OCHA centre for humdata

THE CENTRE FOR HUMANITARIAN DATA

GUIDANCE NOTE SERIES
DATA RESPONSIBILITY IN HUMANITARIAN ACTION

NOTE #8: RESPONSIBLE APPROACHES TO DATA SHARING

KEY TAKEAWAYS:

- Open sharing of timely and accurate data is essential to effective and efficient humanitarian response. How humanitarian organisations approach data sharing directly relates to trust and cooperation in the sector.
- As the humanitarian data ecosystem grows, the opportunities and risks of sharing data become clearer, prompting organisations to explore more limited approaches to data sharing.
- Humanitarian organisations widely recognize the sensitivity of personal data: its exposure has a high likelihood of causing harm. While the majority of non-personal data is safe to share openly, non-personal data can also be sensitive and should be handled with caution.
- Humanitarian organisations should take into account four factors when deciding whether to share non-personal data: (i) utility; (ii) sensitivity; (iii) human and technical capacity; and (iv) governance.
- Humanitarian organisations should identify and compare all available approaches for data sharing, considering the most open approach first and working down to more limited approaches as necessary.

INTRODUCTION

The use and exchange of data have become core functions of humanitarian organisations. Staff regularly need to decide whether and how to share their organisation's data, even if their role is not primarily focused on data or information management. Beyond individual organisations, the interest in the sharing and use of data generated in humanitarian action has also grown. In response to this interest, the humanitarian sector has seen a surge in data generation and sharing in recent years.

> "Accurate data is the lifeblood of good policy and decision-making. Obtaining it, and sharing it across hundreds of organizations, in the middle of a humanitarian emergency, is complicated and time-consuming – but it is absolutely crucial."
>
> - United Nations Secretary-General António Guterres at the opening of the OCHA Centre for Humanitarian Data in The Hague in December 2017

Open sharing of timely and accurate data is essential to effective and efficient humanitarian response and should remain a key objective for the sector. For example, the COVID-19 epidemiological data[1] compiled and shared daily by the Johns Hopkins University Center for Systems Science and Engineering has been

[1] Access the data on HDX: **https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases**.

integrated in a number of dashboards and situation reports for decision-makers across the humanitarian sector since the outset of the pandemic. This dataset has been downloaded more than 320,000 times from the Humanitarian Data Exchange (HDX) since it was first published on the platform in January 2020. The utility of open data to humanitarian practitioners is further underlined by the fact that usage of HDX in countries with a Humanitarian Response Plan (HRP) in place has grown much faster than usage in other locations.[2]

How humanitarian organisations approach data sharing directly relates to trust and cooperation in the sector. Maintaining trust within the data ecosystem is critical to the sustainability of data sharing, and relates to issues such as the quality of the data, the level to which the data will be secured after sharing, and the responsible use of data by the recipient. Because data in the humanitarian sector often relates to the most at-risk populations, managing and sharing it warrants caution.

Many humanitarian organisations have developed or updated their guidance, governance and practices to support different aspects of data responsibility: the safe, ethical and effective management of data. The sector has also seen an increasing number of collaborative efforts to improve data responsibility beyond individual organisations.[3] Still, as the humanitarian system learns more about the risks associated with data sharing, organisations face more complex challenges in sharing this data responsibly.[4]

This guidance note aims to support decision-making around the sharing of non-personal data in humanitarian settings. It explains data sensitivity, provides common examples of sensitive non-personal data, and explains an approach to information and data sensitivity classification in humanitarian settings. It also offers a framework that organisations can use to weigh four factors that help determine whether data can be shared and explains common approaches for doing so responsibly.

Options for Data Sharing on the Humanitarian Data Exchange

When the Humanitarian Data Exchange (HDX) was launched in 2014, it held close to 900 datasets, shared by a handful of 'early adopter' organisations. By the end of 2020, that number had grown to over 18,000 datasets. Only approved organisations are able to share data on the platform. They can make data available publicly to anyone who visits the site or privately to only the members of their organisations.

In 2017, the HDX team added another option for data sharing: HDX Connect. This feature enables organisations to only publish the metadata, with the underlying data available upon request. If access is granted, the data is shared bilaterally without passing through the HDX platform. For example, Ground Truth Solutions use HDX Connect to provide access to COVID-19 Community Perceptions Data collected in Iraq.

