour of the *mean survival rate* depending on disaster group, geographic region and data veracity. To close, it is argued that the propositions of the study's utilities theory is supported in this iteration.

Chapter 7: AID & POPULATION (ITERATION 4)

7.1 Overview

This chapter describes the fourth and final iteration of the DSR design cycle for this research. The chapter moves the study beyond the scope of Iteration 3 [*Chapter 6*] to search for Macro-Indicators of Impact (M^iIs) and Macro-Indicators of Effect (M^iEs). It places mean *survival rate* (the M^iO from the previous chapter) in context with humanitarian aid, development aid and population figures in order to identify M^iIs and M^iEs. The intention here is to identify metrics that can indicate a high level relationship between the human outcome of disasters and humanitarian aid (M^iIs) and/or a relationship between disaster outcomes and factors extrinsic to the humanitarian domain such as other international aid or population figures (M^iEs). *Figure 7-1* is a basic schematic of the flow of Iteration 4.

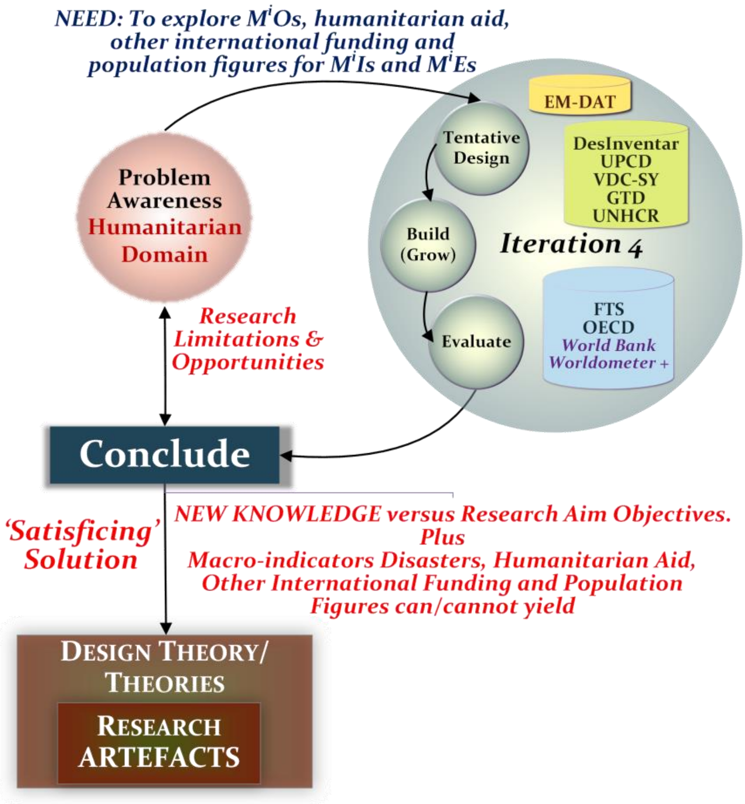


Figure 7-1: Iteration 4 of the Design Cycle

The structure of the chapter is as follows: *Section 7.2* describes the shift in focus to search for M^{i}Is and M^{i}Es and the need to acquire more data to enable this exploration. *Section 7.3* discusses the *Tentative Design* of sourcing the additional data needed, the DVi weightings to be applied to any new datasets evaluated and the design principle applied to searching for M^{i}Is and M^{i}Es. *Section 7.4* describes the *Build (Grow)* step to acquire, prepare and evaluate the additional data needed, as well as the identification of anomalies in the flow of aid versus the occurrence of disasters. The chapter goes on to discuss the search for and identification of M^{i}Is and M^{i}Es. *Section 7.5* is the *Conclusion* of this iteration, and as this is the last iteration, it also includes a reflection on the study as a whole. Finally, *Section 7.6* closes with a summary of the chapter.

7.2 Problem Awareness

The problem space for this iteration moves the study beyond issues of establishing fundamental data scaffolds (i.e. the MSGD and MDC) and identifying meaningful changes in disaster outcomes (i.e. the M^{i}O) to search for Macro-Indicators of Impact (M^{i}Is) and Macro-Indicators of Effect (M^{i}Es). The previous iteration identified *mean survival rate* as a Macro-Indicator of Outcome (M^{i}O) [*Figure 7-2*]; a metric that takes an isolated view of changes in the human effect outcome of disasters over time. The challenge here is to place *mean survival rate* (M^{i}O) in context by exploring its relationship with humanitarian aid to search for M^{i}Is, before going on to search for M^{i}Es. M^{i}Es being an indicator that may suggest M^{i}Os and/or M^{i}Is have a relationship with factors extrinsic to the humanitarian domain, such as other international aid and population size. This of course requires more data, including humanitarian aid, other flows of international financial support and population figures.

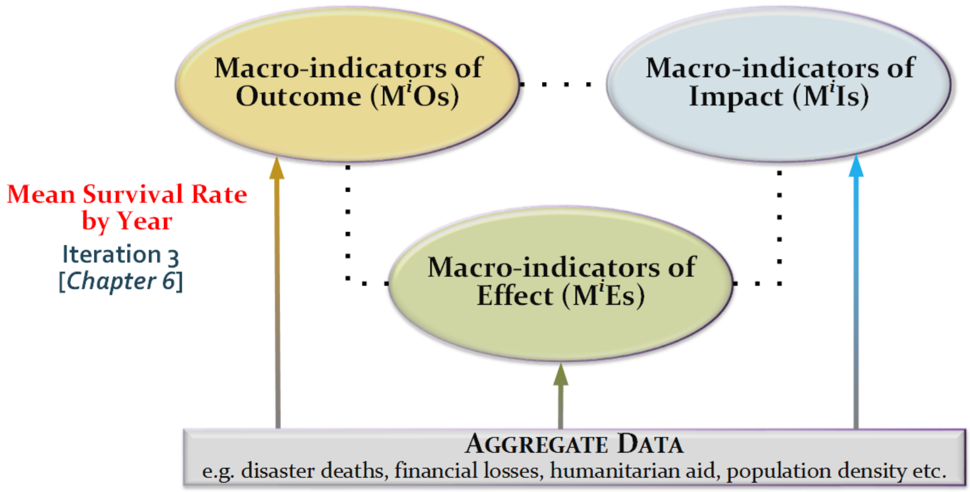


Figure 7-2: Macro-indicators Triad (M^{i}O → Mean Survival Rate)

7.3 Tentative Design

The *Tentative Design* for this iteration includes selection of:

- a humanitarian aid dataset;
- a development aid dataset;
- a data source for global and country population size by year.

The design principle applied to the search for M^{i}Is and M^{i}Es.

7.3.1 Humanitarian Aid

Data for humanitarian aid to disaster-affected countries is obtained from the ***Financial Tracking Services*** (FTS, 2017a). FTS tracks the flow of international humanitarian aid and is managed by the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA, 2017; FTS, 2017a). Furthermore, this UNOCHA managed humanitarian aid tracking database is referenced on the EM-DAT site as the source of its "*disaster-specific international aid contributions*" (Guha-Sapir et al., 2017a; FTS, 2017a, 2017i, 2017b; UNOCHA, 2017). Unfortunately, this information is not in fact available via EM-DAT; therefore, during this iteration, humanitarian aid contribution data is obtained directly from the FTS site (FTS, 2017d).

FTS is updated on a daily basis after cross-checking and reconciling data submissions from numerous actors and sources including the

European Disaster Response Information System (EDRIS); governmental donors; UN agencies; Central Emergency Response Fund (CERF); Country-based pooled funds (CBPF); UNDP's Multi-Partner Trust Fund Office (MPTF); United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA); the International Red Cross and Red Crescent Movement (ICRC), NGOs and private donors (FTS, 2017e; EDRIS, 2017; UN, 2017; CERF, 2017; CBPF, 2017; MPTF, 2017; UNOCHA, 2017; ICRC, 2017). FTS monitors and publishes the flows of humanitarian funding – reporting the totals as well as those assigned to specific appeals and plans – and makes this data available as a searchable database (FTS, 2017e). Of note is that it does not include concessional financing; soft loans; development aid; overseas remittances; a government's self-funding of a crisis or a country's support of refugees within its borders (ibid).

7.3.2 Development Aid

The ***International Development Statistics*** online databases (IDS, 2017) managed by the Organisation for Economic Co-operation and Development (OECD, 2017b) is selected as the source of development aid data. As discussed in *Chapter 2* there is ambiguity as to the *actual* usage of humanitarian aid versus development aid during the disaster management life-cycle (Safran, 2004; Cozzolino, 2012): Humanitarian aid may not be exclusively used for post-disaster *response* as intended and can be stretched to support activities in other disaster management phases or even development initiatives (Scholten et al., 2010; Riddell, 2008, 2014a). Similarly, development aid which has no obvious home in the disaster management life-cycle, can also be used to 'prop up' *mitigation* and *preparedness* activities (Riddell, 2014a). Acknowledging that it is not clear exactly where and how each type of aid is applied, in looking for M*i*Es this study expands its perspective of aid to include other types of international financial assistance given to disaster-affected countries. As such it sources data from the following four data tables available from the OECD's International Development Statistics (IDS) online databases [*Table 7-1*] (OECD, 2017b; IDS, 2017):

Table 2a [DAC2a]	**Aid (ODA) disbursements to countries and regions** This depicts annualised totals for the destination of Official Development Assistance disbursements. It includes geographic breakdowns by donor, recipient and some types of aid (e.g. grant, loan, technical co-operation) on a disbursement basis (i.e. actual expenditures).
Table 2b [DAC2b]	**Other official flows (OOF) – disbursements** *Other official flows* are transactions that do not meet the ODA criteria, e.g. grants; qualifying official bilateral transaction; net securities' acquisitions multilateral development banks; private sector subsidies, private funds for investment.
Table 3a [DAC3a]	**Aid (ODA) commitments to countries and regions** A *commitment* is a firm written substantiated obligation by a government or official agency, to provide resources of a specified amount under specified financial terms and conditions and for specified purposes for the benefit of a recipient country or a multilateral agency.
Table 4 [DAC4]	**Private flows** Private transactions are those undertaken by firms and individuals resident in the reporting country.

Table 7-1: International Development Statistics (IDS) – Aid tables
(OECD.Stat, 2017)

7.3.3 Population

Population size is of interest at two levels for each year from 1990 to 2015: (1) for each disaster-affected country; (2) the world. The ***World Bank's Databank*** is identified as the likely source for annualised population size, at both the country level and aggregated globally (World Bank, 2017).

7.3.4 Aid Data DVi Weightings

The L3 dimension weightings for all aid datasets to be evaluated during this iteration of the DSR design cycle are shown in the diagram below [*Figure 7-3*]. All L3 dimensions are weighted ***3*** (*needed*). It is assumed that the transactional movement of significant sums of money is tightly controlled and meticulously maintained. Therefore the expectation of veracity is accordingly high.

So much so that even L3 dimensions that are more leniently weighted for MSGD data sources, i.e. '*4. No omitted metadata* and '*5. No irrelevant entries*', are weighted as *needed* for aid data.

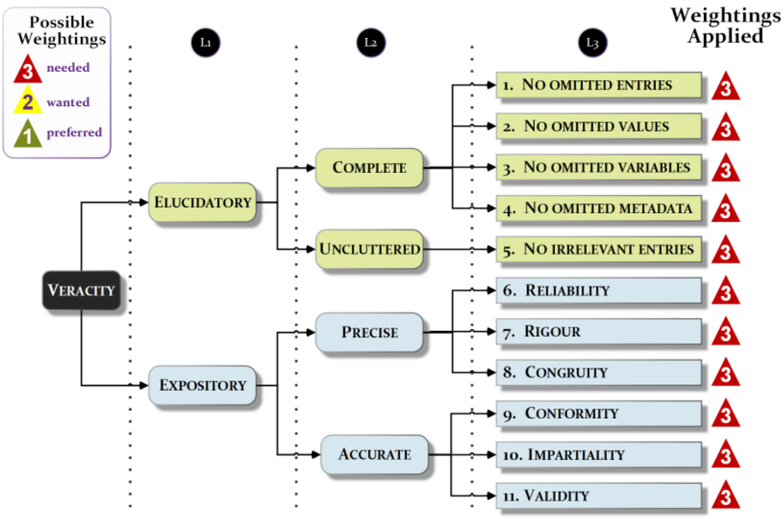


Figure 7-3: Aid Datasets Veracity Profile DVi – Applied Weightings

7.3.5 Macro-Indicators

The search here is for Macro-Indicators of Impact (M^{i}Is) and Macro-Indicators of Effect (M^{i}Es). Once again in the search for macro-indicators no prescriptive definition of what may or may not constitute M^{i}Is or M^{i}Es is specified here. The intent is to allow these to emerge through the data exploration process, with the following guidelines:

- M^{i}Is will place *mean survival rate* (the M^{i}O from the previous chapter) in context with humanitarian aid; *and*
- M^{i}Es will place the M^{i}O, and/or any identified M^{i}Is, in context with other aid and population figures.

7.4 Build (Grow)

This section describes the acquisition, preparation, examination and data veracity evaluation of the aid datasets of disaster-affected countries, but unlike disaster datasets, which are amalgamated to build the MSGD, aid datasets are not combined. As with the last iteration the DVf toolset is used to evaluate the veracity of each of the downloaded aid datasets. Of note is that data veracity evaluation is applied only to aid datasets and not to population values. The acquisition and preparation of population size, at both global and country level, is sourced as single variables, as such veracity evaluation is considered to have limited applicability or benefit. Finally, analytical models and visualisations are created using the MSGD and data from this iteration to search for Macro-indictors of Impact (M^iIs) and Macro-indictors of Effectiveness (M^iEs).

7.4.1 Financial Tracking Services (FTS)

As per the *tentative design* the UNOCHA's Financial Tracking Services (FTS) is the identified data source used for humanitarian aid to disaster-affected countries (UNOCHA, 2017; FTS, 2017a). FTS offers humanitarian aid data via two websites, one is an older, now archived, site that is still accessible (FTS, 2017c), the other a newer version launched at the start of 2017 (FTS, 2017a). As the new FTS site promises to offer clearer financial flows this is chosen as the source of FTS data (FTS, 2017a; FTS, 2017c). The datasets from both *sites* are expected to hold data for 1992–2015 as FTS commenced data collection in 1992 [*Figure 7-4*]. In reality, neither FTS website provides data that predates 1999; no explanation is offered.

"FTS aims to present a complete picture of all international humanitarian funding flows. Since 1992, it has collected reports on humanitarian funding flows submitted by Government donors, UN-administered funds, UN agencies, NGOs and other humanitarian actors and partners, including the private sector."

Figure 7-4: FTS – 1992 Launch Year
(FTS, 2017b)

(a) Acquiring the Data

Detail FTS data can only be acquired based on specific requests – e.g. by organisation, response plan, country etc. – for a stated date range. Ultimately, two types of data by country and year, are obtained: *detailed humanitarian aid flows*; and *summarised financial assistance to countries*.

FTS detail humanitarian aid flows

This represents each movement or commitment of funds into, out of, and within a country (FTS, 2017d). FTS humanitarian aid flows are categorised as *Incoming flows*, *Outgoing flows* and *Internal flows (FTS, 2017f)*. When looking at humanitarian aid flows with respect to a country, the *Incoming flow* is the aid the country received, the *Outgoing flow* is aid the country gave to others and the *Internal flows* represent the movement of aid within the country. The FTS site stipulates that *Internal flows* and *Incoming flows* should not be added together as there is a risk of double-counting. *Outgoing* flows appear to be mirror images of the *Incoming flows* , i.e. the same numbers from a different perspective *(FTS, 2017f)*.

To obtain most of the detailed FTS dataset 197 countries listed as 'affected' on the FTS site are individually requested for all years available, and each dataset downloaded (FTS, 2017g). Notably, of the affected country names explicitly listed on the site, two locations – West Bank and the Gaza Strip – have no FTS data, while data is found for ten countries that are not mentioned on the site's 'affected country' list (ibid). In total, aid flow entries are obtained for 205 countries. These FTS detail-level Excel workbooks each contain three worksheets – *Incoming*, *Internal* and *Outgoing* flows (FTS, 2017f). In the end 202,041 detailed FTS entries are obtained for year range 1999–2015 inclusive.

FTS summary financial assistance to countries (by year)

The process to obtain detailed humanitarian aid flow data relies heavily on the site's list of 'affected' countries and the happenstance of identifying relevant countries not explicitly mentioned on this list (FTS, 2017g). It is therefore considered prudent to utilise the site's facility of providing totals of the annual funding to each country as a way of verifying the previously acquired detailed flow data (FTS, 2017h). Of note is that even these annualised per country summaries require multiple request/download cycles (ibid). This laborious process however does serve to confirm that although FTS (new) allows the selection of all years from 1980, data is not found for years prior to 1999. Therefore, only seventeen years, 1999–2015, of FTS summary data is obtained, coinciding and confirming the date range available for FTS detailed data. Concatenated, these seventeen 'year' datasets provide 1959 entries.

FTS Detail vs FTS Summary

As a cross-check that no relevant data is missed the US$ totals of the FTS detail and summary downloads are compared (FTS, 2017d; FTS, 2017h). The totals do not match, the FTS summary total is US$11.55bn greater than FTS detail total [*Table 7-2*]. *Note that this value is not adjusted to 2015 values using the US CPI* (BLS, 2016).

	Detail	Summary
Incoming	$145,783,929,732	
Internal	$23,105,557,699	
Total	$168,889,487,431	$180,439,758,293
Shortfall	**$11,550,270,862**	

Table 7-2: FTS Detail Flow vs FT Annual Summary (US$s)

To find the root cause of this discrepancy a meticulous comparison between the FTS detail and summary downloads is carried out. This exercise identifies 231 entries in the FTS summary data that do not reconcile to their equivalent year/country subtotals in the FTS detail data (FTS, 2017d; FTS, 2017h). While the total mismatch between the summary and detail may be US11.55bn (when entire datasets are netted), the 231 summary entries that are discrepant span 57 countries (plus country 'Not Specified') and the total net value of just these entries is in fact more than US$127.7bn (adjusted to 2015 figures) [*Table 7-3*].

Year	# of Entries	US$ Total	US$ Total_2015
1999	1	-$27,141,731	-$38,613,755
2000	2	$595,918,807	$820,225,829
2001	2	$689,612,265	$922,923,942
2002	2	$697,823,644	$919,377,802
2003	2	-$537,070,278	-$691,819,490
2004	3	$1,492,265,574	$1,872,378,558
2005	5	$6,818,844,205	$8,275,381,448
2006	9	$3,424,946,109	$4,026,639,146
2007	19	$3,468,617,159	$3,965,049,210
2008	16	$8,615,099,219	$9,483,959,681
2009	15	$8,442,516,510	$9,327,155,389
2010	19	$14,805,953,112	$16,093,400,727
2011	19	$11,206,567,173	$11,808,298,835
2012	25	$9,932,217,496	$10,253,335,864
2013	27	$11,650,428,611	$11,853,473,549
2014	29	$19,147,908,112	$19,170,636,222
2015	36	$19,638,349,937	$19,638,349,937
	231	$120,062,855,924	$127,700,152,894

Table 7-3: FTS – 231 Summary totals ≠ Detail year/country totals

More information about the identification and breakdown of these 231 entries can be found in Appendix H.5.

The significant discrepancy between FTS summary and FTS detail makes a compelling argument for setting aside FTS's detailed download as this is where the total net shortfall resides (FTS, 2017d; FTS, 2017h). This study therefore focuses on the annualised country summaries and flags the mismatched 231 entries in the summary dataset as soft numbers, i.e. of weaker veracity.

(b) Preparing the Data

Two main types of data preparation are needed to enable analysis of humanitarian assistance US$s from FTS in context with the MSGD. One is to make the aid flow values congruous so that comparison can be drawn; the other is to ensure that countries are identified by their ISO codes so that FTS entries can be matched to MSGD countries.

FTS Aid Flow (US $s)

The Year/Country summary for each year selected lists two values per country, Funding US$s and Pledges US$s (FTS, 2017h), pledges being an intention or allocation by a donor (FTS, 2017f). For the purposes of this study, unless explicitly stated otherwise, pledges are assumed to be fulfilled and funding is therefore considered the sum of Funding US$s and Pledges US$s (FTS, 2017h; FTS, 2017f). FTS US$s values are recorded based on the year the aid was given. To enable comparisons over time these values are adjusted to 2015 levels using USA CPI [*Figure 7-5*] (BLS, 2016).

YEAR	US$s (Original)	US$s (CPI 2015)
1999	$269,458,499	$383,350,811
2000	$1,819,636,682	$2,504,557,650
2001	$3,577,193,285	$4,787,439,982
2002	$4,768,242,366	$6,282,126,185
2003	$6,254,453,385	$8,056,585,748
2004	$4,499,825,322	$5,646,030,163
2005	$13,841,995,393	$16,798,710,815
2006	$8,282,911,932	$9,738,050,283
2007	$7,724,672,416	$8,830,235,460
2008	$13,474,723,789	$14,833,693,018
2009	$12,132,801,005	$13,404,121,885
2010	$17,010,203,516	$18,489,321,123
2011	$14,423,396,979	$15,197,854,893
2012	$12,601,548,742	$13,008,969,215
2013	$13,993,098,777	$14,236,972,026
2014	$24,743,339,573	$24,772,709,330
2015	$21,022,256,632	$21,022,256,632

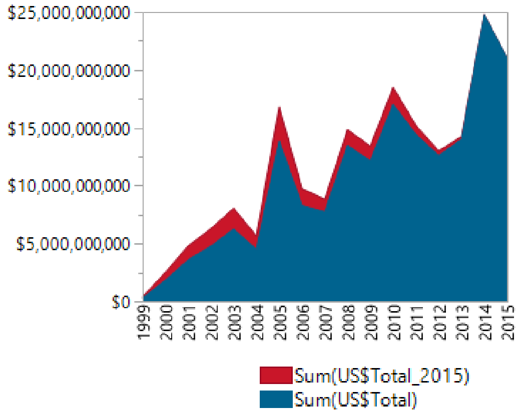


Figure 7-5: FTS US$ Annual Totals (including CPI Adjustment)

There is also the anomaly of negative numbers found in the annualised summary dataset, but explanation of these negative values is not found in their equivalent detail flows [*Table 7-4*].

		From Summary		From Detail Flow		
Year	Country	Funding	Pledges	Incoming	Internal	Outgoing
1999	Not specified	-$27,141,731	$0	.	.	.
2003	Not specified	-$902,373,170	$7,441,660	.	.	$473,362,933
2009	Not specified	$1,582,413,663	-$3,391,420	.	.	$2,094,767,614
2012	Kuwait	-$2,000,000	$0	.	.	.
2013	Kuwait	-$1,250,000	$0	.	.	.
2015	Brazil	-$2,350,620	$0	.	.	.
2015	Canada	-$1,966,955	$0	.	.	.
2015	United States	-$6,653,389	$0	.	.	.

Table 7-4: FTS Negative Summaries not explained by the Detail

No explanation of negative values is found on either the new or archived FTS sites (FTS, 2017a; FTS, 2017c). The only explanation identified is from a 2011 newspaper blog that states that FTS negative entries "*are balancing entries to avoid double-counting*" (The

Guardian, 2011). This is undoubtedly about the data on the old FTS site as the blog entry predates the new FTS site and it is difficult to grasp the cogency of this explanation for data in *Table 7-4*.

Countries

There are 213 unique entries for country name in the summary FTS datasets. Some of these are *blank*, '*Not specified*', '*Multiple Locations (shared)*'and '*Various/Unknown*'. The explicit non-countries are consolidated to a single '*Not specified*' and all *blank* entries are filtered out. This leaves 188 actual countries plus one '*Not specified*'. The 188 country names are checked against the ISO 3166 standard and their ISO alpha3 codes are added (ISo-3166, 2017). 33 country names do not comply with ISO 3166 naming convention, each of these are checked and their ISO alpha3 codes individually added.

(c) Examining the Data

Examining first the overall humanitarian aid figures held in FTS [*Table 7-5*], and the proportion of these that are considered soft because FTS summary totals do not match FTS detail totals [*Table 7-6*] (FTS, 2017d; FTS, 2017h). Interestingly, these 231 soft FTS entries amount to less than 11.8% of the entries and 64.5% of the US$ Total (2015 adjusted).

Recipient Region	# of Entries for Recipient	US$Total_2015 to Recipient	%age of US$Total_2015 Total
Africa	722	$83,816,874,811	42.33%
Asia	571	$76,752,422,215	38.77%
Americas	339	$8,769,546,699	4.43%
Europe	219	$1,914,533,176	0.97%
Oceania	91	$201,950,762	0.10%
Various/Unknown	17	$26,537,657,556	13.40%
Total Aid (FTS)	1,959	$197,992,985,219	100%

Table 7-5: FTS Humanitarian Aid by Region

Recipient Region	# of Firm Entries	US$Total_2015 [Firm]	# of Soft Entries	%age of Entries [Soft]	US$Total_2015 [Soft]	%age US$Total_2015 [Soft]
Africa	614	$34,816,323,731	108	15%	$49,000,551,080	58.46%
Asia	492	$31,440,491,435	79	13.84%	$45,311,930,780	59.04%
Americas	316	$2,280,342,143	23	6.78%	$6,489,204,556	74.00%
Europe	216	$1,595,980,222	3	1.37%	$318,552,954	16.64%
Oceania	90	$159,694,794	1	1.10%	$42,255,968	20.92%
Various/Unknown			17	100.00%	$26,537,657,556	100.00%
Total Aid (FTS)	1,728	$70,292,832,325	231	11.79%	$127,700,152,894	64.50%

Table 7-6: FTS Humanitarian Aid by Region (Firm vs Soft Data)

Figure 7-6 shows the number of annual recipients of humanitarian aid each year together with the mean and sum of CPI (2015) adjusted annual humanitarian aid. The mean and sum charts are of nested bars, with the most transparent outer bars being the soft values. Note that:

- Although FTS start year is 1992, no data is available until 1999.
- 1999 and 2000 may not be complete as most years show funding to at least 100 named countries.
- Chart 3 suggests that US$ value of the 231 mismatched entries are spread over the most recent eight years.

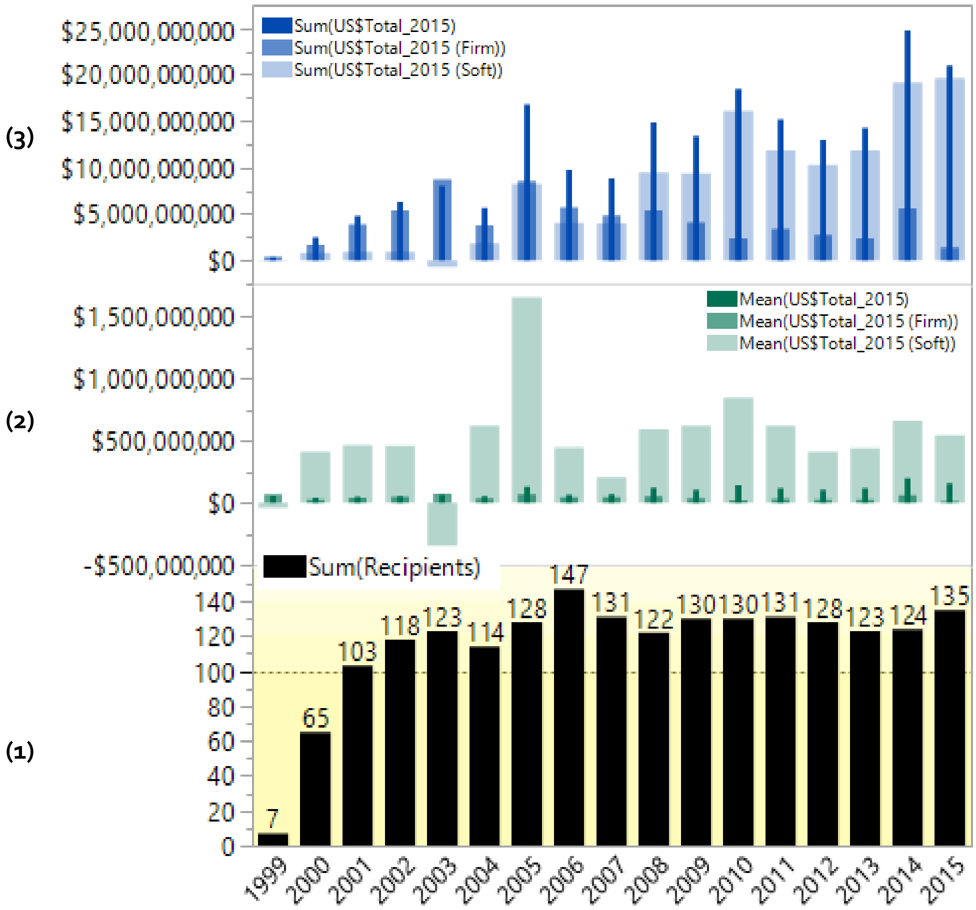


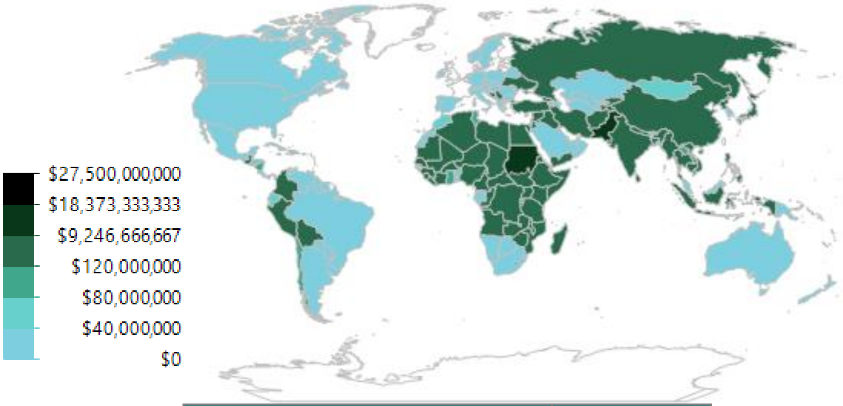
Chart	
(1)	Humanitarian funding recipients per year– *lowermost chart*.
(2)	Mean humanitarian fund per year – *middle chart.*
(3)	Total humanitarian fund per year – *topmost chart.*

Figure 7-6: FTS Recipients and US$s per YEAR

Another curious finding is that over USD26.5bn net has been given to unspecified recipients [*Table 7-7*]. This is almost US10bn more than the amount given to Sudan, the largest <u>named</u> recipient of ***all*** according to FTS [*Figure 7-7*], and over US$20bn more than Iraq the highest recipient according to FTS ***firm*** (reconcilable) data [*Figure 7-8*]. Admittedly, almost US$13.25bn of the money to Sudan has been flagged as soft because it cannot be reconciled to FTS details totals [*Figure 7-9*].

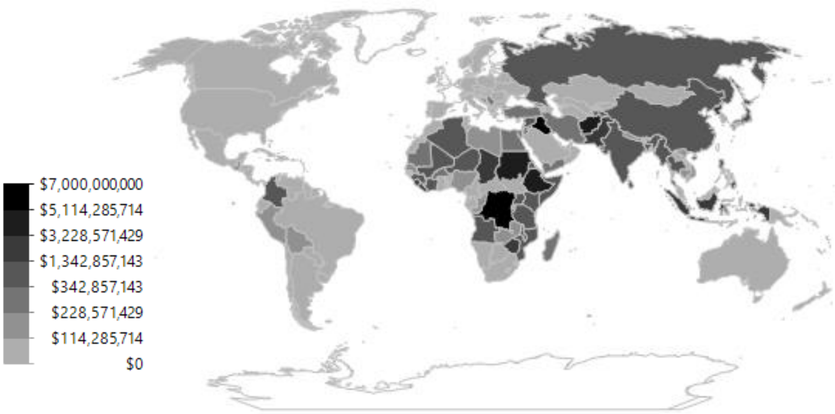
	Recipient = Not Specified		
	US Funding (2015)	US Pledges (2015)	US Total (2015)
1999	-$38,613,755	$0	-$38,613,755
2000	$641,140,369	$0	$641,140,369
2001	$609,368,318	$0	$609,368,318
2002	$566,697,945	$0	$566,697,945
2003	-$1,162,379,248	$9,585,869	-$1,152,793,379
2004	$351,355,445	$32,024,653	$383,380,098
2005	$5,035,436,488	$689,391,317	$5,724,827,805
2006	$1,058,466,571	$70,129,208	$1,128,595,779
2007	$707,066,701	$3,761,689	$710,828,390
2008	$1,249,036,034	$1,704,983,092	$2,954,019,126
2009	$1,748,224,964	-$3,746,786	$1,744,478,178
2010	$1,063,672,505	$2,472,324	$1,066,144,829
2011	$2,125,762,386	$174,532,992	$2,300,295,378
2012	$1,922,600,281	$888,181	$1,923,488,462
2013	$1,316,484,421	$112,537,213	$1,429,021,634
2014	$3,574,581,527	$603,379,519	$4,177,961,046
2015	$2,312,291,393	$56,525,940	$2,368,817,333
Total	**$23,081,192,345**	**$3,456,465,211**	**$26,537,657,556**

Table 7-7: FTS Funding per Year to Unspecified Recipients (US$ 2015)



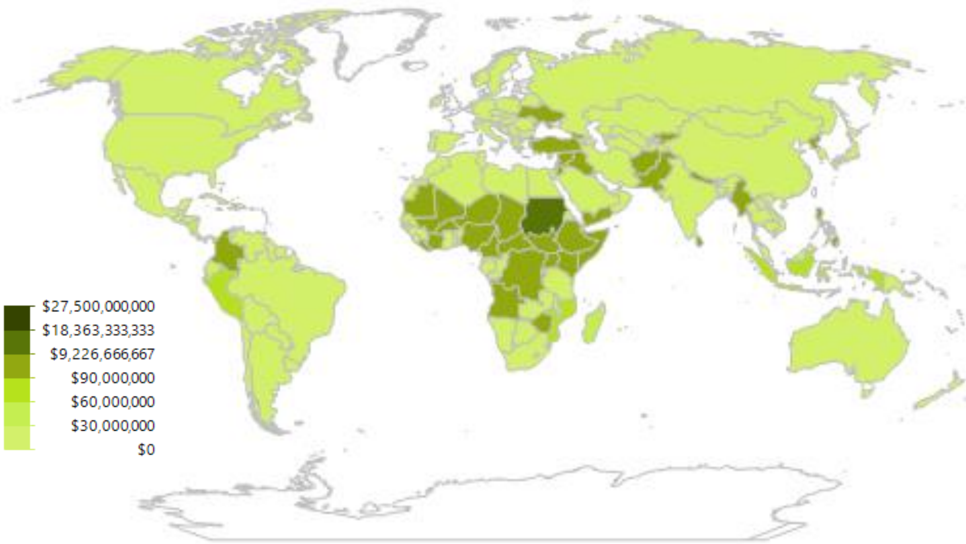
Recipient	Total
Not specified	$26,537,657,556
Sudan	$16,750,395,818
Pakistan	$10,022,110,340
Iraq	$9,118,099,960
Afghanistan	$8,760,161,061
Somalia	$8,423,809,824
Ethiopia	$8,411,170,564
occupied Palestinian territory	$8,180,092,597
Congo, The Democratic Republic of the	$7,752,254,740
Syrian Arab Republic	$7,555,349,306
Haiti	$6,500,664,051

Figure 7-7: FTS 1999-2015 Top 10 Recipients



Recipient	Total (Firm)
Iraq	$6,414,302,015
Congo, The Democratic Republic of the	$6,319,773,572
Afghanistan	$4,262,693,521
Ethiopia	$4,069,774,445
Sudan	$3,503,302,627
Pakistan	$3,222,663,080
occupied Palestinian territory	$2,610,300,285
Korea, Democratic People's Republic of	$2,533,906,116
Liberia	$2,368,264,680
Indonesia	$1,860,782,109

Figure 7-8: FTS Firm data (reconciles with FTS detail records)



Recipient	Total (Soft)
Not specified	$26,537,657,556
Sudan	$13,247,093,191
Somalia	$7,759,247,482
Syrian Arab Republic	$6,881,012,428
Pakistan	$6,799,447,260
South Sudan	$6,194,617,259
Haiti	$5,961,203,605
occupied Palestinian territory	$5,569,792,312
Lebanon	$5,023,259,257
Afghanistan	$4,497,467,540
Ethiopia	$4,341,396,119

Figure 7-9: FTS Soft data (does not reconcile with FTS detail records)

(d) FTS Veracity

A detailed DVp for FTS can be found in *Appendix A.8*, while FTS's DVi scores and indices are included here [*Figure 7-10*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions			Weighting	FTS		
L1	L2	L3		Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	No entries 1992-1998. Few entries 1999 & 2000. Summary US$11.5bn>Detailed.	1	0.27
		2. No omitted values	3	Detailed FTS= 3,753 'empty' entries. Summary 17 entries unspecified recipients ~US$26.5bn net.	1	0.27
		3. No omitted variables	3	There is no known variable that links each funding flow to specific disasters.	3	0.82
		4. No omitted metadata	3	Missing explanations include: -ve or zero contribution; unspecified recipient; summary total logic and data source etc.	2	0.55
		Complete				0.48
	Uncluttered	5. No irrelevant entries	3	Zero value clutter (3,753 entries->Detailed & 794 entries->Summary). Inconsistent/Unclear Countries.	2	0.55
		Uncluttered				0.55
	Elucidatory Index					0.49
Expository	Precise	6. Reliability	3	Detail-> *'Creation'* & *'Last Updated'* dates. No visibility-> changes or deletions. Summary->no admin dates.	3	0.82
		7. Rigour	3	Rigour at FTS data entry is plausible, but rigour at FTS data sources is not assumed.	3	0.82
		8. Congruity	3	Detailed-> extremes of US$/entry (US$1-US$1bn).All US$ values are relevant to the year given.	3	0.82
		Precise				0.82
	Accurate	9. Conformity	3	231 entries of Summary cannot be reconciled Detailed==> flagged as soft.	3	0.82
		10.Impartiality	3	Only 17 of the 26 years of interest represented with tacit bias towards 15 years (2001 - 2015).	2	0.55
		11.Validity	3	Cited by EM-DAT->valid, but NO "*disaster-specific" information* contrary to EM-DAT claim	3	0.82
		Accurate				0.73
	Expository Index					0.78
Data Veracity index (DVi)			33			0.65

Figure 7-10: FTS DVi

FTS DVi of 0.65 places it just below the mid-point of 0.68 on the Dataset DVi Scale [*Figure 7-11*]; near UNHCR, the other UN dataset.

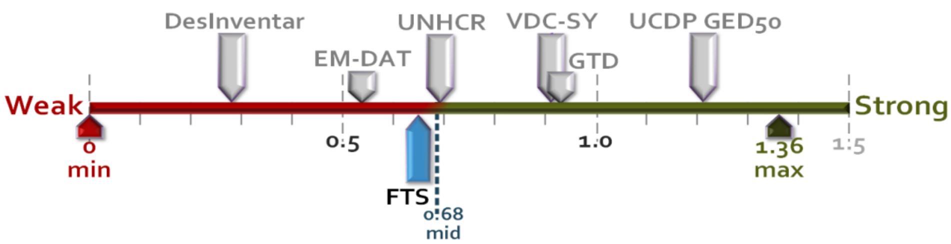


Figure 7-11: FTS on the Dataset DVi Scale

Unsurprisingly, the DVi L1 and L2 dimension scores are such that they also congregate below their respective mid-points [*Figure 7-12*]. The drag on FTS's veracity comes primarily from the clutter of unexplained zero and negative value entries and the inability to match over US11.55bn country/year summary totals from the site with totals calculated from detailed recorded sourced from the site.

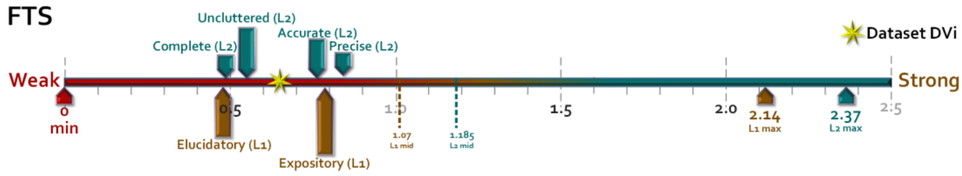


Figure 7-12: FTS on the L2 & L1 DVi Scale

FTS humanitarian funding comes to a total of almost US$198bn, and as stated earlier, 231 FTS summary entries are flagged as being of weaker veracity (soft) because their totals cannot be confirmed by the FTS detail year/country totals.

7.4.2 International Development Statistics (IDS)

To gain perspective of other types of international financial assistance given to disaster-affected countries the following four data tables from OECD's International Development Statistics (IDS) online databases are selected (OECD, 2017b; IDS, 2017):

(1) Official Development Assistance (ODA) Disbursements [Table2a] – *i.e. actual expenditures.*

(2) Other Official Flows (OOF) [Table2b]

(3) Official Development Assistance (ODA) Commitments [Table3a] – *i.e. not actual expenditure.*

(4) Private flows [Table4]

Four IDS tables are believed to hold data pertinent to this study and these tables, similar to the FTS summary data, are only available as annual summaries (IDS, 2017; FTS, 2017h). Of note is that these are annualised summaries and as such detecting a relationship between a specific humanitarian crisis and the flow of these funds is infeasible.

(a) Acquiring the Data

Acquisition of the four IDS data tables is relatively simple; each table is selected from the *Development* 'theme' menu tree on the OECD.Stat web page, viewed and downloaded (OECD.Stat, 2017).

(1) Official Development Assistance (ODA) Disbursements [Table2a]

With Table 2a a total of **7,280,692 records** are downloaded for years **1960 to 2015**. Values are in both current price in millions of US$s (based on the year of occurrence), and constant price in millions of US$s (made consistent to 2015 values of the US$).

(2) Other Official Flows (OOF) [Table2b]

With Table 2b a total of **1,430,532 records** are downloaded for years **1960 to 2015**. Values are in both current price in millions of US$s (based on the year of occurrence), and constant price in millions of US$s (made consistent to 2015 values of the US$).

(3) Official Development Assistance (ODA) Commitments [Table3a]

With Table 3a a total of **1,683,902 records** are downloaded for years **1966 to 2015**. Values are in both current price in millions of US$s (based on the year of occurrence), and constant price in millions of US$s (made consistent to 2015 values of the US$)

(4) Private flows [Table4]

With Table 4 a total of **494,753 records** are downloaded for years **1968 to 2015** with values in current price in millions of US$s (based on the year of occurrence) only.

(b) Preparing the Data

The four IDS data tables contain both detailed (donor/recipient) entries and subtotal entries (for regional, local funds, development banks etc.). Additionally, in three of the tables, Table 2a, Table 2b and Table 3a, each entry is effectively duplicated to show both current value in millions of US$s and constant 2015 value in millions of US$s. The records in each table are also distinguished from each other by the type of aid that they detail, e.g. grants, debts etc., not all of which are needed for this study. For all four data tables the following steps are taken:

- Only entries in the date range 1990–2015 are retained
- Subtotal and summary entries are removed.
- All recipients that are not a country are removed to eliminate the risk of double counting pass-through aid, i.e. aid that is given to a country via organisations.
- ISO alpha3 codes are added for each recipient country.

Additionally for Table 2a, Table 2b and Table 3a all duplicate entries of values in current US$s are removed, leaving only those entries with 2015 US$ values. For Table 4, where constant 2015 US$ value entries are not provided, US$ values are adjusted using 2015 USA CPI (BLS, 2016).

Aid Types

There are fifty-three aid types available in the four downloaded tables from OECD's International Development Statistics (IDS) (OECD.Stat, 2017; IDS, 2017). Of these only five aid types are considered pertinent to his research. These five aid type are shown below [*Table 7-8*] and can be seen in context with the 48 aid types not selected from IDS Tables 2a, 2b, 3a and 4 in *Appendix H.6*. In addition to the selected 5 variables a sixth aid value is calculated

#	Source	Aid	Type ID
(i)	Table 2a	ODA: Total Net	206
(ii)	Table 2a	Humanitarian Aid	216
(iii)	Table 2b	Total OOF, Net	206
(iv)	Table 3a	Total Commitments	305
(v)	Table 4	Total Private Net	420
(vi)	Calculated [Table 2a]	ODA: Total Net *minus* Humanitarian Aid	

Table 7-8: IDS, OECD Aid Types selected/calculated

(i) ODA: Total Net [206]

This aid type is from **Table2a**–Official Development Assistance (ODA) Disbursements and is the total of the other 23 aid types in this table (IDS, 2017).

(ii) Humanitarian Aid [216]

This aid type is also from **Table2a–Official Development Assistance (ODA) Disbursements** (IDS, 2017). Although it is included in **ODA: Total Net**, identifying its value separately allows comparison, and where appropriate selection, of IDS versus FTS humanitarian aid. It can also be deducted from **ODA: Total Net** to distinguish between the flow humanitarian aid and 'other' aid (ibid).

(iii) Total OOF, Net [206]

This is the total of 11 aid types held in **Table2b**–Other Official Flows (OOF) (ibid)(IDS, 2017).

(iv) Total Commitments [305]

This is the total of 5 aid types held in **Table3a–Official Development Assistance (ODA) Commitments** (ibid).

(v) Total Private Net [420]

This is the total of 10 private beneficial flow of funds held in **Table4–Private Flows** (ibid).

(vi) ODA: Total Net *minus* Humanitarian Aid [206 – 216]

A variable created by deducting Humanitarian Aid [216] from ODA: Total Net [206], both aid types are from **Table2a**–Official Development Assistance (ODA) Disbursements.

Finally, all values are multiplied by a million for ease of comparison with FTS figures (FTS, 2017a).

A table of annual totals from 1990 to 2015 from each of these aid types can be found in *Appendix H.7*.

(c) Examining the Data

A simple scan of the IDS (OECD) aid type values reveals that they all contain some negative entries, even humanitarian aid. While extensive searches uncover rudimentary descriptions that may be explanations of this phenomenon – i.e. loan repayments, asset selling etc. – but no firm definition is found (OECD, 2017d; OECD, 2017c). In **Table2a–Official Development Assistance (ODA) Disbursements** 120 negative flow entries equalling -US$26.56M (2015 values) and spanning 21 years, 1995–2015, for 61 countries are found for humanitarian aid [*Figure 7-13*] (IDS, 2017).

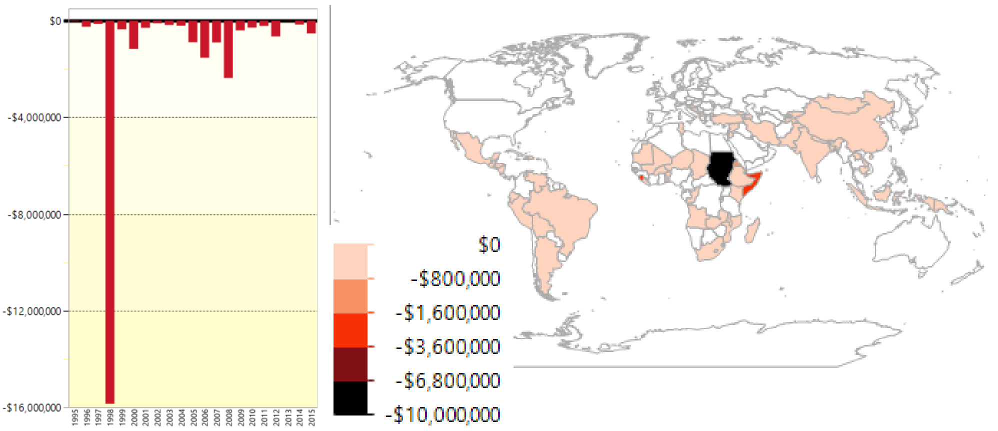


Figure 7-13: IDS (OECD) Table 2a Negative Flow of Humanitarian Aid

All but -US$60,000 of the -US$26.56M humanitarian aid flow nets to positive in the composite aid dataset; absorbed in the Year/Recipient Country subtotals of aid flows. The -US$60,000 relates to one entry in the raw dataset, which is from Bahrain to USA. Interestingly, there are no 'event' disasters per se in the MSGD for Bahrain during 1998, but in the year before 5 were killed in terrorist attacks, and between 1997 and 1998, 199 people left Bahrain with the help of UNHCR. With the data available it is not possible to tell if the money returned to USA is related to these occurrences. Due to the absence of relevant detail, in the case of negative flows, this study makes the assumption that for humanitarian aid, negative flows are a 'return' of humanitarian funds, and for all other types of aid, negative subtotals are deficits in aid to each recipient country.

IDS (OECD) unexpectedly also holds humanitarian aid flows. No explanation can be found as to the relationship between IDS's Humanitarian Aid [Table2a] and FTS's humanitarian aid figures (IDS, 2017; FTS, 2017h). These are not considered equivalents, especially as IDS has values for years that are gaps in FTS (ibid). *Figure 7-14* offers three charts. The first (*lowermost*) bar chart displays the annual number of recipient of all 5 IDS aid types selected for study, aid types (i) – (v) in *Table 7-8*. The second (*middle*) stacked bar chart shows the annual flow of monies other than humanitarian aid, aid types (iii) – (vi) in *Table 7-8*. Finally, the third (*topmost*) bar chart illustrates the annual flow of IDS's humanitarian aid, aid type (ii) in *Table 7-8*.

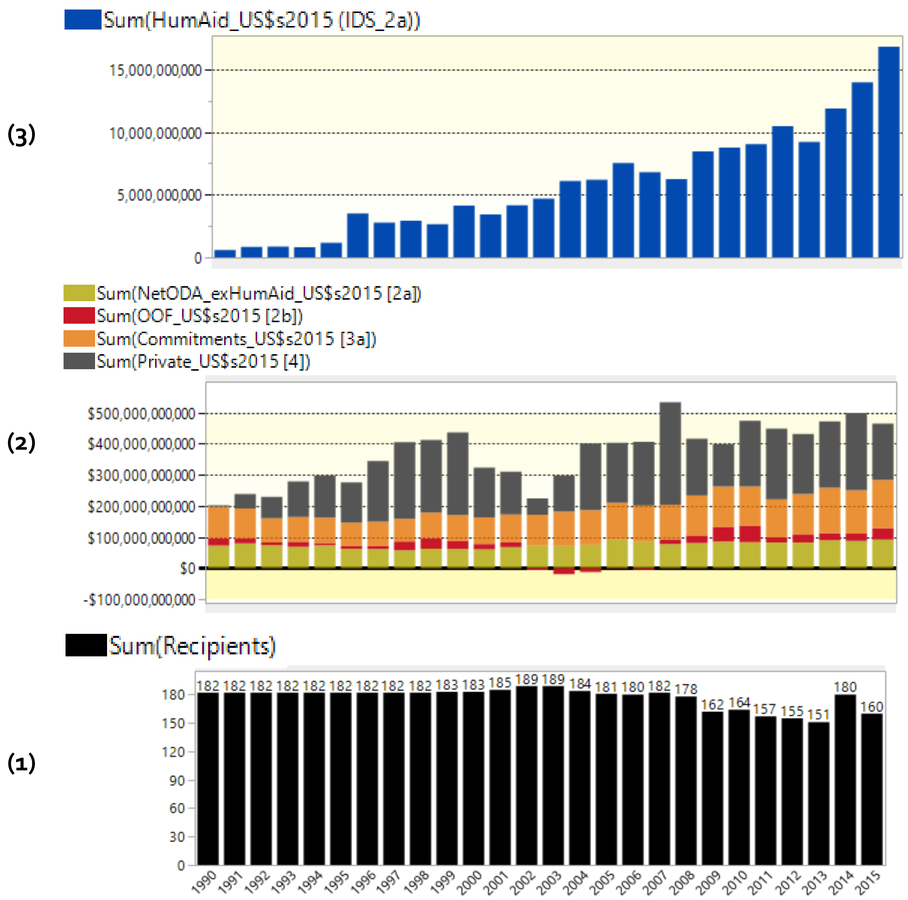
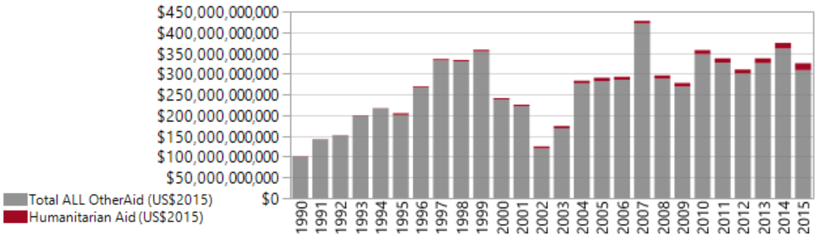


Chart	
(1)	IDS (OECD) total recipients per year
(2)	IDS (OECD) *other* aid per year (US$_2015)
(3)	IDS (OECD) total humanitarian aid per year (US$_2015)

Figure 7-14: IDS (OECD) Recipients and US$s per YEAR

As can be seen in Chart (1) of *Figure 7-14* the number of recipients remains between 180 and 189 until 2008 then dips as low as 151, before rising for one year to 180 in 2014, then falling back to 160 in 2015. The shape of the number of recipients' bar chart is not mimicked by the other two charts. Chart (2) of *Figure 7-14* suggests that ODA disbursements (excluding humanitarian aid) have remained level for 26 years (at 2015 US$$ rates); whereas ODA commitments have gradually increased, and other official flows (OOFs) and private flows have fluctuated inexplicably year by year. Most interesting is the near steady increase in humanitarian aid from 1990 to 2015. The *Figure 7-15*, provides a better view of the barely visible flow of humanitarian aid vs. all other aid per year.



Year	Humanitarian Aid			All Other Aid (US$000s_2015)	
	as %age of All Aid	US$	%age of its 26-year Total	US$	%age of its 26-year Total
1990	0.58%	$584,130,000	0.38%	$100,325,120,000	1.47%
1991	0.59%	$832,550,000	0.54%	$141,412,400,000	2.07%
1992	0.58%	$877,630,000	0.57%	$150,695,530,000	2.21%
1993	0.40%	$807,670,000	0.52%	$198,981,640,000	2.91%
1994	0.53%	$1,159,170,000	0.75%	$215,859,720,000	3.16%
1995	1.71%	$3,500,270,000	2.27%	$200,988,550,000	2.94%
1996	1.03%	$2,784,390,000	1.81%	$267,333,800,000	3.91%
1997	0.87%	$2,935,160,000	1.90%	$332,635,620,000	4.87%
1998	0.80%	$2,665,310,000	1.73%	$330,227,630,000	4.83%
1999	1.16%	$4,139,600,000	2.69%	$353,449,900,000	5.17%
2000	1.43%	$3,445,250,000	2.24%	$238,012,520,000	3.48%
2001	1.85%	$4,171,780,000	2.71%	$221,101,400,000	3.24%
2002	3.76%	$4,698,030,000	3.05%	$120,311,000,000	1.76%
2003	3.50%	$6,092,310,000	3.95%	$168,121,360,000	2.46%
2004	2.19%	$6,197,520,000	4.02%	$276,822,110,000	4.05%
2005	2.60%	$7,542,760,000	4.89%	$282,742,880,000	4.14%
2006	2.33%	$6,813,610,000	4.42%	$285,935,040,000	4.18%
2007	1.46%	$6,258,660,000	4.06%	$421,438,740,000	6.17%
2008	2.85%	$8,451,440,000	5.48%	$287,752,740,000	4.21%
2009	3.15%	$8,757,790,000	5.68%	$269,061,350,000	3.94%
2010	2.53%	$9,031,440,000	5.86%	$347,706,410,000	5.09%
2011	3.11%	$10,485,560,000	6.80%	$326,533,430,000	4.78%
2012	2.97%	$9,224,140,000	5.98%	$300,839,660,000	4.40%
2013	3.52%	$11,887,390,000	7.71%	$325,389,510,000	4.76%
2014	3.73%	$13,962,280,000	9.06%	$360,505,180,000	5.28%
2015	5.17%	$16,823,290,000	10.92%	$308,517,560,000	4.52%
Total	2.21%	$154,129,130,000	100.00%	$6,832,700,800,000	100.00%

Figure 7-15: IDS (OECD) Humanitarian vs. All Other Aid per YEAR

The annual totals of each of the five IDS (OECD) aid types are charted as nested bars to gain a sense of the relative scale of Humanitarian Aid [Table2a] [*Figure 7-16*](IDS, 2017). It is worthy of note that Humanitarian Aid [Table2a] is so comparatively smaller than other flows of money that it is barely visible in more recent years. At the other end of the spectrum, the Private (Table4) flow of funds dominates for at least 23 of the 26 years scrutinised.

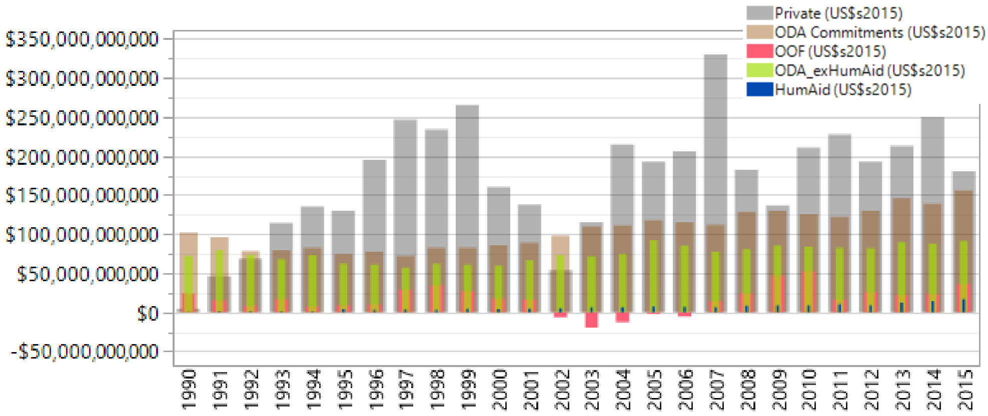


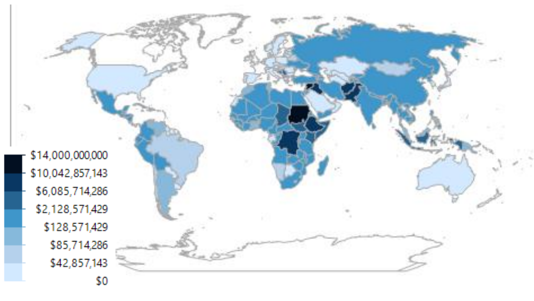
Figure 7-16: IDS (OECD) Nested Bars of Key Aid US$s per YEAR

Figure 7-17 explores each of the IDS (OECD) aid types in a little more detail, by looking at the top ten countries receiving each of the aid types' 26-year totals. Points of interest include:

- Countries in Africa and Asia are the greatest recipients of IDS's version of humanitarian aid.
- Only two of the top ten countries, Iraq and Pakistan, receiving humanitarian aid are also in the top ten recipients of OOD.
- There is no coincidence between the top ten countries in receipt of humanitarian assistance and the top ten recipients of private financial flows.
- At least two of the top ten countries in receipt of private financial flows, Bermuda and Cayman Islands, are considered tax havens.
- Four out of five BRICS economies, all but South Africa, are in the top ten recipients of OOF. BRICS (Brazil, Russia, India, China and South Africa) being economies that were considered to have tremendous growth potential (Esposito et al., WEFORUM, 2016).

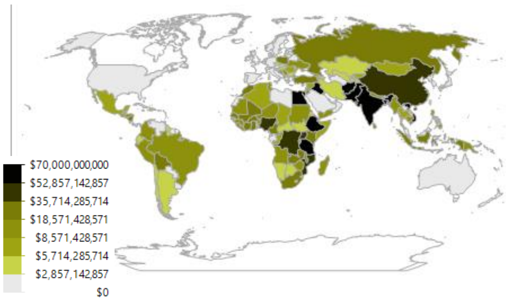
IDS (OECD) Humanitarian Aid US$s [Table 2a]

Country	Humanitarian Aid
Sudan	$13,232,690,000
Syrian Arab Republic	$12,350,550,000
Afghanistan	$8,409,470,000
Ethiopia	$8,286,610,000
Iraq	$8,129,840,000
Pakistan	$7,061,010,000
Democratic Republic of the Congo	$6,150,410,000
Somalia	$5,369,620,000
West Bank and Gaza Strip	$5,105,720,000
Kenya	$3,827,380,000



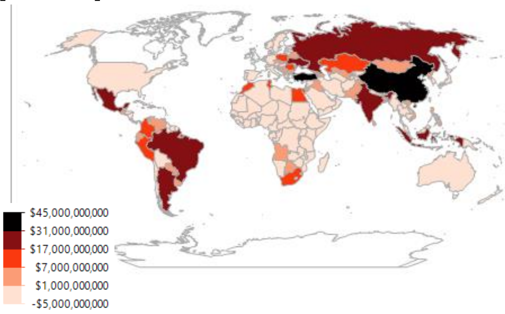
IDS (OECD) Net ODA exc. Humanitarian Aid US$s [Table2a]

Country	Net ODA (ex HumAid)
Iraq	$68,466,040,000
India	$67,974,940,000
Egypt	$65,395,630,000
Viet Nam	$60,550,630,000
Afghanistan	$58,291,300,000
Tanzania	$56,834,220,000
Ethiopia	$55,356,000,000
Pakistan	$54,829,370,000
Bangladesh	$53,186,630,000
Nigeria	$46,884,560,000



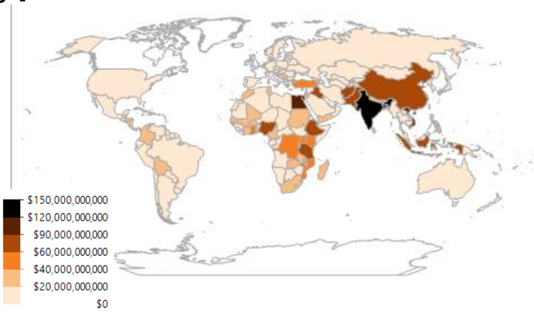
IDS (OECD) Other Official Flows (OOF) US$s [Table 2b]

Country	OOF
China (People's Republic of)	$44,708,810,000
Turkey	$34,151,450,000
India	$30,809,120,000
Brazil	$29,825,900,000
Mexico	$27,356,850,000
Argentina	$23,468,980,000
Russia	$20,079,370,000
Indonesia	$19,692,890,000
Ukraine	$17,026,600,000
Colombia	$16,146,660,000



IDS (OECD) ODA Commitments US$s [Table3a]

Country	Commitments
India	$145,658,200,000
Egypt	$102,083,120,000
Bangladesh	$90,987,990,000
Viet Nam	$90,177,920,000
Pakistan	$89,513,490,000
Iraq	$86,752,300,000
China (People's Republic of)	$77,321,600,000
Afghanistan	$75,390,730,000
Ethiopia	$74,949,110,000
Indonesia	$73,520,140,000



IDS (OECD) Private US$s [Table4]

Country	Private
Brazil	$533,753,000,000
China (People's Republic of)	$455,563,000,000
Mexico	$350,963,000,000
Turkey	$209,886,000,000
India	$188,526,000,000
Cayman Islands	$143,934,000,000
Bermuda	$143,399,000,000
Singapore	$137,492,000,000
Thailand	$133,116,000,000
Democratic People's Republic of Korea	$126,112,000,000

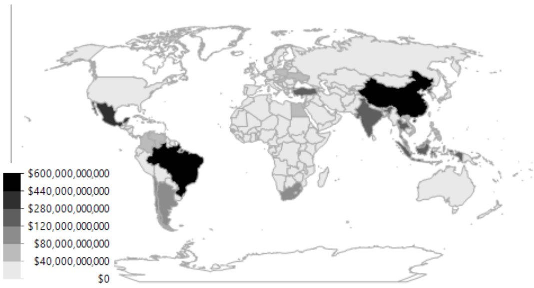


Figure 7-17: IDS Top 10 Countries for each Funding Flow

(d) IDS (OECD) Veracity

A detailed DVp for IDS (OECD) can be found in *Appendix A.9* and the DVi table for IDS (OECD) has been included here [*Figure 7-18*]. Indices for the dataset and its L1 and L2 veracity dimensions are shades of red, amber or green depending on their proximity to their dimensional midpoint.

Veracity Dimensions L1	L2	L3	Weighting	IDS (OECD) Reasoning for Score	Score	Index
Elucidatory	Complete	1. No omitted entries	3	NO disaggregate entries. Some flows missing, date range 1990-2015 is covered.	3	0.82
		2. No omitted values	3	Of the 10,889,879 entries from the 4 downloaded tables, 946,877 were either zero or missing.	3	0.82
		3. No omitted variables	3	All variables that are visible on the site appear to be provided in the downloads.	5	1.36
		4. No omitted metadata	3	High level explanations. No information aggregation logic or -ve, missing or zero values.	1	0.27
		Complete				0.82
	Uncluttered	5. No irrelevant entries	3	416,024 of 10,889,879 entries, only ~3.8% relevant identified as relevant. RISK: removing 'good' with 'bad'.	2	0.55
		Uncluttered				0.55
	Elucidatory Index					0.76
Expository	Precise	6. Reliability	3	No change management information provided.	2	0.55
		7. Rigour	3	Cannot find any information to confirm that the data is meticulously collected and maintained.	2	0.55
		8. Congruity	3	The data is at too high a level to identify incongruities that may/may not be an underlying issue.	2	0.55
		Precise				0.55
	Accurate	9. Conformity	3	Unable to tell as the data is only visible in the aggregate.	2	0.55
		10.Impartiality	3	Not enough information to gauge bias. Its scope is OECD members, therefore it fulfils its scope.	3	0.82
		11.Validity	3	The 416,024 usable entries retrieved from the 4 tables are valid for the study.	5	1.36
		Accurate				0.91
	Expository Index					0.73
Data Veracity index (DVi)			33			0.75

Figure 7-18: IDS (OECD) DVi

IDS DVi of 0.75 places it above the mid-point of 0.68 on the Dataset DVi Scale and positions it higher than FTS [*Figure 7-19*].

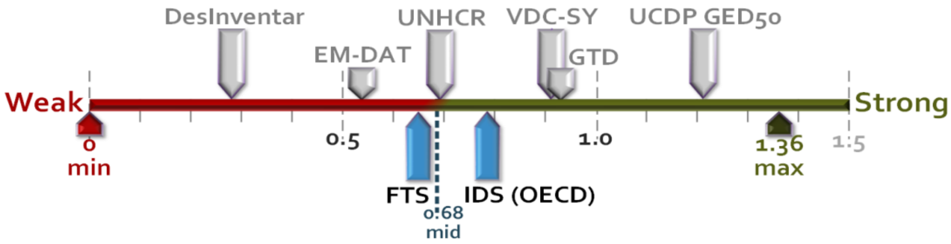


Figure 7-19: IDS (OECD) on the Dataset DVi Scale

Considering the scale of noise that is filtered out of IDS tables, the low DVi for L2 dimension Uncluttered is not surprisingly. The lack of detail, sparse explanation and limited metadata places more downward pressure on the DVi [*Figure 7-20*].

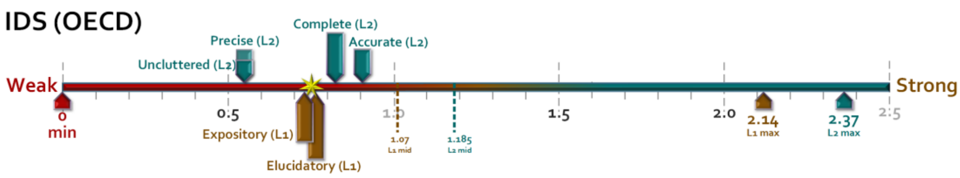


Figure 7-20: IDS (OECD) on the L2 & L1 DVi Scale

The data veracity evaluation of IDS did not identify any entries that can be flagged as soft. A standalone view of IDS does not provide enough information to distinguish firm entries from soft entries

7.4.3 Population

The population figures required for further analysis are at two levels: the first being the population of each disaster-affected country in the year that the disaster(s) occurred; the second being the global population from 1990 to 2015. As this exercise is expected to be the simple acquisition of two sets of numbers, and not the acquisition of a dataset, the DVf toolset is not applied to population values.

The *tentative design* of this iteration identified the World Bank's Databank as the key source for population data (World Bank, 2017). Before country/year level population figures are obtained the following compromises are made to align disaster countries to the population data:

- Tibet and Taiwan are assigned China's ISO 3166 alpha3 code as their populations are counted as China (ISo-3166, 2017).
- Kosovo's independence from Serbia in 2008 is not internationally recognised, therefore for population figures it is assigned Serbia's ISO 3166 alpha3 code.

In total 4,950 population figures are needed to accommodate all per country per year entries, however 153 country/year combinations cannot be found in the population numbers from World Bank's Databank [*Appendix H.8*] (World Bank, 2017). For these 153 combinations alternative sources are identified:

- 115 are obtained from the United Nations Conference on Trade and Development statistics database (UNCTAD, 2017);
- 34 are obtained from a website called Worldometers (Worldometers, 2017); and
- 4 are obtained from a website called Population Pyramid (Population Pyramid, 2017).

The last two of these sites collate population data from a number of credible sources including the UN and WHO (UN, 2017; WHO, 2017). Finally, considering the shortfall in population data availability in the World Bank's Databank (World Bank, 2017); world population figures are obtained from Worldometers (Worldometers, 2017). Interestingly, obtaining Worldometer's annual world population size also provides access to two other potentially useful values, annual population density, people per square kilometre, and urban population (ibid).

7.4.4 Data Landscape Anomalies

Before searching for M^iIs and M^iEs it is worth scanning for anomalies in the alignment of the datasets acquired, prepared, evaluated and, as in the case of the MSGD and MDC, created so far. Any misalignment can only be identified at the level of country/year at best as the datasets and values acquired during this iteration cannot facilitate cross-referencing specific disasters with the flow of funds. These anomalies are either obvious where no corresponding entries between MSGD country/year combinations and aid funding exist; or are slightly less obvious disproportionate funding flows.

MSGD country/years with No FTS Humanitarian Aid

There are 4,950 year/country combinations in MSGD 1990–2015. FTS has humanitarian aid data for only 1,943 of these combinations, less than 40%. Of the 3,059 MSGD year/country combinations with no corresponding FTS entries 1,616 are for years 1990–1998, which provides an explanation, given that FTS does not have data for years before 1999. The lack of FTS recorded funding for the remaining 1,443 MSGD country/year combinations – equalling more than forty thousand disasters and over four hundred million people affected – are more difficult to explain [*Appendix H.9*].

MSGD country/years with No IDS Humanitarian Aid

Of the 4,950 year/country combinations in MSGD 1990–2015, 3,383 combinations have non-missing values for IDS humanitarian aid. Five of these can be considered anomalous; four have an explicit zero value and one entry has an inexplicable -$60K as a value [*Appendix H.9*]. Another 1,619 country/year combinations have no IDS recorded humanitarian aid numbers. These country/year combinations represent over 250k deaths and 143m people affected and in excess of US$2 trillion in financial losses [*Appendix H.9*].

MSGD country/years with No IDS Other Aid

Here one of the MSGD 4,950 country/year combinations has an explicit zero value for all other IDS documented aid. This is for Croatia in 1992, for 37 disaster entries that show 213 deaths and 168,595 people affected. A further 1,151 MSGD country/year combinations do not have any IDS records showing that any other aid is provided [*Appendix H.9*].

MSGD country/years – FTS AND IDS Humanitarian Aid

1,770 of the MSGD 4,950 country/year combinations have both FTS and IDS humanitarian aid figures. It is clear that the numbers if added together inflate humanitarian aid to levels that are a far cry from those quoted in official reports (Lattimer et al., 2016). For example, in 2015 the UN stated need is reported to be US$19.8bn of which US$8.9bn is said to be unmet (ibid) [*Figure 4-12*]. If IDS and FTS humanitarian values for 2015 are added the resulting US$35.4bn is excessively high, therefore considered improbable [*Appendix H.9*]. The question is then which of the two humanitarian aid values, FTS or IDS, should be considered the more credible [*Figure 7-21*].

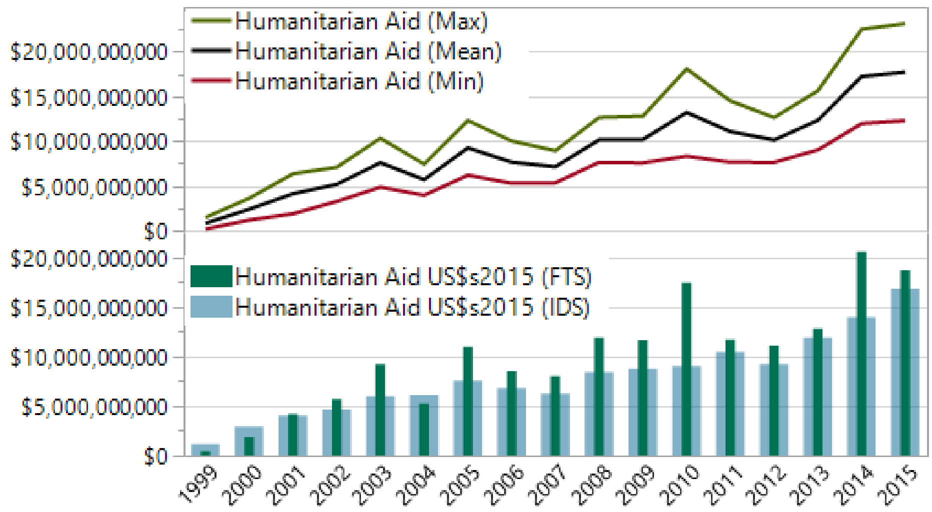


Figure 7-21: FTS vs IDS Humanitarian Aid

The bar chart shows that from 2005 onwards FTS holds higher figures for humanitarian aid than IDS. The line chart shows the change in aid when the highest, lowest and mean values between FTS and IDS are taken. Unfortunately, none of this information provides guidance as to which is the most appropriate to pursue, therefore both FTS and IDS humanitarian aid figures are retained for comparison (IDS, 2017; FTS, 2017h). Choropleth maps of five-year totals of maximum humanitarian aid, i.e. time-phased snapshots of the recipients of the best case scenario of humanitarian assistance, can be found in *Appendix H.10*.

Disproportionate Humanitarian Aid

There are MSGD country/year combinations where humanitarian aid figures exist but the flow or values are incongruent or nonsensical, such as:

- **8** country/year MSGD combinations representing 367 disasters in which 379 people died and almost 2.25 million people were affected, but the flow of aid is negative totalling over US$1.2bn [*Appendix H.9*]. One combination also has details of other aid, but 7 do not therefore are excluded from analysis as anomalous data.
- **30** country/year MSGD combinations that appear to have received aid to the tune of over US$250,000 per person affected [*Appendix H.9*]. These 30 combinations are flagged as suspect, 'soft', anomalous data.
- **146** country/year MSGD combinations that appear to have received aid of less than US$1 person affected. These combinations are significant: almost 28.5K disasters, over one million deaths and almost four billion people affected, yet on average US$0.38 per person in humanitarian aid [*Appendix H.9*]. These less than a $1 humanitarian aid per person values for these also do not appear to show any corresponding pattern with mean survival rate [*Figure 7-22*]. All 146 combinations are flagged as suspect, 'soft', anomalous data.

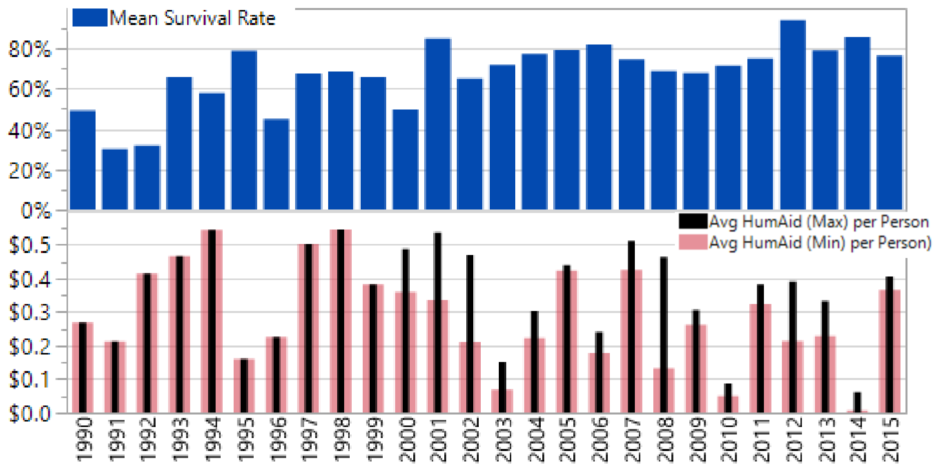


Figure 7-22: MSGD Country/Years Less than US$1 per Person Survival Rates

Humanitarian Aid but No Disasters

Both FTS and IDS have entries that show the flow of humanitarian aid for country/year combinations that have no disasters in MSGD (IDS, 2017; FTS, 2017h). There are **16** FTS entries equalling over ***US$14.4m*** humanitarian aid for countries that do not have corresponding entries for disasters in the MSGD [*Table 7-9*].

Year	Recipient Country	US$Total_2015
2003	Aruba	$5,378,437
	Vanuatu	$33,615
2006	Anguilla	$11,757
	Cook Islands	$117,568
	Marshall Islands	$2,680,549
	Micronesia, Federated States of	$693,651
	Palau	$364,461
	Saint Helena, Ascension and Tristan da Cunha	$105,811
2009	Marshall Islands	$154,084
2010	Palau	$104,040
2011	Cook Islands	$900,881
	Samoa	$800,588
2012	Palau	$337,656
	Vanuatu	$515,991
2013	Cook Islands	$466,392
	Vanuatu	$1,743,784
	Total	**$14,409,265**

Table 7-9: FTS Humanitarian Aid – no MSGD Disasters

There are also **116** IDS humanitarian aid flows equating to almost ***US$144m*** for which there are no disaster entries [*Table 7-10* & *Figure 7-23*].

Year	Countries	US$s2015	Year	Countries	US$s2015
1990	8	$24,390,000	2003	4	$2,030,000
1991	1	$50,000	2004	3	$560,000
1992	7	$1,450,000	2005	8	$770,000
1993	5	$8,760,000	2006	7	$3,970,000
1994	1	$60,000	2007	4	$720,000
1995	4	$1,660,000	2008	4	$710,000
1996	3	$570,000	2009	9	$6,850,000
1997	3	$16,860,000	2010	6	$3,970,000
1998	6	$24,510,000	2011	4	$7,430,000
1999	1	$10,920,000	2012	6	$9,030,000
2000	3	$8,660,000	2013	6	$3,800,000
2001	5	$2,890,000	2014	3	$1,040,000
2002	1	$2,090,000	2015	4	$180,000
		Total	$143,930,000		

Table 7-10: IDS Humanitarian Aid – no MSGD Disasters

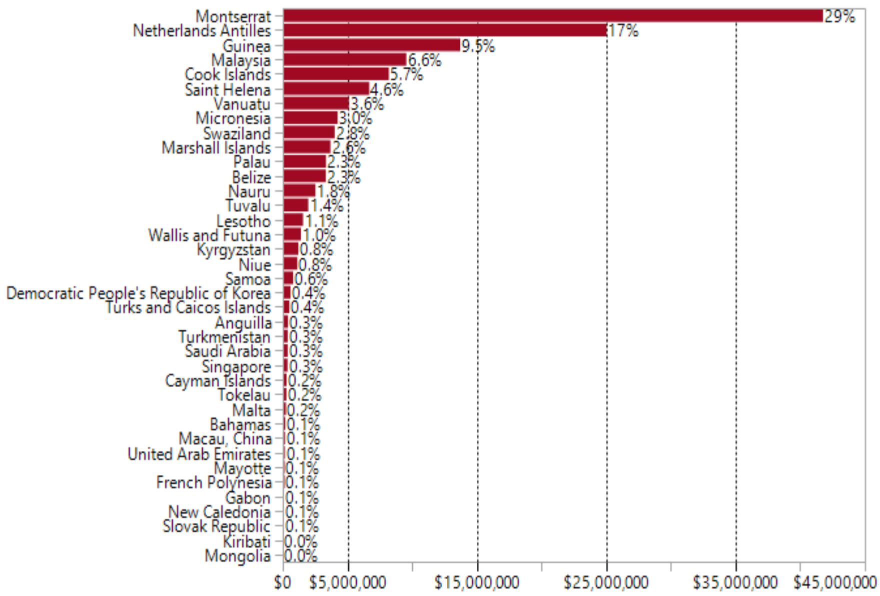


Figure 7-23: IDS Humanitarian Aid – no MSGD Disasters

Finally, the extent of the anomalies at this high level of year/country totals is unexpected and there are extraordinarily significant gaps and mismatches where:

- humanitarian aid is being sent to countries and in years when none of the six disaster databases can give evidence of any disasters in need of aid; or

- millions of people are recorded to have been affected by disasters but there is no flow of humanitarian aid into those countries in the years in which these disasters occurred; or
- there are disproportionate flows of aid. For example, during 1991 in Guinea the only incidents on record are two refugees leaving the country, one to Denmark, the other to Sweden, but according to IDS Guinea receive US$20.99m of humanitarian aid.

If all anomalous year/country combinations are excluded from analysis only 1,721 year/countries remain; less than 35% of the year/countries that have experienced disasters 1990–2015. This inherent weakness may be related to inconsistencies in the attribution of aid versus disasters to countries. Regardless of its cause, not enough information is available to resolve such inconsistencies. Therefore a workaround is needed, which in this case is to raise the level of analysis to global and annual figures.

7.4.5 The Search for M^iIs and M^iEs

This section returns to the M^iO identified in the previous chapter, namely the *Mean Survival Rate by Year,* and places it in context with humanitarian aid to search for M^iIs, and development aid and population figures to search for M^iEs.

Search for Macro-Indicators of Impact (M^iIs)

From 1999 to 2015 humanitarian aid data from FTS and IDS overlap and it is not clear which of the two sources take precedence (IDS, 2017; FTS, 2017h). A pragmatic approach is needed to ensure the possible effects of any significant variation between these figures is not lost, which would be the case if either FTS or IDS figures are taken to the exclusion of the other. As such, three variables are created:

- HumAid (MAX) = whichever is the higher of the two values;
- HumAid (MIN) = whichever is the lower of the two values;
- HumAid (MEAN) = the average of the two values.

Of course where only one of the two, FTS or IDS, humanitarian aid values is available it is used as is. Additionally, to provide perspective, each of these versions of the annual global humanitarian figure is divided by the annual global total human effect of disasters (i.e. deaths+survivors) to obtain a per person figure for humanitarian aid.

Figure 7-24 includes three charts. The lowest nested bar chart shows the growth in humanitarian aid over the 26-year period of interest, 1990–2015. From 1999 onwards the difference between HumAid (MAX) and HumAid (MIN) is noticeable. The central nested bar chart depicts the amount of humanitarian aid per person affected by disaster each year. The increase of humanitarian aid per person is almost in line with the increase in total annual humanitarian aid. As all US$ values are adjusted to 2015 levels this parallel increase is not related to the increasing value of money nor does it appear to be related to greater numbers of people in need of assistance. Note that the changes in humanitarian aid over time (total and per person) is not reflected in the trend pattern of survival rates (topmost chart).

