Considerations *for* Using Data Responsibly at USAID

USAID
FROM THE AMERICAN PEOPLE

fhi360
THE SCIENCE OF IMPROVING LIVES

mSTAR

Acknowledgements

This document was produced by the Development Informatics team at the U.S. Global Development Lab, with essential input from:

» Center for Digital Development: Anna Arnaudo (Former AAAS Fellow), Subhashini Chandrasekharan (Former AAAS Fellow), Craig Jolley, Rebecca Saxton-Fox, Vivian Ranson, Aubra Anthony, Amy Paul, and Megan Cagle

» M/CIO: Data Services Team and Albert Bullock

» USAID Privacy Council

» General Counsel: Megan Metcalf and Gayle Girod

» Bureau for Policy, Planning, and Learning: Elizabeth Roen, Jessica Pomerantz and Hoang Anh Lam-Vieira

Research Team + Advisory Committee

This document was based on research conducted under the Mobile Solutions Technical Assistance and Research (mSTAR) project, United States Agency for International Development Cooperative Agreement No. AID-OAA-A-12-0073. The content and views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

The research team consisted of: Siobhan Green (Sonjara, Inc.), Subhashini Chandrasekharan (Former AAAS Fellow, USAID), Claudia Schwegmann (Open Knowledge Foundation), Julie Cohen (Georgetown University), Clare Sullivan (Georgetown University), Linda Raftree, Abdul Bari Farahi (FHI 360), and Nina Getachew (FHI 360).

The research team would like to give special thanks to the members of the Advisory Committee who were a great resource to the project. These include Zara Rahman (The Engine Room), Joel Urbanowicz (Catholic Relief Services), Haneen Malallah (Oxfam America), Josh Woodard (FHI 360), and Norman Shamas (Digital Security/Privacy Expert).

Eric Bond/Elizabeth Glaser Pediatric AIDS Foundation

Imagine an HIV-positive mother and her newborn live in a rural area where HIV diagnostic services are not easily accessible. A community health worker (CHW) is their only tie to the formal health system. The CHW visits the family, taking a blood sample to test for mother-to-child transmission of HIV. The health worker sends the sample to a testing facility and then picks up the paper-based test results, which show that, sadly, the baby has the virus. It is critical for the child to start treatment quickly, but it may be several days before the health worker can return to that family. The family is left waiting for potentially life-saving information–time that may cost them the health or even life of the child.

Recognizing that faster transmission of information could have lifesaving results, the USAID program manager on a regional maternal and child health project wants to digitize this process. An implementing partner has suggested sending the test results by a simple text message to the family and the treatment center to increase the likelihood of initiating treatment on time.

This could be an ideal opportunity for digital technologies to enable USAID's programming to have even greater impact. But the premise is not so clear cut–it is also possible that increased access to information could inadvertently create complicated or even harmful outcomes. What if someone else in the extended family or community also has access to that phone? What if the mother's extended family does not know of her HIV status? Could disclosure of that information harm the woman or her baby?

As we increasingly digitize our programs and activities, enormous opportunities emerge for time savings, cost savings, and even saving lives. However, digital tools can introduce the potential for harm if used without appropriate attention to critical issues like privacy, security, or the many unique ways people interact with digital technologies across cultural, social, and gender lines. Many of the risks associated with data use may not come from what many of us envision as traditional threats, such as hackers or intentionally nefarious actors. Ultimately, we must acknowledge that some of the greatest risks may come from the false sense that issues of privacy and security exist solely in the realms of checklists and compliance. We must address the risk posed by even well-intentioned actors if we do not appreciate how to responsibly leverage the powerful tools we wield.

Table of Contents

Introduction

Data and digital technologies are changing the way that international development programs are implemented. For years, donors, implementers, and host-country governments have turned toward evidence-driven programming to maximize the impact of development efforts.

