Responsible Development Data – Practitioner’s Guide

(V1)

Sept 2014

Table of Contents

[Introduction 3](#_Toc399194089)

[What is the Responsible Development Data - Practitioner’s Guide? 3](#_Toc399194090)

[Who is this guide for? 3](#_Toc399194091)

[How to contribute to this content 3](#_Toc399194092)

[Why is ‘responsible data’ relevant in humanitarian and development projects? 4](#_Toc399194093)

[Key Concepts 7](#_Toc399194094)

[What is considered ‘data’ within a project? 7](#_Toc399194095)

[Data responsibility 8](#_Toc399194096)

[Holistic security 8](#_Toc399194097)

[List of Ethical Questions and Considerations 9](#_Toc399194098)

[Planning 9](#_Toc399194099)

[Collecting 11](#_Toc399194100)

[Storing 15](#_Toc399194101)

[Analyzing 17](#_Toc399194102)

[Sharing 17](#_Toc399194103)

[Risk Assessment and Safety Planning Guide 19](#_Toc399194104)

[Case Studies 21](#_Toc399194105)

[Case Study 1: Biometric identification system for refugee registration 21](#_Toc399194106)

[Case Study 2: SMS for crowd-sourcing data in an emergency 23](#_Toc399194107)

[Case Study 3: Mobile application to report child abuse cases 25](#_Toc399194108)

[History of this document 27](#_Toc399194109)

[About the Authors 27](#_Toc399194110)

[Special thanks 28](#_Toc399194111)

# Introduction

## What is the Responsible Development Data - Practitioner’s Guide?

Increasing civil society trends toward digital data collection, storage and communications, and contributing to ‘big data’ and ‘open data’ raise responsible data challenges and ethical questions for development practitioners. The Responsible Development Data Practitioner’s Guide is intended to be an introduction to ethical and technical considerations when engaging in work that will gather or use human data in development or emergency relief contexts. This guide presents key questions, tools and tactics for responsible data practices organized by project cycle (planning and design, implementation, monitoring and evaluation) and across the data lifecycle (plan; acquire; process; analyze; preserve; publish/share), so that practitioners can carefully consider each step. This is merely a starting point to raise awareness among practitioners and meant to be a living project that will evolve alongside a community of responsible data practitioners.

## Who is this guide for?

Do you implement or fund projects with a data collection component (this could be any project that includes beneficiary registration, case management, monitoring & evaluation, indicator tracking, etc.)? Do you work with datasets containing personally identifiable information (e.g. medical/health records, address and contact information, etc.)? Do you care about data ethics but are not sure where to start? Are you involved in any of the following groups:

* Humanitarian response organizations and partners
* International development organizations and partners
* Research, monitoring, and evaluation organizations and specialists
* Volunteers / people collecting or using data

• International aid donors

If you answered yes to any of the above, this guide is for you.

## How to contribute to this content

As mentioned, this guide is meant to be a living project and we welcome your feedback on how to improve and expand the content. This is the first iteration of the document and additional sections have been planned. We are also interested in digitizing this guide to create an interactive element to the content and share up-to-date resources. If you would like to contribute, please send your thoughts and comments to [dataguide@responsibledata.io](mailto:dataguide@responsibledata.io).

## Why is ‘responsible data’ relevant in humanitarian and development projects?

Humanitarian and development practitioners collecting and communicating project data are facing shifting technological and political terrain. Their work rapidly evolves and threats arise from formal and inform actors in crisis situations and contexts of relative security and peace. Like human rights defenders and journalists, project practitioners and participants face a range of risks and threats that go beyond data co-option and digital attacks to include personal intimidation, psychosocial distress and physical harm. Humanitarian and development practitioners need to build up a minimum of knowledge in responsible data practices in order to effectively identify and minimize risks to themselves and to project participants, and to develop safety plans for collecting, handling and communicating project-related data.  
  
*Source: Adapted from Romero, C. (2014) “What Next? The Quest to Protect Journalists and Human Rights Defenders in a Digital World,” Conference Report, Freedom House and USAID, http://www.freedomhouse.org/report/special-reports/what-next-quest-protect-human-rights-defenders-and-journalists-digital-world#.VCBsKl4k9g1.*

Collecting and using data to gain insights into how to address the world’s most pressing issues is certainly nothing new in the fields of international development and humanitarian response. Mass amounts of data are now easier to collect and archive, accelerating interest in data collection and popularizing digital data collection tools. While digital data collection methods and tools purport to increase efficiencies and data accuracy, donors are also demanding more data collection and analysis for funded projects.

This has led to a situation where there are copious amounts of data collected, produced, and made available by individuals or organizations that may not be fully aware of how to responsibly protect these data and the people these data might identify. This is potentially endangering as data is rapidly becoming a precious commodity and those with access to data can create new avenues of power, while those without control of their own data become increasingly vulnerable to risks and threats of harm. Here are some recent examples[[1]](#footnote-1) illustrating some dangers:

*“A project working to improve public health services, including HIV treatment centers, develops an online map based on individual reports, and uses patient data to map treatments sites. Failing to understand the local context, the map inadvertently “exposes” individuals who are receiving HIV treatment in a country where HIV is associated with homosexuality and homosexuality is strongly taboo.”*

*“A project advocating for the rights of a marginalized community conducts surveys with that community to document their satisfaction and needs as related to public service delivery. Survey results are used in advocacy and the raw data is released under an open license in keeping with the project’s aspiration towards transparency. Shortly thereafter, the municipal government launches an infrastructure project that clearly excludes the marginalized community, apparently on the basis of GIS data released by the project.”*

*“A campaign organization working in a highly oppressive context saves the names and personal information of people reporting abuse of public service on a desktop, and keeps backups on a USB. Though similar information is protected in a safe when collected in paper formats, the organization does not have strong protocols for passwords and access to data. One night the organization’s offices are burglarized and a USB with these names goes missing. That week, several of the people who had reported corporate abuse are arrested.”*