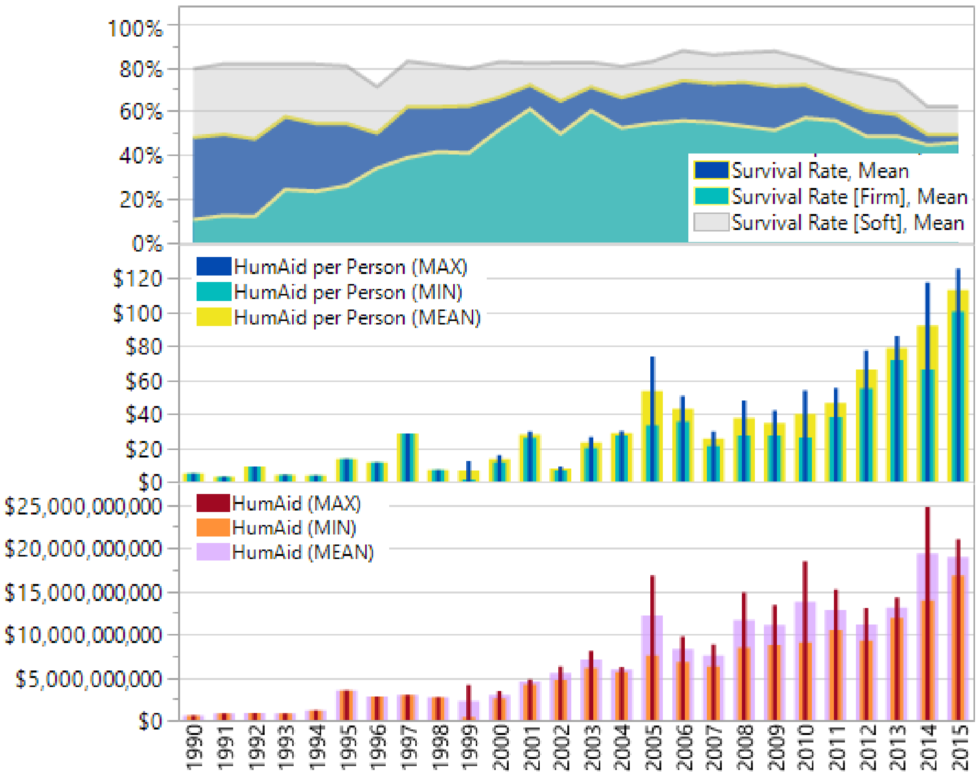


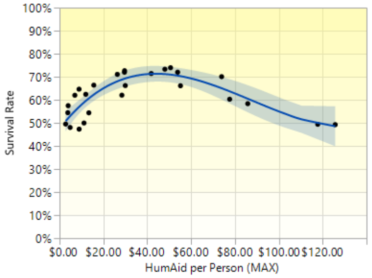
Figure 7-24: Annual Survival Rate and Humanitarian Aid Charts

To explore for M^iIs the null hypothesis is based on humanitarian aid per person as this provides a perspective of aid versus human effect that cannot be gained by simply using humanitarian aid. Here the following null hypothesis, H_0, is tested by plotting mean survival rate against US$ humanitarian aid per person [*Figure 7-25*].

H_0 There is no relationship between *humanitarian aid per person* and *mean survival rate*.

Note, the charts and analyses for *mean survival rate* and HumAid (MAX) are shown here, equivalent charts for HumAid (MIN) and HumAid (MEAN) can be founding *Appendix I.1*. As before when searching for M^iOs three sets of charts are created using the combined (undifferentiated) *mean survival rates*, the firm *mean survival rates* and the soft *mean survival rates*.

Survival Rate (Undifferentiated)



Survival Rate = 0.6580567 + 0.0013653*HumAid per Person (MAX) - 0.0001059*(HumAid per Person (MAX)-37.1636)^2 + 7.0026e-7*(HumAid per Person (MAX)-37.1636)^3

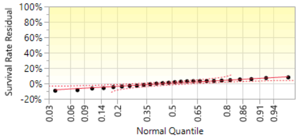
Summary of Fit

RSquare	0.725858
RSquare Adj	0.688475
Root Mean Square Error	0.050709
Mean of Response	0.61737
Observations (or Sum Wgts)	26

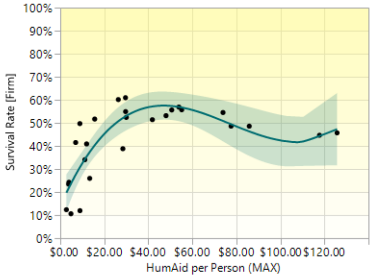
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.14978646	0.049929	19.4168
Error	22	0.05657140	0.002571	Prob > F
C. Total	25	0.20635786		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Survival Rate [Firm] = 0.4518508 + 0.0029524*HumAid per Person (MAX) - 0.0001723*(HumAid per Person (MAX)-37.1636)^2 + 1.4424e-6*(HumAid per Person (MAX)-37.1636)^3

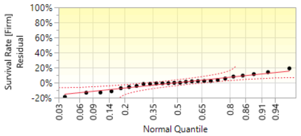
Summary of Fit

RSquare	0.688567
RSquare Adj	0.646099
Root Mean Square Error	0.09168
Mean of Response	0.426508
Observations (or Sum Wgts)	26

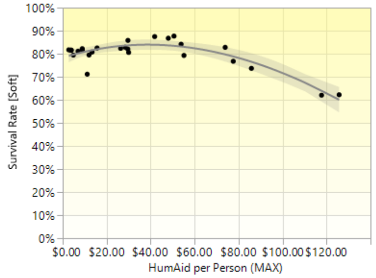
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.40883880	0.136280	16.2137
Error	22	0.18491424	0.008405	Prob > F
C. Total	25	0.59375303		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Survival Rate [Soft] = 0.8354347 + 0.0001112*HumAid per Person (MAX) - 3.5886e-5*(HumAid per Person (MAX)-37.1636)^2 + 4.7314e-8*(HumAid per Person (MAX)-37.1636)^3

Summary of Fit

RSquare	0.778565
RSquare Adj	0.748369
Root Mean Square Error	0.032197
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.08018788	0.026729	25.7840
Error	22	0.02280660	0.001037	Prob > F
C. Total	25	0.10299448		<.0001*

Residual Normal Quantile Plot

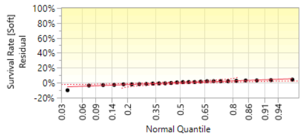


Figure 7-25: Mean Survival Rate x Humanitarian Aid Per Person (MAX)
Polynomial Line of Fit (Degree=3)

All nine incarnations of these scatter plots and polynomial lines of fit (i.e. the three in *Figure 7-25* and the six in *Appendix I.1*) are statistically significant and the null hypothesis can be rejected. Note that the R^2, F-Ratio, and Prob>F are most significant for the humanitarian aid using the maximum values of either FTS or IDS – i.e. HumAid (MAX) per Person [*Figure 7-25*] (IDS, 2017; FTS, 2017h). Note that all variants of FTS/IDS humanitarian aid per person affected by disasters confirm that it has a statistically significant relationship with *mean survival rate* at the macro-level (annualised aggregates). As a result, ***Mean Survival Rate by Humanitarian Aid per Person*** is taken to be a **M^{i}I** – in that it provides a high level view of the relationship between survival and humanitarian aid.

From *Figure 7-25* it can be seen that plots vary depending on the veracity of the survival data, but in all three plots survival peaks at around US$50 per person. As the amount of humanitarian aid per person increases towards US$120:

- the soft *mean survival rate*, which peaked at just below 90% gradually slopes to around 60%;
- the firm *mean survival rate*, peaks at around 60% falling to around 40% at US$100 per person, before rising slightly between US$100 and US$120 per person;
- the undifferentiated plot shows the effect on the blended soft and firm data with the *mean survival rate* peaking at 70% at circa US$50 per person and then steadily falling to 50% as humanitarian aid moves towards US$120 per person.

In all cases, *Mean Survival Rate by Humanitarian Aid per Person (*M^{i}I) signposts that increasing humanitarian aid above US$50 per person does not necessarily return a higher likelihood of surviving a disaster. It is not argued here that more money is the cause of poorer survival rates. Note that this does, however, offer a strong indication that disasters receiving high humanitarian aid funding are worthy of investigation as there may be legitimate or correctible underlying reasons why more aid does not translate to more lives saved.

Search for Macro-Indicators of Effect (M^{i}Es)

This exploration attempts to place mean survival rate in context with other aid and also population figures to identify any statistically significant relationships. With international (other) aid this is seen as an opportunity to test the assertion that there is a blurring of lines between development (other) aid and humanitarian aid. If this is the case there should be a statistically significant relationship between other aid and survival (Riddell, 2014a). With population the intent is to test the assertion that disaster outcomes worsen as population figures increase (HERR, 2011).

(a) Other International Aid

First a cursory view of other financial aid [*Figure 7-26*]:

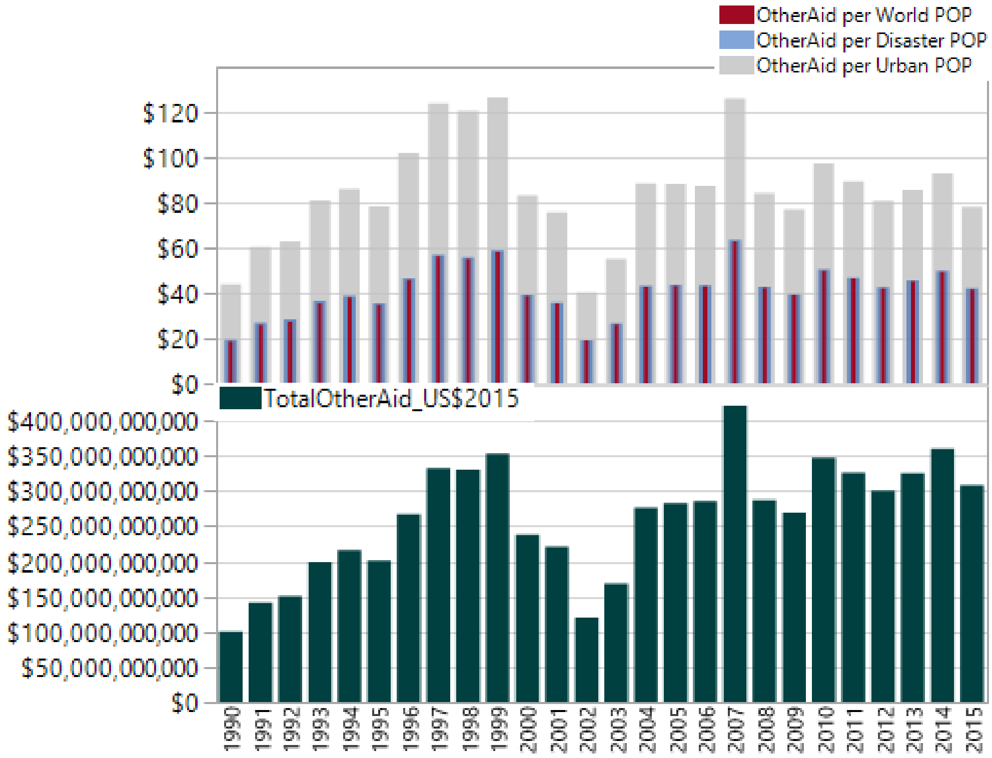
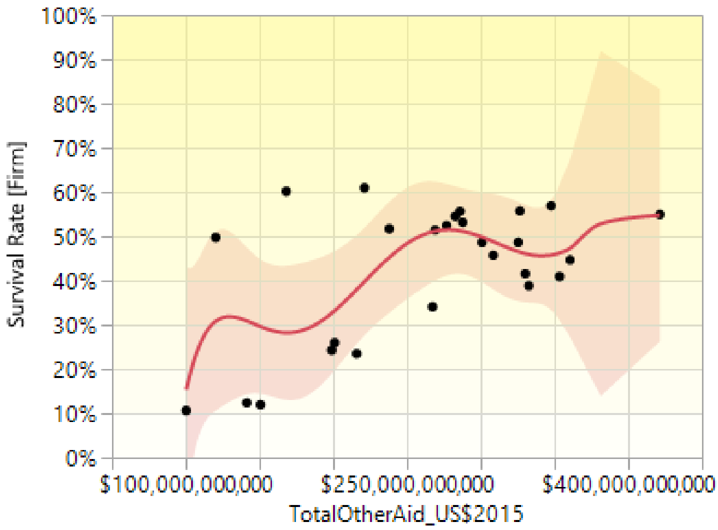


Figure 7-26:IDS Other Aid by Year

These charts show the scale of all financial assistance, other than humanitarian aid, as recorded in IDS (IDS, 2017). The lower bar chart depicts the annual total, the upper nested bar chart shows the per capita US$s spent on aid based on annual *urban* population, *world* population values, and the total population of disaster-affected countries (*Disaster POP*). Moving forward, as before a null hypothesis is constructed to search for a relationship between other aid and mean survival rate:

H_o There is no relationship between *other aid* and *mean survival rate*.

Note, every variation of other aid, either as a total or at the per population level, fails to reject this null hypothesis. As an example, even the very best 'fit' of all the possible permutations, *Mean Survival Rate [Firm]* by *Total Other Aid* [*Figure 7-27*] shows no statistical significance.



Survival Rate [Firm] = 0.1881884 + 1.208e-12*TotalOtherAid_US$2015[IDS_YR_ID] - 4.091e-23*(TotalOtherAid_US$2015[IDS_YR_ID]-2.6e+11)^2 - 3.639e-35*(TotalOtherAid_US$2015[IDS_YR_ID]-2.6e+11)^3 + 3.778e-45*(TotalOtherAid_US$2015[IDS_YR_ID]-2.6e+11)^4 + 1.249e-57*(TotalOtherAid_US$2015[IDS_YR_ID]-2.6e+11)^5 - 9.366e-68*(TotalOtherAid_US$2015[IDS_YR_ID]-2.6e+11)^6

Summary of Fit

RSquare	0.404871
RSquare Adj	0.216936
Root Mean Square Error	0.136374
Mean of Response	0.426508
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.24039351	0.040066	2.1543
Error	19	0.35335952	0.018598	Prob > F
C. Total	25	0.59375303		0.0941

Figure 7-27: Mean Survival Rate x Total Other Aid
Polynomial Line of Fit (Degree=6)

As a result, other aid is abandoned from further searches for M^{i}Es. This also challenges, at the aggregate level, the perception that other (development) aid is used to supplement humanitarian aid (Riddell, 2014a). If this use of international development aid does take place it does not appear to be wide-spread enough to percolate up to a macro-level relationship with disaster survival.

(b) Populations Figures

Before exploring population data for M^{i}Es it is worth gaining a sense of the population data. *Figure 7-28* depicts bar charts of population growth and density and mortality by population.

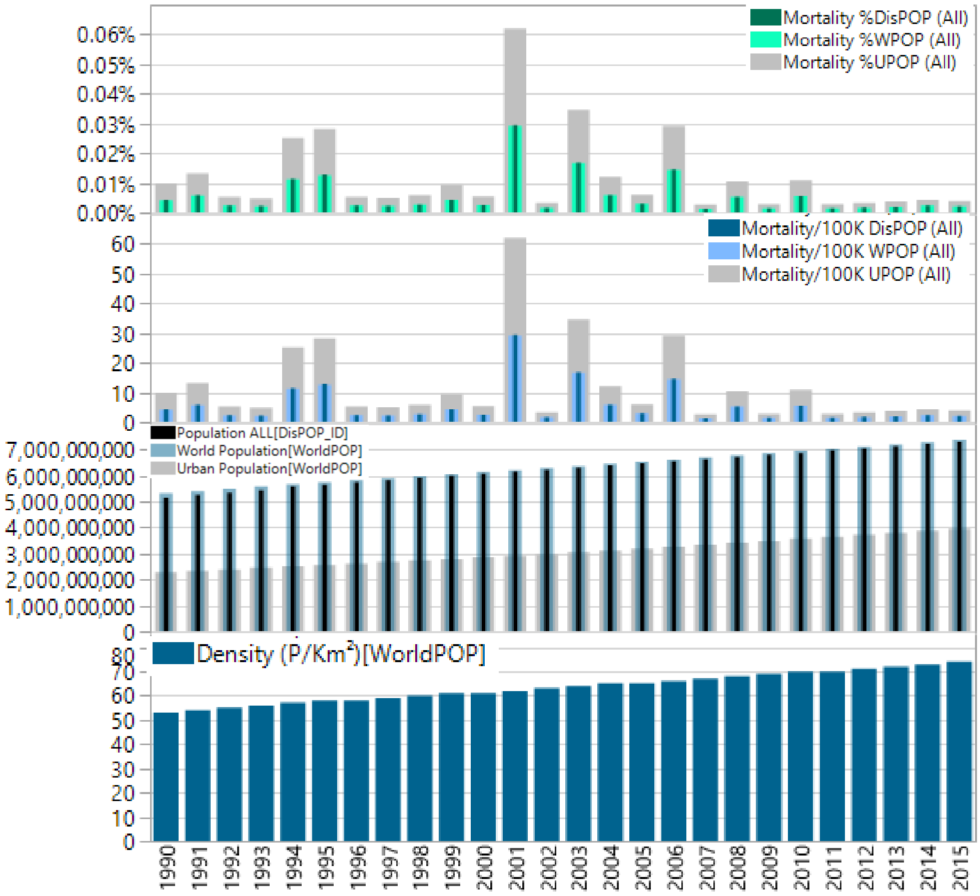


Figure 7-28: Population & Mortality Bar Charts

The lowest chart shows the annual increase in global population density. The remaining three nested bar charts are based on three types of global populations: (1) **DisPOP** a calculated annual total based on the population figures obtained from the World Bank's

Databank and other sources for each disaster-affected country (World Bank: HN&P, 2017); (2) **WPOP** a world population figures obtained from Worldometers (Worldometers, 2017); and (3) **UPOP** the annual urban population figures (ibid).

Taking guidance from the Sendai framework, which views mortality per 100K population as a measure of progress in disaster risk reduction (Wahlström, 2015), the upper two charts depict the numbers of disaster-related deaths per 100K of each population type and as a percentage of the total of each population type. Notably, neither mortality charts show a pattern of decline in deaths, or any particular pattern at all. Nevertheless it is worth testing two null hypotheses:

(i) Population Mortality and Mean Survival Rate

> **H_0 There is no relationship between *population disaster mortality rates* and *mean survival rate*.**

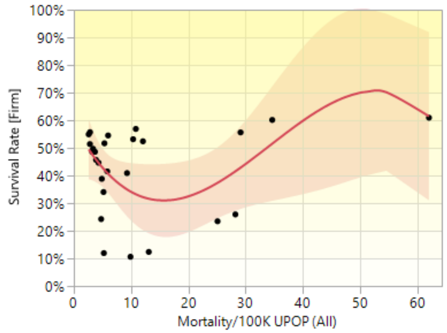
All variation of mortality rate, as a percentage and per 100K, are plotted against mean survival rate. All variations fail to reject this null hypothesis. The best of these 'bad fits' is *mean survival rate [Firm]* by *Mortality per 100K Urban Population* [*Figure 7-29*], which serves to confirm the lack of statistical significance between these factors.

(ii) Humanitarian Aid by Population Mortality

> **H_0 There is no relationship between *humanitarian aid* and *population disaster mortality rates*.**

All variation of mortality rate, as a percentage and per 100K, are plotted against each level – MAX, MIN and MEAN – of humanitarian aid. All variations fail to reject this null hypothesis. The best of these 'bad fits' is *Mortality per 100K Urban Population* by *Humanitarian Aid (Mean)* [Figure 7-30], which serves to confirm the lack of statistical significance between these factors.

(i) Population Mortality and Mean Survival Rate



Survival Rate [Firm] = 0.411096 - 0.0073554*Mortality/100K UPOP (All) + 0.0010492*(Mortality/100K UPOP (All)-11.8685)^2 - 1.5681e-5*(Mortality/100K UPOP (All)-11.8685)^3

Summary of Fit

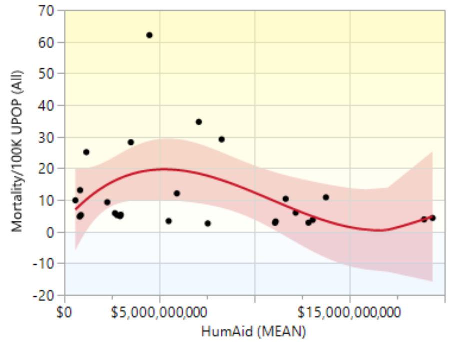
RSquare	0.201028
RSquare Adj	0.092077
Root Mean Square Error	0.146844
Mean of Response	0.426508
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.11936080	0.039787	1.8451
Error	22	0.47439223	0.021563	**Prob > F**
C. Total	25	0.59375303		0.1685

Figure 7-29: Mean Survival Rate x Mortality/100K UPOP

(ii) Humanitarian Aid by Population Mortality



Mortality/100K UPOP (All) = 28.148532 - 1.3857e-9*HumAid (MEAN) - 3.076e-19*(HumAid (MEAN)-7.08e+9)^2 + 2.693e-29*(HumAid (MEAN)-7.08e+9)^3

Summary of Fit

RSquare	0.171798
RSquare Adj	0.058861
Root Mean Square Error	13.28011
Mean of Response	11.86846
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	804.8373	268.279	1.5212
Error	22	3,879.9490	176.361	**Prob > F**
C. Total	25	4,684.7863		0.2369

Figure 7-30: Mortality/100K UPOP x Humanitarian Aid (Mean)

Polynomial Line of Fit (Degree=3)

As neither of the mortality related null hypotheses can be rejected, mortality rate is abandoned from further searches. The study now moves to exploring population in relation to mean survival rate for possible M^iEs. Here the null hypothesis tested is:

H_o There is no relationship between *population* and *mean survival rate*.

There are at least four sets of population figures that can be used to test this hypothesis:

(1) ***DisPOP***: *global population totals* calculated by adding the populations of disaster-affected countries for each year, 1990–2015, based on annual country population figures from a variety of sources (World Bank: HN&P, 2017; UNCTAD, 2017; Population Pyramid, 2017);

(2) ***WorldPOP***: *world population* figures, 1990–2015 (Worldometers, 2017);

(3) ***UrbanPOP***: *world <u>urban</u> population* figures, 1990–2015 (Worldometers, 2017);

(4) ***DensityPOP***: *world population <u>density</u>* figures, 1990–2015 (Worldometers, 2017).

Charts and reports testing the null hypotheses based on each of these population figures can be found in *Appendix I.2*. All four variations of population figures confirm the null hypotheses can be rejected.

Of note is that all population charts bear a striking resemblance to the plots of *Mean Survival Rate by Year* [*Figure 6-33*]; which is illustrated by *Figure 7-31*. This is not altogether surprising considering all four versions of population figures edge up each year [*Figure 7-28*].

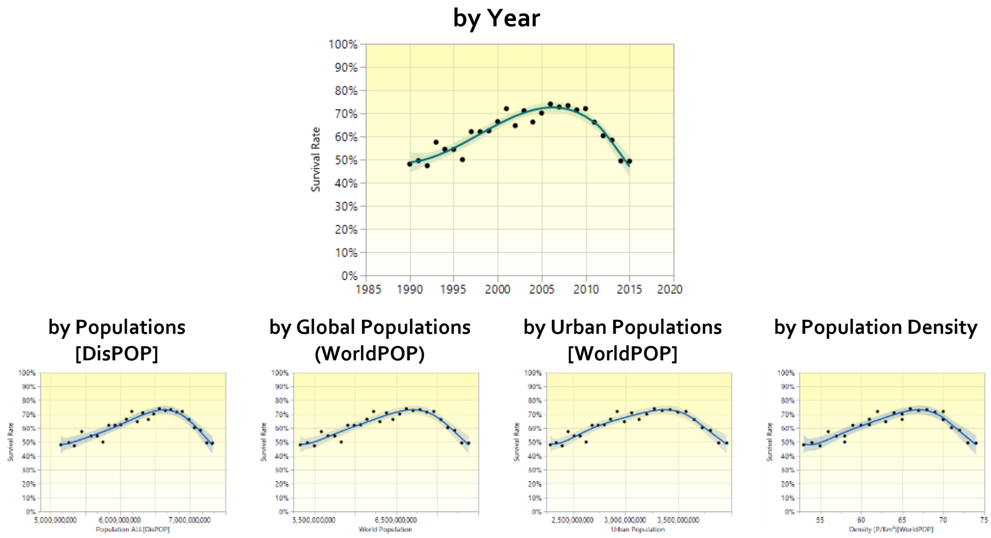


Figure 7-31: Mean Survival Rate [Undifferentiated] by Year vs by Population

To examine if this behaviour continues at the regional level equivalent charts are created for calculated totals of regional populations [Figure 7-32]. The full suite of these plots and supporting reports can be found in *Appendix I.3*.

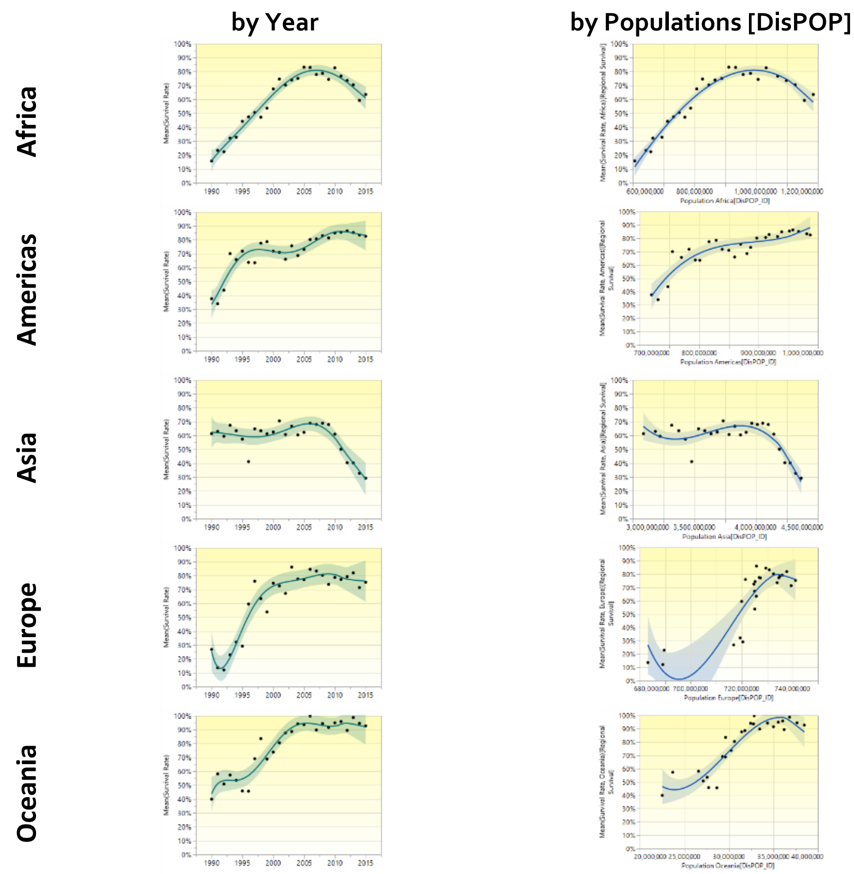


Figure 7-32: Mean Survival Rate [Undifferentiated] by Year vs by Population (Regional)

Even at the regional level the plots *by year* and *by population* look remarkably similar. That said, there are differences in the R^2, F Value and Pro>F values between the sets of charts, as well in the confidence shading and the shapes of the Americas, Europe and Oceania charts vary subtly. Notably, the similarities between the two sets of charts continue when comparing different, i.e. firm or soft, veracity levels of mean survival rate [*Appendix G.4.2* and *Appendix I.3*].

Even though *mean survival rate by population* plots shadow mean *survival rate by year,* plots ***mean survival rate by population*** is considered a viable ***MⁱE***, as there is a statistically significant relationship. Though these charts may not provide insight into a direct causal relationship between population and survival, they do challenge a key assertion that as populations grow, particularly urban populations, the effect of disasters will worsen (HERR, 2011). That is, they suggest, that a perceived and oft quoted effect of population growth is not borne out by the data. In all versions of the population plots increasing population numbers do <u>not</u> have a blanket negative effect on survival.

7.5 Conclusion

This section brings the study to a conclusion and also constitutes its final evaluation step. It follows a similar structure to the evaluation steps of previous iterations in that it discusses the alignment of the *build (grow)* step with the *tentative design*; the *DSR artefacts* created; the *knowledge* gained and the *consequences* of this knowledge; and the *design theory/theories* relevant to this iteration. (Hevner et al., 2004; Simon, 1996; Hooker, 2004; Venable et al., 2012; Venable, 2013; Gregor, 2006; Vaishnavi and Kuechler, 2004b). As this section concludes the study, included here are reflections of this research as a whole and an acceptance that the outcome of this iteration is '*good enough*' to complete the study and finalise the knowledge contribution of this work (March and Smith, 1995; Simon, 1996; Gregor and Jones, 2007).

7.5.1 Tentative Design⇔Build (Grow) Alignment

The *tentative design* for this iteration identifies sources of humanitarian aid, development aid and population figures as follows:

- For humanitarian aid UNOCHA's Financial Tracking Services (FTS), which tracks the international flow of humanitarian aid is selected (UNOCHA, 2017; FTS, 2017a).
- For development aid dataset OECD's International Development Statistics (IDS) online databases is selected (OECD, 2017b; IDS, 2017).
- For annual global and country-level population figures the World Bank's Databank *Population Statistics* are selected (World Bank: HN&P, 2017).

A number of Issues are experienced with these datasets including:

- Detail level FTS data do not reconcile with summary level FTS data (FTS, 2017d; FTS, 2017h). Humanitarian funding flow at the detail level falls shorts of the humanitarian funding flow at the summary level by over US$11.5bn (ibid).
- FTS does not contain any humanitarian aid flow data prior to 1999, even though it is documented as starting its data collection in 1992 (UNOCHA, 2017; FTS, 2017a).
- The IDS dataset contains fifty-three aid variables of which five are considered useful. These five variables include humanitarian aid. Not that even though high level descriptions are available for each IDS data table, of which four are acquired, there is little information about each of the variables. As a result it is not clear how the humanitarian aid values in IDS compare to the humanitarian aid values in FTS (OECD, 2017b; IDS, 2017).
- Comparison of countries and the years in which they received humanitarian aid funding with the years in which they

experienced disasters highlighted some intriguing anomalies. Numerous year/country combinations with recorded disasters do not have any record of humanitarian aid flowing to that country in that year. Similarly, there are numerous flows of humanitarian aid to countries in years in which they did not record any disasters of any type. There are also incongruous entries of the type that show the flow of over USD$20m of aid in support of two individuals seeking refuge from Guinea to the Nordic region. In the end only 35% of year/country combinations in MSGD did not have obviously anomalous aid flow.

- The Worldbank population statistics proved to be an insufficient source of data as it did not contain population figures for 153 year/country combinations of the 4,950 combinations needed (World Bank, 2017). This missing population figures needed to be sourced from elsewhere (UNCTAD, 2017; Worldometers, 2017; Population Pyramid, 2017). The gap in Worldbank population data raised concerns that global population totals from this source may not be complete, therefore Worldometer global population totals were obtained as cross-check (World Bank, 2017; Worldometers, 2017)

One key consequence of the misalignment between aid data and disaster data is that aid related analysis is restricted to the use of global annual figures. The misalignments of aid and disasters are assumed to represent less than clear attribution of the flow of funds, therefore examining aid flow at the country-level is abandoned. Thus the search for any M^iIs and M^iEs cannot be taken to any geographical level of detail as the gaps in country-level funding information are significant enough to arbitrarily skew results.

7.5.2 DSR Artefacts

Table 7-11 maps out the incremental additions to the pool of research artefacts created through this last iteration of the DSR design cycle. By way of actualising the *Mean Survival Rate by Humanitarian Aid* *per Person* and *Mean Survival Rate by Regional Population* the study argues that these are viable new constructs of $M^{i}I$ and $M^{i}E$ respectively. These final artefacts also serve to offer a 'satisficing' response to the overall research question – *Can exploration of curated web-available data yield macro-indicators of humanitarian intervention in the aftermath of disasters?*

Research Framework			Research Activities			
			Design Science		Natural Science	
			Build (grow)	Evaluate	Theorise	Justify
Research Outputs	Artefacts	Constructs	[a][f]	→	[c]	- - ->
		Models	[d][g][h]	→		
		Methods	[e]	→		
		Instantiations	[b][i][j][k][l]	→		

[a]	**Macro-indicators of** disaster outcome and **the impact and effectiveness of humanitarian intervention** ($M^{i}O$s, **$M^{i}I$s and $M^{i}E$s).**
[b]	**Data analysis outputs and visualisations**
[c]	A (behavioural science) hypothesis relating the availability, or lack thereof, of humanitarian data and the flow of humanitarian aid that emerges from the domain knowledge and may be worthy of future research.
[d]	Data Veracity framework (DVf) and Data Veracity model (DVm)
[e]	Data Veracity profile (DVp) and Data Veracity index (DVi)
[f]	Expansion of the construct of '*data scaffolds*' for the humanitarian domain
[g]	Data structure of the Master Set of Global Disasters (MSGD)
[h]	Classification structure of Master Disaster Classification Model (MDC)
[i]	Master Set of Global Disasters dataset (MSGD)
[j]	Master Disaster Classification reference dataset.
[k]	**Data Veracity profile (DVp) and Data Veracity index (DVi) instances for** each of the six datasets amalgamated for the MSGD and **the FTS and IDS aid datasets**
[l]	Mean Survival Rate by Year as an actualised $M^{i}O$ **Mean Survival Rate by Humanitarian Aid per Person as an actualised $M^{i}I$** **Mean Survival Rate by Population as an actualised $M^{i}E$**

Table 7-11: DSR Output to Research Framework Mapping v.4
(Vaishnavi and Kuechler, 2004b; Hevner, 2007; March and Smith, 1995)

Notably, many attempts to identify a relationship between the outcome of disasters and the various forms of development, private, or other international funding failed to yield a result. Finally, numerous charts and diagrams are created to better understand the data that may be reusable artefacts and this iteration also further tests the DVf toolset by creating instantiations of DVps and DVis for the FTS and IDS datasets (IDS, 2017; FTS, 2017h).

7.5.3 Knowledge ⇔ Consequence

Nuggets of knowledge that emerge from this iteration include:

(a) Misalignment of Humanitarian aid and Disasters

Even at the highly aggregated level of year and country totals, reconciliation of aid to disasters is not possible. The best that can be achieved is a sense of how many US$s per disaster victim (dead or surviving) is spent globally each year.

(b) Aid and Disaster Data exist in Data Silos

Disaster and humanitarian crises identification and naming conventions and the flow of humanitarian aid identification and naming conventions are not created to allow accurately linking of aid to disasters.

(c) The flow of Aid is Opaque

FTS summary data do not reconcile to FTS detail data and no information is available as to how the summary totals are calculated or what data may be included in the summary but not provided in the detail (FTS, 2017d; FTS, 2017h). Inexplicably, and unexplained, humanitarian aid flow starts from 1999 in FTS even though the site claims data collection commenced in 1992. No trackable information or metadata is provided for IDS humanitarian aid data most of which overlaps with the humanitarian funding flow data from FTS. No information is available for either FTS or IDS that assists in the selection of one or the other as the system of record of humanitarian aid flow (IDS, 2017; FTS, 2017h).

<table>
<tr><th>Knowledge</th><th>Evaluate</th><th>Consequence</th></tr>
<tr><td>(a) Misalignment of Humanitarian aid and Disasters</td><td>As the primary drivers for humanitarian aid are humanitarian disasters and crises, there is an expectation that at the very least there is some alignment between countries/years of disasters and countries/years of humanitarian aid. This is not the case.

There are countries/years with few of no recorded disasters/victims yet in receipt of considerable humanitarian aid. Similarly, there are countries/years with significant disasters/victims yet no or very little humanitarian aid is provided.</td><td rowspan="2">No geographic alignment of humanitarian aid to disasters can be achieved let alone an alignment of aid to specific disasters. As a consequence, the only options available are to
• Use either the annual total; or
• Annual per capita of humanitarian aid calculated by dividing the annual total by the total number of people affected by disaster each year.</td></tr>
<tr><td>(b) Aid and Disaster Data exist in Data Silos</td><td>There is an expectation that as the total sums humanitarian aid involved equates to US$bns each year auditable links between aid and disasters exist. At the very least some similarity in disaster naming conventions may allow tying funds to humanitarian crises. Humanitarian aid data and disaster loss data appear to be maintained in vacuums.</td></tr>
<tr><td>(c) The flow of Aid is Opaque</td><td>Little information is available as to how humanitarian aid total are calculated; or why data is not available for years it is expected to be available; or how IDS values relate to FTS values.</td><td>For humanitarian aid figures from both FTS and IDS are retained. Three variables are created MIN, MAX and MEAN. For each of these variables for years where one or the other has a value that value is used. Where both FTS and IDS have values the higher of the two values is used for MAX, the lower of the two values is used for MIN and the average of the two values is used for MEAN.</td></tr>
</table>

Table 7-12: Iteration 3 Knowledge ⇔Consequence Mapping

7.5.4 The Utility Theory

Restating the utility theory statement of this study [*Table 7-13*]:

	Solution Space	Utility	Problem Space
	Form	Function	Purpose
	Artefact [*What*]	Efficacy [*How*]	to Address [*Why*]
STATEMENT	*Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector…*	*…when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level…*	*…the inability to gauge the consequences of monies spent and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.*

Table 7-13: Structure of the Utility Theory Statement

Mean survival rate when plotted against humanitarian aid per person affected or regional level population figures exhibits a plot that suggests that humanitarian aid and population figures have relationships with the impact and effect of disasters that is worthy of further investigation. While the macro-indicators found in this study are very high-level and cannot confirm causal relationships they provide visibility where previously there was none and therefore, it is argued, support the propositions of the design theory [*Table 7-13*].

Figure 7-33 illustrates the dependency flow and relationships between the constructs relevant to this iteration within the utility theory. The diagram shows that the macro-indicator constructs of *M^iI* and *M^iE* (*from this iteration*) rely on the construct of the *Data Veracity framework* (*from Iteration 2*) as well as the construct of *Data Scaffolds*, specifically the MSGD and MDC (*from Iteration 3*).

.

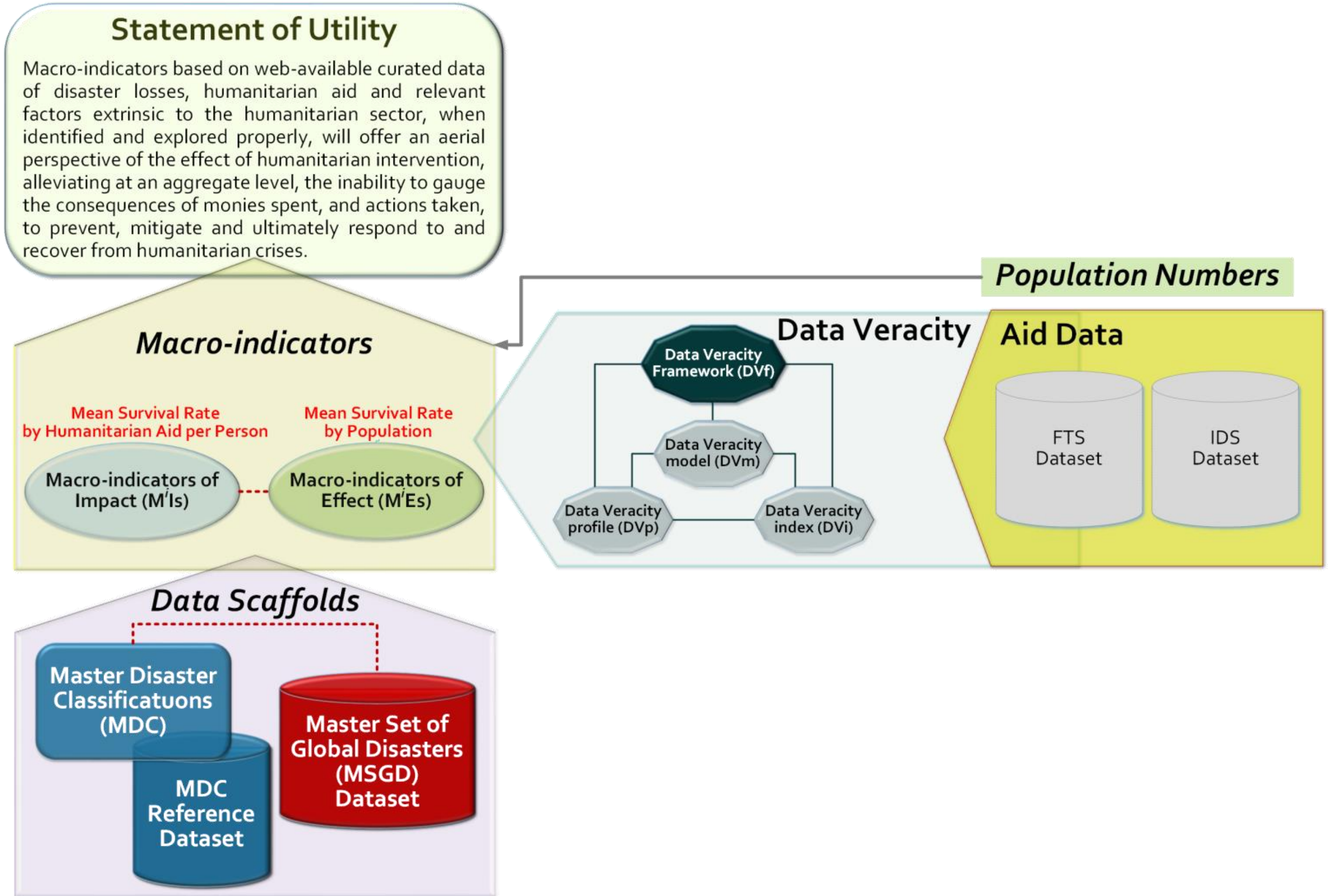


Figure 7-33: Constructs Contributing to the Utility Theory (Iteration 4)

7.5.5 Final Reflections

A key benefit of adopting DSR methodology for this research is the ability to adapt to emergent knowledge and limitations (Simon, 1996; March and Smith, 1995; Venable, 2006a; Hevner, 2007; Vaishnavi, 2008). This has proven to be invaluable in the journey of discovery completed by this study. Assumptions at the start of the research were challenged and therefore merited a detour to meet the overall aim of the research. For example, two fundamental preconceptions of available resources were found to be completely incorrect:

(1) *Data and classifications for most disasters are readily available.*
It was assumed that the humanitarian sector, which exists solely because of disasters, must have a standard definition and classification of all disasters types and at the very least veraciously documents most, if not all, disasters that have occurred in recent history. This assumption of the availability of key data resources stemmed from the certitude implied in the research presented by various dominant and credible actors in the humanitarian domain (HERR, 2011; GAR/UNISDR, 2015; HYOGO, 2005; Wahlström, 2015; Voigt et al., 2016; Guha-Sapir et al., 2017j; Toya and Skidmore, 2007; Kourosh and Richard, 2008; Strmberg, 2007; Alcántara-Ayala, 2002; Pears-Piggott and Muir-Wood, 2016; Blaikie et al., 2014; Sodhi, 2016; Corey et al., 2016; Raschky and Schwindt, 2016).

(2) *Data veracity has a definition and there are methods to assess it.*
It was assumed that the burgeoning interest in the study of data, with commentators frequently advocating the need for veracity, must mean that there is some shared understanding of what data veracity is and how to assess it (IBM; Janssen et al., 2017; Saporito, 2014; Berti-Equille and Lamine Ba, 2016; Lukoianova and Rubin, 2014; Powers Dirette, 2016; Normandeau, 2013).

The realisation that these assumptions were false required a change of tack and as such the originally planned 3 iterations for this DSR study were changed to 4 iterations [*Figure 7-34*].

Iteration	Original	Revised
1	Explore Disaster Losses for Macro-indicators of Outcome (M^iOs)	
2	Explore M^iOs and Humanitarian Aid for Macro-indicators of Impact (M^iIs)	Develop a Mechanism to Equitably and Consistently Evaluate the Veracity of all utilised datasets
3	Explore M^iOs, M^iIs, other international funding and population figures for Macro-indicators of Effect (M^iEs)	Create an amalgamated Master Set of Global Disasters, a Master Disaster Classification system, then explore the amalgamated Disaster Losses for Macro-indicators of Outcome (M^iOs)
4		Explore M^iOs, humanitarian aid, other international funding and population figures for Macro-indicators of Impact and Effect (M^iIs & M^iEs).

Figure 7-34: Comparison of Original and Revised Iterations

This adapted shape and flow of the study allowed the creation of usable alternatives to missing resources before returning to its initial course. Numerous additional artefacts are therefore created en route to creating the artefacts originally envisaged for this study [*Table 3-5*], and for the sake of completeness these have been mapped to the DSR research framework at the end of each iteration [*Table 4-4*, *Table 5-2*, *Table 6-14* and *Table 7-11*]. Note that these artefacts are not considered to have equal prominence as knowledge contributions from this work. The core knowledge-contributing artefacts of this research are those directly related to the research question – '*Can exploration of curated web-available data yield macro-indicators of humanitarian intervention in the aftermath of disasters?*' – as well as those created in lieu of the missing resources in the humanitarian and data science domains, namely:

M^{i}Os, M^{i}Is and M^{i}Es

The introduced constructs of M^{i}Os, M^{i}Is and M^{i}Es align to the research aim of this study and through simple bivariate modelling are identified as *mean survival rate* plotted by *year*, *humanitarian aid per person affected* and *population size* respectively [*Figure 7-35*]. While these relationships cannot be argued as causal in nature, there is sufficient correlation to justify that they provide visibility where previously there was none and signpost where further investigation is merited.

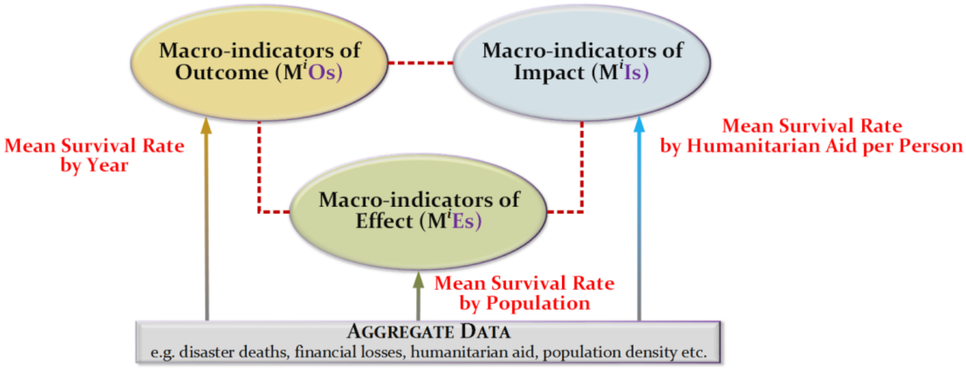


Figure 7-35: The Triad of Macro-indicators (Constructs)

These macro-indicators are not considered exclusive or exhaustive, just a beginning. Much of the time and effort of this research was necessarily expended on the basic groundwork of creating crucial resources absent for the humanitarian and data science domains. As such, the data analysis and models developed are rudimentary, sufficient to provide some understanding and insight, but not of the statistical sophistication that can fully explore the messages that may be contained in the data resources created. This is an innate compromise of a *satisficing* completion to this research.

Master Disaster Classification (MDC)

This is a model and corresponding dataset instantiation to classify all humanitarian crises that can attract humanitarian intervention [*Figure 7-36*]. The classification model builds on the glossary referenced by the EM-DAT and DesInventar databases, the IRDR, the Integrated Research on Disaster Risk Peril Classification and

Hazard Glossary (IRDR, 2014; Guha-Sapir et al., 2017g; DesInventar, 2017c). It also expands the naturogenic disaster classifications from the IRDR base to ensure that classifications exist for all other disaster types that have occurred, i.e. aggression-based events from UCDP, VDC-SY and GTD or where people are uprooted from their homes or countries, i.e. UNHCR *persons-of-concern* (UCDP, 2017d; VDC-SY, 2016a; GTD, 2017c; UNHCR, 2017b).

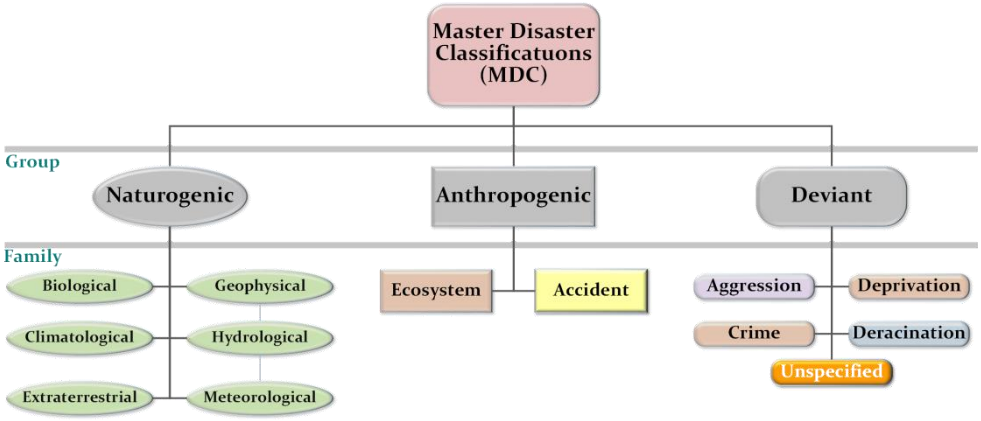


Figure 7-36: Master Disaster Classifications (MDC) – Top Tiers

The MDC is a key '*data scaffold*' of the humanitarian domain as without this constructing useful models of disaster variants is severely inhibited. The created MDC artefact set is sufficient for the needs of this study and its strength is that it covers all disaster types amalgamated for this work. This aspect is also its weakness. To be more robust the MDC needs further development and refinement through more congruous modelling across the disaster groups and the future-proofing inclusion of anthropogenic and deviant disaster events that have not yet occurred but are reasonably possible (for example).

Master Set of Global Disasters (MSGD)

This is a master set of global disasters losses created by amalgamating humanitarian crises events form six source datasets, EM-DAT, DesInventar, UCDP, VDC_SY, GTD and UNHCR [*Figure 7-37*] (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017d; VDC-SY, 2016a; GTD, 2017c; UNHCR, 2017b).

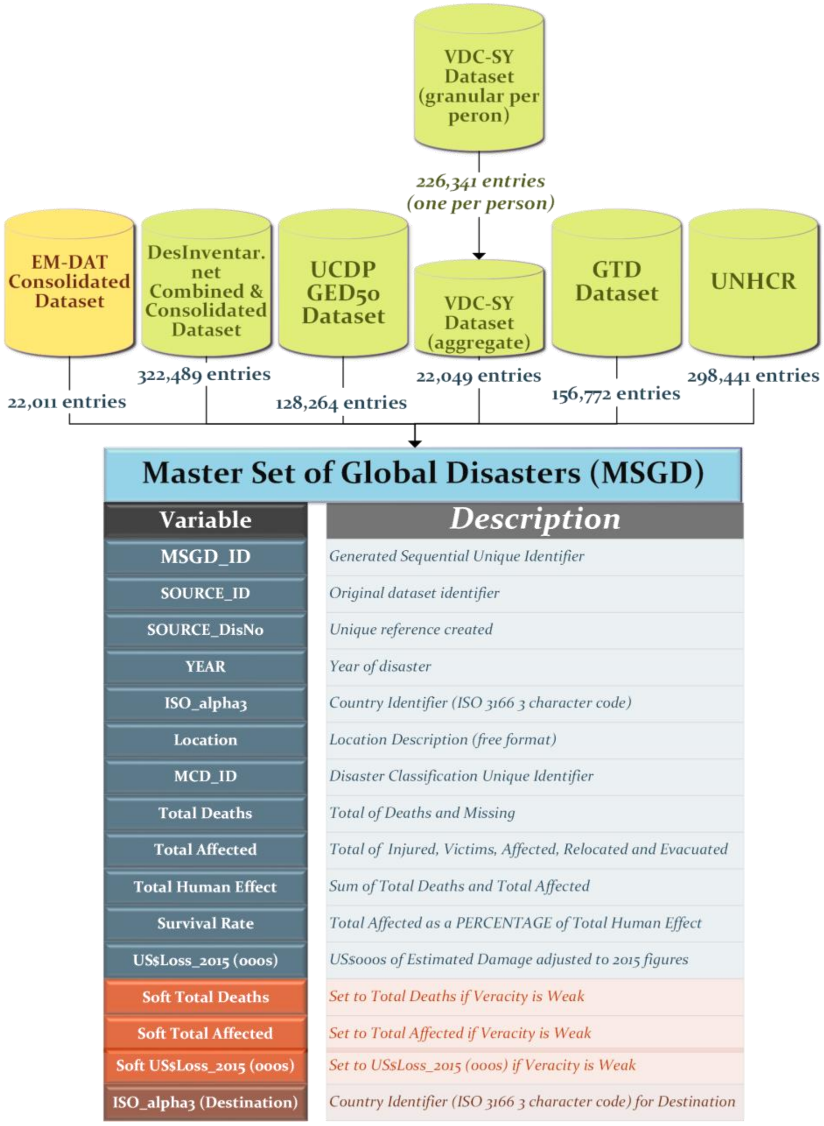


Figure 7-37: The Composition of the MSGD

This MSGD artefact set is also considered a core '*data scaffold*' of the humanitarian domain as without it there is no way of knowing the full scale of demand (needs) driving the humanitarian domain. The amalgamated dataset created here is to compensate for the absence of a credible comprehensive system of record for all disasters that have occurred in recent history. This is adequate for this study and is considered reusable by other studies, but is still a first instantiation of this artefact and therefore considered a prototype that would benefit from further development, refinement and scrutiny.

Data Veracity framework (DVf)

This includes the definition, model (DVm), and methods of evaluating data veracity (DVp and DVi) [*Figure 7-38*]. The DVf is considered a key knowledge contribution from this study plugging a conspicuous gap in useful toolsets when working with data.

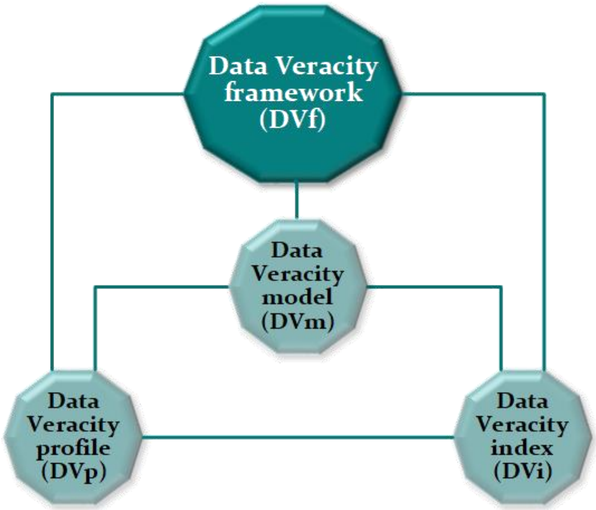


Figure 7-38: The Data Veracity framework (DVf)

For this study, utilising the DVf toolset facilitates discipline and consistency to the evaluation of data veracity of acquired datasets. It also enables the distinguishing of less veracious entries from the more reliable ones. The distinction of these entries, coined 'soft' and 'firm' respectively, adds value to the analysis when bivariate models of 'soft' entries are found to have no. or less, statistically significant fit when compared to models of their 'firm' counterparts. This phenomenon in turn also helps to validate the data veracity toolsets.

The DVf created here is in its first incarnation and created as a means to an end with a specific purpose in mind; therefore it is untested beyond the scope of this work. It is an embryonic product that would benefit from further research to help mature it and develop it for broader use. In particular, it is believed that further development of the DVf could include a context-setting component; a review of the dimensions of the DVm; improved, more refined, DVi algorithms; and detailed instructional documentation.

In summary, to maintain consistency the core artefacts created during this research are highlighted in this final mapping of created artefacts to the research framework [*Table 7-14*].

Research Framework			Research Activities			
			Design Science		Natural Science	
			Build (grow)	Evaluate	Theorise	Justify
Research Outputs	Artefacts	Constructs	[a][f]		[c]	
		Models	[d][g][h]			
		Methods	[e]			
		Instantiations	[b][i][j][k][l]			

[a]	**Macro-indicators of disaster outcome and the impact and effectiveness of humanitarian intervention (M^iOs, M^iIs and M^iEs).**
[b]	Data analysis outputs and visualisations
[c]	A (behavioural science) hypothesis relating the availability, or lack thereof, of humanitarian data and the flow of humanitarian aid that emerges from the domain knowledge and may be worthy of future research.
[d]	**Data Veracity framework (DVf) and Data Veracity model (DVm)**
[e]	**Data Veracity profile (DVp) and Data Veracity index (DVi)**
[f]	Expansion of the construct of '*data scaffolds*' for the humanitarian domain
[g]	Data structure of the Master Set of Global Disasters (MSGD)
[h]	**Classification structure of Master Disaster Classification Model (MDC)**
[i]	**Master Set of Global Disasters (MSGD) dataset**
[j]	**Master Disaster Classification (MDC) reference dataset**
[k]	Data Veracity profile (DVp) and Data Veracity index (DVi) instances for each of the six datasets amalgamated for the MSGD and the FTS and IDS aid datasets
[l]	**Mean Survival Rate by Year as an actualised M^iO** **Mean Survival Rate by Humanitarian Aid per Person as an actualised M^iI** **Mean Survival Rate by Population as an actualised M^iE**

Table 7-14: DSR Output to Research Framework Mapping v.5
(Vaishnavi and Kuechler, 2004b; Hevner, 2007; March and Smith, 1995)

The Utility Theory

	Solution Space	Utility	Problem Space
	Form	Function	Purpose
	Artefact [*What*]	Efficacy [*How*]	to Address [*Why*]
STATMENT	*Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector…*	*…when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level…*	*…the inability to gauge the consequences of monies spent and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.*

Table 7-15: The Utility Theory Statement

The macro-indicators identified support the utility theory of this research to an extent [*Table 7-15*]. The patterns of plots show that meaningful relationships exist that provide aerial views of changes in survival rate as time, humanitarian aid and population changes occur. Causal relationships, however, are not established through the analysis carried out here. A *Master Disaster Classification* (*MDC*) model and a *Master Set of Global Disasters* (*MSGD*) were created as prerequisite *data scaffolds*, without which the theory could not be tested. Similarly the MSGD would have been less meaningful if it had been developed without the use of the *Data Veracity framework* (*DVf*)

Finally, it is believed the artefacts designed and created by this research open the realms of possibilities and provide a launch point from which numerous trajectories of future research can embark, for example:

- The created '*data scaffolds*' of MSGD and MDC can be further developed to include more variables and more disasters, if any were missed. Processes and solutions can be put in place to maintain these centrally with global organisations having the ability to contribute, correct and use. This may in turn allow more macro-indicators and relationships between macro-indicators to be identified, thus further testing the utility theory of this study.

- The MSGD and MDC can be augmented and expanded using proprietary data and/or social media data to create not only more elaborate and detailed historical exploratory models but also more geographically specific predictive models to aid decision-making. These artefacts may even be useful in supporting research into the behavioural science hypothesis suggested earlier in this work [*Section 3.3.3*, *Table 3-5, [c]*], which posits that the collection of detailed supply/demand and financial data is absent from the flow of humanitarian supply networks (HSNs) because it is an ‘unnatural’ act, i.e. the data is not needed to secure the ongoing funds that fuel the HSNs.
- The DVf toolset is also ripe for further research and development with other broader types of data that can be used to gain insights, e.g. Big Data, transactional data, signal data etc. and data from a variety of subject areas. Thus enabling other design theories to emerge and be tested.
- There is also an opportunity to extend the DVf by carrying out research to develop complementary frameworks for other *aspirational qualities* of data (ref: Section 5.5.3), i.e. data virtue and data value, to create a usable suite of tools for data practitioners.
- Lastly, as mentioned earlier, the construct of ‘*data scaffolds*’ is worthy of additional research and testing in its applicability to other research domains and other forms of artefacts that are crucial to the study of data

7.6 Summary

This chapter discusses Iteration 4 of this work and constitutes the final iteration of the DSR design cycle. The chapter builds on the work of Iteration 3 by moving beyond disaster losses to explore aid funding and population figures to identify patterns of change in disaster outcomes that may yield macro-indicators of impact (M^iIs)

or macro-indicators of effect (M^{i}Es). Three main data sources are identified namely, UNOCHA's FTS for humanitarian aid data (FTS, 2017d); OECD's IDS for other international financial aid (IDS, 2017); and the World Banks Databank for population figures (World Bank: HN&P, 2017).

Limitations and anomalies of the aid data inhibit geographic analysis and restrict analysis to global annual levels. Additionally, not all population figures for years and countries in which disasters occurred are available from the World Bank, and alternative sources are used (World Bank: HN&P, 2017; UNCTAD, 2017; Worldometers, 2017; Population Pyramid, 2017). Despite these challenges analysis is carried out and a candidate M^{i}I of *mean survival rate by humanitarian aid per person* is identified and a M^{i}E *of mean survival rate by population* is also found.

Finally, the *Conclusion* step discusses the alignment of the design to the artefacts built in this iteration, before mapping the created artefacts to the DSR research framework. The knowledge and design theory contributions are outlined. The chapter closes with final reflections on the research and artefacts created.

Chapter 8: CONCLUSION

This final chapter of the thesis discusses to what extent the work has managed to meet its original objectives [*Section 8.1*]. It then goes on to summarise the core knowledge contribution of this research [*Section 8.2*] before discussing its limitations [*Section 8.3*] and the opportunities it presents for future research [*Section 8.4*] The chapter, and therefore the thesis, closes with concluding remarks [*Section 8.5*]. *Figure 8-1* is a simple schematic depicting the relationship between the main themes of this chapter.

Figure 8-1: Conclusion (Structure)

8.1 Research Objectives

To answer the overall research question of – *Can exploration of curated web-available data yield macro-indicators of humanitarian intervention in the aftermath of disasters?* – three objectives are set. Therefore in this concluding chapter it is important to first assess to what extent this research has met its objectives.

Objective 1:
Review current practices and research relevant to assessing the effect of humanitarian intervention.

This research objective is met in *Chapter 2*. The literature review detailed in this chapter highlights the increasing demands and diminishing resources faced by the humanitarian sector in fulfilling its role (HERR, 2011). *Chapter 2* goes on to describe the need to gauge the effect of humanitarian intervention in disasters and discusses existing efforts to obtain such measures as well as the limitations of these efforts. The study also acknowledges that some guidance as to the selection of measurable factors are gleaned from the only initiatives found that attempt to obtain objective quantifiable measures of progress, in this case by disaster risk reduction programmes (HYOGO, 2008; Wahlström, 2015).

Additionally, the knowledge gained through the literature review seeds a behavioural science hypothesis [*Section 3.3.3, Table 3-5, [c]*] that suggests the absence of crucial reconcilable data connecting the needs of disaster victims (demand) and the flow of aid (supply) is caused by the lack of a mandatory transactional link between victim need and donor funds. Therefore, data collection becomes a superimposed activity that is an 'unnatural act' that is left undone or inadequately done. This hypothesis is captured and mapped in the research framework for future reference, but not pursued as it is outside the scope of this work.

Objective 2:
Explore curated web-available data of global disaster losses, humanitarian aid and other factors extrinsic to the humanitarian sector for macro-indicators may signpost the consequences of humanitarian intervention.

This is achieved through Iteration 3 [*Chapter 6*] and Iteration 4 [*Chapter 7*] of the DSR design cycle [*Figure 8-2*].

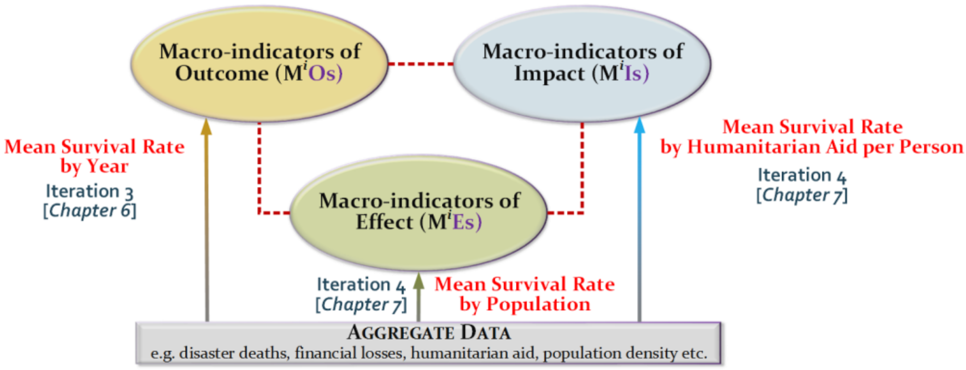


Figure 8-2: The Triad of Macro-indicators – Iteration/Chapter Map

Iteration 3 identifies a statistically significant relationship between mean survival rate and the years in which disasters occurred. Mean survival rates variations by year also exhibit statistically significant relationships at the more detail levels of disaster groups and geographic regions. The study therefore identifies ***mean survival rate by year*** as a macro-indicator of outcome (**M^{i}O**). M^{i}Os being a construct coined here to refer to any metrics that signposts change in the outcome of disasters over time. When embarking on the search for M^{i}Os, based on the data believed to be available, the factors considered feasible measures of the outcome of disasters are identified as either human effects or financial losses. Financial loss data acquired by this study is too sparse for analysis. As for human effect, no meaningful pattern of change can be identified using sums or means of deaths, people affected, or total human effect (*deaths + people affected*). Of note is that the mean of the calculated variable of survival rate (*people affected as a percentage of total human effect*) exhibits a statistically significant relationship that may signpost the effect of humanitarian intervention.

Iteration 4, builds on the identification of mean survival rate by year as a M^{i}O, to search for factors that may have a relationship with this macro-indicator. It is found that the global flow of humanitarian aid per person (i.e. annual totals divided by total human effect) exhibits a statistically meaningful relationship with mean survival rate per year. ***Mean survival rate by humanitarian aid per person*** is therefore deemed a viable **M^{i}I**. Notably, gaps and misalignments in humanitarian aid data renders investigating geographic patterns of change infeasible, and aggregate humanitarian aid values (i.e. annual totals) do not exhibit any useful pattern of change.

Iteration 4 goes on to search for M^{i}Es using other international financial funding data acquired from the OECD to test if any relationship can be found that indicates other forms of aid to countries may affect the outcome of disasters (OECD, 2017b; IDS, 2017). Notably, no relevant relationship can be identified between mean survival rate, the M^{i}O and other international flows of funds. Relationships are found to exist between *population* *__density__*, *__global__* *population size*, *__global urban__* *population size*, *__regional__* *population size* and mean survival rate. The patterns of these plots closely mimic the plot of mean survival rate by year. This is not surprising as populations increase as years increase. The pattern of the plots suggests that survival improves as populations grow and only drops after a peak is reached. This suggests that the asserted effect of increased population related to poorer disaster outcomes may not be a given (HERR, 2011). In any case, as statistical significance exists in all variations of population and mean survival rate bivariate models ***__mean survival rate by population__*** is identified as a **M^{i}E**.

__Objective 3:__
Evaluate the artefacts, theories and findings from this study in the context of the research domain, identifying knowledge contribution, research limitations and the potential for future research.

Evaluation of the design, developed artefacts and theories is a part of the DSR design cycle iteration [*Section 4.5*, *Section 5.5*, *Section 6.5* and *Section 7.5*]. Also included in the evaluation process is a harvesting of nuggets of knowledge as they emerge and mapping these to the consequences these have on the research. This is particularly important as artefacts and theories are developed that are not explicitly identified in the aim and objectives of the study, but are in fact the result of improved knowledge of the humanitarian and data science domains and the absence of needed resources. These additional artefacts and theories, as products of the unexpected groundwork needed to enable this research to continue, constitute a bonus yield of knowledge from this work. This yield, however, is not without cost, as more detailed and sophisticated data analytics in the search for macro-indicators are sacrificed in order to come to a 'satisficing' conclusion to this work. This acknowledgement of a compromise needed to complete this work provides a useful segue to the limitations of this research, and by association the opportunities for future research. These are touched upon in the *Evaluate* step of the iterations and expanded to some degree in *Section 7.5.5*.

8.2 Research Knowledge Contribution

The knowledge contribution of this research essentially takes three forms, the *developed artefacts*, the *tested utility theory* and the n*uggets of knowledge* that emerge from the work as harvested at the end of each iteration.

8.2.1 Developed Artefacts

To ensure completeness all developed artefacts are routinely mapped to the DSR Research framework, either in explicit or general terms, as part the *Evaluate* step of the iterations in which they are created [*Table 4-4*, *Table 5-2*, *Table 6-14* and *Table 7-11*]. Not all artefacts, however, are deemed to be of equal prominence. Therefore, the numerous data models and visualisations; the instantiations of DVps and DVis for each evaluated dataset; and the model of the MSGD dataset are not presented here as core knowledge-contributing artefacts. The artefacts considered to be 'core' contributions from this study are [*Figure 8-3*]:

- the constructs and instantiations of macro-indicators, as these address the overall research question that drives this work;
- the model and instantiation of the Master Disaster Classification (MDC) system, as it facilitates the search for macro-indicators;
- the instantiation of the Master Set of Global Disasters (MSGD), as it is crucial for the exploration for macro-indicators;
- the data veracity models and methods, as these are needed for consistent data veracity evaluation of all acquired datasets.

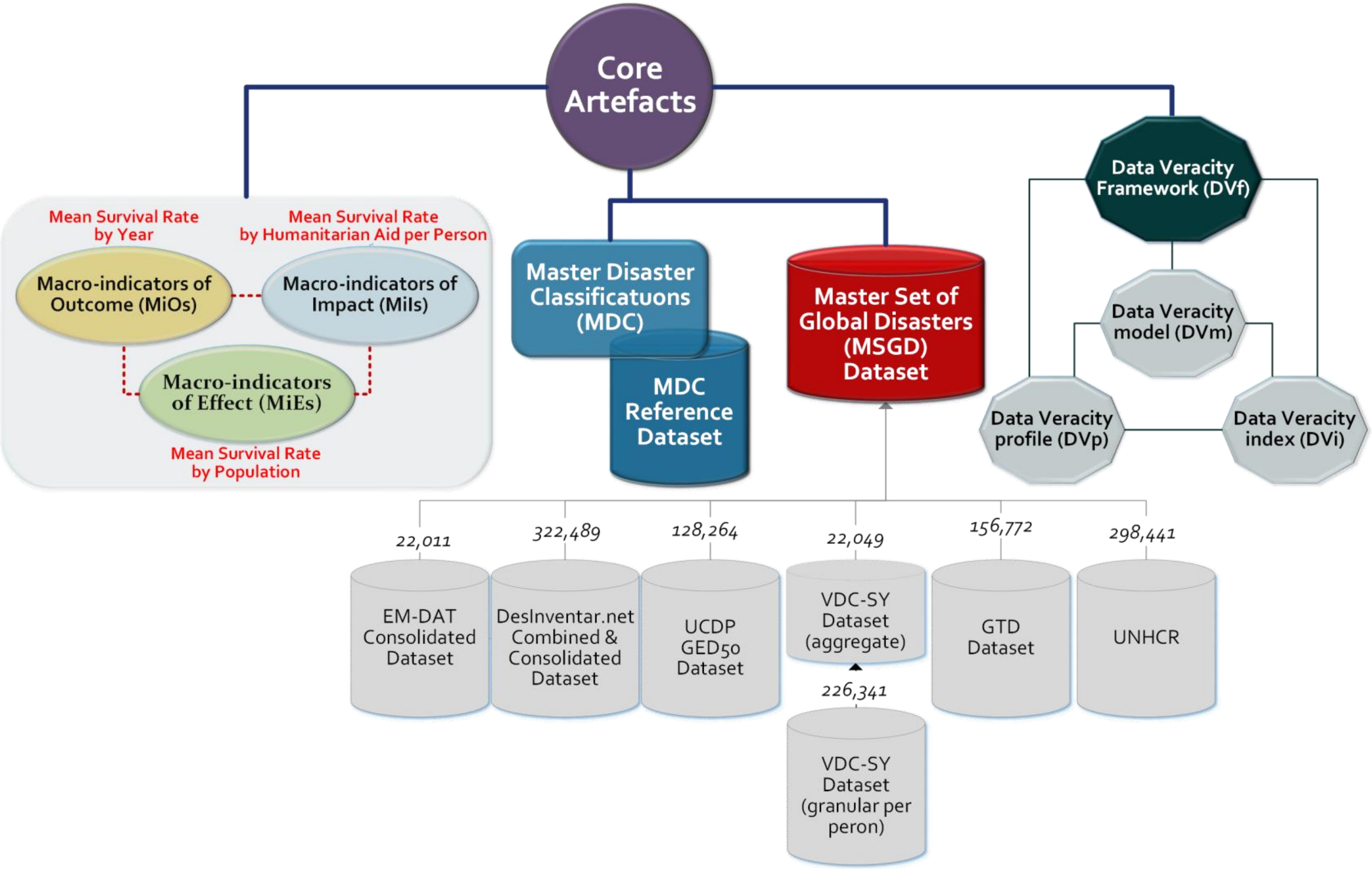


Figure 8-3: Core Developed Artefacts (Knowledge Contribution)

8.2.2 The Utility Theory

The utility theory presented and tested as part of this research [*Figure 8-4]* aligns with the research question – *Can exploration of curated web-available data yield macro-indicators of humanitarian intervention in the aftermath of disasters?*

Statement of Utility

Macro-indicators based on web-available curated data of disaster losses, humanitarian aid and relevant factors extrinsic to the humanitarian sector, when identified and explored properly, will offer an aerial perspective of the effect of humanitarian intervention, alleviating at an aggregate level, the inability to gauge the consequences of monies spent, and actions taken, to prevent, mitigate and ultimately respond to and recover from humanitarian crises.

Figure 8-4: Statement of the Utility Theory

The theory is tested by this study and is supported by the identification of *Mean Survival Rate by Year* as a *Macro-indicator of Outcome* ($M^{i}O$); *Mean Survival Rate by Humanitarian Aid per Person* as a *Macro-indicator of Impact* ($M^{i}I$); and *Mean Survival Rate by Population* as a *Macro-indicator of Effect* ($M^{i}E$).

*Figure C.*2- illustrates the full suite of dependency flows and relationships between the constructs of this study that constitute the utility theory.

- The macro-indicator constructs of $M^{i}O$ (*from Iteration 3*) and $M^{i}I$ and $M^{i}E$ (*from Iteration 4*) depend on the construct of *Data Scaffolds* (*from Iteration 3*), namely the artefacts of the MSGD and MDC.
- The macro-indicator constructs of $M^{i}I$ and $M^{i}E$ (*from Iteration 4*).are also directly dependent on the construct of the *Data Veracity framework* (*from Iteration 2*).
- The construct of *Data Scaffolds* (*from Iteration 3*), i.e. the artefacts of the MSGD and MDC, relies on the construct of the *Data Veracity framework* (*from Iteration 2*).

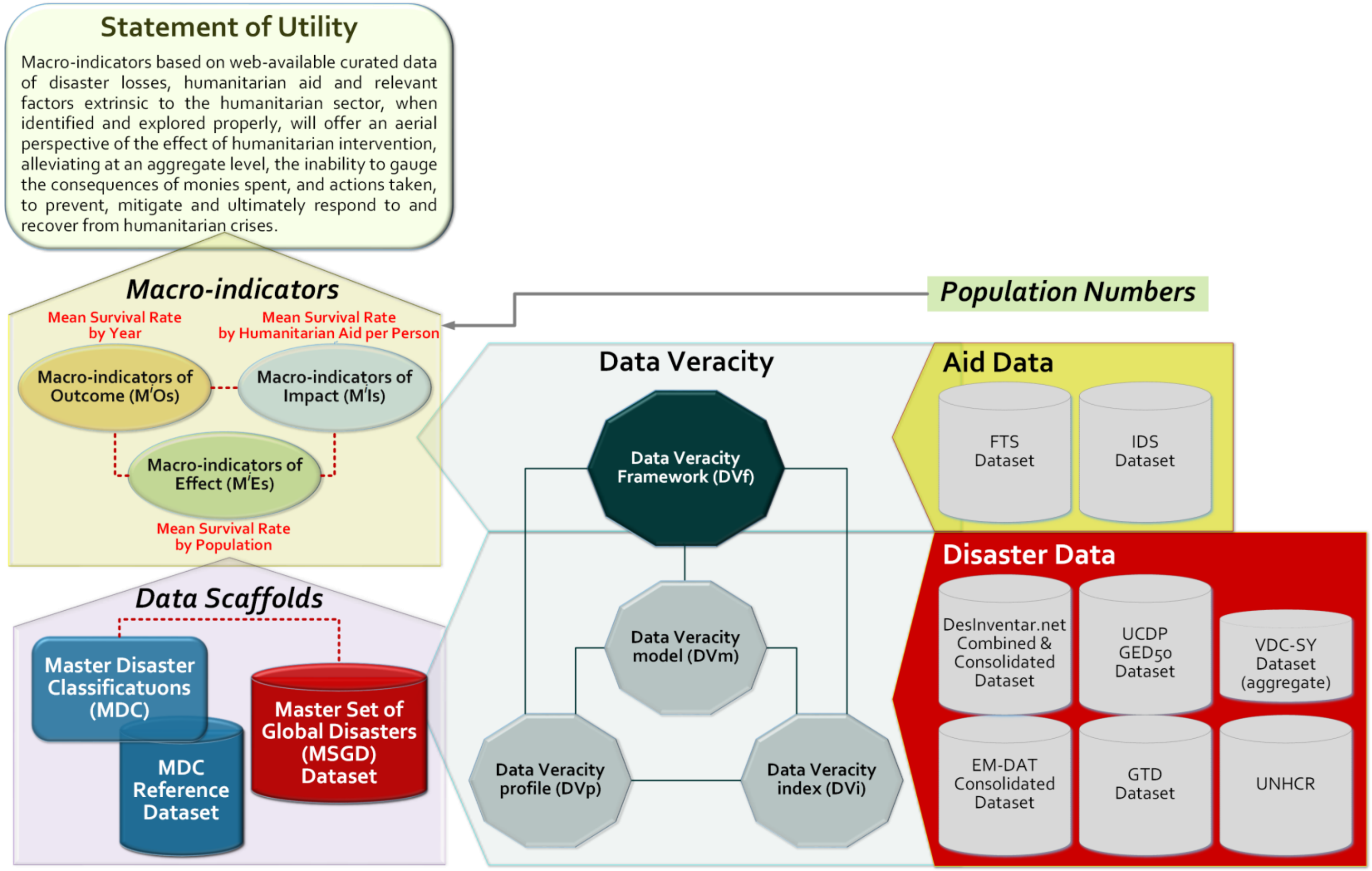


Figure 8-5: Structure of the Utility Theory

8.2.3 Knowledge '*Nuggets*'

During the *Evaluate* step of the iterations of the DSR design cycle effort is made to formally document 'nuggets' of knowledge that have emerged that not only have consequences on the path and shape of the research but also constitute new awareness of the problem space as a result of this work.

The knowledge gained from **Iteration 1 [*Chapter 4*]** includes an understanding that the there is no universally agreed definition of, or classifications for, '*disasters*' and that the previously assumed comprehensive and veracious source of disasters, the Emergency Events Database (EM-DAT), is neither as comprehensive nor as veracious as assumed (Guha-Sapir et al., 2017l). In fact, there is no single source of disaster loss data that includes all humanitarian crises eligible for international humanitarian aid funding; noting that not all humanitarian crises are given the moniker of '*disaster*'.

Furthermore, new knowledge of the limitations of EM-DAT is instrumental in shaping the remainder of the research and the structure and scope of the disaster loss data analysed. Over and above other issues of veracity – e.g. arbitrary estimation methods; the exclusion of a complete group of disasters; and missing or partial disaster dates – data before 1990 is sparse and financial loss data is inconsistently and incompletely populated. Consequently, more disaster data must be acquired and amalgamated and disaster loss analysis can only really focus on human effects and is limited to annualised values from 1990 onwards. It also creates a realisation that a credible source does not equate to reliable data; therefore data veracity must be evaluated and not assumed. Finally, calculation and modelling of the human effect variable of survival rate – calculated as people affected, i.e. surviving victims, expressed as a percentage of the total human effect (deaths + people affected) – may be worth investigating as a potential macro-indicator of outcome (M^iO).

Knowledge gained from **Iteration 2 [*Chapter 5*]** includes a realisation that there is no standard definition of data veracity, nor is there a model or methodology to equitably and consistently evaluate data veracity. Therefore, if this study is to proceed with a balanced understanding of the veracity of the datasets acquired, a means of evaluating data veracity must be developed. This becomes the focus of *Chapter 5*. Additionally, as offering potential for future research, it is noted that data veracity is not the only aspirational quality of data sans a definition and evaluation toolset. Other key aspirational qualities, such as value and virtue, lack a standard definition let alone any agreed mechanisms to evaluate data as having 'value' or 'virtue' (Laney, 2001, 2013; Normandeau, 2013; Grimes, 2013; Floridi and Taddeo, 2016; Boyd and Crawford, 2012; Glikman and Glad, 2015; Baldwin, 2015).

Over and above the identification of ***Mean Survival Rate by Year*** as a ***M^iO***, the knowledge gained from **Iteration 3 [*Chapter 6*]** includes the awareness that:

- all disaster datasets sourced have gaps in their data;
- disaster data related to conflict or deracination are likely to be more reliable than data for other disaster types;
- the majority of human effect data, across all disaster types, is found to be of weak veracity;
- weaker veracity data present more optimistic survival rates than firm data;
- weaker veracity data also results in weaker, or non-existent, statistical significance in most of the bivariate models created.

Finally, in addition to identifying ***Mean Survival Rate by Humanitarian Aid per Person*** as a ***M^iI*** and annual ***Mean Survival Rate by Population*** as a ***M^iE***, other knowledge is gained during **Iteration 4 [*Chapter 7*].** This includes the realisation that the documented flow of aid is not clearly aligned to disasters. In particular, there is little or no sharing of nomenclature to connect humanitarian crises to the flow of humanitarian aid. Furthermore, the dataset that is considered to be the system of record for humanitarian aid, FTS, has unexplained data gaps; with both FTS and IDS offering data at aggregations that obfuscate underlying flows and calculations (FTS, 2017h; IDS, 2017).

8.2.4 Data Science Domain Contributions

As a result of the path and evolution of this research specific artefacts are created that are conceptual contributions to the data science domain. While these artefacts have been referenced earlier in this chapter – *in context with other knowledge contributions by the study* – they are further listed below to make explicit their place within the data science domain:

- The data veracity ***models*** and ***methods***; and
- The ***construct*** of '*data scaffolds*'

Additionally, there is a dearth of literature regarding the use of DSR for data science research. Only one article was found and this was specific to Big Data research and the opportunities and challenges for design science (and other science) (Abbasi et al., 2016). As a result, it can be argued that this study also contributes to the data science domain as an example of the applicability of the DSR approach for data-centric research.

8.3 Research Limitations

The limitations of this work are discussed here. For the most part this research has adapted to obstacles encountered by reshaping the path of study and realigning efforts in order to accommodate the need for unexpected groundwork. This need to plug gaps and prepare toolsets as prerequisites to satisfying the original research question necessitated previously unplanned artefacts and utility theories to be created. The ad hoc creation of these material and abstract artefacts can also be considered to be at the root of their limitations, e.g.:

- The Data Veracity framework (DVf) is designed, built, and utilised specifically for and within the confines of this study; as such it lacks the honing, enrichment and robustness that comes from broader input and application.
- The Master Set of Global Disasters (MSGD) and Master Disaster Classification (MDC) system are pragmatically built as necessary resources, balancing analytical need and data availability. These are considered domain *'data scaffolds* that are prototypes, untested and unvalidated beyond the confines of this work, which would benefit from further in-depth study.

Finally, the necessitated detour to create the data resources to search for macro-indicators has limited the time and effort available to dedicate to carrying out more sophisticated data analytics, therefore the work has focussed on exploratory statistics and simple bivariate analysis to examine the data and identify basic macro-indicators. Of note is that no analysis is done for causal relationships and the identified M^iO, M^iI and M^iE exemplify the concept of a satisficing solution in conclusion of a DSR study.

8.4 Research Opportunities

While the limitations of this research presents opportunities for further research, there are also avenues of study that extend beyond simply improving what is achieved here. Examples of future research could include:

Humanitarian Domain

- The Master Disaster Classification (MDC) model can be:
    - Refined to create more congruous classifications across all disaster types; remove redundant entries; and add missed disaster types that have occurred or could occur.
    - Developed to include clear, detailed and quantified definitions and identification criteria to enable the accurate matching of events to disaster type.

