



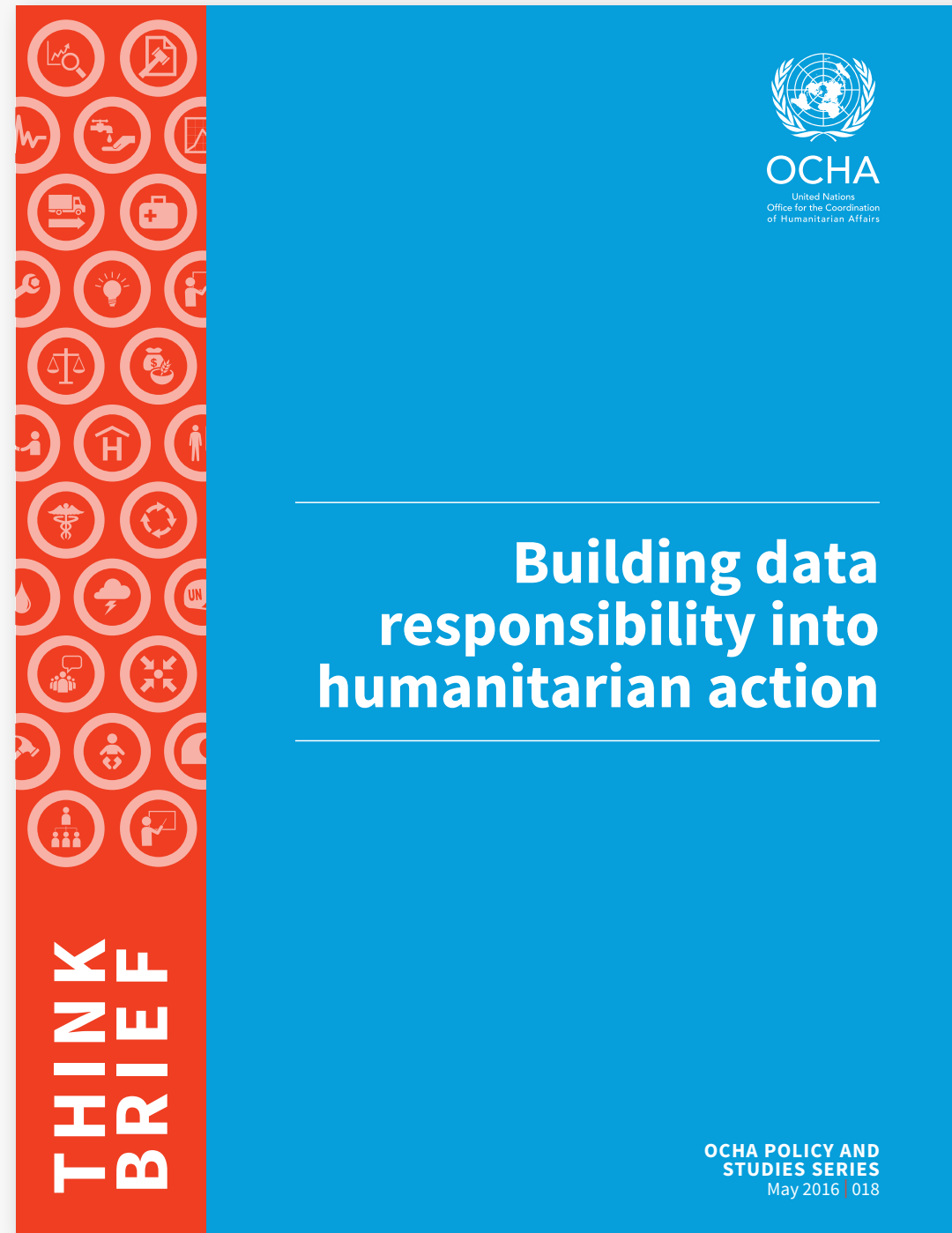
A FRAMEWORK FOR UNDERSTANDING DATA RISK

RESPONSIBLE DATA USE IN UNDERSTANDING RISK

STEFAN G. VERHULST




OCHA



The image shows the cover of a report. The left side is a vertical red bar with a grid of white icons representing various humanitarian and data-related concepts. The main body is blue. At the top right is the OCHA logo. The title is centered in white text between two horizontal lines. At the bottom right is the series information.