These exciting developments have also brought new tensions. A push for greater openness—epitomized by data sharing requirements within USAID's Development Data Policy[1]—promises to make evidence and information available to a wider audience than ever before. At the same time, such data sharing has reignited debates about data ownership, privacy, and informed consent. In both developed and developing countries, high-profile privacy incidents have eroded public trust in the ability of governments and private companies to keep data secure. All of this happens against a backdrop of increasing private-sector involvement in development, where differing norms and cultures of data use can come into conflict.

This document aims to provide USAID staff and local partners with a framework for identifying and understanding risks associated with development data. It is meant as a conversation starter—to highlight important concerns and provide actionable advice—to help those who use data in development programs maximize utility while also managing risk. By starting to have conversations around responsible data practices, staff and partners will begin to build competency in this area. USAID's Journey to Self Reliance includes supporting countries to build their own technological capacity and readiness by taking ownership of their data and being held accountable that it is kept safe.

Purpose

The primary goal of this document is to help all USAID staff and partners have better conversations about data—more specifically, how to balance the tremendous opportunity presented by data with the associated risks. The goal is to increase the capacity of staff and local partners to implement responsible data practices. In addition, it includes references to useful resources, such as official USAID policy guidance about privacy, open data, and data quality. Lastly, this document is intended to supplement overall understanding of the "Address Privacy and Security" aspect of the Principles for Digital Development for the broader digital development community.[2]

This document is not a compliance checklist or a comprehensive data management guide. There is no list of tasks that must be done in order to "be responsible." While this document aims to provide context and thought-provoking questions, it cannot offer all the answers; this will depend on the unique circumstances of your work. The goal of this report is to promote thoughtful conversations rather than establish rigid policies.

It is also important to note that this document does not present official USAID policy and should not be taken as legal guidance. The authoritative source for USAID's operational policy is the Automated Directives System (ADS), and your source for legal advice should be either the Office of General Counsel (GC) (in Washington) or your Resident Legal Officer (RLO) (in Missions). In addition, the Office of the Chief Information Officer (M/CIO) is a valuable resource for issues around information management, security, and open data compliance.

How to Use This Document

Each section of this document provides resources for specific questions that may arise during your work using data. There are also tools and tips for how to help guide discussions or navigate areas of responsible data practice that may be unclear or a point of tension. Tools are provided to help you think through hard questions—not necessarily to answer those questions.

1 ADS Chapter 579: USAID Development Data.
2 Principles for Digital Development: https://digitalprinciples.org/.

Responsible Data Overview

What Does It Mean to Use Data Responsibly?

The term "responsible data" can be used to describe a number of specific practices in data collection, management, and analysis. However, using data responsibly in development programs ultimately requires balancing three broad thematic areas:

» ***Data use*** — Data can be used to maximize the efficiency and effectiveness of programs and activities, with the goal of improving outcomes. They can help us target interventions to the people or communities who will benefit the most. When shared with others, data can help us to build support and consensus by communicating challenges and possible solutions. Some programs share data directly with the people we serve, helping them to make more informed decisions about their own health or livelihoods. Intended use should guide our decisions about data collection and management in order to avoid needless expense or risk.

» ***Privacy and security*** — Data carry substantial risk, both for data subjects and for data stewards (e.g., implementers and donors). We are responsible for taking appropriate measures to minimize the risk to individuals based on data that is collected.

» ***Transparency and accountability*** — We are also responsible for sharing information with the people affected by our projects, host-country governments, and the U.S. taxpayers who fund our work. We need to be transparent about our programs and whether they are effective. We should also use data to improve the planning, effectiveness, and efficiency of future programs and activities. We should make data openly available for the public good. We should encourage the application of existing data to new purposes with the goal of achieving breakthroughs and insights.

These three areas are frequently in tension with one another (Figure 1). Understanding these tensions and working to balance them can help us work responsibly and highlight questions about risk and benefit surrounding our data. For example, a single-minded focus on data use might lead to over-collection of sensitive data, leading to potential privacy risks. On the other hand, if we prioritize privacy and security above all else, then we might delete data soon after collection. If done carelessly, this could compromise the records retention necessary for transparency and accountability.