*“An international organization advocating for the rights of a local community determines that mobile survey technology, thanks to its savings of cost and time, would allow for much more regular data collection from that community. They secure funding to increase data collection using mobiles, also by supplying community members with hardware to conduct surveys. The community experiences the increase in data collection, but does not see an increase in advocacy impact (services and accountability do not improve over time). To the contrary, the allocation of novel technologies to specific individuals exacerbates social tensions within the community, and some community members express a feeling of exploitation, provoking general resentment towards the project. The relationship between the project and the community deteriorates, worsening data quality and limiting the project’s scope of activity.”*

These anecdotes represent different kinds of harm. Some are more obvious and damaging than others. Similar stories have circulated in different contexts and with different types of tools. What they have in common is that projects were not able to anticipate the dynamics, exchanges and risks implied by digital information. Whether this is a failure to translate traditional security practices into a digital context (as with the USB drive), or to understand how technology can impact local social relationships (as in the mobile surveys), it is challenging and dangerous despite the fact that most projects are smart, motivated, thoughtful and well-intentioned. It’s simply that with new technologies come new risks, and most of us aren’t familiar with them yet.

*Potential Threats to Data Privacy*

Data can be compromised both directly and indirectly. Three common threats to privacy are: surveillance, data intrusion and providing too much information. We must address these threats throughout the data lifecycle to mitigate risk and ensure privacy.

* *Surveillance:* Governments, telecommunications companies, and other adversaries can intercept transmitted information to gather intelligence. While this can be mitigated by ensuring information is sent securely, metadata cannot always be controlled or secured. In heavily surveilled contexts, it is important to consider what information can unintentionally be collected and is outside of your control whenever a digital solution is used.
* *Data Intrusion*: Militant organizations, such as the Syrian Electronic Army, have begun to illegally compromise computers to steal data. In these environments it is important not only to consider what data are being stored and how they are being stored, but the larger information security system. It is not always appropriate or necessary to store data digitally, where they are more vulnerable to being remotely compromised.
* *Providing Too Much Information*: When sharing data it is important to aggregate to a scale where an individual cannot be identified by the information provided. In the US, census data are aggregated to the census tract, even though they are collected at a household level, to ensure privacy for respondents. Anytime data are shared it is important to consider what data can reveal an individual.

Humanitarian projects rarely have built-in plans for how to limit data collection to what is necessary and proportionate, control access to data across its lifecycle, or contingency procedures to carry out in the case of an information security breach. Project design rarely considers what information needs to be stored, the metadata that is automatically collected, and which actors have inherent access to this data.

It is time that humanitarian and development practitioners and project-based research, monitoring and evaluation communities start asking and answering hard ethical questions about collecting and managing their data. It is imperative that these communities draw up data management policies and articulate practices that are consistent with an organization’s values and minimize the potential for harm.

This document attempts to serve as a guide to understanding what ethical considerations should be included in undertaking development projects that use data: from the collection and storage of data, through the management of their life cycle to its presentation. The guide provides suggested tools, tactics and further resources for integrating responsible data practices.

# Key Concept

## What is considered ‘data’ within a project?

Data is any information about the project plans, context or participants that is manually or digitally captured, stored, shared, or published either via physical paper, on a computer, on a mobile device, and/or on the internet.

**Raw data** is data collected from observation, document review, interviews or automatically through mobile devices or the web that has not yet been manipulated in any way. Some examples:

* Information captured by hand-written means (e.g. filling out paper forms)
* Information captured by digital means (e.g. biometrics, online surveys, audio or video content, e-mail text, text messages or app usage, etc.)
* **Meta data** is "data about data". This can include dates of interaction, GPS locations, IP, duration of calls, etc.

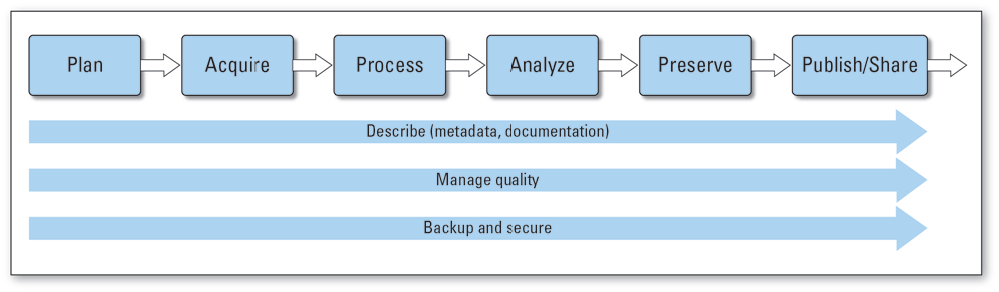
**Processed data** is raw data that has been manipulated in any way. This could include removing columns or anonymizing data, extracting entities from long text, coding free-form text, etc.

**Aggregated data** is raw data that has been summarized around certain fields for example summarizing the data around time periods such as totals by day, by week, by year or geographic locations such as averages by village.

All of this data can be stored in paper, laptops, digital devices, the cloud, and can take the form of spreadsheets, databases, or media files.

**Visualized data** is representing data in summarized, visual ways. This can be as simple as producing graphs or infographics, or can take interactive forms. Visualizations inherently tell a specific story both because of what, and also how, it is presented

**Data Lifecycle** describes what happens to data across a humanitarian or development project cycle. Project cycle related data must be proactively managed in each stage of the data lifecycle to prevent and respond to risks to people and to organizations. Incorporating responsible data management principles and practices is essential to processes of planning, implementing monitoring and evaluating activities across a humanitarian or development project cycle.