  - Developed as an automated rule-based decision tool that guides practitioners to assign the correct MDC classification to each new event.
- The Master Set of Global Disasters (MSGD) can be:
  - Systematically checked at the detail level to identify which humanitarian crises 'events' coincide with the movement of UNHCR 'persons of concern' (UNHCR, 2017b).
  - Corrected, augmented and expanded working with one or more humanitarian aid agencies to add their proprietary data and develop additional data models and identify improved measurements of progress
  - Developed as a central repository that can be used, corrected or contributed to by global humanitarian actors.
  - Combined with social media data to create more detailed and elaborate historical exploratory models and geographically specific predictive models to assist in decision-making.
- Macro-indicators of Outcome/Impact/Effect (M^iOs/M^iIs/M^iEs)

  The analysis and model developed here are considered rudimentary. They are prototypes that serve to illustrate possibilities. It is believed that more sophisticated analysis can be carried out using the datasets created here, or improved versions of these datasets, to develop macro-indicators that can suggest causal relationships between factors. There is also a possibility of creating models that reflect relationships between macro-indicators or are able to help pinpoint best and worst outcomes that can help drive improvements in humanitarian intervention.
- HSNs and the '*unnatural act*' of data collection

  There is also the opportunity to examine and address the behavioural science hypothesis that emerged from the literature review [*Chapter 2*]. This hypothesis posits that the collection of detailed supply/demand and financial flow data is absent from humanitarian supply networks (HSNs) because it is an 'unnatural' act not needed to secure donor funds. As the lack of accurate and detailed data to show the flow of money, goods and services is a considerable hindrance to assessing the effectiveness of humanitarian interventions, finding ways to counteract the 'unnatural' aspect of data collection may be worth pursuing.

- Improving compatibility across humanitarian sector datasets
  by developing and maintaining foundational and domain-level ontologies for the humanitarian domain by building on initiatives already started or established (Liu et al., 2013; HDX, 2017).

- Building more '*data scaffolds*'
  Defining and seeding other datasets that could serve as data scaffolds in the humanitarian sector, e.g. a comprehensive registry of iNGOs/NGOs; a catalogue of international early warning systems; a list of non-profit and commercial suppliers of humanitarian goods and services.

- An International Humanitarian Market Hub
  There may be an opportunity to explore the possibility of creating an international market hub of voluntary, donated and commercial goods and services. This could allow unprecedented visibility and accessibility to all crucial and available services. It could also offer the potential to improve the visibility, cost and effect humanitarian interventions. The added benefit of this could be improved transparency and better supply/demand data.

- Process changes to make funding contingent on the supply of data
  There may be merit in researching the implications of making some or all future humanitarian funding given to actors contingent on the detail and veracity of the data they share.

Data Science Domain

- The DVf toolset developed here is considered an initial offering. It serves its purpose for this work but offers the potential to be developed and tested further:

- Across different types of data, e.g. Big Data, transactional data, signal data etc.;
- Across different subject areas;
- Through improved algorithms for the DVi;
- By the development of a method that allows the inclusion of context and relevance to the outputs created;
- With improved documentation and instructional guides;
- Through the refinement of DVf outputs, i.e. completed DVp and DVi, so that they can be used as metadata.

- The DVf toolset can also be expanded to be a toolset for data *veracity*, *value* and *virtue* DV^{3}f. Thereby creating a suite of definitions and toolsets that enable the assessment of each of these aspirational qualities of data and how they interact for datasets or within a subject area.
- The constructs of '*data scaffolds*' can be explored further, for example to identify other qualifying artefacts, or to investigate the effect on research domains when such artefacts are absent.

8.5 Concluding Remarks

This aim of this research was to explore web-available curated data from credible (official) sources for macro-indicators of humanitarian intervention. Unexpectedly this required the completion of fundamental groundwork before any data analysis could be conducted. A classification model (MDC) for most known types of humanitarian crises as well as a baseline master dataset of the human and financial effects of these humanitarian crises (MSGD) needed to be created. The creation of the MSGD was through the amalgamation of six 'disaster' dataset, which in turn necessitated a toolset to equitably and consistently evaluate the veracity of each sourced datasets (DVf).

Ultimately the proposed constructs of macro-indicators of outcome impact and effect (M^iO, M^iI and M^iE) are supported by the findings of this work, where *mean survival rate by year*, *mean survival rate by humanitarian aid per person* and *mean survival rate by population* are identified as M^iO, M^iI and M^iE respectively. Furthermore, an interesting relationship between the veracity of data and the level of statistical significance of bivariate plots emerges that appears to validate the data veracity evaluations carried out.

Finally, even though extensive effort has been made by this study to contribute new knowledge to the humanitarian and data science domains, this work is still considered a foundational offering. As such, it is hoped that one of the most significant contributions this research has made is to provide inspiration for future opportunities for further study.

GLOSSARY OF TERMS

AIT	Aid Transparency Index An independent measure of aid transparency among the world's major development agencies
ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action A global network of NGOs, UN agencies, members of the Red Cross/Crescent Movement, donors, academics and consultants learning to improve response to humanitarian crises.
APICS SCC	American Production and Inventory Control Society Supply Chain Council APICS Supply Chain Council is a non-profit organisation that maintains the Supply Chain Reference model (SCOR).
BLS	Bureau of Labor Statistics The Bureau of Labor Statistics of the U.S. Department of Labor and the source of USA-CPI data.
CBPF	Country-based pooled funds UNOCHA's mechanism to allow donors to pool their contributions into single, '*unearmarked*' funds to support local humanitarian efforts.
CERF	Central Emergency Response Fund UNOCHA's global emergency response fund since 2006 to deliver funding quickly to humanitarian responders and kick-start life-saving action whenever and wherever crisis hit.
CHS Alliance	Core Humanitarian Standard Alliance A network of organisations committed to improving humanitarian and development work through the application of standards.
CRED	Centre for Research on the Epidemiology of Disasters Active in the fields of international disaster and conflict health studies, with activities linking relief, rehabilitation and development, promoting research, training and technical expertise on humanitarian emergencies, particularly in public health and epidemiology.
CRISP-DM	Cross-Industry Standard Process for Data Mining A data mining process model that describes commonly used approaches that data mining experts use to tackle problems.

CSR-SY	Syrian Center for Statistics and Research An NGO providing advisory services and research studies relevant to the Syrian situation, including the collecting and analysis of data and the production of relevant reports.
DAC	Development Assistance Committee The committee of the OECD that deals with development co-operation matters. This currently constitutes 30 member countries based on the membership criteria - the existence of appropriate strategies, policies and institutional frameworks that ensure capacity to deliver a development co-operation programme; an accepted measure of effort; and the existence of a system of performance monitoring and evaluation. Note: While individual European Union countries are members, the European Union as an entity is also counted as one of the 30 DAC members. Furthermore not all OECD members are DAC members, as is the case with Chile, Estonia, Israël, Latvia, Mexico and Turkey.
DCD	Development Co-operation Directorate Supports the OECD Development Assistance Committee (DAC), to help set international principles and standards for development co-operation, and monitor how donors deliver on their commitments. Promotes coordinated, innovative international action to accelerate progress towards the Sustainable Development Goals (SDGs) in developing countries and improve their financing.
DCHRS	Damascus Center for Human Rights Studies An independent NGO established in 2005 focussed on human rights in Syria.
DfID	Department for International Development The UK Government's foreign aid department.
DRR	Disaster Risk Reduction Aims and actions to reduce the damage caused by natural hazards through prevention.
DSR	Design Science Research Uses design as research method. As research approach of "learning through the act of building".
DVf	Data Veracity framework A framework that can be used to evaluate data veracity. It consists of 3 parts: (1) a hierarchical model (DVm) of determining characteristics; (2) a profiling template (DVp) to capture information pertinent to each component of the hierarchical model; (3) a scoring and indexing (DVi) method for the components of the hierarchical model to provide a method to compare the relative veracity of different datasets.

DVi	Data Veracity index Part of the DVf. A scoring mechanism that enables the measurement of data veracity to create indices that can be used to compare the relative veracity afforded by different datasets.
DVm	Data Veracity model Part of the DVf. A hierarchical model of component attributes/characteristics of a dataset that can be used to establish the subjective veracity the said dataset.
DVp	Data Veracity profile Part of the DVf. A profile created by documenting the findings and associated actions taken when evaluating a the veracity of a dataset.
ECHO	The Directorate-General for European Civil Protection and Humanitarian Aid Operations – *formerly known as the* ***European Community Humanitarian Aid Office*** The European Commission's department for overseas humanitarian aid and for civil protection.
EDRIS	European Disaster Response Information System A database containing real-time information on ECHO and Member States' contributions to Humanitarian Aid.
EM-DAT	Emergency Events Database Launched in 1988 by the Centre for Research on the Epidemiology of Disasters (CRED) with the initial support of the World Health Organisation (WHO) and the Belgian Government. Currently fund by USAID
EU	The European Union An economic and political union between 28 European countries encompassing most of the European continent.
FEMA	Federal Emergency Management Agency The department that leads and coordinates the U.S. government's response to national disasters
FTS	Financial Tracking Service This service tracks humanitarian aid and is managed by the UN Office for the Coordination of Humanitarian Affairs (OCHA).
GAR	Global Assessment Report on Disaster Risk Reduction UNISDR biennial review and analysis of the natural hazards that affect humanity. Closely linked to the HFA and SFDRR.

GDP	Gross Domestic Product The monetary value of all the finished goods and services produced within a country's borders, typically calculated on an annual basis.
GED50	Georeferenced Event Dataset Global version 5.0 UCDP's 2015 version of their most disaggregated dataset, which covers individual events of organized violence, namely phenomena of lethal violence occurring at a given time and place.
GHAR	Global Humanitarian Assistance Report An annual report considered a resource for understanding humanitarian financing and related aid flows
GTD	Global Terrorism Database An open-source database of international terrorist events.
HAP	Humanitarian Accountability Partnership International (HAP International). A now dissolved organisation set-up as the humanitarian sector's first international self-regulatory body. It merged with People In Aid in 2015 to form the CHS Alliance, but now no longer exists
HAR	Humanitarian Accountability Report A periodic report published by the CHS Alliance
HDX	Humanitarian Data Exchange An open platform for sharing data, launched in July 2014. The goal of HDX is to make humanitarian data easy to find and use for analysis.
HFA	Hyogo Framework for Action The global blueprint for disaster risk reduction efforts between 2005 and 2015. Its successor is the Sendai Framework for Disaster Risk Reduction (SFDRR).
HRDAG	Human Rights Data Analysis Group A non-profit, non-partisan organization focussed on the analysis of world-wide human rights violations.
HSN	Humanitarian Supply Networks The entirety of the humanitarian response to a disaster – from securing funds, in-kind aid and specialist skills and services, to sourcing and delivering urgently needed goods and services to alleviate the suffering of the victims.
IATI	International Aid Transparency Initiative A voluntary, multi-stakeholder initiative seeking to improve the transparency of aid, development, and humanitarian resources in order to increase their effectiveness in tackling poverty.

ICRC	The International Red Cross and Red Crescent Movement An independent and neutral humanitarian organisation that operates worldwide helping people affected by conflict and armed violence and promoting the laws that protect victims of war. Established in 1863 and based in Geneva, the ICRC, its mandate stems essentially from the Geneva Conventions of 1949.
ICT	Information and Communication Technology (or Technologies) No formal universally accepted definition was found but generally accepted as a broad-brush term for all computing and communications *technologies* that enable the modern digital world.
IDS	International Development Statistics online databases These are databases managed by the OECD that cover bilateral, multilateral aid (ODA) and private providers' aid and other resource flows to developing countries.
IFRC	International Federation of Red Cross and Red Crescent Societies The world's largest humanitarian network. A movement constituting almost 100 million members, volunteers and supporters in 190 National Societies.
IGOs	Intergovernmental Organizations (IGOs) Groups such as the United Nations or the International Labour Organization.
iNGO	International Non-Governmental Organization Like an non-governmental organization (NGO), this is a not-for-profit organization that is independent from states and international governmental organizations, but it is international in scope and has outposts around the world to deal with specific issues in many countries.
IRC	International Rescue Committee A global humanitarian aid, relief, and development international non-governmental organization (iNGO).
IRDR	Integrated Research on Disaster Risk A global, multi-disciplinary research programme focussing on the challenges brought by natural disasters, mitigating their impacts, and improving related policy-making mechanisms. Co-sponsored by the International Council for Science (ICSU), the International Social Science Council (ISSC), and the United Nations Office for Disaster Risk Reduction (UNISDR).

IS	Information Systems A subset of ICT. This is the professional and academic discipline that focusses on strategic, managerial and operational activities of information flow and processing using complementary networks of hardware and software.
ISO 3166	International Organization for Standardization, Country Codes - ISO 3166 International coding standard for countries, dependent territories and special areas of geographical interest.
IT	Information Technology A subset of ICT. This is the professional and academic discipline that focusses on the application of computers to store, study, retrieve, transmit, and manipulate information. Often used as a synonym for computers and computer networks, but can encompass other converging technologies such as television and phones.
JHSPH	Johns Hopkins Bloomberg School of Public Health Founded in 1916, the Bloomberg School conducts advanced research, education and practice to create solutions to public health problems around the world.
KDD	Knowledge Discovery in Databases A data mining technique that includes data preparation and selection, data cleansing, incorporating prior knowledge on data sets and interpreting solutions from the observed results.
LOG	Logistics Operational Guide An on-line collection of information such as best practices, templates, guidelines and standard operating procedures for logisticians operating in the field. The content is based on manuals and documents from over 28 humanitarian organisations, academia and the private sector.
MDC	Master Disaster Classifications The name given in this study to the disaster classification system created in iteration 2 of the DSR Design Cycle to support the MSGD. Both MDC and MSGD are *data scaffolds* in the humanitarian data ecosystem.
M^{i}E (M^{i}Es if plural)	Macro-indicator of Effectiveness The bringing together of macro-indicators of outcome (M^{i}Os) and macro-indicators of impact (M^{i}Is).

M^{i}I (M^{i}Is if plural)	Macro-indicator of Impact Based on guidance from "*Indicators of Progress: Guidance on Measuring the Reduction of Disaster Risks and the Implementation of the Hyogo Framework for Action*". Defined here as indicators of the aggregate impact of aid taking into consideration societal measures, e.g. population growth/density, urbanisation, etc.
M^{i}O (M^{i}Os if plural)	Macro-indicator of Outcome Based on guidance from "*Indicators of Progress: Guidance on Measuring the Reduction of Disaster Risks and the Implementation of the Hyogo Framework for Action*". Defined here as indicators of the aggregate outcome of disasters as relevant to those affected, e.g. survival rate, financial to human loss ratio, etc.
MPTF	Multi-Partner Trust Fund Office A UN center of expertise on pooled financing mechanisms.
MSGD	Master Set of Global Disasters The name given in this study to the disaster loss dataset built through iteration 1 and 2 of the DSR Design Cycle. This study creates a basic MSGD as a crucial *data scaffold* in the humanitarian data ecosystem.
NGO	Non-Governmental Organization A not-for-profit organization that is independent from states and international governmental organizations.
ODA	Official Development Aid (OECD) A term used by DAC as "*those flows to countries and territories on the DAC List of ODA Recipients and to multilateral institutions which are:* **i.** *provided by official agencies, including state and local governments, or by their executive agencies; and* **ii.** *each transaction of which:* **(a)** *is administered with the promotion of the economic development and welfare of developing countries as its main objective; and* **(b)** *is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 per cent).*"
OECD	Organisation for Economic Co-operation and Development An organisation dedicated to economic development. It constitutes 35 member countries. It originated in 1960 with 18 European countries together with United States and Canada.
OFDA	Office of U.S. Foreign Disaster Assistance The department that leads and coordinates the U.S. government's response to disasters overseas.

OOF	Other Official Flows (OECD) "*Other official flows (OOF) are defined as official sector transactions that do not meet official development assistance (ODA) criteria.*"
PoC	Persons/Populations of Concern These are individuals or groups of individuals that are of concern to UNHCR based on their status (refugees, asylum seekers, internally displaced persons, etc.).
SCM	Supply Chain Management The management of the flow and transformation of products and services from source to customer.
SCOR	Supply Chain Operations Reference Model The most widely accepted framework for evaluating and comparing supply chain activities and performance.
SEMMA	Sample, Explore, Modify, Model, Assess A list of sequential steps developed by the SAS Institute to guide the implementation of data mining applications.
SFDRR	Sendai Framework for Disaster Risk Reduction The successor instrument to the Hyogo Framework for Action (HFA)
SIPRI	Stockholm International Peace Research Institute (SIPRI) Established in 1966, SIPRI's focus is on the research of conflict, armaments, arms control and disarmament.
sn4hr	Syrian Networks for Human Rights An independent NGO focussed on defending human rights in Syria and documenting human rights violations.
SOHR	Syrian Observatory for Human Rights Established in 2006, a body that documents and reports the human rights violations that occur in Syria.
START	The National Consortium for the Study of Terrorism and Responses to Terrorism A university-based research and education center for the causes and human consequences of terrorism in the United States and around the world.
UNCDATstat	United Nations Conference on Trade and Development Statistics Data source for ready-to-use analytical data for countries and products – particularly developing and transition economies.

UCDP	Uppsala Conflict Data Program Established in the mid-1980s, this programme collects information on a large number of aspects of armed violence and is managed by Department of Peace and Conflict Research, Uppsala University, Sweden.
UN	United Nations An international organization of 193 Member. Founded in 1945 with a view to maintaining international peace and security, developing friendly relations among nations and promoting social progress, better living standards and human rights.
UNDP	United Nations Development Programme The United Nations' office for the advance human development. Also the source of developing regions' classifications.
UNESCO	United Nations Educational, Scientific, and Cultural Organization Responsible for coordinating international cooperation in education, science, culture and communication.
UNHCR	United Nations High Commissioner for Refugees The UN Refugee Agency.
UNISDR	United Nations Office for Disaster Risk Reduction Mandated by United Nations General Assembly Resolutions. Most notably "*to serve as the focal point in the United Nations system for the coordination of disaster reduction and to ensure synergies among the disaster reduction activities of the United Nations system and regional organizations and activities in socio-economic and humanitarian fields*".
UNOCHA	United Nations Office for the Coordination of Humanitarian Affairs OCHA is part of the United Nations Secretariat and is responsible for bringing together humanitarian actors to ensure a coherent response to emergencies and a framework within which each actor can contribute to the overall response effort.
USA-CPI	USA Consumer Price Index USA's Consumer Price Indexes (CPI) for monthly changes in the prices paid by urban consumers for a representative basket of goods and services.
USAID	US AID *(agency)* The US Government's foreign aid agency.
USD or US$	USA Dollars United States' currency

VDC-SY	Violations Documentation Center in Syria A NGO that monitors and documents violations in human rights in Syria.
WDI	World Development Indicators These are the primary World Bank collection of development indicators. Compiled from officially recognized international sources. Considered to present the most current and accurate global development data available, and includes national, regional and global estimates.
WHO	World Health Organisation Directs and coordinates international health within the United Nations' system.
xSCM	Extended Supply Chain Management Extends the boundary of the supply chain beyond its primary relationships to incorporate the ecosystem that reaches from customer demand to all tiers of suppliers and contributors that serve to meet that demand.

References

Abbasi, A., Sarker, S. & Chiang, R. H. L. 2016. Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association of Information Systems,* Vol:17, pp.1-32.

Alcántara-Ayala, I. 2002. Geomorphology, Natural Hazards, Vulnerability and Prevention of Natural Disasters in Developing Countries. *Geomorphology,* Vol:47, pp.107-124.

Ali, A., Adnan, H., Nathan, K., Kuan Yew, W. & Sander De, L. 2017. Toward a Dynamic Balanced Scorecard Model for Humanitarian Relief Organizations' Performance Management. *Journal of Humanitarian Logistics and Supply Chain Management,* Vol:7, pp.194-218.

Ayankoya, K., Calitz, A. & Greyling, J. 2014. Intrinsic Relations between Data Science, Big Data, Business Analytics and Datafication. *Proceedings of the Southern African Institute for Computer Scientist and Information Technologists Annual Conference 2014 on SAICSIT 2014 Empowered by Technology.* Centurion, South Africa: ACM.

Baldwin, H. 2015. *Drilling into the Value of Data* [Online]. Forbes. Available:http://www.forbes.com/sites/howardbaldwin/2015/03/23/drilling-into-the-value-of-data/#4c4d177365fa [Accessed:2016].

Baskerville, R. & Pries-Heje, J. 2010. Explanatory Design Theory. *Business & Information Systems Engineering,* Vol:2, pp.271-282.

Beamon, B. M. & Balcik, B. 2008. Performance Measurement in Humanitarian Relief Chains. *International Journal of Public Sector Management,* Vol:21, pp.4-25.

Belanger, J., Sharrock, D., Araki, H., Asfaw, T., Matern, D., Papaioannou, E., Pilegaard, L. & Subki, S. K. A. 2016. Too Important to Fail—Addressing the Humanitarian Financing Gap. *World Humanitarian Summit.* Available:http://interagencystandingcommittee.org/system/files/hlp_report_too_important_to_failgcoaddressing_the_humanitarian_financing_gap.pdf

Below, R., Vos, F. & Guha-Sapir, D. 2010. Moving Towards Harmonization of Disaster Data (a Study of Six Asian Databases). Brussels, Belgium. *Centre for Research on the Epidemiology of Disasters (CRED).* Available:http://www.cred.be/sites/default/files/WP272.pdf

Below, R., Wirtz, A. & Guha-Sapir, D. 2009. Disaster Category Classification and Peril Terminology for Operational Purposes. *common accord CRED and MunichRe.* Available:http://cred.be/sites/default/files/DisCatClass_264.pdf

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V. & Szalay, A. S. 2018. Realizing the Potential of Data Science. *Commun. ACM,* Vol:61, pp.67-72.

Berti-Equille, L. & Borge-Holthoefer, J. 2015. Veracity of Data: From Truth Discovery Computation Algorithms to Models of Misinformation Dynamics. *Synthesis Lectures on Data Management,* Vol:7, pp.1-155.

Berti-Equille, L. & Lamine Ba, M. 2016. Veracity of Big Data: Challenges of Cross-Modal Truth Discovery. *Journal of Data and Information Quality (JDIQ),* Vol:7, pp.1-3.

Blaikie, P., Cannon, T., Davis, I. & Wisner, B. 2014. At Risk: Natural Hazards, People's Vulnerability and Disasters. *Routledge*. Available:http://www.microdis-eu.be/sites/default/files/D4.4.5%20-%20Combined%20Literature%20Review.pdf

Blei, D. M. & Smyth, P. 2017. Science and Data Science. *Proceedings of the National Academy of Sciences of the United States of America,* Vol:114, pp.8689-8692.

Bloor, R. 2013. *A Data Science Rant* [Online]. The Bloor Group. Available:http://insideanalysis.com/2013/08/a-data-science-rant/ [Accessed:2018].

Bogoshi, J., Naidoo, K. & Webb, J. 1987. The Oldest Mathematical Artefact. *The Mathematical Gazette,* Vol:71, pp.294.

Boucher, D. 2012. *Would a Large-Scale Reforestation Effort Help Counter the Global Warming Impacts of Deforestation?* [Online]. Available:http://www.ucsusa.org/publications/ask/2012/reforestation.html#.WwclaCAh3Dc [Accessed:2018].

Boyd, D. & Crawford, K. 2012. Critical Questions for Big Data. *Information, Communication & Society,* Vol:15, pp.662-679.

Buchanan-Smith, M. & Fabbri, P. 2005. Linking Relief, Rehabilitation and Development. A Review of the Debate. Available:http://www.alnap.org/pool/files/lrrd-review-debate.pdf

Cairo, A. 2012. *The Functional Art: An Introduction to Information Graphics and Visualization*, New Riders. California, USA.

Cao, L. 2017. Data Science: A Comprehensive Overview. *ACM Comput. Surv.,* Vol:50, pp.1-42.

Castrounis, A. 2017. *What Is Data Science, and What Does a Data Scientist Do?* [Online]. Available:http://www.kdnuggets.com/2017/03/data-science-data-scientist-do.html [Accessed:2018].

Chambre, A. 2017. *Priti Patel: Britain Could Slash UN Aid Funding If Alliance Fails to Reform* [Online]. Available:http://www.politicshome.com/news/uk/foreign-affairs/news/89131/priti-patel-britain-could-slash-un-aid-funding-if-alliance-fails [Accessed:2017].

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. 2000. *CRISP-DM 1.0 Step-by-Step Data Mining Guide* [Online]. SPSS. Available:http://www.the-modeling-agency.com/crisp-dm.pdf [Accessed:2018].

Cheung, K. & Chung, T. 2014. Master Data Management Fundamental and Integrated Solution. *Proceedings of 24th Annual International Conference on Computer Science and Software Engineering*. Markham, Ontario, Canada: IBM Corp.

Chonghaile, C. N. 2015. Red Cross Says Conflict in Syria and Iraq Causing 'Alarming' $82m Hole in Finances. *The Guardian*, 30 September 2015. Available:http://www.theguardian.com/global-development/2015/sep/30/red-cross-says-conflict-in-syria-and-iraq-causing-alarming-82m-hole-in-finances

Clarke, R. 2016. Big Data, Big Risks. *Information Systems Journal,* Vol:26, pp.77-90.

Claverie-Berge, I. 2012. *Solutions Big Data* [Online]. IBM. Available:http://www-05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf [Accessed:2017].

Coburn, A. W., Bowman, G., Ruffle, S. J., Foulser-Piggott, R., Ralph, D. & Tuveson, M. 2014. A Taxonomy of Threats for Complex Risk Management. Cambridge, United Kingdom. *University of Cambridge*. Available:http://cambridgeriskframework.com/page/20

Conklin, J. 2001 *Wicked Problems and Social Complexity* [Online]. CogNexus Institute. Available:http://cognexus.org/wpf/wickedproblems.pdf [Accessed:2016].

Corey, L., Pedram, R. & Navin, R. 2016. Influence of Extreme Weather Disasters on Global Crop Production. *Nature,* Vol:529, pp.84.

Costa, C. & Santos, M. Y. 2017. The Data Scientist Profile and Its Representativeness in the European E-Competence Framework and the Skills Framework for the Information Age. *International Journal of Information Management,* Vol:37, pp.726-734.

Cozzolino, A. 2012. Humanitarian Logistics and Supply Chain Management. *In: Humanitarian Logistics*. Ch.2. pp.5-16. Springer-Verlag Berlin Heidelberg.

Croicu, M. & Sundberg, R. 2015. UCDP GED Codebook Version 5.0. Sweden. *Department of Peace and Conflict Research, Uppsala University*. Available:http://ucdp.uu.se/downloads/ged/ucdp-ged-50-codebook.pdf

Cukier, K. & Mayer-Schoenberger, V. 2013. The Rise of Big Data: How It's Changing the Way We Think About the World. *Foreign Affairs,* Vol:92, pp.28-40.

D'mello, N. 2016. 5 Cases Where Big Data Was a Big Flop. *CIO*, 14th July 2016 Available:www.cio.in/feature/5-cases-where-big-data-was-big-flop

Dabelstein, N. 1999. Guidance for Evaluating Humanitarian Assistance in Complex Emergencies. *OECD*. Available:http://www.oecd.org/dac/evaluation/2667294.pdf

Darcy, J. & Hofmann, C. A. 2003. According to Need? Needs Assessment and Decision-Making in the Humanitarian Sector. London, United Kingdom. *Overseas Development Institute*. Available:http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/285.pdf

De Groeve, T., Poljansek, K. & Ehrlich, D. 2013. Recording Disaster Losses, Recommendations for a European Approach. *European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen*. Available:http://ipsc.jrc.ec.europa.eu/; http://www.jrc.ec.europa.eu/

De Groeve, T., Poljansek, K., Ehrlich, D. & Corbane, C. 2014. Current Status and Best Practices for Disaster Loss Data Recording in EU Member States. Luxembourg. Available:http://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/current-status-and-best-practices-disaster-loss-data-recording-eu-member-states

De Waal, A. 2005. *Famine That Kills: Darfur, Sudan (Revised Edition)* Oxford University Press. New York, USA.

Duffield, M. 1994. Complex Emergencies and the Crisis of Developmentalism. *IDS Bulletin,* Vol:25, pp.37-45.

Ergun, Ö., Keskinocak, P., Swann, J., Stamm, J. L. H. & Villarreal, M. 2010. *How to Improve Humanitarian Logistics* [Online]. Institute for Operations Research and the Management Sciences. Available:http://www.analytics-magazine.org/january-february-2010/168-relief-operations-how-to-improve-humanitarian-logistics [Accessed:2012].

Everett, J. & Friesen, C. 2010. Humanitarian Accountability and Performance in the Théâtre De L'absurde. *Critical Perspectives on Accounting,* Vol:21, pp.468-485.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996. The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Commun. ACM,* Vol:39, pp.27-34.

Ferrari, B. 2015. *"Extended Supply Chain" Is the New Supply Chain* [Online]. . Available:http://www.theferrarigroup.com/supply-chain-matters/2015/01/21/extended-supply-chain-is-the-new-supply-chain/ [Accessed:2017].

Ferris, E. 2007. The Global Humanitarian Platform: Opportunity for NGOs? *Forced Migration Review,* Vol:29, pp.6-8.

Few, S. 2013. *Information Dashboard Design: Displaying Data for at-a-Glance Monitoring*, 2nd ed. Analytics Press. California, USA.

Fleming, M. 2015. The Appeal for Humanitarian Funding Is Urgent as Syrians Continue to Flee. *The Guardian*, 7 December 2015. Available:http://www.theguardian.com/global-development-professionals-network/2015/dec/07/syria-refugees-urgent-appeal-humanitarian-funding

Floridi, L. & Taddeo, M. 2016. What Is Data Ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* Vol:374, pp.1-5.

Gates, S., Nygård, H. M., Strand, H. & Urdal, H. 2016. Trends in Armed Conflict, 1946–2014. Oslo, Norway. *Peace Research Institute Oslo (PRIO)*. Available:http://www.prio.org/Publications/Publication/?x=8937

Gleditsch, N. P., Nordås, R. & Urdal, H. 2017. Peer Review and Replication Data: Best Practice from Journal of Peace Research. *College & Research Libraries,* Vol:78, pp.267-271.

Glikman, P. & Glad, N. 2015. What's the Value of Your Data? , 27 July 2017. Available:http://techcrunch.com/2015/10/13/whats-the-value-of-your-data/

Gortney, W. E. 2016. Department of Defense Dictionary of Military and Associated Terms. *Joint Chiefs of Staff Washington United States*. Available:http://fas.org/irp/doddir/dod/jp1_02.pdf

Goteti, G. 2016. *Cran.R-Project, Em-Dat Disaster Dataset, Cran.R-Project* [Online]. Available:cran.r-project.org/src/contrib/Archive/emdatr/ [Accessed:2016].

Granville, V. 2016. *40 Techniques Used by Data Scientists* [Online]. Available:http://www.datasciencecentral.com/profiles/blogs/40-techniques-used-by-data-scientists [Accessed:2018].

Granville, V. 2017. *Difference between Machine Learning, Data Science, Ai, Deep Learning, and Statistics* [Online]. Available:http://www.datasciencecentral.com/profiles/blogs/difference-between-machine-learning-data-science-ai-deep-learning [Accessed:2018].

Gregor, S. 2006. The Nature of Theory in Information Systems. *MIS Quarterly,* Vol:30, pp.611-642.

Gregor, S. & Hevner, A. R. 2013. Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly,* Vol:37, pp.337-355 (Appendices A1-A6).

Gregor, S. & Jones, D. 2007. The Anatomy of a Design Theory. *Journal of the Association for Information Systems,* Vol:8, pp.312-323,325-335.

Grimes, S. 2013. *Big Data: Avoid 'Wanna V' Confusion* [Online]. Information Week. Available:http://www.informationweek.com/big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077 [Accessed:2017].

Guha-Sapir, D. & Below, R. 2002. The Quality and Accuracy of Disaster Data: A Comparative Analyse of 3 Global Data Sets. *Disaster Management Facility, World Bank*. Available:http://www.cred.be/node/288

Guha-Sapir, D., Below, R. & Hoyois, P. 2017a. *Emergency Events Database (Em-Dat), About* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://www.emdat.be/about [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017b. *Emergency Events Database (Em-Dat), Country Entry* [Online]. EM-DAT. Available:http://emdat.be/country-entry [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017c. *Emergency Events Database (Em-Dat), Disaster List* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be/disaster_list/index.html [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017d. *Emergency Events Database (Em-Dat), Disaster Trends* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be/disaster_trends/index.html [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017e. *Emergency Events Database (Em-Dat), Explanatory Notes* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be/explanatory-notes [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017f. *Emergency Events Database (Em-Dat), Frequently Asked Questions* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be/frequently-asked-questions [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017g. *Emergency Events Database (Em-Dat), General Classification* [Online]. EM-DAT. Available:http://www.emdat.be/classification [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017h. *Emergency Events Database (Em-Dat), Glossary* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be/Glossary [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017i. *Emergency Events Database (Em-Dat), Guidelines* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be/guidelines [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017j. *Emergency Events Database (Em-Dat), Publications* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://www.emdat.be/Publications [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017k. *Emergency Events Database (Em-Dat), Sources* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://www.emdat.be/source-entry [Accessed:2017].

Guha-Sapir, D., Below, R. & Hoyois, P. 2017l. *Emergency Events Database (Em-Dat): The CRED/Ofda International Disaster Database* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:http://emdat.be [Accessed:2017].

Guha-Sapir, D. & Hoyois, P. 2012. Measuring the Human and Economic Impact of Disasters. *CRED, Government Office of Science*. Available:http://www.cred.be/node/1310

Head, B. & Alford, J. 2008 Wicked Problems: The Implications for Public Management. Panel on Public Management, 26-28 March Brisbane, Australia. International Research Society for Public Management.

Hella, A., Sander De, L. & Matthias, K. 2014. Humanitarian Supply Chain Performance Management: A Systematic Literature Review. *Supply Chain Management: An International Journal,* Vol:19**,** pp.592-608.

Hevner, A. R. 2007. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems,* Vol:19**,** pp.87-92.

Hevner, A. R. & Chatterjee, S. 2010. Design Science Research in Information Systems. *In: Design Research in Information Systems: Theory and Practice.* Ch.2. pp.9-22. New York, USA.

Hevner, A. R., March, S. T., Park, J. & Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly,* Vol:28, pp.75-105.

Hlava, M. 2012. *What Is a Taxonomy?* [Online]. KMWorld. Available:http://www.kmworld.com/Articles/Editorial/What-Is-.../What-is-a-Taxonomy-81159.aspx [Accessed:2017].

Hooker, J. N. 2004. Is Design Theory Possible? *JITTA: Journal of Information Technology Theory and Application,* Vol:6, pp.73-82.

Hoven, J. V. D. 2003. Data Architecture: Principles for Data. *Information Systems Management,* Vol:20, pp.93-96.

Huberty, M. 2015. Awaiting the Second Big Data Revolution: From Digital Noise to Value Creation. *Journal of Industry, Competition and Trade,* Vol:15, pp.35-47.

Ibrahim Mohamed, S. 2007. Disaster Types. *Disaster Prevention and Management,* Vol:16, pp.704-717.

Janssen, M., Van Der Voort, H. & Wahyudi, A. 2017. Factors Influencing Big Data Decision-Making Quality. *Journal of Business Research,* Vol:70, pp.338-345.

John, L. & Thangamani, G. 2015. A New Classification Schema for Disasters. *The Second World Congress on Disaster Management (2nd WCDM).* Visakhapatnam, Andhra Prdesh, India.

Kourosh, E. & Richard, C. L. 2008. Disasters: Lessons from the Past 105 Years. *Disaster Prevention and Management,* Vol:17, pp.62-82.

Kuechler, B. & Vaishnavi, V. 2008. On Theory Development in Design Science Research: Anatomy of a Research Project. *European Journal of Information Systems,* Vol:17, pp.489-504.

Laney, D. 2001. 3d Data Management: Controlling Data Volume, Velocity, and Variety. *Meta Group*. Available:http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Laney, D. 2012. *Deja VVVu: Others Claiming Gartner's Construct for Big Data* [Online]. Gartner. Available:http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/ [Accessed:2017].

Laney, D. 2013. *Batman on Big Data* [Online]. Gartner. Available:http://blogs.gartner.com/doug-laney/batman-on-big-data/ [Accessed:2017].

Lattimer, C., Swithern, S., Sparks, D., Tuchel, L., Beecher, J., Cecilia Caio, Collins, R., Dalrymple, S., Galinié, A., Kenei, S., Knox, D., Miller, A., Shaikh, L., Simenon, M., Parrish, C., Tew, R. & Watts, R. 2016. Global Humanitarian Assistance Report 2016. *Global Humanitarian Assistance (GHA).* Available:www.globalhumanitarianassistance.org/wp.../2016/07/GHA-report-2016-full-report.pdf

Lavalle, S., Lesser, E., Shockley, R., Hopkins, M. S. & Kruschwitz, N. 2011. Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Management Review,* Vol:52, pp.21-32.

Lee, Y. W., Strong, D. M., Kahn, B. K. & Wang, R. Y. 2002. AIMQ: A Methodology for Information Quality Assessment. *Information Management,* Vol:40, pp.133-146.

Lemire, S. & Lefebvre-Naré, F. 2017. What Is Big Data? *In: Cyber Society, Big Data, and Evaluation.* Ch.2. pp.27-42. Routledge. New Brunswick, USA.

Lerner, M. 2016. Alternative Classification Schemes for Man-Made Hazards in the Context of the Implementation of the Sendai Framework. *United Nations Environment Programme (UNEP), Office for the Coordination of Humanitarian Affairs (OCHA), Joint Environment Unit (JEU).* Available:http://goo.gl/PFA5lw

Liu, S., Brewster, C. & Shaw, D. 2013 Ontologies for Crisis Management: A Review of State of the Art in Ontology Design and Usability. Proceedings of the 10th International ISCRAM Conference. pp.349-359.

Lukoianova, T. & Rubin, V. L. 2014. Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances in Classification Research Online; 24th ASIS SIG/CR Classification Research Workshop,* Vol:24, pp.4-15.

Lycett, M. 2013. 'Datafication': Making Sense of (Big) Data in a Complex World. *European Journal of Information Systems,* Vol:22, pp.381–386.

Maiers, C., Reynolds, M. & Haselkorn, M. 2005 Challenges to Effective Information and Communication Systems in Humanitarian Relief Organizations. Professional Communication Conference, 2005. IPCC 2005. Proceedings. International, 10-13 July 2005. pp.82-91.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. *McKinsey Global Institute.* Available:http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx

March, S. T. & Smith, G. F. 1995. Design and Natural Science Research on Information Technology. *Decision Support Systems,* Vol:15, pp.251-266.

McCarthy, J. 1980. Circumscription—a Form of Non-Monotonic Reasoning. *Artificial Intelligence,* Vol:13, pp.27-39.

McGilvray, D. & Thomas, G. 2008. Definitions of Data Categories. *In: Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information.* Ch.Appendix - Quick Reference. pp.293-294. Morgan Kaufmann Publishers.

Mikalef, P., Pappas, I. O., Krogstie, J. & Giannakos, M. 2017. Big Data Analytics Capabilities: A Systematic Literature Review and Research Agenda. *Information Systems and E-Business Management,* Vol:Online, pp.1-32.

Mithas, S., Lee, M. R., Earley, S., Murugesan, S. & Djavanshir, R. 2013. Leveraging Big Data and Business Analytics [Guest Editors' Introduction]. *IT Professional,* Vol:15, pp.18-20.

Moore, D. M., Carvalho, D. & Taylor, D. H. 2011. Humanitarian Logistics Professionalism. *In: Humanitarian Logistics*. Ch.10. pp.189-234. Kogan Page. London, UK.

Moorhead, J. & Sandler Clarke, J. 2015. The Global Humanitarian Crisis 'Can't Depend on a Continuous Increase in Funding'. *The Guardian*, 10 December 2015. Available:http://www.theguardian.com/global-development-professionals-network/2015/dec/10/global-humanitarian-crisis-continuous-increase-funding

Nemeth, S. 2013. The Effect of Competition on Terrorist Group Operations. *Journal of Conflict Resolution,* Vol:58, pp.336-362.

Normandeau, K. 2013. *Beyond Volume, Variety and Velocity Is the Issue of Big Data Veracity* [Online]. InsideBigData. Available:http://insidebigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/ [Accessed:2017].

Normann, R. 2001. *Reframing Business: When the Map Changes the Landscape*, John Wiley & Sons. Chichester, United Kingdom.

Nowrasteh, A. 2016. Terrorism and Immigration: A Risk Analysis. *CATO Institute*. Available:http://object.cato.org/sites/cato.org/files/pubs/pdf/pa798_2.pdf

Nunamaker, J. F., Chen, M. & Purdin, T. D. M. 1990. Systems Development in Information Systems Research. *Journal of Management Information Systems,* Vol:7, pp.89-106.

Opentracker *Definitions of Big Data* [Online]. Opentracker. Available:http://www.opentracker.net/article/definitions-big-data [Accessed:2014].

Otto, B. 2011. Data Governance. *Business & Information Systems Engineering,* Vol:3, pp.241-244.

Pawlak, P. & Dietrich, C. 2015. Global Terrorism: Trends in 2014/2015. *European Parliamentary Research Service (EPRS).* Available:http://www.europarl.europa.eu/RegData/etudes/BRIE/2015/571321/EPRS_BRI%282015%29571321_EN.pdf

Pears-Piggott, M. I. B. & Muir-Wood, R. 2016. What Constitutes a Global Baseline for Worldwide Casualties from Catastrophes? *International Journal of Disaster Risk Reduction,* Vol:17, pp.123-127.

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems,* Vol:24, pp.45-77.

Piatetsky, G. 2014. *CRISP-DM, Still the Top Methodology for Analytics, Data Mining, or Data Science Projects* [Online]. KDnuggets. Available:http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html [Accessed:2018].

Pizzi, M. 2014. UN Abandons Death Count in Syria, Citing Inability to Verify Toll. *Al Jazeera America*, 7 January, 2014. Available:http://america.aljazeera.com/articles/2014/1/7/un-abandons-deathcountinsyria.html

Polman, L. 2010. *The Crisis Caravan: What's Wrong with Humanitarian Aid?*, Metropolitan Books New York, USA.

Polman, L. 2011. *War Games: The Story of Aid and War in Modern Times*, UK ed. Viking. New York, USA.

Popovič, A., Hackney, R., Tassabehji, R. & Castelli, M. 2018. The Impact of Big Data Analytics on Firms' High Value Business Performance. *Information Systems Frontiers,* Vol:20, pp.209-222

Powers Dirette, D. 2016. Why the Veracity of Data Matters in Health Care Research. *The Open Journal of Occupational Therapy,* Vol:4, pp.1-4.

Press, G. 2013a. Data Science: What's the Half-Life of a Buzzword? *Forbes TECH*, 19 August 2013. Available:http://www.forbes.com/sites/gilpress/2013/08/19/data-science-whats-the-half-life-of-a-buzzword/#26896aff7bfd

Press, G. 2013b. A Very Short History of Big Data. *Forbes TECH*, 9 May 2013. Available:http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/

Press, G. 2013c. A Very Short History of Data Science. *Forbes TECH*, 28 May 2013. Available:http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#31ce836755cf

Price, M., Gohdes, A. & Ball, P. 2014. Updated Statistical Analysis of Documentation of Killings in the Syrian Arab Republic. On-line. *Commissioned by Office of the UN High Commissioner for Human Rights (UNHCR)*. Available:http://www.ohchr.org/Documents/Countries/SY/HRDAGUpdatedReportAug2014.pdf

Price, M., Gohdes, A. & Ball, P. 2016. Technical Memo for Amnesty International Report on Deaths in Detention. On-line. *Human Rights Data Analysis Group (HRDAG)*. Available:http://hrdag.org/wp-content/uploads/2016/07/HRDAG-AI-memo.pdf

Pries-Heje, J. & Baskerville, R. 2008. The Design Theory Nexus. *MIS Quarterly,* Vol:32, pp.731-755.

Provost, F. & Fawcett, T. 2013. Data Science and Its Relationship to Big Data and Data-Driven Decision Making. *Big Data,* Vol:1, pp.51-59.

Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y. & Brück, T. 2015. What Methods May Be Used in Impact Evaluations of Humanitarian Assistance? Germany. *Institute for the Study of Labor (IZA)*. Available:http://repec.iza.org/dp8755.pdf

Purvis, K. 2015. Delivering Aid: Agencies Struggle with Increasing Needs, Costs and Insecurity. *The Guardian*, 16 December 2015. Available:http://www.theguardian.com/global-development-professionals-network/2015/dec/16/delivering-humanitarian-aid-increasing-needs-costs-insecurity-conflict-

Raschky, P. A. & Schwindt, M. 2016. Aid, Catastrophes and the Samaritan's Dilemma. *Economica,* Vol:83, pp.624-645.

Rayner, S. 2006 Wicked Problems: Clumsy Solutions. Jack Beale Memorial Lecture on Global Environment, Sydney, Australia. James Martin Institute for Science and Civilization.

Reva, B. 2004. Consideration of the Origin of Herbert Simon's Theory of "Satisficing" (1933-1947). *Management Decision,* Vol:42, pp.1240-1256.

Riddell, R. C. 2008. The Growth of Emergencies and the Humanitarian Response. *In: Does Foreign Aid Really Work?* , Ch.18. pp.311-324. Oxford University Press. Oxford, United Kingdom.

Riddell, R. C. 2014a Does Foreign Aid Really Work? Australasian Aid and International Development Workshop, Canberra, Australia.

Riddell, R. C. 2014b. Does Foreign Aid Really Work? An Updated Assessment. *Australian National University, Development Policy Centre Discussion* Available:http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409847

Rittel, H. W. J. & Webber, M. M. 1973. Dilemmas in a General Theory of Planning. *Policy Sciences,* Vol:4, pp.155-169.

Safran, P. 2004. Disaster and Emergency Policy. Manila. *Asian Development Bank*. Available:http://www.adb.org/documents/disaster-and-emergency-assistance-policy

Saporito, P. 2014. *2 More Big Data V's — Value and Veracity* [Online]. SAP. Available:http://blogs.sap.com/innovation/big-data/2-more-big-data-vs-value-and-veracity-01242817 [Accessed:2014].

Sauer, A. 2016. *Humanitarian Supply Chain Performance Management: Development and Evaluation of a Comprehensive Performance Measurement Framework Based on the Balanced Scorecard*. Munich Business School.

Schleussner, C.-F., Donges, J. F., Donner, R. V. & Schellnhuber, H. J. 2016. Armed-Conflict Risks Enhanced by Climate-Related Disasters in Ethnically Fractionalized Countries. *Proceedings of the National Academy of Sciences,* Vol:113, pp.9216-9221.

Scholten, K., Scott, P. S. & Fynes, B. 2010. (Le)Agility in Humanitarian Aid (NGO) Supply Chains. *International Journal of Physical Distribution & Logistics Management,* Vol:40, pp.623-635.

Schroeck, M. 2012. Analytics: The Real-World Use of Big Data. How Innovative Enterprises Extract Value from Uncertain Data. IBM Global Business Services. Available:http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=GBE03519USEN

Schuemer-Cross, T. & Taylor, B. H. 2008. The Right to Survive. *Oxfam International*. Available:http://www.oxfam.org/sites/www.oxfam.org/files/right-to-survive-report.pdf

Seddon, M. 2017. Trump's Next Target Is Humanitarian Aid – a Devastating Blow for Millions. *The Guardian*, 10 April 2017. Available:http://www.theguardian.com/commentisfree/2017/apr/10/donald-trump-un-cuts-federal-budget-humanitarian-aid

Sharma, R., Mithas, S. & Kankanhalli, A. 2014. Transforming Decision-Making Processes: A Research Agenda for Understanding the Impact of Business Analytics on Organisations. *European Journal of Information Systems,* Vol:23, pp.433-441.

Shukla, A. 2010. *First Official Estimate: An NGO for Every 400 People in India* [Online]. The Indian Express. Available:http://archive.indianexpress.com/news/first-official-estimate-an-ngo-for-every-400-people-in-india/643302/ [Accessed:2017].

Simon, H. 1996. *Sciences of the Artificial*, 3rd ed. MIT Press. United States of America.

Smillie, I. & Minear, L. 2003. The Quality of Money. Donor Behaviour in Humanitarian Financing. On-line. *The Feinstein International Famine Center, Tufts University*. Available:http://reliefweb.int/sites/reliefweb.int/files/resources/FAE568BBFA9AE8A4C1256D3C0031BA16-tuft-donor-apr03-2.pdf

Smith, B. 2003. Ontology. *In: The Blackwell Guide to the Philosophy of Computing and Information*. Ch.11. pp.153-166. Wiley Online Library. Oxford, United Kingdom.

Sodhi, M. S. 2016. Natural Disasters, the Economy and Population Vulnerability as a Vicious Cycle with Exogenous Hazards. *Journal of Operations Management,* Vol:45, pp.101-113.

Spence, R. 2006. *Information Visualization: Design for Interaction*, 2nd ed. Prentice Hall. United Kingdom.

Stephenson, J. M. 2005. Making Humanitarian Relief Networks More Effective: Operational Coordination, Trust and Sense Making. *Disasters,* Vol:29, pp.337-350.

Strmberg, D. 2007. Natural Disasters, Economic Development, and Humanitarian Aid. *Journal of Economic Perspectives,* Vol:21, pp.199-222.

Strong, D. M., Lee, Y. W. & Wang, R. Y. 1997. Data Quality in Context. *Commun. ACM,* Vol:40, pp.103-110.

Tatham, P. & Houghton, L. 2011. The Wicked Problem of Humanitarian Logistics and Disaster Relief Aid. *Journal of Humanitarian Logistics and Supply Chain Management,* Vol:1, pp.15-31.

Tatham, P. & Hughes, K. 2011. Humanitarian Logistics Metrics : Where We Are and How We Might Improve. *In: Humanitarian Logistics : Meeting the Challenge of Preparing for and Responding to Disasters*. 1st ed, Ch.4. pp.Pages 65-84. Kogan Page Limited. London, United Kingdom.

Tatham, P. H. & Pettit, S. J. 2010. Transforming Humanitarian Logistics: The Journey to Supply Network Management. *International Journal of Physical Distribution & Logistics Management,* Vol:40, pp.609-622.

Themnér, L. & Wallensteen, P. 2014. Armed Conflicts, 1946–2013. *Journal of Peace Research,* Vol:51, pp.541-554.

Thomas, A. S. & Kopczak, L. R. 2005. From Logistics to Supply Chain Management: The Path Forward in the Humanitarian Sector. *Fritz Institute*. Available:http://www.fritzinstitute.org/PDFs/WhitePaper/FromLogisticsto.pdf

Tomasini, R. M. & Van Wassenhove, L. N. 2009. From Preparedness to Partnerships: Case Study Research on Humanitarian Logistics. *International Transactions in Operational Research,* Vol:16, pp.549-559.

Toya, H. & Skidmore, M. 2007. Economic Development and the Impacts of Natural Disasters. *Economics Letters,* Vol:94, pp.20-25.

Tschoegl, L., Below, R. & Guha-Sapir, D. 2006. An Analytical Review of Selected Data Sets on Natural Disasters and Impacts. Bangkok, Thailand. *UNDP/CRED*. Available:http://www.emdat.be/sites/default/files/TschoeglDataSetsReview.pdf

Vaishnavi, V. & Kuechler, W. 2004a. *Design Research in Information Systems* [Online]. Association for Information Systems. Available:http://ais.affiniscape.com/displaycommon.cfm?an=1&subarticlenbr=279 [Accessed.

Vaishnavi, V. & Kuechler, W. 2004b. *Design Science Research in Information Systems* [Online]. Available:http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf [Accessed:2017].

Vaishnavi, V. K. K., William Jr. 2008. *Design Science Research Methods and Patterns*, Auerbach Publications. United States of America.

Van Der Zee, B. 2015. UN Launches Its Biggest Ever Annual Appeal for Humanitarian Funding. *The Guardian*, 7 December 2015. Available:http://www.theguardian.com/global-development-professionals-network/2015/dec/07/un-launches-biggest-ever-annual-appeal-for-humanitarian-funding

Van Wassenhove, L. N. 2005. Humanitarian Aid Logistics: Supply Chain Management in High Gear. *The Journal of the Operational Research Society,* Vol:57, pp.475-489.

Vasilescu, L., Khan, A. & Khan, H. 2008. Disaster Management Cycle–a Theoretical Approach. *Management & Marketing-Craiova,* Vol:1, pp.43-50.

Velasquez, A., Rosales, C. & Ramirez, F. 2002. Comparative Analysis of Disaster Databases. Available:http://www.unisdr.org/2005/task-force/working%20groups/wg3/Comparative_Analysis_of_Disaster_Database.pdf

Venable, J. 2006a. A Framework for Design Science Research Activities. *Information Resource Management Conference*. Washington DC, USA.

Venable, J. 2006b The Role of Theory and Theorising in Design Science Research. First International Conference on Design Science Research in Information Systems, Feb 24 -25 Claremont, California, USA. Design Science Research in Information Systems and Technology (DESRIST).

Venable, J. 2013. Rethinking Design Theory in Information Systems. *Design Science at the Intersection of Physical and Virtual Design: 8th International Conference, DESRIST 2013, Helsinki, Finland, June 11-12, 2013. Proceedings*. Berlin, Germany: Springer Berlin Heidelberg.

Venable, J. 2015 Five and Ten Years On: Have Dsr Standards Changed? International Conference on Design Science Research in Information Systems, Dublin, Ireland. Springer, pp.264-279.

Venable, J., Pries-Heje, J. & Baskerville, R. 2012 A Comprehensive Framework for Evaluation in Design Science Research. Design Science Research in Information Systems. Advances in Theory and Practice, Dublin, Ireland. Springer, pp.423-438.

Venable, J., Pries-Heje, J. & Baskerville, R. 2016. Feds: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems,* Vol:25, pp.77-89.

Voigt, S., Giulio-Tonolo, F., Lyons, J., Kučera, J., Jones, B., Schneiderhan, T., Platzeck, G., Kaku, K., Hazarika, M. K. & Czaran, L. 2016. Global Trends in Satellite-Based Emergency Mapping. *Science,* Vol:353, pp.247-252.

Völz, C. 2005. Humanitarian Coordination in Indonesia: An NGO Viewpoint. Available:http://www.fmreview.org/sites/fmr/files/FMRdownloads/en/FMRpdfs/Tsunami/11.pdf

Wahlström, M. 2015. Sendai Framework for Disaster Risk Reduction 2015-2030. *Third United Nations World Conference on Disaster Risk Reduction*. Sendai, Miyagi, Japan.

Wailgum, T. 2007. *Supply Chain Management Definition and FAQs* [Online]. CIO. Available:http://www.cio.com/article/2439493/supply-chain-management/supply-chain-management-supply-chain-management-definition-and-solutions.html?page=2#toc-6 [Accessed:2017].

Walls, J. G., Widermeyer, G. R. & El Sawy, O. A. 2004. Assessing Information System Design Theory in Perspective: How Useful Was Our 1992 Initial Rendition? *Journal of Information Technology Theory and Application (JITTA),* Vol:6, pp.43-58.

Walls, J. G., Widmeyer, G. R. & El Sawy, O. A. 1992. Building an Information System Design Theory for Vigilant Eis. *Information Systems Research,* Vol:3, pp.36-59.

Wand, Y. & Wang, R. 1996. Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM,* Vol:39, pp.86-95.

Wang, R. Y. & Strong, D. M. 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems,* Vol:12, pp.5-33.

Ward, J. S. & Barker, A. 2013. Undefined by Data: A Survey of Big Data Definitions. Available:http://arxiv.org/abs/1309.5821?context=cs

Wattegama, C. 2007. ICT for Disaster Management. On-line. *United Nations Development Programme (UNDP)*. Available:http://www.unapcict.org/ecohub/resources/ict-for-disaster-management

Young, J. K. & Dugan, L. 2011. Veto Players and Terror. *Journal of Peace Research,* Vol:48, pp.19-33.

REFERENCES — *no author(s) named*

ADRC. 2017. *Asian Disaster Reduction Center (Adrc)* [Online]. ADRC. Available:http://www.adrc.asia/index.php [Accessed:2017].

ALNAP. 2017. *Active Learning Network for Accountability and Performance in Humanitarian Action (Alnap) Evaluation Reports* [Online]. ALNAP. Available:http://www.alnap.org/resources/results.aspx?type=22 [Accessed:2017].

APICS SCC. 2012. *Supply Chain Operations Reference Model (Scor) - Version 11.0* [Online]. SCC. Available:http://www.apics.org/apics-for-business/scor-and-your-organization, http://www.apics.org/docs/default-source/scc-non-research/apicsscc_scor_quick_reference_guide.pdf [Accessed:2016].

APSC. 2007. Tackling Wicked Problems: A Public Policy Perspective. Canberra, Australia. *Australian Government, Australian Public Service Commission*. Available:

ATI. 2017. *Aid Transparency Index (Ati)* [Online]. Global Campaign for Aid Transparency. Available:http://ati.publishwhatyoufund.org/ [Accessed:2017].

BBC. 2003. *World Drowning in Oceans of Data* [Online]. BBC. Available: [Accessed:2017].

BBC. 2016. *Syria: The Story of the Conflict* [Online]. Available: http://www.bbc.co.uk/news/world-middle-east-26116868.

BLS. 2016. *USA Consumer Price Index (CPI)* [Online]. United States Department of Labor. Available:www.bls.gov/cpi/ [Accessed:2016].

Care International. 2017. *Care International* [Online]. Available:http://www.care-international.org/ [Accessed:2017].

CBPF. 2017. *Country-Based Pooled Funds (CBPF)* [Online]. Available:http://www.unocha.org/our-work/humanitarian-financing/country-based-pooled-funds-cbpfs [Accessed:2017].

CERF. 2017. *Central Emergency Response Fund (CERF)* [Online]. Available:http://www.unocha.org/cerf/about-us/cerf-at-a-glance [Accessed:2017].

Christian Aid. 2015. *Corporate Evaluations* [Online]. Available:http://www.christianaid.org.uk/resources/corporate_reports/corporate-evaluations.aspx [Accessed:2017].

CHS Alliance. 2015. *Humanitarian Accountability Report* [Online]. Available:http://www.chsalliance.org/what-we-do/policy-advocacy-and-learning/published-research [Accessed:2017].

CHS Alliance. 2017. *Core Humanitarian Standard Alliance* [Online]. Available:http://www.chsalliance.org/ [Accessed:2017].

CRED. 2017. *Centre for Research on the Epidemiology of Disasters (CRED)* [Online]. Centre for Research on the Epidemiology of Disasters (CRED). Available:www.cred.be/ [Accessed:2017].

CSR-SY. 2017. *Syrian Center for Statistics and Research* [Online]. Available:http://www.csr-sy.org/index.php?l=1 [Accessed:2017].

Data Foundation. 2016. *The State of the Union of Open Data* [Online]. Data Foundation. Available:http://www.datafoundation.org/state-of-the-union-of-open-data-2016/#data-standardization [Accessed:2017].

DCHRS. 2017. *Damascus Center for Human Rights Studies* [Online]. Available:http://dchrs.org/ [Accessed:2017].

DesInventar. 2017a. *Desinventar, Disaster Information Management System, Database Downloads* [Online]. Network of Social Studies in the Prevention of Disasters in Latin America (Red de Estudios Sociales en Prevención de Desastres en América Latina - LA RED). Available:http://www.desinventar.net/DesInventar/index.jsp; http://www.desinventar.net/download.html [Accessed:2017].

DesInventar. 2017b. *Desinventar, Disaster Information Management System, Definition of Basic Effects,* [Online]. Network of Social Studies in the Prevention of Disasters in Latin America (Red de Estudios Sociales en Prevención de Desastres en América Latina - LA RED). Available:http://www.desinventar.net/effects.html [Accessed:2017].

DesInventar. 2017c. *Desinventar, Disaster Information Management System, Definition of Disaster Types* [Online]. Network of Social Studies in the Prevention of Disasters in Latin America (Red de Estudios Sociales en Prevención de Desastres en América Latina - LA RED). Available:http://www.desinventar.net/definitions.html [Accessed:2017].

DesInventar. 2017d. *Desinventar, Disaster Information Management System, Documentation* [Online]. Network of Social Studies in the Prevention of Disasters in Latin America (Red de Estudios Sociales en Prevención de Desastres en América Latina - LA RED). Available:http://www.desinventar.net/documentation.html [Accessed:2017].

DesInventar.NET. 2017. *Desinventar, Disaster Information Management System (.Net)* [Online]. UNISDR - Risk Knowledge Section. Available:http://www.desinventar.net [Accessed:2017].

DesInventar.ORG. 2017. *Desinventar, Disaster Information Management System (.Org)* [Online]. UNISDR - Risk Knowledge Section. Available:http://www.desinventar.org [Accessed:2017].

EDRIS. 2017. *European Disaster Response Information System (EDRIS)* [Online]. Available:http://ec.europa.eu/echo/ [Accessed:2016].

EU. 2017. *The European Union* [Online]. Available:http://europa.eu/european-union/about-eu/eu-in-brief_en [Accessed:2017].

Euromonitor. 2011. *Q&A: The Importance and Prospects of Mobile Telephony* [Online]. Euromonitor Global Market Research Blog. Available:http://blog.euromonitor.com/2011/05/qa-the-importance-and-prospects-of-mobile-telephony.html [Accessed.

FEMA. 2017. *Glossary of Terms* [Online]. Available:www.fema.gov/pdf/plan/glo.pdf [Accessed:2017].

FTS. 2017a. *Financial Tracking Services* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/ [Accessed:2017].

FTS. 2017b. *Financial Tracking Services, About* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/content/about-fts-1 [Accessed:2017].

FTS. 2017c. *Financial Tracking Services, Archive* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://ftsarchive.unocha.org/ [Accessed:2017].

FTS. 2017d. *Financial Tracking Services, Data Search* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/data-search [Accessed:2017].

FTS. 2017e. *Financial Tracking Services, FAQs About FTS* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/content/faqs-about-fts [Accessed:2017].

FTS. 2017f. *Financial Tracking Services, Glossary* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/glossary [Accessed:2017].

FTS. 2017g. *Financial Tracking Services, List of Affected Countries* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/countries/overview [Accessed:2017].

FTS. 2017h. *Financial Tracking Services, Total Reported Funding (by Year)* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://fts.unocha.org/global-funding/countries/2015 [Accessed:2017].

GAR/UNISDR. 2015. Global Assessment Report for Disaster Risk Reduction (GAR). On-line. *United Nations.* Available:http://www.preventionweb.net/english/hyogo/gar/2015/en/home/documents.html; http://www.preventionweb.net/english/hyogo/gar/

GHA. 2012. Global Humanitarian Assistance Report 2012. Somerset, UK. *Global Humanitarian Assistance.* Available:http://www.globalhumanitarianassistance.org/reports

GLIDE. 2017. *Global Identifier Number (GLIDE)* [Online]. ADRC & UNISDR. Available:http://www.glidenumber.net/glide/public/search/search.jsp [Accessed:2017].

GTD. 2017a. *Global Terrorism Database (GTD), About* [Online]. National Consortium for the Study of Terrorism and Responses to Terrorism. Available:http://www.start.umd.edu/about/about-start [Accessed:2017].

GTD. 2017b. *Global Terrorism Database (GTD), Codebook* [Online]. National Consortium for the Study of Terrorism and Responses to Terrorism. Available:http://www.start.umd.edu/gtd/downloads/Codebook.pdf [Accessed:2017].

GTD. 2017c. *Global Terrorism Database (GTD), Download* [Online]. National Consortium for the Study of Terrorism and Responses to Terrorism. Available:http://www.start.umd.edu/gtd/contact/ [Accessed:2017].

GTD. 2017d. *Global Terrorism Database (GTD), Frequently Asked Questions* [Online]. National Consortium for the Study of Terrorism and Responses to Terrorism. Available:http://www.start.umd.edu/gtd/faq/ [Accessed:2017].

GTD. 2017e. *Global Terrorism Database (GTD), National Consortium for the Study of Terrorism and Responses to Terrorism (Start)* [Online]. National Consortium for the Study of Terrorism and Responses to Terrorism. Available:http://www.start.umd.edu/gtd/ [Accessed:2017].

HDX. 2017. *Humanitarian Data Exchange (HDX)* [Online]. United Nations Office for the Coordination of Humanitarian Affairs (OCHA). Available:http://data.humdata.org/ [Accessed:2017].

HERR. 2011. *Humanitarian Emergency Response Review* [Online]. Department for International Development, UK Gov. Available:http://www.dfid.gov.uk/what-we-do/key-issues/humanitarian-disasters-and-emergencies/how-we-respond/humanitarian-emergency-response-review/ [Accessed:2013].

HRDAG-Syria. 2017. *Human Rights Data Analysis Group - Syria* [Online]. Available:http://hrdag.org/syria/ [Accessed:2017].

HSRP. 2013. Human Security Report 2013: The Decline in Global Violence: Evidence, Explanation, and Contestation. Vancouver, Canada. *Simon Fraser University.* Available:http://www.hsrgroup.org/docs/Publications/HSR2013/HSRP_Report_2013_140226_Web.pdf; http://www.hsrgroup.org/human-security-reports/2013/overview.aspx

Humanitarian Coalition. 2017. *From Humanitarian to Development Aid, Important Differences between Humanitarian Aid and Development Aid* [Online]. HumanitarianCoalition. Available:http://humanitariancoalition.ca/media-resources/factsheets/from-humanitarian-to-development-aid [Accessed:2017].

HYOGO. 2005 Hyogo Framework for Action (HFA) 2005 - 2015. World Conference on Disaster Reduction, January 2008 Kobe, Hyogo, Japan. Geneva, Switzerland: UN/ISDR & UN/OCHA

HYOGO. 2008 Indicators of Progress: Guidance on Measuring the Reduction of Disaster Risks and the Implementation of the Hyogo Framework for Action. World Conference on Disaster Reduction, January 2008 Kobe, Hyogo, Japan. Geneva, Switzerland: UN/ISDR.

IATI. 2017. *International Aid Transparency Initiative (IATI)* [Online]. International Aid Transparency Initiative. Available:http://www.aidtransparency.net/ [Accessed:2017].

IBM. *What Is Big Data?* [Online]. Available:http://www-01.ibm.com/software/data/bigdata/ [Accessed:2013].

IBM. 2013. *The Four V's of Big Data* [Online]. Available:http://www.ibmbigdatahub.com/infographic/four-vs-big-data [Accessed:2014].

ICRC. 2017. *The International Red Cross and Red Crescent Movement* [Online]. Available:http://www.icrc.org/en [Accessed:2017].

IDS. 2017. *International Development Statistics (IDS) Online Databases (OECD)* [Online]. Available:http://www.oecd.org/development/stats/idsonline.htm [Accessed:2017].

IFRC. 2017. *What Is a Disaster?* [Online]. Available:http://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/what-is-a-disaster/ [Accessed:2017].

IRC. 2017. *International Rescue Committee* [Online]. Available:http://www.rescue.org/ [Accessed:2017].

IRDR. 2014. Peril Classification and Hazard Glossary. Beijing, China. *International Social Science Council (ISSC); United Nations International Strategy for Disaster Reduction (UNISDR)*. Available:http://www.irdrinternational.org/wp-content/uploads/2014/04/IRDR_DATA-Project-Report-No.-1.pdf

IRDR. 2017. *Integrated Research on Disaster Risk (IRDR)* [Online]. International Social Science Council (ISSC); United Nations International Strategy for Disaster Reduction (UNISDR). Available:http://www.irdrinternational.org/ [Accessed:2017].

ISo-3166. 2017. *International Organization for Standardization, Country Codes - ISO 3166* [Online]. Available:http://www.iso.org/iso-3166-country-codes.html [Accessed:2017].

ISO. 2017. *International Organization for Standardization* [Online]. ISO. Available:www.iso.org/iso/home.html [Accessed:2017].

JHSPH. 2008. Disaster Definitions (Ch1). Geneva, Switzerland. *Johns Hopkins and Red Cross Red Crescent Societies*. Available:http://reliefweb.int/sites/reliefweb.int/files/resources/Forward.pdf

La Red. 2017. *Network of Social Studies in the Prevention of Disasters in Latin America (Red De Estudios Sociales En Prevención De Desastres En América Latina)* [Online]. Available:http://www.desenredando.org/ [Accessed:2017].

Lancet 2010. Growth of Aid and the Decline of Humanitarianism. *The Lancet*, Vol:375, pp.253 (Editorial).

LOG. 2017. *Intervention Types, Logistics Operational Guide (LOG)* [Online]. Available:dlca.logcluster.org/display/LOG/Intervention+Types [Accessed:2017].

MPTF. 2017. *UNDP, Multi-Partner Trust Fund Office* [Online]. Available:http://mptf.undp.org/ [Accessed:2017].

MSF. 2017. *Médecins Sans Frontières* [Online]. Available:http://www.msf.org/ [Accessed:2017].

OECD. 2017a. *OECD, Net ODA (Indicator)* [Online]. Available:http://data.oecd.org/oda/net-oda.htm [Accessed:2017].

OECD. 2017b. *Organisation for Economic Co-Operation and Development (OECD)* [Online]. Available:http://www.oecd.org/ [Accessed:2017].

OECD. 2017c. *Organisation for Economic Co-Operation and Development (OECD), DAC Glossary* [Online]. Available:http://www.oecd.org/dac/dac-glossary.htm [Accessed:2017].

OECD. 2017d. *Organisation for Economic Co-Operation and Development (OECD), FAQs - Development Finance Statistics* [Online]. Available:http://www.oecd.org/dac/stats/faq.htm [Accessed:2017].

OECD.Stat. 2017. *Organisation for Economic Co-Operation and Development (OECD), OECD.Stat* [Online]. Available:http://stats.oecd.org/ [Accessed:2017].

OECD/DCD-DAC/HA. 2017. *Humanitarian Assistance* [Online]. Organisation for Economic Co-operation and Development (OECD). Available:http://www.oecd.org/dac/stats/humanitarian-assistance.htm [Accessed:2017].

OECD: DAC. 2017. *OECD, Development (DAC) Members* [Online]. Available:http://www.oecd.org/dac/dacmembers.htm [Accessed:2017].

OECD: Korea. 2017. *OECD, Korea* [Online]. Available:http://www.oecd.org/dac/korea.htm [Accessed:2017].

ONS. 2013. A Short History of the Census. United Kingdom: Office for National Statistics.

Oxfam. 2017. *Oxfam International* [Online]. Available:http://www.oxfam.org/ [Accessed:2017].

Oxford Dictionaries. 2017. *Third Sector* [Online]. Available:http://en.oxforddictionaries.com/definition/third_sector [Accessed:2017].

PGIS. 2017. *Pinkerton Global Intelligence Services (PGIS)* [Online]. Pinkerton Global Corporate Intelligence Services. Available:http://www.pinkerton.com/intelligence-services/ [Accessed:2017].

Population Pyramid. 2017. *Population Pyramids* [Online]. Population Pyramids. Available:http://www.populationpyramid.net/ [Accessed:2017].

ReliefWeb. 2017. *Reliefweb* [Online]. UN Office for the Coordination of Humanitarian Affairs (OCHA). Available:http://reliefweb.int/disasters [Accessed:2017].

SAS. 2014. *Data Mining and Semma* [Online]. Available:http://support.sas.com/documentation/cdl/en/emcs/66392/HTML/default/viewer.htm#n0pejm83csbja4n1xueveo2uoujy.htm [Accessed:2018].

Save The Children. 2017. *Save the Children* [Online]. Available:http://www.savethechildren.net/ [Accessed:2017].

SIPRI. 2016. *Sipri Yearbook: Armaments, Disarmament and International Security* [Online]. Stockholm International Peace Research Institute (SIPRI). Available:www.sipri.org/yearbook [Accessed:2017].

sn4hr. 2017. *Syrian Network for Human Rights* [Online]. Available:http:\\sn4hr.org [Accessed:2017].

SOHR. 2017. *Syrian Observatory for Human Rights* [Online]. Available:http://www.syriahr.com/en/ [Accessed:2017].

START. 2017. *National Consortium for the Study of Terrorism and Responses to Terrorism* [Online]. Available:http://www.start.umd.edu/ [Accessed:2017].

Syrian Shuhada. 2017. *Syrian Shuhada, Syrian Revolution Martyr Database* [Online]. Available:http://translate.google.co.uk/translate?hl=en&sl=ar&u=http://syrianshuhada.com/&prev=search [Accessed:2017].

The European Commission. 2002. Guidelines for the Use of Indicators in Country Performance Assessment. Brussels, Belgium. *The EUROPEAN COMMISSION*. Available:http://ec.europa.eu/commission/index_en

The Fritz Insitute. 2007. *Fritz Institute Surveys* [Online]. Fritz Instite. Available:http://www.fritzinstitute.org/rescenSurveys.htm [Accessed:2017].

The Guardian. 2011. *Update: Aid for the Food Crisis in the Horn of Africa – Get the Data* [Online]. Available: http://www.theguardian.com/global-development/poverty-matters/2011/aug/01/drought-food-crisis-africa-data.

U.S. Dept. of State. 2017. *Non-Governmental Organizations (NGOs) in the United States* [Online]. Bureau of Democracy, Human Rights, and Labor, U.S. Dept. of State. Available:http://www.state.gov/j/drl/rls/fs/2017/266904.htm [Accessed:2017].

UCDP. 2017a. *Department of Peace and Conflict Research, Uppsala University* [Online]. Department of Peace and Conflict Research, Uppsala University. Available:http://www.pcr.uu.se/ [Accessed:2017].

UCDP. 2017b. *Uppsala Conflict Data Program (UCDP)* [Online]. Department of Peace and Conflict Research, Uppsala University. Available:http://ucdp.uu.se/ [Accessed:2017].

UCDP. 2017c. *Uppsala Conflict Data Program (UCDP), Definitions* [Online]. Department of Peace and Conflict Research, Uppsala University. Available:http://www.pcr.uu.se/research/ucdp/definitions/ [Accessed:2017].

UCDP. 2017d. *Uppsala Conflict Data Program (UCDP), Downloads* [Online]. Department of Peace and Conflict Research, Uppsala University. Available:http://ucdp.uu.se/downloads/ [Accessed:2017].

UCDP. 2017e. *Uppsala Conflict Data Program (UCDP), FAQs* [Online]. Department of Peace and Conflict Research, Uppsala University. Available:http://pcr.uu.se/research/ucdp/faq [Accessed:2017].

UN. 2017. *United Nations* [Online]. Available:http://www.un.org/en/index.html [Accessed:2017].

UNCTAD. 2017. *United Nations Conference on Trade and Development* [Online]. Available:http://unctadstat.unctad.org/EN/Index.html; http://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx [Accessed:2017].

UNDP. 2013. A Comparative Review of Country-Level and Regional Disaster Loss and Damage Databases. *United Nations Development Programme (UNDP)*. Available:

UNDP. 2017a. *United Nations Development Programme (UNDP)* [Online]. United Nations Development Programme (UNDP). Available:http://www.undp.org/ [Accessed:2017].

UNDP. 2017b. *United Nations Development Programme, Developing Regions* [Online]. Available:http://hdr.undp.org/en/content/developing-regions [Accessed:2017].

UNESCO-UIA. 2017. *Yearbook of International Organizations (YBIO)* [Online]. Available:http://portal.unesco.org/en/ev.php-URL_ID=35293&URL_DO=DO_TOPIC&URL_SECTION=201.html; http://www.uia.org/ybio/ [Accessed:2017].

UNHCR. 2017a. *UNHCR Statistical Online Population Database: Sources, Methods and Data Considerations* [Online]. UNHCR. Available:http://www.unhcr.org/uk/statistics/country/45c06c662/unhcr-statisti#OTHER_DATA_CONSIDERATIONS [Accessed:2017].

UNHCR. 2017b. *United Nations High Commissioner for Refugees (UNHCR), Persons of Concern* [Online]. UNHCR. Available:http://popstats.unhcr.org/en/persons_of_concern [Accessed:2017].

UNHCR. 2017c. *United Nations High Commissioner for Refugees (UNHCR), Population Statistics Database Overview* [Online]. UNHCR. Available:http://popstats.unhcr.org/en/overview [Accessed:2017].

UNHCR. 2017d. *United Nations High Commissioner for Refugees (UNHCR), Time Series* [Online]. UNHCR. Available:http://popstats.unhcr.org/en/time_series [Accessed:2017].

UNISDR. 2009. *Terminology, the United Nations Office for Disaster Risk Reduction (UNISDR)* [Online]. The United Nations Office for Disaster Risk Reduction (UNISDR). Available:www.unisdr.org/we/inform/terminology#letter-d [Accessed:2016].

UNISDR. 2017. *The United Nations Office for Disaster Risk Reduction (UNISDR)* [Online]. The United Nations Office for Disaster Risk Reduction (UNISDR). Available:http://www.unisdr.org/ [Accessed:2016].

UNISDR DRR. 2017. *What Is DRR?* [Online]. Available:http://www.unisdr.org/who-we-are/what-is-drr [Accessed:2017].

UNOCHA. 2016. UNOCHA Humanitarian Programme Cycle, Guide to Valuation of in-Kind Contributions from the Private Sector. On-line. *United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA).* Available:http://www.humanitarianresponse.info/en/programme-cycle/space/document/guide-valuation-kind-contributions-private-sector

UNOCHA. 2017. *United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA)* [Online]. Humanitarian Programme Cycle Information Services Unit (HPC-IS). Available:http://www.unocha.org/ [Accessed:2017].

VDC-SY. 2016a. *Database, Violations Documentation Center in Syria* [Online]. VDC-SY. Available:http://www.vdc-sy.info/index.php/en/martyrs, http://www.vdc-sy.info/index.php/en/detainees, http://www.vdc-sy.info/index.php/en/missing [Accessed:2016].

VDC-SY. 2016b. *Violations Documentation Center in Syria* [Online]. VDC-SY. Available:http://vdc-sy.net/en/ [Accessed:2016].

Esposito, M., Kappor, A. & Mathur, D. WEFORUM. 2016. *What Is the State of the BRICS Economies?* [Online]. World Economic Forum. Available:http://www.weforum.org/agenda/2016/04/what-is-the-state-of-the-brics-economies/ [Accessed:2017].

WHO. 2017. *World Health Organization (WHO)* [Online]. Available:http://www.who.int/en/ [Accessed:2017].

World Bank. 2017. *World Bank Databank* [Online]. World Bank. Available:http://databank.worldbank.org/data/home.aspx [Accessed:2017].

World Bank: HN&P. 2017. *World Bank Databank, Health Nutrition and Population Statistics* [Online]. World Bank. Available:http://databank.worldbank.org/data/reports.aspx?source=health-nutrition-and-population-statistics [Accessed:2017].

Worldometers. 2017. *Worldometers* [Online]. Worldometers. Available:http://www.worldometers.info/ [Accessed:2017].

WVI. 2017. *World Vision International* [Online]. Available:http://www.wvi.org/ [Accessed:2017].

APPENDICES

Appendix A: Data Veracity profiles (DVps)

A.1: DVp Template

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		*Enter Name of Dataset*	Action(s)
L2	L3		
Complete	**1. No omitted entries**		
Are there missing entries/observations?			
Complete	**2. No omitted values**		
Are there empty/incomplete fields?			
Complete	**3. No omitted variable**		
Are the observations incomplete? Are any known dataset variables inaccessible or corrupted?			
Complete	**4. No omitted metadata**		
Is there missing information about the data? Is there documentation that explains the structure and content of the dataset?			
Uncluttered	**5. No irrelevant entries**		
Is dataset is free from 'noise' or does it include inappropriate, spurious or misleading entries?			

Expository Dimensions			
Data Veracity Sub-Dimensions		*Enter Name of Dataset*	Action(s)
L2	L3		
Precise	**6. Reliability**		
Is the data volatile or uncertain, subject to unexplained change? To what extent does it deviate from the correct, intended or original values? Are there abnormalities in the data? Is it vague or confusing?			
Precise	**7. Rigour**		
Was the data meticulously collected or measured as opposed to being estimated or assumed. Is the data a product of scrupulous data gathering or assumptions/guesswork?			
Precise	**8. Congruity**		
Is there consistency and congruity in the data? Are the values of equivalent granularity and detail, are they consistent in there measures and representation?			
Accurate	**9. Conformity**		
To what extent does the data conform to facts and is a reflection of reality? What is the scale and depth of 'softness' in the data?			

Expository Dimensions			
Data Veracity Sub-Dimensions		*Enter Name of Dataset*	Action(s)
L2	L3		
Accurate	**10. Impartiality**		
Is the data biased or skewed? Can an underlying agenda be implies from the dataset?			
Accurate	**11. Validity**		
Is the data applicable, appropriate and relevant to the problem? Is the data current and up-to-date? Does the data contain obsolete information, if so to what extent?			

A.2: EM-DAT DVp

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">EM-DAT</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Complete</td><td colspan="3">1. No omitted entries</td></tr>
<tr><td colspan="3">Of the 21,982 clean entries in the finalised EM-DAT dataset, 13,810 were for natural disasters, 8,192 were for technological disasters and 14 were for complex disasters. As almost 63% of the entries are for natural disasters it is possible that there may be more emphasis placed on recording natural disasters than technological ones, rather than there were 26% fewer technological disasters? The 14 entries for complex disasters is a strong indication that complex disasters are not recorded. This also aligns to the lack of entries for conflict-related disasters; the plight of refugees; or the suffering of internally displaced persons. This appears to be the case even when these events may qualify as disasters according to the UNISDR definition and meet the EM-DAT inclusion criteria, neither of which include or exclude based on disaster group (UNISDR, 2009; Guha-Sapir et al., 2017e). Furthermore, the EM-DAT human impact criteria of at least 10 deaths or 100 people affected could risk 'blind spots' e.g. when each individual disaster in a series of related consecutive disasters does not match the EM-DAT inclusions criteria, but collectively equate to heavy losses. Additionally, the majority of entries are for the most recent 26 years in the dataset, it is highly likely that there are disasters missing from the first 90 years covered by the dataset.</td><td><u>Accepted</u> entries for certain disaster groups and years may be missing by focussing analysis primarily on Natural and Technological disasters, noting that entries for the latter may be light.
<u>Restricted</u> the focus of the study to a time-period starting 2-years after EM-DAT was launched, 1990-2015.</td></tr>
<tr><td>Complete</td><td colspan="3">2. No omitted values</td></tr>
<tr><td colspan="3">• 893 entries have no values for human and financial impact;
• 17.342 entries do not have any estimated damage;
• 934 have insured losses, of which 181 entries had no estimated damage;
• 17,191 entries have no values for estimated damaged or insured loss, which considering 4,386 of these were for transport accidents is incongruous.</td><td><u>Accepted</u> that it is only feasible to analyse human effect as financial information does not appear to be fully populated.
<u>Excluded</u> any human impact entries with missing or zero values for TotDeaths and TotAffected.</td></tr>
</table>

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		EM-DAT	Action(s)
L2	L3		
Complete	**3. No omitted variable**		
Less than half of the fields documented as held in the EM-DAT database were obtained via the historic EM-DAT R datasets and the screen scrape (Goteti, 2016; Guha-Sapir et al., 2017c).			**Corrected** for missing EM-DAT Region by manually creating a cross-reference table to use EM-DAT Country code to update the final dataset with the Region. **Accepted** that all other inaccessible variables will remain so.
Complete	**4. No omitted metadata**		
No description is provided in the guidelines of the use of zeros versus blanks (Guha-Sapir et al., 2017i). Typically zero is taken to mean a value was available and it was explicitly zero, while blank is taken to mean that a value was not available e.g. for TotDeaths, zero could be interpreted as no one died, whereas blank could mean that we don't know if anyone died. Furthermore, no explanation is provided of the relationship between estimated financial loss figures and insured financial loss figures.			**Accepted** the impossibility of telling which values of zero should be missing, therefore set the default to zero and toggled between zeros and missing for each relevant analysis to check if the results differ. **Interpreted** insured losses to be a proportion of estimated losses, but stored whichever of the two figures is greater in MSGD to avoid the risk of understating what $loss information is available.
Uncluttered	**5. No irrelevant entries**		
893 EM-DAT entries were for disasters that had no values for human or financial impact. It is not clear if these are placeholders or erroneous entries, but nonetheless they inflate the number of disasters without documenting exactly why they were disasters.			**Flagged** entries for 'no effect' disasters to be excluded from analysis.