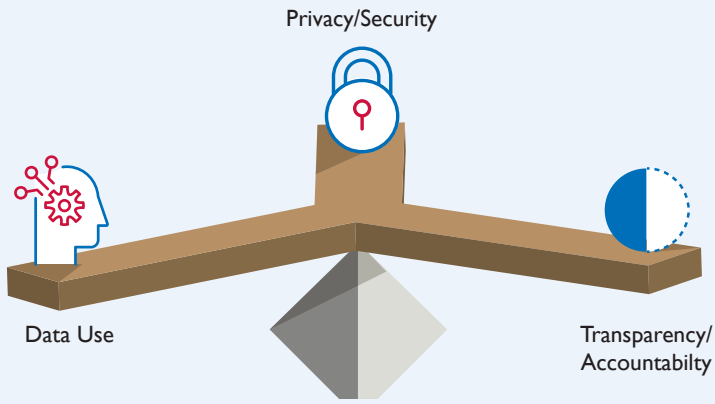


Figure 1. Responsible data practices balance key tensions

There is no one-size-fits-all solution to balancing these tensions. Extremely sensitive or unstable contexts may require prioritizing security. Outside scrutiny may lead to an increased need for transparency. All three areas of concern consume resources in projects with limited time, funding, or human capacity. Projects may need to engage in "data triage" by meeting their most urgent responsibilities first.

Our Responsibilities

As your project team talks through your data management plans, it may be helpful to ask the question: "To whom are we responsible?" One way to think about using data responsibly is in terms of three primary groups, with distinct needs and requirements, to whom we must commit certain responsibilities:

To Data Subjects

Data subjects are the people from whom data are collected. In most development projects, these people are part of the communities that are affected by our work. We are responsible for:

- **Protecting privacy.**
- **Respecting the agency** of our data subjects by insisting on informed consent, allowing for correction of data, and seeking redress for any data-related harms.
- **Improving our interventions.** This includes both the direct use of project data for adaptive management and the use of evaluation results to inform future projects and shape country-level strategies.
- Using the data we collect to **promote social equity** by seeking a better understanding of disparities and development gaps within the communities where we work.

To Ourselves

We define "ourselves" for the purposes of this paper as the people who are handling data: donors, implementing partners, and host-country governments. The project team will work within institutional structures, possibly including government agencies, private companies, and academic institutions. We are responsible for:

- Meeting our **legal and ethical commitments.**
- Managing **reputational risk** (privacy breaches, etc.).
- Avoiding **physical risk** to our staff, partners, and data subjects.
- Ensuring that **adequate funding** is available for data collection, analysis, use, and curation.
- **Being good resource stewards** through accountability, transparency, and using data to improve our programs.

To the Broader Development Community

Beyond our own institutions, we work as part of a broader development community that may have an interest in the data that we collect and analyze. We are responsible for:

- **Making our data usable** by complying with data formatting standards, following best practices for clearly documenting our collection and analysis methods, and recording metadata.
- **Documenting the limitations of the data collected,** including omissions and biases.
- **Creating public goods**. The highest-value datasets to share will be the ones that help others learn about the effectiveness of our interventions or plan their own development projects. One way to think about this is to imagine the data sources you *wish* you had when designing a project. If you end up creating a similar data source during the course of a project, then someone else might find it useful. You may also reduce the likelihood of duplication of data collection, datasets, and data sets.

Each of these responsibilities are further explained in the next section of the document.

Responsible Data Considerations

This section will walk through a set of recommendations and tools for the responsible use of data in development projects.

Data Policy and Planning

Here, we are concerned with the establishment and implementation of Mission-level data processes and the planning of data-enabled programming. Establishing clear data processes is a way to maintain consistency by ensuring that some issues—such as compliance with applicable laws and USAID operational policies—are handled the same way for all interventions.