Data responsibility— While data collection can be a very powerful in aiding the work of the development community, if mismanaged it may cause harm to the very people it is trying to protect. Even the most well intentioned data handlers can easily risk compromising the security of their data without realizing it. Data responsibility is an overall concept that draws attention to the often overlooked or ignored issues of accountability that are involved in data management. These include, but are not limited to, drafting and implementing formal information security policies that do the following:

* Classify which data (and meta data) are collected and which are not. If possible, this information should be shared with the community from which you are collecting the data. This can be part of an informed consent process (see glossary for definition).
* Define roles of data handlers within an organization. This generally means who has access to which data. Logically, more sensitive data should only be accessible to a limited group of people on the project, not volunteers or other individuals with whom your organization may have a less formal relationship.
* Establish rules for handling these data, especially their lifecycle (i.e. how long they will be kept, if they will be passed on to another organization, etc.) and how information is stored in a database (i.e. coded/encrypted, as a hash, clear text).
* Clearly define how you will and will not share data on a project-by-project basis.

Holistic security – Humanitarian and development practitioners must adapt security protocols and tactics to encompass: 1) digital information security; 2) physical and operational security; and 3) psychosocial well-being required for good security implementation. These aspects comprise a new three-part approach to “holistic security.” Digital security is not only a question of a focus on software or tools. It requires integrating emotional well-being, personal and organizational security. Good implementation of digital security tools and tactics requires attending to the practitioners’ psychosocial capacities to recognize and respond dynamically to different threats to themselves and to participants related to project data collection and communications. Humanitarian and development practitioners face high stress situations combined with new risks related to the use of new technologies across the data lifecycle in project planning, implementation, monitoring and evaluation. Practitioners’ understanding of “Do No Harm” related to responsible data in humanitarian and development projects needs to be deepened and expanded to include digital, physical, and psychosocial aspects for a holistic approach to security. Please see the Tactical Technology Collective’s note, “Towards Holistic Security for Rights Advocates” for further discussion of this key concept: <https://tacticaltech.org/holistic-security>

# List of Ethical Questions and Considerations

The following sections outline lists of ethical questions, considerations and some definitions grouped by project stages across the data lifecycle: Planning, Collecting, Storing, Analyzing, and Sharing.

## Planning

*Define the scope*

* Describe the group from whom you intend to collect data (age/gender/locations/religious/cultural identities/other specific descriptors etc.)
  + Describe the ways in which this group is potentially vulnerable, digitally, physically and psychosocially.
  + Describe persons/groups/entities that could potentially cause harm to the described group and how this harm might be caused
* Describe the specific data you intend to collect over a specified timeframe.
  + How will the data be used? What goals will the data help to achieve?
  + Are you collecting only what is necessary for the current project scope?
  + Are you balancing future possible scope with risks associated with data collection?
  + Could any of the information you plan to collect lead to harm, especially if fallen into the wrong hands?
  + How will the subject group directly and/or indirectly benefit from the information collected?

*Plan for participation*

***Informed Consent*** – (human subjects research ethics, purpose of data collection, potential risks and benefits to participants, risks of metadata leakage to third parties, mitigating steps taken) Standard components of data collection informed consent include plain language, easy-to-understand explanations of the types of data to be collected, the purposes of collecting those data, the intended and potential unintended uses of that data, who has access to and control over the data, risks of data leakage to third parties, and any benefits to participation in data collection. A “responsible data” approach to informed consent would discuss specific steps taken to mitigate risks to all aspects of holistic security and participants’ well-being: digital, operational-physical and psychosocial.

* How will you involve local groups/partners/subjects in the planning process?
* What is the process of informed consent among relevant group members (informed adult consent; Informed caregiver consent; Informed children’s assent[[2]](#footnote-2))?
* How will the group interact with the data during various stages of the project?
* How will the group provide individual feedback/concerns and what is the plan to respond to feedback/concerns? Is there confidential reporting mechanism and follow-up plan for fear or incidence of harm to participants?

*Plan for data collection*

***Pre-collection*** – As part of a responsible approach to data in humanitarian and development projects, holistic security, risk assessment and mitigation planning needs to be built in to project planning activities, including any formative research, feasibility studies, or rapid assessments used to generate grounded insights on the basis of which to design project activities, and a framework to monitor and evaluate the projects’ objectives, activities, outcomes and impacts. Project planning should include holistic (digital; operational-physical; psychosocial) security planning across the project cycle and throughout the data lifecycle.

* Who will be performing data collection?
* Have they done informed consent, data collection and data management work before?
* What training will be made available to data collectors? Ensure training is comprehensive especially if this is the first time they are involved in data collection work.
* What may be some potential challenges between data collectors and responders (language/gender/age/etc.)?
* Consider existing cultural norms that may present challenges to data collection – how will you work with the community to address these challenges?
* Anticipate risks to subjects/respondents: how is it appropriate to anonymize and/or aggregate the data collected?
* Consider the technology landscape / mobile access / electricity situation.
  + How does the above inform tool choices? How does the tool choice create or limit digital, physical, or psychosocial security threats? Does this require you to reconsider what data is collected or how the data is collected?
  + Do you need to plan for local human resources to support ethical use of the selected technology? If so, ensure also there is local technical capacity to support the selected technology.
* Is the selected technology appropriate for use in front of the respondent group? Why/why not?
* Is the data collector group familiar with the selected technology? If not, ensure comprehensive training is made available or reconsider selected technology for an option that is more familiar to the data collector group.
* How is the internet infrastructure regulated in local context?
* How technically savvy are potential aggressors?
* How do you plan to ensure your choice of technology doesn’t compromise yourself or the subject/group?