Expository Dimensions			
Data Veracity Sub-Dimensions		EM-DAT	Action(s)
L2	L3		
Precise	**6. Reliability**		
		The fact that the data is regularly reviewed and corrected in and of itself indicates that it is not definitive and final, therefor subject to correction (Guha-Sapir et al., 2017f). Looking at the changes across the 3 EM-DAT datasets each new dataset retrospectively added or removed entries to past years, with the majority of these new entries affecting the most recent years. Similarly, values of entries were retrospectively changed, e.g. for 20% of the entries the values of human and financial impact were changed.	**Corrected** the data by excluding double entries for the same disasters. **Accepted** the data is subject to revision, but maintained an audit trail of the changes from one historic dataset to the next in case anomalous analytical results need to be investigated.
Precise	**7. Rigour**		
		When specific numbers for human impact are not provided, rough figures are entered based on either the language used by the source of the information, e.g. 'thousands' translates to 2000, 'hundreds' to 200; or by multiplying the number of houses destroyed by 3 or 5 depending on the development status of the country (Guha-Sapir et al., 2017i). Investigation indicates some 38% of the human impact values may be guestimates of this type.	**Accepted** that a significant proportion of disaster entries may have guestimated human impact values **Flagged** entries that conform to EM-DAT's estimating methodology as possible 'soft' numbers'.
Precise	**8. Congruity**		
		EM-DAT classification of disasters is detailed down to different degrees of specificity for *natural* disasters versus *technological* disasters. *Natural* disaster can be defined using up to 46 variations down to sub-subtype. In comparison, *technological* disasters are limited to 16 possible classifications, including 'miscellaneous accident' of type 'other' as a catch all. Complex disasters are not included in EM-DAT classifications tables. US$ values are relevant to the year of the entry.	**Accepted** this limitation and restricted any comparative analysis across disaster groups to the lowest level of sub-classification available for *Technological* disasters. **Adjusted** US$s to 2015 levels using US CPI numbers to enable comparison (BLS, 2016)

Expository Dimensions			
Data Veracity Sub-Dimensions		EM-DAT	Action(s)
L2	L3		
Accurate	**9. Conformity**		
Almost 4.5% of deaths and 36% of people affected are likely to be guestimated numbers.			**Accepted**, but tagged all potentially guestimated values 'soft'.
Accurate	**10. Impartiality**		
62.7% of the entries are for *Natural* disasters; 37.2% are for *Technological* disaster and only 14 entries are for Complex disasters. It is highly unlikely that only 14 disaster events were Complex over a period 115 years. Also, the 2:1 ratio of *Natural* to *Technological* disasters may indicate a bias in the data towards *Natural* Disasters. Disaster entries for the first 90 years, 1900 to 1989 inclusive, equate to 26% of the dataset, where the subsequent 26 years, from 1990 – 2015 inclusive, represent 74% of the entries. While there may be an argument that disasters have increased in recent years, it is far more likely that the data is more complete for the years after EM-DAT was launched.			**Compensated** for potential disaster group bias focussing only on *Natural* and *Technological* disasters, setting aside Complex disasters and accepting *Technological* disasters may be underrepresented and less refined in detail. **Accepted** the data prior to EM-DAT launch may be not be complete, therefore focus analysis on the time-period 1990 to 2015.
Accurate	**11. Validity**		
The data is relevant to the problem and it is historic and appropriate to the time period under consideration, therefore not obsolete.			None required.

A.3: DesInventar DVp

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">DesInventar</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Complete</td><td colspan="3">1. No omitted entries</td></tr>
<tr><td colspan="3">Although wider international usage of DesInventar has enabled the capture and availability of considerable new disaster information, thousands of entries in the raw data downloads were not usable, therefore had to be omitted. All countries are not represented in the dataset, e.g. Haiti, nor are all countries fully represent e.g. India and Nigeria. Coverage over time is also sporadic and inconsistent, e.g. Guam and Kiribati seem not to have experienced disasters since the mid-1800s; whereas American Samoa, Rwanda and Swaziland appear not have experienced any disasters prior to 2015. Additionally, disasters that do not originate in nature are not fully represented, e.g. only 758 entries could be attributed to conflict none of which represent the conflict in Syria or the genocide in Rwanda.</td><td><u>Accepted</u> entries that were usable as this dataset is used a supplement to disaster loss data sourced from EM-DAT. Nevertheless, these shortcomings must be taken into consideration as part of any analysis using this data.
<u>Reduced</u> focus to time period 1990-2015, guided by the time period used for EM-DAT.</td></tr>
<tr><td>Complete</td><td colspan="3">2. No omitted values</td></tr>
<tr><td colspan="3">Of the 322,489 entries retained
5,404 have no values for human and financial impact;
33,823 have no values for human impact;
279,982 were blank or zero for Losses $Local;
318,918 were blank or zero for Losses $USD
1,261 had an Event that was blank or ‘other’ or ‘unknown’
53,983 had a Cause that was blank or ‘other’ or ‘unknown’</td><td><u>Accepted</u>. For all but 88 entries the disaster type was interpreted from other fields. The focus will need to remain on analysing human impact as financial information is poorly populated.
<u>Excluded</u> any human impact entries with missing or zero values for Tot Deaths and Tot Affected.</td></tr>
</table>

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		DesInventar	Action(s)
L2	L3		
Complete	**3. No omitted variable**		
Of the 22 core variables used to populate the MSGD from DesInventar only 4 appeared consistently in all 79 downloaded DesInventar datasets. Another 4 could be found in 70 or more datasets. The remaining 14 variables appeared in fewer than 70 datasets, with 3 variables appearing in 7 or fewer datasets,			**Accepted** that unavailable variables will remain so, therefore set these to 'missing value' and processed them accordingly through the data preparation process.
Complete	**4. No omitted metadata**		
Although considerable documentation accompanies the DesInventar methodology and system (DesInventar, 2017d) it preparing the data it became obvious that local implementation varied greatly. Flexibility of structure and the creation and use of variables in local implementations could not be found. This was obvious in the variety of values that could be found in Event, Cause, Description of Cause and Comments fields. Similarly, some implementations appear to use the Losses $USD and Losses $Local as incremental, others populate one or the other, and yet others (84 entries) populate the fields with exact same value.			**Accepted** the impossibility of know what was actually intended by each of the 78 database owners. **Used** whichever field provided a clue as to the event that caused the losses. **Abandone**d the possibility of using Losses $Local for analysis.
Uncluttered	**5. No irrelevant entries**		
The 78 separate DesInventar databases were concatenated to create a single dataset with 1,178,753 entries. This dataset was carefully checked and 856,264 spurious, unintelligible, EM-DAT-sourced and duplicate entries removed. There remain, however, over 33,000 entries that contain no values for human effect and 88 that are for indeterminate disasters. Also, some entries were not necessarily for disasters, e.g. theft, rape etc.; but if on a large scale could be a sign of 'anarchy' or social unrest that is disastrous.			**Flagged** entries for 'no effect' disasters to be excluded from analysis. **Retained and classified** entries that may not be disasters accordingly to enable these to be excluded from analysis if appropriate.

<table>
<tr><th colspan="4">Expository Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">DesInventar</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Precise</td><td colspan="3">6. Reliability</td></tr>
<tr><td colspan="3">It is not obvious as to why or when data is updated on, or new databases added to, DesInventar's download page (DesInventar, 2017a). While the consolidated database, GAR2015 (with entries up to 2013) appears to have remained unchanged since first published, databases have been added and possibly updated, after its creation. There is no obvious process of correction or cross-referencing of changes between individual databases and the consolidated database.</td><td><u>Checked</u> manually for changes to suite of databases over the years and recreated the entire dataset 3 times to reduce the risk of omission and duplication.
<u>Maintained</u> an audit trail while creating the dataset and retained old versions for reference.</td></tr>
<tr><td>Precise</td><td colspan="3">7. Rigour</td></tr>
<tr><td colspan="3">Many thousands of the entries originally downloaded had to be entirely discarded because of issues of extremely poor quality. No systemic or methodological validations seems to be carried out when data is added to or maintained in DesInventar databases (Guha-Sapir and Hoyois, 2012). Trust appears to be placed entirely with those entering the data to enter it correctly. This is why errors such as location information could be found in the data field for a number of the original records.</td><td><u>Accepted</u> as even after considerable cleaning and elimination over 300,000 entries remain to supplement the ~22,000 entries from EM-DAT.
<u>Flagged</u> all DesInventar entries as possible 'soft' numbers'.</td></tr>
<tr><td>Precise</td><td colspan="3">8. Congruity</td></tr>
<tr><td colspan="3">IRDR compliant classifications were found in only 80,392 of the 322,489 retained DesInventar entries (IRDR, 2014). For another 27,665 entries, IRDR compliant classifications appear to be randomly entered as a cause, description of cause, or comment. For over 66% of entries the disaster classification constituted varied localised 'free range' descriptions of what happened, often local language, and entered into a field of choice. US$ values are relevant to the year of the entry.</td><td><u>Checked</u>, and if necessary translated, and interpreted 27,665 entries and either assigned a disaster classification already in the foundational MDC or created a suitable classification that could be added to the MDC.
<u>Adjusted</u> US$s to 2015 levels using US CPI numbers to enable comparison (BLS, 2016).</td></tr>
</table>

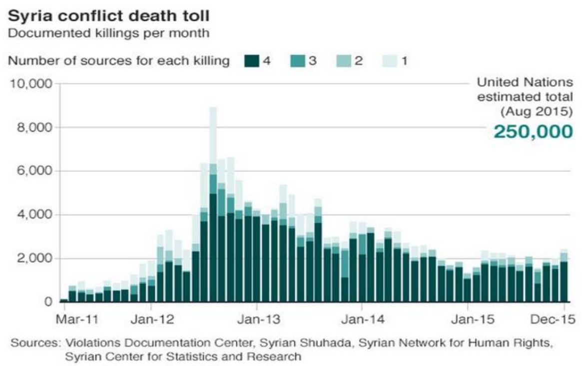
Expository Dimensions			
Data Veracity Sub-Dimensions		DesInventar	Action(s)
L2	L3		
Accurate	**9. Conformity**		
It is difficult to tell the accuracy of the quantifiable numbers for variables recording effect. From a visual of the databases, it is know that check marks are shown where the numbers for human effect variable are not known. The existence of these check marks is lost during the download. Therefore it is assumed for some, if not all, DesInventar databases the numbers of total deaths or total affected will be understated because of lack and loss of information. Similar caveats may apply to financial loss numbers, in that the inability to quantify has meant the information is not held.			**Accepted**, as there is no action that can be taken. That said, all analysis utilising human/financial effect values from DesInventar should be assumed to be based on 'rough' numbers.
Accurate	**10. Impartiality**		
Over 78% of the entries are for disasters originating from nature and 99.7% of entries are for the period 1970 to 2017, with 84% for the period 1990 to 2017. Therefore, it would be reasonable to consider the database information to be slanted towards naturogenic disasters over the past 27 years.			**Accepted.** DesInventar supplements data obtained from EM-DAT and is also 'skewed' to similar types of disasters and the bulk of its data extends over similar time periods. This gap in disaster types is compensated by adding 4 datasets that provide data other than for naturogenic disasters.
Accurate	**11. Validity**		
The data is relevant to the problem and it is historic and appropriate to the time period under consideration, therefore not obsolete. Different implementations however, of the DesInventar database may be used to record all kinds of loss data as is indicated by existence of over 1500 customised fields.			**Assigned** classifications to help identify any entries that may not be considered disasters to enable exclusion as and when necessary.

A.4: UCDP GED50 DVp

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		UCDP GED50	Action(s)
L2	L3		
Complete	**1. No omitted entries**		
Based on the scope of UCDP, there is no obvious evidence of missing entries, apart from entries for Syria, which the UCDP site and GED50 documentation state clearly are intentionally omitted from the dataset as UCDP is not confident in the data available to it for Syria.			**Identified** alternative source for Syria conflict data.
Complete	**2. No omitted values**		
Only two of the 128,264 entries in the dataset were missing information on fatalities.			**Accepted** that these values are missing.
Complete	**3. No omitted variable**		
All variables are available in the download.			None required.
Complete	**4. No omitted metadata**		
Full and detailed documentation is available and accessible and the data is well explained.			None required.
Uncluttered	**5. No irrelevant entries**		
Only entries relevant to the scope of UCDP appear to be held, with only 2 of the 128,264 entries with no fatalities, therefore these 2 entries may be spurious.			**Flagged** two 'empty' entries for exclusion during analysis.

Expository Dimensions			
Data Veracity Sub-Dimensions		UCDP GED50	Action(s)
L2	L3		
Precise	**6. Reliability**		
NEW GED versions are released as more conflict data is collated, this version is therefore static..			None required.
Precise	**7. Rigour**		
Even though fatality values are labelled as estimates the dataset appears to be meticulously maintained and its sources are cited in detail. That said, 4,573 entries are designated to Yugoslavia (912 entries) and Soviet Union (3,661 entries).			**Accepted** that UCDP do not claim their numbers are definitive. **Corrected** the outdated country assignments.
Precise	**8. Congruity**		
3 numeric (integer) codes are used in UCDP's type_of_violence. These are explained in detail and appear to be used congruously.		**1** state-based conflict **2** non-state conflict **3** one-sided violence	**Accepted** these classifications and created equivalent entries in the MDC.
Accurate	**9. Conformity**		
best_est is perceived to be the firmest of the values for total fatalities in the UCDP dataset. best_est is populated in all but 13,092 of the 128,264 entries. For the approximate 10% of entries with zero best_est, high_est is populated and can be used. It could be argued that for 13,090 (2 entries have zero values for all fatality variables) where high_est is used are 'softer'.			**Accepted** the softness of using high_est when best_est is zero. **Flagged** entries that use high_est as possible 'soft' numbers'.
Accurate	**10. Impartiality**		
No evidence of bias was found			None required.
Accurate	**11. Validity**		
UCDP can be considered a relevant source of conflict-related disasters.			None required.

A.5: VDC-SY DVp

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		**VDC-SY**	**Action(s)**
L2	**L3**		
Complete	**1. No omitted entries**		
VDC-SY totals less than 152K deaths from 2011 to end 2015, which is almost 100K fewer than the 250K deaths depicted in the graph below, which is displayed in coverage by the BBC of the Syrian Civil War (BBC, 2016). These are UN calculations based on numbers from fours sources (1) VDC-SY; (2) Syrian Shuhada; (3) Syrian Network for Human Rights and (4) the Syrian Center for Statistics and Research (VDC-SY, 2016b; Syrian Shuhada, 2017; sn4hr, 2017; CSR-SY, 2017). Based on this, it is reasonable to assume that the VDC-SY as a single source for data for the Syrian conflict is not complete and there are missing entries. 			**<u>Recognised</u>** that ~100K discrepancy in number of deaths between the VDC-SY and the UN graph used by the BBC may not take account of possible double-counting by the UN as one of their sources, Syrian Shuhada cites another of their sources VDC-SY as a data source. **<u>Accepted</u>** any shortfall in entries as all other sources, including the UN raw data for the graph shown by the BBC are not available, accessible or usable. The alternative would be to have no data for the Syria conflict.
Complete	**2. No omitted values**		
• 758 detail entries (resulting in 16 summary entries) had no date • for 6,009 detail entries (rolling-up to 988 summary entries) a location was not specified • 14,881 entries were for unidentified persons (affecting 4,046 summary entries)			**<u>Accepted</u>** this information is not available. **<u>Flagged</u>** entries without location and known individuals as possible 'soft' numbers'.
Complete	**3. No omitted variable**		
All known variables were available in the datasets.			None required.

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		VDC-SY	Action(s)
L2	L3		
Complete	**4. No omitted metadata**		
Most of the variables are self-explanatory, which is good as very little detailed documentation is available for the data or the rationale/validation applied when populating the database.			**Accepted** this limitation as the fields that were of interest for the MSGD were self-explanatory.
Uncluttered	**5. No irrelevant entries**		
The 758 entries with no dates are effectively 'clutter' as it is not possible to tell when they occurred therefore they are limited in their usefulness.			**Flagged** the undated entries for exclusion from analysis.

Expository Dimensions			
Data Veracity Sub-Dimensions		VDC-SY	Action(s)
L2	L3		
Precise	**6. Reliability**		
It is not obvious if the database is subject to change. On the surface it appears to be added to rather than modified.			**Accepted** that this it cannot easily be determined.
Precise	**7. Rigour**		
Notwithstanding the missing values, the data is meticulously collected, often down to the place of birth of the victims, and at great risk to those collecting it.			**Accepted** excluded entries with missing dates from analysis and flagged missing location and identification as 'soft'.

Expository Dimensions			
Data Veracity Sub-Dimensions		VDC-SY	Action(s)
L2	L3		
Precise	**8. Congruity**		
Only one MDC disaster classification applies to all entries and as each entry represents one person there is little room for a incongruity in granularity or aggregation.			None required.
Accurate	**9. Conformity**		
The data is a count of dead, missing and detained at a per person level but some vagueness/blanks in qualifiers exist; which may be an indication of some 'softness' in these numbers.			Flagged missing location and identification as ‘soft’.
Accurate	**10. Impartiality**		
The 6:1 ratio between those Killed (non-regime) and Regime Fatalities may indicate recording bias in favour of non-regime fatalities.			**Accepted** the data is likely unbiased, but as the study is not comparing victim’s affiliation, the main effect on the study will be understated fatalities.
Accurate	**11. Validity**		
The dataset was selected because it is a valid fit to the study.			None required.

A.6: GTD DVp

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">GTD</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Complete</td><td colspan="3">1. No omitted entries</td></tr>
<tr><td colspan="3">All entries for 1993 are missing and changes data collection methodology in 1998, 2008 and 2012 have made a difference to the data recorded (GTD, 2017d). This, together with the observation that of the data that are available, and populated with the effect on people, are predominantly for deaths, which may mean that there may be entries missing for incidents when people are wounded or kidnapped.</td><td><u>Accepted</u>, as even with the possibility of missing data, this still improves the coverage of the MSGD.</td></tr>
<tr><td>Complete</td><td colspan="3">2. No omitted values</td></tr>
<tr><td colspan="3">103,219 entries have no values for human and financial impact;
112,557 entries do not show any human effect;
146,035 entries do not record any financial losses;
5,478 entries were for an 'unknown' type of terror attack
1,050 entries included an undefined number of hostages
12 entries did not have type of attack or number of hostages defined</td><td><u>Accepted</u> the dataset contains entries that are surplus to those relevant to this study.
<u>Assigned</u> disaster classifications to 5 unknown attack type entries based on information in the secondary attack descriptor field. No other unknown attacks had information in the secondary and tertiary attach descriptor fields.
<u>Flagged</u> entries with 'unknown' type and undefined number of hostages as possible 'soft' numbers'.</td></tr>
<tr><td>Complete</td><td colspan="3">3. No omitted variable</td></tr>
<tr><td colspan="3">All known variables were available in the datasets.</td><td>None required.</td></tr>
</table>

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		GTD	Action(s)
L2	L3		
Complete	**4. No omitted metadata**		
All data, including reasoning for populating the variable in a particular way, very well documented.			None required.
Uncluttered	**5. No irrelevant entries**		
103,219 GTD entries have no values for human or financial impact. These hold other information pertinent to the database, but for the purposes of this study these are surplus entries.			**Flagged** these 'empty' entries for exclusion.

Expository Dimensions			
Data Veracity Sub-Dimensions		GTD	Action(s)
L2	L3		
Precise	**6. Reliability**		
GDT changes to coding and recording processes can be retrospectively applied, as can corrections to the data. Process changes are documented; data changes do not appear to be tracked.			**Accepted** that there have been changes to process and data can and will take place. Will work with the last downloaded of dataset knowing that there is potential for change.
Precise	**7. Rigour**		
Beyond 1998, whatever data collection processes are current appear to be meticulously applied and information sources are cited.			None required.

Expository Dimensions			
Data Veracity Sub-Dimensions		GTD	Action(s)
L2	L3		
Precise	**8. Congruity**		
Care appears to be taken to maintain congruity, to the extent that people related fields, such as deaths, wounded and kidnapped can be entered as fractions across linked events for 'statistical accuracy' (GTD, 2017b). Notably, changes to data collection methodology in 1998, 2008 and 2012, may mean an imbalance in which events and the level of detail recorded. US$ values are relevant to the year of the entry.			**Accepted** the possibility of incongruity in data that may have been caused by 3 collection methodology changes. **Adjusted** US$s to 2015 levels using US CPI numbers to enable comparison (BLS, 2016).
Accurate	**9. Conformity**		
There is importance placed on accuracy and sources, but some unknowns in numbers are identified. In particular for hostage/kidnapping incidents, where numbers are not know -99 is entered.			**Flagged** entries with 'unknown' numbers as 'soft'.
Accurate	**10. Impartiality**		
While as whole the dataset shows no partiality to any one group or country, but of 70,988 entries (before netting with UCDP) that record either deaths or people affected, 70,911 entries record deaths, only 761 record people affected. This may indicate a preference to record incidents that result in fatalities, rather than those that may only result in injuries or kidnappings.			**Accepted** as there is no obvious additional source of information. Potential weakness recognised.
Accurate	**11. Validity**		
Where losses (human and financial) are recorded, entries can be considered valid for this study.			None required.

A.7: UNHCR DVp

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">UNHCR</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Complete</td><td colspan="3">1. No omitted entries</td></tr>
<tr><td colspan="3">The UNHCR Time Series dataset is aggregated to subtotal to Year, country of Origin and Country of destination. Detail entries are missing therefore this data can only be compared, reconciled and analysed at this aggregated level. the UNHCR site clearly highlights the potential for underrepresentation of affected people being an issue (UNHCR, 2017a). Two of the reasons cited include the inability to register some individuals who unlawfully settle outside official camps and the lack of refugee registers in industrialised countries (ibid).</td><td><u>Accepted</u>. Analysis will need to be restricted to Year and Country (Origin/Destination) level when this dataset is used.</td></tr>
<tr><td>Complete</td><td colspan="3">2. No omitted values</td></tr>
<tr><td colspan="3">• 129,177 have zero values for people affected
• 4,687 entries are assigned Various/Unknown instead of country of Origin
• 4,157 entries where * was used instead of a value between 1-4</td><td><u>Flagged</u> zero/* value entries for exclusion.
<u>Flagged</u> all entries where country of origin is unknown as ‘soft’ numbers.</td></tr>
<tr><td>Complete</td><td colspan="3">3. No omitted variable</td></tr>
<tr><td colspan="3">More variables are held in UNHCR than made available though the Time Series dataset. There appears to be no way to access these additional variables such that they can be reconciled to the aggregated time series data download.</td><td><u>Accepted</u> the limitations of the dataset.</td></tr>
</table>

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">UNHCR</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Complete</td><td colspan="3">4. No omitted metadata</td></tr>
<tr><td colspan="3">A high level overview of the data is made available via an 'Overview' web page (UNHCR, 2017a). It explains the definitions of each UNHRC Person of Concern (PoC) status and provides some background to the data collection process at a very high level view, but not detailed metadata.</td><td><u>Accepted</u> this limitation.</td></tr>
<tr><td>Uncluttered</td><td colspan="3">5. No irrelevant entries</td></tr>
<tr><td colspan="3">Over and above the 133,334 'empty' entries (129,177 0-value & 4,157 with *) , the dataset included:
• one entry with a spurious negative one value
• 952 entries for 1 person, 1 origin, 1 year – highly unlikely to be as a result of a disaster
• an additional 6,873 entrees for returnees – not relevant to the aftermath of a disaster</td><td><u>Flagged</u> all such entries for exclusion.</td></tr>
</table>

Expository Dimensions			
Data Veracity Sub-Dimensions		UNHCR	Action(s)
L2	L3		
Precise	**6. Reliability**		
No change management information provided. The best on offer is a separate download of mid-year statistics, but it is not clear if the data is retrospectively improved or corrected.			**Accepted** as the data will be used a snapshot in time, but future use could be influenced by this.
Precise	**7. Rigour**		
The site documents that UNHCR can employing estimates in their figures, particularly where their system does not have information of 'illegally camped' refugees or they cannot obtain definitive figures from 'industrialised' countries who do not hold refugee registers (UNHCR, 2017a).			**Accepted** as this is the best information currently available.
Precise	**8. Congruity**		
The data is at too high a level to identify incongruities that may/may not be an underlying issue.			None required.
Accurate	**9. Conformity**		
14,710 entries (~ 5%) were identified as potentially guestimated numbers.			**Flagged** as potentially 'soft' numbers.
Accurate	**10. Impartiality**		
The data does not provide any indication of bias.			None required.
Accurate	**11. Validity**		
49,197 entries were for returnees; therefore about 16.5% of the dataset had no relevance.			None required.

A.8: FTS DVp

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		FTS	Action(s)
L2	L3		
Complete	**1. No omitted entries**		
FTS was established in 1992, however the earliest humanitarian funding data available is from 1999 (FTS, 2017b) – no explanation is provided. The earliest two years of data, 1999 and 2000 have significantly lower values and fewer entries that subsequent years, therefore it is conceivable that these two year do not reflect the full complement of humanitarian assistance provided. Additionally, the detailed flows and the annualised summary cannot be completely reconciled; therefore there exists the possibility that the acquired detail of contribution flows is missing entries. – 231 (> US$11.55bn) of the annual summary entries cannot be matched to the detailed flow subtotal for the same country/year.			**Accepted**, as there is no choice here other than to work with what information is available. **Flagged** the 231 irreconcilable entries in the year/country summary as potentially 'soft' numbers.

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		FTS	Action(s)
L2	L3		
Complete	**2. No omitted values**		

Detail flows

3,753 entries of the detailed flows have no value for the aid provided. Although most of these are for in-kind contributions without valuations (UNOCHA, 2016), 22% are for financial or unspecified types of contributions.

With the numbers of nil contribution entries falling below 100 in only 2 of the 16 years that they are found. The 146 that had no *Contribution Type* also have no *Description*; most – 131 entries – are for Haiti.

No US$	Contribution type			
Flow	No info	Financial	In Kind	Total
Incoming	137	387	1,848	**2,372**
Internal	0	10	42	**52**
Outgoing	9	284	1,036	**1,329**
Total	**146**	**681**	**2,926**	3,753

2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
259	383	270	211	235	407	277	129	147	85	315	215	89	146	320	265

Year	to 'Unspecified'
1999	-$38,613,755
2000	$641,140,369
2001	$609,368,318
2002	$566,697,945
2003	-$1,152,793,379
2004	$383,380,098
2005	$5,724,827,805
2006	$1,128,595,779
2007	$710,828,390
2008	$2,954,019,126
2009	$1,744,478,178
2010	$1,066,144,829
2011	$2,300,295,378
2012	$1,923,488,462
2013	$1,429,021,634
2014	$4,177,961,046
2015	$2,368,817,333
Total	$26,537,657,556

Year/Country Summary

Each of the 17 years in the FTS summary has one entry showing contribution values associated with 'unspecified' recipient countries, the net total of these exceed US$26.5bn at 2015 values. No information is provided as to how this should be interpreted.

Action(s):

__Accepted__ as there is little else that can be done. The annualised summary dataset will be the primarily basis for analysis and this 'missing' information in the detail is assumed to be absorbed (lost) in the calculations.

__Accepted__ that the recipient information for a net of over $26.5bn cannot be viewed. As the existence of these is now known, they may provide some explanation to any anomalies in analysis results.

Elucidatory Dimensions			
Data Veracity Sub-Dimensions		**FTS**	**Action(s)**
L2	**L3**		
Complete	**3. No omitted variable**		
The EM-DAT site refers to FTS as its source for "*disaster-specific international aid contributions*" (FTS, 2017a; Guha-Sapir et al., 2017a). Therefore, prior to obtaining FTS detailed flow data it is assumed that there is at least one variable in FTS that links funding to specific disasters. This is not the case. The only way any full mapping of funding to disaster can be attempted is to read the free-format Description field for each of the 204,278 populated entries (of 209,162 total entries) and manually attempt to reconcile these to disasters that occurred in each country in the equivalent year. Alternative variables that may be useful (see table), are either inconsistently populated or do not hold meaningful information. In short, there is no variable visible or accessible in FTS that appears to definitively link all flow entries to specific disasters.			**Accepted** as this is now a moot point. The detail flow cannot be used for analysis because it cannot be reconciled to the FTS year/country summaries. Therefore inability to link a contribution to a disaster is a level of detail that is not needed for now.
Complete	**4. No omitted metadata**		
There is no metadata that explains: • Negative contribution values in flow totals. • The calculation logic for summary totals versus detail entries. • 'Empty' (zero-value) Entries. • When, and with what justification, 'unspecified recipient' is used..			**Accepted** that the only solution for now is to use the year/country summary. Will need to investigate the effect of negative values before considering how best to address them.

# empty entries of the 209,162 in the dataset		
Variable	**Destination**	**Source**
Emergency:	150,504	207,896
Project:	125,320	208,897
Plan:	114,468	199,002

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">FTS</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Uncluttered</td><td colspan="3">5. No irrelevant entries</td></tr>
<tr><td colspan="3">
Detail flows
The 3,753 zero value contribution are clutter, in that they provide no relevant funding information.
Year/Country Summary
• 794 of the 2,756 entries originally downloaded were for zero funding.
• In 2001, 2006 and 2008 entries exist for both Serbia and Serbia and Montenegro.
• In 2015 there are two entries that are not for a specific country.
<table>
<tr><th></th><th>Recipient Country</th><th>Funding US$</th><th>Pledges US$</th></tr>
<tr><td>2001</td><td>Serbia</td><td>208,719</td><td>0</td></tr>
<tr><td></td><td>Serbia and Montenegro (until 2006-2009)</td><td>334,001</td><td>0</td></tr>
<tr><td>2006</td><td>Serbia</td><td>1,396,153</td><td>0</td></tr>
<tr><td></td><td>Serbia and Montenegro (until 2006-2009)</td><td>2,628,187</td><td>0</td></tr>
<tr><td>2008</td><td>Serbia</td><td>5,050,877</td><td>0</td></tr>
<tr><td></td><td>Serbia and Montenegro (until 2006-2009)</td><td>0</td><td>0</td></tr>
<tr><td>2015</td><td>Multiple Locations (shared)</td><td>5,201,374</td><td>.</td></tr>
<tr><td></td><td>Not specified</td><td>2,307,090,019</td><td>56,525,940</td></tr>
</table>
</td><td>
<u>Flagged</u> all zero-contribution in the detail flow for exclusion if the dataset is used for any analysis.
<u>Removed</u> all zero-contribution entries from the Year/Country Summary.
<u>Merged</u> the 2015 'Multiple Locations (shared)' into 'Not specified' in the Year/Country summary dataset to consolidate all entries that do not relate to a specific country for the same year into one entry.
<u>Merged</u> Serbia and Serbia and Montenegro (until 2006-2009) in 2001, 2006 and 2008 to simplify ISO assignment and regional calculations
</td></tr>
</table>

Expository Dimensions			
Data Veracity Sub-Dimensions		FTS	Action(s)
L2	L3		
Precise	**6. Reliability**		
Detail flows The detail flow dataset provided dates of when an entry is created and modified. It shows that of the 209,162 entries from period 1999–2015, 4298 were last updated after 2015 and 3049 were newly created after 2015. There is no information as to what changes are made and there Is no visibility of any deleted entries. **Year/Country Summary** This does not have any audit trail for changes.			**Revised** and renewed the detail flow data as best possible adding 7,121 entries to the detail data already downloaded, but the detail still does not reconcile to the Year/Country summary. **Focussed** the study on the Year/Country summary dataset as the detail one falls short of the numbers in the Year/Country summary and there is no obvious way of ensuring that both can be kept in sync.
Precise	**7. Rigour**		
Data is updated from information supplied by sources to FTS, where FTS staff validate and reconcile incoming information and update the database. Here it is assumed that there is rigour at data entry, but no assumptions can be made of the quality of the data supplied by the sources.			None required.
Precise	**8. Congruity**		
All US$ values are held at the value of the US$ of the year in which given. **Detail flows** There are extremes of US$ contribution per entry; with one entry for over US$1bn co-existing with 40 entries of US$1 each. **For Detail flows & Year/Country Summary** US$ values are relevant to the year the contribution was made.			**Adjusted** US$s to 2015 levels using US CPI numbers to enable comparison (BLS, 2016)

Expository Dimensions			
Data Veracity Sub-Dimensions		FTS	Action(s)
L2	L3		
Accurate	**9. Conformity**		
Although it is not possible to tell if any numbers have been estimated rather than simply entered as received from the source, 231 of the Year/Country summary values cannot be reconciled to the detail flows. The calculations for these subtotals are not published and various permutations of logic to recreate these values from the detail flows are unsuccessful.			**Flagged** these 231 entries as possible 'soft' numbers'.
Accurate	**10. Impartiality**		
The database should cover at least 24 years (1992–2015) it only covers 17 years (1999–20 15); 1999 & 2000 may not be fully represented. As the documented scope of the dataset is not 15 years (2001–2015) an assumption of tacit bias towards these more recent years is assumed.			None required.
Accurate	**11. Validity**		
FTS is cited by EM-DAT as the their source for disaster-specific funding therefore is considered valid (Guha-Sapir et al., 2017a). Note that the "*disaster-specific international aid contributions*" information that EM-DAT allude to on their site is not found on FTS, and there is no cross-referencing with entries in EM-DAT(ibid).			None required.

A.9: IDS (OECD) DVp

<table>
<tr><th colspan="4">Elucidatory Dimensions</th></tr>
<tr><th colspan="2">Data Veracity Sub-Dimensions</th><th rowspan="2">IDS (OECD)</th><th rowspan="2">Action(s)</th></tr>
<tr><th>L2</th><th>L3</th></tr>
<tr><td>Complete</td><td colspan="3">1. No omitted entries</td></tr>
<tr><td colspan="3">Aid data from the OECD is only available as annual summaries and not at the detail level, therefore it is not possible to drill down to search for anomalies or trace back the source of negative flows. Not all countries report through the OECD, only the 30 DAC members, which equates to 37 sovereign states once the 28 members of the EU are identified and rationalised (EU, 2017).
Additionally, there other funds given that are known by the OECD but there is no matching flow in the database, e.g. US$13.3m given by Korea to the Democratic People's Republic of Korea (DPRK) in 2014 (OECD: DAC, 2017; OECD: Korea, 2017)</td><td><u>Accepted</u> as this is the only other credible source of aid data, apart from FTS.</td></tr>
<tr><td>Complete</td><td colspan="3">2. No omitted values</td></tr>
<tr><td colspan="3">659,953 entries in Table 2a are missing values, 10 of these are for Humanitarian Aid
76,034 entries in Table 2a are zero value, 682 of these are for Humanitarian Aid
1,575 entries in Table 2b are zero value, 163 of these are for Total OOF, Net
4,035 entries in Table 3a are zero value, 663 of these are for Total Commitments
9 entries in Table 4 are missing values
5,926 entries in Table 4 are zero value, 76 of these are for Total Private Net</td><td><u>Accepted</u> these null entries exist and they are filtered out.</td></tr>
<tr><td>Complete</td><td colspan="3">3. No omitted variable</td></tr>
<tr><td colspan="3">All variables that are visible on the site appear to be provided in the downloads.</td><td>Accepted</td></tr>
</table>

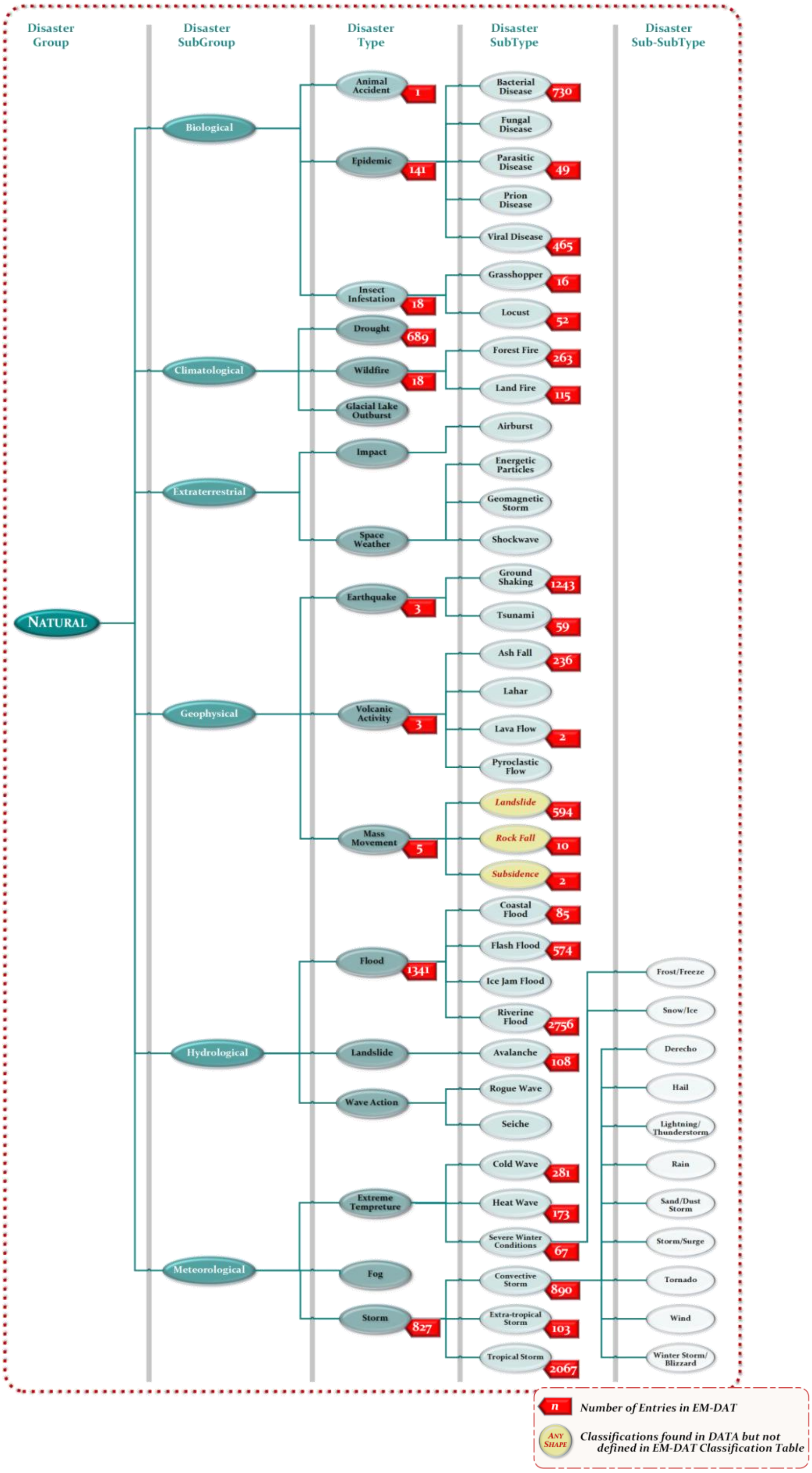
Elucidatory Dimensions			
Data Veracity Sub-Dimensions		IDS (OECD)	Action(s)
L2	L3		
Complete	**4. No omitted metadata**		
There is no comprehensive detailed information about negative value flows. Information provided with tables is usually only a high level description and no detail of how each variable or total is captured or calculated is provided calculation of totals. Missing/zero value entries are not explained.			**Accepted** this limitation.
Uncluttered	**5. No irrelevant entries**		
Of the 10,889,879 entries downloaded via Tables 2a, Table2b, Table 3 and Table 4, only 416,024 – ~3.8% – are relevant to this study. Some were duplicates, some were missing values or were for zero values, some were for banks, funds and NGOs therefore were 'pass through' flows, including them would result in double counting.			**Flagged** all superfluous entries for exclusion.

Expository Dimensions			
Data Veracity Sub-Dimensions		IDS (OECD)	Action(s)
L2	L3		
Precise	**6. Reliability**		
There is no visibility of what, how or if any retrospective changes are made.			**Accepted** this limitation.
Precise	**7. Rigour**		
Cannot find any information to confirm that the data is meticulously collected and maintained.			**Accepted** this limitation.
Precise	**8. Congruity**		
The data is at too high a level to identify incongruities that may/may not be an underlying issue.			None required.
Accurate	**9. Conformity**		
Unable to tell as the data is only visible in the aggregate.			None required.
Accurate	**10. Impartiality**		
Not enough information to gauge bias. Its scope is OECD members, therefore the dataset fulfils its scope.			None required.
Accurate	**11. Validity**		
The 416,024 usable entries retrieved from the 4 tables are valid for the study.			None required.

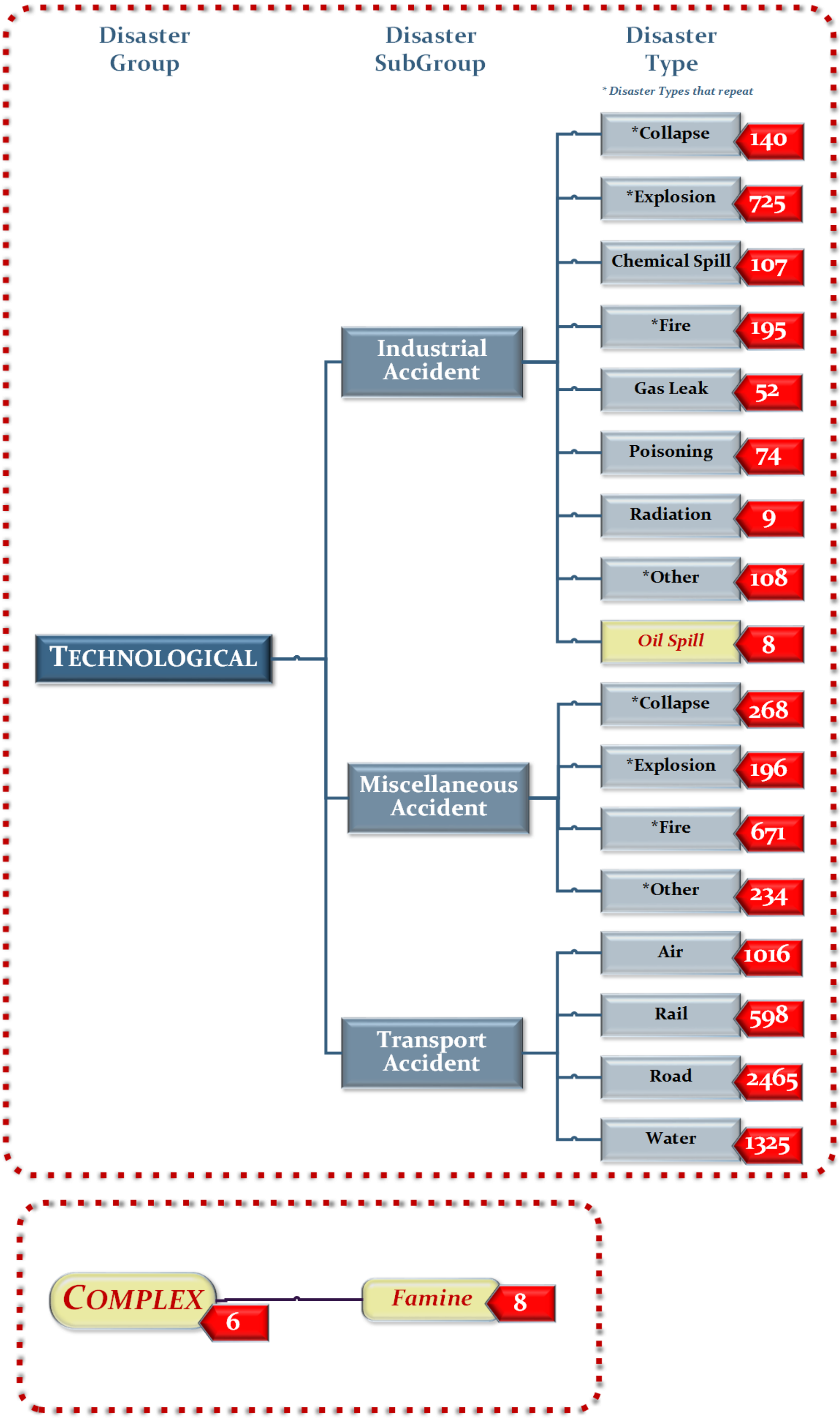
Appendix B: Disaster Classifications

B.1: EM-DAT Disaster Classifications

B.1.1: Natural Disasters



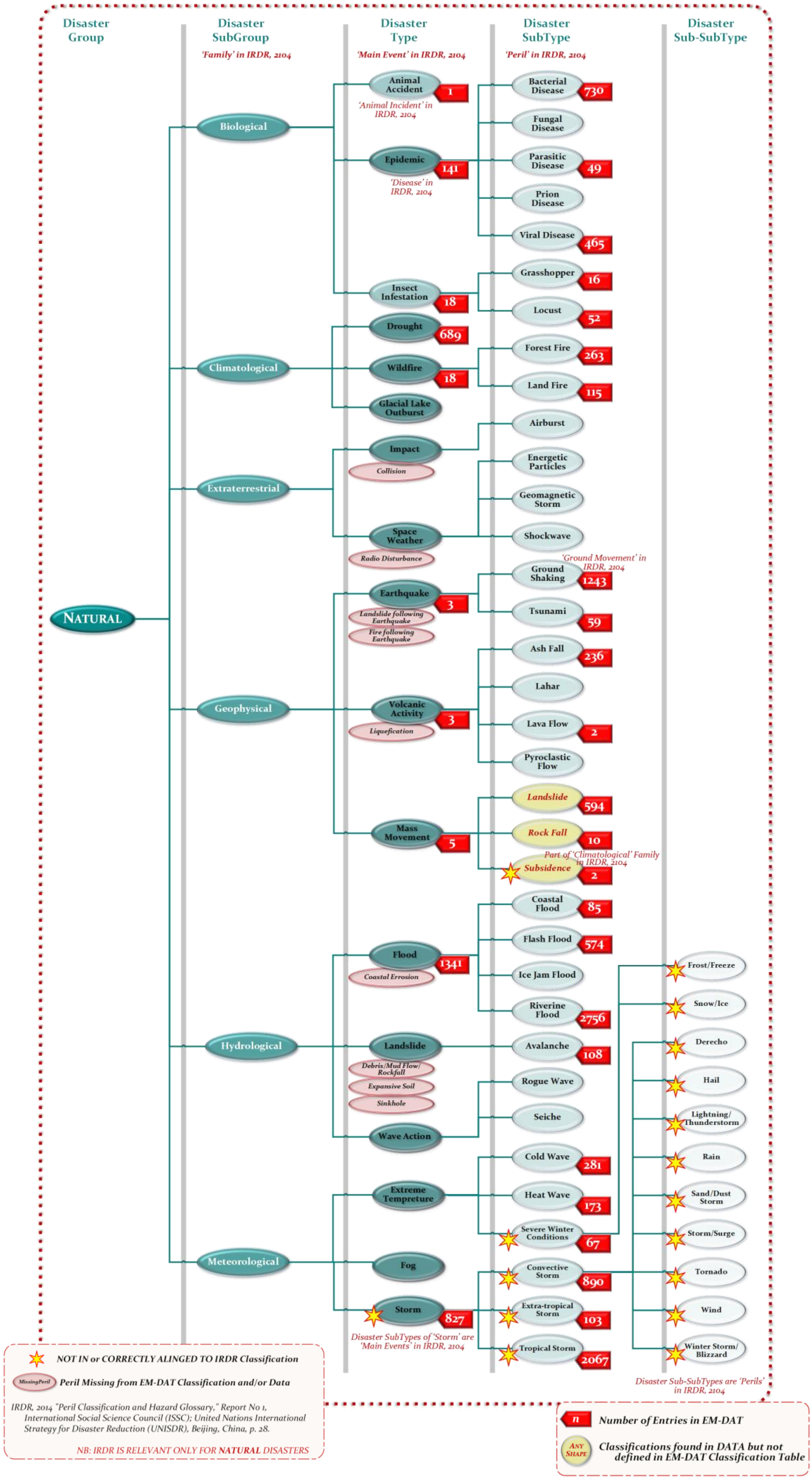
B.1.2: Technological and Complex Disasters



n *Number of Entries in EM-DAT*

ANY SHAPE *Classifications found in DATA but not defined in EM-DAT Classification Table*

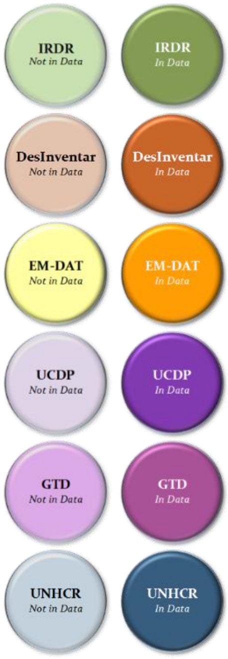
B.1.3: Natural Disaster EM-DAT vs IRDR



B.2: Master Disaster Classification (MDC)

This appendix includes schematics of the MDC hierarchy created for this study. It colour coded to identify the source/inspiration for each entry as defined in the colour key below. Lighter shades of each colour indicate MDC entries that do not in and of themselves represent an entry in the Master Sect of Global Disasters. This is typical of nomenclature at the Group or Family level. There are occasions, however, when an Event, with one or more Peril below it, may be shaded darker because it has been explicitly assigned to MSGD entries. Conversely, a Peril may be shaded light because, even though it appears in a cited classification, it was never assigned to an MSGD entry. The last section in this appendix contains tables detailing each MDC entry, its number of MSGD entries, a breakdown of the number of its occurrences per data source and its origination.

MDC Colour Key



IRDR, Integrated Research on Disaster Risk Peril Classification and Hazard Glossary (IRDR, 2014)

DesInventar, Disaster Inventory System (DesInventar.NET, 2017)

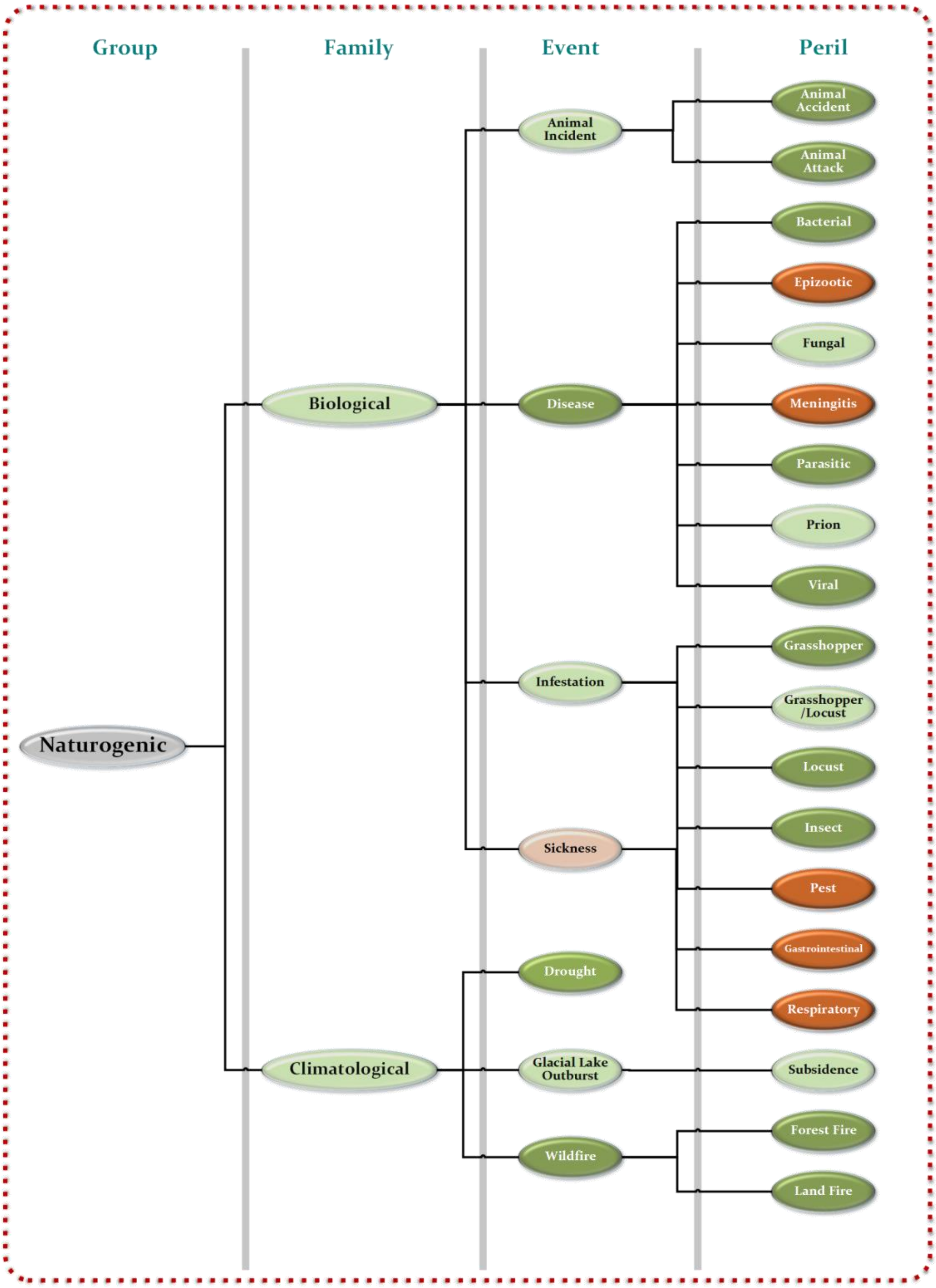
EM-DAT, Emergency Events Database (Guha-Sapir et al., 2017l)

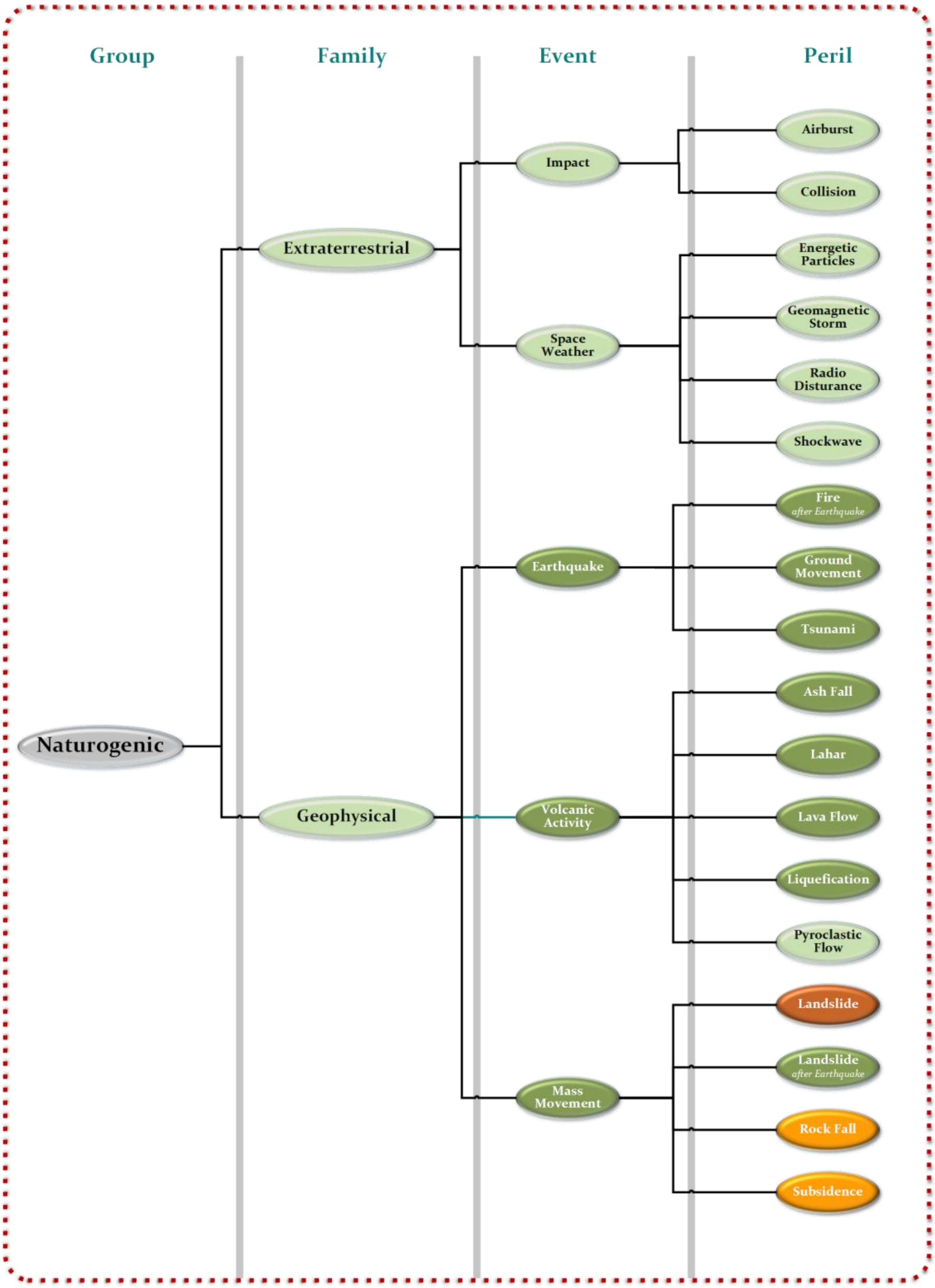
UCDP, Uppsala Conflict Data Program (UCDP, 2017b)

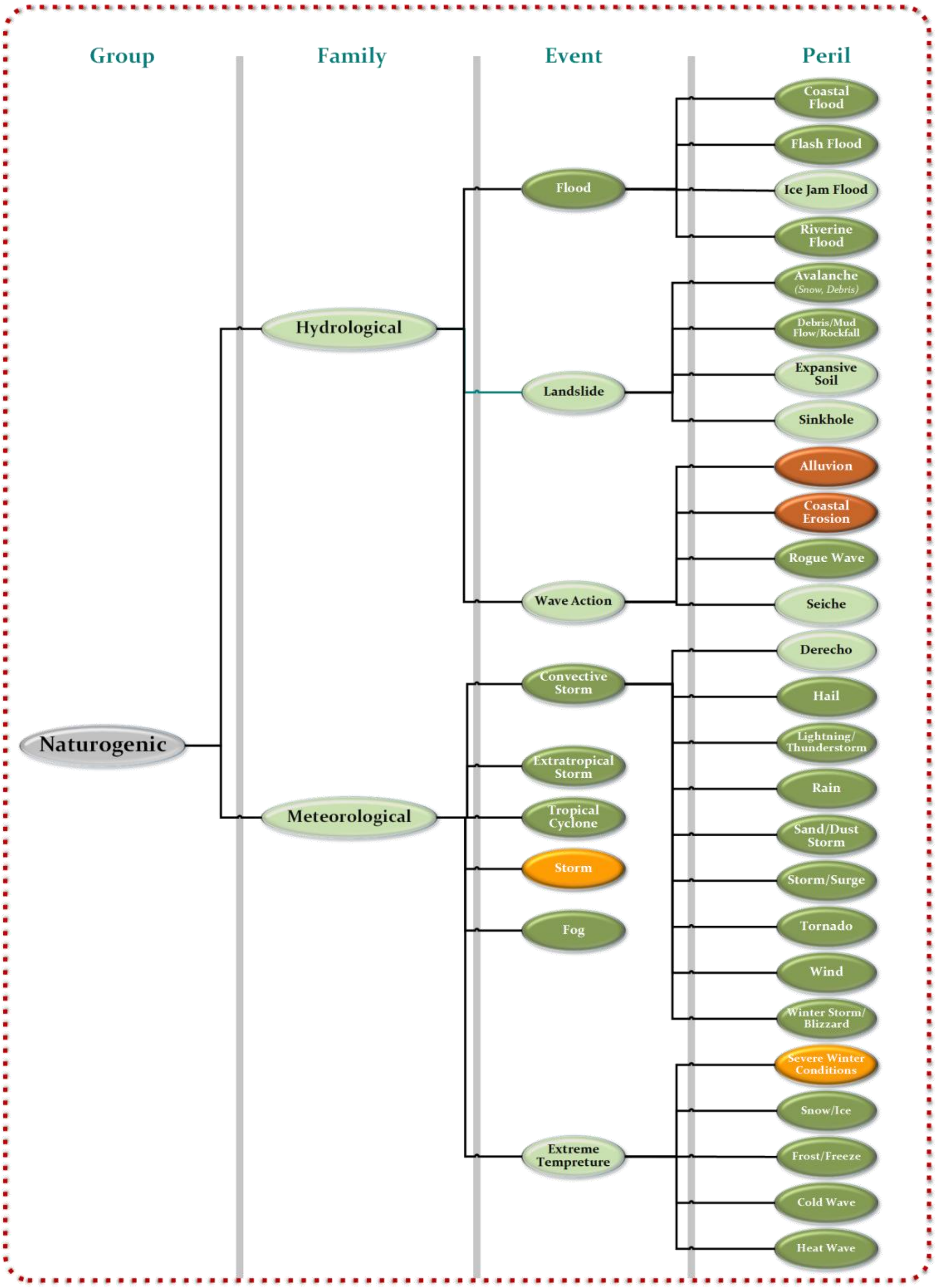
GTD, Global Terrorism Database (GTD, 2017e)

UNHCR, United Nations High Commissioner for Refugees (UNHCR, 2017c)

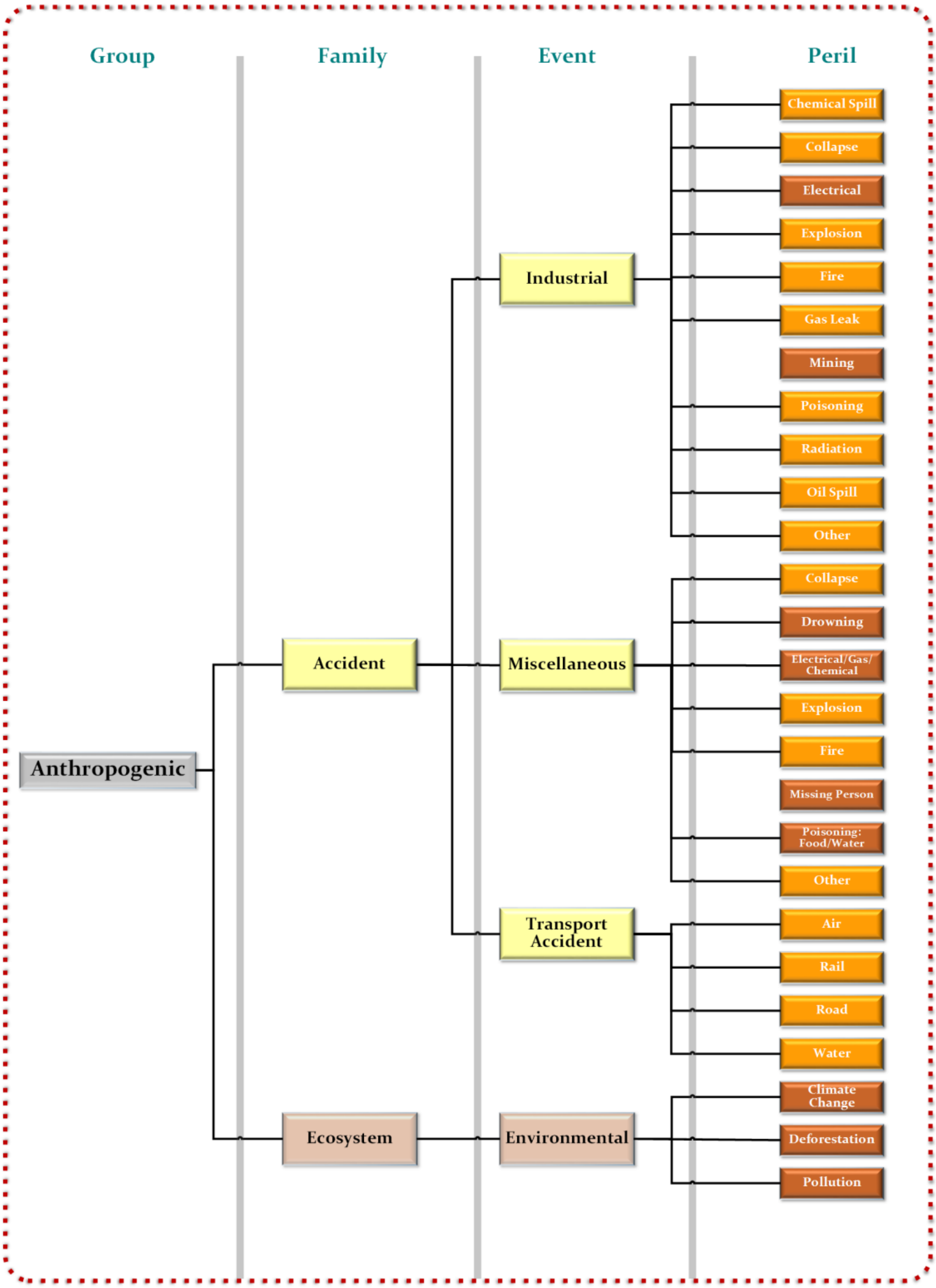
B.2.1: Naturogenic Disasters



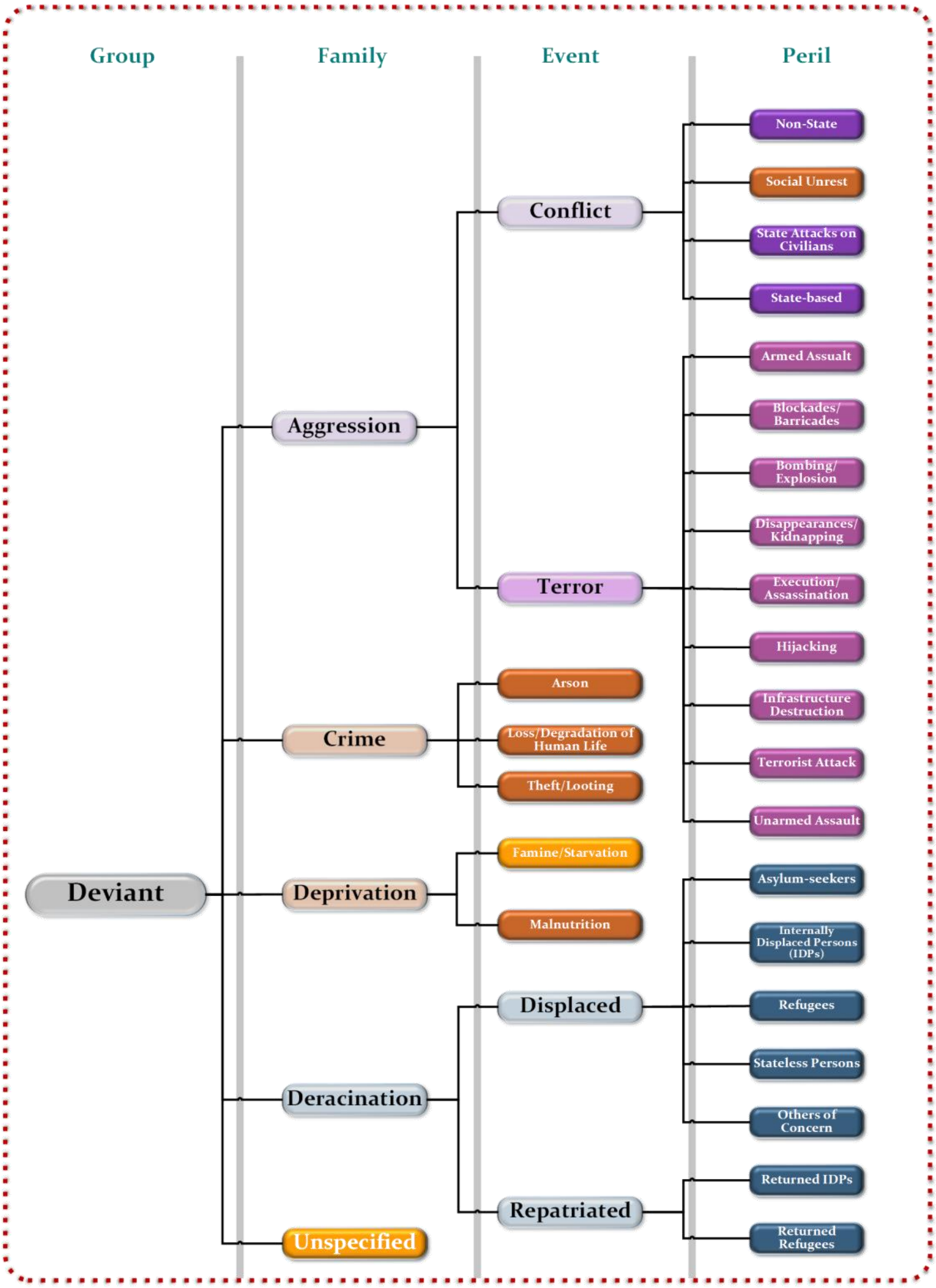

Group
Family
Event
Peril
Naturogenic
Extraterrestrial
Geophysical
Impact
Space Weather
Earthquake
Volcanic Activity
Mass Movement
Airburst
Collision
Energetic Particles
Geomagnetic Storm
Radio Disturance
Shockwave
Fire after Earthquake
Ground Movement
Tsunami
Ash Fall
Lahar
Lava Flow
Liquefication
Pyroclastic Flow
Landslide
Landslide after Earthquake
Rock Fall
Subsidence


Group
Family
Event
Peril
Naturogenic
Hydrological
Meteorological
Flood
Landslide
Wave Action
Convective Storm
Extratropical Storm
Tropical Cyclone
Storm
Fog
Extreme Temprature
Coastal Flood
Flash Flood
Ice Jam Flood
Riverine Flood
Avalanche (Snow, Debris)
Debris/Mud Flow/Rockfall
Expansive Soil
Sinkhole
Alluvion
Coastal Erosion
Rogue Wave
Seiche
Derecho
Hail
Lightning/ Thunderstorm
Rain
Sand/Dust Storm
Storm/Surge
Tornado
Wind
Winter Storm/ Blizzard
Severe Winter Conditions
Snow/Ice
Frost/Freeze
Cold Wave
Heat Wave

B.2.2:Anthropogenic Disasters



B.2.3:Deviant Disasters



B.2.4: MDC: MSGD Entries (+ Numbers from each data source)

Naturogenic										
Family	**Event**	**Peril**	**MSGD**	**EM-DAT**	**DesInventar**	**UCDP**	**VDC-SY**	**GTD**	**UNHCR**	**Origination**
Biological	Animal Incident	Animal Accident	3	1	2					IRDR to Main Event
		Animal Attack	9,874		9,874					IRDR to Main Event
	Disease		7,059	141	6,918					IRDR to Main Event
		Bacterial	34,252	730	33,522					IRDR to Peril
		Epizootic	1,467		1,467					DesInventar Data
		Fungal								IRDR to Peril
		Meningitis	261		261					DesInventar Data
		Parasitic	2,088	49	2,039					IRDR to Peril
		Prion								IRDR to Peril
		Viral	30,752	465	30,287					IRDR to Peril
	Infestation	Grasshopper	25	16	9					IRDR to Main Event
		Grasshopper/Locust								IRDR to Main Event
		Insect	363	18	345					IRDR to Main Event
		Locust	171	52	119					IRDR to Main Event
		Pest	15		15					DesInventar Data
	Sickness	Gastrointestinal	23,060		23,060					DesInventar Data
		Respiratory Difficulties	366		366					DesInventar Data
Climatological	Drought		15,550	689	14,861					IRDR to Main Event
	Glacial Lake Outburst	Subsidence								IRDR to Peril
	Wildfire		83	18	65					IRDR to Main Event
		Forest Fire	4,935	263	4,672					IRDR to Peril
		Land Fire: Brush, Bush, Pasture	553	115	438					IRDR to Peril
Extraterrestrial	Impact	Airburst								IRDR to Peril
		Collision								IRDR to Peril
	Space Weather	Energetic Particles								IRDR to Peril
		Geomagnetic Storm								IRDR to Peril
		Radio Disturbance								IRDR to Peril
		Shockwave								IRDR to Peril

Naturogenic										
Family	**Event**	**Peril**	**MSGD**	**EM-DAT**	**DesInventar**	**UCDP**	**VDC-SY**	**GTD**	**UNHCR**	**Origination**
Geophysical	Earthquake		192	3	189					IRDR to Main Event
		Fire after Earthquake	2		2					IRDR to Peril
		Ground Movement	4,118	1,243	2,875					IRDR to Peril
		Tsunami	483	59	424					IRDR to Peril
	Mass Movement		5	5						IRDR to Main Event
		Landslide	17,347		17,347					DesInventar Data
		Landslide after Earthquake	594	594						IRDR to Peril
		Rock Fall	258	10	248					EM-DAT Data
		Subsidence	276	2	274					EM-DAT Data
	Volcanic Activity		352	3	349					IRDR to Main Event
		Ash Fall	237	236	1					IRDR to Peril
		Lahar	4		4					IRDR to Peril
		Lava Flow	2	2						IRDR to Peril
		Liquefication	371		371					IRDR to Peril
		Pyroclastic Flow								IRDR to Peril
Hydrological	Flood		29,563	1,341	28,222					IRDR to Main Event
		Coastal Flood	92	85	7					IRDR to Peril
		Flash Flood	3,760	574	3,186					IRDR to Peril
		Ice Jam flood								IRDR to Peril
		Riverine Flood	12,857	2,576	10,281					IRDR to Peril
	Landslide	Avalanche: Snow, Debris	594	108	486					IRDR to Peril
		Debris/Mud Flow/Rockfall	98		98					IRDR to Peril
		Expansive Soil								IRDR to Peril
		Sinkhole								IRDR to Peril
	Wave Action	Alluvion	453		453					DesInventar Data
		Coastal Erosion	92		92					IRDR to Peril
		Rogue Wave	40		40					IRDR to Peril
		Seiche								IRDR to Peril

2

Naturogenic										
Family	**Event**	**Peril**	**MSGD**	**EM-DAT**	**DesInventar**	**UCDP**	**VDC-SY**	**GTD**	**UNHCR**	**Origination**
Meteorological	Convective Storm		890	890						IRDR to Main Event
		Derecho								IRDR to Peril
		Hail	1,231		1,231					IRDR to Peril
		Lightning	4,917		4,917					IRDR to Peril
		Rain	27,534		27,534					IRDR to Peril
		Sandstorm/Dust Storm	17		17					IRDR to Peril
		Storm Surge	652		652					IRDR to Peril
		Tornado	212		212					IRDR to Peril
		Wind	9,970		9,970					IRDR to Peril
		Winter Storm/Blizzard	1,158		1,158					IRDR to Peril
	Extratropical Storm		103	103						IRDR to Main Event
	Extreme Temperature	Cold Wave	1,433	281	1,152					IRDR to Peril
		Frost/Freeze	1,147		1,147					IRDR to Peril
		Heat Wave	747	173	574					IRDR to Peril
		Severe Winter Conditions	78	67	11					EM-DAT Data
		Snow/Ice	22		22					IRDR to Peril
	Fog		150		150					IRDR to Main Event
	Storm		6,386	827	5,559					EM-DAT Data
	Tropical Cyclone		6,835	2,067	4,768					IRDR to Main Event

Anthropogenic											
Family	**Event**	**Peril**	**MSGD**	**EM-DAT**	**DesInventar**	**UCDP**	**VDC-SY**	**GTD**	**UNHCR**	**Origination**
Accident	Industrial	Chemical Spill	137	107	30					EM-DAT Classification
		Collapse	589	140	449					EM-DAT Classification
		Electrical	576		576					DesInventar Data
		Explosion	1,078	725	353					EM-DAT Classification
		Fire	224	195	29					EM-DAT Classification
		Gas Leak	68	52	16					EM-DAT Classification
		Mining	185		185					DesInventar Data
		Oil Spill	22	8	14					EM-DAT Data
		Other	140	108	32					EM-DAT Classification
		Poisoning	132	74	58					EM-DAT Classification
		Radiation	22	9	13					EM-DAT Classification
	Miscellaneous	Collapse	1,745	268	1,477					EM-DAT Classification
		Drowning	3,037		3,037					DesInventar Data
		Electrical/Gas/Chemical	421		421					DesInventar Data
		Explosion	1,254	196	1,058					EM-DAT Classification
		Fire	41,822	671	41,151					EM-DAT Classification
		Missing Persons	92		92					DesInventar Data
		Other	2,242	234	2,008					EM-DAT Classification
		Poisoning: Food/Water	3,057		3,057					DesInventar Data
	Transport	Air	1,232	1,016	216					EM-DAT Classification
		Rail	641	598	43					EM-DAT Classification
		Road	13,652	2,465	11,187					EM-DAT Classification
		Water	3,136	1,325	1,811					EM-DAT Classification
Ecosystem	Environmental	Climate Change	16		16					DesInventar Data
		Deforestation	534		534					DesInventar Data
		Pollution	435		435					DesInventar Data

Deviant										
Family	**Event**	**Peril**	**MSGD**	**EM-DAT**	**DesInventar**	**UCDP**	**VDC-SY**	**GTD**	**UNHCR**	**Origination**
Aggression	Terror	Armed Assault	37,554					37,554		GTD Data
		Blockades/Barricades	835					835		GTD Data
		Bombing/Explosion	75,982		19			75,963		GTD Data
		Disappearances/Kidnapping	9,120		5			9,115		GTD Data
		Execution/Assassination	17,582					17,582		GTD Data
		Hijacking	556					556		GTD Data
		Infrastructure Destruction	8,854					8,854		GTD Data
		Terrorist Attack	5,500		15			5,485		GTD Data
		Unarmed Assault	828					828		GTD Data
	Conflict	Non-State	10,976			10,976				UCDP Classification
		Social Unrest	756		756					DesInventar Data
		State Attacks on Civilians	27,347		2	27,345				UCDP Classification
		State-based	111,992			89,943	22,049			UCDP Classification
Crime	Arson		228		228					DesInventar Data
	Loss/Degradation of Human Life		254		254					DesInventar Data
	Theft/Looting		318		318					DesInventar Data
Deprivation	Famine/Starvation		16	8	8					EM-DAT Data
	Malnutrition		155		155					DesInventar Data
Deracination	Displaced	Asylum-seekers	80,885						80,885	UNHCR Classification
		Internally Displaced Persons (IDPs)	21,811						21,811	UNHCR Classification
		Others of Concern	22,072						22,072	UNHCR Classification
		Refugees (incl. refugee-like situations)	102,484						102,484	UNHCR Classification
		Stateless Persons	21,992						21,992	UNHCR Classification
	Repatriated	Returned IDPs	21,609						21,609	UNHCR Classification
		Returned Refugees	27,588						27,588	UNHCR Classification
Unspecified			6	6						EM-DAT Data

NOTE: Over and above the 6 unspecified disasters in the Deviant Disaster group, there are 88 Unspecified disasters in MSGD from DesInventar (1991– 2012, 25 deaths, 1,115 people affected, no financial impact) that offer no information that could aid assigning them to any disaster group.

Appendix C: OTHER MSGD DATASETS

C.1: The Selection of other MSGD Datasets

Disaster Inventory System (DesInventar)

DesInventar started life in the mid-1990s when it was developed by the Network of Social Studies in the Prevention of Disasters, an NGO consortium in Latin America (La Red, 2017; UNDP, 2013). It was subsequently sponsored by the UNDP and UNISDR for implementation in developing countries across Latin America, Asia and Africa (UNISDR, 2017; UNDP, 2017a; UNDP, 2013; De Groeve et al., 2013). DesInventar is not a centrally curated database like EM-DAT, but a suite of open source tools, which includes a methodology and a customisable database with data management and analysis software (DesInventar.NET, 2017; DesInventar.ORG, 2017).

DesInventar is designed to be deployed by institutions, states, countries and regions to capture and analyse disaster loss data from a variety of sources, therefore DesInventar datasets were not historically categorised as global or international (Tschoegl et al., 2006; Velasquez et al., 2002). That said, the deployment of DesInventar to capture disaster data in numerous countries across continents and the availability of a consolidated 'global' dataset makes it a feasible source of disaster loss data to augment data sourced from EM-DAT (De Groeve et al., 2013).

The appropriateness of selecting DesInventar as an additional disaster loss data source alongside EM-DAT is reinforced by the knowledge that DesInventar and EM-DAT are both sources used for UNISDR's Global Assessment Report (GAR) on Disaster Risk Reduction (GAR/UNISDR, 2015). Furthermore DesInventar is cited as an important source of disaster losses in the guidance notes for the Hyogo Framework for Action (HFA) disaster risk reduction plan and in a 2013 study undertaken by the European Commission's Joint Research Centre (HYOGO, 2008; De Groeve et al., 2013).

Uppsala Conflict Data Program (UCDP)

The Uppsala Conflict Data Program (UCDP) is maintained by the Department of Peace and Conflict Research, Uppsala University, Sweden and is considered the main source of conflict data (UCDP, 2017a; UCDP, 2017b). Notably, the assertion that the UCDP is a credible and reputable source of conflict data is borne out by the use of its definitions of conflicts as de facto standards by various studies and reports (HSRP, 2013; Themnér and Wallensteen, 2014; Schuemer-Cross and Taylor, 2008; GHA, 2012; Gleditsch et al., 2017; SIPRI, 2016).

The Uppsala Conflict Data Program (UCDP) is therefore selected as the primary source for disasters that can be categorised as violent conflict (UCDP, 2017b; UCDP, 2017a). UCDP offers disaggregated conflict data for all countries other than Syria, which it excludes as the detail data available to the programme does not reach the clarity and consistency required to meet the standard applied to data sources for the UCDP (Croicu and Sundberg, 2015; Themnér and Wallensteen, 2014).

Violations Documentation Center in Syria (VDC-SY)

The absence of conflict data for Syria is more of a challenge to address. The UN abandoned updating its record of fatalities in the Syrian conflict in July 2013 because of difficulties in obtaining neutral verifiable accounts of victims (Themnér and Wallensteen, 2014; Pizzi, 2014). In recent years, however, the needs of the victims of the conflict in Syria have been placing considerable pressure on the humanitarian system (Fleming, 2015; Chonghaile, 2015) – this gap in data needs to be addressed for this work therefore. In investigating potential sources of data the following are examined and considered:

(a) Human Rights Data Analysis Group–Syria (HRDAG-Syria, 2017)
(b) Syrian Observatory for Human Rights (SOHR, 2017)
(c) Syrian Network for Human Rights (sn4hr, 2017)
(d) Damascus Center for Human Rights Studies (DCHRS, 2017)

(e) Syrian Center for Statistics and Research (CSR-SY, 2017)
(f) Violations Documentation Center in Syria (VDC-SY, 2016b)
(g) Syrian Shuhada Revolution Martyr Database (Syrian Shuhada, 2017)

HRDAG – (a) above – has used data sourced from (b) through (f), as well as from the Syrian Government, to compile reports on deaths caused by the Syrian conflict for UNHCR and Amnesty International (Price et al., 2016, 2014). Notably, all but two of these sourced datasets had to be given to HRDAG either directly by the data curator or via UNHCR. This is because (b), (c) and (d) do not make their source data available, they provide just charts and reports based on the data; eliminating these as sources of data for this study. Similarly, HRDAG, present the results of its analysis, but does not make accessible any of the data used for this analysis (HRDAG-Syria, 2017); thus also eliminating this as a source of data for this work.