Legal and Policy Issues

All USAID interventions need to comply with relevant laws and operational policies. USAID staff are bound by the laws of the United States and should expect implementers to follow the laws of the countries in which they operate. There may also be numerous internal operational policies or standard operating procedures, both from USAID and implementing partners, that require adherence. Understanding which legal and policy regulations (including which sections of the ADS) are relevant for a particular intervention can be a challenge. The following table can be used as a starting point.

Assessment of Existing Legal and Policy Regulations

	Consideration	Resource to Address Consideration
USAID[3]	Will your data collection support a USAID deliverable? If so, you may need to take additional steps to curate your data.	See USAID Development Data Policy (ADS 579).
	Is this activity considered human subjects research? If so, you may need extra protections on your data.	See Protection of Human Subjects in Research Supported by USAID (ADS 200mbe).
	Will your data collection and analysis support USAID Program Cycle processes, such as country strategic planning, project or activity design and implementation, or monitoring and evaluation?	See USAID's Program Cycle Operational Policy (ADS 201) for requirements and expectations around data-informed planning and decision-making throughout the Program Cycle.

3 Implementing partners should consult their A/CORs for help with ADS documents. Questions from the A/COR should be directed to the responsible office of the relevant ADS functional series (https://www.usaid.gov/who-we-are/agency-policy/about-ads). General ADS questions can be sent to ads@usaid.gov.

	Consideration	Resource to Address Consideration
	Will you be reporting performance monitoring data externally (i.e., via the annual Performance Plan and Report)? If so, USAID must conduct a Data Quality Assessment (DQA).	For DQA applicability, see USAID Program Cycle Operational Policy (ADS 201), especially section 201.3.5.8. For DQA requirements, see USAID Recommended Data Quality Assessment Checklist (ADS 201sae) and DQA How-To Note.
	Will your data collection happen within a U.S. Government institution?	See USAID Risk Assessment Guidelines (ADS 545may) and Protection of Human Subjects in Research Supported by USAID (ADS 200mbe).
	Will your data collection involve U.S. citizens?	See USAID Privacy Policy Program (ADS 508), Paperwork Reduction Act.
National (host-country governments)	Which host-country laws or policies govern (and are applicable to) the collection, use, and sharing of personal data? Relevant policies may be general or sector-specific (especially in finance and health). How will you translate these laws or policies into specific guidelines?	Country Specific—Consult your RLO for guidance on local laws as well as bilateral agreements, data sovereignty issues, etc.
International and Broader U.S. Government	Does the planned data processing potentially concern European Union (EU) citizens or the transfer of personal data from an EU country to the U.S.?	See EU General Data Protection Regulation (GDPR).[4]
	Does the data include financial transactions?	See PCI compliance.
	Does the intervention create any potential right-to-privacy issues?	See ethical standards on rights to privacy by the UN Commission on Human Rights.
	Does your intervention have any access to information or transparent government requirements?	See Sustainable Development Goals, Development Effectiveness Consensus, and Open Government Partnerships.

4 USAID and the USG have not yet established an official position regarding the GDPR. You should consult with GC regarding if and how to factor GDPR into data planning and management.

Sometimes, the demands of USAID policies and local laws create perceived conflicts for implementers, who must be able to comply with the terms of their agreement with USAID, while also respecting host-country governments. For example, a tuberculosis program in an African country worked to integrate digital systems into clinics and the Ministry of Health, hoping to enable more timely sharing of test results. Data protection and privacy are taken very seriously in this country's health sector, and the National TB Program (NTP) routinely denied requests to share information with donors. At the same time, implementers felt that USAID and other donors expected on-demand access to aggregated data in order to meet their reporting and monitoring requirements. In the absence of a written agreement between USAID and the NTP, the implementer felt caught in the middle, resulting in frustration on all sides.

One of the key findings of the research supporting this document is that issues of *data sovereignty* are often highlighted as a barrier to responsible data use. In the context of government-operated systems (such as the tuberculosis program described above), national data sovereignty refers to a government's exclusive authority and control over virtual public assets.[5] Data sovereignty can be complicated significantly when government data are hosted on cloud services, or when international donors are involved in data creation.