As part of its quality assurance process, the HDX team also runs a disclosure risk assessment on any resource added to the platform that contains microdata. The HDX team does this because it may be possible to re-identify individuals or expose confidential information even after direct identifiers have been removed from microdata.[5]

Some organisations on HDX have become more oriented towards controlled access to their data, either due to the sensitive nature of the data, increased pressure to track and report on how the data is used, or resource constraints related to operational sustainability. While HDX will always support different ways of sharing data, open access remains the best option for the majority of data that is generated for humanitarian response.

[2] "From August 2016 through August 2020 (the period for which the data is available), growth in monthly users from HRP+ countries was 943% compared to 566% across all countries." From the HDX Case Study, September 2020, available here: **https://centre.humdata.org/hdx-case-study/**.

[3] These include, for example, the IASC Sub-Group on Data Responsibility in Humanitarian Action, the Protection Information Management initiative, and the Responsible Data for Children initiative, among others.

[4] For a better understanding of the challenge facing humanitarian organisations when sharing data specifically in protracted humanitarian crises, see ALNAP, Data Collection, Analysis and Use in Protracted Humanitarian Crises, June 2020, available here: **https://www.alnap.org/system/files/content/resource/files/main/Humanitarian-Research-Brief-2.pdf**.

[5] Learn more about the Centre's risk mitigation process for microdata, 'Statistical Disclosure Control' or SDC in the Learning Path on the topic, available here: **https://centre.humdata.org/learn-how-to-conduct-a-disclosure-risk-assessment/**.

UNDERSTANDING DATA SENSITIVITY

Humanitarian organisations widely recognize the sensitivity of **personal data**[6]: its exposure has a high likelihood of causing harm. This understanding does not yet widely exist for **non-personal data**, which typically covers the following three categories in humanitarian settings:

- Data about the context in which a response is taking place (e.g. legal frameworks, political, social and economic conditions, infrastructure, etc.) and the humanitarian situation (e.g security incidents, protection risks, drivers of the situation or crisis).
- Data about the people affected by the situation and their needs, the threats and vulnerabilities they face, and their capacities.
- Data about humanitarian response actors and their activities (e.g. as reported in 3W/4W/5W).

While the majority of this data is safe to share openly, non-personal data can also be sensitive. Examples of sensitive non-personal data include data on groups experiencing gender-based violence or the location of ethnic minorities in conflict settings. Such data is considered sensitive because it enables the identification of groups of individuals by demographically defining factors, such as ethnicity, gender, age, occupation, religion, or location of origin. Non-personal data can also create risk in other ways, for example by exposing the location of medical facilities in areas where they are prone to attack. As the awareness of the risk associated with sharing such data continues to grow, some organisations are turning from a focus on open data to more controlled sharing.

Many organisations have specific classifications regarding what constitutes sensitive data in order to facilitate responsible data management practices. An information and data sensitivity classification (see figure 1 below) may also be developed as a collective exercise to help organisations align around what constitutes sensitive data in their context and identify the appropriate disclosure or dissemination methods for different data types depending on their sensitivity.

Information and Data Sensitivity Classification		
Sensitivity	Definition	Information and Data Sensitivity Classification
Low or No	Information or data that, if disclosed or accessed without proper authorization, are unlikely to cause any harm or negative impacts to affected people and/or humanitarian actors.	**Public**
Moderate	Information or data that, if disclosed or accessed without proper authorization, are likely to cause minor harm or negative impacts and/or be disadvantageous for affected people and/or humanitarian actors.	**Restricted**
High	Information or data that, if disclosed or accessed without proper authorization, are likely to cause serious harm or negative impacts to affected people and/or humanitarian actors and/or damage to a response.	**Confidential**
Severe	Information or data that, if disclosed or accessed without proper authorization, are likely to cause severe harm or negative impacts and/or damage to affected people and/or humanitarian actors and/or impede the conduct of the work of a response.	**Strictly Confidential**

Figure 1. Sample Information and Data Sensitivity Classification[7]

[6] Personal data should not be shared openly, and management of personal data should always comply with national and regional data protection laws, or with internal data protection policies in the case of organisations covered by privileges and immunities.

[7] UNOCHA (2019), Working Draft Data Responsibility Guidelines, available here: **https://centre.humdata.org/introducing-the-working-draft-of-the-ocha-data-responsibility-guidelines/**.

Organisational and collective governance instruments such as policies and Information Sharing Protocols (ISPs) often include a sensitivity classification and should be the primary points of reference for determining how to manage sensitive data. However, these documents tend to leave room for discretion in deciding whether and how to share. This means data sharing can be influenced by personal preferences and skills and may vary across organisations. By taking a more consistent approach to data sharing and ensuring adequate safeguards for sensitive data, organisations can build trust and contribute to more efficient and effective humanitarian response.