Examination of (e) CSR-SY, identified pages of victims that could be scraped (CSR-SY, 2017), but apart from an identifying number, location and date, all other data are in Arabic, making data veracity evaluation for this study infeasible. The Syrian Shuhada Revolution Martyr Database is also set aside as it cites as its primary data sources to be (d) and (f) above (Syrian Shuhada, 2017; VDC-SY, 2016b; DCHRS, 2017). Also, it recognises only the deaths of those martyred in the revolutionary cause and as such contains fewer entries than (f) VDC-SY, as a result it provides an explicitly one-sided view of the conflict in Syria.

After completing this process of elimination (f) VDC-SY remains the only dataset that is accessible, and (relatively) credible. Additionally VDC-SY has been used in studies for UNHCR and Amnesty International. Established in 2011, the Violation Documentation Center in Syria is an NGO that monitors and documents violations in human rights in Syria (VDC-SY, 2016b). It details, down to the individual, those who are missing, detained or killed as a result of the conflict (ibid).

Global Terrorism Database (GTD)

To provide a perspective of humanitarian crises caused by terrorist attacks the Global Terrorism Database (GTD) is selected (GTD, 2017e). This is an open source database of domestic, transnational and international terrorist events from around the world spanning a period of around 45 years. The database is curated and made available by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), which is a US Department of Homeland Security Center of Excellence headquartered at the University of Maryland (START, 2017). It is also a credible source dataset used by numerous studies and reports (Pawlak and Dietrich, 2015; Nowrasteh, 2016; Nemeth, 2013; Young and Dugan, 2011).

United Nations High Commissioner for Refugees (UNHCR)

A final type of disaster identified as missing from view is the disruption and uprooting of communities from their homes and/or countries through conflict, famine or other forms of hazards. While the event that caused their inability to remain in their home, locale or country may in and of itself be a disaster, their plight as refugees, asylum-seekers or internally displaced persons (IDPs) is also a trigger for humanitarian assistance. The data for these people are obtained from the United Nations High Commissioner for Refugees (UNHCR) Persons of Concern population statistics (UNHCR, 2017c).

C.2: DesInventar (Acquire, Prepare, Examine)

Two sites are found for DesInventar: DesInventar.NET and Desinventar.ORG (DesInventar.NET, 2017; DesInventar.ORG, 2017). As literature and research can often refer to either of the two sites, both sources are examined and compared before DesInventar.NET is selected as the only source of DesInventar data to be added to the MSGD. (De Groeve et al., 2014; Wattegama, 2007; Below et al., 2010; Guha-Sapir and Hoyois, 2012). The reasons for this are:

- The DesInventar.ORG site offers only 30 country-specific disaster datasets whereas, DesInventar.NET provides access to 78 region/state/country/city databases as well as a large consolidated multi-country database.
- Quantifiable values that allow calculations for total deaths and total affected are predominantly answered simply with the word 'YES'. For example, the first dataset retrieved, Argentina, with 19,513 entries from 1970 to 2009, of which 7 variables should be quantifiable variables defining the human effect of each event – Deaths, Missing, Wounded, Victims, Affected, Evacuees and Relocated, but 17,318 entries (almost 89% of the dataset) contained the word 'YES' for one or more of these fields.
- For numerous countries (e.g. Ecuador, Iran) datasets from DesInventar.NET contain more entries than their equivalents from DesInventar.ORG.

In the end data is acquired from 79 DesInventar.NET databases. These datasets are prepared, examined and evaluated using the DVf toolset then added to the MSGD (DesInventar, 2017a).

(a) Acquiring the data

The selected DesInventar site allows downloads as a suite of file downloads but issues with spurious character sets and mismatched fields rendered this method problematic (DesInventar, 2017a). Data is therefore obtained using the more time-consuming method of sequentially running queries and creating downloadable xls/csv files for each of the following 79 DesInventar databases (ibid):

- One international consolidated database, GAR2015, named to correspond with the UNISDR 2015 issue of Global Assessment Report on Disaster Risk Reduction (GAR/UNISDR, 2015)
- One multi-country database for the Pacific Islands
- Seventy-three country specific databases – Albania, Angola, Antigua and Barbuda, Argentina, Barbados, Belize, Bhutan, Bolivia, Cambodia, Chile, Colombia, Comoros, Costa Rica, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Grenada, Guatemala, Guinea, Guyana, Honduras, Indonesia, I.R. Iran, Jordan, Kenya, Laos, Lebanon, Liberia, Madagascar, Maldives, Mali, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Myanmar, Nepal, Nicaragua, Niger, Pakistan, Palestine, Panama, Paraguay, Peru, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Senegal, Serbia, Seychelles, Sierra Leone, Spain, Sri Lanka, Swaziland, Syria, Timor Leste, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, Uruguay, Venezuela, Vietnam, Yemen, Zanzibar
- Three individual databases for three states in India – Orissa, Tamil Nadu and Uttarakhand
- One database for a city in Nigeria – Ibadan Metropolis

It should be noted that the single international dataset, GAR2015, cannot be used in isolation as it does not cover all years and countries individually listed on the site (DesInventar, 2017a).

(b) Preparing the data

The DesInventar methodology and open source solution provides wide latitude and flexibility in implementation and usage (DesInventar.NET, 2017). This means that each implementation of DesInventar can take a different shape, making consolidation and comparison across databases a considerable challenge. This challenge is heightened by local language usage; limited, if any, validation of data; and individual interpretations of what each field of data is expected to contain (Guha-Sapir and Hoyois, 2012). After considerable scrutiny and effort the 1,178,753 entries and 1,592 uniquely named variables acquired from DesInventar.NET are reduced to 322,489 entries and 22 variables that have the potential to populate the MSGD [*Figure C.2-1*].

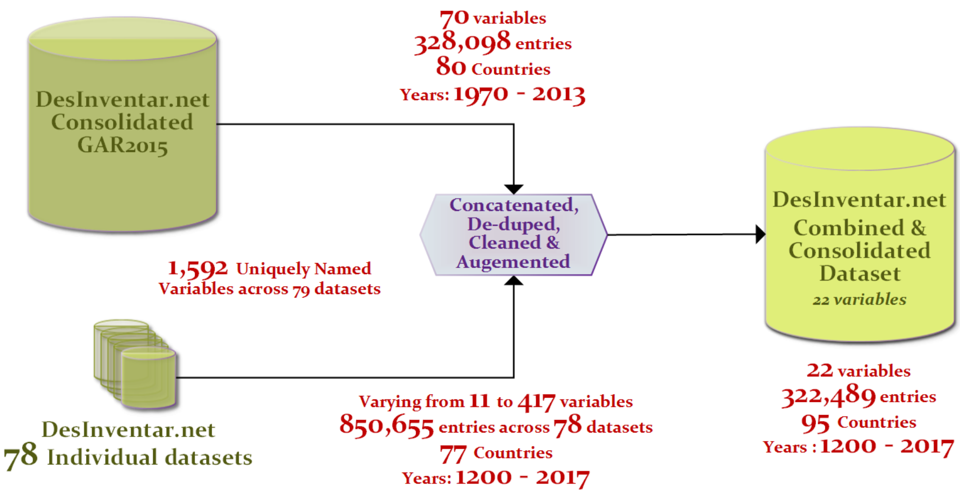


Figure C.2-1: Building a DesInventar dataset to map to MSGD

This rationalisation process involved:

- Eliminating any entries that are duplicates of EM-DAT entries.
- Removing spurious, unintelligible entries, these are entries that:
    - Contain unrecognisable characters and words that defy translation; or
    - Are mostly blank; or
    - Offer no viable information i.e. no type of disaster, when it occurred and what it affected.
- Discarding irrelevant variables. – 1,475 variables are ignored as they appeared only once each in different databases. Another 69 variables are found in less than 50% of the databases. Of the remaining 48 variables, only those that potentially hold information about *when*, *what*, *where* and to *what human or financial effect* are retained.
- Eliminating duplicates and entries sourced from EM-DAT.

Preparation of the remaining 322,489 entries is as follows:

Year (of occurrence)

Year is extracted from the date field (Date (YMD)) where possible, but even after the elimination of spurious records, 80 entries have no date values. For these a reference to the Year of occurrence is found elsewhere (e.g. Serial, Cause, Comments) and manually entered.

Country

Correction are made to 6,091 entries for errors in country referencing and ISO codes (IS0-3166, 2017). Most of these errors are for states referenced as countries (e.g. Tamil Nadu instead of India) and incorrect ISO codes for countries such as Cambodia, Timor-Leste, Morocco and Zanzibar.

Disaster Classification

DesInventar entries are mapped to the foundational Master Disaster Classification model (MDC) which at this stage is based on IRDR disaster classifications with additions to support EM-DAT [*Section 6.3.4* & *Appendix B.2*] (IRDR, 2014; Guha-Sapir et al., 2017g; DesInventar, 2017c). As DesInventar documentation states its classification models are based on IRDR, aligning DesInventar to the MDC is not expected to be problematic, however (DesInventar, 2017c; IRDR, 2014):

- Only 80,392 entries have an Event that match an IRDR originated classification in the MDC and another 27,665 had values in Cause, Description of Cause or Comments that matched IRDR originated classifications. In total, only around a third of the entries aligned to the IRDR convention.
- For the remaining 214,432 entries all the Event, Cause, Description of Cause and Comments fields are scrutinised and interpreted, *often after translation from a local language*, before a classification is assigned as a result:

- 95,225 entries are assigned an MDC disaster classification that originated from IRDR;
- 66,045 entries are assigned an MDC disaster classifications that originated from EM-DAT;
- 88 entries representing, 25 deaths and people and 1,115 people affected, are classed as '*unspecified*' events;
- 41 entries are assessed as acts of conflict or terror, but left unclassified pending the inclusion of other conflict and terror related data;
- 53,033 DesInventar entries requiring the creation of 21 MDC classifications in order to maintain visibility of nuanced distinctions in disaster types [*Table C.2-1, Appendix B.2*]:

Group	Family	Event	Peril	MSGD
Naturogenic	**Biological**	Disease	Epizootic	**1,467**
			Meningitis	**261**
		Infestation	Pest	**15**
		Sickness	Gastrointestinal	**23,060**
			Respiratory	**366**
	Geophysical	Mass Movement	Landslide	**17,347**
	Hydrological	Wave Action	Alluvion	**453**
Anthropogenic	**Accident**	Industrial	Electrical	**576**
			Mining Accident	**185**
		Miscellaneous	Drowning	**3,037**
			Electrical/Gas/Chemical	**421**
			Missing Persons	**92**
			Poisoning: Food/Water	**3,057**
	Ecosystem	Environmental	Climate Change	**16**
			Deforestation	**534**
			Pollution	**435**
Deviant	**Aggression**	Conflict	Social Unrest	**756**
	Crime	Arson	.	**228**
		Loss/Degradation of Human Life	.	**254**
		Theft/Looting	.	**318**
	Deprivation	Malnutrition	.	**155**

Table C.2-1: MDC Classifications (DesInventar)

Disaster Effect – Human

To maintain consistency with EM-DAT, DesInventar values for Deaths and Missing are added to create a new variable, Total Deaths (Guha-Sapir et al., 2017i). Similarly, all other human effect values – injured, victims, affected, relocated and evacuated – are summed to create the new variable Total Affected. It should be noted that a number of entries in DesInventar simply acknowledge that some human effect, not how much, is caused by displaying a tick [*Figure C.2-2*]. Additionally, while this information is visible on the Query screens, it is lost when downloading.

Deaths	Injured	Missing	Houses Destroyed	Houses Damaged	Victims
	72				
					✓
					✓
				✓	✓
					✓
				✓	✓
	6				
	10				
	✓				✓
			✓		✓
	✓		✓	✓	1000

Figure C.2-2: Extract from DesInventar.NET query view (Argentina)

Disaster Effect – Financial

Two of the most prevalent financial loss variables in DesInventar are:

- **Losses $USD**, which is found in 54 DesInventar databases and populated with non-zero values in 3,571 entries, 1.1% of the 322,489 retained DesInventar entries. These values are retained for the MSGD after being adjusted to 2015 levels using USA CPI and converted to $000s to remain consistent with figures held for EM-DAT (Guha-Sapir et al., 2017h; BLS, 2016).
- **Losses $Local**, which is found in 65 databases and populated with non-zero values in 42,707 entries, 13.2% of the 322,489 retained DesInventar entries. This variable is not added to MSGD, as:
    - In the case of 1,804 entries spanning 42 countries that coincided with non-zero values in Losses $USD, it is not clear if these value are equivalent or incremental to Losses $USD, or whether the approach taken varies by country.

- To facilitate comparison these values would need to be converted. As these entries span 68 countries and range from 1905 to 2017, the identification of acceptable average exchange rates for each year and currency combination is not feasible. Especially as currency conversion is further confused in countries where historic devaluations, or transitions to new currencies, have occurred

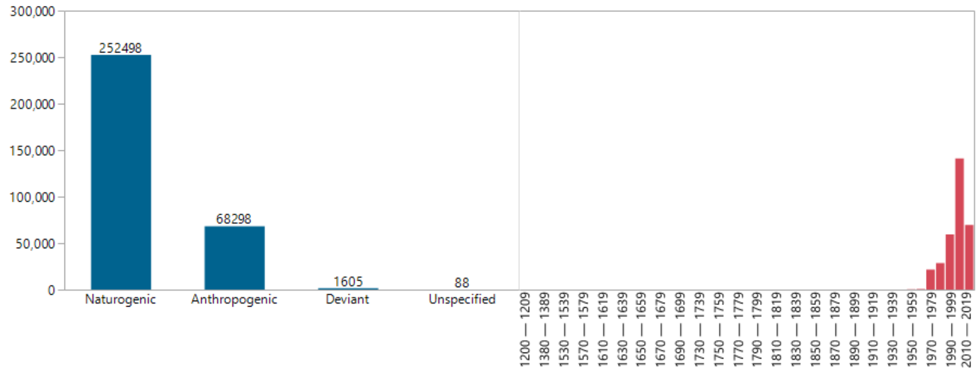
Note: the **Losses $Local** variable is retained in the consolidated and prepared DesInventar dataset, which can be referenced from MSGD as and when required.

Finally, as with EM-DAT, a unique identifier field is created for all DesInventar entries to allow a reference link to remain between the MSGD and a clean version of the original dataset.

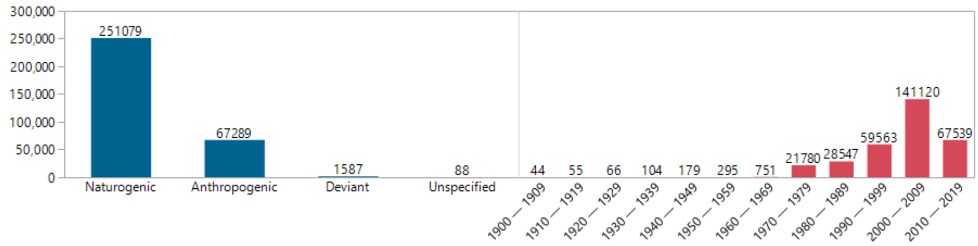
(c) Examining the data

DesInventar, like EM-DAT, predominantly holds information of disasters that originate in nature, grouped as 'naturogenic' disasters in the MDC. Although the date range of the dataset extends over centuries, 1200 – 2017, the vast majority of the entries, 298,973 or 92.7%, are for events that occurred from 1980 onwards. Notably, the emphasis on naturogenic disasters remains unwaveringly at over 78% even if the time period examined is one that coincides with EM-DAT (1900 – 2015) or even the mostly richly populated EM-DAT time period (1990 – 2015) [*Figure C.2-3*].

DesInventar: 1200 – 2017 *(full dataset)*



DesInventar: 1900 – 2015 (to coincide with EM-DAT's date range)



DesInventar: 1990 – 2015 (to coincide with EM-DAT's most populated date range)

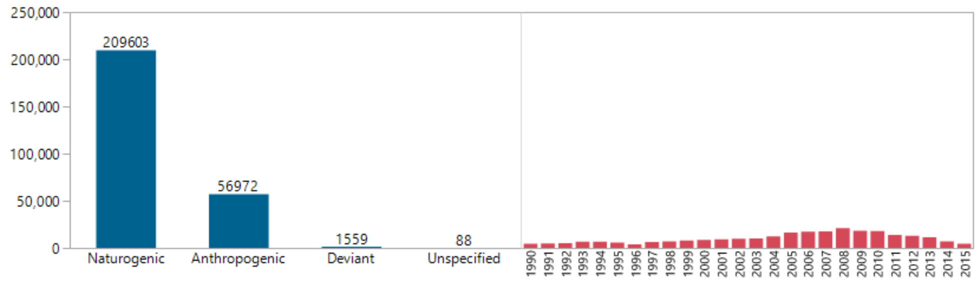
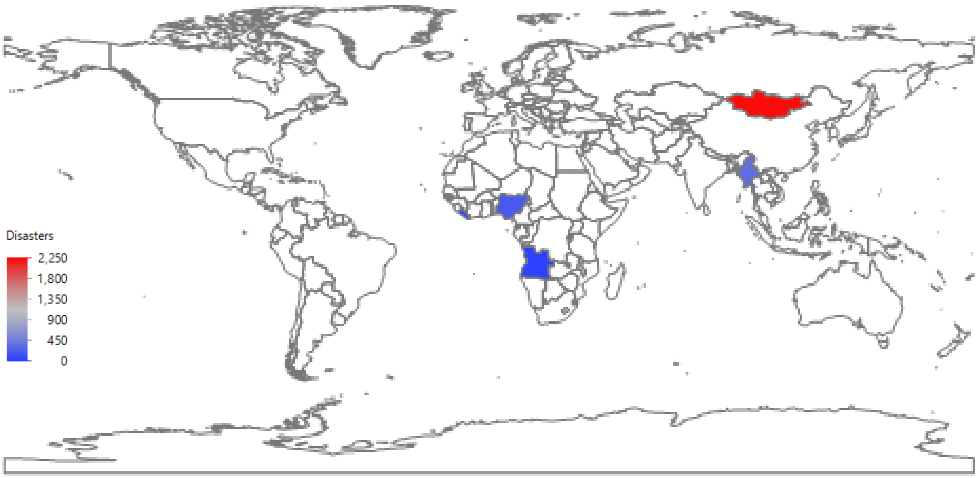


Figure C.2-3: DesInventar by Disaster Groups & by Decades

Even after eliminating numerous spurious entries, 5,404 zero-effect disasters remain in the dataset. This may be because these entries are simply unquantified 'place marks'; *or* used to record information in one or more of the 1,570 variables not retained for analysis. These 'empty' entries spanned nine countries, and for three of these countries – Rwanda, Mongolia and Dominican Republic – the 'empty' entries represent over 66% of their contribution to the MSGD [*Figure C.2-4*].



Country	Empty Entries	Entries in Dataset	Empty	# of Variables
Rwanda	277	323	86%	41
Mongolia	2,275	3,174	72%	108
Dominican Republic	1,392	2,112	66%	121
Barbados	908	2,199	41%	50
Equatorial Guinea	21	119	18%	33
Myanmar	328	2,977	11%	417
Nigeria	186	2,427	8%	48
Liberia	4	77	5%	32
Angola	13	349	4%	92

Figure C.2-4: DesInventar Countries with 'empty' entries in MSGD

Another point of note is that 33,823 entries, spanning 80 countries, have a zero or missing value in all the human effect variables – Deaths, Missing, Injured, Victims, Affected, Relocated and Evacuated – with India's 11,696 'no human effect' entries (8,420 of which are for Tamil Nadu) representing almost 35% of this number [*Figure C.2-5*].

There is also the question of the coverage of the disaster records. Visually scanning a chart of the years in which disasters are recorded for each country, the picture looks varied and sporadic [*Figure C.2-6*]. It is unclear if this chart truly depicts the occurrence of disasters. It may be more a reflection of when individual DesInventar databases started and stopped; local disaster recording practices and inclusion criteria; and/or retrospective updates

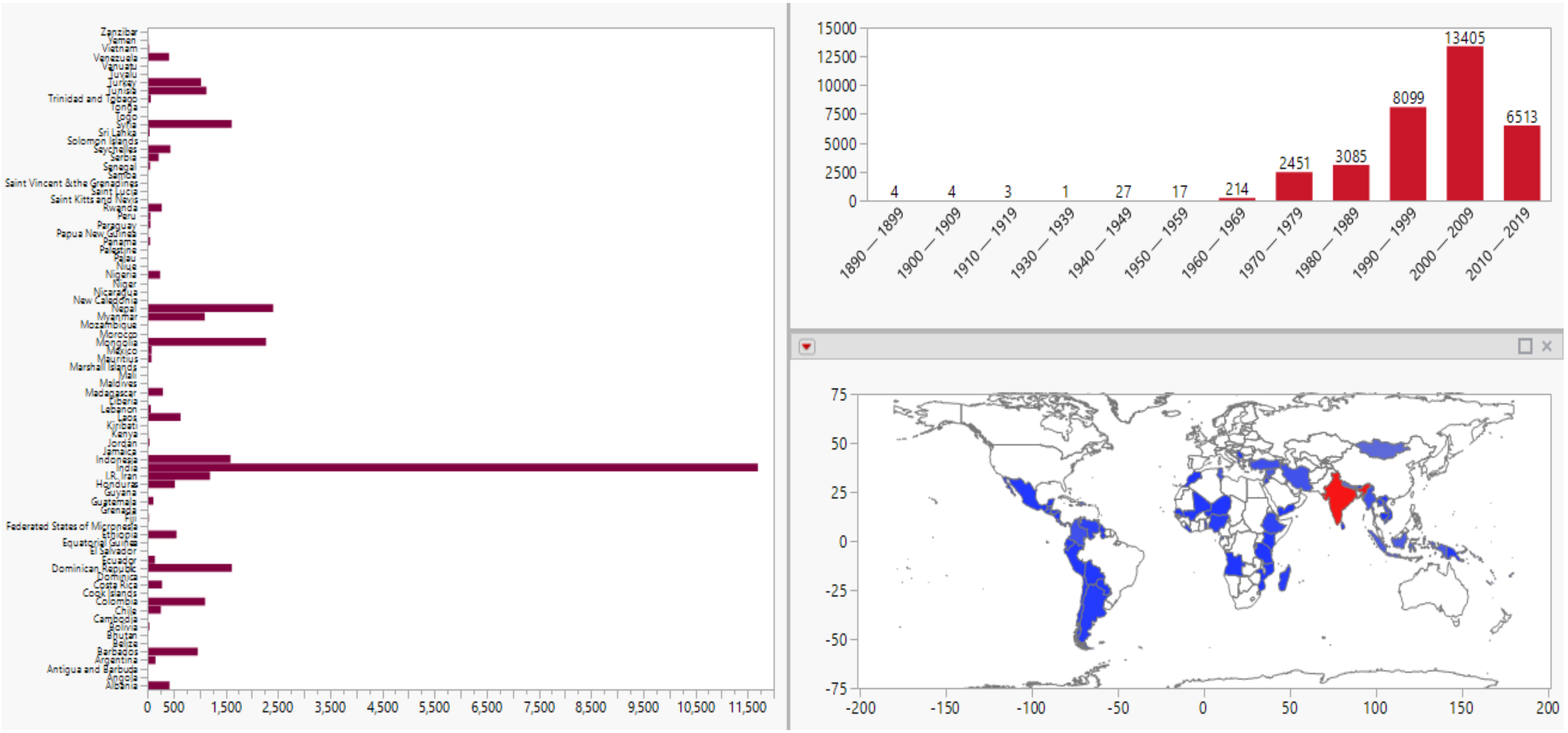


Figure C.2-5: DesInventar - No Human Effect Entries

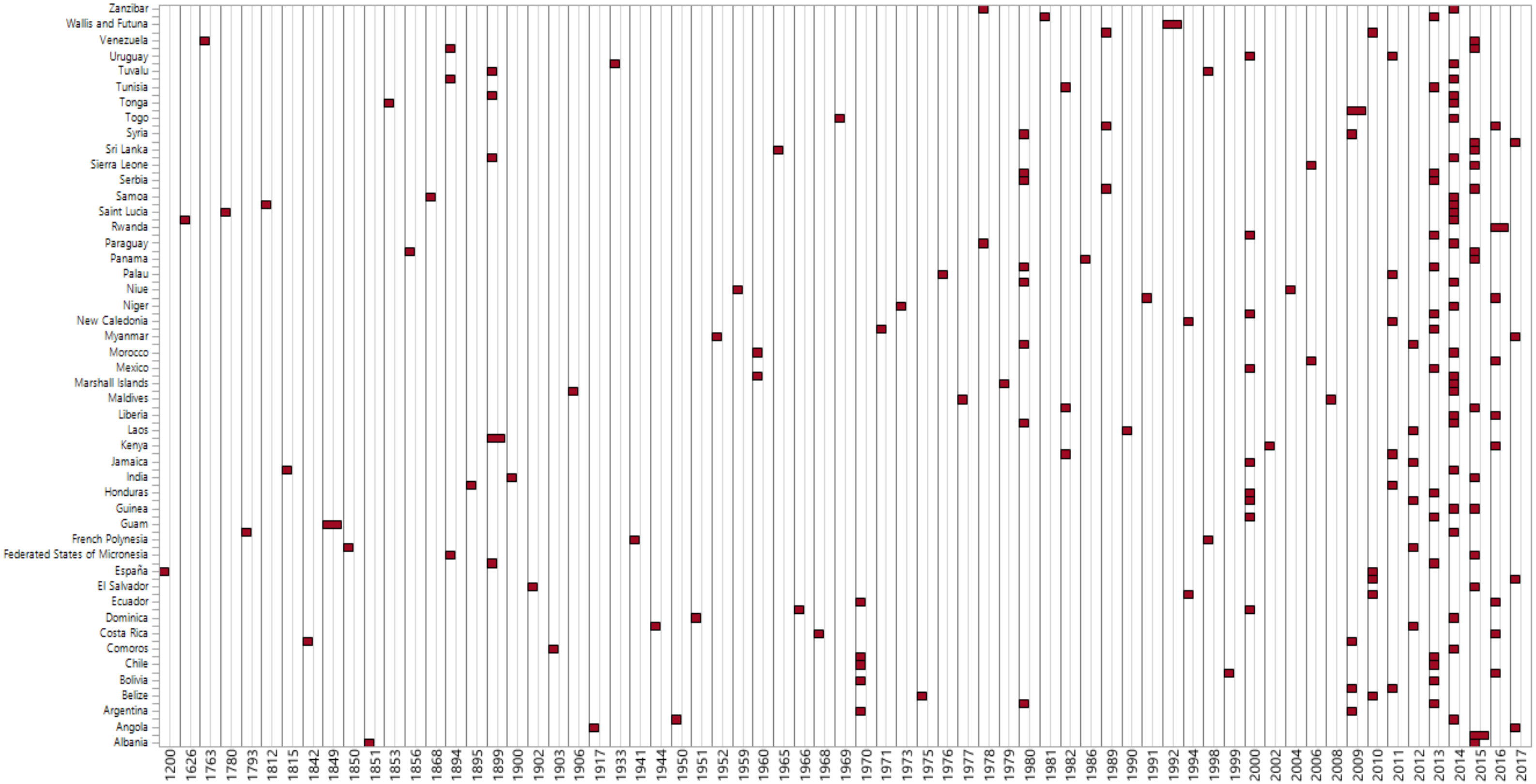


Figure C.2-6: DesInventar: Years in which Disasters are Recorded

C.3: UCDP GED50 (Acquire, Prepare, Examine)

Over fourteen sets of conflict data from the Uppsala Conflict Data Program (UCDP) are made available by the Department of Peace and Conflict Research of Uppsala University, Sweden (UCDP, 2017d; UCDP, 2017a). These datasets range from those that hold detailed conflict data (down to the date and geolocation) to those that are either aggregated by year or segregated by conflict type or reference (e.g. geography, actor or conflict). The most granular dataset, UCDP's 2015 Georeferenced Event Dataset Global version 5.0 (**GED50**), is selected for this study.

(a) Acquiring the data

All of the UCDP datasets are very easy to acquire and are made available for download in multiple formats. The UCDP datasets also come equipped with comprehensive downloadable documentation, referred to as c*odebooks*. The GED50 dataset obtained from the site contained 128,264 entries (UCDP, 2017d).

(b) Preparing the data

The database did not require a lot of cleaning per se and appears to be meticulously maintained with adherence to standards. The primary focus of preparing UCDP for consolidation into the MSGD is to ensure the country identifiers are compatible, assign appropriate disaster classifications, and identify the relevant fatality variables.

Country

4,573 values in the isocc field are outdated and had to be corrected. Of these 3,661 still referenced to the Soviet Union and 912 referenced Yugoslavia.

Disaster Classification

UCDP categorises its conflict data using an integer – 1, 2 or 3 – in its type_of_violence variable [*Table C.3-1*].

1	**state-based conflict**	'This is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year. "Armed conflict" is also referred to as "state-based conflict", as opposed to "non-state conflict", in which none of the warring parties is a government.'
2	**non-state conflict**	"The use of armed force between two organised armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year. "
3	**one-sided violence**	"The use of armed force by the government of a state or by a formally organised group against civilians which results in at least 25 deaths in a year. Extrajudicial killings in government facilities are excluded."

Table C.3-1: UCDP, Types of Violence
(UCDP, 2017c)

As no classifications in the MDC are appropriate to the types of events recorded in UCDP, 3 new entries are created under the '*Deviant*' group of disasters (*Appendix B.2.3*). *Figure C.3-1* illustrates the distribution of entries across these new MDC classifications across the three datasets acquired to this point, i.e. EM-DAT, DesInventar, UCDP (Guha-Sapir et al., 2017l; DesInventar.NET, 2017; UCDP, 2017b).

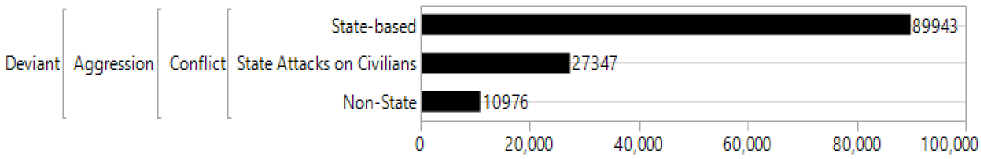


Figure C.3-1: MDC Classifications created to accommodate UCDP

Notably, one of the new UCDP originated additions to MDC classifications, State Attacks on Civilians, is also applicable to 2 of the 41 DesInventar entries left unclassified pending the inclusion of more conflict related disaster data.

Disaster Effect - Human

UCDP records only fatalities and contains 7 variables to distinguish between the deaths it records, such as how many deaths on each 'side' of the conflict and how many civilians are killed. For this study only total fatality numbers are of interest, therefore the follow two variables are considered relevant (Croicu and Sundberg, 2015):

best_est The best (most likely) estimate of total fatalities resulting from an event – always the sum of deaths of combatants on both sides (**deaths_a** and **deaths_b**) as well as civilian and unknown deaths (**deaths_civilians** and **deaths_unknown**).

high_est The highest reliable estimate of total fatalities integer

If **best_est** contains a non-zero value it is used to populate **Total Deaths** in the MSGD, if not, then high_est is used. Of note is that 2 entries for Columbia 2001 contained zero for all fatality fields, these entries were retained in MSGD for reference but flagged for exclusion during analysis. Finally, as with previously added disaster datasets, a unique identifier field is created to link between the MSGD and a cleaned version of the original dataset.

(c) Examining the data

Examination of the data is carried out acknowledging UCDP GED50 is a detailed dataset of fatalities caused by conflict, therefore it does not contain information of people affected by conflict. Points of note include:

- The 126,264 entries obtained in the GED50 represents conflicts in 115 countries with the highest number recorded for Afghanistan (20,306 entries), India (14,045 entries) and Iraq (5,891 entries).
- The top three countries with the highest total fatalities versus those with highest mean fatalities differ noticeably [*highlighted in Table C.3-2*]. For the USA 3 entries in GED50 for 9/11, recorded as 2,986 deaths in total, positions it as the highest average losses per incident above Rwanda, which has over half a million deaths attributable to conflict.

Total	Mean Fatalities	Total Fatalities	# of GED50 Entries
USA	995	2,986	3
Rwanda	933	549,728	589
Afghanistan	10	198,191	20,306
Ethiopia	119	197,355	1,663
Eritrea	456	19,172	42

Table C.3-2: UCDP Top 3 Countries – Mean vs Total Fatalities

- The number of conflicts recorded in UCDP have risen over the years to 7,304 in 2015 [*Figure C.3-2*], but the 4,213 conflicts in 1994 resulted in the highest number of deaths, 592.992 in total, in any one year. Remarkably this outlier year for fatalities related to conflict does not appear to be attributable to any one conflict of exceptional scale [*Figure C.3-3*].

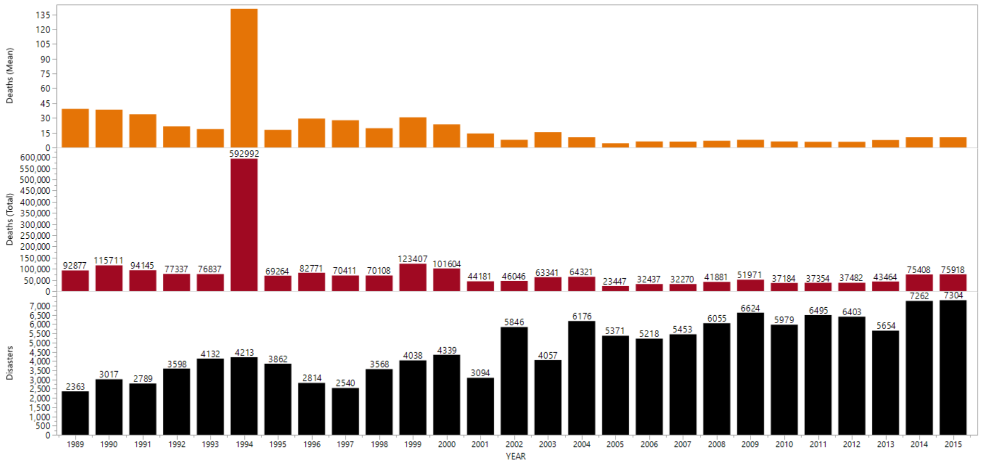


Figure C.3-2: UCDP - Disasters and Deaths by Year

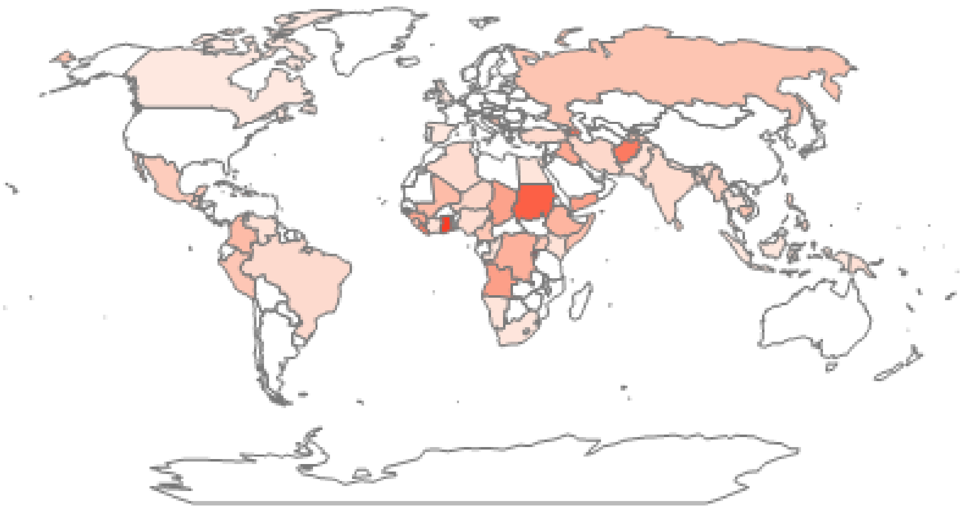


Figure C.3-3: UCDP 1994 - Map of Fatalities

C.4: VDC-SY (Acquire, Prepare, Examine)

The VDC-SY database is designed to help identify individuals killed, missing or detained during the conflict in Syria. VDC-SY does not provide mechanisms to download data for analysis, therefore web-scraping techniques are employed (VDC-SY, 2016b; VDC-SY, 2016a).

(a) Acquiring the data

Web-scraping techniques are used to obtain VDC-SY four datasets – Killed, Missing, Detainees and Regime's Casualties – via four views of the data (VDC-SY, 2016a). The 226,341 entries are obtained in total, with the spread of entries per dataset, each representing a victim, as follows:

Killed:	138,375 entries
Regime Fatalities:	21, 078 entries
Missing:	2,640 entries
Detainees:	64,248 entries

(b) Preparing the data

The four VDC-SY datasets are inconsistent with the other disaster entries obtained so far, in that there is one record per person and each datasets signifies a different human cost. These datasets are therefore consolidated and summarised to show all human effects that are the result of an incident on the same day in the same Syrian province as follows:

Step 1: Anonymise and concatenate the four VDC-SY datasets flagging deaths and affected.

Deaths: Killed, Regime Fatalities, and Missing (***NO*** Date Found)

Affected: Detainees and Missing (***WITH*** Date Found)

Step 2: Summarise by Date and Province

The resultant dataset has 22,049 entries. Other preparation includes:

- **Year** *(of occurrence)* is extracted from the Date.
- **Country** is set to the ISO 3166 Country Code SYR (for Syria)
- **Disaster Classification** is set to the UCDP originating MDC class:

Group	Family	Event	Peril
Deviant	Aggression	Conflict	State-based

Finally, each of the 22,049 summary entries is assigned a unique identifier to enable a link between the MSGD and a cleaned VDC-SY dataset for future look-up and reference.

(c) Examining the data

First, examining the data for missing information:

- Dates are missing for 758 of the 226,341 detail entries. This means that 16 summary entries, totalling 162 deaths and 596 people affected, cannot be attributed to a year.
- 6,009 detail entries are for incidents that took place at an unidentified or unknown location, rolling-up to 988 summary entries and equating to 1,611 deaths and 4,398 people affected.
- Although VDC-SY is a detailed database, painstakingly recording names of victims and even location of birth, 14,881 entries have victim as a blank or documented as 'unidentified'.

Looking at the spread of incidents over time, it becomes obvious that VDC-SY primarily holds records of the victims of the current conflict in Syria, as 225,276 of the 226,341 (~99.7%) of the original entries are for the period 2011-2016 [*Figure C.4-1*].

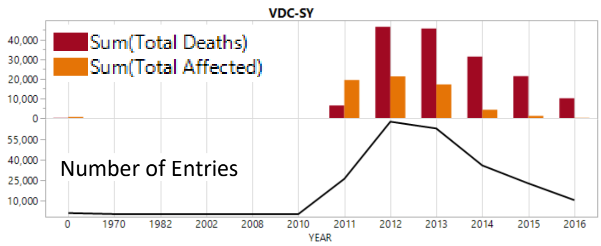


Figure C.4-1: VDC-SY by YEAR

Also noticeable is the in excess of 6:1 ratio of non-regime deaths versus regime fatalities [*Figure C.4-2*].

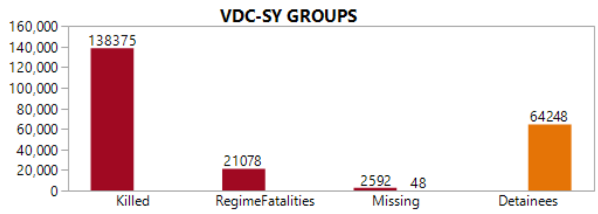


Figure C.4-2: VDC-SY Deaths and Affected by VDC-SY Groups

C.5: GTD (Acquire, Prepare, Examine)

The Global Terrorism Database (GTD) from the National Consortium for the Study of Terrorism and Responses to Terrorism (START) is a collection of domestic, transnational and international terrorism related data and with entries dating back to 1970 (GTD, 2017e; START, 2017). GTD is open source and made readily available through the completion of a simple form. It is a well-documented database with detailed descriptions of its fields and data collection methodology (GTD, 2017b; GTD, 2017d).

(a) Acquiring the data

The GTD data download process provides a full set of data. The version of the dataset used by this study (dated: June 2016) contains 156,772 entries and 142 field variables. It provides considerable detail for each incident, including sources of its data and the reasoning behind selection of values used for its variables.

(b) Preparing the data

This dataset, as with UCDP GED50, does not require much in the way of 'cleaning' and also appears to strictly adhere to its own standards. The focus here, therefore, is to understand, select and align whichever of the 142 GTD variables are pertinent to this work and can be used to populate the MSGD.

Country

Two GTD fields identify the country affected by the terrorist attack, country (a numeric) and country_txt (the name of the country). The numeric variable of country does not refer to numeric ISO 3166 country identifiers, but to GTD's own country numeric identification scheme (IS0-3166, 2017; GTD, 2017b). GTD's country_txt field is therefore used to identify each country's ISO 3166 code.

For 23.365 entries for 53 countries, the country_txt did not match ISO country names. In most cases this was because of small differences in naming conventions, e.g. GTD contains **Ivory Coast** where the ISO refers to **Côte d'Ivoire**. In some instances, however, the name of the country is outdated, e.g. **Rhodesia** is now **Zimbabwe**. For these 53 countries all 23,265 were manually updated.

Disaster Classification

Disaster classifications already defined in the MDC were not applicable to the incidents described in GTD therefore 9 new classifications were created in the MDC based on the contents of GTD field attacktype1 [*Figure C.5-1*]. For 5 entries attacktype1 is **Unknown** but information in attacktype2 facilitated MDC classification.

	GTD attacktype1	GTD attacktype2	MDC Event	MDC Peril	GTD Entries
1	Assassination		Terror	Execution/Assassination	17,582
2	Armed Assault		Terror	Armed Assault	37,554
3	Bombing/Explosion		Terror	Bombing/Explosion	75,963
4	Hijacking		Terror	Hijacking	556
5	Hostage Taking (Barricade Incident)		Terror	Blockades/Barricades	835
6	Hostage Taking (Kidnapping)		Terror	Disappearances/Kidnapping	9,115
7	Facility/Infrastructure Attack		Terror	Infrastructure Destruction	8,849
8	Unarmed Assault		Terror	Unarmed Assault	828
9	Unknown		Terror	Terrorist Attack	5,485
		Facility/Infrastructure Attack	Terror	Infrastructure Destruction	5

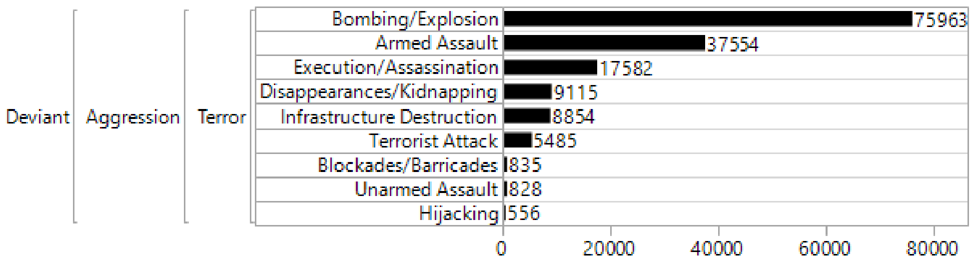


Figure C.5-1: GTD Attacks and MDC Classifications

Furthermore, the creation of MDC classifications to accommodate GTD enabled the 39 pending assignment of classifications to DesInventar entries to be completed [*Figure C.5-2*]:

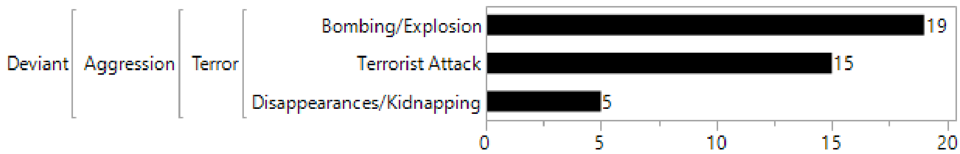


Figure C.5-2: GTD originated MDC classifications assigned to 39 DesInventar entries

Disaster Effect – Human

The two human effect fields Total Deaths and Total Affected needed the MSGD are calculated as follows:

- Total Deaths = nkill, which is the total all killed in the incident, victim and attackers. Furthermore, GTD incident fatalities are cross-checked with UCDP incidents to eliminate double-counting. The logic applied is to compare incidents that are recorded in both datasets to have occurred on the same day in the same country; as a result 59,466 overlapping entries are found. For these entries if GTD deaths are fewer or equal to UCDP fatalities, GTD fatalities are not populated in the MSGD; if they are greater than UCDP fatalities, the difference between the two numbers is used as GTD fatalities.
- Total Affected = nwound + nhostkid, which totals of all injured and/or kidnapped in each incident. If nhostkid is unknown (the field has the value -99), then nhostkid is set to 0.

Note: GTD does not restrict the human effect values of terror attacks to integers. Where there are linked incidents, human effect is distributed across entries, sometimes resulting in fractional values in these fields.

Disaster Effect – Financial

The loss variables of propvalue, the value of damaged property, and ransomamt, the amount of ransom demanded are held in GTD in US dollars based on the year in which the incident occurred. These financial effects of terrorism must be made consistent with other financial losses held in the MSGD, which are in US$1000s adjusted to 2015 levels. This entails adding propvalue and ransomamt; adjusting the total to US 2015 values using the USA CPI (BLS, 2016); and then dividing the resultant value by 1000.

(c) Examining the data

The GTD database appears to be strictly maintained, with each entry containing considerable nuanced information about each event, but some information is still missing or not definitive:

- There are no entries for 1993 [*Figure C.*5-3]. This is because hard copy data for 1993 was lost when the data was donated by the Pinkerton Global Intelligence Services (PGIS) who maintained this terror data from 1990 to 1997 (PGIS, 2017). The missing 1993 data is explained on the GTD site (GTD, 2017d).
- The fluctuation in GTD data also reflects changes in data collection methods, which differs for the periods 1970 – 1998, 1998 – 2008, 2008 – 2012 and 2012 – 2015 [*Figure C.5-3*].

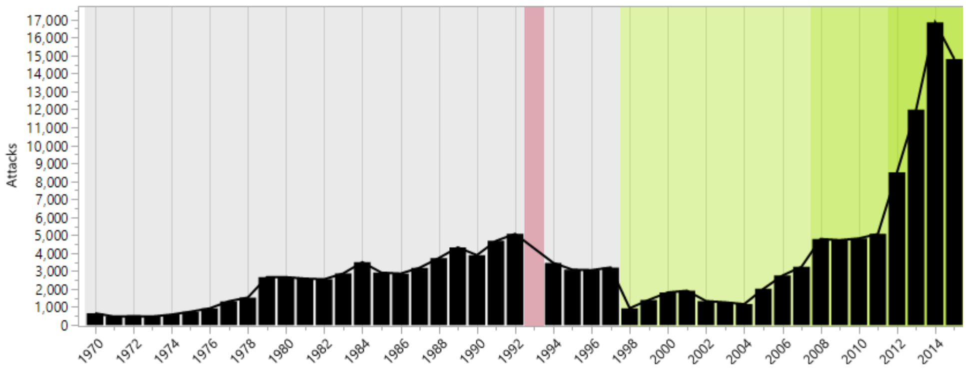


Figure C.5-3: GTD Terror Attacks from 1970 to 2015

- Excluding GTD vs UCDP double-counting values that are zeroed a further 55,732 GTD entries are found to be 'empty' in that there is no recorded effect, human or financial, of the terror attack. This is because GTD includes entries for failed attacks and is used to capture a broad spectrum of information e.g. weapon type, target type, ransom notes etc. (GTD, 2017d). As only variables relevant to losses are needed for this study, many of GTD's 142 variables are set aside and the variables that remain (relating to effect) may therefore legitimately be 'empty'.

- For 5,490 entries the type of terrorist attack is not known. These entries are spread over 132 countries and can be found in most years. They represent around 15,557 deaths with 158 people wounded or kidnapped [*Figure C.5-4*]. Of these only five entries could be identified to any degree using the secondary variable used to hold the description of attack. Notably, just over 34% of these unknown attacks occurred in the period 2013 to 2015.

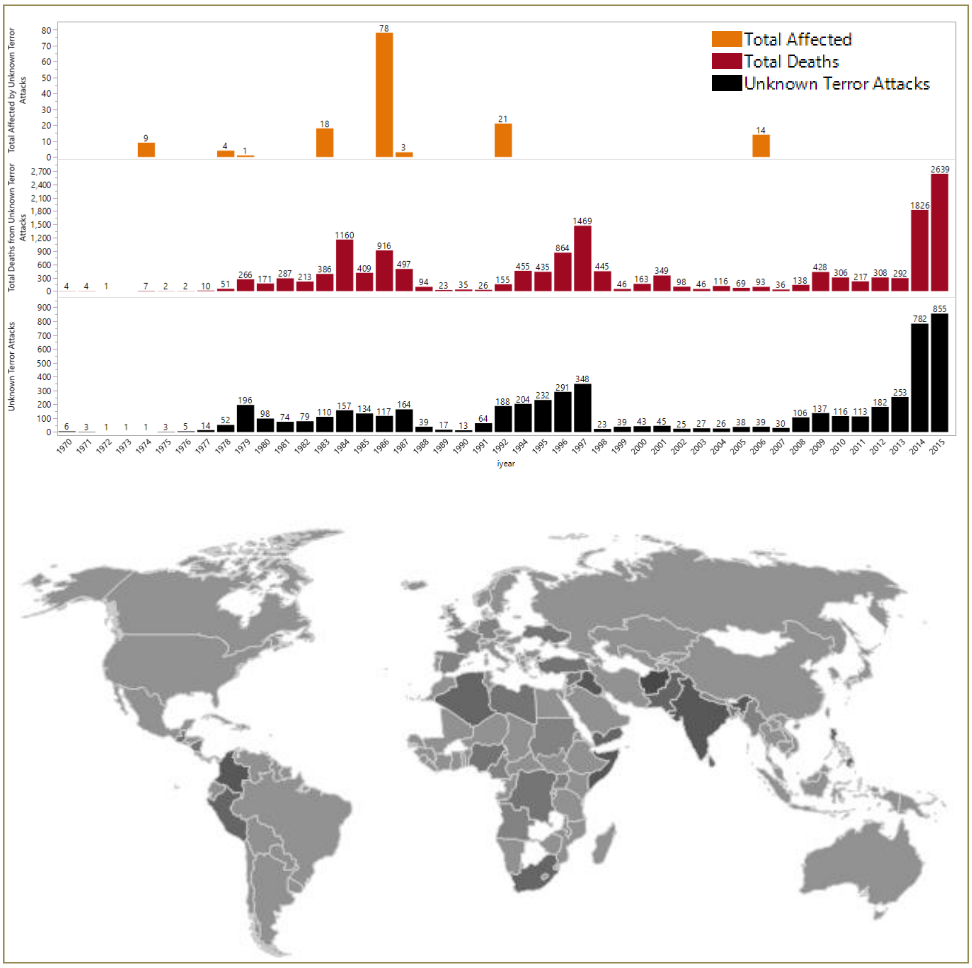


Figure C.5-4: GTD - Unknown Terror Attacks

- GTD contains more entries for deaths than for people affected. If GTD disaster losses are charted [*Figure C.5-5*], the numbers show 223,779 deaths, 28,233 people affected and over $23.5bn of losses, spread over 9 disaster classifications.

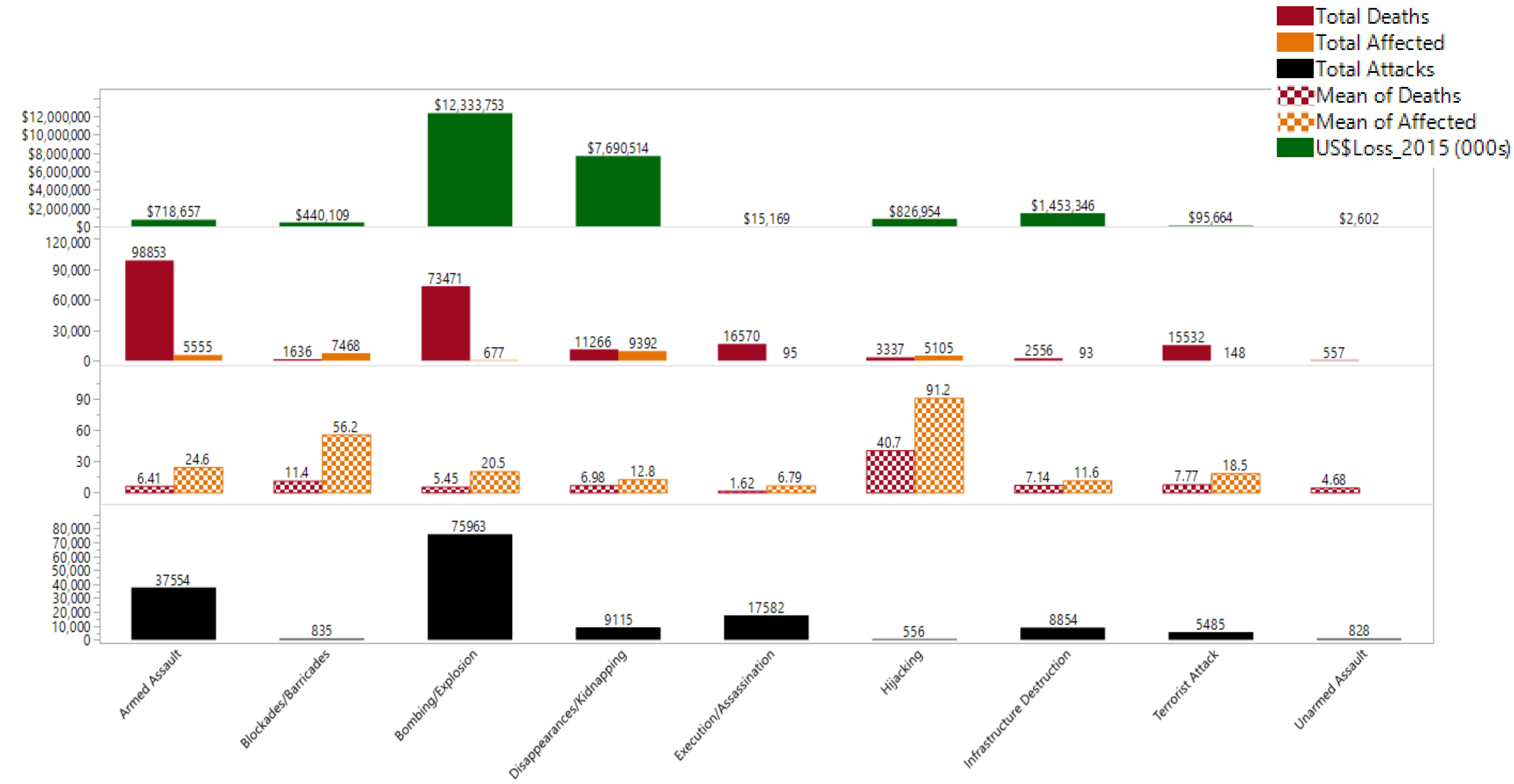


Figure C.5-5: GTD Disaster & Losses by Disaster Classification

C.6: UNHCR (Acquire, Prepare, Examine)

The United Nations High Commissioner for Refugees (UNHCR) maintains and provides access to a statistics database containing information about the persons/populations of concern (PoC) based on their *UNHCR PoC status* [*Table C.6-1*] (UNHCR, 2017c):

Refugees	Those recognised in accordance with the UNHCR Statute; individuals granted complementary forms of protection; or those enjoying temporary protection. Includes people in a refugee-like situation.
Asylum-seekers	Individuals who have sought international protection and whose claims for refugee status have not yet been determined.
Internally displaced persons (IDPs)	People who have been forced to leave their homes or places of habitual residence, in particular as a result of, or in order to avoid the effects of armed conflict, situations of generalised violence, violations of human rights, or natural or man-made disasters, and who have not crossed an international border.
Stateless persons	Persons who are not considered as nationals by any State under the operation of its law; they do not possess the nationality of any State.
Others of concern	Individuals who do not necessarily fall directly into any of the other groups, but to whom UNHCR extends its protection and/or assistance services, based on humanitarian or other special grounds.
Returned refugees	Former refugees who have returned to their country of origin spontaneously or in an organised fashion but are yet to be fully integrated.
Returned IDPs	IDPs who were beneficiaries of UNHCR's protection and assistance activities and who returned to their areas of origin or habitual residence during the year.

Table C.6-1: UNHCR Persons of Concern (PoC) Status
(UNHCR, 2017c)

It is this information that is used here to fill the gap in the MSGD pertaining to people **affected** by conflict and terrorism, where currently most of the conflict-related data is for fatalities.

(a) Acquiring the data

UNHCR provides access to its population statistic through a set of seven 'select and search' pages (UNHCR, 2017c). Of these only the *Time Series* download best suited this study (UNHCR, 2017d). This enabled a download of 298,441 entries from 1951 to 2016 including variables – Year, Country/territory of asylum/residence (*aka* Destination), Origin (*country of*), PoC Status, *and* Value (*aka* Number of People). It should be noted that, with annualised subtotals of populations per PoC status, this is the most aggregate of the humanitarian crises datasets acquired so far. For this reason each of the six other UNHCR population statistics data downloads were explored for opportunities to add granularity to this *Time Series* data [*Table C.6-2*] (UNHCR, 2017c).

UNHCR Page	Description	Reason Not Suited
Persons of Concerns	Same information as Time Series page, but each UNHCR PoC is a variable.	Would need restructuring.
Demographics	Breakdown of PoCs by sex, age and location in destination country, with no information of country of origin.	This level of details is not required and data only collected since 2000.
Asylum-Seekers (Refugee Status Determination)	Details numbers of individuals going through each stage of the refuge determination process.	This level of details is not required. Data is specific to asylum-seekers and only collected since 2000.
Asylum-Seekers (Monthly)	Breakdown of asylum applications lodged detailing origin and asylum country for each monthly total	This level of details is not required. Data is specific to asylum-seekers and 44 destination countries.
Resettlement	Data sourced from Governments about refugees (with or with UNHCR assistance) settling in their country each year and from which country.	Data excludes humanitarian admissions and specific to UNHCR PoC of refugee only.
Mid-Year Statistics	Data collected during the first half of the last full calendar year.	Only a small snapshot of data is available.

Table C.6-2: UNHCR Populations Statistics – Other Downloads

Examining the six other UNHCR data download pages, it becomes obvious that although the population statistics database holds more detail than made available via the *Time Series* dataset, this detail is neither complete nor is it reconcilable with the *Times Series* data. Therefore, a search for supplementary UNHCR downloads is exhausted and the study continues with the *Time Series* dataset.

(b) Preparing the data

UNHCR's *Time Series* population statistics dataset is lean in the variables it offers, but some preparation is still needed to allow this data to be added to MSGD.

Country

Each UNHCR *Time Series* data entry contains two country fields – Origin and Country/territory of asylum/residence. For the purposes of this research the MSGD the Origin field is considered the country where the event that caused the uprooting of the population occurred. The Country/territory of asylum/residence is considered the Destination country of the uprooted population. Destination is retained in recognition of the possibility that there may be a need for assistance in coping with refugees by these host countries.

Country information in the original dataset is in name format only, therefore ISO 3166 country codes are added to ensure alignment with other entries in the MSGD (IS0-3166, 2017). Due to differences in naming conventions, the ISO 3166 country code for 39 Origin countries (59.942 entries) and 34 destination countries (50,428 entries) cannot be automatically updated, therefore these entries are individually checked and updated with the correct ISO 3166 country code (ibid).

Disaster Classification

It can be argued that the deracination of populations is not in and of itself a disaster, but typically caused by disastrous events. Here it is deemed that the humanitarian aid-requiring needs of uprooted people is justification to classify the plight of UNHCR PoCs as a form of humanitarian crises; hence the creation of seven new entries, one for each UNHCR PoC status [*Table C.*6-3 & *Figure C.6-1*].

Group	Family	Event	Peril	UNHCR Entries
Deviant	Deracination	Displaced	Refugees (incl. refugee-like situations)	102,484
Deviant	Deracination	Displaced	Asylum-seekers	80,885
Deviant	Deracination	Repatriated	Returned Refugees	27,588
Deviant	Deracination	Displaced	Others of Concern	22,072
Deviant	Deracination	Displaced	Stateless Persons	21,992
Deviant	Deracination	Displaced	Internally Displaced Persons (IDPs)	21,811
Deviant	Deracination	Repatriated	Returned IDPs	21,609

Table C.6-3: MDC/UNHCR Classifications

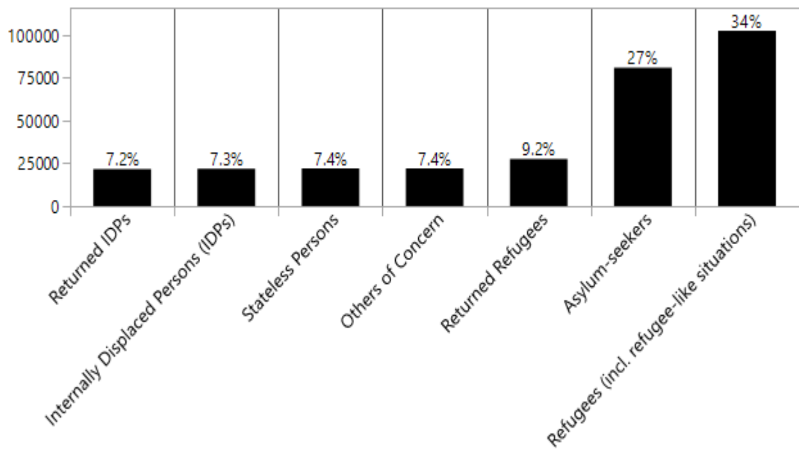


Figure C.6-1: UNHCR Distribution of Entries for MDC Perils

Of note is that classifications for returnees – IDPs and refugees – are created in the MDC. The reason for this is that, while repatriation implies resolution of the events that would have caused this population to be uprooted, there may be aid required to enable these individuals to settle and re-assimilate on their return, therefore excluding this group entirely from visibility via the MDC is not deemed appropriate. To avoid double-counting, returning populations are excluded from analysis of disaster effects. Of the 49,197 UNHCR entries for returning refugees and IDPs, 42,324 entries are empty and another 51 were marked by an asterisk, which is used to ensure anonymity when fewer than 4 people are involved. The remaining 6,822 entries equate to a population of 62,341,754 returnees that are excluded from the analysis of disaster effects [*Figure C.6-2*].

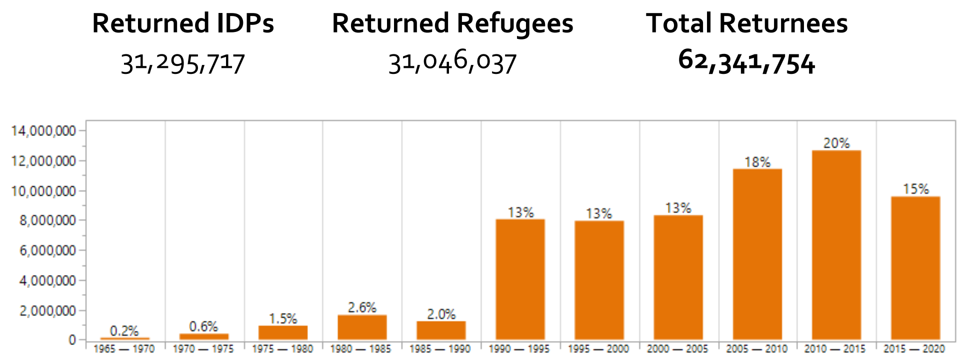


Figure C.6-2: UNHCR: Total Population of Returnees

Total Affected

This is the number of people quantified in each UNHCR *Time Series* dataset entry. This field is labelled Value in the downloaded dataset. Expected to be of type numeric, 4,208 entries for 2016, the latest year in the dataset, contain an asterisk in this field to maintain anonymity if 4 or fewer people are involved. Asterisked entries are not reflected in the totals used by UNHCR. As the likelihood of 4 or fewer people a year being displaced because of a humanitarian crises is considered low, all 4.208 entries are flagged for exclusion.

(c) Examining the data

A number of anomalies are identified with the UNHCR data:

Empty Entries

Of the 298,441 UNHCR entries obtained, 129,177 have zero values for the movement of people (Total Affected). No explanation for this is found in the documentation. Visualising the occurrence of these empty entries by UNHCR PoC status, year or geography provides no obvious answers [*Figure C.6-3* & *Figure C.6-4*]. It is, however, noticeable that empty entries occur least for PoC status Asylum-seekers and Refugees and for the first time in 2007, rising significantly from 2014 and most are for destinations in North America or Europe. These entries are also flagged for exclusion.

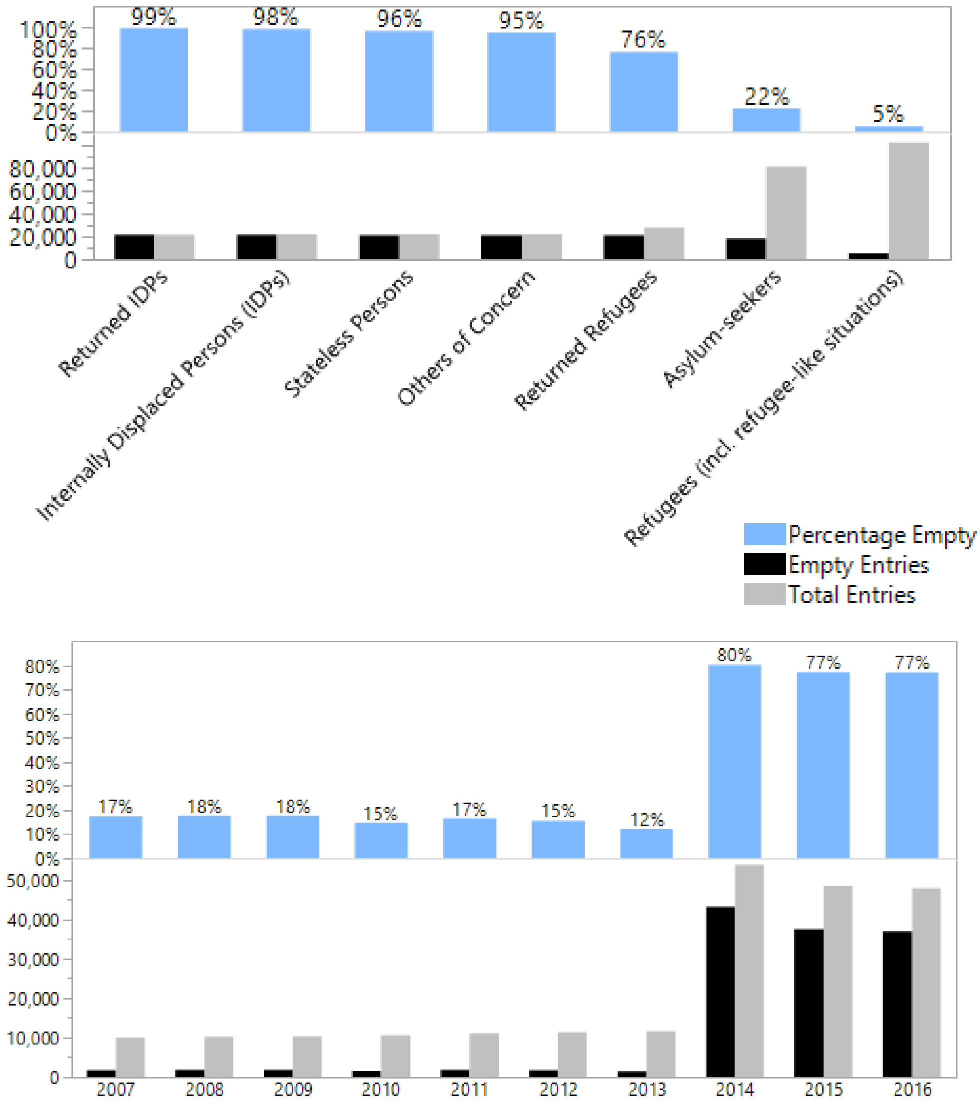


Figure C.6-3: UNHCR Empty Entries by UNHCR PoC Status and by Year

Note: a single entry with a spurious negative one value of people movement was also excluded found in the data – 2013 – from Gambia to Angola – Asylum-seeker – minus one person – This entry does not appear on the UNHCR web page, but only in the data download.

ORIGIN: 221 Countries

Empty Entries: TOP 10 ORIGINS		
1	Syrian Arab Rep.	2,202
2	Somalia	2,084
3	Sudan	1,958
4	Iraq	1,897
5	Eritrea	1,862
6	Ethiopia	1,859
7	Dem. Rep. of the Congo	1,817
8	Nigeria	1,803
9	Pakistan	1,773
10	Iran (Islamic Rep. of)	1,743

DESTINATION:192 Countries

Empty Entries: TOP 10 DESTINATIONS		
1	United Kingdom	3,449
2	United States of America	3,120
3	France	2,911
4	Canada	2,892
5	Italy	2,850
6	Germany	2,681
7	Belgium	2,642
8	Australia	2,612
9	Netherlands	2,485
10	Spain	2,471

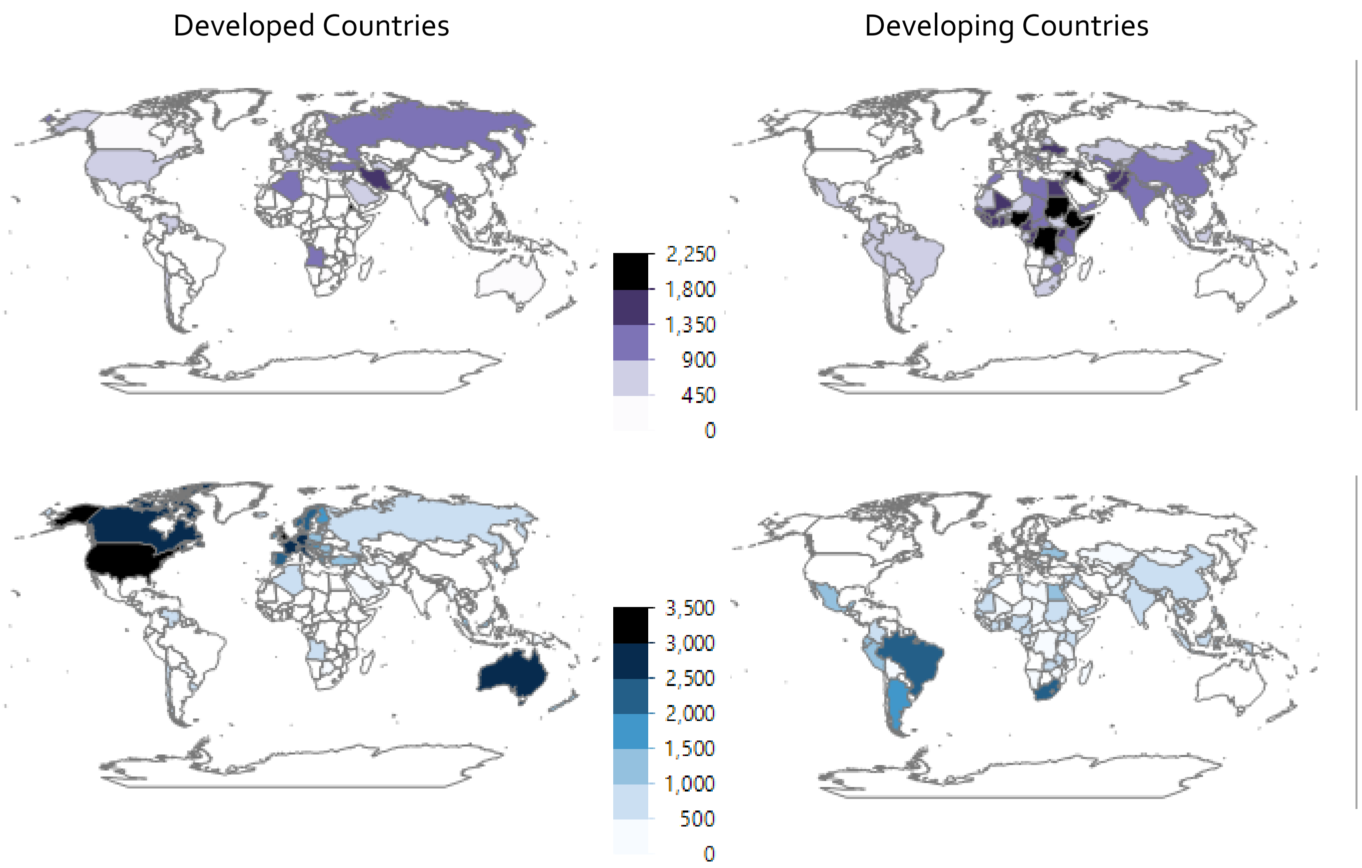


Figure C.6-4: UNHCR Empty Entries (Origin/Destination – Developed/Developing)

Fewer than 10 People from One Country in One Year

Of the 158,233 entries not excluded from analysis, 47% register the movement of single individuals, or groupings of fewer than 10 people [*Figure C.6-5*].

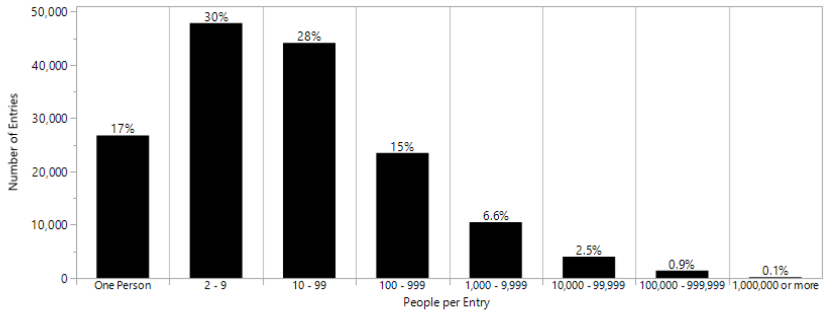


Figure C.6-5: UNHCR Scale of PoCs per Entry

The question is then – of the 47% (74,550) entries for the movement of small groups of people, how many can realistically be assumed to represent the uprooting of people as a result of disasters? Or is it a more realistic assumption that the movement of larger groups of people is likely to be because a significant event made remaining untenable? 952 entries representing only one person in one year and from one country are identified. These entries are marked for exclusion in order to minimise the risk of interpreting them to be a sign of crisis in their country of origin.

Nevertheless, examining these 952 solo entries is interesting. The highest volume are for destinations USA (128 people) and Italy (108 people). The Origin countries however show no real patterns other than many are often from the same country one a year over sequential years [*Figure C.6-6*].

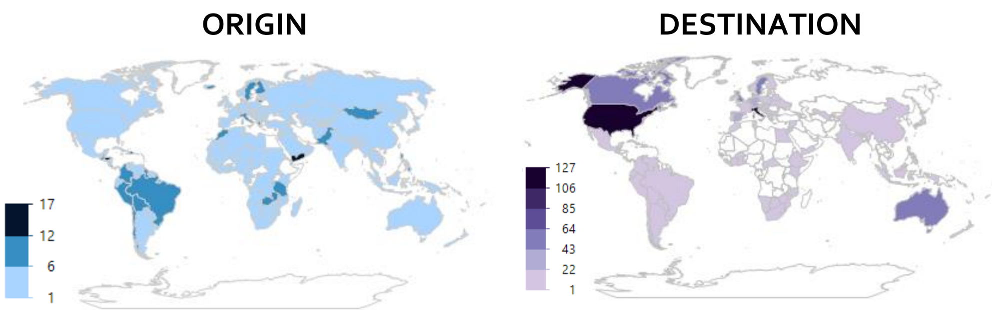


Figure C.6-6: UNHCR Movement of one person per year

Further suspect entries are found and marked a ‘soft’ as follows:

- 9,499 one-person entries that represent less than 10 people from the same Origin country in any one year.
- 1,727 low value entries, equating to 6,356 people, representing less than 10 people from the same Origin country in any one year.

Origin Unknown/Various

Of the now remaining 157,281 entries (not excluding from analysis), 3,565 entries the country of origin is Various/Unknown. These entries of ambiguous origin equate to 89,871, 341 people with most of these people (~73%) destined for Northern America or Western Europe [*Figure C.6-7*].

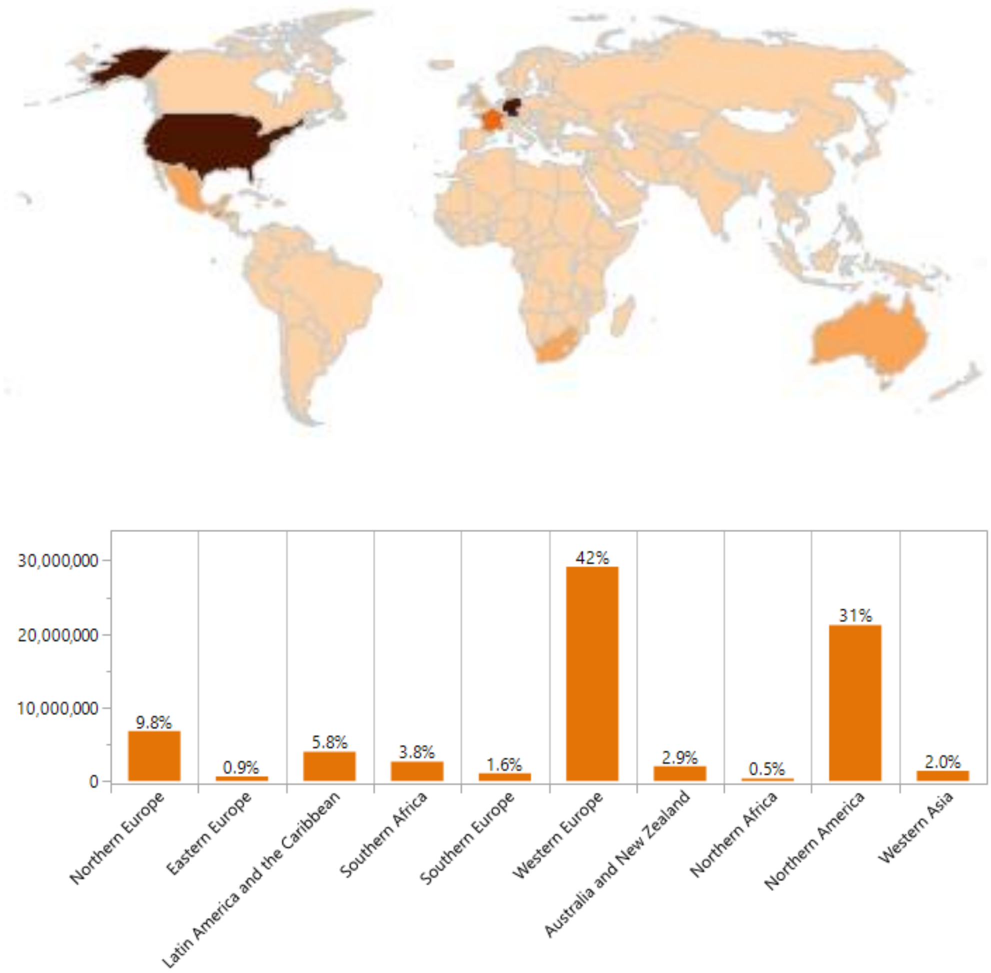


Figure C.6-7: UNHCR Destination of People of Unknown/Various Origin

In excess of 61% of these entries of Unknown/Various origin are dated before the year 2000 and represent over 88% of people in this grouping. The remaining 38%+ entries are after 2000 and represent over 11% of the people of Unknown/Various origin [*Figure C.6-8*].

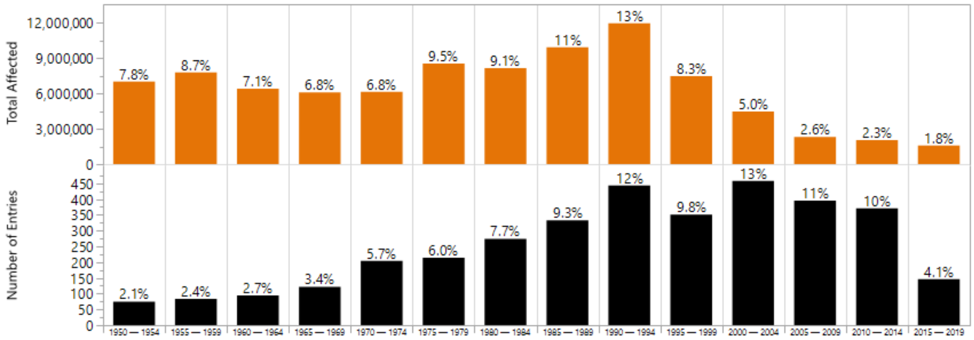


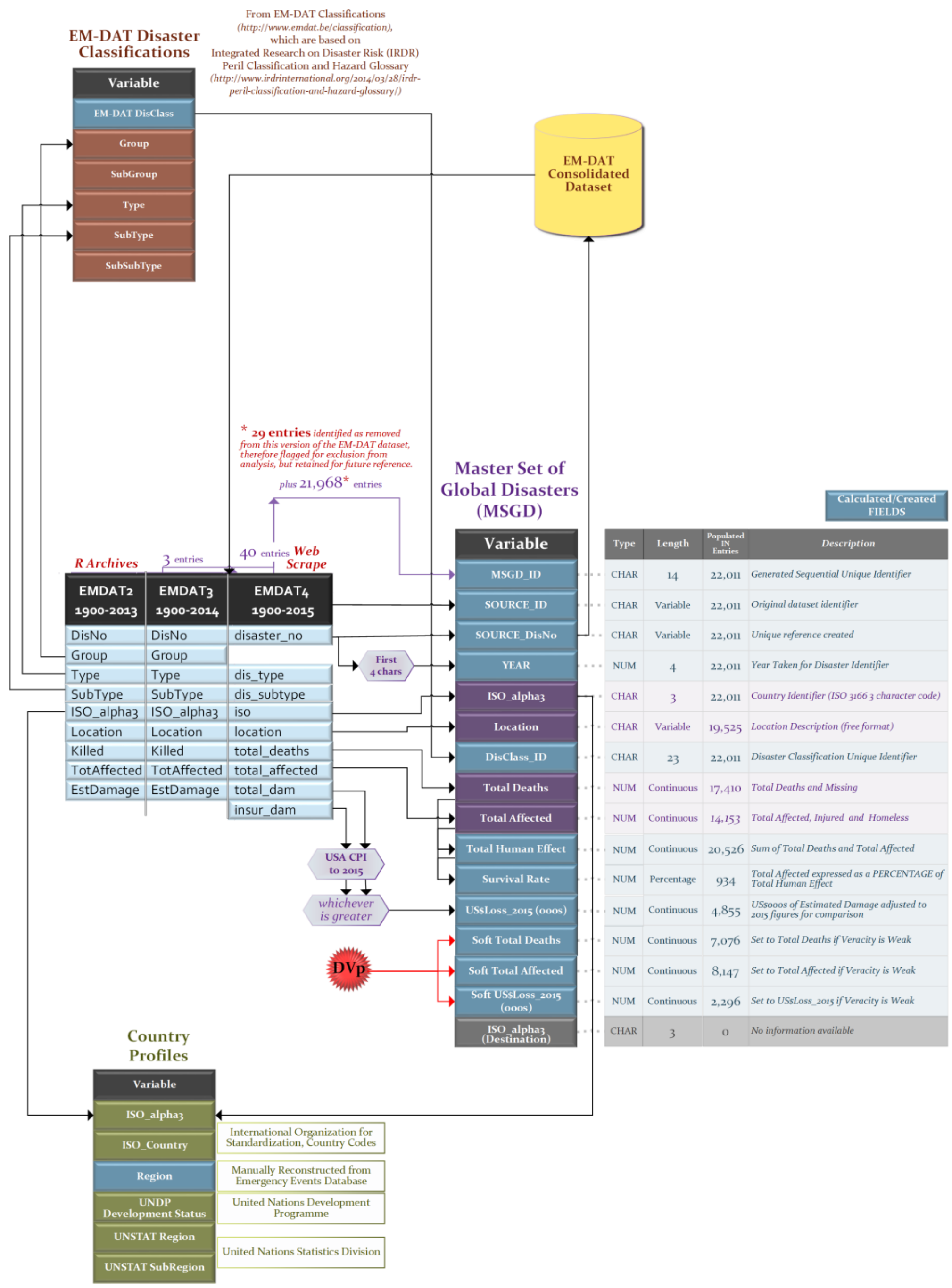
Figure C.6-8: UNHCR, People and Entries of Unknown/Various

The spread and extent of these entry/people numbers raises the possibility that these entries are less definitive. They may be a form of ‘catch-all’ used when information is less certain. Regardless, these entries are considered suspect and flagged as ‘soft’ data.