When a donor-funded project supports the development of official government systems, overlapping claims of data ownership may result in conflict and confusion. These problems can sometimes be averted by avoiding assumptions about which data can be shared. Any existing data sharing agreements between the U.S. Government and the local host-country government require compliance. For assistance addressing this challenge, reference Tool 1, Data Ownership and Data Sovereignty on the next page.

Kashish Das Shrestha/USAID

5 Irion, Kristina (2013). Government Cloud Computing and National Data Sovereignty. *Policy and Internet, 4(3-4)*, 40-71.

TOOL

1 Data Ownership and Data Sovereignty

How will conflicting claims be managed? This flowchart is designed to help think about data ownership and sovereignty in the context of an intervention. Who owns the data, which groups can access them and whose laws regulate them? This worksheet can be used as a basis for conversation with your RLO, GC, or other Agency legal representative.[6]

Who can claim data sovereignty?

Data sovereignty defines which country's laws apply to data during processing. For example, if data are processed within the borders of a country, generally that country's laws are paramount, regardless of who has done the processing, for what reason, or by whom the data were funded.

List Claimants

Who can claim data ownership?

Data ownership considers who has final, legal authority over access and use of the data. Data funded by USAID is de facto owned by the organization that collects them, but the USG retains an unrestricted right to access and use those data.

List Claimants

Who needs to be informed of conflicting claims?

The **stakeholders and decision makers** for conflicting claims on data depend in part on the country context, nature of the data, and other sensitivities. For example, in a country with strong data sovereignty laws or where data on vulnerable populations is involved, it may be necessary to involve higher levels of Mission leadership.

List Stakeholders

Who is responsible for prioritizing?

What formal documents are being written?

Capture in **formal documents** any decisions already made on resolving data conflicts. Also review existing contracts or the memorandum of understanding (MOU) to see if data conflicts are already identified.

List Documents

6 For full definitions, see Annex I.

Planning for Data Use

It is important to be clear during the activity design stage how data will be collected, transmitted, stored, analyzed, and curated and how all interventions will be tracked and documented. It is also necessary to:

» **Outline** the roles of different actors.

» **Estimate** what resources (human, financial, and technical) are necessary for data-related tasks and ensure that they are available. Resource allocation needs to take into account what is needed to collect, process, analyze, and use data throughout and after the intervention to ensure sustainability.

» **Plan** for how events such as staff departure will impact data use (see Tool 2, Key Events Planning Table, page 10).

» **Investigate** the benefits and risks associated with data collection and how to mitigate risk (see Tool 3, Benefit Risk Assessment, page 11).

These factors are typically documented in the Mission-wide Performance Management Plan (for strategic planning and implementation across a country strategy)[7], Project Monitoring, Evaluation and Learning Plans (PMELPs), and Activity Monitoring, Evaluation and Learning Plans (AMELPs).[8]

One best practice at the activity planning level is to develop a Data Management Plan (DMP) to outline the resources and data needs discussed above in a greater level of detail than what is already required in the AMELP. The DMP can be a section of (or annex to) the AMELP. A DMP should:

» **Be grounded** in the activity's theory of change and AMELP.

» **Identify** data needs related to the following:

- achieving desired outputs and outcomes,
- monitoring an activity's performance against results and adapting as necessary, and
- evaluating outcomes and impacts.

» **Identify** if third-party data sources are available or necessary (e.g., demographic information, household surveys, and geospatial information).

» **Describe** how activity managers and implementers will store, manage, process, analyze, and document these data throughout the activity.

» **Describe** plans for curation.

» **Identify** which data can or cannot be released publicly due to privacy concerns.

» **Describe** costs and benefits of collecting and using these data, considering the administrative and beneficiary burden.