THINK BRIEF


OCHA
United Nations
Office for the Coordination
of Humanitarian Affairs

**Building data
responsibility into
humanitarian action**

OCHA POLICY AND
STUDIES SERIES
May 2016 | 018

When used responsibly, data has the potential to provide more targeted and efficient humanitarian interventions

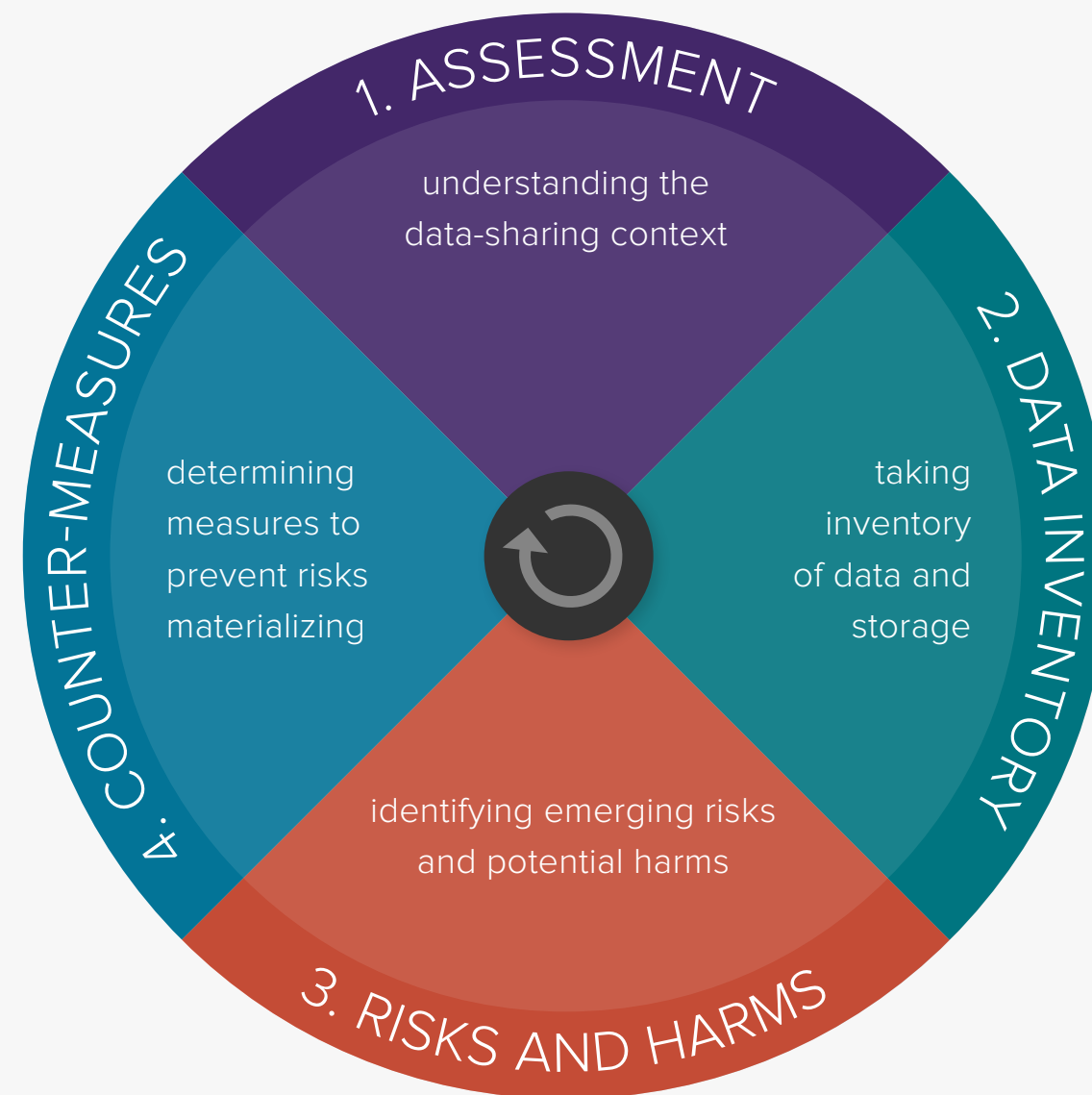
- ▶ **Existing data sources** (statistics, satellite, open data), when aggregated, triangulated and linked, can enable tracking of resource allocation and enable evidence based interventions. Example: HDX.
- ▶ **Crowdsourced data** can provide accurate, real-time information from communities most affected by disaster, supporting risk-reduction strategies both prior to and following crisis events. Example: Open Seventeen and SDG.
- ▶ **‘Crisis Mapping’** can help better coordinate emergency services, reducing risks associated with ad hoc ad mismanaged aid deployment. Example: Flood Monitoring in Pakistan and Philippines.
- ▶ By accessing and scraping **data from private actors including social media, telecom operators and others** humanitarian organizations have the potential to more accurately predict disaster events. Example: Mobility Patterns in Haiti.