*Perform a risk analysis*

* Do the potential risks to project participants of data collection outweigh the benefits to participants of data collection?
* Have you conducted a risk/threat assessment prioritizing probable risks and developed a safety plan of steps to mitigate those risks? Consider threats to respondents, data collectors, and software/tools/servers. See also sections below, “Risk Assessment Guide,” and “Case Studies,” for tools and tactics concerning risk assessment, safety planning, and case study examples with suggested mitigation steps to provoke further learning.

## Collecting

*Tools*

* Consider the pros and cons of using open source tools for the project (pro: may be low/no cost; con: lack of technical support, etc).
* Will you need to lock down other features within your selected technology to ensure it’s only used for appropriate data collection? Using other features not related to data collection could present additional security challenges.
* Will data collectors need to edit their data/information? If so, ensure the technology supports the desired level of data editing.
* Ensure selected tools allow for your desired range of data access (will you need persons with various levels of permissions to interact with specified sets of data?).
* Establish a response plan for potential technical challenges (phones breaking, no wifi, etc.)
* Establish a response plan in case tools are confiscated or lost (remote wipe, etc.)

*Method*

* Establish a clear data collection process – will data collectors perform data entry in front of responders or after a conversation with responders? Consider the impact of the data collection method on the desired conversational outcomes between the data collector and respondents (e.g. conversations about sensitive topics shouldn’t be interrupted by the data collection tool).
* Does the data collected need to be centralized? If so, consider the data flow process from data collection to centralization. What is the desired frequency of data centralization? How secure is the data centralization process?

***Data anonymization*** – Digital anonymization of personally identifiable information is fraught with limits and privacy risks. “Anonymized” names, contact information or “merely demographic” information can easily provide clues that can lead to bootstrapped searches that reveal personal identities. Many seemingly innocuous, neutral or “common” data points could potentially de-anonymize individual identities. Privacy laws have not kept pace with new technological realities and possibilities of triangulating across multiple digital data sources. Deanonymization research results have shown that it is no longer sustainable to make hard distinctions between “personally-identifiable information” (or PII) and “non-personally identifiable information” (or not PII). Classifying data points as PII or not PII has become increasingly questionable as new technologies and data proliferate. Paul Ohm has argued that it is more appropriate to think of identifiability as a continuum. It is imperative that practitioners handling personal data think critically through their data collection, storage, processing and sharing practices, and how effective the anonymization steps are that they are using. While all usual steps should be taken to anonymize data and aggregate information such that it becomes more difficult to identify individuals, digital data and metadata collection, tracking and government surveillance have further shown the limits of digital anonymization. Encrypting data and minimizing metadata leakage have become imperative for reducing risks of data de-anonymization to an acceptable level given the specific challenges or threats to digital, operational-physical and psychosocial security associated with participation in a given humanitarian or development project.

For further discussion, see: Ohm, P. (2009) “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,” UCLA Law Review, Vol, 57, P. 1701, 2010, University of Colorado Law Legal Studies Research Paper No. 9-12, http://papers.ssrn.com/sol3/papers.cfm?abstract\_id=1450006

*Security*

* Consider using strong encryption software with good end-to-end implementation to protect data and communications content.
* Consider built-in protection features within your selected technology and their potential vulnerabilities – how will you mitigate potential vulnerabilities?

Data encryption methods make your data unreadable to those who have not been granted access. There are many ways to encrypt data and the method chosen should aim to address specific risks. Setting unnecessarily strict encryption standards has the potential of overburdening and frustrating data handlers, who may opt to ignore policies in favor of maintaining productivity. Below are some examples of simple ways to encrypt data. Some are more user-friendly and stable than others. This list is by no means exhaustive.

* **Hard Drives** 
  + Mac: go to “System Preferences” and then click on “Security & Privacy.” You will see an option to turn on the “FireVault” which encrypts the hard drive.
  + PCs: for Windows 7 or 8, encrypt by turning on BitLocker, which is a built in feature of these operating systems. (Previously, Windows XP and non-Windows users were recommended to use TrueCrypt for encrypting their hard drives. However, recently potential security holes were discovered in TrueCrypt and it is not recommended at this time. )
  + Linux (e.g. Ubuntu): use the built-in disk encryption features of these systems.
* **Mobile Devices**
  + Android: use built-in security feature in the settings menu.
  + iPhone: enable the pass lock to encrypt. Users may select a simple 4-digit code or a more complex one, which is recommended. To add a bit more security, you can select the option to wipe the iPhone after a password is entered incorrectly 10 times. For still more security, it is also wise to encrypt any backups of your mobile device data. Once your hardware is encrypted, if sensitive data will be transmitted over voice calls, text or mobile chat, you may want to consider encrypting these communications. There are several open source options that offer this services such as RedPhone, Silent Circle, Securechat, Securetext, and Cryptocat.
* **SMS****or Short Message Service** (e.g. text messaging) is typically sent in plain text and is very difficult to encrypt (in apart because it can potentially travel through a variety of service networks, encryption is optional, and when there is encryption it is easily broken). Messages can easily be captured by someone snooping on the line.
* **Email and Instant Messaging** — One basic thing to remember is that most people use the email or IM services of private companies, which have varying degrees of transparency regarding how they manage your data and whom they give access to it. As a general rule, no matter which email provider you are using, you can encrypt the content of your messages with PGP or GPG keys. These are often used in conjunction with email clients such as Enigmail, Thunderbird, or Apple Mail. However, if you are accessing your email directly through the web browser, you can still use Mailvelope with Google Chrome to send PGP encrypted emails. Just make sure that your computer is encrypted and protected by a password because the PGP keys (the ones that unlock the emails) are saved in your browser.
  + One terrible inconvenience about using PGP encryption is that both communications partners need to use it in order for it to work. Sometimes this is not possible because local partners may have low-tech skills and explaining and installing PGP tools from a distance can be challenging. If you determine that email and IM encryption will be necessary for your project, it is recommended that you build in training into the project planning and practice using it before the start of the project.
  + Another option for improving the security of email is enabling two-factor authentication. Gmail offers this with the Google authenticator App or you can also use Yubikeys, which generate a unique 52 character password from a thin plastic key that fits on your keyring.
  + For encrypted video calls and chat, Jit.si is an open-source encrypted alternative to Skype. You can create an account on Jabber using XMPP, but you can also use Jit.si to encrypt other chat services like G-chat. However, Jit.si is pretty finicky and many people report that it can work poorly at times. There is no support documentation available online and new versions (stable and unstable) come out all the time without notification. It may be difficult to wean people off Skype, especially since it may be more reliable and popular. If your project will require encrypted video calls and IM, it is recommended that you install and test Jit.si or another option with your partners before the kick-off.
* **Zip or Rar** — Encrypting Zip or Rar files with a password of your choice is easy and does not require your communications partner to download additional software. Simply select “encrypt with a password” under the settings when creating the Zip or Rar archive. Remember to share the password with your partner through a secure channel.
* **Flash or USB drives** — These days encrypted USB drives are readily accessible and inexpensive. The device allows you to encrypt all the data on the drive with a password. Usually if the password is entered incorrectly after a certain number of times, the data will be wiped.