Appendix D: DATA SOURCE MAPPINGS TO MSGD

D.1: EM-DAT → MSGD

D.1.1: EM-DAT Variables Mapped

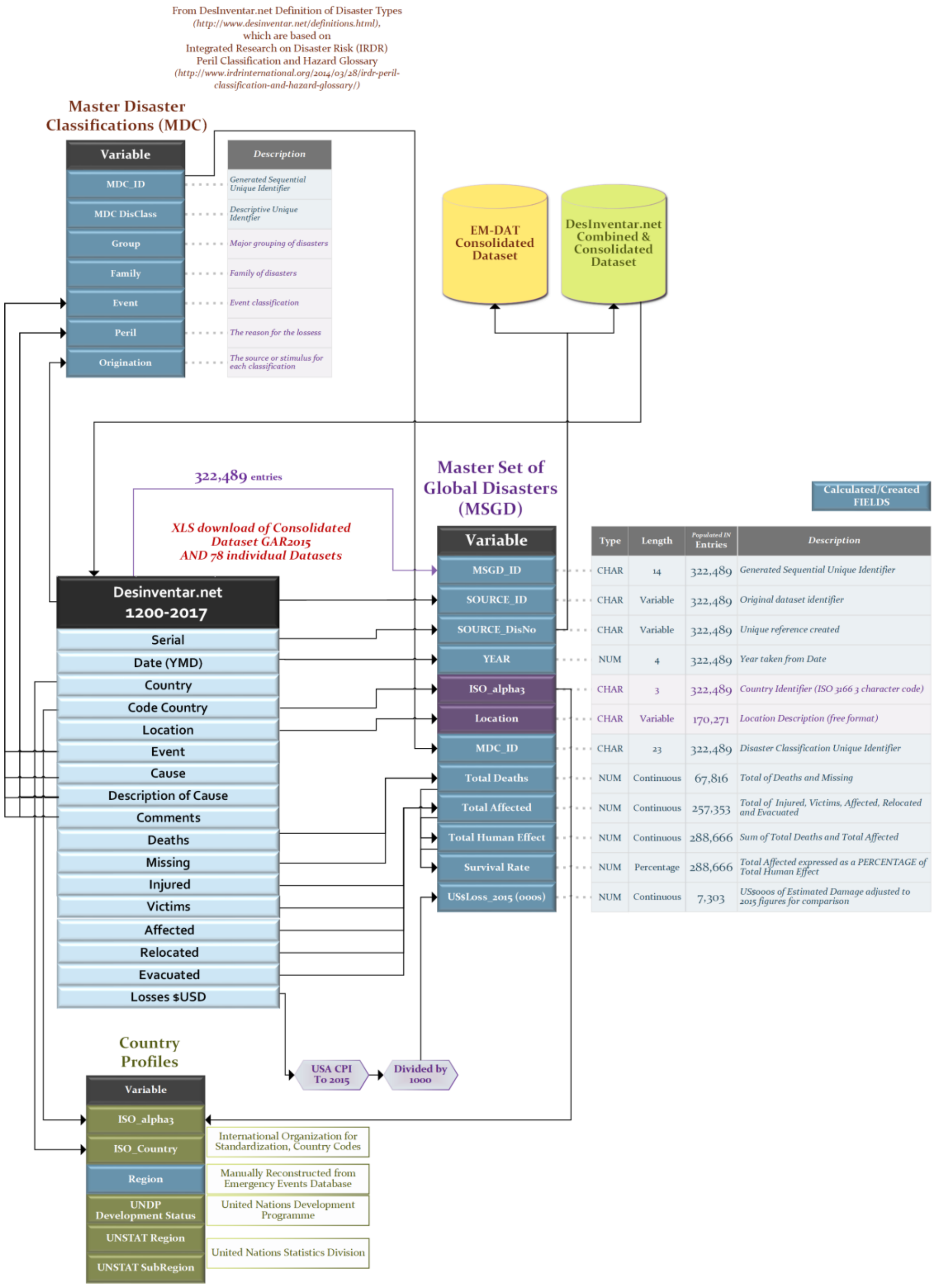


D.1.2: EM-DAT Variables NOT Mapped

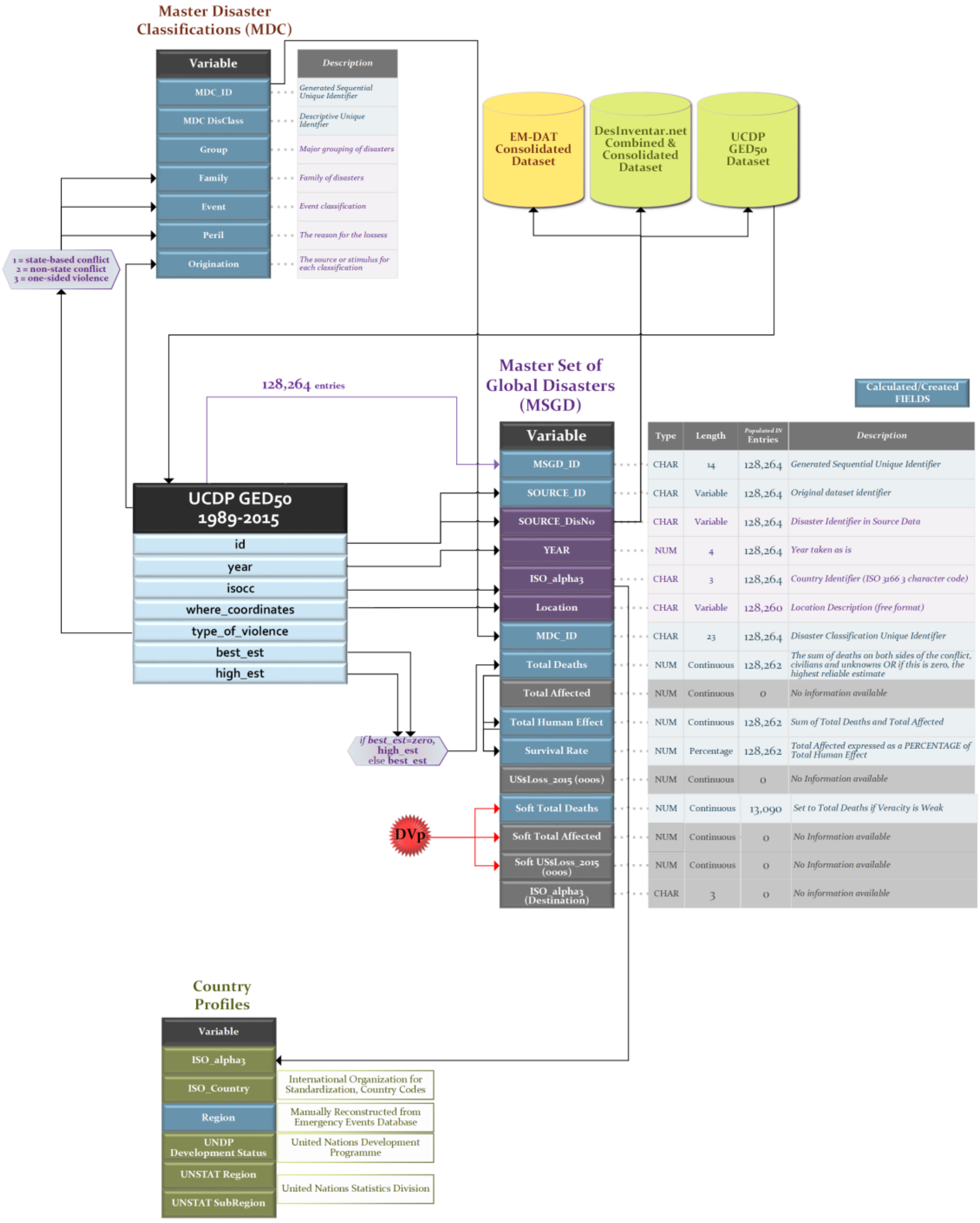
EM-DAT Variable	Reason for Exclusion
Date entered	*Not Accessible*
Entered by	*Not Accessible*
Last updated	*Not Accessible*
Entered by	*Not Accessible*
Disaster sub-group	*Not Accessible*
Disaster sub-sub-type	*Not Accessible*
Deaths	*Not Accessible*
Affected	*Not Accessible*
Declaration/international appeal	*Not Accessible*
Event name	*Not Accessible*
Glide Number	*Not Accessible*
Country	*Not used, contains errors, ISO Country used instead*
Region	*R Archives only, manually recreated for EMDAT4*
Continent	*Not Accessible*
River basin	*Not Accessible*
Latitude	*Not Accessible*
Longitude	*Not Accessible*
Start day/month/year	*Not used as poorly or incompletely populated*
End day/month/year	*Not used as poorly or incompletely populated*
Local time	*Not Accessible*
Origin	*Not Accessible*
Disaster magnitude scale and value	*Not Accessible*
Aid contribution	*Not Accessible*
OFDA response	*Not Accessible*
Appeal for international assistance+date	*Not Accessible*
Declaration of disaster + date	*Not Accessible*
Source type and name	*Not Accessible*
Reporting date	*Not Accessible*
Reliability score (1/5)	*Not Accessible*
Deaths	*Not Accessible*
Missing	*Not Accessible*
Injured	*Not Accessible*
Affected	*Not Accessible*
Homeless	*Not Accessible*
Total affected	*Not Accessible*
Total estimated damages (000'US$ current value)	*Not Accessible*
Reconstruction cost (000'US$ current value)	*Not Accessible*
Insured losses (000'US$ current value)	*Not Accessible*
Disaster impact	*Not Accessible*
Infrastructure	*Not Accessible*
Comments	*Not Accessible*

Fields shaded yellow can be accessed in EM-DAT cconsolidated and cleaned dataset

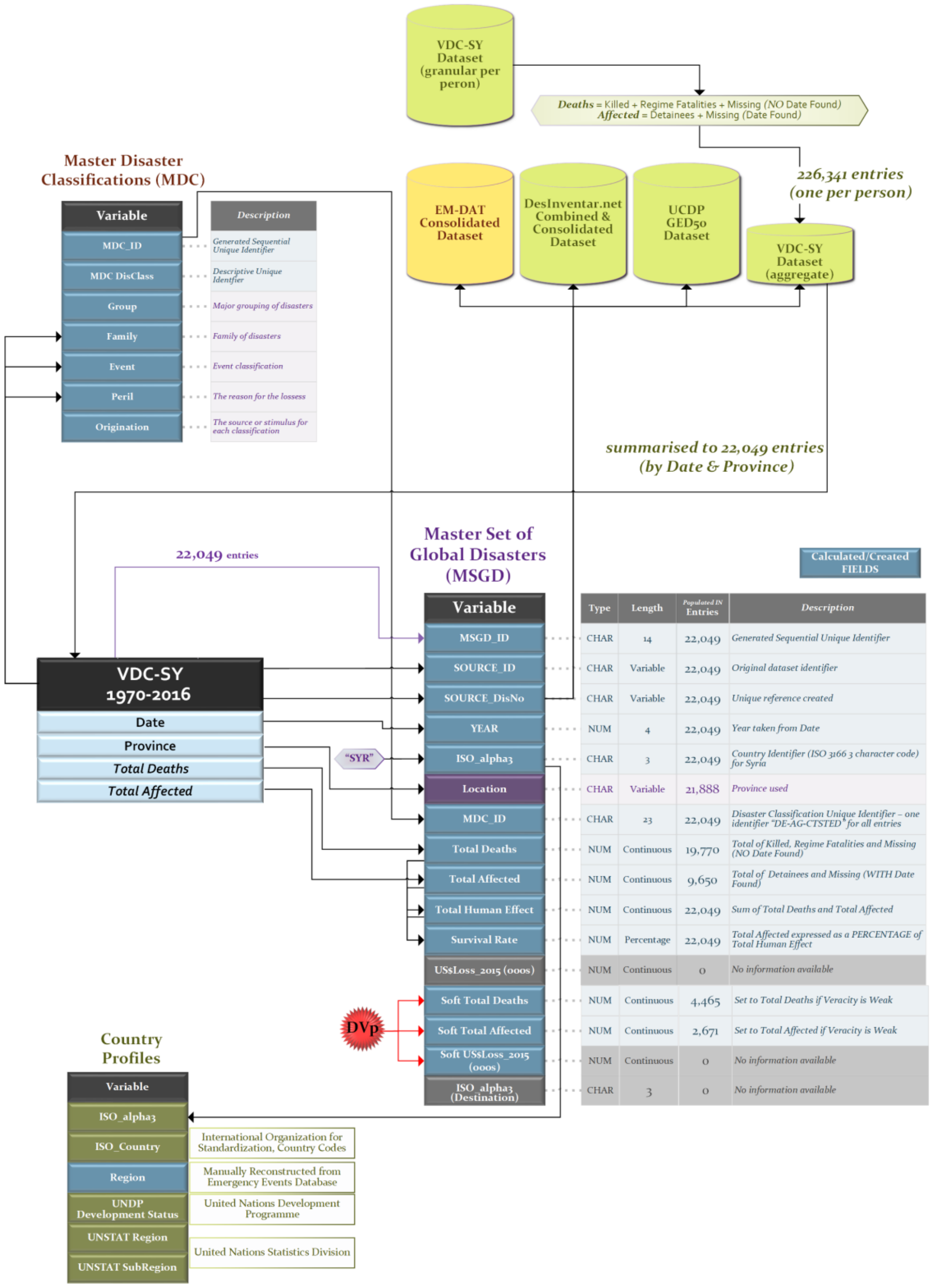
D.2: DesInventar → MSGD



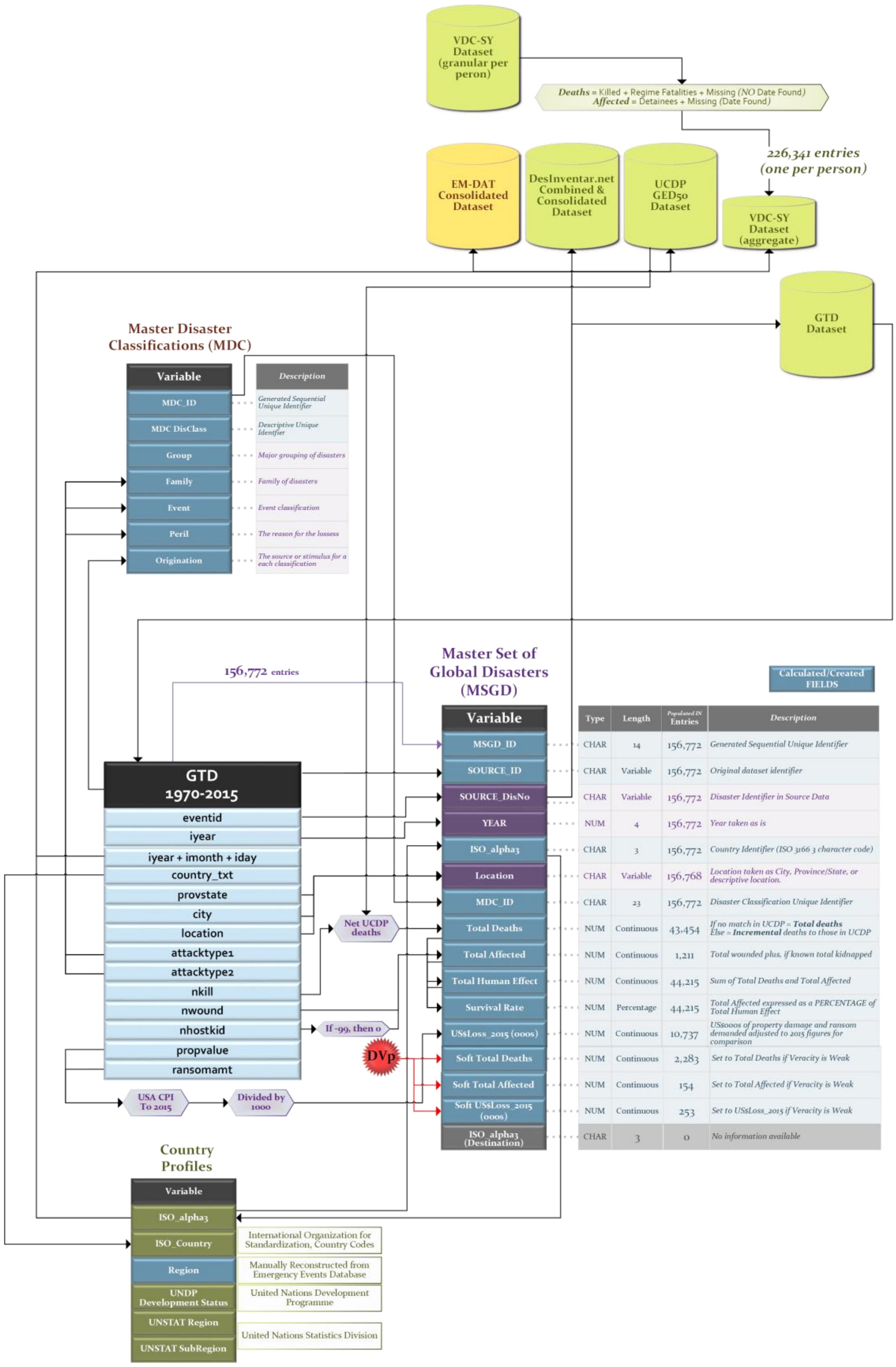
D.3: UCDP GED50 → MSGD



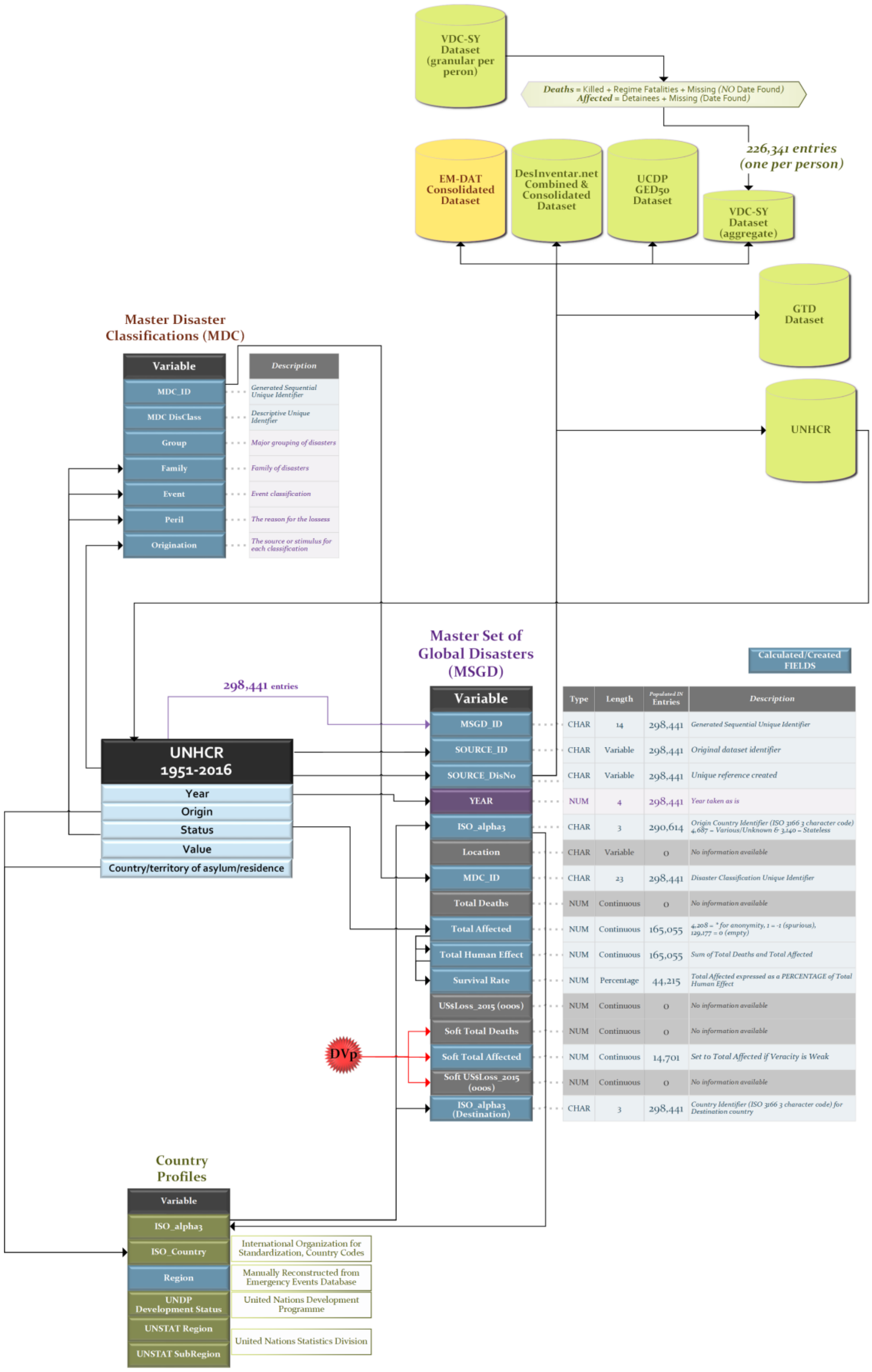
D.4: VDC-SY → MSGD



D.5: GTD → MSGD


VDC-SY Dataset (granular per peron)
Deaths = Killed + Regime Fatalities + Missing (NO Date Found)
Affected = Detainees + Missing (Date Found)
226,341 entries (one per person)
EM-DAT Consolidated Dataset
DesInventar.net Combined & Consolidated Dataset
UCDP GED50 Dataset
VDC-SY Dataset (aggregate)
GTD Dataset
Master Disaster Classifications (MDC)
Variable
Description
MDC_ID
Generated Sequential Unique Identifier
MDC DisClass
Descriptive Unique Identfier
Group
Major grouping of disasters
Family
Family of disasters
Event
Event classification
Peril
The reason for the losses
Origination
The source or stimulus for a each classification
156,772 entries
Master Set of Global Disasters (MSGD)
Calculated/Created FIELDS
GTD 1970-2015
eventid
iyear
iyear + imonth + iday
country_txt
provstate
city
location
attacktype1
attacktype2
nkill
nwound
nhostkid
propvalue
ransomamt
Net UCDP deaths
If -99, then 0
DVp
USA CPI To 2015
Divided by 1000
Variable
MSGD_ID
SOURCE_ID
SOURCE_DisNo
YEAR
ISO_alpha3
Location
MDC_ID
Total Deaths
Total Affected
Total Human Effect
Survival Rate
US$Loss_2015 (000s)
Soft Total Deaths
Soft Total Affected
Soft US$Loss_2015 (000s)
ISO_alpha3 (Destination)
Type
Length
Populated IN Entries
Description
CHAR 14 156,772 Generated Sequential Unique Identifier
CHAR Variable 156,772 Original dataset identifier
CHAR Variable 156,772 Disaster Identifier in Source Data
NUM 4 156,772 Year taken as is
CHAR 3 156,772 Country Identifier (ISO 3166 3 character code)
CHAR Variable 156,768 Location taken as City, Province/State, or descriptive location.
CHAR 23 156,772 Disaster Classification Unique Identifier
NUM Continuous 43,454 If no match in UCDP = Total deaths Else = Incremental deaths to those in UCDP
NUM Continuous 1,211 Total wounded plus, if known total kidnapped
NUM Continuous 44,215 Sum of Total Deaths and Total Affected
NUM Percentage 44,215 Total Affected expressed as a PERCENTAGE of Total Human Effect
NUM Continuous 10,737 US$000s of property damage and ransom demanded adjusted to 2015 figures for comparison
NUM Continuous 2,283 Set to Total Deaths if Veracity is Weak
NUM Continuous 154 Set to Total Affected if Veracity is Weak
NUM Continuous 253 Set to US$Loss_2015 if Veracity is Weak
CHAR 3 0 No information available
Country Profiles
Variable
ISO_alpha3
ISO_Country
Region
UNDP Development Status
UNSTAT Region
UNSTAT SubRegion
International Organization for Standardization, Country Codes
Manually Reconstructed from Emergency Events Database
United Nations Development Programme
United Nations Statistics Division

D.6: UNHCR → MSGD



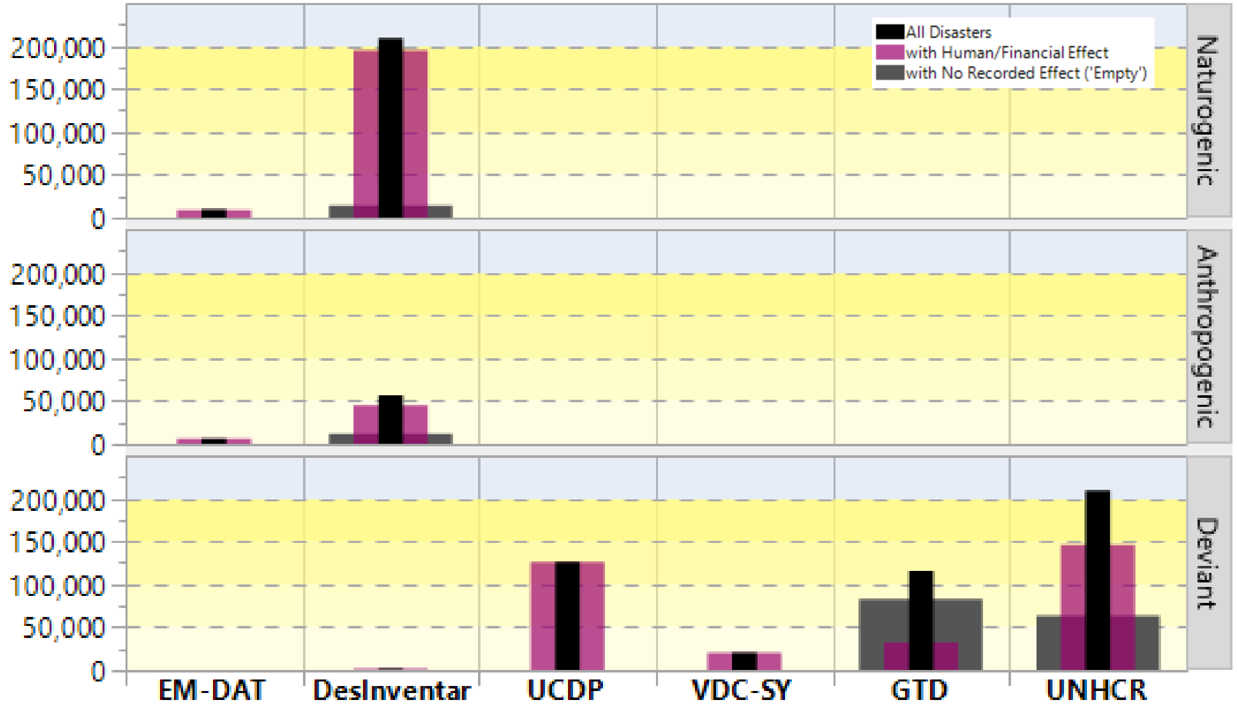
Appendix E: No Effect ('Empty') Disaster Entries

E.1: Disaster Groups x Data Source

These charts illustrate the 'empty 'entries (in grey) found in each MSGD contribution dataset for the years studied (1990–2015).

These are disaster entries which do not contain any values for human financial losses. 'Empty' entries are filtered out of the data used for the study.

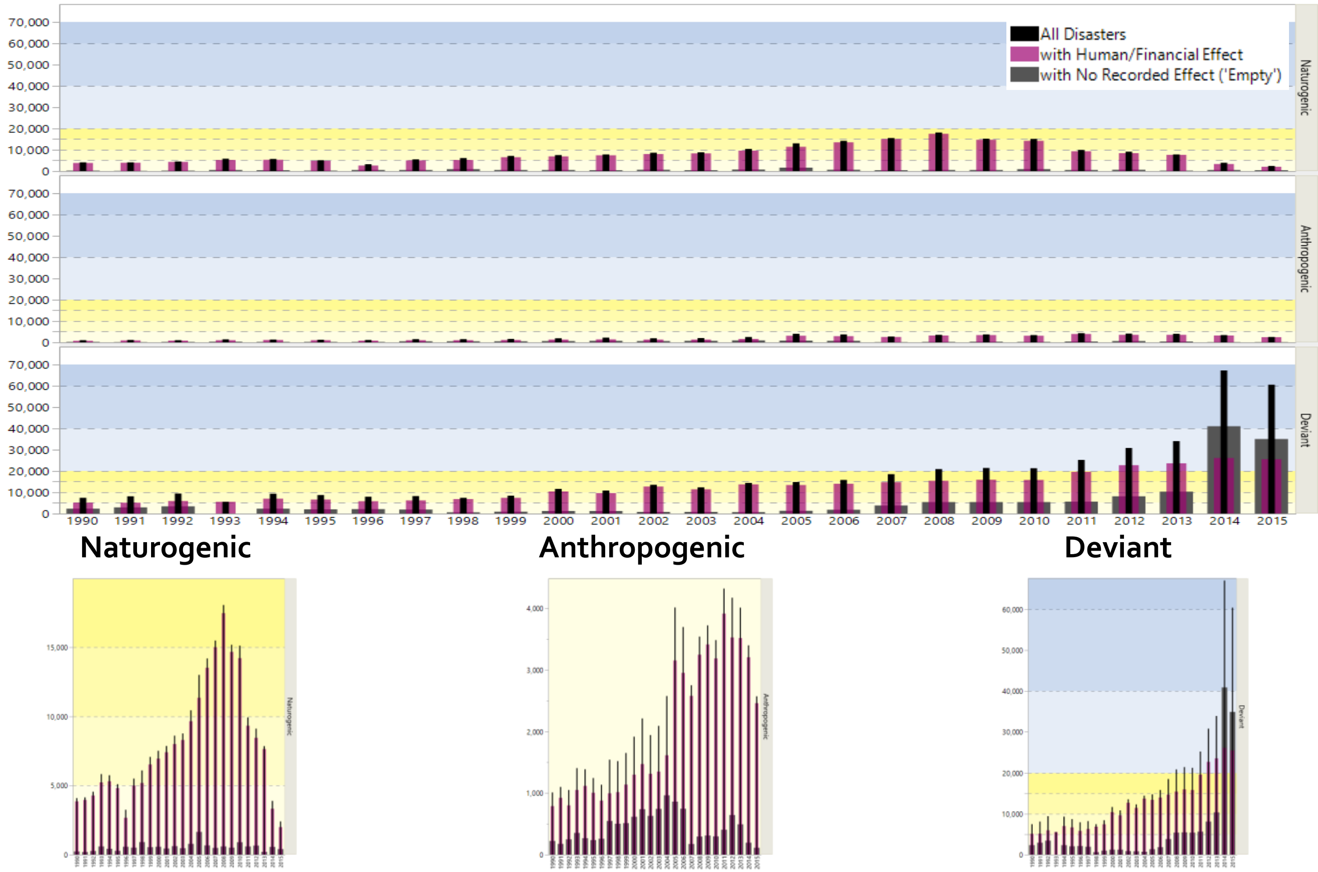
As can been seen from the bars, most of the 'empty' entries encountered were from GTD and UNHCR.



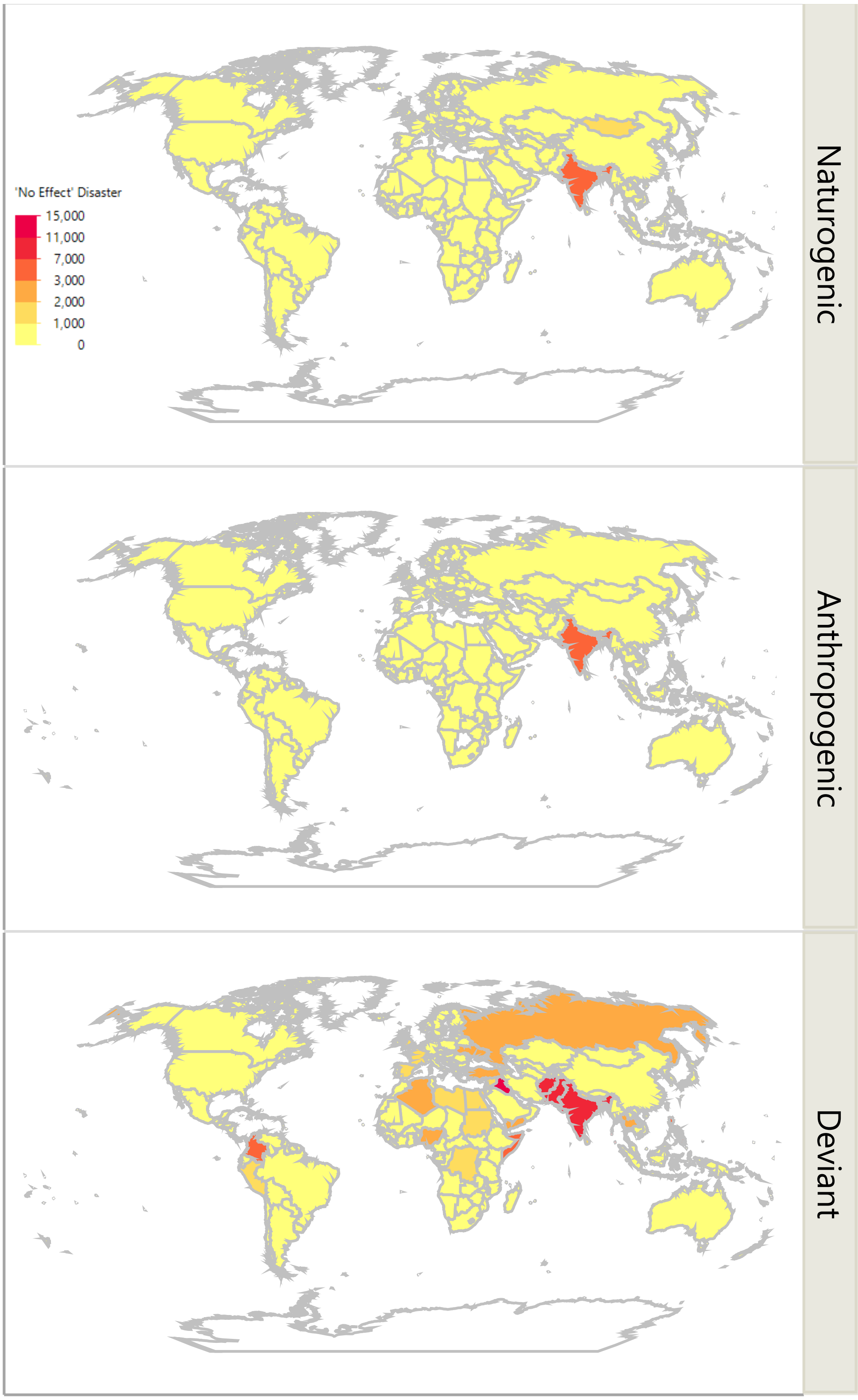
E.2: Disaster Groups x Year

These charts illustrate the ‘empty ‘entries (in grey) found for the years studied (1990–2015).

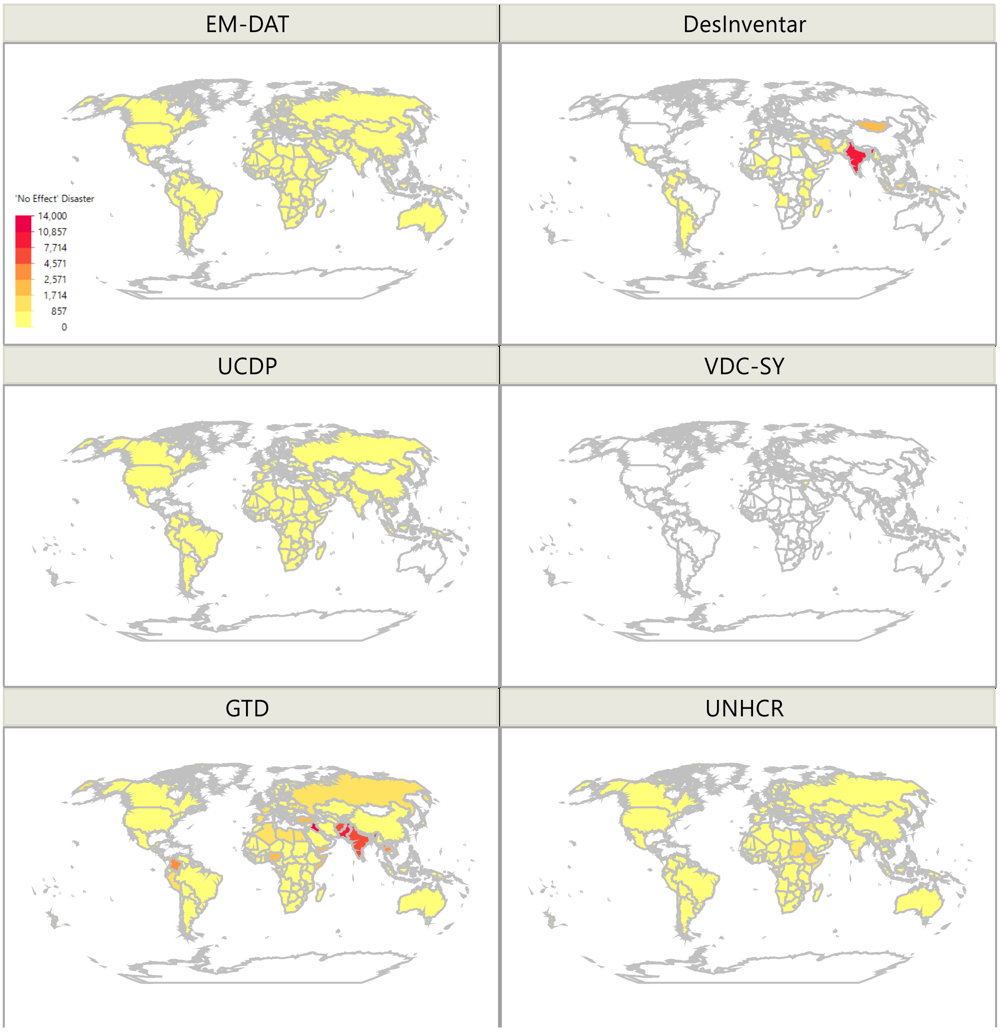
The predominance of such entries in more recent years, when the expectation is that data would be more accurate and complete is noteworthy.



E.3: Disaster Groups (Maps)



E.4: Data Source (Maps)



Appendix F: FINANCIAL EFFECTS OF DISASTERS

F.1: EM-DAT Financial Losses

This appendix provides some basic representations of the ~US$4.26 trillion financial losses (2015 figures) found in EM-DAT found in 4,855 EM-DAT entries (~22%) [*Figure F.1-1*, Figure F.1-2 and Figure F.1-3].

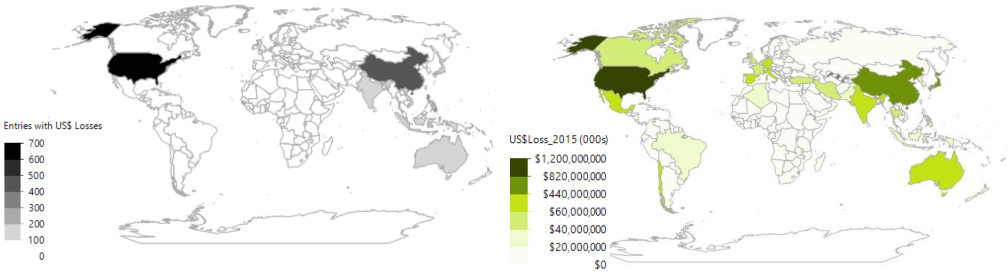


Figure F.1-1: EM-DAT Map of Financial Losses – Entries & 000USs

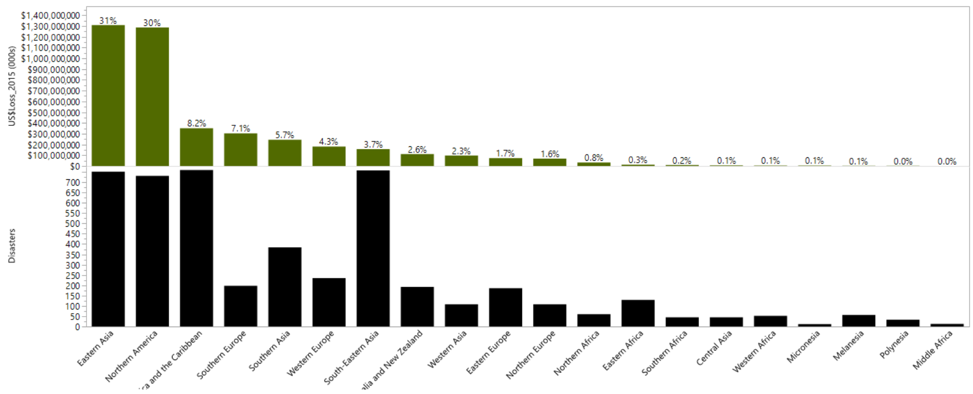


Figure F.1-2: EM-DAT Financial Losses by Sub-Region

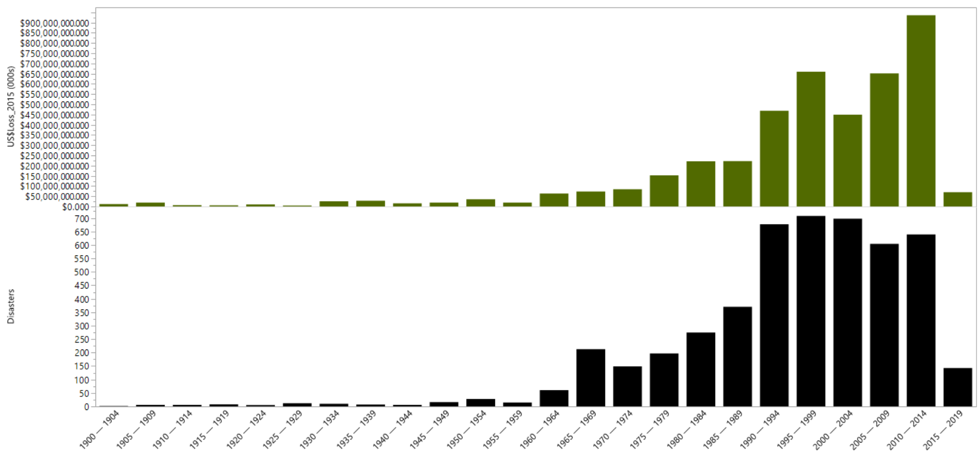


Figure F.1-3: EM-DAT Financial Losses (5-Year periods)

As a reference this map depicts EM-DAT entries with NO financial losses [*Figure F.1-4*]:

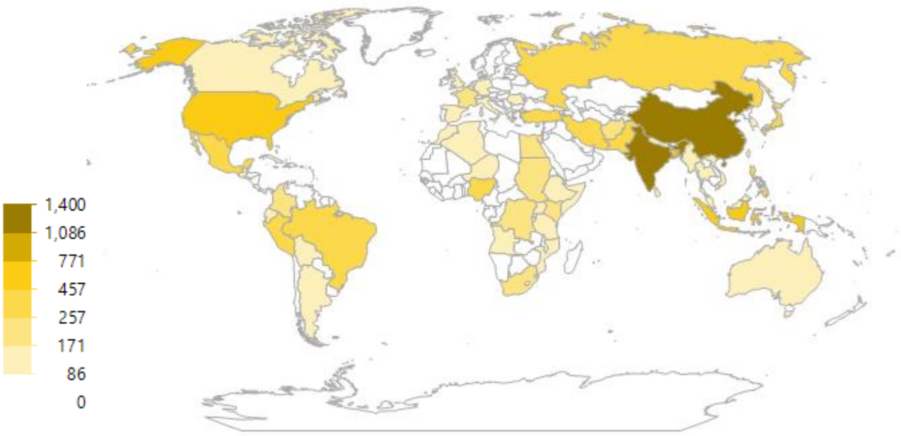


Figure F.1-4: EM-DAT Map of Entries with NO Financial Losses

... and this chart depicts no financial effect entries over time in five year time-slots [*Figure F.1-5*].

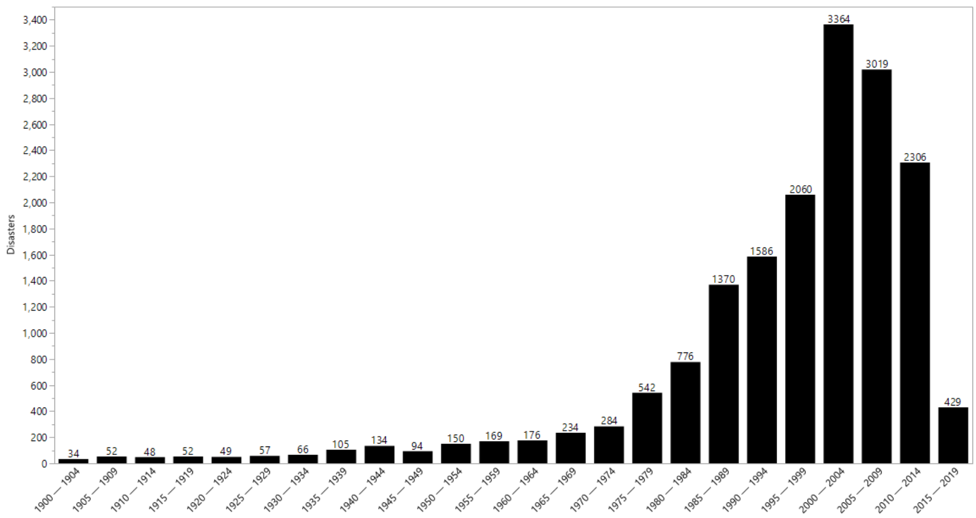


Figure F.1-5: EM-DAT entries with NO Financial Losses (5-Year periods)

F.2: MSGD Financial Losses

This appendix provides a cursory look at the circa $5.7 trillion financial losses (2015 figures) found in the MSGD (1990–2015). Looking at the occurrence of financial disaster entries by sub-region it can be seen that Latin America and the Caribbean, with 3002 entries, is almost 3 times the number of the next highest sub-region [*Figure F.2-1*]. Shifting perspective to countries, the United States, with 701 entries, has the highest number of entries for financial losses [*Table F.2-1*].

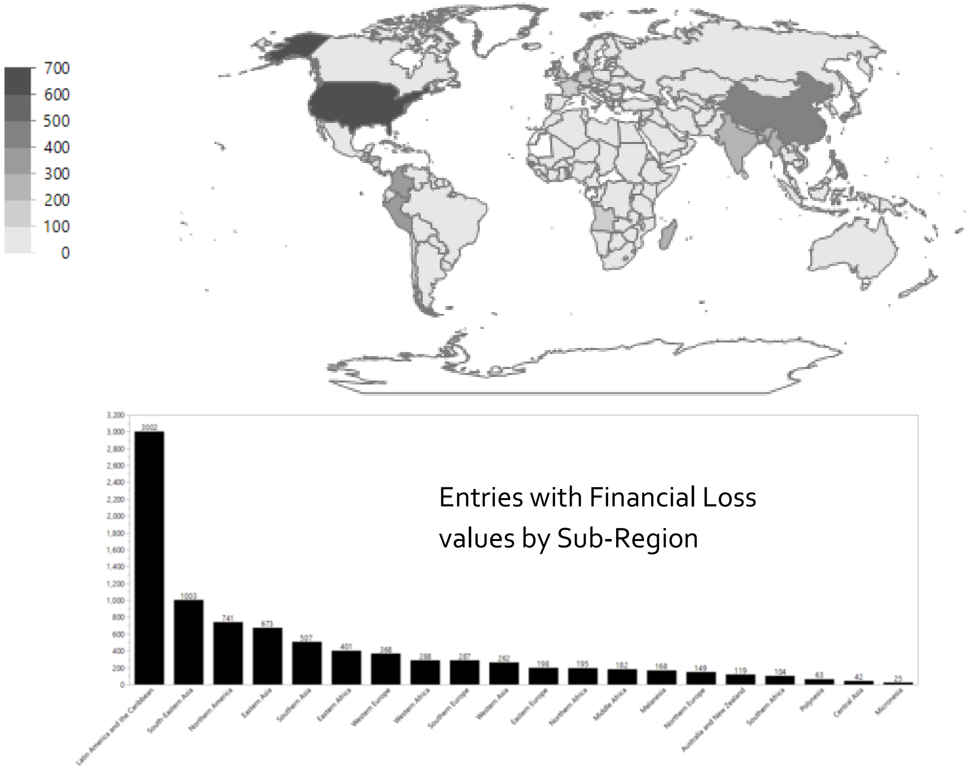


Figure F.2-1: MSGD Number of Entries with US$ Losses

	Country	Entries	%age of Total	US$Loss_2015 (000s)[ALL]	%age of Total US$	US$Loss_2015 (000s) [Firm]	US$Loss_2015 (000s) [Soft]	%age [SOFT]
1	United States of America (the)	701	7.99%	$951,495,265	16.71%	$630,697,316	$320,797,949	33.72%
2	China	429	4.89%	$544,292,531	9.56%	$194,248,959	$350,043,572	64.31%
3	Philippines (the)	423	4.82%	$25,931,653	0.46%	$11,075,254	$14,856,398	57.29%
4	Panama	348	3.96%	$308,733	0.01%	$242,728	$66,004	21.38%
5	Peru	311	3.54%	$2,091,990	0.04%	$1,403,357	$688,633	32.92%
6	El Salvador	306	3.49%	$6,042,441	0.11%	$865,148	$5,177,293	85.68%
7	Colombia	301	3.43%	$9,530,813	0.17%	$6,801,461	$2,729,353	28.64%
8	Guatemala	296	3.37%	$4,148,870	0.07%	$1,992,145	$2,156,725	51.98%
9	Honduras	293	3.34%	$7,275,961	0.13%	$487,517	$6,788,444	93.30%
10	Madagascar	277	3.16%	$2,132,102	0.04%	$842,345	$1,289,757	60.49%
Remaining 187 countries with Financial losses		5,092	58.02%	$4,140,165,259	72.72%	$898,891,067	$3,241,274,192	78.29%
	All	8,777		$5,693,415,616		$1,747,547,297	$3,945,868,320	

Table F.2-1: MSGD 1990–2015 US$ Losses - Top 10 Countries

The geographic spread of undifferentiated, ‘firm’ and ‘soft’ US$ losses can be seen in the maps that form *Figure F.2-2*.

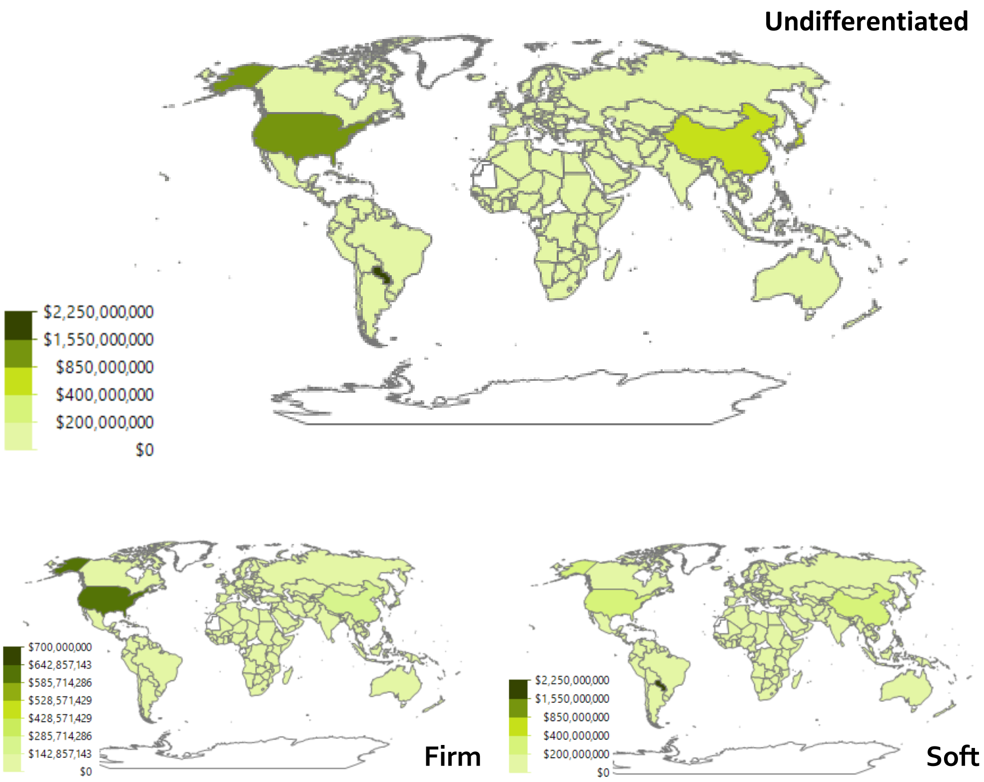


Figure F.2-2: MSGD 1990–2015 US$ Loss Maps

*Figure F.*2-3 shows the distribution of US$ losses totals and means in nested bars – undifferentiated, ‘firm’ and ‘soft’ – by sub-region in descending order of sum and mean respectively for each chart. The first chart helps identify the scale of ‘softness’ of the Latin American and Caribbean losses. Also, notable is that ‘firm’ numbers become more dominant when looking at means as these relatively fewer entries contain higher value losses, whereas ‘soft’ numbers are numerous and hold smaller values.

*Figure F.*2-4 shows the distribution of US$ losses totals and means in nested bars – undifferentiated, ‘firm’ and ‘soft’ – by year, with an inexplicable spike caused by soft data in the year 2000. Further investigation shows that this soft value is because of over US$2.34 trillion of losses recorded DesInventar for a fire in Paraguay.

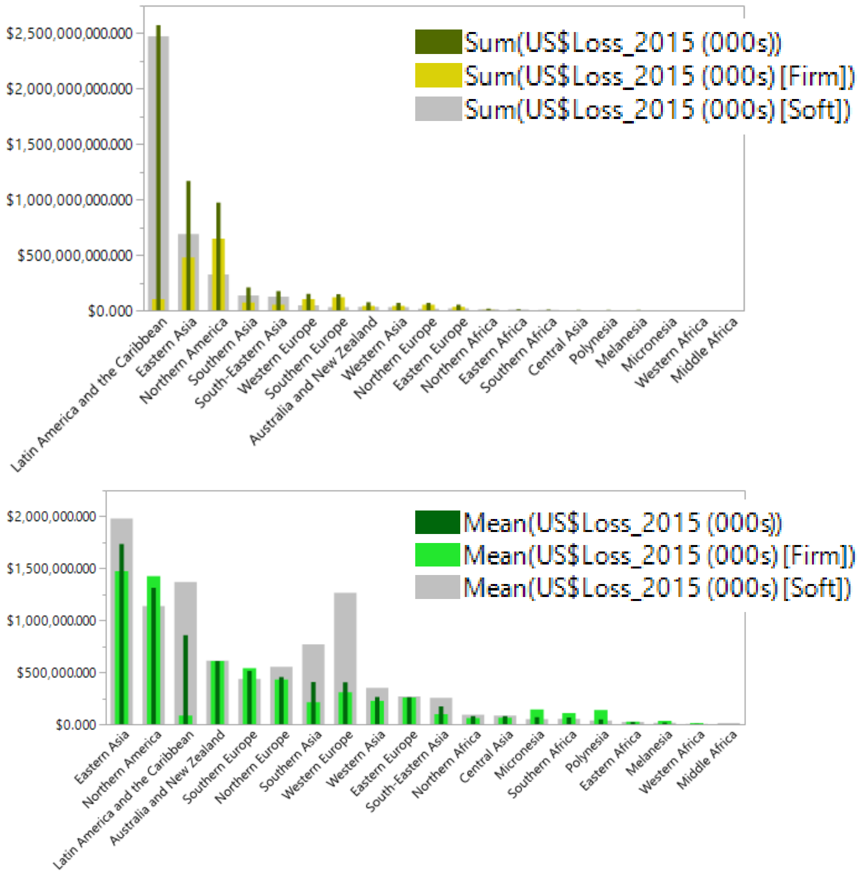


Figure F.2-3: MSGD 1990–2015 US$ Losses by Sub-Region

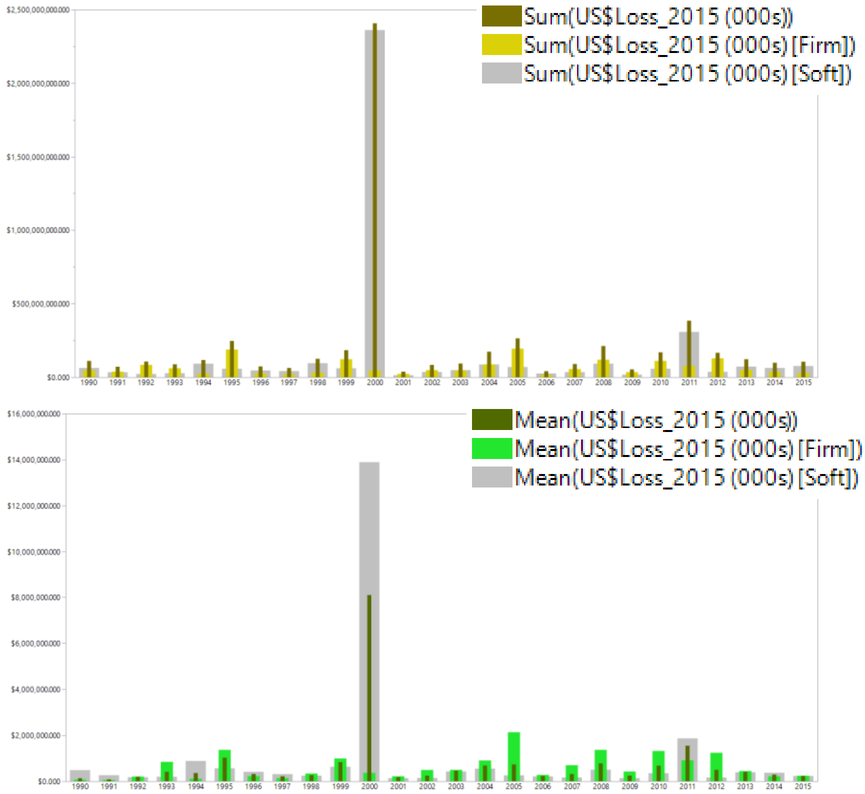
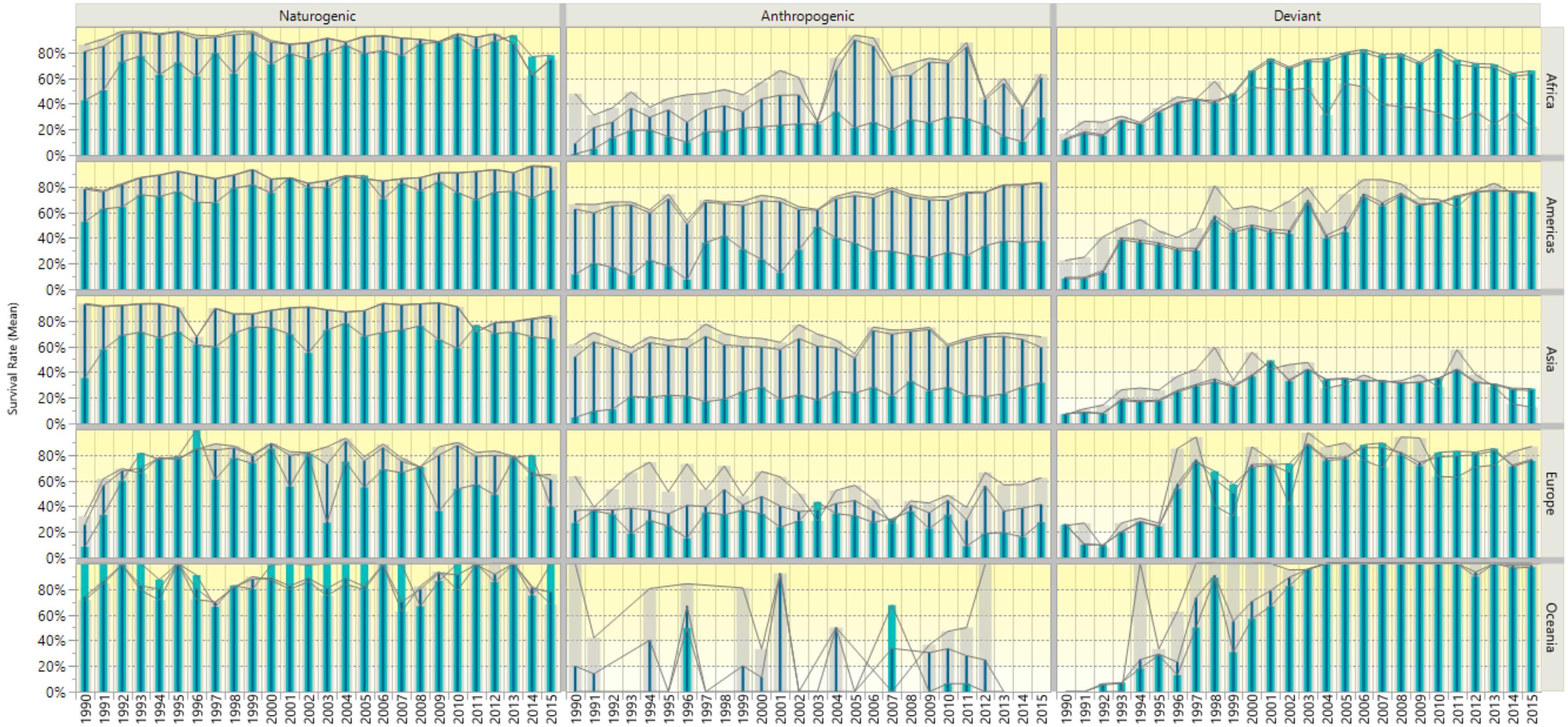


Figure F.2-4: MSGD 1990–2015 US$ Losses by Year

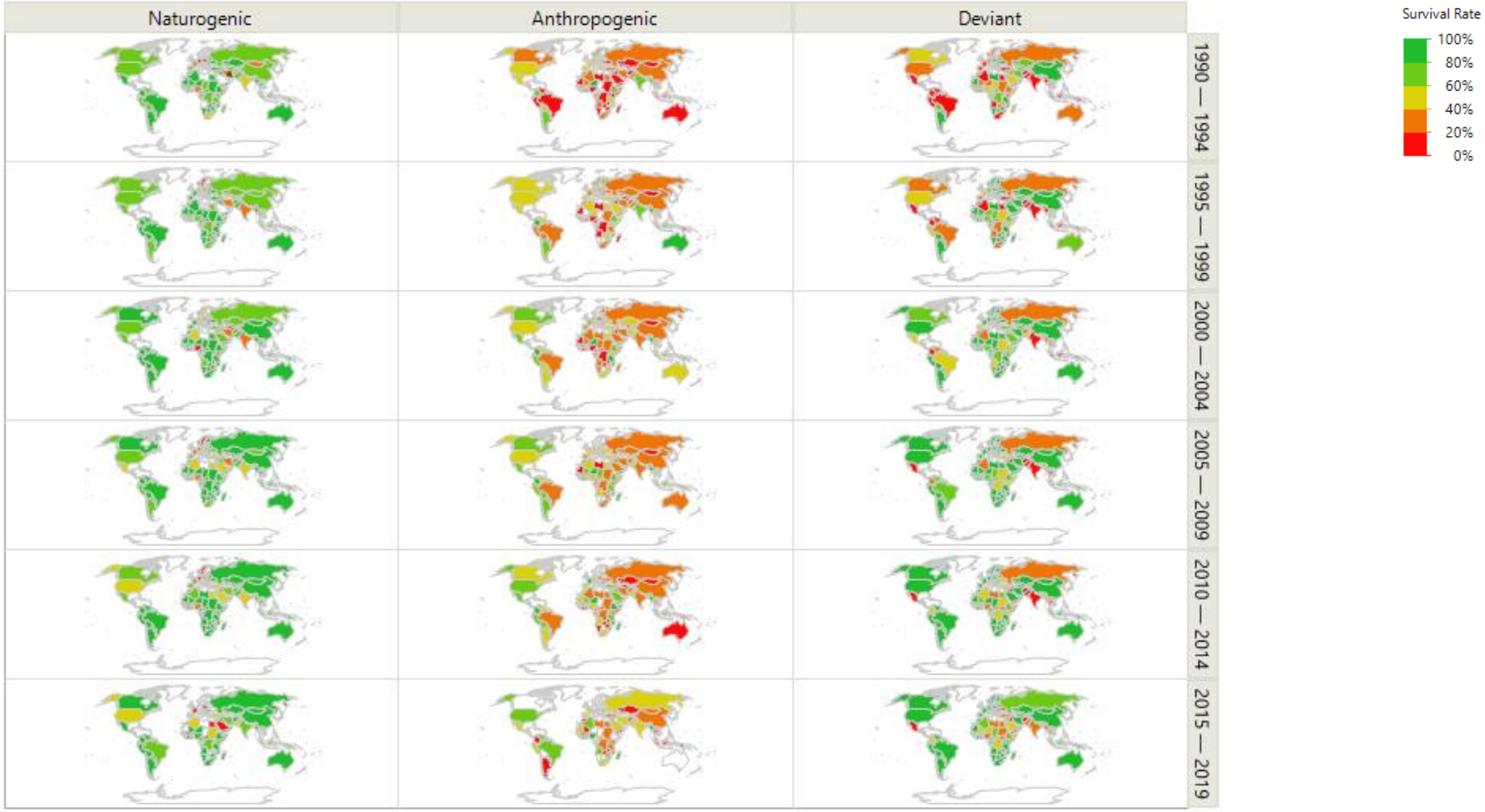
Appendix G: Mean Survival Rates (M^IO)

G.1: Disaster Group x Region (Bar Charts)

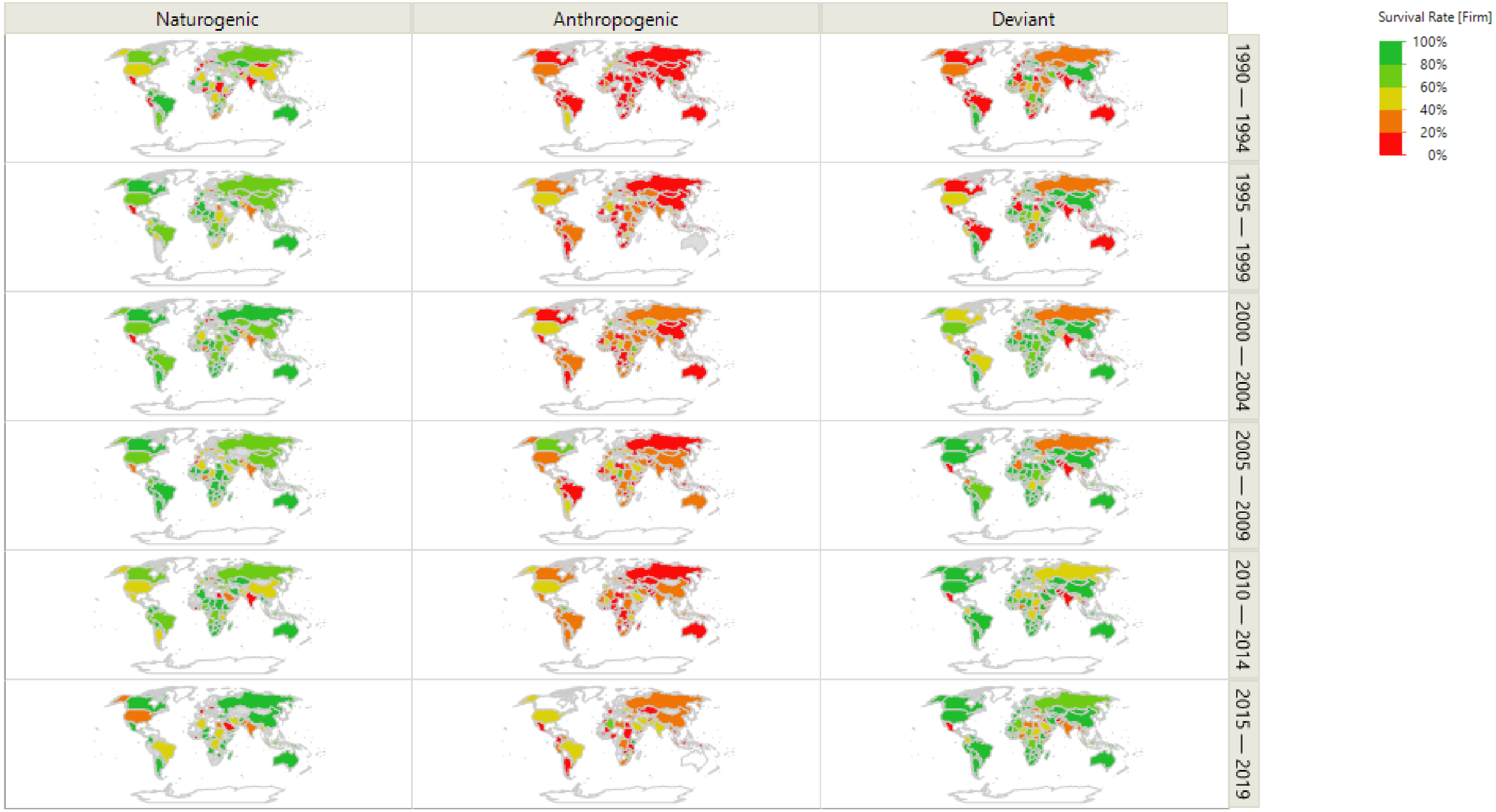


G.2: Survival Rate Maps (Disaster Groups x 5-Year Means)

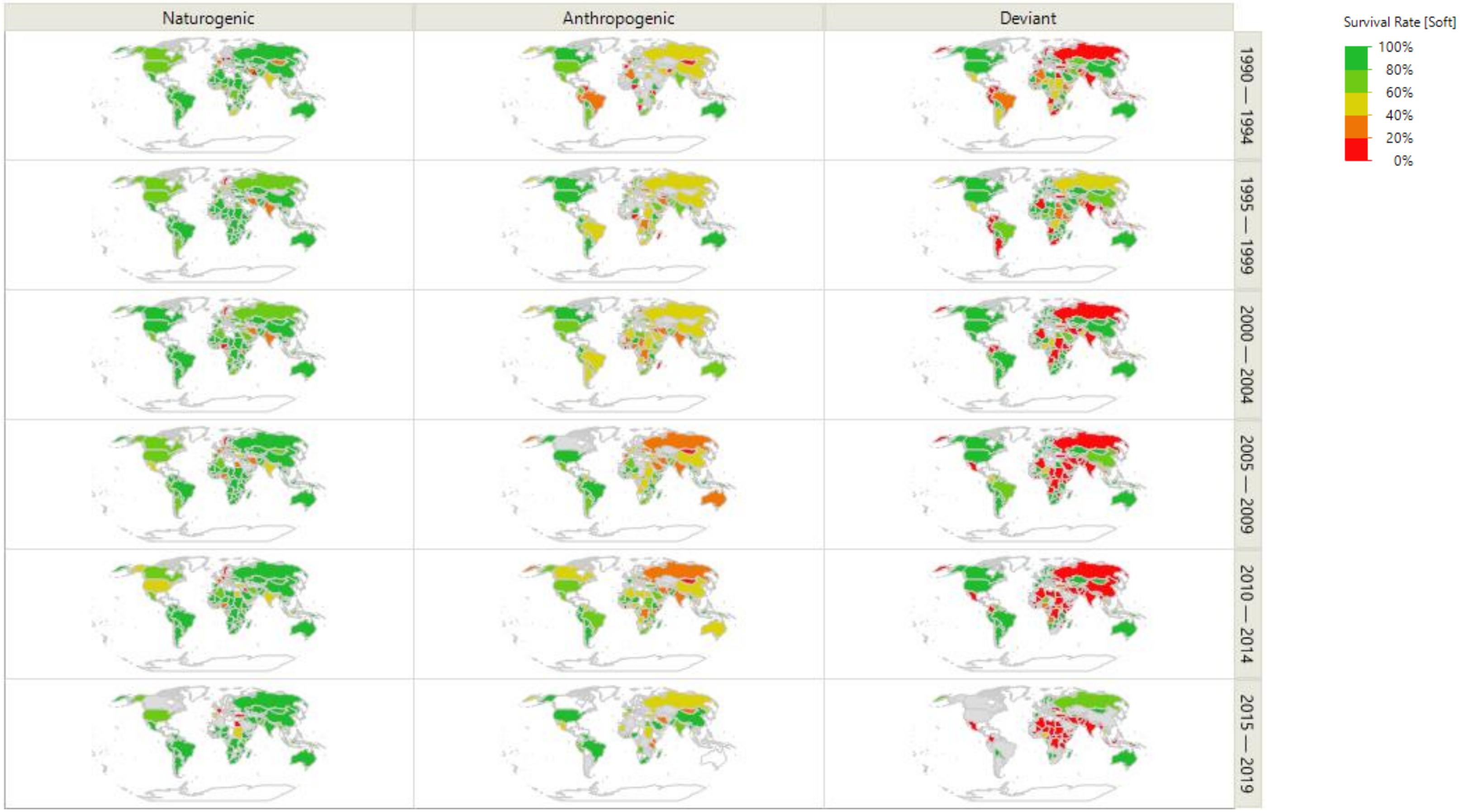
Undifferentiated



Firm


Naturogenic
Anthropogenic
Deviant
1990 — 1994
1995 — 1999
2000 — 2004
2005 — 2009
2010 — 2014
2015 — 2019
Survival Rate [Firm]
100%
80%
60%
40%
20%
0%

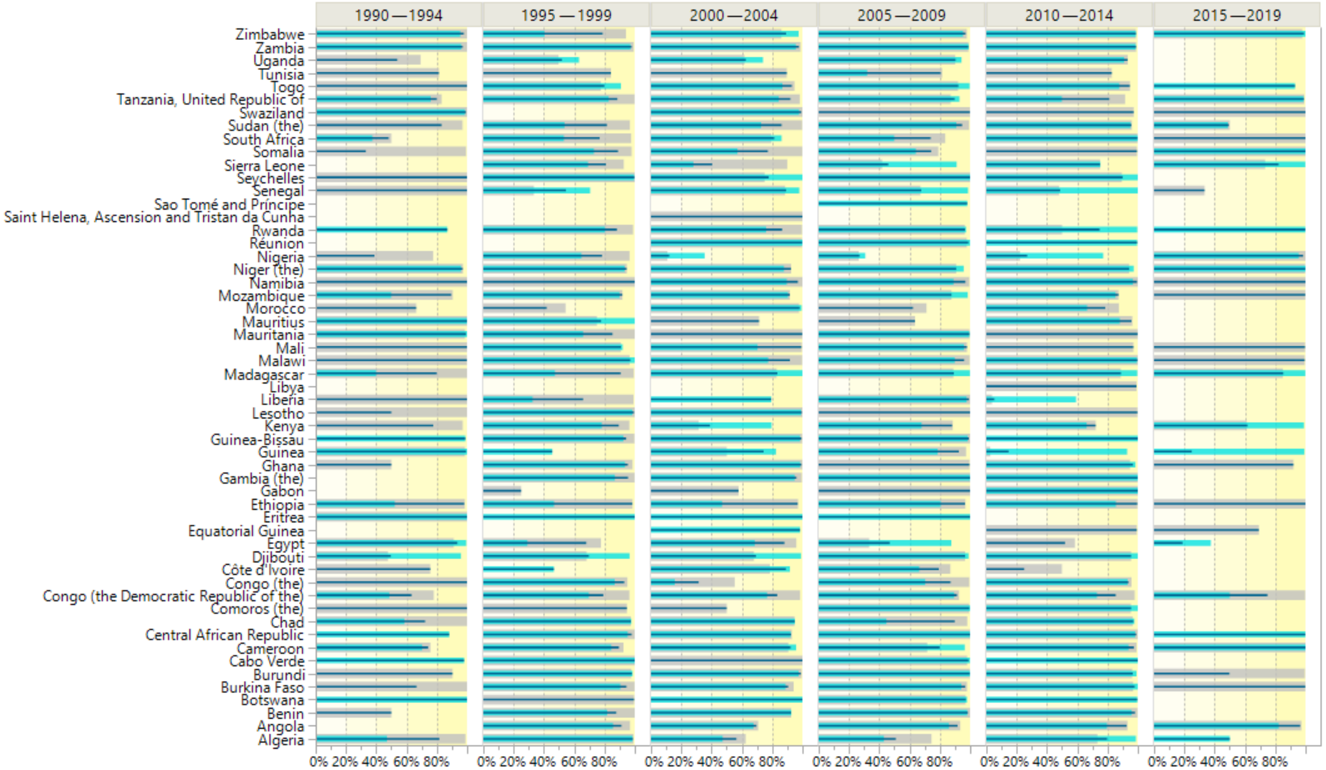
Soft


Naturogenic
Anthropogenic
Deviant
1990 — 1994
1995 — 1999
2000 — 2004
2005 — 2009
2010 — 2014
2015 — 2019
Survival Rate [Soft]
100%
80%
60%
40%
20%
0%

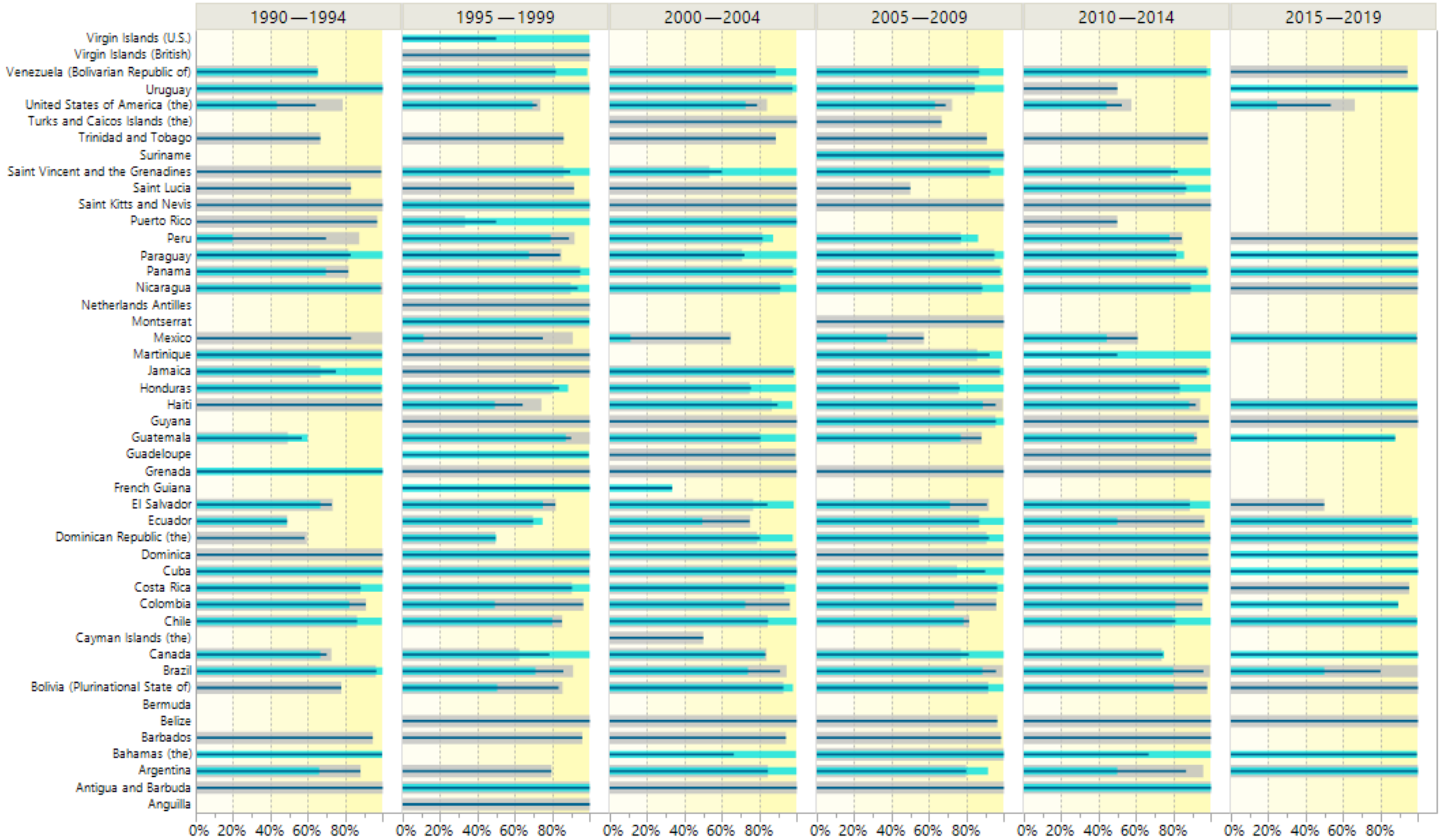
G.3: Disaster Group x Region x 5-Year Means (Bar Charts)