Data also introduces new risks.

By building a data-risk framework, organizations can more effectively anticipate, prevent and manage these emerging “data risks”.

A DATA-RISK FRAMEWORK

A four-step data risk process to think through the potential risks and harms for data collection, storage and use.



WHAT IS HIGH-RISK DATA?

DISCLOSING IDENTIFIABLE DATA

- ▶ PII: Data that includes attributes about individuals. This is commonly referred to as **PII** or **personally identifiable information**.
- ▶ CII: Data that identifies **communities** or demographics within a group and ties them to a place (i.e., women of a certain age group in a specific location).

WHAT IS HIGH-RISK DATA?

RISK COMES WHEN, FOR INSTANCE

- ▶ Data is collected and shared without proper authorization from the individual or the organization acting as the data steward;
- ▶ Data is not be securely stored and thus vulnerable to hacks;
- ▶ Data universe that is being analysed is inaccurate and/or not representative;
- ▶ Data is being used for purposes other than what was initially stated during collection.

HARMS IN COLLECTING HIGH-RISK DATA

- ▶ Individuals and communities may feel exploited or vulnerable as the result of how data is used.
- ▶ Crowd-sourced information can be messy and inaccurate, and risks compromising data quality. This can lead to inaccurate and poorly designed relief strategies which can be ineffective and counter-proactive.
- ▶ Relief efforts based on data can underserve groups that lack access to or proficiency with digital technologies. This can in turn exacerbate existing social and economic inequalities following disasters, leading to what some call “digital discrimination”.

A DATA-RISK FRAMEWORK

1. ASSESSMENT

What is the anticipated benefit of using the data?

Who has access to the data?

What constitutes the actionable information for a potential perpetrator?

What could set off the threat to the data being used inappropriately?

A DATA-RISK FRAMEWORK

2. DATA INVENTORY

Where is the data – is it stored locally or hosted by a third party?

Where could the data be housed later?

Who might gain access to the data in the future?

How will we know – is data access being monitored?

A DATA-RISK FRAMEWORK

3. RISKS AND HARMS

Thinking through various risk-producing scenarios will help prepare staff for incidents. Examples of risks include:

- ▶ Your organization's data being correlated with other data sources that expose individuals;
- ▶ Your organization's raw data being publicly released; and/or your organization's data system being maliciously breached.