*Training*

* Ensure appropriate training on the established informed consent process and human subjects research ethics
* Ensure appropriate training on data security for secure data management – password best practices, anonymized data, device security, etc.

## Storing

*Where/How*

* Where should data be stored given the type of data and the potential vulnerabilities?
* Should data be stored on-line or off-line or both? Within borders or in another country? Or is it open data?
* Consider pros/cons of third party data storage vs local data storage (e.g. do you have local capacity for local data storage; what is the desired uptime; etc.)
* What are the implications of storing data locally or outside of the borders?
* What data back-ups are required?
* Ensure storage method allows for the necessary levels of access.
* What is the data lifecycle and who will manage it until the kill date?

*Security*

* Consider local internet regulations and potential challenges or harm from storing sensitive data.
* What is the data access / data saving policy, especially for sensitive data (e.g. if data must be downloaded locally it must be encrypted, etc.)

There are many ways to store data that vary in terms of convenience and security. How you store your data should be measured against the particular risks you or your partners may be facing. How you choose to store your data will vary depending on your most acute risks. Are you more concerned about data loss or surveillance? These are extremely important questions. Maybe it is not really important if authorities see your data, but it would be disastrous if they were somehow destroyed. It is important to understand that no method of storing data is 100% safe from risk. This is why it is essential that you backup your data no matter how/where it is stored.

***Physical storage*** — This includes paper documents, electronic versions of documents stored on physical drives (hard or external/flash or a local server) on location. The advantage to storing data in this format is that it can’t be hacked because they are not online. It is also much easier to control who has access to this data. Many organizations choose to keep their most sensitive data offline in undisclosed locations only known by one or two select people. The downside is that with all other physical things, they can be stolen or destroyed very easily or with the case of drives, they can simply break.

***Offsite servers*** — Websites, databases, document archives, and other data can easily be stored in servers off site or out of country. This is a practical solution for organizations, which may be subject to physical raids, or in countries where hostile authorities can get access to data in domestic centers or companies without legal warrants. Usually, for an increased fee, server companies offer additional services that can enhance security such as DDoS protection, regular security maintenance and monitoring traffic. If this option is chosen, be careful to check the political relationships and legal agreements between the country where the data is stored and the country or countries where you are working. Feel free to test the companies where you choose to store your data. Call them up and ask them what would happen if the security services of country X asked for data or make a phony request for data without proving your authorization. The downside of this method of data storage is that the data is transmitted over the internet, presenting a whole slew of new risks. Sensitive data should only be sent over secure, encrypted channels. Setting up a private VPN connected is highly recommended for very sensitive data. Additional security measures, such as defining specific IP address that can connect to the server or monitoring a connection log, may be essential in securing the integrity of the data. Offsite storage may also be unrealistic for organizations working in countries with unreliable internet access or frequent and unpredictable power shortages.

***Cloud storage*** — This method of storage is similar to storage on off-site storage, but differs in the sense that data are not stored in one place but rather pooled onto several servers usually owned by a commercial hosting company like Google, Apple, Amazon, etc. While this form of storage can be very inexpensive and a great option for not-so-sensitive large datasets, it is often less secure than purchasing your own server space because the client/service provider relationship is murkier and less accountable (insecure for high risk, sensitive data and advanced threat models).

## Analyzing

*Who*

* What are the multiple levels of data access permissions?
* Who should have access to the data and how will they access it?
* Have the persons performing the data analysis been provided the proper data security training?
* Who should have access to raw data, who can be trusted to anonymize data, and who should only have access to subsets of data?
* Who should sign an NDA when accessing the data? What should the NDA contain?
* How will the data be accessed? How should the data be sent and how should questions about the data be communicated?
* Do the people conducting the analysis properly understand data security?

*Method*

* How will data be analyzed/presented?
* Who should have access to raw data, who can be trusted to anonymize data, and who should only have access to subsets of data?
* Who should sign an NDA when accessing the data? What should the NDA contain?
* How will the data be accessed? How should the data be sent and how should questions about the data be communicated?
* Do the people conducting the analysis properly understand data security?

*Data uses/impact*

* What are potentials risks associated with the end results of the analysis?
* Are there enough data points that analysis will definitely keep the individuals anonymous?
* Have you considered all the worst (and best) case scenarios that the results may lead to?
* Have the results, and methodology, of the analysis been used to improve the data literacy of the population?

## Sharing

*Who/How*

* If sharing data with third parties, consider establishing a “terms of use” guide. What should happen if the “terms of use” are breeched? Who should monitor for breeches and how?
* What is the selection/screening process for third parties who want to access the data?
* Do you have complete knowledge of who the data is being shared with?
* How can you ensure that data consumers don't misread and misinterpret data?
* How will the data/information be shared with the originator/respondent groups?
* How can originators of data access this information and associated analysis/reports on their own?