Data management planning can serve as a useful tool to help USAID staff anticipate and manage concerns about security, privacy, and data use that are prominent throughout the Program Cycle. AMELPs and accompanying DMPs are a valuable tool and there are many resources available to develop them (see chart page 9). USAID Mission staff and partners might ask their Agreement or Contracting Officer's Representative (A/COR) about including data management planning as a part of the work plan process.

7 USAID Learning Lab (2017). How-To Note: Prepare and Maintain a Performance Management Plan (PMP).
8 USAID Learning Lab (2016). How-To Note: Activity Monitoring, Evaluation, and Learning Plan.

Consideration	Resource to Address Consideration
How will you develop your AMELP?	See How-to Note: Activity MEL Plan.
Which data needs are implied by your activity's theory of change?	Key People: MEL points of contact, AOR/CORs, and Communities of Practice.
How will you assess benefits and risks associated with your data and how will you mitigate risk?	See Tool 3, Benefit Risk Assesment, page 11. See Developing a Risk Management Tool.
Are you applying standard data model structures to your intervention?	There are numerous resources including: USAID Performance Indicator Reference Sheet, International Aid Transparency Initiative, and Demographic Health Surveys.
Have you identified all potential data sources? Can you obtain publicly available data or data from other stakeholders?	Data banks including: World Bank Open Data, UN Data, and The DHS Program: Demographic and Health Surveys. USAID's EADS is a central source for all of these data banks.
Are you designing your data collection to consider sharing and reuse?	FAIR is a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. See FAIR Principles.

Even the best plans will be constrained by some external factors. For example, resource limitations may require you to collect, analyze, or use smaller than optimal amounts of data. Development projects are typically targeted at a particular location or demographic; this means that data collected will not be representative of the broader population and it would be unwise to use such limited data to make inferences about other locations or groups of people. Most importantly, not all data needs can be anticipated in advance. A flexible DMP will leave space to adapt data collection or analysis to emerging needs.

Afandi Djauhari/ NetHope

TOOL

2 Key Events Planning Table

This chart is meant to help think through the key events that could happen during an intervention and plan how to responsibly handle data during those events. Some examples are given to help start.

Key Event	Relevant Procedures	Resources Allocated
Staff Training	-Informed Consent Training -Handling PII Training -Archival, Retention, and Disposal Training	-Time -Budget
IT Transitions (Maintenance or Device Retirement)	-Track Sensitive Information (see worksheet page 17)	-Time -Trained Staff

Key Consideration: Staff Autonomy and Resources

One essential consideration when thinking about data in intervention planning is: Do the people who will need to use these data have the resources and autonomy they need? Considerations could include:

- **Authority:** Do they have clear roles and responsibilities?
- **Incentives:** Are there proper incentives for them to perform their roles?
- **Capacity:** Do they have the skills and knowledge required?
- **Resources:** Do they have the proper resources such as time and money?

Key Consideration: Bias and Existing Inequalities

Given the context-specific nature of international development, an important thing to consider is:

What are the context-specific biases and existing inequalities which may impact how you collect, analyze, use, and share your data? Are there ways to protect against these biases? Examples of these kinds of assessments can be found in Reflecting the Past, Shaping the Future: Making AI Work for International Development.

TOOL

3 Benefit Risk Assessment

This tool[9] is designed to help assess potential benefits and risks of data collection, use, and sharing. A key to properly assessing risks and benefits is including relevant stakeholders in this process, including those from whom you are collecting data. Please note, a separate risk assessment may be necessary for data submission to a digital repository.

Complete the tables below to assess the benefits and risks of collecting, using, and storing each data item required for your activity.