A DATA-RISK FRAMEWORK

4. COUNTER-MEASURES

Methods and tools include for instance

- ▶ Developing data handling policies and decision trees (Fair Information Practices)
- ▶ Implementing access controls to the data,
- ▶ Technological solutions,
- ▶ Training staff on how to use data responsibly, and
- ▶ Data Ethics Framework

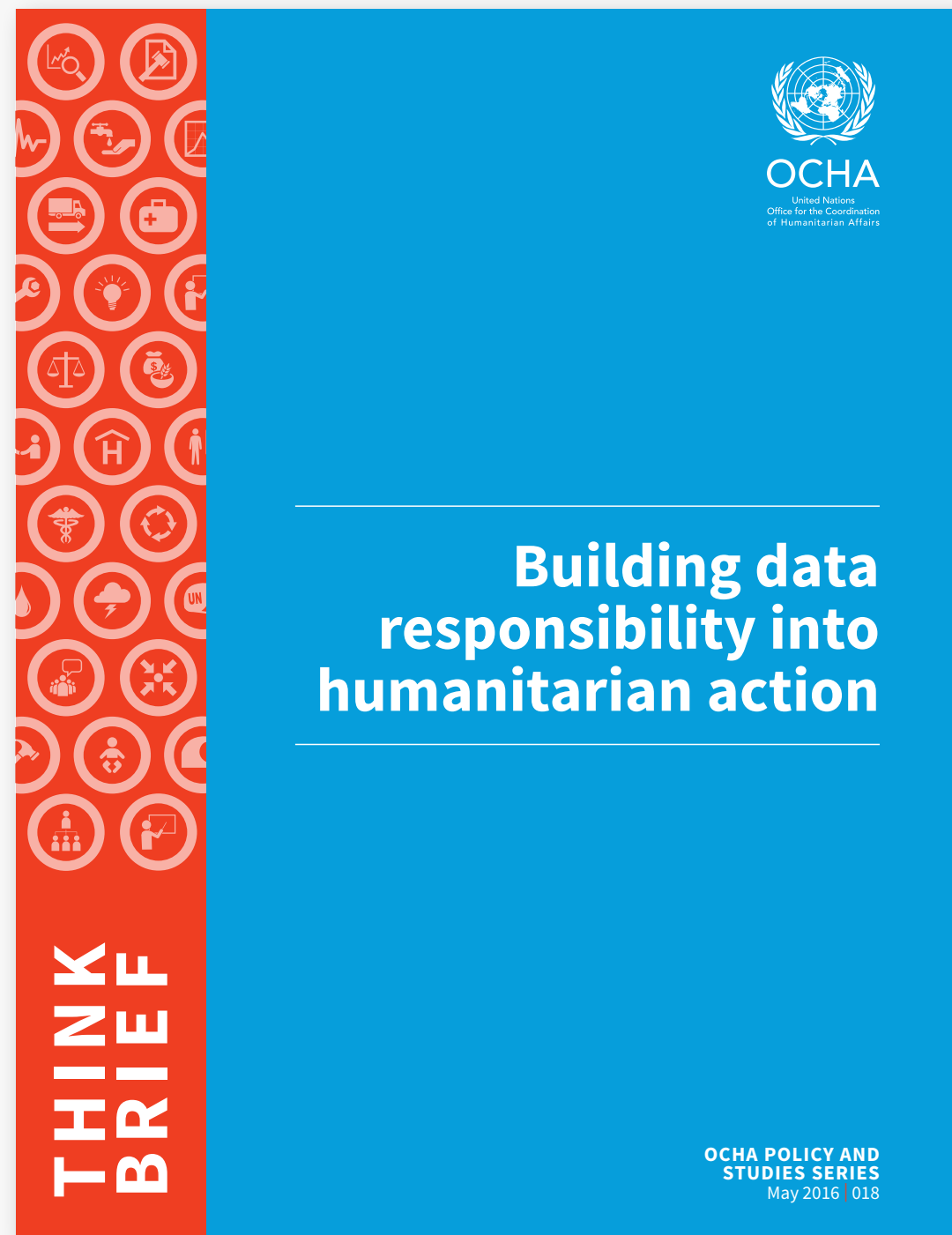
QUESTIONS

What can we learn from risk assessment in other areas?

What are current (best) practices that can inform the emerging field of data risk?

How to establish a responsible data culture?

How to share data responsibly in a systemic manner beyond bilateral agreements?





THANK YOU

@sverhulst

@thegovlab | thegovlab.org