*Lifecycle*

* What is the lifecycle of the data? Is there a plan to dispose of the data after it’s no longer useful?
* What will happen to the data if the team can no longer manage it? Should it be destroyed? How can you ensure that it will not be sold or end up in the wrong hands?

# Risk Assessment and Safety Planning Guide

# 

Responsible data practices require risk assessment. This is a process that involves:[[3]](#footnote-3)

* Identifying capacities: valuable assets (e.g., contact lists, project documents, research data, interview notes, audiovisual files).
* Determining vulnerabilities: what threatens those assets?
* Assessing risk probability: assessing when and where the threats are likely to hit.
* Considering impacts: weighing the potential consequences likely for different staff and project participants.

Answering these questions not only provides a full picture of what hardware and information is at risk, it also helps a practitioner prioritize what’s most important. No practitioner wants to lose the work they’ve completed on project datasets or documents, for example, but they also cannot do their work without their raw data or interview notes.

When conducting a risk assessment, it may help to think of your environment in layers:

* Neighborhood: Do your neighbors share your concerns about safety? Are there ways you can help one another to make your homes or offices more secure?
* Outside the office: Can anyone walk into the office? Can people reach your Internet or phone equipment from a window? Is your office Internet access point visible to people immediately outside
* From the front door: From the front door of your home or office, can you see potential vulnerabilities? Are you sharing your project details or your ideas with visitors or people walking by the window? Could someone walking by have physical access to your network cables or to a PC?
* At your desk: Is your PC locked down with a cable or padlock, or can anyone walk off with it? Is it protected with a password?
* Your digital ‘space’: Are your devices protected with passwords? Do you have any policies or guidelines that you follow when sharing materials or communicating with others? How secure is your data storage?
* Your ‘human network’: Who do you trust? Who should have access to project documents and data collected on the participants and context—and who shouldn’t?

A practitioner can, with colleagues, and with project participants, brainstorm the possible and probable threats each faces, and consider how they might relate to the use of technology in the project – e.g., staff or participants’ mobile phones, computers, the internet, wifi, cloud storage, satellite communications, GPS, location tracking devices, drones. These threats can be organized in a matrix such as this:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Threats** | **Who?** | **Vulnerabilities** | **Capacities** | **Capacities required** |
|  |  |  |  |  |
|  |  |  |  |  |

Responsible data practices require safety planning. This identifies actions you can take to address the threats. Questions that may help formulate your plan include:

* What risks can be eliminated entirely and how?
* Which ones can be mitigated?
* Based on their likelihood and significance, which risks should be addressed first?
* How can those risks be mitigated?

It is assumed that practitioners and managers won’t be able to address all threats at once. They should be prepared to schedule work on project risk assessment and safety planning, alongside project design, implementation, and monitoring and evaluation activities.

Things to keep in mind:

* Be inclusive in your planning: A practitioner’s own or participants’ risks may depend on other peoples’ habits. Having confidential group discussions about organizational safety policies and practices is important.
* Be judicious with permissions and access to digital data, software and hardware: Does everyone in the office have access to all the data or devices in that office? Should they?

**Safety Planning Risk Matrix Example: Probability and Impact[[4]](#footnote-4)**

It may help in safety planning to first prioritize risks through considering their probability for different staff and project participants, and the severity of the likely impacts. Separate Risk Matrices could be completed to map out risk profiles related to different staff roles and to diverse project participants.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Probability** |  |  |  |  |
| **Very High** |  |  |  |  |
| **High** |  |  |  |  |
| **Medium** |  |  |  |  |
| **Low** |  |  |  |  |
| **Impact** | **Low** | **Medium** | **High** | **Catastrophic** |

# Case Studies

The following case studies are amalgamations of real projects that have been genericized to illustrate program risks in humanitarian and development project contexts, and offer suggested steps practitioners can take to prevent or respond to foreseen or unforeseen safety threats. The case studies are intended to spark ideas on how to apply responsible data practices of risk assessment and safety planning in humanitarian and development project design, implementation, monitoring and evaluation, and across the data lifecycle. The suggested safety planning tips are not exhaustive. Please refer to further resources referenced in this guide for more information.

Please note that there is no one-size-fits-all or 100% always-reliable safety plan for any given project context or participant. Each safety plan should be developed based upon careful project-specific risk assessment.

## Case Study 1: Biometric identification system for refugee registration

*Project overview*

A joint initiative of the UN, the local government and an NGO uses biometric identification systems to register refugeeswith the aim of better coordinated camp-based services and reduced inefficiencies. Refugee registration involves collecting biometric data from facial recognition and iris scanning software, and other personally identifiable data, including demographics, security threats, physical locations, needs in crisis, protection reports, cross-links with family members and associates, etc. Refugee registration officers and government staff working on registration routinely use their personal mobile phones and email accounts **to verify the backgrounds and identities of persons seeking refugee status**. The local context is characterized by armed unrest, organized crime, mass trafficking of young women and girls, and covert counter-terrorism operations by US military, security and intelligence contractors and national military known to surveil and target all young men as suspected “terrorists.”