FOUR FACTORS FOR DETERMINING WHETHER TO SHARE NON-PERSONAL DATA

There are four factors humanitarian organisations should take into account when deciding whether to share non-personal data.

1. What is the utility of the data for other stakeholders?
The utility of data depends on the level of detail, the number of people or the geographical area covered, its timeliness, and its relevance to analysis and decision-making in humanitarian response. To determine the utility of specific data, conduct a Data Impact Assessment (DIA).[8]

2. How sensitive is the data?
The sensitivity of data is based on the risk associated with its exposure in a particular context.[9] In some response contexts, organisations, clusters/sectors, and system-wide coordination structures also have established data and information sensitivity classifications (see above) that can inform this determination. Conducting a DIA can also help determine the sensitivity of data. For survey results and other forms of microdata, sensitivity is closely linked to the risk of re-identification, which can be determined by applying a disclosure risk assessment.

3. What human and technical capacity do the organisations sharing and using the data have?
Both the organisation sharing data and the organisation(s) receiving and using the data should have sufficient human and technical capacity for responsible data management. This includes staff availability, data literacy, technical infrastructure and related resources. In environments with low connectivity, bandwidth-heavy data sharing methods may not be appropriate. For contexts with known security risks, data should typically be shared through more limited approaches.

4. Which governance instruments apply?
Common data governance instruments include ISPs, data sharing agreements for bilateral data sharing and licenses or terms of use for public data sharing.[11] These instruments should inform how data is shared in a safe, ethical and effective manner.[12] In some situations, governance will need to be developed for the selected data sharing approach. Governance instruments can address a range of topics and special provisions, but should always include the following elements: (a) clearly state the purpose and scope of sharing; (b) specify any limitations to how data should be managed after sharing; (c) define roles and responsibilities throughout the sharing process; and (d) establish procedures for data incident management.[13]

[8] See the Guidance Note on Data Impact Assessments here: **https://centre.humdata.org/guidance-note-data-impact-assessments/**.

[9] Risk is defined by the International Standardisation Organisation (ISO) as 'the effect of uncertainty on objectives', and is 'usually expressed in terms of risk sources, potential events, their consequences and their likelihood.' For data management in the humanitarian sector, risk can be defined as the likelihood and impact of harm resulting from data management. For more information on the ISO definition of risk see the ISO 31000:2018 Risk Management Guidance, available here: **https://www.iso.org/obp/ui#iso:std:iso:31000:ed-2:v1:en**.

[10] The World Economic Forum, together with **Washington University Centre for Information Assurance and Cybersecurity**, the **Sustainable Development Solutions Network TReNDS** and the **NYU GovLab**, have begun building a repository of data sharing agreements to support the professionalisation of this practice through their **Contracts for Data Collaboration** (C4DC) project.

[11] The licenses recommended for data sharing via HDX are listed here: **https://data.humdata.org/about/license**.

[12] To supplement directly applicable governance instruments, various policy and regulatory initiatives have taken place to develop global guidance and standards. Some guidance is specific to data management in the sector, such as the ICRC Handbook on Data Protection in Humanitarian Action, available here: **https://www.icrc.org/en/publication/handbook-data-protection-humanitarian-action**.

[13] For more information about data incident management in humanitarian response, see our Guidance Note on Data Incident Management: **https://centre.humdata.org/guidance-note-data-incident-management/**.

IDENTIFY AND COMPARE APPROACHES

Based on the four factors above, organisations should determine the best approach to data sharing. These approaches range from open sharing to maximise the benefit of data, to more limited approaches such as bilateral data sharing or only sharing data insights. The table below contains an overview of different approaches to data sharing and offers examples of some commonly used tools and platforms.

Approaches, Tools and Platforms for Data Sharing in Humanitarian Response [14]

Approaches to Sharing	**Open Access** Sharing data publicly is the most open approach to sharing, allowing unmediated access to anyone.	**Limited Access** Limiting access to data still allows select partners to use the data as long as they meet certain requirements.	**Bilateral Sharing** Bilateral sharing is the most limited way in which data can be shared, directly with one partner.	**Limited Access** Data that should not be shared at all can still offer value to partners, if they are allowed to query the data remotely or benefit from it indirectly.
Common Tools and Platforms	Organisation data platforms HDX Open Listservs	HDX Private UNHCR Microdata Library[15] IFRC GO[16] Cluster/Sector Sharing Closed mailing lists	HDX Connect Email[17] Dropbox	OPAL[18] Aircloak[19] Homomorphic encryption[20] Multi-Party Computation[21]

In comparing these different approaches, always consider the most open approach first and work down to more limited approaches as necessary. Different data types will require different ways of sharing. For example, large datafiles will require specialized infrastructure, and Application Programming Interfaces (APIs) are suitable for data that is published in the same format on a regular basis. Because technologies for data sharing continue to evolve, organisations should regularly revisit and compare available data sharing approaches.