G.3.1: Naturogenic

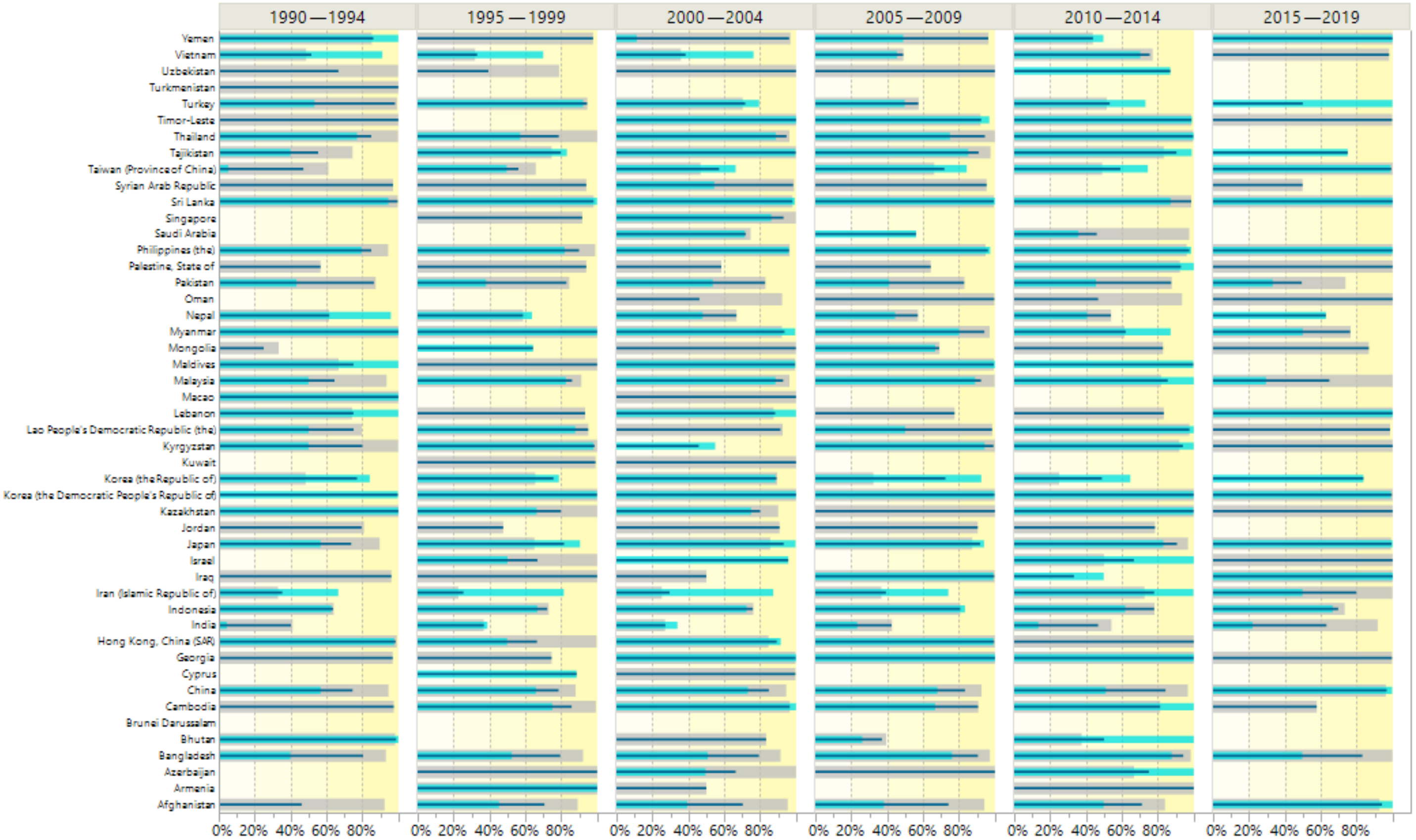
Africa



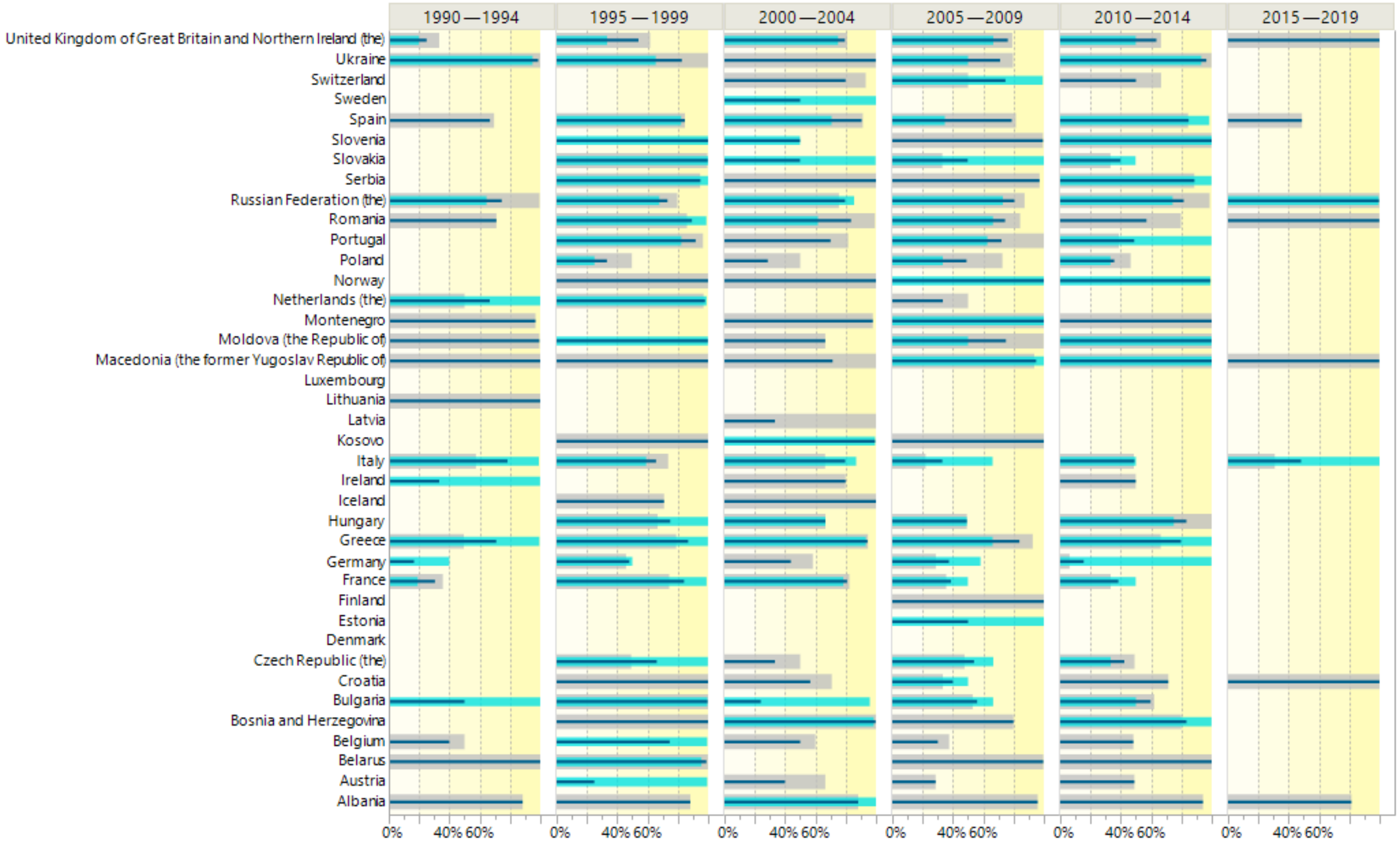
Americas


1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
Virgin Islands (U.S.)
Virgin Islands (British)
Venezuela (Bolivarian Republic of)
Uruguay
United States of America (the)
Turks and Caicos Islands (the)
Trinidad and Tobago
Suriname
Saint Vincent and the Grenadines
Saint Lucia
Saint Kitts and Nevis
Puerto Rico
Peru
Paraguay
Panama
Nicaragua
Netherlands Antilles
Montserrat
Mexico
Martinique
Jamaica
Honduras
Haiti
Guyana
Guatemala
Guadeloupe
Grenada
French Guiana
El Salvador
Ecuador
Dominican Republic (the)
Dominica
Cuba
Costa Rica
Colombia
Chile
Cayman Islands (the)
Canada
Brazil
Bolivia (Plurinational State of)
Bermuda
Belize
Barbados
Bahamas (the)
Argentina
Antigua and Barbuda
Anguilla
0% 20% 40% 60% 80%

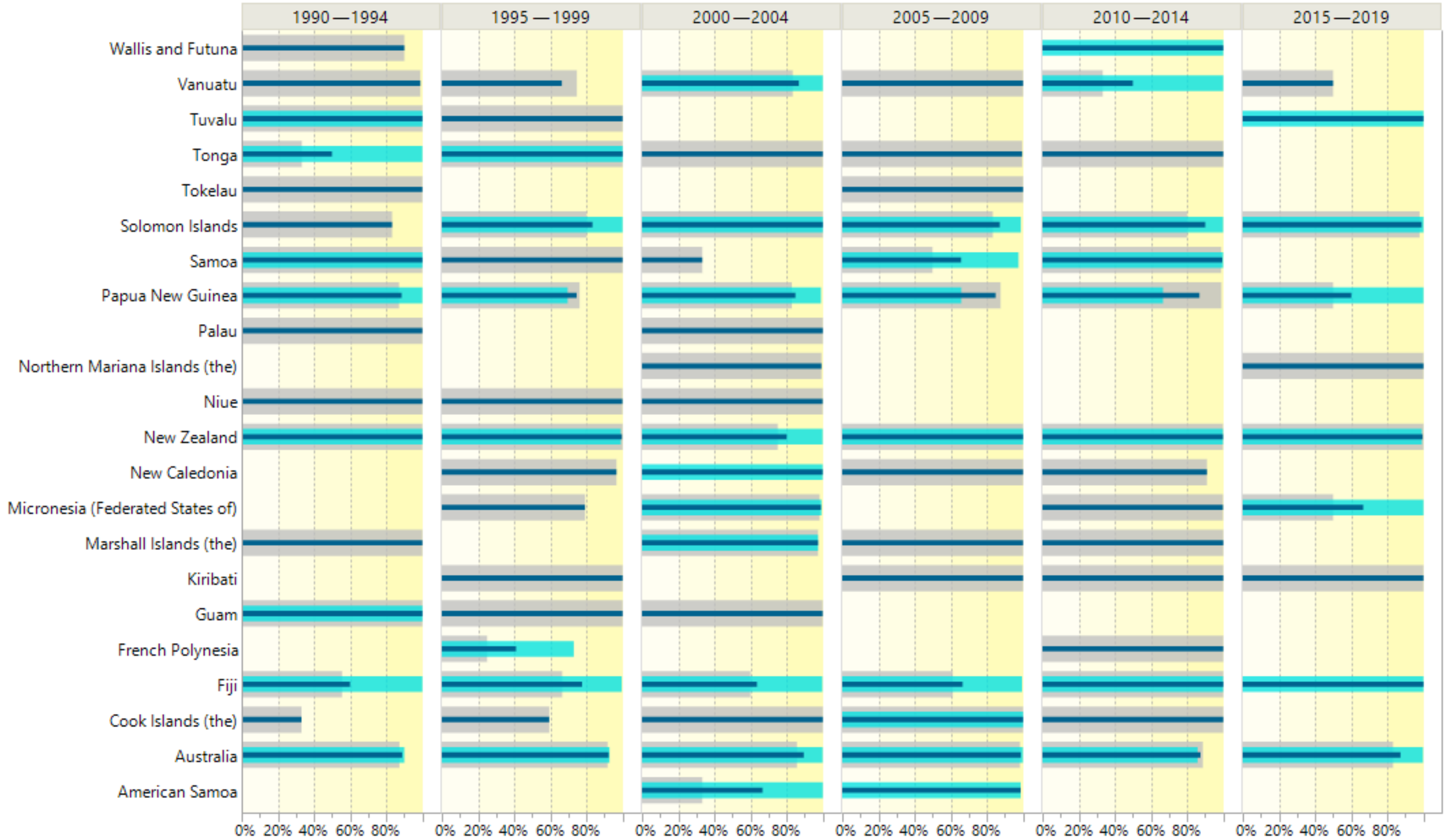
Asia


1990 — 1994
1995 — 1999
2000 — 2004
2005 — 2009
2010 — 2014
2015 — 2019
Yemen
Vietnam
Uzbekistan
Turkmenistan
Turkey
Timor-Leste
Thailand
Tajikistan
Taiwan (Province of China)
Syrian Arab Republic
Sri Lanka
Singapore
Saudi Arabia
Philippines (the)
Palestine, State of
Pakistan
Oman
Nepal
Myanmar
Mongolia
Maldives
Malaysia
Macao
Lebanon
Lao People's Democratic Republic (the)
Kyrgyzstan
Kuwait
Korea (the Republic of)
Korea (the Democratic People's Republic of)
Kazakhstan
Jordan
Japan
Israel
Iraq
Iran (Islamic Republic of)
Indonesia
India
Hong Kong, China (SAR)
Georgia
Cyprus
China
Cambodia
Brunei Darussalam
Bhutan
Bangladesh
Azerbaijan
Armenia
Afghanistan
0% 20% 40% 60% 80%

Europe


1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
United Kingdom of Great Britain and Northern Ireland (the)
Ukraine
Switzerland
Sweden
Spain
Slovenia
Slovakia
Serbia
Russian Federation (the)
Romania
Portugal
Poland
Norway
Netherlands (the)
Montenegro
Moldova (the Republic of)
Macedonia (the former Yugoslav Republic of)
Luxembourg
Lithuania
Latvia
Kosovo
Italy
Ireland
Iceland
Hungary
Greece
Germany
France
Finland
Estonia
Denmark
Czech Republic (the)
Croatia
Bulgaria
Bosnia and Herzegovina
Belgium
Belarus
Austria
Albania
0%
40% 60%

Oceania



1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
Wallis and Futuna
Vanuatu
Tuvalu
Tonga
Tokelau
Solomon Islands
Samoa
Papua New Guinea
Palau
Northern Mariana Islands (the)
Niue
New Zealand
New Caledonia
Micronesia (Federated States of)
Marshall Islands (the)
Kiribati
Guam
French Polynesia
Fiji
Cook Islands (the)
Australia
American Samoa
0% 20% 40% 60% 80%

G.3.2: Anthropogenic

Africa

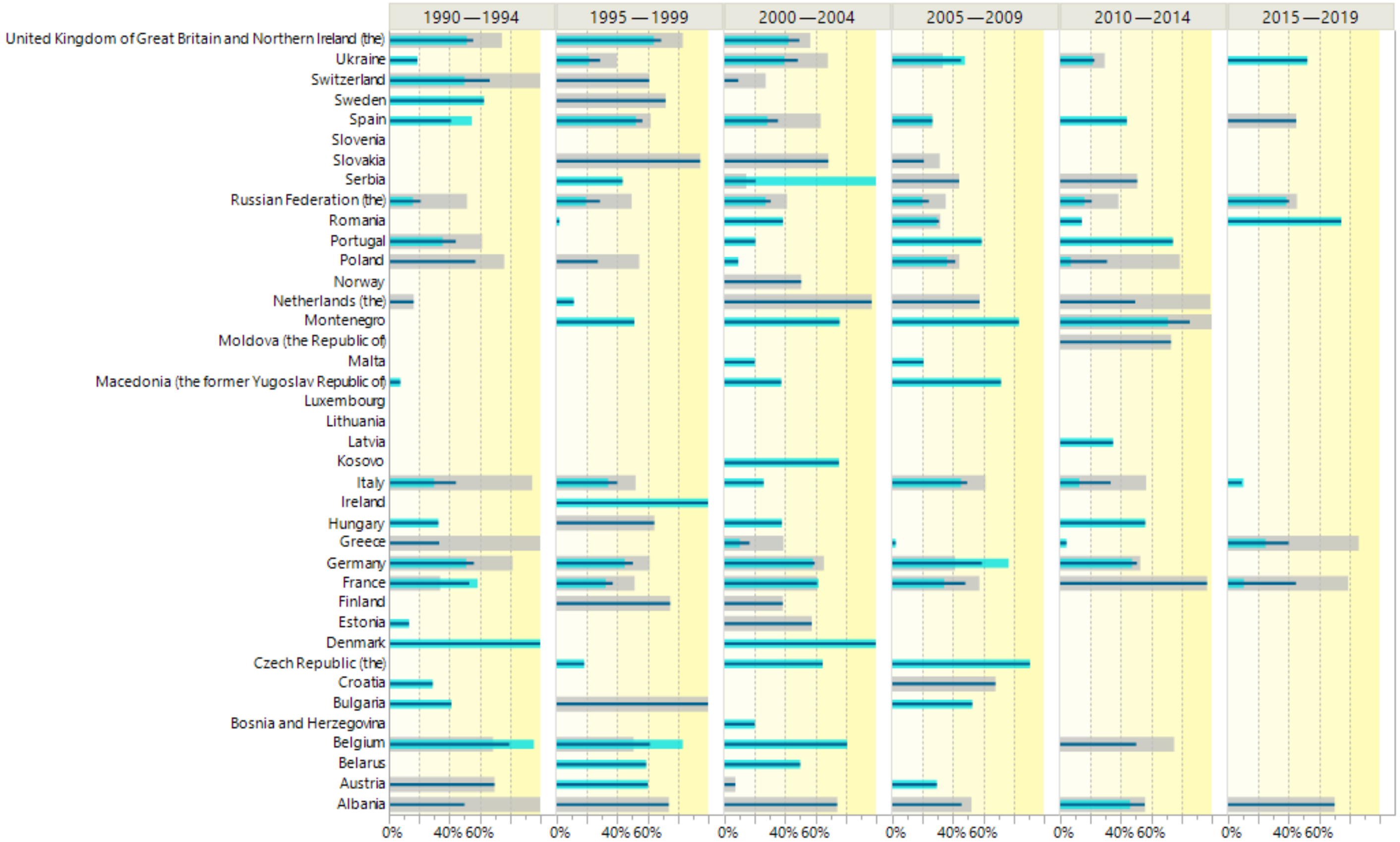
Americas


1990 — 1994
1995 — 1999
2000 — 2004
2005 — 2009
2010 — 2014
2015 — 2019
Venezuela (Bolivarian Republic of)
Uruguay
United States of America (the)
Turks and Caicos Islands (the)
Trinidad and Tobago
Suriname
Saint Vincent and the Grenadines
Saint Lucia
Saint Kitts and Nevis
Puerto Rico
Peru
Paraguay
Panama
Nicaragua
Netherlands Antilles
Mexico
Jamaica
Honduras
Haiti
Guyana
Guatemala
Guadeloupe
Grenada
El Salvador
Ecuador
Dominican Republic (the)
Dominica
Cuba
Costa Rica
Colombia
Chile
Canada
Brazil
Bolivia (Plurinational State of)
Bermuda
Belize
Barbados
Bahamas (the)
Argentina
Antigua and Barbuda
0% 20% 40% 60% 80%

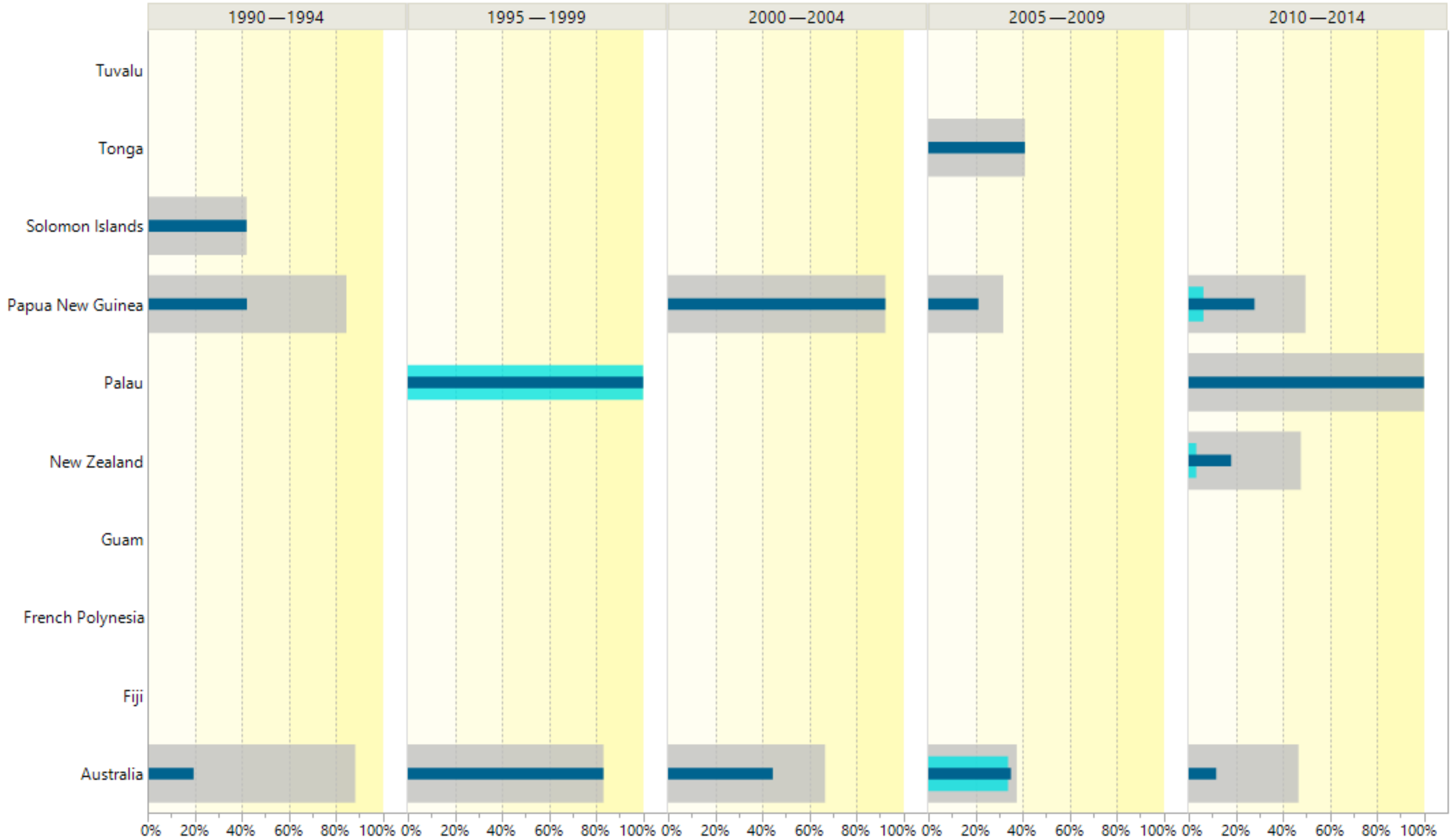
Asia


1990 — 1994
1995 — 1999
2000 — 2004
2005 — 2009
2010 — 2014
2015 — 2019
Yemen
Vietnam
Uzbekistan
United Arab Emirates (the)
Turkmenistan
Turkey
Timor-Leste
Thailand
Tajikistan
Taiwan (Province of China)
Syrian Arab Republic
Sri Lanka
Singapore
Saudi Arabia
Qatar
Philippines (the)
Palestine, State of
Pakistan
Oman
Nepal
Myanmar
Mongolia
Maldives
Malaysia
Macao
Lebanon
Lao People's Democratic Republic (the)
Kyrgyzstan
Kuwait
Korea (the Republic of)
Korea (the Democratic People's Republic of)
Kazakhstan
Jordan
Japan
Israel
Iraq
Iran (Islamic Republic of)
Indonesia
India
Hong Kong, China (SAR)
Georgia
Cyprus
China
Cambodia
Bhutan
Bangladesh
Bahrain
Azerbaijan
Armenia
Afghanistan
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%

Europe

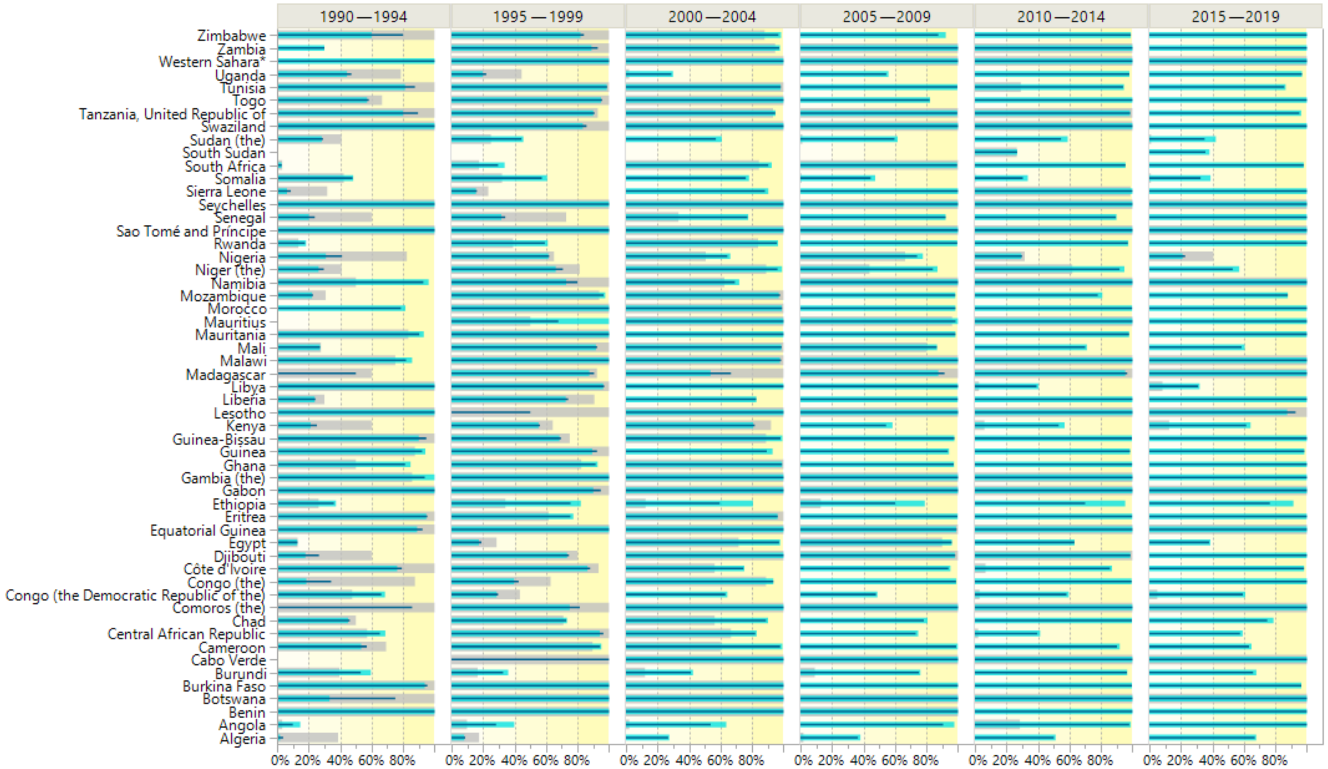

1990 — 1994
1995 — 1999
2000 — 2004
2005 — 2009
2010 — 2014
2015 — 2019
United Kingdom of Great Britain and Northern Ireland (the)
Ukraine
Switzerland
Sweden
Spain
Slovenia
Slovakia
Serbia
Russian Federation (the)
Romania
Portugal
Poland
Norway
Netherlands (the)
Montenegro
Moldova (the Republic of)
Malta
Macedonia (the former Yugoslav Republic of)
Luxembourg
Lithuania
Latvia
Kosovo
Italy
Ireland
Hungary
Greece
Germany
France
Finland
Estonia
Denmark
Czech Republic (the)
Croatia
Bulgaria
Bosnia and Herzegovina
Belgium
Belarus
Austria
Albania
0%
40% 60%

Oceania

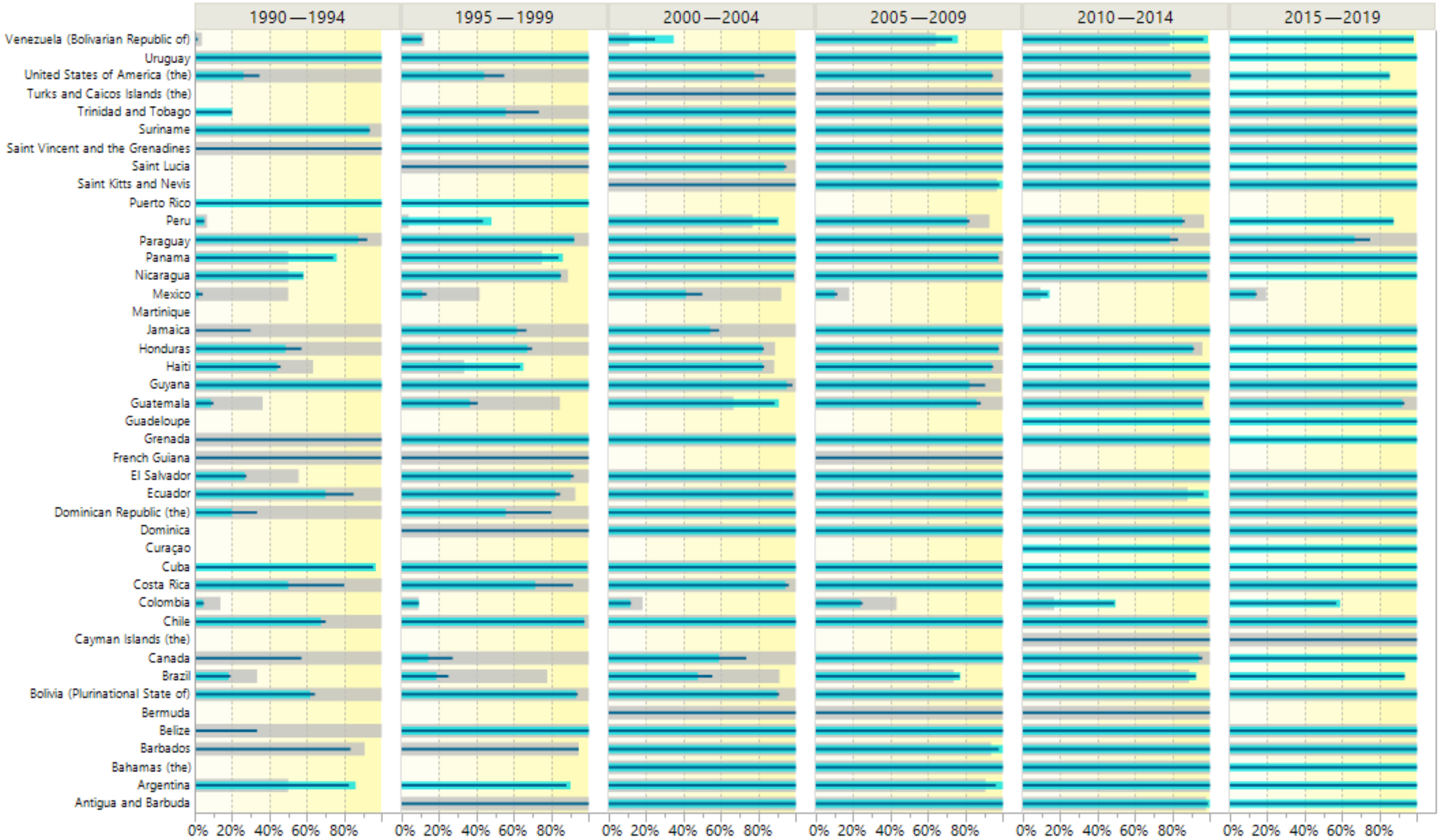

1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
Tuvalu
Tonga
Solomon Islands
Papua New Guinea
Palau
New Zealand
Guam
French Polynesia
Fiji
Australia
0%
20%
40%
60%
80%
100%

G.3.3: Deviant

Africa

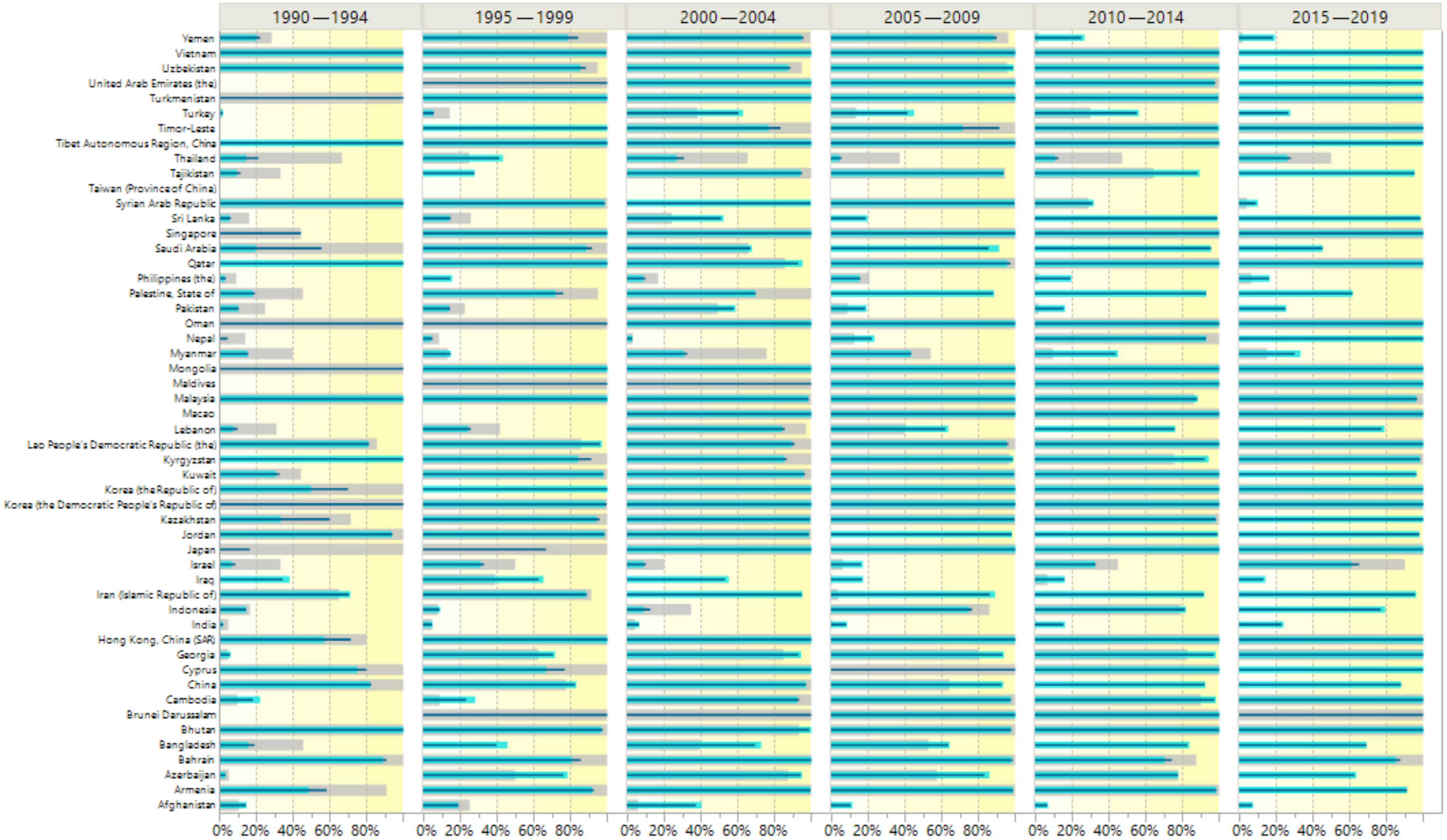


Americas

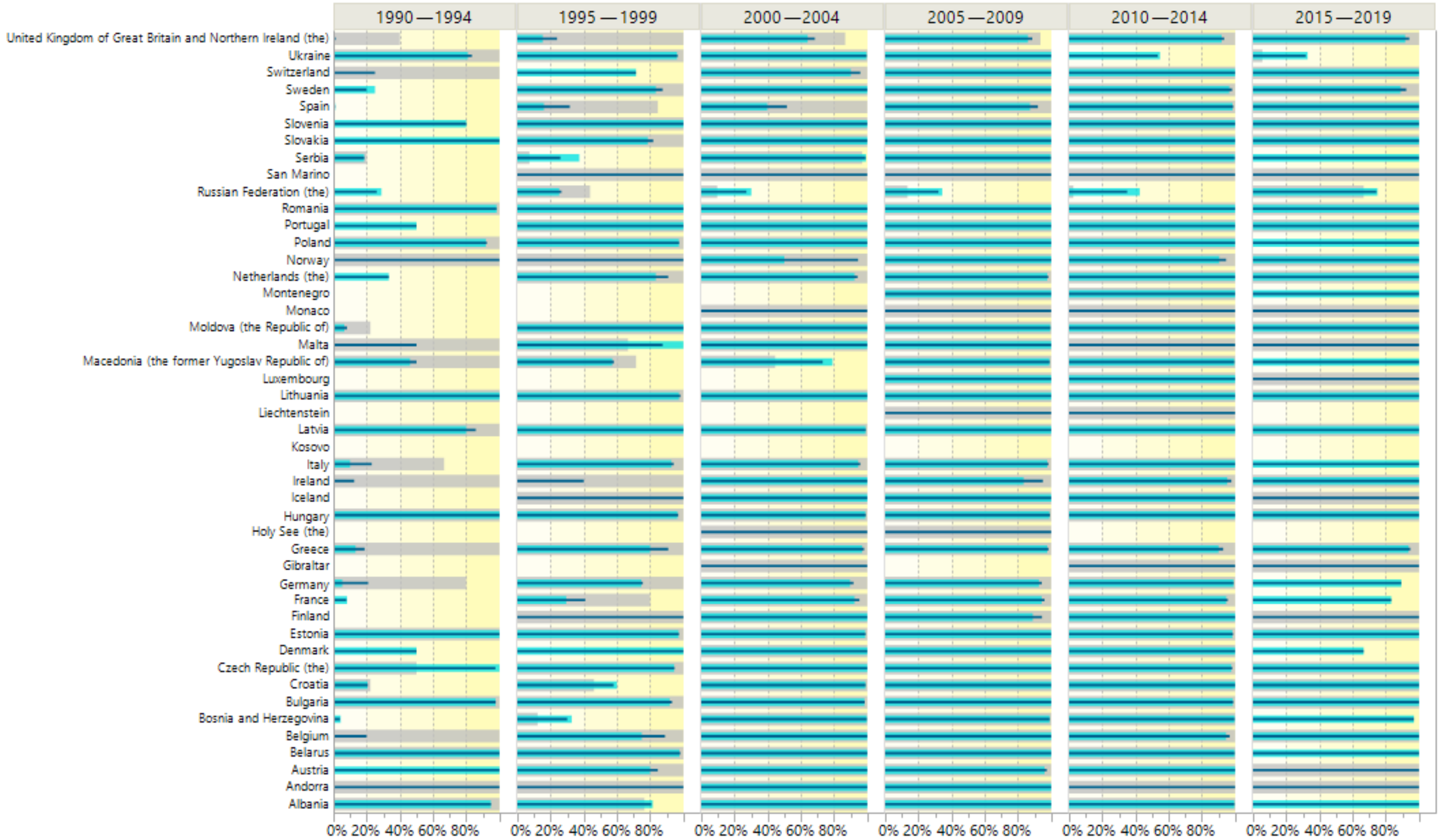


1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
Venezuela (Bolivarian Republic of)
Uruguay
United States of America (the)
Turks and Caicos Islands (the)
Trinidad and Tobago
Suriname
Saint Vincent and the Grenadines
Saint Lucia
Saint Kitts and Nevis
Puerto Rico
Peru
Paraguay
Panama
Nicaragua
Mexico
Martinique
Jamaica
Honduras
Haiti
Guyana
Guatemala
Guadeloupe
Grenada
French Guiana
El Salvador
Ecuador
Dominican Republic (the)
Dominica
Curaçao
Cuba
Costa Rica
Colombia
Chile
Cayman Islands (the)
Canada
Brazil
Bolivia (Plurinational State of)
Bermuda
Belize
Barbados
Bahamas (the)
Argentina
Antigua and Barbuda
0% 20% 40% 60% 80%

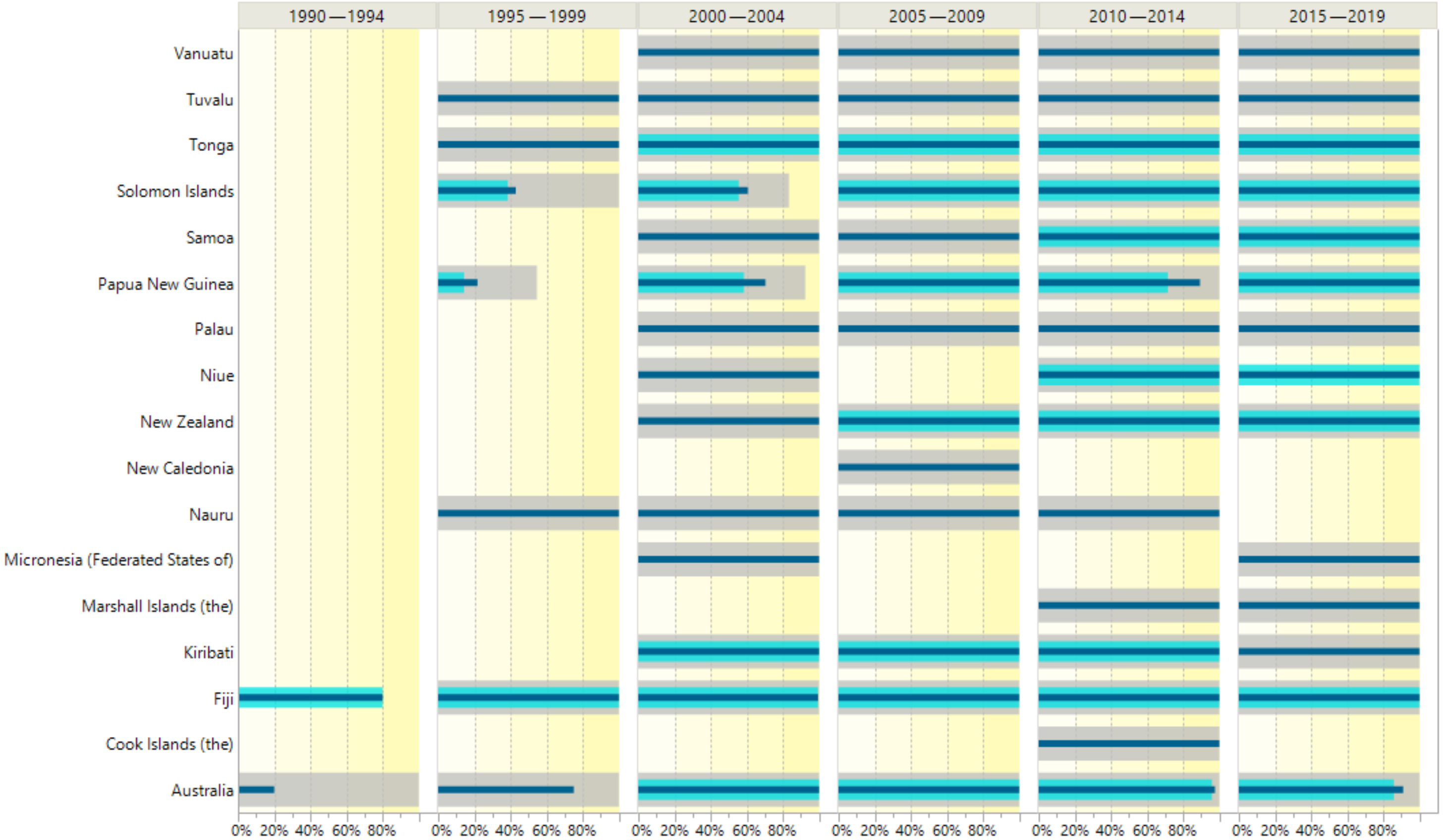
Asia


1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
Yemen
Vietnam
Uzbekistan
United Arab Emirates (the)
Turkmenistan
Turkey
Timor-Leste
Tibet Autonomous Region, China
Thailand
Tajikistan
Taiwan (Province of China)
Syrian Arab Republic
Sri Lanka
Singapore
Saudi Arabia
Qatar
Philippines (the)
Palestine, State of
Pakistan
Oman
Nepal
Myanmar
Mongolia
Maldives
Malaysia
Macao
Lebanon
Lao People's Democratic Republic (the)
Kyrgyzstan
Kuwait
Korea (the Republic of)
Korea (the Democratic People's Republic of)
Kazakhstan
Jordan
Japan
Israel
Iraq
Iran (Islamic Republic of)
Indonesia
India
Hong Kong, China (SAR)
Georgia
Cyprus
China
Cambodia
Brunei Darussalam
Bhutan
Bangladesh
Bahrain
Azerbaijan
Armenia
Afghanistan
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%
0% 20% 40% 60% 80%

Europe


1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
United Kingdom of Great Britain and Northern Ireland (the)
Ukraine
Switzerland
Sweden
Spain
Slovenia
Slovakia
Serbia
San Marino
Russian Federation (the)
Romania
Portugal
Poland
Norway
Netherlands (the)
Montenegro
Monaco
Moldova (the Republic of)
Malta
Macedonia (the former Yugoslav Republic of)
Luxembourg
Lithuania
Liechtenstein
Latvia
Kosovo
Italy
Ireland
Iceland
Hungary
Holy See (the)
Greece
Gibraltar
Germany
France
Finland
Estonia
Denmark
Czech Republic (the)
Croatia
Bulgaria
Bosnia and Herzegovina
Belgium
Belarus
Austria
Andorra
Albania
0% 20% 40% 60% 80%

Oceania



1990—1994
1995—1999
2000—2004
2005—2009
2010—2014
2015—2019
Vanuatu
Tuvalu
Tonga
Solomon Islands
Samoa
Papua New Guinea
Palau
Niue
New Zealand
New Caledonia
Nauru
Micronesia (Federated States of)
Marshall Islands (the)
Kiribati
Fiji
Cook Islands (the)
Australia
0% 20% 40% 60% 80%

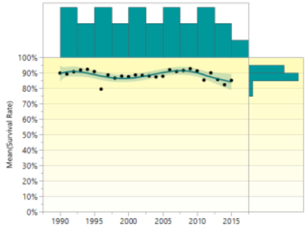
G.4: Mean Survival Rate (Fit Line Charts)

G.4.1: By Disaster Group

Naturogenic Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -15.75337 + 0.0083063*YEAR + 0.000686*(YEAR-2002.41)^2 - 0.0001709*(YEAR-2002.41)^3 - 8.0561e-6*(YEAR-2002.41)^4 + 6.6852e-7*(YEAR-2002.41)^5 + 2.0168e-8*(YEAR-2002.41)^6

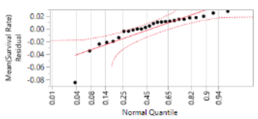
Summary of Fit

RSquare	0.454107
RSquare Adj	0.281719
Root Mean Square Error	0.026792
Mean of Response	0.885549
Observations (or Sum Wgts)	26

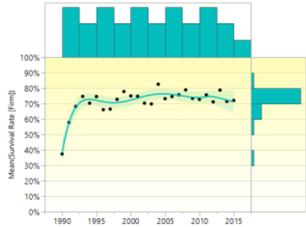
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.01134563	0.001891	2.6342
Error	19	0.01363888	0.000718	Prob > F
C. Total	25	0.02498452		0.0496*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -18.65086 + 0.0096902*YEAR - 0.0016977*(YEAR-2002.41)^2 - 0.0002543*(YEAR-2002.41)^3 + 0.0000357*(YEAR-2002.41)^4 + 1.8198e-6*(YEAR-2002.41)^5 - 2.156e-7*(YEAR-2002.41)^6

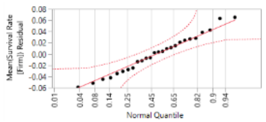
Summary of Fit

RSquare	0.840663
RSquare Adj	0.790346
Root Mean Square Error	0.038694
Mean of Response	0.714793
Observations (or Sum Wgts)	26

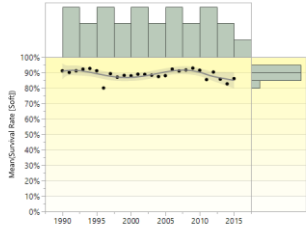
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.15008567	0.025014	16.7073
Error	19	0.02844687	0.001497	Prob > F
C. Total	25	0.17853254		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -14.87966 + 0.0078713*YEAR + 0.0007433*(YEAR-2002.41)^2 - 0.0001647*(YEAR-2002.41)^3 - 9.5257e-6*(YEAR-2002.41)^4 + 6.3701e-7*(YEAR-2002.41)^5 + 2.9513e-8*(YEAR-2002.41)^6

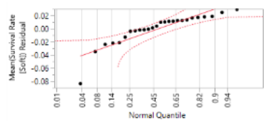
Summary of Fit

RSquare	0.443829
RSquare Adj	0.268196
Root Mean Square Error	0.026758
Mean of Response	0.889575
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.01085560	0.001809	2.5270
Error	19	0.01360336	0.000716	Prob > F
C. Total	25	0.02445896		0.0571

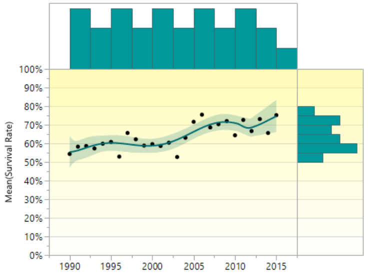
Residual Normal Quantile Plot



Anthropogenic Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -31.43918 + 0.0160055*YEAR + 0.0024456*(YEAR-2002.41)^2 - 0.0002141*(YEAR-2002.41)^3 - 3.713e-5*(YEAR-2002.41)^4 + 1.0171e-6*(YEAR-2002.41)^5 + 1.4711e-7*(YEAR-2002.41)^6

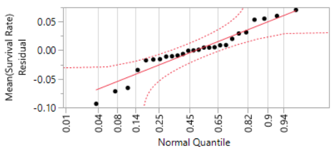
Summary of Fit

RSquare	0.681799
RSquare Adj	0.581314
Root Mean Square Error	0.044403
Mean of Response	0.638101
Observations (or Sum Wgts)	26

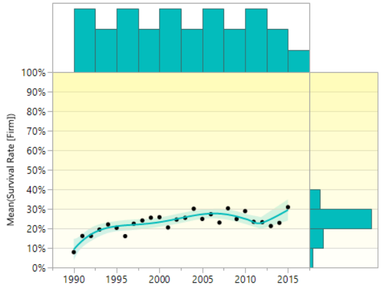
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.08026797	0.013378	6.7851
Error	19	0.03746170	0.001972	**Prob > F**
C. Total	25	0.11772968		0.0006*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -18.07931 + 0.0091549*YEAR + 0.0000031*(YEAR-2002.41)^2 - 0.0001773*(YEAR-2002.41)^3 - 9.218e-6*(YEAR-2002.41)^4 + 1.0898e-6*(YEAR-2002.41)^5 + 4.2015e-8*(YEAR-2002.41)^6

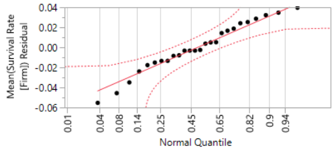
Summary of Fit

RSquare	0.772979
RSquare Adj	0.701289
Root Mean Square Error	0.027866
Mean of Response	0.229917
Observations (or Sum Wgts)	26

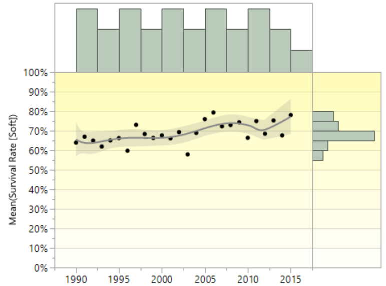
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.05023420	0.008372	10.7821
Error	19	0.01475356	0.000777	**Prob > F**
C. Total	25	0.06498776		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -20.09831 + 0.0103764*YEAR + 0.0015027*(YEAR-2002.41)^2 - 0.0001369*(YEAR-2002.41)^3 - 0.0000296*(YEAR-2002.41)^4 + 6.2433e-7*(YEAR-2002.41)^5 + 1.3642e-7*(YEAR-2002.41)^6

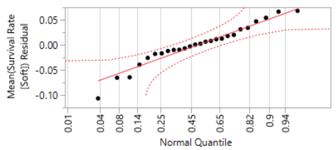
Summary of Fit

RSquare	0.450127
RSquare Adj	0.276483
Root Mean Square Error	0.046167
Mean of Response	0.689314
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.03315003	0.005525	2.5922
Error	19	0.04049594	0.002131	**Prob > F**
C. Total	25	0.07364598		0.0524

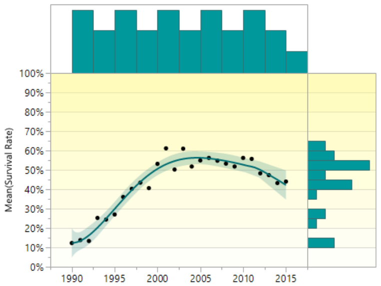
Residual Normal Quantile Plot



Deviant Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -21.86713 + 0.0111957*YEAR - 0.0029816*(YEAR-2002.41)^2 + 0.0001189*(YEAR-2002.41)^3 + 9.261e-6*(YEAR-2002.41)^4 - 7.2322e-7*(YEAR-2002.41)^5 - 9.5876e-9*(YEAR-2002.41)^6

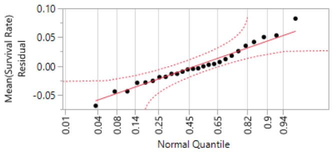
Summary of Fit

RSquare	0.948717
RSquare Adj	0.932522
Root Mean Square Error	0.038734
Mean of Response	0.430611
Observations (or Sum Wgts)	26

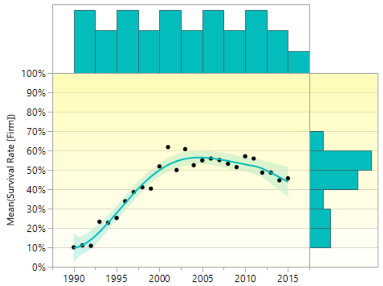
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.52733581	0.087889	58.5816
Error	19	0.02850548	0.001500	Prob > F
C. Total	25	0.55584129		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -25.28078 + 0.0129003*YEAR - 0.0033786*(YEAR-2002.41)^2 + 0.0001136*(YEAR-2002.41)^3 + 1.4249e-5*(YEAR-2002.41)^4 - 6.9034e-7*(YEAR-2002.41)^5 - 2.6351e-8*(YEAR-2002.41)^6

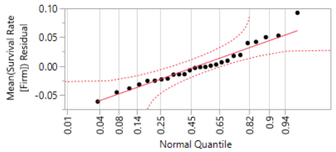
Summary of Fit

RSquare	0.953389
RSquare Adj	0.93867
Root Mean Square Error	0.039346
Mean of Response	0.425018
Observations (or Sum Wgts)	26

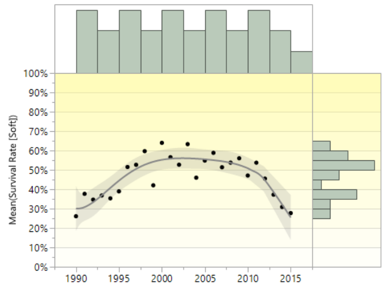
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.60164542	0.100274	64.7720
Error	19	0.02941412	0.001548	Prob > F
C. Total	25	0.63105954		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = 0.7270926 - 8.3211e-5*YEAR - 0.0010449*(YEAR-2002.41)^2 + 0.0001349*(YEAR-2002.41)^3 - 1.0173e-5*(YEAR-2002.41)^4 - 9.2308e-7*(YEAR-2002.41)^5 + 3.504e-8*(YEAR-2002.41)^6

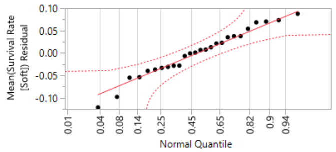
Summary of Fit

RSquare	0.775329
RSquare Adj	0.704381
Root Mean Square Error	0.059722
Mean of Response	0.467439
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.23386196	0.038977	10.9280
Error	19	0.06776721	0.003567	Prob > F
C. Total	25	0.30162917		<.0001*

Residual Normal Quantile Plot

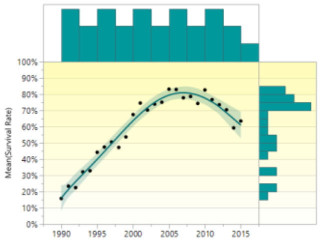


G.4.2: by Region

Africa Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -64.15623 + 0.0324047*YEAR - 0.0030168*(YEAR-2002.5)^2 - 0.0001505*(YEAR-2002.5)^3 + 8.586e-6*(YEAR-2002.5)^4 + 3.7342e-7*(YEAR-2002.5)^5 - 2.2959e-8*(YEAR-2002.5)^6

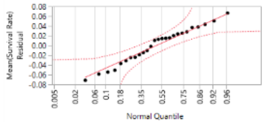
Summary of Fit

RSquare	0.969052
RSquare Adj	0.959279
Root Mean Square Error	0.041815
Mean of Response	0.597659
Observations (or Sum Wgts)	26

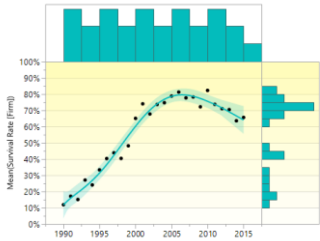
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	1.0402259	0.173371	99.1565
Error	19	0.0332207	0.001748	**Prob > F**
C. Total	25	1.0734466		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -75.15419 + 0.037891*YEAR - 0.0046401*(YEAR-2002.5)^2 - 0.0001735*(YEAR-2002.5)^3 + 0.0000277*(YEAR-2002.5)^4 + 4.1493e-7*(YEAR-2002.5)^5 - 7.716e-8*(YEAR-2002.5)^6

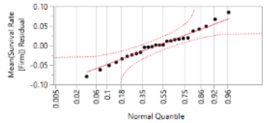
Summary of Fit

RSquare	0.972667
RSquare Adj	0.964035
Root Mean Square Error	0.043874
Mean of Response	0.566425
Observations (or Sum Wgts)	26

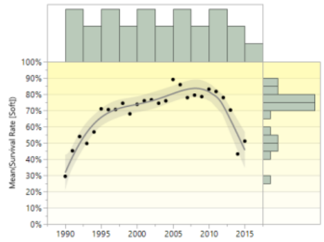
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	1.3015145	0.216919	112.6873
Error	19	0.0365743	0.001925	**Prob > F**
C. Total	25	1.3380888		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -26.24855 + 0.0134901*YEAR + 0.0010862*(YEAR-2002.5)^2 - 5.4753e-5*(YEAR-2002.5)^3 - 3.1958e-5*(YEAR-2002.5)^4 + 2.4557e-8*(YEAR-2002.5)^5 + 6.0756e-8*(YEAR-2002.5)^6

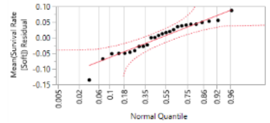
Summary of Fit

RSquare	0.887984
RSquare Adj	0.85261
Root Mean Square Error	0.057074
Mean of Response	0.686274
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.49062346	0.081771	25.1030
Error	19	0.06189057	0.003257	**Prob > F**
C. Total	25	0.55251403		<.0001*

Residual Normal Quantile Plot

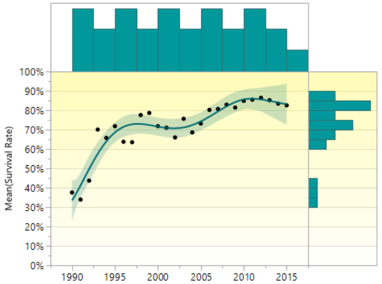


Appendix G: Mean Survival Rates (M^iO)

Americas Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -8.526564 + 0.0046118*YEAR + 0.0036839*(YEAR-2002.5)^2 + 0.0001324*(YEAR-2002.5)^3 - 5.5244e-5*(YEAR-2002.5)^4 - 2.2365e-7*(YEAR-2002.5)^5 + 1.6977e-7*(YEAR-2002.5)^6

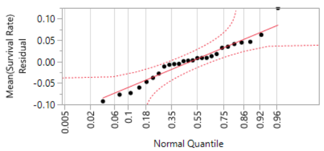
Summary of Fit

RSquare	0.888774
RSquare Adj	0.85365
Root Mean Square Error	0.054508
Mean of Response	0.717677
Observations (or Sum Wgts)	26

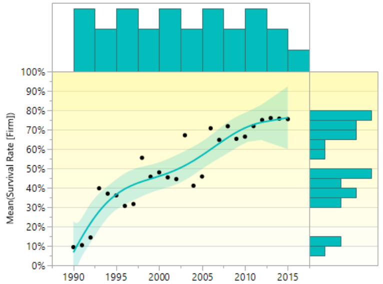
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.45108905	0.075182	25.3040
Error	19	0.05645161	0.002971	Prob > F
C. Total	25	0.50754065		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -42.50745 + 0.0214788*YEAR + 0.0018331*(YEAR-2002.5)^2 + 1.6524e-5*(YEAR-2002.5)^3 - 2.5642e-5*(YEAR-2002.5)^4 + 1.5289e-7*(YEAR-2002.5)^5 + 6.5518e-8*(YEAR-2002.5)^6

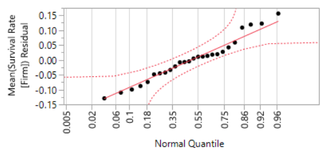
Summary of Fit

RSquare	0.878112
RSquare Adj	0.839621
Root Mean Square Error	0.083473
Mean of Response	0.505993
Observations (or Sum Wgts)	26

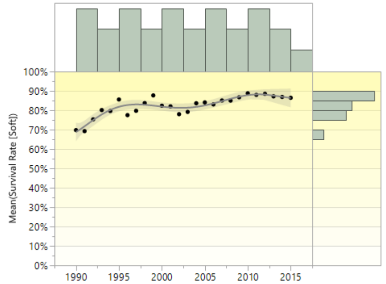
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.9537595	0.158960	22.8135
Error	19	0.1323881	0.006968	Prob > F
C. Total	25	1.0861476		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -0.882095 + 0.0008464*YEAR + 0.0016827*(YEAR-2002.5)^2 + 0.0000655*(YEAR-2002.5)^3 - 0.0000223*(YEAR-2002.5)^4 - 1.6661e-7*(YEAR-2002.5)^5 + 6.4197e-8*(YEAR-2002.5)^6

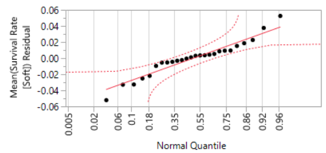
Summary of Fit

RSquare	0.832161
RSquare Adj	0.779159
Root Mean Square Error	0.024896
Mean of Response	0.824468
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.05839041	0.009732	15.7006
Error	19	0.01177681	0.000620	Prob > F
C. Total	25	0.07016721		<.0001*

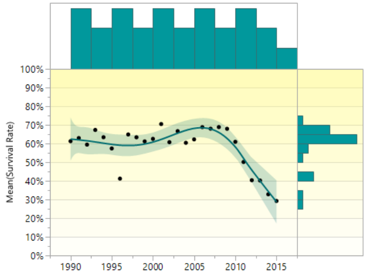
Residual Normal Quantile Plot



Asia Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -33.43051 + 0.0170177*YEAR - 0.0002971*(YEAR-2002.5)^2 - 0.0003895*(YEAR-2002.5)^3 - 1.7378e-5*(YEAR-2002.5)^4 + 1.2435e-6*(YEAR-2002.5)^5 + 7.3124e-8*(YEAR-2002.5)^6

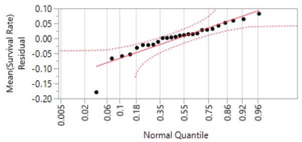
Summary of Fit

RSquare	0.800276
RSquare Adj	0.737206
Root Mean Square Error	0.059746
Mean of Response	0.581914
Observations (or Sum Wgts)	26

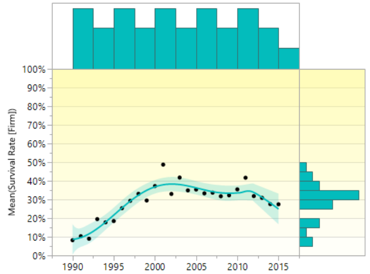
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.27175403	0.045292	12.6886
Error	19	0.06782122	0.003570	Prob > F
C. Total	25	0.33957525		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = 5.6572074 - 0.0026341*YEAR - 0.0028313*(YEAR-2002.5)^2 + 0.0002599*(YEAR-2002.5)^3 + 2.088e-5*(YEAR-2002.5)^4 - 1.288e-6*(YEAR-2002.5)^5 - 7.39e-8*(YEAR-2002.5)^6

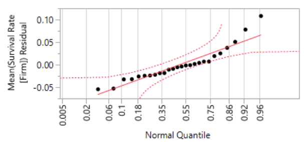
Summary of Fit

RSquare	0.865476
RSquare Adj	0.822994
Root Mean Square Error	0.042292
Mean of Response	0.291433
Observations (or Sum Wgts)	26

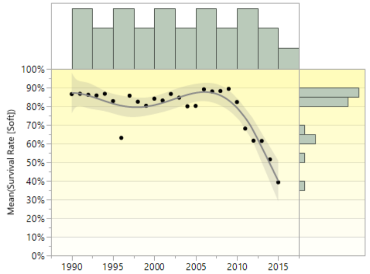
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.21864015	0.036440	20.3731
Error	19	0.03398413	0.001789	Prob > F
C. Total	25	0.25262428		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -29.62247 + 0.0152123*YEAR - 0.0001021*(YEAR-2002.5)^2 - 0.0003419*(YEAR-2002.5)^3 - 1.1262e-5*(YEAR-2002.5)^4 + 7.9175e-7*(YEAR-2002.5)^5 + 2.2603e-8*(YEAR-2002.5)^6

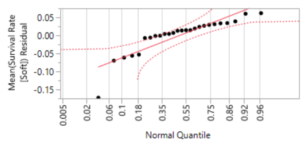
Summary of Fit

RSquare	0.856173
RSquare Adj	0.810754
Root Mean Square Error	0.056275
Mean of Response	0.785925
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.35818153	0.059697	18.8505
Error	19	0.06017031	0.003167	Prob > F
C. Total	25	0.41835184		<.0001*

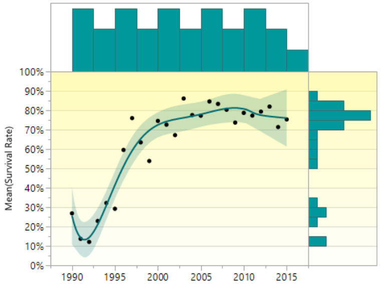
Residual Normal Quantile Plot



Europe Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -15.71109 + 0.0082241*YEAR - 0.0005469*(YEAR-2002.5)^2 + 0.0004557*(YEAR-2002.5)^3 - 5.3556e-5*(YEAR-2002.5)^4 - 2.4281e-6*(YEAR-2002.5)^5 + 2.9982e-7*(YEAR-2002.5)^6

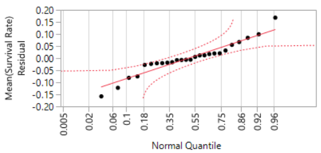
Summary of Fit

RSquare	0.921465
RSquare Adj	0.896665
Root Mean Square Error	0.076267
Mean of Response	0.627156
Observations (or Sum Wgts)	26

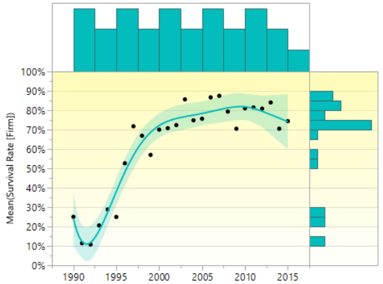
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	1.2966968	0.216116	37.1551
Error	19	0.1105152	0.005817	**Prob > F**
C. Total	25	1.4072120		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -18.22788 + 0.0094836*YEAR - 0.0014258*(YEAR-2002.5)^2 + 0.0005084*(YEAR-2002.5)^3 - 4.0053e-5*(YEAR-2002.5)^4 - 2.8251e-6*(YEAR-2002.5)^5 + 2.4366e-7*(YEAR-2002.5)^6

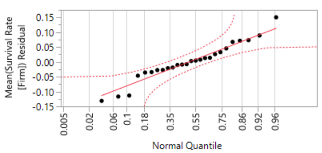
Summary of Fit

RSquare	0.935447
RSquare Adj	0.915062
Root Mean Square Error	0.07273
Mean of Response	0.621472
Observations (or Sum Wgts)	26

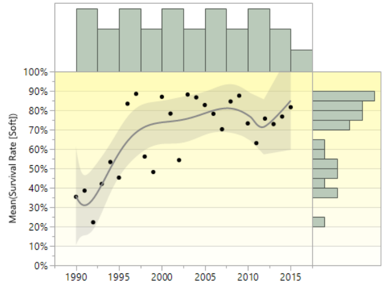
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	1.4564110	0.242735	45.8889
Error	19	0.1005030	0.005290	**Prob > F**
C. Total	25	1.5569139		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -17.12826 + 0.0089292*YEAR + 0.0019062*(YEAR-2002.5)^2 + 7.8432e-5*(YEAR-2002.5)^3 - 7.9877e-5*(YEAR-2002.5)^4 - 6.1698e-8*(YEAR-2002.5)^5 + 3.9432e-7*(YEAR-2002.5)^6

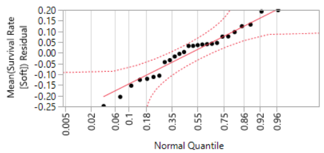
Summary of Fit

RSquare	0.653295
RSquare Adj	0.543809
Root Mean Square Error	0.131032
Mean of Response	0.674669
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.61469532	0.102449	5.9669
Error	19	0.32622006	0.017169	**Prob > F**
C. Total	25	0.94091538		0.0012*

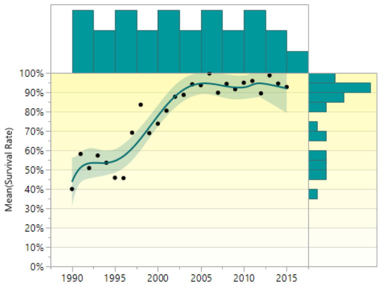
Residual Normal Quantile Plot



Oceania Disasters

Polynomial Line of Fit (Degree=6)

Survival Rate (Undifferentiated)



Mean(Survival Rate) = -68.64168 + 0.0347263*YEAR - 0.0062552*(YEAR-2002.5)^2 - 0.0002045*(YEAR-2002.5)^3 + 7.6252e-5*(YEAR-2002.5)^4 + 6.7723e-7*(YEAR-2002.5)^5 - 2.8931e-7*(YEAR-2002.5)^6

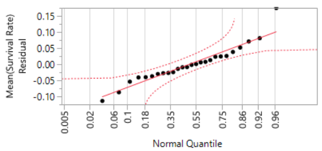
Summary of Fit

RSquare	0.91432
RSquare Adj	0.887263
Root Mean Square Error	0.064885
Mean of Response	0.7819
Observations (or Sum Wgts)	26

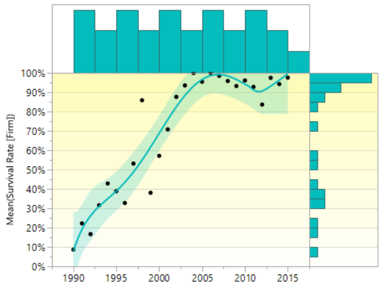
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.85362825	0.142271	33.7927
Error	19	0.07999241	0.004210	Prob > F
C. Total	25	0.93362065		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm]) = -116.9806 + 0.0588445*YEAR - 0.0047514*(YEAR-2002.5)^2 - 0.0005311*(YEAR-2002.5)^3 + 2.8551e-5*(YEAR-2002.5)^4 + 2.4926e-6*(YEAR-2002.5)^5 - 7.1444e-8*(YEAR-2002.5)^6

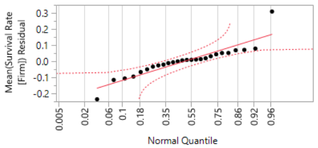
Summary of Fit

RSquare	0.908414
RSquare Adj	0.879492
Root Mean Square Error	0.107547
Mean of Response	0.701833
Observations (or Sum Wgts)	26

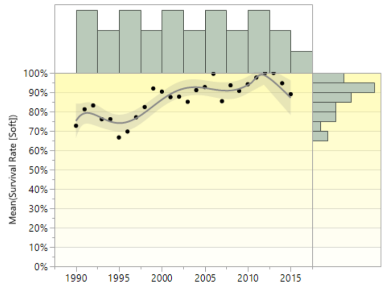
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	2.1797586	0.363293	31.4093
Error	19	0.2197621	0.011566	Prob > F
C. Total	25	2.3995207		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft]) = -22.30829 + 0.011599*YEAR - 0.0047008*(YEAR-2002.5)^2 + 6.6587e-5*(YEAR-2002.5)^3 + 7.6913e-5*(YEAR-2002.5)^4 - 7.0536e-7*(YEAR-2002.5)^5 - 3.2773e-7*(YEAR-2002.5)^6

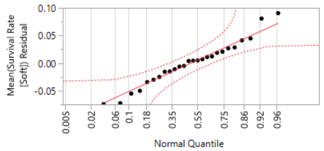
Summary of Fit

RSquare	0.813713
RSquare Adj	0.754885
Root Mean Square Error	0.046378
Mean of Response	0.86781
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	6	0.17851142	0.029752	13.8322
Error	19	0.04086746	0.002151	Prob > F
C. Total	25	0.21937888		<.0001*

Residual Normal Quantile Plot



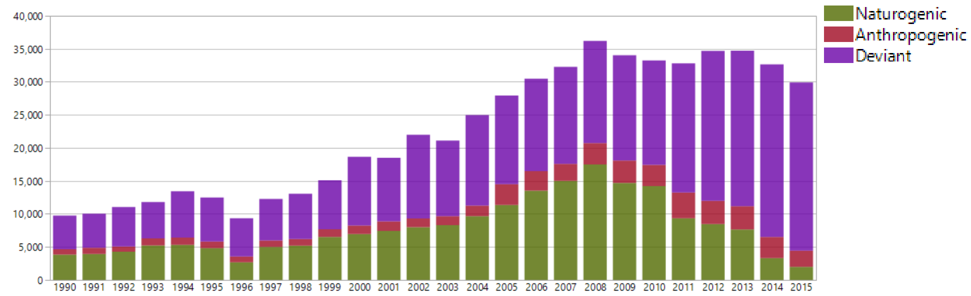
Appendix H: Miscellaneous Supplemental Material

H.1: MAX DVi, Weighting Permutations

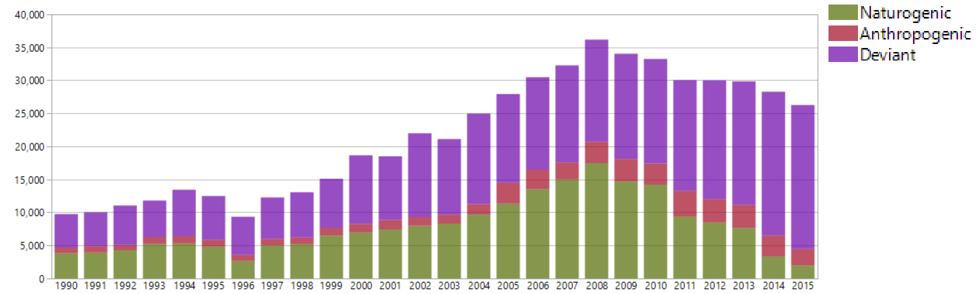
Veracity DVi - Max Permutations			Dataset MAX			L1 MAX			L2 MAX			L3 MAX		
L1	L2	L3	Weigting	Score	Index	Weigting	Score	Index	Weigting	Score	Index	Weigting	Score	Index
Elucidatory	Complete	1. No omitted entries	3	5	1.36	3	5	2.14	3	5	2.37	3	5	3.46
		2. No omitted values	3	5	1.36	3	5	2.14	3	5	2.37	1	5	0.38
		3. No omitted variables	3	5	1.36	3	5	2.14	3	5	2.37	1	5	0.38
		4. No omitted metadata	3	5	1.36	3	5	2.14	3	5	2.37	1	5	0.38
		Complete			1.36			2.14			2.37			1.15
	Uncluttered	5. No irrelevant entries	3	5	1.36	3	5	2.14	1	5	0.26	1	5	0.38
		Uncluttered			1.36			2.14			0.26			0.38
	Elucidatory Index				1.36			2.14			1.95			1
Expository	Precise	6. Reliability	3	5	1.36	1	5	0.24	1	5	0.26	1	5	0.38
		7. Rigour	3	5	1.36	1	5	0.24	1	5	0.26	1	5	0.38
		8. Congruity	3	5	1.36	1	5	0.24	1	5	0.26	1	5	0.38
		Precise			1.36			0.24			0.26			0.38
	Accurate	9. Conformity	3	5	1.36	1	5	0.24	1	5	0.26	1	5	0.38
		10. Impartiality	3	5	1.36	1	5	0.24	1	5	0.26	1	5	0.38
		11. Validity	3	5	1.36	1	5	0.24	1	5	0.26	1	5	0.38
		Accurate			1.36			0.24			0.26			0.38
	Expository Index				1.36			0.24			0.26			0.38
Data Veracity index (DVi)			33		1.36	21		1.1	19		1.03	13		0.66

H.2: MSGD Disaster Entries x Year

Stacked by Disaster Group (all sources)

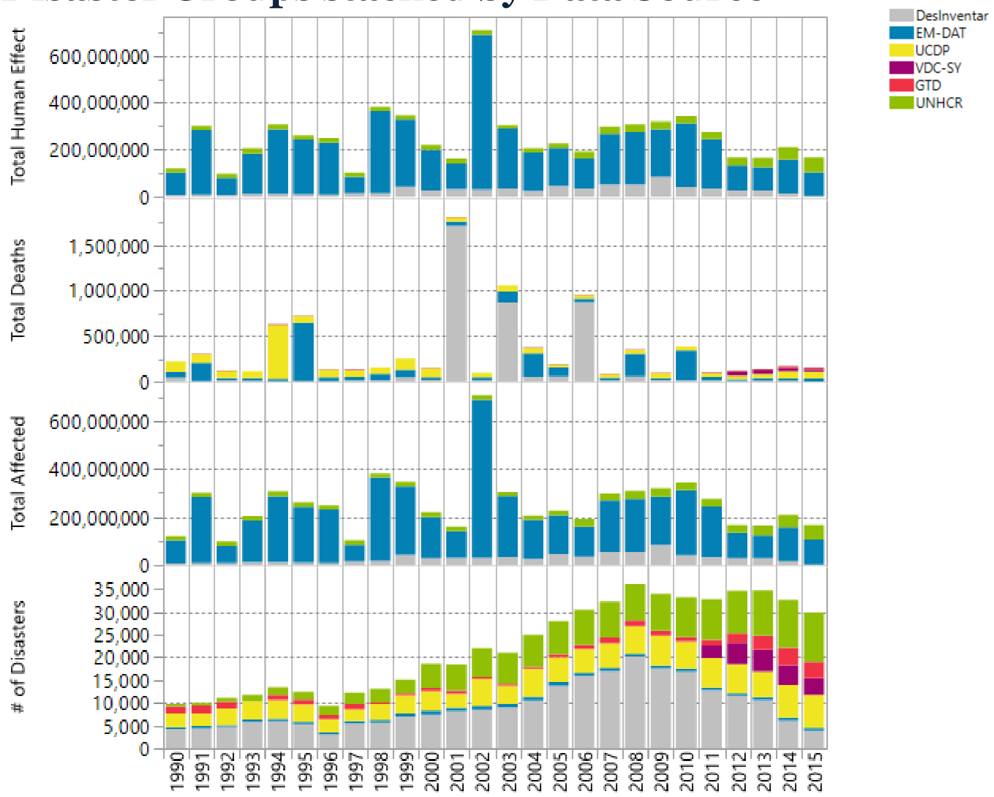


Stacked by Disaster Group (ex VDC-SY)



H.3: Disasters and Human Effect

Disaster Groups stacked by Data Source



H.4: Human Effect 'Spikes' – 2001,2002,2003 & 2006

YEAR	Country	Event	Entries	Total Human Effect	Total Human Effect [Firm]	Total Human Effect [Soft]	People Affected	People Affected [Firm]	People Affected [Soft]	Deaths	Deaths [Firm]	Deaths [Soft]
2001	India	Earthquake	1	6,341,817	6,341,817	-	6,321,812	6,321,812		20,005	20,005	
2001	Pakistan	Flood	2	1,693,014	-	1,693,014	945	-	945	1,692,069	-	1,692,069
2002	Afghanistan	Disease	1	200,000	-	200,000	200,000	-	200,000		-	
2002	Albania	Convective Storm	1	125,006	-	125,006	125,000	-	125,000	6	-	6
2002	Argentina	Flood	1	250,000	-	250,000	250,000	-	250,000		-	
2002	Bangladesh	Flood	1	1,500,010	-	1,500,010	1,500,000	-	1,500,000	10	-	10
2002	Bangladesh	Tropical Cyclone	1	100,431	-	100,431	100,400	-	100,400	31	-	31
2002	Brazil	Disease	1	317,787	-	317,787	317,730	-	317,730	57	-	57
2002	Cambodia	Drought	9	3,719,905	-	3,719,905	3,719,905	-	3,719,905		-	
2002	Cambodia	Flood	3	2,079,126	-	2,079,126	2,079,097	-	2,079,097	29	-	29
2002	Chile	Flood	1	221,856	221,856	-	221,842	221,842		14	14	
2002	China	Convective Storm	2	100,220,137	-	100,220,137	100,220,110	-	100,220,110	27	-	27
2002	China	Drought	3	64,560,000	-	64,560,000	64,560,000	-	64,560,000		-	
2002	China	Flood	6	113,187,520	81,035,539	32,151,981	113,186,368	81,034,568	32,151,800	1,152	971	181
2002	China	Storm	1	7,000,066	-	7,000,066	7,000,040	-	7,000,040	26	-	26
2002	China	Tropical Cyclone	1	180,025	-	180,025	180,000	-	180,000	25	-	25
2002	Congo (the Democratic Republic of the)	Disease	1	502,000	-	502,000	500,000	-	500,000	2,000	-	2,000
2002	Congo (the Democratic Republic of the)	Volcanic Activity	1	110,600	-	110,600	110,400	-	110,400	200	-	200
2002	Cuba	Tropical Cyclone	1	281,473	281,473	-	281,470	281,470		3	3	
2002	Czech Republic (the)	Flood	1	200,018	-	200,018	200,000	-	200,000	18	-	18
2002	Ecuador	Volcanic Activity	1	128,150	-	128,150	128,150	-	128,150		-	
2002	Ethiopia	Drought	3	1,766,372	-	1,766,372	1,766,372	-	1,766,372		-	
2002	Germany	Flood	1	330,135	-	330,135	330,108	-	330,108	27	-	27
2002	Guinea-Bissau	Drought	1	100,000	-	100,000	100,000	-	100,000		-	
2002	Honduras	Flood	1	100,000	-	100,000	100,000	-	100,000		-	
2002	India	Drought	1	300,000,000	-	300,000,000	300,000,000	-	300,000,000		-	
2002	India	Flood	1	42,000,549	-	42,000,549	42,000,000	-	42,000,000	549	-	549
2002	Indonesia	Flood	3	1,106,637	-	1,106,637	1,106,487	-	1,106,487	150	-	150
2002	Iran (Islamic Republic of)	Earthquake	1	111,527	-	111,527	111,300	-	111,300	227	-	227
2002	Iran (Islamic Republic of)	Flood	1	200,039	-	200,039	200,000	-	200,000	39	-	39
2002	Jamaica	Flood	3	682,000	-	682,000	682,000	-	682,000		-	
2002	Japan	Tropical Cyclone	1	100,023	100,023	-	100,018	100,018		5	5	
2002	Kenya	Flood	1	150,061	150,061	-	150,008	150,008		53	53	
2002	Lao People's Democratic Republic (the)	Flood	1	150,002	-	150,002	150,000	-	150,000	2	-	2
2002	Lesotho	Drought	1	500,000	-	500,000	500,000	-	500,000		-	
2002	Madagascar	Drought	1	600,000	-	600,000	600,000	-	600,000		-	
2002	Madagascar	Tropical Cyclone	1	526,220	-	526,220	526,200	-	526,200	20	-	20
2002	Malawi	Drought	1	2,829,935	-	2,829,935	2,829,435	-	2,829,435	500	-	500
2002	Malawi	Flood	2	396,349	-	396,349	396,340	-	396,340	9	-	9
2002	Mexico	Drought	1	962,000	-	962,000	962,000	-	962,000		-	

YEAR	Country	Event	Entries	Total Human Effect	Total Human Effect [Firm]	Total Human Effect [Soft]	People Affected	People Affected [Firm]	People Affected [Soft]	Deaths	Deaths [Firm]	Deaths [Soft]
2002	Mexico	Tropical Cyclone	4	2,927,574	-	2,927,574	2,927,480	-	2,927,480	94	-	94
2002	Mongolia	Storm	1	665,003	-	665,003	665,000	-	665,000	3	-	3
2002	Mozambique	Drought	12	1,876,040	-	1,876,040	1,876,040	-	1,876,040		-	
2002	Mozambique	Flood	2	200,022	-	200,022	200,016	-	200,016	6	-	6
2002	Mozambique	Tropical Cyclone	3	665,054	-	665,054	665,000	-	665,000	54	-	54
2002	Namibia	Drought	1	345,000	-	345,000	345,000	-	345,000		-	
2002	Nepal	Convective Storm	1	162,000	-	162,000	162,000	-	162,000		-	
2002	Nepal	Mass Movement	1	266,337	-	266,337	265,865	-	265,865	472	-	472
2002	Niger (the)	Drought	2	243,874	-	243,874	243,874	-	243,874		-	
2002	Pakistan	Convective Storm	1	143,557	-	143,557	143,509	-	143,509	48	-	48
2002	Pakistan	Earthquake	1	140,801	140,801	-	140,782	140,782		19	19	
2002	Philippines (the)	Storm	1	194,529	194,529	-	194,472	194,472		57	57	
2002	Philippines (the)	Tropical Cyclone	1	700,074	700,074	-	700,041	700,041		33	33	
2002	Russian Federation (the)	Flood	1	330,704	330,704	-	330,613	330,613		91	91	
2002	Senegal	Drought	1	284,000	-	284,000	284,000	-	284,000		-	
2002	Senegal	Flood	1	179,028	-	179,028	179,000	-	179,000	28	-	28
2002	South Africa	Convective Storm	1	100,022	-	100,022	100,000	-	100,000	22	-	22
2002	Sri Lanka	Flood	1	500,002	500,002	-	500,000	500,000		2	2	
2002	Sudan (the)	Flood	1	100,000	-	100,000	100,000	-	100,000		-	
2002	Thailand	Drought	1	5,000,000	-	5,000,000	5,000,000	-	5,000,000		-	
2002	Thailand	Flood	1	3,289,574	-	3,289,574	3,289,420	-	3,289,420	154	-	154
2002	Turkey	Earthquake	1	252,369	-	252,369	252,327	-	252,327	42	-	42
2002	Uganda	Drought	1	655,079	-	655,079	655,000	-	655,000	79	-	79
2002	United States of America (the)	Flood	1	144,010	-	144,010	144,000	-	144,000	10	-	10
2002	Vietnam	Drought	1	1,300,000	-	1,300,000	1,300,000	-	1,300,000		-	
2002	Vietnam	Flood	2	1,429,953	291,671	1,138,282	1,429,816	291,616	1,138,200	137	55	82
2003	Algeria	Earthquake	1	212,527	-	212,527	210,261	-	210,261	2,266	-	2,266
2003	Argentina	Flood	1	160,023	-	160,023	160,000	-	160,000	23	-	23
2003	Argentina	Storm	1	400,696	-	400,696	400,690	-	400,690	6	-	6
2003	Bangladesh	Flood	1	500,065	-	500,065	500,000	-	500,000	65	-	65
2003	Brazil	Flood	1	175,631	-	175,631	175,470	-	175,470	161	-	161
2003	Chile	Earthquake	1	156,158	-	156,158	156,158	-	156,158		-	
2003	Chile	Wildfire	1	128,578	-	128,578	128,578	-	128,578		-	
2003	China	Convective Storm	1	200,164	200,164	-	200,162	200,162		2	2	
2003	China	Drought	2	51,000,000	-	51,000,000	51,000,000	-	51,000,000		-	
2003	China	Earthquake	5	3,036,493	1,570,652	1,465,841	3,036,192	1,570,627	1,465,565	301	25	276
2003	China	Flood	4	155,917,607	4,371,024	151,546,583	155,916,986	4,370,986	151,546,000	621	38	583
2003	China	Tropical Cyclone	2	9,120,040	-	9,120,040	9,120,020	-	9,120,020	20	-	20
2003	Ethiopia	Drought	17	14,750,772	-	14,750,772	14,750,772	-	14,750,772		-	
2003	Ethiopia	Flood	4	420,183	-	420,183	420,000	-	420,000	183	-	183

YEAR	Country	Event	Entries	Total Human Effect	Total Human Effect [Firm]	Total Human Effect [Soft]	People Affected	People Affected [Firm]	People Affected [Soft]	Deaths	Deaths [Firm]	Deaths [Soft]
2003	Guyana	Drought	1	100,000	-	100,000	100,000	-	100,000		-	
2003	Haiti	Flood	1	150,038	-	150,038	150,000	-	150,000	38	-	38
2003	Honduras	Drought	1	185,000	-	185,000	185,000	-	185,000		-	
2003	India	Convective Storm	1	485,940	-	485,940	485,910	-	485,910	30	-	30
2003	India	Flood	8	8,605,019	-	8,605,019	8,604,802	-	8,604,802	217	-	217
2003	Indonesia	Flood	1	350,148	-	350,148	350,000	-	350,000	148	-	148
2003	Indonesia	Mass Movement	1	229,624	229,624	-	229,548	229,548		76	76	
2003	Iran (Islamic Republic of)	Earthquake	1	294,424	294,424	-	267,628	267,628		26,796	26,796	
2003	Madagascar	Tropical Cyclone	2	382,962	162,175	220,787	382,786	162,086	220,700	176	89	87
2003	Mexico	Convective Storm	1	160,060	-	160,060	160,060	-	160,060		-	
2003	Mexico	Earthquake	1	178,632	178,632	-	178,603	178,603		29	29	
2003	Mexico	Tropical Cyclone	1	303,500	-	303,500	303,500	-	303,500		-	
2003	Mozambique	Convective Storm	1	100,003	-	100,003	100,000	-	100,000	3	-	3
2003	Mozambique	Drought	5	715,019	-	715,019	715,010	-	715,010	9	-	9
2003	Mozambique	Flood	2	500,039	100,007	400,032	500,003	100,003	400,000	36	4	32
2003	Nicaragua	Storm	1	155,000	-	155,000	155,000	-	155,000		-	
2003	Nigeria	Flood	1	210,016	-	210,016	210,000	-	210,000	16	-	16
2003	Pakistan	Convective Storm	4	1,480,080	-	1,480,080	1,479,959	-	1,479,959	121	-	121
2003	Pakistan	Flood	4	2,739,588	1,266,453	1,473,135	2,739,243	1,266,223	1,473,020	345	230	115
2003	Peru	Extreme Temperat	1	1,840,227	1,840,227	-	1,839,888	1,839,888		339	339	
2003	Philippines (the)	Mass Movement	1	218,243	218,243	-	217,988	217,988		255	255	
2003	Philippines (the)	Tropical Cyclone	3	398,900	271,757	127,143	398,879	271,749	127,130	21	8	13
2003	Portugal	Wildfire	1	150,014	-	150,014	150,000	-	150,000	14	-	14
2003	Russian Federation (the)	Drought	1	1,000,000	1,000,000	-	1,000,000	1,000,000			-	
2003	Rwanda	Drought	1	1,000,000	-	1,000,000	1,000,000	-	1,000,000		-	
2003	Sri Lanka	Flood	1	695,235	695,235	-	695,000	695,000		235	235	
2003	Sudan (the)	Flood	1	325,076	-	325,076	325,056	-	325,056	20	-	20
2003	Tanzania, United Republic of	Drought	1	1,900,000	-	1,900,000	1,900,000	-	1,900,000		-	
2003	Thailand	Flood	1	104,706	-	104,706	104,700	-	104,700	6	-	6
2003	Turkey	Earthquake	1	290,697	-	290,697	290,520	-	290,520	177	-	177
2003	Uganda	Drought	2	550,045	-	550,045	550,045	-	550,045		-	
2003	United States of America (the)	Tropical Cyclone	1	225,016	-	225,016	225,000	-	225,000	16	-	16
2003	Vietnam	Flood	2	415,936	415,936	-	415,823	415,823		113	113	
2006	Afghanistan	Drought	1	1,900,000	-	1,900,000	1,900,000	-	1,900,000		-	
2006	Afghanistan	Mass Movement	1	300,013	-	300,013	300,000	-	300,000	13	-	13
2006	Argentina	Convective Storm	2	260,354	-	260,354	260,350	-	260,350	4	-	4
2006	Argentina	Flood	1	150,000	-	150,000	150,000	-	150,000		-	
2006	Argentina	Storm	2	308,000	-	308,000	308,000	-	308,000		-	
2006	Bangladesh	Flood	1	135,775	-	135,775	135,775	-	135,775		-	
2006	Bolivia (Plurinational State of)	Flood	1	126,121	126,121	-	126,096	126,096		25	25	

YEAR	Country	Event	Entries	Total Human Effect	Total Human Effect [Firm]	Total Human Effect [Soft]	People Affected	People Affected [Firm]	People Affected [Soft]	Deaths	Deaths [Firm]	Deaths [Soft]
2006	Brazil	Flood	1	116,008	-	116,008	116,000	-	116,000	8	-	8
2006	Cambodia	Drought	1	222,678	-	222,678	222,674	-	222,674	4	-	4
2006	Cambodia	Flood	1	222,678	-	222,678	222,674	-	222,674	4	-	4
2006	Chile	Storm	1	157,000	-	157,000	157,000	-	157,000		-	
2006	China	Convective Storm	1	105,000	-	105,000	105,000	-	105,000		-	
2006	China	Drought	1	18,000,134	-	18,000,134	18,000,000	-	18,000,000	134	-	134
2006	China	Earthquake	1	265,128	265,128	-	265,106	265,106		22	22	
2006	China	Flood	8	14,881,233	6,975,090	7,906,143	14,881,075	6,975,040	7,906,035	158	50	108
2006	China	Tropical Cyclone	5	55,227,832	-	55,227,832	55,226,350	-	55,226,350	1,482	-	1,482
2006	Colombia	Flood	1	221,615	-	221,615	221,465	-	221,465	150	-	150
2006	Ecuador	Volcanic Activity	1	300,018	300,018	-	300,013	300,013		5	5	
2006	Ethiopia	Drought	6	4,142,210	-	4,142,210	4,142,210	-	4,142,210		-	
2006	Ethiopia	Flood	3	831,360	-	831,360	831,200	-	831,200	160	-	160
2006	India	Flood	10	7,561,474	-	7,561,474	7,560,697	-	7,560,697	777	-	777
2006	India	Tropical Cyclone	1	150,414	-	150,414	150,300	-	150,300	114	-	114
2006	Indonesia	Earthquake	5	5,091,450	3,183,701	1,907,749	5,080,023	3,177,923	1,902,100	11,427	5,778	5,649
2006	Indonesia	Flood	1	618,722	618,722	-	618,486	618,486		236	236	
2006	Iran (Islamic Republic of)	Earthquake	2	324,962	-	324,962	324,836	-	324,836	126	-	126
2006	Kenya	Flood	1	723,114	-	723,114	723,000	-	723,000	114	-	114
2006	Malaysia	Flood	1	100,006	100,006	-	100,000	100,000		6	6	
2006	Mexico	Drought	4	1,350,000	-	1,350,000	1,350,000	-	1,350,000		-	
2006	Mexico	Extreme Temperat	2	2,783,000	-	2,783,000	2,783,000	-	2,783,000		-	
2006	Mexico	Flood	1	160,002	-	160,002	160,000	-	160,000	2	-	2
2006	Mexico	Mass Movement	1	127,000	-	127,000	127,000	-	127,000		-	
2006	Mexico	Tropical Cyclone	4	784,707	-	784,707	784,700	-	784,700	7	-	7
2006	Mozambique	Disease	1	441,823	-	441,823	441,585	-	441,585	238	-	238
2006	Mozambique	Drought	2	419,180	-	419,180	419,180	-	419,180		-	
2006	Nepal	Drought	1	200,000	-	200,000	200,000	-	200,000		-	
2006	Niger (the)	Drought	4	556,440	-	556,440	556,440	-	556,440		-	
2006	Pakistan	Convective Storm	2	437,837	-	437,837	437,827	-	437,827	10	-	10
2006	Pakistan	Flood	1	100,273	-	100,273	100,264	-	100,264	9	-	9
2006	Paraguay	Disease	1	100,017	100,017	-	100,000	100,000		17	17	
2006	Philippines (the)	Flood	2	672,263	672,263	-	672,259	672,259		4	4	
2006	Philippines (the)	Tropical Cyclone	6	7,693,583	7,493,224	200,359	7,691,868	7,491,513	200,355	1,715	1,711	4
2006	Somalia	Flood	2	454,587	-	454,587	454,500	-	454,500	87	-	87
2006	Sri Lanka	Flood	1	333,027	333,027	-	333,002	333,002		25	25	
2006	Sudan (the)	Flood	1	150,000	-	150,000	150,000	-	150,000		-	
2006	Tanzania, United Republic of	Drought	1	3,700,000	-	3,700,000	3,700,000	-	3,700,000		-	
2006	Thailand	Flood	2	2,555,588	2,212,577	343,011	2,555,308	2,212,413	342,895	280	164	116
2006	Vietnam	Tropical Cyclone	3	3,294,655	-	3,294,655	3,294,285	-	3,294,285	370	-	370

H.5: FTS Reconciliation Mismatch

The first table in this appendix maps out for each year the entries in FTS summary data download that do not reconcile to the FTS detail data download data (FTS, 2017d; FTS, 2017h). The second table details the 57 countries (plus the Not Specified catchall) and values involved in the irreconcilable summary entries.