|  |  |
| --- | --- |
| **Probable Risks** | **What could have been done better (safety planning)** |
| This open sharing of personally identifiable information over communications technologies that have weak and proprietary encryption (mobile telecommunications networks, Skype, Gchat, Yahoo email) increase risks to refugees as this information is being intercepted by government, military, intelligence and organized crime groups. | **Establish a strong verification process within more secure information and communications systems that are well encrypted** (this could become a workflow supported within a secure system storing the data). If you must send data over IM, voice, or SMS consider using, and getting colleagues to use the following technologies\* with proven encryption:   * Jitsi (more secure alternative to Skype) * OTR (using XMPP/Jabber), CryptoCat (more secure alternatives to Gchat) * Thunderbird with Enigmail (more secure alternative to Gmail and Yahoo) * Silent Circle and Silent Text (more secure services for voice and text)   \*It is important to note there will be a learning curve to consider when introducing new technology mid-project; training sessions along with application support and mentoring should be offered. |
| Many government staff use Yahoo! As their email accounts, over which they **regularly share personal details and identifying information** about women, girls, boys and men seeking refugee status throughout the determination process. | There should be **better policies and practices to securely share personal identifying information** (e.g. do not use commercial email accounts for any work-related, refugee verification process-related information or communications.) |
| Glitches in biometrics data collection technologies lead to time delays and deter local staff from using the technologies as planned. This may result in longer waits for persons to obtain IDs and access food and non-food aid and protection services. Rumors are that women, girls and boys are being trafficked for labor and sex out of the processing zone and camps, as they have nowhere else to go and have not been issued IDs to enable them to access aid and services. Many displaced persons are not registered via the planned information systems, or at all. | Either drop biometrics registration technologies, or invest in time and resources for upfront training and strong on-the-ground support of the registration process using biometric data to ensure processes are consistent and secure. Without this level of support, staff may find alternative ways of registering persons that are less secure and could increase risks to individuals. |
| Targeting of young men and women refugees of particular ethnic origins on the basis of facial recognition software algorithms and data. | Ensure that data capture, data storage and data sharing of information recorded by the biometric system is strongly encrypted and stored in a facility with good operational security protocols. As much as possible, do not plan to share personally identifiable data outside of this system.  If data security measures cannot be met, consider an alternative registration process that does not facilitate potential racial and ethnic discrimination and targeting for denial of status, services, and human rights violations. |

## Case Study 2: SMS for crowd-sourcing data in an emergency

*Project overview*

In an emergency setting, those affected use mobile phones to submit geo-tagged information about immediate needs and emergency incidents via SMS over private sector telecommunications networks to a cloud-based platform. Volunteers from all over the world access the platform freely to translate, parse and map the data. The overall goal is to map translated, geo-located data to help aid agencies, the Coast Guard and military first responders to find and assist survivors, and deliver targeted aid based on expressed needs. The country concerned is one of mass surveillance of the entire population by the national government and by international intelligence agencies with military and private security operations in the region.

Consent to post personal information in public maps is considered “implied” through use of the mobile phone texting service. Data collected include GPS coordinates, mobile phone account holder’s name, their address and other personally identifiable information, and sensitive information about vulnerabilities, needs and incidences of violence. Data are communicated via private telecommunications networks with no data security or data privacy policies. Mobile phone service providers routinely share customers’ information with government agencies and law enforcement. Rumors are that organized crime groups have hacked into this private sector telecommunications data.

The crowd-sourcing data platform remote managed from a university in a high-income country by PhD students. Volunteers off-site translating and processing texted end-user data hold regular Skype chats with government and military agencies requesting the data.

While the data hold the potential for geospatial locating and targeting of people in need for aid and assistance, only 2% of the data have been mapped so far and accuracy of the data remains a challenge. There has been a tendency for certain groups to engage in crowd-sourced reporting over others, potentially leaving out minority groups and creating an imbalance of emergency response services. When some mobile phone users from vulnerable groups do report needs or incidents, their sensitive data may place them at risk of further harm as their data are not adequately encrypted and protected.

|  |  |
| --- | --- |
| **Probable risks** | **What could be done better (safety planning)** |
| Data surveilled and hacked can lead to formal and informal actors targeting crisis survivors for economic, labor and sexual exploitation and abuse, or denial of emergency response services to politically oppressed or otherwise marginalized segments of national population. | Develop app for encrypted texts or run SMS service over SecureChat. Build capacity in-country among all staff and practitioners for SecureChat usability on Android phones. |
| Users of SMS service are un- or under-informed about the potential uses and risks to public posting of their reports and metadata | Provide informed consent information notifying end-users that they can opt out of their information being publicly posted |
| First responders and law enforcement may be unprepared to respond to the unique needs of survivors of gender-based violence | Provide response options to code any gender-based violence incidences to inform humanitarian responder planning and preparedness. Ensure that each survivor reporting gender-based violence holds the right of disclosure to authorities. Do not report to authorities if a survivor does not give consent to do so. |
| Lack of data security compromises the privacy, confidentiality and right of disclosure that belong to survivors of gender-based violence | Provide explicit option for whether survivors of gender-based violence wish their report to be shared with local law enforcement, or not. |
| Data reported concerning children, elderly or other vulnerable survivors can make them more vulnerable to targeting for exploitation and abuse | Provide an option noting whether the affected include children, elderly, functionally (physically or mentally) disabled persons, prisoners, or other vulnerable populations. |

## Case Study 3: Mobile application to report child abuse cases

*Project overview:*

This is a story of an NGO’s “landmark achievement in today’s world of technology” and example of ICT4D. Staff of a local field office developed a mobile-based application to map and report cases of violence against children in local language. The app is a GPS-based model run through an Android or Java-enabled mobile phone. The NGO plans to provide the first set of phones to volunteer children officers, paralegal personnel and community-based leaders. The NGO developed the infrastructure and prototype for the application. The app would be used to report any case of any type of violence against children to police, judiciary and the county level child protection officer, based on the nature of the case reported. When an alert is sent to the police/judiciary, the sender also receives a message on what should be done by the people present at the site of crime. For example, in the case of rape, the clothes should be wrapped in paper and not plastic, and so on. The system can integrate audio, photos and video. This project is funded from NGO child sponsorship funds and funding from a prominent bilateral and is a local office initiative with support from the national child protection specialist. The NGO considers the app a very innovative step towards systems strengthening.