[14] Not all tools and platforms in this overview have been vetted by the UN Secretariat. Always consult the relevant Information Technology advisors before using a new tool.

[15] UNHCR's MicroData Library is available here: **https://microdata.unhcr.org/index.php/home**.

[16] IFRC's ' GO Platform is available here: **https://go.ifrc.org**.

[17] Within humanitarian responses, one of the most common ways to share data is via email attachments. When sharing data via email, always take the necessary security precautions. While this way of sharing is responsible in some cases, there are often more suitable ways to share data. For information on how to encrypt email, see for example: **https://www.cloudwards.net/how-to-encrypt-your-emails/**.

[18] For more information about the Open Algorithms Project, see here: **https://www.opalproject.org**.

[19] For more information about Aircloak Insights, see the company webpage here: **https://aircloak.com/solutions/features-en/**.

[20] To learn more about homomorphic encryption as a way of sharing the value of sensitive data, see here: **https://www.microsoft.com/en-us/research/project/homomorphic-encryption/** and here: **https://www.wired.com/story/google-private-join-compute-database-encryption/**.

[21] To learn more about multi-party computation see here: **https://www.tno.nl/en/focus-areas/information-communication-technology/roadmaps/data-sharing/secure-multi-party-computation/**.

Unlocking Value from Data Without Sharing It: The Querying Approach

A relatively new approach to utilizing data without transferring the data itself is 'querying'. Querying allows third parties to formulate specific questions to be asked of the data without accessing it directly. The resulting insights can then be checked for sensitivity and any other issues by the holder of the data. This approach avoids transfers of data which can cause legal and ethical concerns, while still allowing for valuable insights to be used for public good.

In implementing a querying approach, it is critical to establish governance in the form of instructions and boundaries regarding the queries that may be sent, in order to prevent retrieval of sensitive information by posing a combination of questions.[22] Vetting users as well as their questions should always be a key step in the process around this type of approach.

Commercial solutions to set up querying approaches include Aircloak Insights, which acts as a 'proxy between analysts and the sensitive data they need to work with.' Another querying tool is the **Open Algorithms** (OPAL) platform. This tool was specifically developed for the humanitarian and development sectors, and is currently being piloted in Colombia.

Data Sharing Methods at the Joint IDP Profiling Service
In 2019, the **Joint IDP Profiling Service** (JIPS) was awarded a grant by the UNHCR Innovation Fund to research advanced data science methods for data anonymization. JIPS studied methods such as Multi-Party Computation and Homomorphic Encryption and worked with technical experts at the Johns Hopkins University Applied Physics Laboratory, Flowminder and the Government of Colombia National Statistics Office.

In close collaboration with Flowminder and building on their **Flowkit**, JIPS developed a prototype querying approach to enable humanitarian and development actors to safely access and query sensitive individual-level data without needing to share it. The team developed a technical workflow to demonstrate the viability of this approach with one single data provider, and mapped the problems and limitations in case of multiple data providers.

Organisations are encouraged to share their experience in promoting responsible data sharing with the Centre for Humanitarian Data via **centrehumdata@un.org**.

COLLABORATORS: **JOINT IDP PROFILING SERVICE (JIPS).**

The **Centre for Humanitarian Data** ('the Center'), together with key partners, is publishing a series of eight guidance notes on Data Responsibility in Humanitarian Action over the course of 2019 and 2020. The Guidance Note series follows the publication of the **working draft OCHA Data Responsibility Guidelines** in March 2019. Through the series, the Centre aims to provide additional guidance on specific issues, processes and tools for data responsibility in practice. This series is made possible with the generous support of the Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO).

This project is co-funded by the European Union

This document covers humanitarian aid activities implemented with the financial assistance of the European Union. The views expressed herein should not be taken, in any way, to reflect the official opinion of the European Union, and the European Commission is not responsible for any use that may be made of the information it contains.

[22] For an explanation of this risk, see for example: **https://www.usenix.org/conference/usenixsecurity19/presentation/gadotti**.