FTS Annual Summary to FTS Flow Detail Reconciliation

Year	Incoming	Incoming & Outgoing	Incoming + Internal	Internal	Internal-Pooled	Cannot be Reconciled	Total
1999	2	3	1	0	0	1	7
2000	25	22	16	0	0	2	65
2001	40	42	19	0	0	2	103
2002	61	36	19	0	0	2	118
2003	66	36	18	1	0	2	123
2004	70	20	21	0	0	3	114
2005	68	25	28	1	1	5	128
2006	63	50	19	3	3	9	147
2007	59	27	20	3	3	19	131
2008	49	26	29	1	1	16	122
2009	59	27	27	1	1	15	130
2010	52	31	26	1	1	19	130
2011	55	27	29	1	0	19	131
2012	79	8	15	0	1	25	128
2013	64	18	13	0	1	27	123
2014	73	8	13	0	1	29	124
2015	80	8	10	1	0	36	135
Total	965	414	323	13	13	231	1959

Incoming The number of entries in the summary dataset that had totals that matched equivalent year and country subtotals of incoming flows in the detail dataset.

Incoming & Outgoing The number of entries in the summary dataset that had totals that matched equivalent year and country subtotals of outgoing flows as well as the subtotals for incoming flows in the detail dataset.

Incoming *plus* Internal The number of entries in the summary dataset that had totals that matched equivalent year and country subtotals of incoming and internal flows in the detail dataset, if the values are added together. Note: the site discourages adding Internal flows to Incoming flows for risk of double-counting (FTS, 2017f).

Internal The number of entries in the summary dataset that reconciled to internal flows in the detail dataset.

Internal *minus* Pooled The number of entries in the summary dataset that reconciled to internal flows once any such entry with "pooled fund" in its description is deducted.

Cannot be Reconciled The number of entries in the summary dataset for which no calculations that can enable reconciliation can be found.

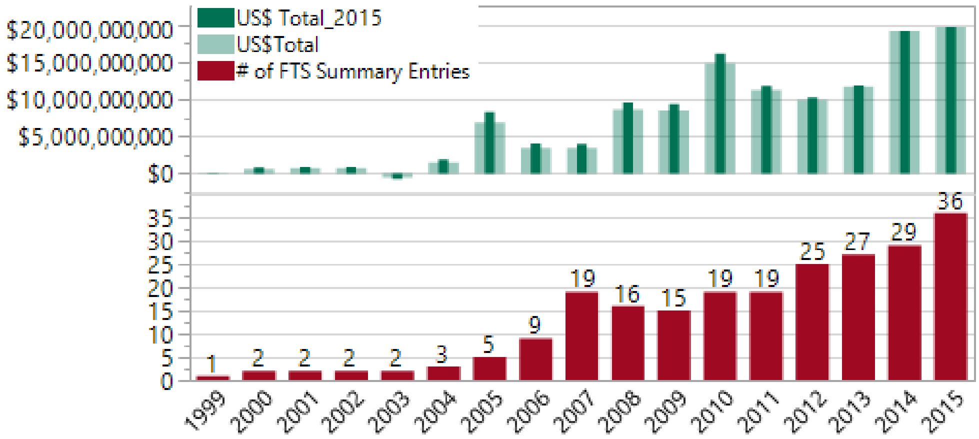
FTS Reconciliation Mismatch

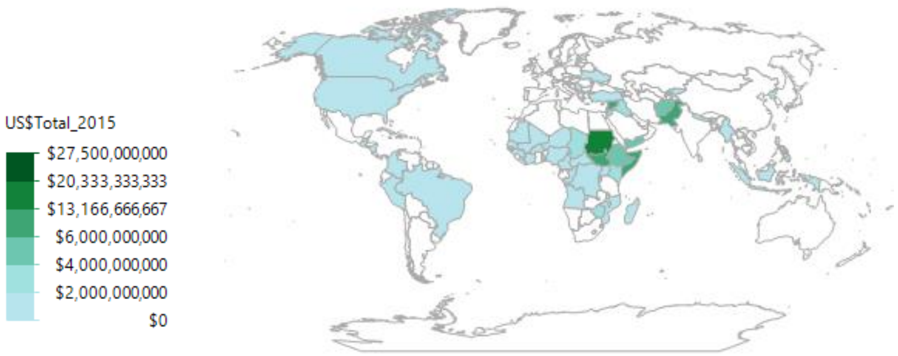
Country and Humanitarian Aid

Recipient Country	US$ Total	US$ Total_2015
Afghanistan	$4,286,277,843	$4,497,467,540
Angola	$170,360,174	$213,754,671
Brazil	$71,115	$125,178
Burkina Faso	$384,913,030	$391,625,005
Burundi	$130,786,920	$149,505,278
Cameroon	$276,055,723	$276,183,178
Canada	-$1,966,955	-$1,966,955
Central African Republic	$1,692,697,500	$1,747,073,889
Chad	$2,136,696,624	$2,288,515,590
Colombia	$392,297,572	$405,277,501
Congo	$13,712,095	$15,674,584
Congo, The Democratic Republic of the	$1,342,724,959	$1,432,481,168
Côte d'Ivoire	$85,581,393	$95,939,190
Djibouti	$24,446,674	$24,475,692
Ethiopia	$4,139,868,377	$4,341,396,119
Georgia	$150,564,477	$165,749,389
Guatemala	$11,398,855	$11,398,855
Guinea	$21,496,761	$25,273,303
Haiti	$5,535,569,344	$5,961,203,605
Indonesia	$56,991,902	$60,201,973
Iraq	$2,585,584,498	$2,703,797,945
Jordan	$3,085,439,955	$3,111,316,591
Kenya	$3,353,135,970	$3,595,586,518
Korea, Democratic People's Republic of	$116,980,384	$133,722,736
Kuwait	-$3,250,000	-$3,336,447
Kyrgyzstan	$138,041,408	$150,044,761
Lebanon	$4,790,337,817	$5,023,259,257
Liberia	$278,186,658	$304,207,280
Macedonia, The Former Yugoslav Republic of	$11,120,896	$11,120,896
Madagascar	$33,819,331	$37,230,120
Mali	$590,062,705	$590,515,755
Mauritania	$255,262,301	$257,296,293
Mozambique	$53,036,730	$60,627,402
Myanmar	$1,445,763,302	$1,521,148,261
Nepal	$1,290,910,585	$1,344,132,735
Nicaragua	$40,689,607	$46,513,145
Niger	$1,542,114,578	$1,572,261,999
Nigeria	$160,533,430	$160,533,430
Not specified	$24,269,692,757	$26,537,657,556

Recipient Country	US$ Total	US$ Total_2015
occupied Palestinian territory	$5,284,796,194	$5,569,792,312
Pakistan	$6,349,645,510	$6,799,447,260
Peru	$63,991,763	$73,150,330
Philippines	$1,475,671,787	$1,479,917,625
Senegal	$32,122,689	$32,160,818
Serbia	$22,304,411	$22,304,411
Somalia	$7,328,436,631	$7,759,247,482
South Sudan	$6,122,414,425	$6,194,617,259
Sri Lanka	$774,649,553	$856,914,160
Sudan	$11,806,876,129	$13,247,093,191
Syrian Arab Republic	$6,831,238,479	$6,881,012,428
Timor-Leste	$31,447,745	$35,948,579
Turkey	$969,515,535	$970,078,556
Uganda	$1,778,643,683	$1,983,670,599
Ukraine	$285,127,647	$285,127,647
United States	-$6,501,978	-$6,497,103
Vanuatu	$42,255,968	$42,255,968
Yemen	$3,943,561,553	$4,011,315,119
Zimbabwe	$2,038,650,905	$2,203,605,267
All	**$120,062,855,924**	**$127,700,152,894**

FTS 231 entries Not Reconciled





H.6: IDS (OECD) Aid Types

There are fifty-three aid types available in the four downloaded tables from OECD's International Development Statistics (IDS) (OECD, 2017b; IDS, 2017). Of these only five aid types are considered pertinent to his research. These five aid type are highlighted in the table below [*Figure H.6-1*].

Table	Aid	Type
2a	**ODA: Total Net**	**206**
2a	**Humanitarian Aid**	**216**
2a	AF/Interest Subsidies	208
2a	Capital Subscriptions - Deposits	210
2a	Development Food Aid	213
2a	Equity investment	217
2a	Grants, Total	201
2a	Grants: Debt Forgiveness	212
2a	Grants: Other Debt Grants	221
2a	Imputed Multilateral ODA	106
2a	Interest received	209
2a	Memo: Net debt relief	255
2a	Memo: ODA Total, excl. Debt	250
2a	Memo: ODA Total, Gross disbursements	240
2a	Memo:Capital Subscriptions - Encashments	211
2a	ODA as % GNI (Recipient)	286
2a	ODA Gross Loans	204
2a	ODA Loan Repayments	205
2a	ODA Loans: Total Net	218
2a	ODA per Capita	296
2a	Offsetting entries for debt relief	215
2a	Recoveries	219
2a	Rescheduled debt	214
2a	Technical Cooperation	207

Table	Aid	Type
2b	**Total OOF, Net**	**206**
2b	Equity investment	217
2b	Grants, Total	201
2b	Interest Received	207
2b	Memo: Net Export Credits	250
2b	Memo: Net Other Long-term Loans	255
2b	Off. Export Credits - Amounts Extended	202
2b	Off. Export Credits - Amounts Received	203
2b	Offsetting entries for debt relief	215
2b	Other Long Term - Amounts Extended	204
2b	Other Long Term - Amounts Received	205
2b	Total OOF, Gross	972

Table	Aid	Type
3a	**Total Commitments**	**305**
3a	Capital Subscriptions	310
3a	Grants	301
3a	Loans and Other Long Term Capital	304
3a	of which: Associated Financing	308
3a	of which: Technical Cooperation	306

Table	Aid	Type
4	**Total Private Net**	**420**
4	Foreign Direct Investment	405
4	Memo: Other Private Flows	450
4	of which - Bank Net Export Credits	416
4	of which - Non-Bank Gross Export Credits	408
4	of which - Non-Bank Net Export Credits	410
4	of which - Non-Bank Securities & Other	407
4	of which - Non-Banks Exp. Cred. Amort	409
4	Offsetting Entries for Debt Relief	419
4	Other Private - Total Banks	418
4	Other Private - Total Non-Banks	417

Figure H.6-1: OECD IDS Aid Types

H.7: IDS (OECD) US$ Totals Per Aid Type Per Year

Year	Humanitarian Aid [2a] (US$s2015)	NetODA_exHumAid [2a] (US$s2015)	OOF [2b] (US$s2015)	ODA Commitments [3a] (US$s2015)	Private [4] (US$s2015)
1990	$584,130,000	$71,990,960,000	$24,166,160,000	$101,876,840,000	$4,168,000,000
1991	$832,550,000	$79,602,570,000	$15,219,830,000	$96,545,990,000	$46,590,000,000
1992	$877,630,000	$73,490,420,000	$8,666,110,000	$78,266,270,000	$68,539,000,000
1993	$807,670,000	$67,898,760,000	$16,991,880,000	$79,520,310,000	$114,091,000,000
1994	$1,159,170,000	$73,226,830,000	$7,310,890,000	$82,401,710,000	$135,322,000,000
1995	$3,500,270,000	$61,944,500,000	$8,960,050,000	$74,847,410,000	$130,084,000,000
1996	$2,784,390,000	$61,233,690,000	$10,637,110,000	$77,486,380,000	$195,463,000,000
1997	$2,935,160,000	$56,906,630,000	$29,082,990,000	$72,259,760,000	$246,646,000,000
1998	$2,665,310,000	$61,977,180,000	$34,116,450,000	$82,478,170,000	$234,134,000,000
1999	$4,139,600,000	$60,930,930,000	$27,059,970,000	$82,845,690,000	$265,459,000,000
2000	$3,445,250,000	$60,127,820,000	$17,315,700,000	$85,711,540,000	$160,569,000,000
2001	$4,171,780,000	$67,037,410,000	$16,377,990,000	$88,723,650,000	$137,686,000,000
2002	$4,698,030,000	$73,208,490,000	-$6,348,490,000	$97,950,280,000	$53,451,000,000
2003	$6,092,310,000	$71,824,160,000	-$19,239,800,000	$110,237,630,000	$115,537,000,000
2004	$6,197,520,000	$74,986,880,000	-$12,831,770,000	$111,423,160,000	$214,667,000,000
2005	$7,542,760,000	$92,102,230,000	-$1,920,350,000	$118,121,800,000	$192,561,000,000
2006	$6,813,610,000	$85,495,800,000	-$5,358,760,000	$115,377,600,000	$205,798,000,000
2007	$6,258,660,000	$77,422,670,000	$14,203,070,000	$112,183,750,000	$329,813,000,000
2008	$8,451,440,000	$80,835,470,000	$24,059,270,000	$128,332,090,000	$182,858,000,000
2009	$8,757,790,000	$85,323,060,000	$46,702,290,000	$130,104,080,000	$137,036,000,000
2010	$9,031,440,000	$83,514,650,000	$52,869,760,000	$125,867,430,000	$211,322,000,000
2011	$10,485,560,000	$82,456,230,000	$16,000,200,000	$122,220,710,000	$228,077,000,000
2012	$9,224,140,000	$82,367,660,000	$26,003,000,000	$130,266,170,000	$192,469,000,000
2013	$11,887,390,000	$89,747,720,000	$22,358,790,000	$146,344,400,000	$213,283,000,000
2014	$13,962,280,000	$87,720,200,000	$23,060,980,000	$139,237,080,000	$249,724,000,000
2015	$16,823,290,000	$91,137,840,000	$36,590,720,000	$156,020,380,000	$180,789,000,000
TOTALS	**$154,129,130,000**	**$1,954,510,760,000**	**$432,054,040,000**	**$2,746,650,280,000**	**$4,446,136,000,000**

H.8: World Bank Population Data Not Found

MSGD: Disaster Affected Country	World Bank Population Data Not Found for:
Andorra	2014, 2015
Anguilla	1999
Austria	1993
Bahamas (the)	1995
Barbados	1990, 1991, 1993
Belize	1990
Brunei Darussalam	19,982,015
Cayman Islands (the)	2001
Cook Islands (the)	1990, 1991,1992, 1993, 1995, 1997, 2001, 2005, 2009, 2010
Curaçao	2015
Cyprus	1991, 1992, 2014, 2015
Eritrea	2012, 2013, 2014, 2015
Finland	1990
French Guiana	1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2009
Gibraltar	2014, 2015
Guadeloupe	1991, 1995, 1998, 1999, 2001, 2004, 2007, 2010, 2014, 2015
Holy See (the)	2004, 2007, 2008
Iceland	2015
Kuwait	1992, 1993,, 1994
Luxembourg	1990, 1993, 1995, 2015
Macedonia (the former Yugoslav Republic of)	1990
Malta	2014, 2015
Marshall Islands (the)	1997, 2001
Martinique	1990, 1992, 1993, 1995, 1996, 1999, 2007, 2010, 2011
Mayotte	2008, 2012
Micronesia (Federated States of)	1991, 1992
Montserrat	1995, 1996, 1997, 2006
Netherlands Antilles	1995, 2001
New Caledonia	1992
New Zealand	1992
Niue	1990, 1996, 1998, 1999, 2003, 2004, 2011, 2012, 2013, 2014, 2015
Norway	1993, 2015
Palau	2014, 2015
Réunion	1991, 1993, 2000, 2002, 2005, 2007, 2014
Saint Helena, Ascension and Tristan da Cunha	2001
San Marino	2015
Solomon Islands	1994
Swaziland	1991
Tokelau	1990, 2005, 2009
Tonga	1995
Tuvalu	1992, 2014
Virgin Islands (British)	1995
Wallis and Futuna	1992, 1993, 2012
Western Sahara*	1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015

H.9: MSGD to Aid Anomalies

MSGD/FTS Mismatch – Missing Humanitarian Aid

YEAR	Disasters				Disaster Losses			
	Naturogenic	Anthropogenic	Deviant	ALL	Deaths	Affected	Total Human Effect	US$Loss_2015 (000s)
1999	6,438	1,091	6,165	13,694	202,697	324,018,480	324,221,177	$139,488,810
2000	1,377	479	2,617	4,473	16,547	10,358,114	10,374,661	$2,387,761,894
2001	663	207	960	1,830	11,759	6,527,301	6,539,060	$18,484,857
2002	198	167	977	1,342	3,060	1,001,749	1,004,809	$33,434,051
2003	147	101	1,155	1,403	60,683	1,219,387	1,280,070	$48,667,870
2004	916	180	1,195	2,291	4,131	12,639,931	12,644,062	$76,543,928
2005	204	112	746	1,062	3,469	3,435,622	3,439,091	$204,127,112
2006	109	30	704	843	4,733	966,694	971,427	$14,323,523
2007	117	195	927	1,239	1,950	2,084,144	2,086,094	$34,501,654
2008	110	93	1,039	1,242	1,302	14,617,258	14,618,560	$70,904,986
2009	119	210	843	1,172	1,872	916,580	918,452	$25,518,991
2010	296	272	993	1,561	1,204	2,977,433	2,978,637	$50,936,820
2011	272	243	1,062	1,577	1,347	977,125	978,472	$4,602,912
2012	822	116	1,150	2,088	3,996	2,605,969	2,609,965	$10,207,306
2013	168	149	1,296	1,613	2,619	4,612,388	4,615,007	$34,602,574
2014	113	43	1,377	1,533	1,145	4,212,050	4,213,195	$27,990,324
2015	108	24	1,499	1,631	9,121	7,023,915	7,033,036	$10,814,406
All	12,177	3,712	24,705	40,594	331,635	400,194,140	400,525,775	$3,192,912,018

MSGD/IDS Mismatch – Zero or –ve Humanitarian Aid

YEAR	Disasters				Disaster Losses			
	Naturogenic	Anthropogenic	Deviant	ALL	Deaths	Affected	Total Human Effect	US$Loss_2015 (000s)
1998			8	8		97	97	
2006		1	26	27		1,243	1,243	
2008	3		12	15		865	865	
2011	2		6	8		480	480	$4,419
2012			3	3		33	33	
All	5	1	55	61		2,718	2,718	$4,419

MSGD/IDS Mismatch – Missing Humanitarian Aid

YEAR	Disasters				Disaster Losses			
	Naturogenic	Anthropogenic	Deviant	ALL	Deaths	Affected	Total Human Effect	US$Loss_2015 (000s)
1990	191	104	1,227	1,522	17,636	10,481,111	10,498,747	$66,753,325
1991	152	127	1,496	1,775	12,108	5,873,095	5,885,203	$48,054,580
1992	126	70	1,549	1,745	6,355	16,233,419	16,239,774	$84,256,319
1993	170	94	489	753	4,649	10,633,659	10,638,308	$46,250,067
1994	125	90	633	848	4,566	7,448,230	7,452,796	$72,417,879
1995	123	107	273	503	7,663	6,159,599	6,167,262	$196,907,451
1996	251	264	265	780	2,278	3,465,820	3,468,098	$19,995,539
1997	195	87	273	555	1,716	3,158,206	3,159,922	$26,679,737
1998	111	51	273	435	2,107	2,795,538	2,797,645	$42,101,512
1999	146	94	282	522	2,101	9,392,599	9,394,700	$76,820,556
2000	223	142	542	907	1,881	3,750,704	3,752,585	$45,353,390
2001	135	43	500	678	7,166	2,998,273	3,005,439	$16,607,771
2002	170	45	515	730	1,814	2,109,267	2,111,081	$63,343,948
2003	139	68	555	762	72,395	1,773,030	1,845,425	$59,849,063
2004	114	82	500	696	1,276	6,505,037	6,506,313	$136,415,942
2005	148	73	1,341	1,562	4,827	3,313,987	3,318,814	$215,385,668
2006	194	167	1,437	1,798	6,047	2,224,658	2,230,705	$14,554,730
2007	217	149	1,376	1,742	2,946	2,903,376	2,906,322	$53,130,244
2008	103	70	1,376	1,549	2,538	15,272,408	15,274,946	$71,466,451
2009	81	31	1,504	1,616	4,204	1,899,092	1,903,296	$31,080,054
2010	398	343	1,420	2,161	58,430	3,515,182	3,573,612	$72,762,998
2011	119	54	1,426	1,599	22,764	3,235,571	3,258,335	$313,399,013
2012	254	159	1,502	1,915	3,059	2,317,097	2,320,156	$132,061,353
2013	108	65	1,433	1,606	3,501	5,935,918	5,939,419	$53,528,709
2014	102	36	1,812	1,950	3,837	4,829,933	4,833,770	$29,036,667
2015	75	41	1,543	1,659	8,570	5,546,990	5,555,560	$31,004,272
All	4,170	2,656	25,542	32,368	266,434	143,771,799	144,038,233	$2,019,217,241

MSGD/IDS Mismatch – No Other Aid

YEAR	Disasters				Disaster Losses			
	Naturogenic	Anthropogenic	Deviant	ALL	Deaths	Affected	Total Human Effect	US$Loss_2015 (000s)
1990	128	42	963	1,133	4,278	9,332,456	9,336,734	$65,210,309
1991	77	47	1,238	1,362	6,681	3,722,977	3,729,658	$40,335,511
1992	82	22	1,241	1,345	3,050	11,230,423	11,233,473	$82,567,040
1993	91	18	149	258	1,781	6,189,038	6,190,819	$45,120,339
1994	70	35	297	402	2,046	6,387,595	6,389,641	$71,686,272
1995	105	37	236	378	7,603	6,157,475	6,165,078	$196,870,988
1996	102	37	224	363	1,693	3,267,760	3,269,453	$19,926,642
1997	169	27	224	420	1,367	2,535,824	2,537,191	$26,055,406
1998	90	34	237	361	2,024	2,487,878	2,489,902	$42,052,436
1999	92	33	221	346	2,005	9,295,261	9,297,266	$72,742,307
2000	182	93	395	670	1,605	3,538,653	3,540,258	$45,344,731
2001	115	35	372	522	7,097	2,784,823	2,791,920	$16,046,730
2002	140	41	379	560	1,447	2,049,632	2,051,079	$63,284,499
2003	115	38	413	566	72,275	1,619,872	1,692,147	$59,407,502
2004	103	49	399	551	1,174	6,444,223	6,445,397	$132,103,777
2005	147	65	1,318	1,530	4,819	3,310,830	3,315,649	$215,385,668
2006	190	160	1,402	1,752	6,040	2,213,201	2,219,241	$14,554,730
2007	216	145	1,358	1,719	2,942	2,901,575	2,904,517	$53,130,244
2008	80	27	1,297	1,404	2,427	14,993,157	14,995,584	$71,466,227
2009	79	31	1,442	1,552	4,038	1,818,235	1,822,273	$30,085,749
2010	355	333	1,377	2,065	58,406	3,166,374	3,224,780	$71,662,347
2011	88	50	1,379	1,517	22,763	2,800,466	2,823,229	$313,399,013
2012	191	136	1,448	1,775	3,057	1,987,685	1,990,742	$132,061,250
2013	108	65	1,433	1,606	3,501	5,935,918	5,939,419	$53,528,709
2014	101	35	1,742	1,878	3,837	4,829,065	4,832,902	$29,036,667
2015	75	41	1,543	1,659	8,570	5,546,990	5,555,560	$31,004,272
All	3,291	1,676	22,727	27,694	236,526	126,547,386	126,783,912	$1,994,069,364

MSGD – FTS + IDS Humanitarian Aid

YEAR	FTS + IDS Humanitarian Aid
1999	$1,615,574,566
2000	$4,781,307,585
2001	$8,228,556,639
2002	$10,337,539,379
2003	$15,180,728,783
2004	$11,374,597,043
2005	$18,494,842,801
2006	$15,306,220,358
2007	$14,260,708,180
2008	$20,288,828,887
2009	$20,376,176,844
2010	$26,411,644,656
2011	$22,157,533,812
2012	$20,268,794,583
2013	$24,651,226,353
2014	$34,481,369,697
2015	$35,430,260,274

MSGD – Negative Humanitarian Aid

YEAR	Country	Disasters				Disaster Losses			Humanitarian Aid	
		Naturogenic	Anthropogenic	Deviant	ALL	Deaths	Affected	Total Human Effect	FTS	IDS
1998	Bahrain			8	8		97	97		-$60,000
1999	Various/Unknown			65	65		1,417,864	1,417,864	-$38,613,755	
2003	Various/Unknown			87	87		573,169	573,169	-$1,152,793,379	
2012	Kuwait			43	43		94,337	94,337	-$2,064,662	
2013	Kuwait			42	42		94,167	94,167	-$1,271,785	
2015	Canada	1		23	24		13,463	13,463	-$1,966,955	
	United States of America (the)	28	3	67	98	379	54,804	55,183	-$6,653,389	
	Total	29	3	335	367	379	2,247,901	2,248,280	-$1,203,363,925	-$60,000

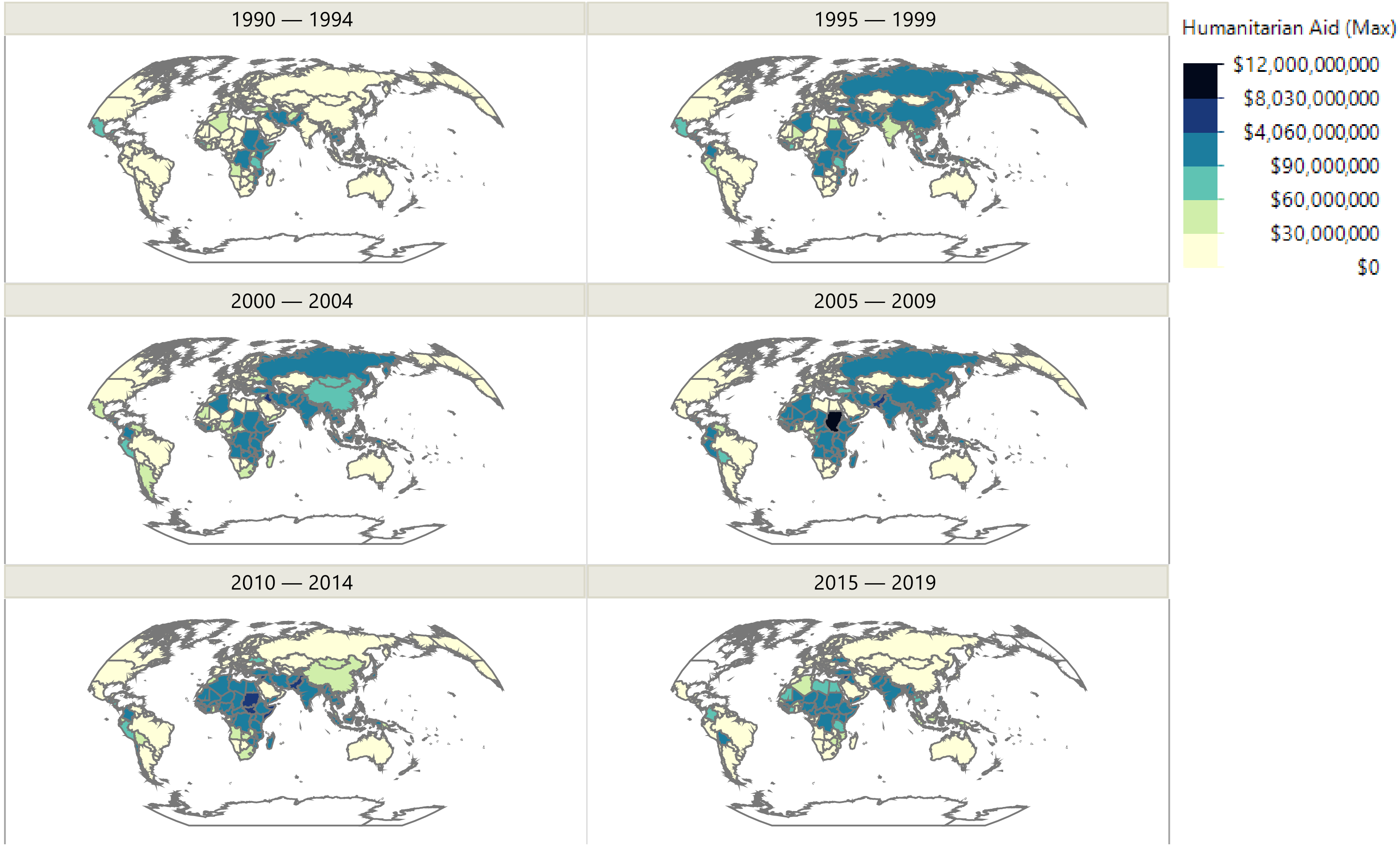
MSGD – US$250K or more Humanitarian Aid per Person

YEAR	Country	Disasters	Disaster Losses			Humanitarian Aid
		ALL	Deaths	Affected	Total Human Effect	Avg per Person
1990	Côte d'Ivoire	3	8	2	10	$826,000
	Cyprus	8		11	11	$1,566,364
	Sierra Leone	1		8	8	$428,750
1991	Belize	1	2		2	$450,000
	Botswana	1		3	3	$280,000
	Guinea	2		2	2	$10,495,000
1992	Belize	1	1		1	$2,450,000
	Botswana	1		3	3	$296,667
	Côte d'Ivoire	5	1	34	35	$313,429
	Malaysia	1		3	3	$2,356,667
1993	Malawi	2		43	43	$849,767
	Zimbabwe	5		17	17	$492,941
1994	Belize	1		8	8	$415,000
	Malawi	3		60	60	$393,167
	Zambia	1		15	15	$406,667
	Zimbabwe	7		42	42	$306,667
1996	Belize	1		3	3	$560,000
	Cabo Verde	2		2	2	$575,000
	Namibia	3	2	11	13	$393,846
1997	Cabo Verde	2		2	2	$315,000
	Namibia	1		2	2	$595,000
2001	Botswana	3		8	8	$280,000
2002	Maldives	3		4	4	$595,000
2005	Lesotho	6		13	13	$488,489
	Maldives	16	1	44	45	$3,047,924
2009	Maldives	3		15	15	$486,667
2010	Vanuatu	3		3	3	$1,000,000
2012	Marshall Islands (the)	1		2	2	$370,000
2014	Jordan	89	1	3,155	3,156	$291,098
2015	Samoa	2		13	13	$453,077
	TOTAL	178	1	3,502	3,544	

MSGD – Less than US$1 Humanitarian Aid per Person

YEAR	Disasters	Disaster Losses			Humanitarian Aid
	ALL	Deaths	Affected	Total Human Effect	Avg per Person
1990	1,257	15,920	68,157,447	68,173,367	$0.27
1991	1,833	181,456	263,188,381	263,369,837	$0.21
1992	1,628	23,518	37,906,336	37,929,854	$0.41
1993	1,779	26,484	147,692,708	147,719,192	$0.47
1994	1,655	20,291	252,963,869	252,984,160	$0.55
1995	743	619,359	194,428,833	195,048,192	$0.16
1996	642	14,514	192,996,399	193,010,913	$0.23
1997	874	10,645	50,585,540	50,596,185	$0.50
1998	878	23,173	312,147,048	312,170,221	$0.55
1999	1,358	39,171	195,238,366	195,277,537	$0.38
2000	2,340	19,385	135,567,305	135,586,690	$0.49
2001	564	2,905	51,857,612	51,860,517	$0.54
2002	1,802	11,809	631,172,447	631,184,256	$0.47
2003	174	2,786	219,818,348	219,821,134	$0.15
2004	1,720	28,050	103,404,782	103,432,832	$0.30
2005	601	4,703	97,265,360	97,270,063	$0.44
2006	858	5,042	97,057,166	97,062,208	$0.24
2007	2,325	8,598	177,669,190	177,677,788	$0.51
2008	936	1,739	10,859,608	10,861,347	$0.46
2009	683	3,856	184,042,027	184,045,883	$0.31
2010	1,015	13,212	193,032,625	193,045,837	$0.09
2011	951	6,069	130,944,424	130,950,493	$0.38
2012	305	1,261	50,366,011	50,367,272	$0.39
2013	968	4,214	36,804,831	36,809,045	$0.33
2014	244	2,466	92,855,183	92,857,649	$0.06
2015	336	4,819	4,268,673	4,273,492	$0.41
All	28,469	1,095,445	3,932,290,519	3,933,385,964	$0.38

H.10: Humanitarian Aid (Max)

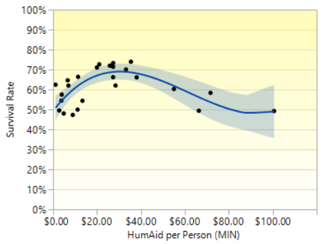


Appendix I: M^I & M^E Supplemental

I.1: M^II: Survival Rate x Humanitarian Aid

I.1.1: Humanitarian Aid – MINIMUM of FTS or IDS

Survival Rate (Undifferentiated)



Survival Rate = 0.6481113 + 0.0014929*HumAid per Person (MIN) - 0.000183*(HumAid per Person (MIN)-26.0946)^2 + 1.7115e-6*(HumAid per Person (MIN)-26.0946)^3

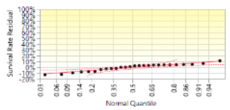
Summary of Fit

RSquare	0.549749
RSquare Adj	0.488351
Root Mean Square Error	0.064987
Mean of Response	0.61737
Observations (or Sum Wgts)	26

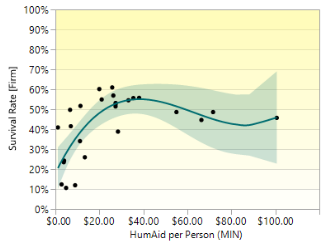
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.11344494	0.037815	8.9539
Error	22	0.09291292	0.004223	Prob > F
C. Total	25	0.20635786		0.0005*

Residual Normal Quantile Plot



Survival Rate (Firm)



Survival Rate [Firm] = 0.3907048 + 0.0050378*HumAid per Person (MIN) - 0.0002487*(HumAid per Person (MIN)-26.0946)^2 + 2.278e-6*(HumAid per Person (MIN)-26.0946)^3

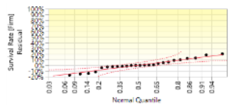
Summary of Fit

RSquare	0.532788
RSquare Adj	0.469077
Root Mean Square Error	0.112292
Mean of Response	0.426508
Observations (or Sum Wgts)	26

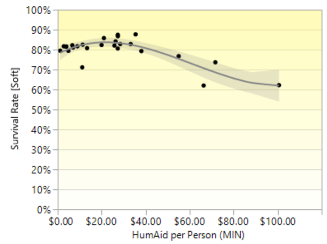
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.31634440	0.105448	8.3626
Error	22	0.27740863	0.012609	Prob > F
C. Total	25	0.59375303		0.0007*

Residual Normal Quantile Plot



Survival Rate (Soft)



Survival Rate [Soft] = 0.8574558 - 0.0008509*HumAid per Person (MIN) - 9.3359e-5*(HumAid per Person (MIN)-26.0946)^2 + 8.8542e-7*(HumAid per Person (MIN)-26.0946)^3

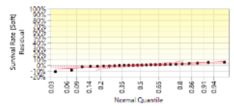
Summary of Fit

RSquare	0.678584
RSquare Adj	0.634755
Root Mean Square Error	0.038791
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.06989045	0.023297	15.4824
Error	22	0.03310403	0.001505	Prob > F
C. Total	25	0.10299448		<.0001*

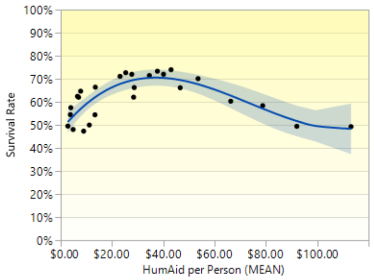
Residual Normal Quantile Plot



*** Polynomial Line of Fit (Degree=3)*

I.1.2: Humanitarian Aid – MEAN of FTS and IDS

Survival Rate (Undifferentiated)



Survival Rate = 0.65535 + 0.001459*HumAid per Person (MEAN) - 0.000143*(HumAid per Person (MEAN)-31.6291)^2 + 1.1323e-6*(HumAid per Person (MEAN)-31.6291)^3

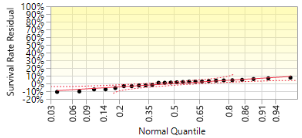
Summary of Fit

RSquare	0.68153
RSquare Adj	0.638103
Root Mean Square Error	0.054655
Mean of Response	0.61737
Observations (or Sum Wgts)	26

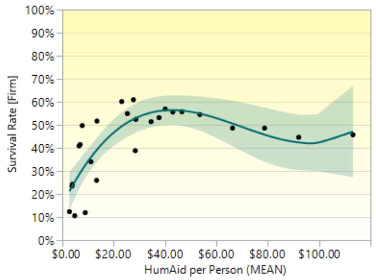
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.14063913	0.046880	15.6935
Error	22	0.06571873	0.002987	Prob > F
C. Total	25	0.20635786		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Survival Rate [Firm] = 0.4192717 + 0.0039415*HumAid per Person (MEAN) - 0.0002087*(HumAid per Person (MEAN)-31.6291)^2 + 1.8329e-6*(HumAid per Person (MEAN)-31.6291)^3

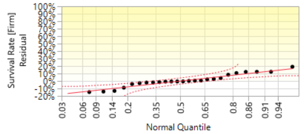
Summary of Fit

RSquare	0.635349
RSquare Adj	0.585624
Root Mean Square Error	0.099204
Mean of Response	0.426508
Observations (or Sum Wgts)	26

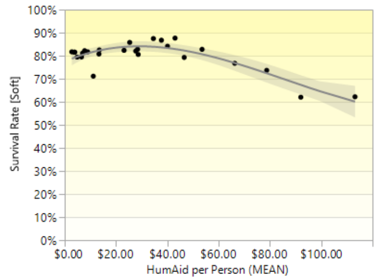
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.37724049	0.125747	12.7772
Error	22	0.21651254	0.009841	Prob > F
C. Total	25	0.59375303		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Survival Rate [Soft] = 0.8512372 - 0.0003298*HumAid per Person (MEAN) - 6.3758e-5*(HumAid per Person (MEAN)-31.6291)^2 + 3.8979e-7*(HumAid per Person (MEAN)-31.6291)^3

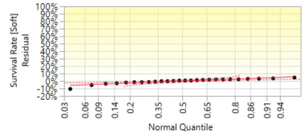
Summary of Fit

RSquare	0.752342
RSquare Adj	0.718571
Root Mean Square Error	0.03405
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.07748711	0.025829	22.2774
Error	22	0.02550737	0.001159	Prob > F
C. Total	25	0.10299448		<.0001*

Residual Normal Quantile Plot

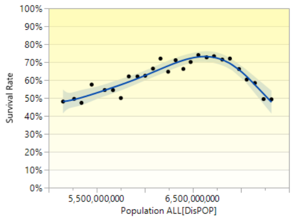


*** Polynomial Line of Fit (Degree=3)*

I.2: M^{i}E: Survival Rate x Population

I.2.1: DisPOP: Population of Disaster-Affected Countries

Survival Rate (Undifferentiated)



Survival Rate =-0.6467 + 2.124e-10*Population ALL[DisPOP_ID] - 1.076e-19*(Population ALL[DisPOP_ID]-6.27e+9)^2 - 3.054e-28*(Population ALL[DisPOP_ID]-6.27e+9)^3 - 1.767e-37*(Population ALL[DisPOP_ID]-6.27e+9)^4 + 9.801e-47*(Population ALL[DisPOP_ID]-6.27e+9)^5 + 9.647e-56*(Population ALL[DisPOP_ID]-6.27e+9)^6

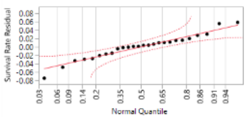
Summary of Fit

RSquare	0.898716
RSquare Adj	0.866732
Root Mean Square Error	0.033167
Mean of Response	0.61737
Observations (or Sum Wgts)	26

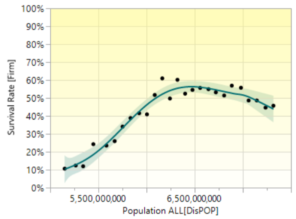
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.18545719	0.030910	28.0987
Error	19	0.02090066	0.001100	Prob > F
C. Total	25	0.20635786		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Survival Rate [Firm] =-0.579266 + 1.796e-10*Population ALL[DisPOP_ID] - 5.174e-19*(Population ALL[DisPOP_ID]-6.27e+9)^2 + 9.347e-29*(Population ALL[DisPOP_ID]-6.27e+9)^3 + 4.393e-37*(Population ALL[DisPOP_ID]-6.27e+9)^4 - 1.071e-46*(Population ALL[DisPOP_ID]-6.27e+9)^5 - 1.734e-55*(Population ALL[DisPOP_ID]-6.27e+9)^6

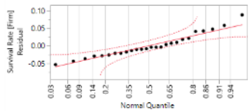
Summary of Fit

RSquare	0.952648
RSquare Adj	0.937695
Root Mean Square Error	0.038467
Mean of Response	0.426508
Observations (or Sum Wgts)	26

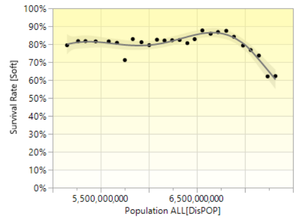
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.56563792	0.094273	63.7090
Error	19	0.02811512	0.001480	Prob > F
C. Total	25	0.59375303		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Survival Rate [Soft] = -0.071281 + 1.425e-10*Population ALL[DisPOP_ID] + 8.365e-20*(Population ALL[DisPOP_ID]-6.27e+9)^2 - 3.466e-28*(Population ALL[DisPOP_ID]-6.27e+9)^3 - 2.764e-37*(Population ALL[DisPOP_ID]-6.27e+9)^4 + 1.168e-46*(Population ALL[DisPOP_ID]-6.27e+9)^5 + 9.438e-56*(Population ALL[DisPOP_ID]-6.27e+9)^6

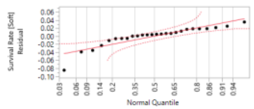
Summary of Fit

RSquare	0.856633
RSquare Adj	0.811359
Root Mean Square Error	0.027878
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.08822848	0.014705	18.9212
Error	19	0.01476600	0.000777	Prob > F
C. Total	25	0.10299448		<.0001*

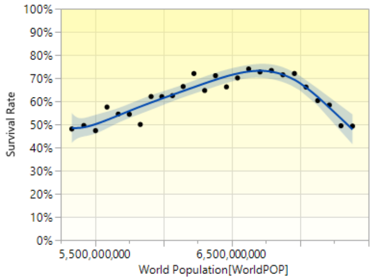
Residual Normal Quantile Plot



****** Polynomial Line of Fit (Degree=6)***

I.2.2: World Populations (Worldometers, 2017)

Survival Rate (Undifferentiated)



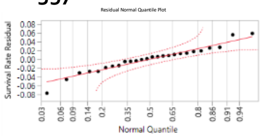
Survival Rate =-0.574266 + 1.986e-10*World Population[WorldPOP] - 8.225e-20*(World Population[WorldPOP]-6.35e+9)^2 - 2.588e-28*(World Population[WorldPOP]-6.35e+9)^3 - 3.105e-37*(World Population[WorldPOP]-6.35e+9)^4 + 6.474e-47*(World Population[WorldPOP]-6.35e+9)^5 + 1.941e-55*(World Population[WorldPOP]-6.35e+9)^6

Summary of Fit

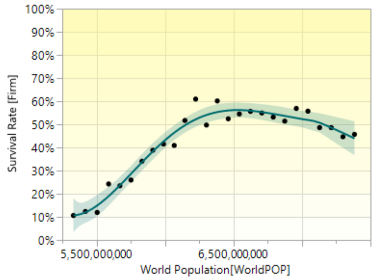
RSquare	0.899495
RSquare Adj	0.867757
Root Mean Square Error	0.033039
Mean of Response	0.61737
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.18561785	0.030936	28.3409
Error	19	0.02074000	0.001092	Prob > F
C. Total	25	0.20635786		<.0001*



Survival Rate (Firm)



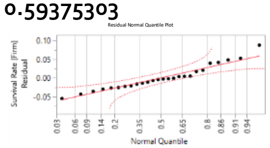
Survival Rate [Firm] = -0.40271 + 1.499e-10*World Population[WorldPOP] - 4.941e-19*(World Population[WorldPOP]-6.35e+9)^2 + 2.111e-28*(World Population[WorldPOP]-6.35e+9)^3 + 3.058e-37*(World Population[WorldPOP]-6.35e+9)^4 - 1.883e-46*(World Population[WorldPOP]-6.35e+9)^5 - 8.014e-56*(World Population[WorldPOP]-6.35e+9)^6

Summary of Fit

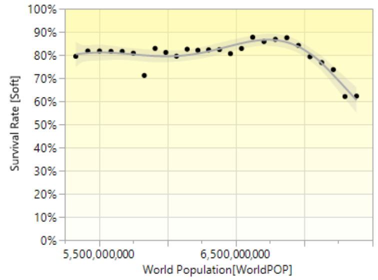
RSquare	0.954544
RSquare Adj	0.94019
Root Mean Square Error	0.03769
Mean of Response	0.426508
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.56676345	0.094461	66.4979
Error	19	0.02698958	0.001421	Prob > F
C. Total	25	0.59375303		<.0001*



Survival Rate (Soft)



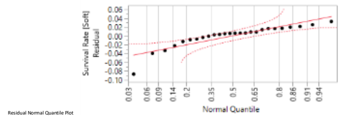
Survival Rate [Soft] = -0.091356 + 1.439e-10*World Population[WorldPOP] + 9.776e-20*(World Population[WorldPOP]-6.35e+9)^2 - 3.668e-28*(World Population[WorldPOP]-6.35e+9)^3 - 3.388e-37*(World Population[WorldPOP]-6.35e+9)^4 + 1.332e-46*(World Population[WorldPOP]-6.35e+9)^5 + 1.331e-55*(World Population[WorldPOP]-6.35e+9)^6

Summary of Fit

RSquare	0.85373
RSquare Adj	0.807539
Root Mean Square Error	0.028158
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

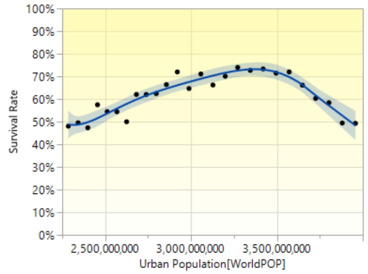
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.08792945	0.014655	18.4828
Error	19	0.01506502	0.000793	Prob > F
C. Total	25	0.10299448		<.0001*



****** Polynomial Line of Fit (Degree=6)***

I.2.3: Urban Populations (Worldometers, 2017)

Survival Rate (Undifferentiated)



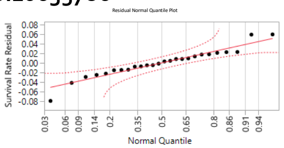
Survival Rate = 0.0785709 + 2e-10*Urban Population[WorldPOP] - 7.966e-20*(Urban Population[WorldPOP]-3.06e+9)^2 - 1.919e-28*(Urban Population[WorldPOP]-3.06e+9)^3 - 9.296e-37*(Urban Population[WorldPOP]-3.06e+9)^4 - 1.381e-46*(Urban Population[WorldPOP]-3.06e+9)^5 + 9.529e-55*(Urban Population[WorldPOP]-3.06e+9)^6

Summary of Fit

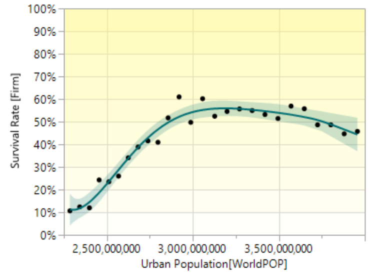
RSquare	0.902091
RSquare Adj	0.871172
Root Mean Square Error	0.03261
Mean of Response	0.61737
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.18615351	0.031026	29.1762
Error	19	0.02020435	0.001063	Prob > F
C. Total	25	0.20635786		<.0001*



Survival Rate (Firm)



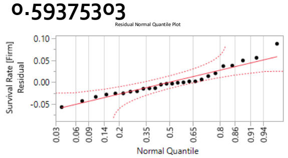
Survival Rate [Firm] = 0.206096 + 1.132e-10*Urban Population[WorldPOP] - 6.029e-19*(Urban Population[WorldPOP]-3.06e+9)^2 + 7.747e-28*(Urban Population[WorldPOP]-3.06e+9)^3 - 4.281e-38*(Urban Population[WorldPOP]-3.06e+9)^4 - 9.44e-46*(Urban Population[WorldPOP]-3.06e+9)^5 + 5.561e-55*(Urban Population[WorldPOP]-3.06e+9)^6

Summary of Fit

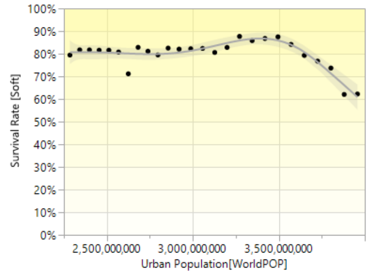
RSquare	0.955164
RSquare Adj	0.941005
Root Mean Square Error	0.037432
Mean of Response	0.426508
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.56713140	0.094522	67.4608
Error	19	0.02662163	0.001401	Prob > F
C. Total	25	0.59375303		<.0001*



Survival Rate (Soft)



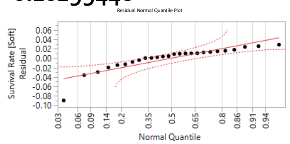
Survival Rate [Soft] = 0.326438 + 1.626e-10*Urban Population[WorldPOP] + 1.607e-19*(Urban Population[WorldPOP]-3.06e+9)^2 - 5.644e-28*(Urban Population[WorldPOP]-3.06e+9)^3 - 8.075e-37*(Urban Population[WorldPOP]-3.06e+9)^4 + 2.713e-46*(Urban Population[WorldPOP]-3.06e+9)^5 + 5.376e-55*(Urban Population[WorldPOP]-3.06e+9)^6

Summary of Fit

RSquare	0.856585
RSquare Adj	0.811296
Root Mean Square Error	0.027882
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

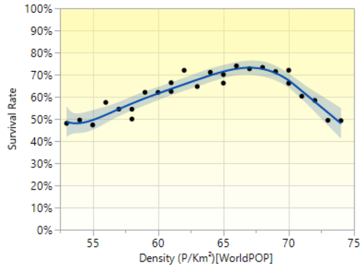
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.08822352	0.014704	18.9138
Error	19	0.01477095	0.000777	Prob > F
C. Total	25	0.10299448		<.0001*



****** Polynomial Line of Fit (Degree=6)***

I.2.4: Population Density

Survival Rate (Undifferentiated)



Survival Rate = -0.566266 + 0.0197433*Density (P/Km^2)[WorldPOP] - 0.0007182*(Density (P/Km^2)[WorldPOP]-63.5)^2 - 0.0002623*(Density (P/Km^2)[WorldPOP]-63.5)^3 - 0.0000368*(Density (P/Km^2)[WorldPOP]-63.5)^4 + 7.2391e-7*(Density (P/Km^2)[WorldPOP]-63.5)^5 + 2.4106e-7*(Density (P/Km^2)[WorldPOP]-63.5)^6

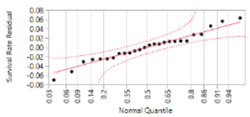
Summary of Fit

RSquare	0.887421
RSquare Adj	0.85187
Root Mean Square Error	0.034967
Mean of Response	0.61737
Observations (or Sum Wgts)	26

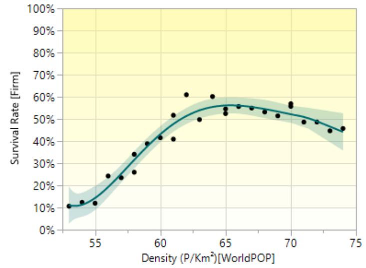
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.18312636	0.030521	24.9618
Error	19	0.02323150	0.001223	Prob > F
C. Total	25	0.20635786		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Survival Rate [Firm] = -0.429907 + 0.0154014*Density (P/Km^2)[WorldPOP] - 0.0047958*(Density (P/Km^2)[WorldPOP]-63.5)^2 + 0.0002011*(Density (P/Km^2)[WorldPOP]-63.5)^3 + 2.461e-5*(Density (P/Km^2)[WorldPOP]-63.5)^4 - 1.7957e-6*(Density (P/Km^2)[WorldPOP]-63.5)^5 - 3.1373e-8*(Density (P/Km^2)[WorldPOP]-63.5)^6

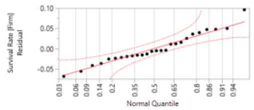
Summary of Fit

RSquare	0.942276
RSquare Adj	0.924048
Root Mean Square Error	0.042472
Mean of Response	0.426508
Observations (or Sum Wgts)	26

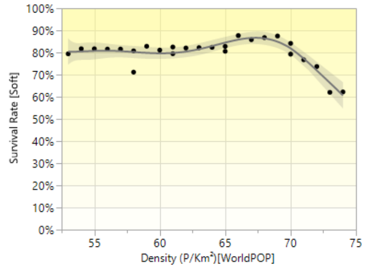
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.55947947	0.093247	51.6925
Error	19	0.03427357	0.001804	Prob > F
C. Total	25	0.59375303		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Survival Rate [Soft] = -0.089174 + 0.0143537*Density (P/Km^2)[WorldPOP] + 0.0012542*(Density (P/Km^2)[WorldPOP]-63.5)^2 - 0.0003799*(Density (P/Km^2)[WorldPOP]-63.5)^3 - 4.3881e-5*(Density (P/Km^2)[WorldPOP]-63.5)^4 + 1.4905e-6*(Density (P/Km^2)[WorldPOP]-63.5)^5 + 2.0778e-7*(Density (P/Km^2)[WorldPOP]-63.5)^6

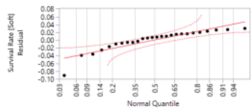
Summary of Fit

RSquare	0.834572
RSquare Adj	0.782331
Root Mean Square Error	0.029946
Mean of Response	0.799947
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.08595627	0.014326	15.9756
Error	19	0.01703821	0.000897	Prob > F
C. Total	25	0.10299448		<.0001*

Residual Normal Quantile Plot

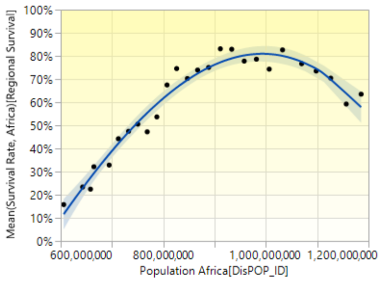


***** Polynomial Line of Fit (Degree=6)***

I.3: M^{i}E: Survival Rate x Regional Population

I.3.1: Africa

Survival Rate (Undifferentiated)



Mean(Survival Rate, Africa)[Regional Survival] = -0.352122 + 1.2423e-9*Population Africa[DisPOP_ID] - 4.702e-18*(Population Africa[DisPOP_ID]-8.73e+8)^2 - 2.751e-27*(Population Africa[DisPOP_ID]-8.73e+8)^3

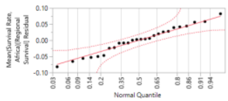
Summary of Fit

RSquare	0.959437
RSquare Adj	0.953905
Root Mean Square Error	0.044488
Mean of Response	0.597659
Observations (or Sum Wgts)	26

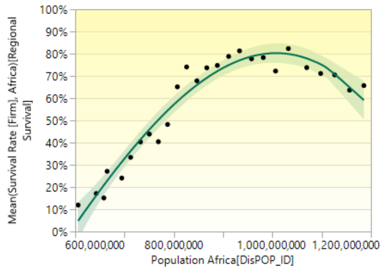
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1.0299039	0.343301	173.4533
Error	22	0.0435427	0.001979	Prob > F
C. Total	25	1.0734466		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm], Africa)[Regional Survival] = -0.561241 + 1.4456e-9*Population Africa[DisPOP_ID] - 4.66e-18*(Population Africa[DisPOP_ID]-8.73e+8)^2 - 3.535e-27*(Population Africa[DisPOP_ID]-8.73e+8)^3

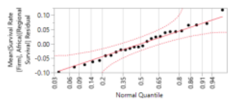
Summary of Fit

RSquare	0.947269
RSquare Adj	0.940078
Root Mean Square Error	0.056632
Mean of Response	0.566425
Observations (or Sum Wgts)	26

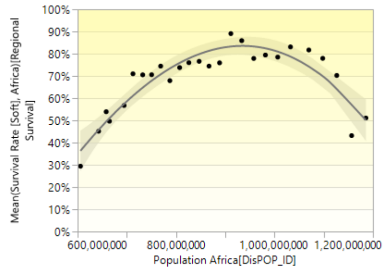
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1.2675297	0.422510	131.7365
Error	22	0.0705592	0.003207	Prob > F
C. Total	25	1.3380888		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft], Africa)[Regional Survival] = 0.3079525 + 5.833e-10*Population Africa[DisPOP_ID] - 4.598e-18*(Population Africa[DisPOP_ID]-8.73e+8)^2 - 1.651e-27*(Population Africa[DisPOP_ID]-8.73e+8)^3

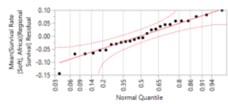
Summary of Fit

RSquare	0.849644
RSquare Adj	0.829141
Root Mean Square Error	0.06145
Mean of Response	0.686274
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.46944013	0.156480	41.4397
Error	22	0.08307390	0.003776	Prob > F
C. Total	25	0.55251403		<.0001*

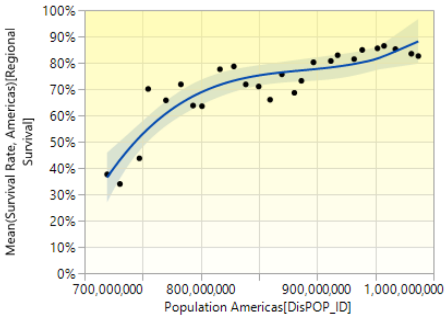
Residual Normal Quantile Plot



**** Polynomial Line of Fit (Degree=3)*

I.3.2: Americas

Survival Rate (Undifferentiated)



Mean(Survival Rate, Americas)[Regional Survival] = 0.2644834 + 5.748e-10*Population Americas[DisPOP_ID] - 5.83e-18*(Population Americas[DisPOP_ID]-8.6e+8)^2 + 7.063e-26*(Population Americas[DisPOP_ID]-8.6e+8)^3

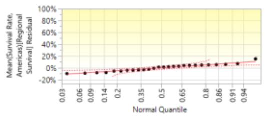
Summary of Fit

RSquare	0.823598
RSquare Adj	0.799543
Root Mean Square Error	0.063793
Mean of Response	0.717677
Observations (or Sum Wgts)	26

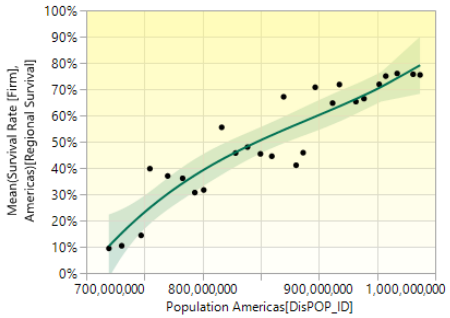
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.41800942	0.139336	34.2384
Error	22	0.08953123	0.004070	Prob > F
C. Total	25	0.50754065		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm], Americas)[Regional Survival] = -1.158976 + 1.9596e-9*Population Americas[DisPOP_ID] - 2.993e-18*(Population Americas[DisPOP_ID]-8.6e+8)^2 + 3.168e-26*(Population Americas[DisPOP_ID]-8.6e+8)^3

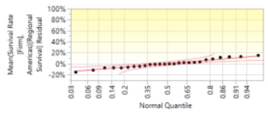
Summary of Fit

RSquare	0.864733
RSquare Adj	0.846287
Root Mean Square Error	0.08172
Mean of Response	0.505993
Observations (or Sum Wgts)	26

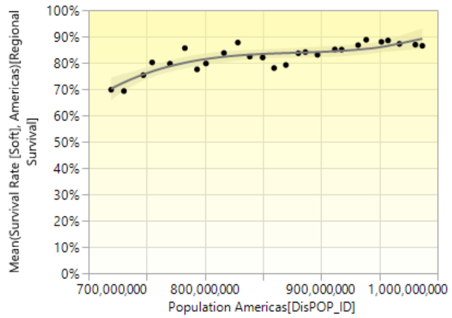
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.9392273	0.313076	46.8803
Error	22	0.1469203	0.006678	Prob > F
C. Total	25	1.0861476		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft], Americas)[Regional Survival] = 0.6950716 + 1.634e-10*Population Americas[DisPOP_ID] - 1.482e-18*(Population Americas[DisPOP_ID]-8.6e+8)^2 + 2.914e-26*(Population Americas[DisPOP_ID]-8.6e+8)^3

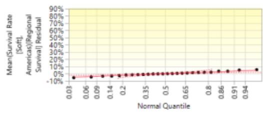
Summary of Fit

RSquare	0.739491
RSquare Adj	0.703967
Root Mean Square Error	0.028825
Mean of Response	0.824468
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.05188804	0.017296	20.8167
Error	22	0.01827917	0.000831	Prob > F
C. Total	25	0.07016721		<.0001*

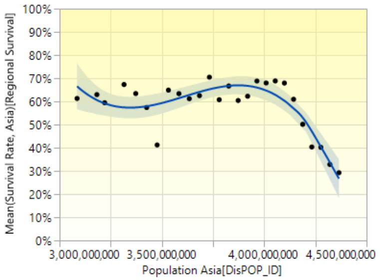
Residual Normal Quantile Plot



**** Polynomial Line of Fit (Degree=3)*

I.3.3: Asia

Survival Rate (Undifferentiated)



Mean(Survival Rate, Asia)[Regional Survival] = 0.0628174 + 1.584e-10*Population Asia[DisPOP_ID] - 6.664e-19*(Population Asia[DisPOP_ID]-3.78e+9)^2 - 1.298e-27*(Population Asia[DisPOP_ID]-3.78e+9)^3

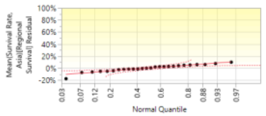
Summary of Fit

RSquare	0.755148
RSquare Adj	0.721759
Root Mean Square Error	0.061476
Mean of Response	0.581914
Observations (or Sum Wgts)	26

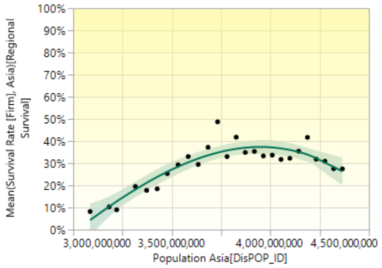
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.25642965	0.085477	22.6168
Error	22	0.08314560	0.003779	Prob > F
C. Total	25	0.33957525		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm], Asia)[Regional Survival] = -0.323474 + 1.803e-10*Population Asia[DisPOP_ID] - 4.969e-19*(Population Asia[DisPOP_ID]-3.78e+9)^2 - 1.51e-28*(Population Asia[DisPOP_ID]-3.78e+9)^3

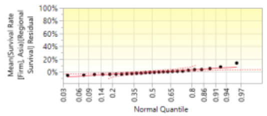
Summary of Fit

RSquare	0.814937
RSquare Adj	0.789701
Root Mean Square Error	0.046098
Mean of Response	0.291433
Observations (or Sum Wgts)	26

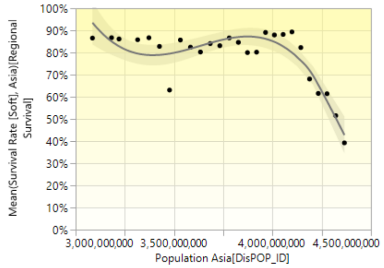
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.20587280	0.068624	32.2928
Error	22	0.04675148	0.002125	Prob > F
C. Total	25	0.25262428		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft], Asia)[Regional Survival] = 0.2460381 + 1.636e-10*Population Asia[DisPOP_ID] - 6.715e-19*(Population Asia[DisPOP_ID]-3.78e+9)^2 - 1.515e-27*(Population Asia[DisPOP_ID]-3.78e+9)^3

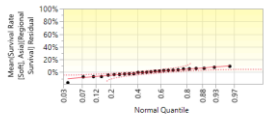
Summary of Fit

RSquare	0.805375
RSquare Adj	0.778836
Root Mean Square Error	0.060836
Mean of Response	0.785925
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.33693027	0.112310	30.3460
Error	22	0.08142156	0.003701	Prob > F
C. Total	25	0.41835184		<.0001*

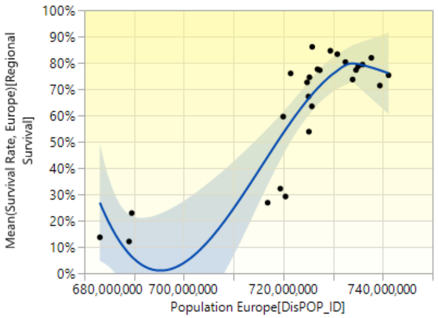
Residual Normal Quantile Plot



****** Polynomial Line of Fit (Degree=3)***

I.3.4: Europe

Survival Rate (Undifferentiated)



Mean(Survival Rate, Europe)[Regional Survival] = -17.20389 + 2.4645e-8*Population Europe[DisPOP_ID] - 5.761e-16*(Population Europe[DisPOP_ID]-7.24e+8)^2 - 2.372e-23*(Population Europe[DisPOP_ID]-7.24e+8)^3

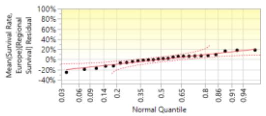
Summary of Fit

RSquare	0.777867
RSquare Adj	0.747576
Root Mean Square Error	0.1192
Mean of Response	0.627156
Observations (or Sum Wgts)	26

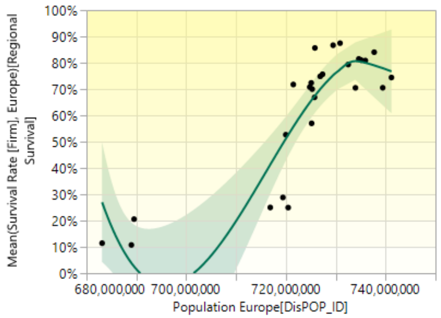
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1.0946232	0.364874	25.6799
Error	22	0.3125888	0.014209	Prob > F
C. Total	25	1.4072120		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm], Europe)[Regional Survival] = -19.12069 + 2.7278e-8*Population Europe[DisPOP_ID] - 6.183e-16*(Population Europe[DisPOP_ID]-7.24e+8)^2 - 2.654e-23*(Population Europe[DisPOP_ID]-7.24e+8)^3

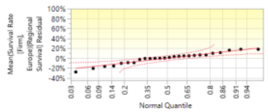
Summary of Fit

RSquare	0.787863
RSquare Adj	0.758935
Root Mean Square Error	0.122526
Mean of Response	0.621472
Observations (or Sum Wgts)	26

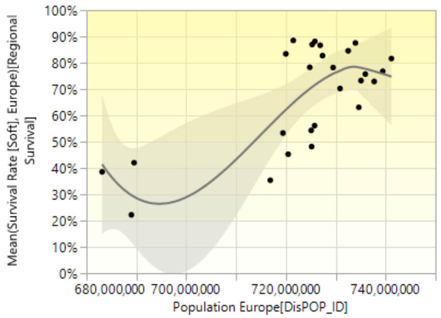
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	1.2266345	0.408878	27.2355
Error	22	0.3302794	0.015013	Prob > F
C. Total	25	1.5569139		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft], Europe)[Regional Survival] = -10.63395 + 1.564e-8*Population Europe[DisPOP_ID] - 4.165e-16*(Population Europe[DisPOP_ID]-7.24e+8)^2 - 1.571e-23*(Population Europe[DisPOP_ID]-7.24e+8)^3

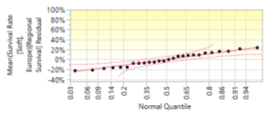
Summary of Fit

RSquare	0.525361
RSquare Adj	0.460637
Root Mean Square Error	0.142477
Mean of Response	0.674669
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.49431983	0.164773	8.1170
Error	22	0.44659555	0.020300	Prob > F
C. Total	25	0.94091538		0.0008*

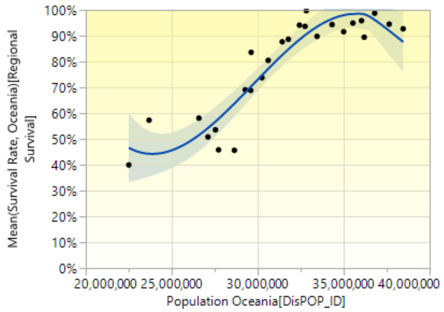
Residual Normal Quantile Plot



****** Polynomial Line of Fit (Degree=3)*

I.3.5: Oceania

Survival Rate (Undifferentiated)



Mean(Survival Rate, Oceania)[Regional Survival] = -1.163976 + 6.332e-8*Population Oceania[DisPOP_ID] - 3.385e-15*(Population Oceania[DisPOP_ID]-3.15e+7)^2 - 6.65e-22*(Population Oceania[DisPOP_ID]-3.15e+7)^3

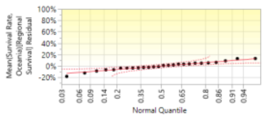
Summary of Fit

RSquare	0.861675
RSquare Adj	0.842813
Root Mean Square Error	0.076617
Mean of Response	0.7819
Observations (or Sum Wgts)	26

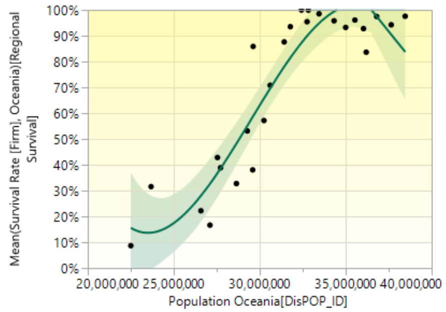
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.80447760	0.268159	45.6819
Error	22	0.12914305	0.005870	Prob > F
C. Total	25	0.93362065		<.0001*

Residual Normal Quantile Plot



Survival Rate (Firm)



Mean(Survival Rate [Firm], Oceania)[Regional Survival] = -2.317077 + 9.8583e-8*Population Oceania[DisPOP_ID] - 5.872e-15*(Population Oceania[DisPOP_ID]-3.15e+7)^2 - 1.012e-21*(Population Oceania[DisPOP_ID]-3.15e+7)^3

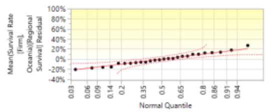
Summary of Fit

RSquare	0.859233
RSquare Adj	0.840038
Root Mean Square Error	0.123908
Mean of Response	0.701833
Observations (or Sum Wgts)	26

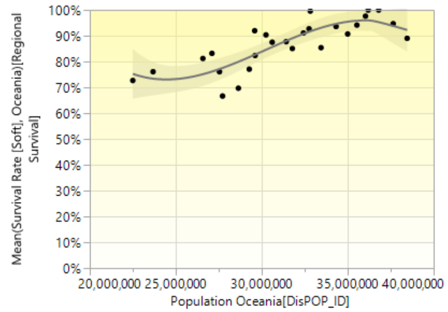
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	2.0617478	0.687249	44.7623
Error	22	0.3377729	0.015353	Prob > F
C. Total	25	2.3995207		<.0001*

Residual Normal Quantile Plot



Survival Rate (Soft)



Mean(Survival Rate [Soft], Oceania)[Regional Survival] = -0.010517 + 2.8358e-8*Population Oceania[DisPOP_ID] - 1.111e-15*(Population Oceania[DisPOP_ID]-3.15e+7)^2 - 2.988e-22*(Population Oceania[DisPOP_ID]-3.15e+7)^3

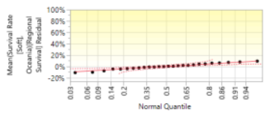
Summary of Fit

RSquare	0.687917
RSquare Adj	0.64536
Root Mean Square Error	0.055786
Mean of Response	0.86781
Observations (or Sum Wgts)	26

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.15091437	0.050305	16.1647
Error	22	0.06846451	0.003112	Prob > F
C. Total	25	0.21937888		<.0001*

Residual Normal Quantile Plot



****** Polynomial Line of Fit (Degree=3)***