BENEFITS

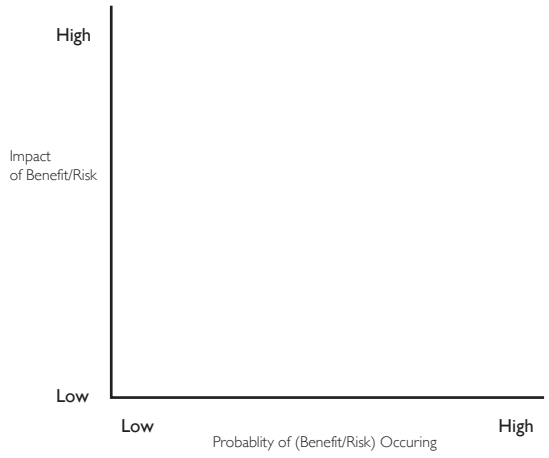
Data item	Who benefits from conducting this exercise in data collection?	How do they benefit?	Likelihood of the benefit being achieved *(Low/Medium/High)*	Positive impact of the benefit to the beneficiary *(Low/Medium/High)*

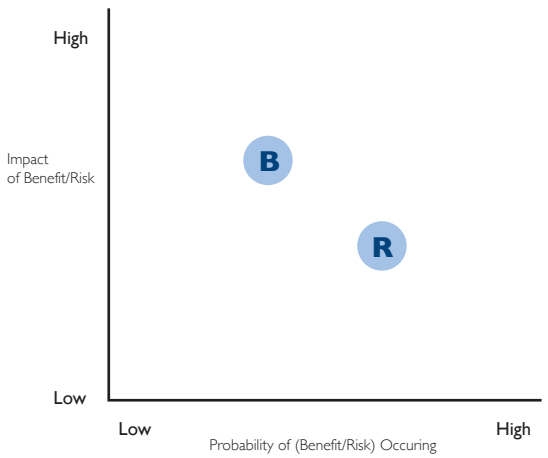
9 Adapted from: Benefit-Risk Analysis for Big Data Projects.

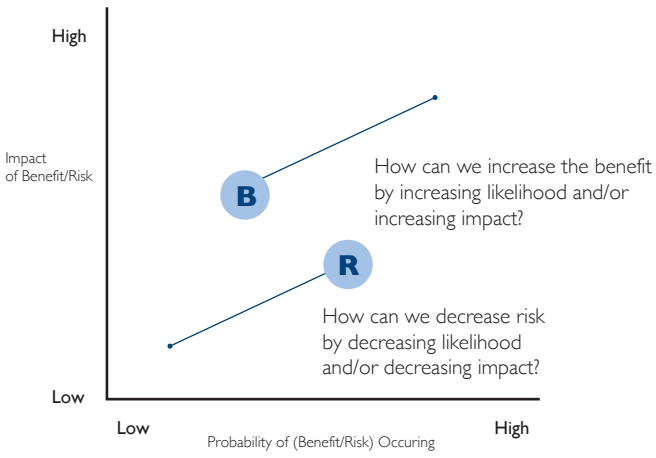
RISKS

Data item	What are the risks of collecting/ holding/sharing the data?	Who causes the risk?	Who bears the risk?	Likelihood of risk happening *(Low/Medium/High)*

Plot each item on a graph as demonstrated below, where the x-axis is probability and the y-axis is impact. Use these charts to start a conversation about the risks versus benefits of each data item using the discussion questions below the charts.







Discussion Questions

- If both the potential risk and benefit are high, how can we mitigate or lower the risk so that we can achieve the benefit? Do we have the capacity and the resources to do so? Do we need to go back to the drawing board and adjust the data plan?
- Are the people on the team capable of fully assessing how to mitigate the risks? Do we need additional consultations or technical expertise?
- What do the intervention participants/beneficiaries and local organizations say about the risks and benefits? Have we asked them? Should we?
- How will we ensure that we are managing the risks? When should we halt the data exercise? What will (or should) we do if we find a risk we had not anticipated or if someone has been harmed by our data practices? What do we do in case of a breach (legally, ethically, etc.)?

For each medium- or high-risk item from the chart above, complete the following:

Data item	What risk minimization or precautions must and should we take to mitigate the risk? (Think about collection, storage, sharing, and maintaining the data).	What impact would the mitigation actions have on the project? (Budget, training, software, data collection procedures, etc.)	How will we monitor/measure/manage the risk? How will we respond in the case of a breach or new risks?