A national Director of Children’s Services visited the local NGO office and worked with local staff to develop the app. On sharing his input, he was keen to receive a full presentation on design and implementation plans for the app. He was also willing to sign a Memorandum of Understanding with the NGO and scale the app up for the entire country. He recommended a national Child Helpline should be included to ensure a comprehensive system. The national Director of Children’s services, and national police chief attended the launch of the mobile-based app. The NGO considers the new project already a scalable best practice example of how the organization is contributing to the strengthening of the National Child Protection System. The project has not yet been evaluated.

|  |  |
| --- | --- |
| **Probable risks** | **What could be done better (safety planning)** |
| Children’s image, voice, surroundings, and testimonial reported with GPS coordinates via unsecured technologies and communications to unknown recipients that can be hacked by formal or informal actors targeting children for exploitation and abuses | Do not collect video testimony from children as it exposes their identities and increases their risks of further or new abuse. |
| Authorities may not cooperate and respond appropriately to reports of child abuse. | Coordination plan should detail ethics protocols for use of information, and all services available to child abuse survivors for their protection, legal, medical and psychosocial needs.  Further, coordination plan should include training and agreement on part of designated authorities staff to use secure communications for mobile phone voice calls, email and file-sharing. |
| Corrupt police or other authorities may try to destroy evidence of child abuse reports. | Backup and decentralize encrypted storage and archive reports on mirror servers outside of the country, or with the support of a human rights information documentation organization (e.g., www.huridocs.org). |
| Notes from verbal testimony typed into unsecured laptops and saved to unencrypted files, and transmitted to national child protection and law enforcement agencies via unsecured public wifi connections at internet cafes | Secure child’s caregiver informed consent (provided primary caregiver is not a perpetrator) and the child’s informed assent.[[5]](#footnote-5) Document child protection issues in writing, encrypt the file, and send using end-to-end encrypted email (PGP/GPG) or file-sharing software (OnionShare) to specific child protection and law enforcement officers tasked with organizing and supervising child abuse investigations, response and followup for medical, legal and psychosocial services as part of child protective services and justice response. |
| GPS metadata and location tracker features on mobile phones can tag children’s homes for targeting by formal or informal actors for exploitation and abuse. | Turn off location services tracking of mobile phone when conducting interviews with children. Do not videotape interviews. Hand write notes documenting child protection case then type notes saving them to an encrypted file and transmit file over encrypted email to the appropriate officers responsible for responding to child protection cases reported. |

# History of this document

This document originated during the Responsible Development Data event in New York on 22 May, 2014. A group of participants from diverse professional backgrounds started by listing out ethical questions and considerations for data management throughout the stages of a project lifecycle. The group wanted to share this list with a wider audience. A smaller team took on the project to turn that content into this introductory guide.

# About the Authors

**Jennifer Schulte** has 15 years of experience in human rights and programs-based research, monitoring and evaluation related to law, economic development, international development and humanitarian response programs, gender-based violence, technology and information ethics. She has designed and carried out research and provided monitoring and evaluation technical support to programs in nearly 50 countries across Sub-Saharan Africa, the Middle East and North Africa, South America and South Asia.Sheholds a Master of Science in the Sociology of Economic Development (Gender Research) from the London School of Economics and Political Science.

**Katherin Machalek** (bio pending)

**Norman Shamas** is an educator and technology for development practitioner. He is currently an online course facilitator at TechChange and wrangles and analyzes data for business development at Creative Associates International. Norman has extensive experience working as an archaeologist and leading dialogue groups in Israel and the West Bank. He speaks Hebrew, Persian and reads numerous dead languages.

**Shubha Bala** is a Master of International Affairs from Columbia University and has a Bachelor in Computer Engineering from the University of Toronto.  From Director of Overseas Sending at Engineers Without Borders Canada, to Impact Officer at the Media Development Investment Fund, she has worked in human rights and international development both in the field and in headquarters.  She was also an enterprise software consultant. Now she assists mission-driven organizations use data science to make positive impact in the world as the Programs Strategist with DataKind.

**Tania Lee** has over 12 years of experience in grassroots community organizing and professional software development project management. She is currently the Program Officer for ICT for Programs at the International Rescue Committee and implements technology projects for humanitarian aid and emergency response. She holds a Masters of International Affairs from The New School, is interested in improving aid accountability, and is (slowly) teaching herself how to code.

Special thanks to the following for their contributions to this guide:

Christopher Wilson, The Engine Room

Tin Geber, The Engine Room

Linda Raftree, Kurante

Emily Jacobi, Digital Democracy

Malavika Jayaram, Fellow at Berkman Center for Internet and Society at Harvard University, and Centre for Internet and Society in Bangalore, India

1. The examples detailed here have been compiled by The Engine Room (<http://www.theengineroom.org>)

   [↑](#footnote-ref-1)
2. See WHO “Informed Consent Form Templates” on informed consent; consent for storage and future use; informed consent for qualitative studies; informed assent for children/minors; informed parental consent for research involving children: http://www.who.int/rpc/research\_ethics/informed\_consent/en/ [↑](#footnote-ref-2)
3. Internews (2014) “Safer Journo: Digital Security Resources for Media Trainers,” <https://saferjourno.internews.org>. [↑](#footnote-ref-3)
4. Tactical Technology Collective and Frontline Defenders (2014) “Community Focus: Tools and Tactics for the LGBTI Community in Sub-Saharan Africa,” https://securityinabox.org/communities/02. [↑](#footnote-ref-4)
5. See WHO “Informed Consent Form Templates” on informed consent; consent for storage and future use; informed consent for qualitative studies; informed assent for children/minors; informed parental consent for research involving children: http://www.who.int/rpc/research\_ethics/informed\_consent/en/ [↑](#footnote-ref-5)