Data Collection and Protection

Informed consent is usually obtained from intervention participants during data collection and often when data are being handled by a larger team of data collectors. Deploying strong data-handling protocols at this stage helps prepare for privacy and security challenges later in the process. Data are collected throughout a project for monitoring purposes, but should also be analyzed and used for adaptive management. In adaptive management, data are used to support experimentation and learning, increase knowledge, and select improved courses of action.

Informed Consent

Any effort to protect the rights of data subjects depends on meaningful informed consent. Data subjects need a clear understanding of what data are being collected about them, who will have access, how those data are to be used, and how long they will be kept. It is also critical that data subjects have a genuine choice whether to offer up their data. In an extreme example, asking hungry people to trade sensitive personal information for food is manipulative and unethical.

Informed consent must also take into consideration the cultural context and social norms in which data are collected. In some social situations, people may not feel comfortable admitting that they do not understand what they are being told. Power differentials associated with age, gender, or social status may lead people to feign understanding rather than appear ignorant. In other cases, people may fear appearing hostile, uncooperative, or ungrateful if they refuse to share their data. There may be special considerations for obtaining informed consent for vulnerable populations, such as sexual violence survivors.[10] People in some cultures are uninhibited about asking and answering highly personal questions, while others are more guarded. Within a single culture, the privacy of all people may not be valued equally, while some (especially women) may be shielded from public view (see Key Consideration: Informed Consent).

Non-traditional data sources can also present challenges for informed consent. Some projects may collect personal information from ground-based sensors (e.g., monitoring the usage of cookstoves or latrines), social media, or mobile phone metadata. The use of "big data" does not circumvent the need for informed consent, and organizations that use personal data without consent may expose themselves to legal risk.[11]

Key Consideration: Informed Consent

Imagine you are an implementing partner obtaining informed consent from sex workers for data collection. Sex work is illegal in the country in which you work, so obtaining written informed consent would cause the sex workers to incriminate themselves. To prevent this, informed consent could be obtained through oral communication.

Imagine you are a program manager working on a project promoting human rights and access to justice for indigenous populations. There are challenges with informed consent due to language, low literacy, and general fear associated with signing a legal-looking document. To overcome these challenges, informed consent language could be shortened and translated into a local or native language. The process could also be altered to allow the interviewer to sign as a witness to informed consent.[12]

As the examples above illustrate, obtaining informed consent is not always a clear-cut process and is largely dependent on the context of your project. Regardless of context, however, the overarching question is: what is your process for obtaining informed consent and what does it include? To help you address this question, discuss the following:

- How will social and cultural contexts impact informed consent needs?
- How will project staff be trained in culturally appropriate consent practices?
- How will you document and track the granting of consent (especially in low-literacy settings)?
- How will you obtain consent for data sharing, reuse, or changes in the purpose of the data's use?
- Will all types of data be included in informed consent processes including photos, video, audio recording, and geospatial data? If not, how will the program specify what data is included and how will it follow up should additional data be required?

10 USAID (2015). Sample Consent Form for Children and Other Vulnerable Survivors.
11 McDonald, Sean (2016). Ebola: A Big Data Disaster.
12 Case study on indigenous populations adapted from: Aguila, Emma, et al. (2016). Culturally Competent Informed-Consent Process to Evaluate a Social Policy for Older Persons With Low Literacy. Sage Open, 6, 1-11.

Key Consideration: Autonomy of Data Subjects

An increasing area of importance is the rights of data subjects in regards to controlling data collected about them. A question to consider is: if data subjects have concerns about their data or how it is being used, what are their rights? To help address this question, discuss the following:

- Are you informing data subjects about their rights of redress?
- Can data subjects retroactively remove data or correct/update data? What are the processes in place for doing so?
- Do any laws concerning the "right to be forgotten"[13] apply to them (i.e., citizens of the EU, Argentina, Russia, etc.)?

Sensitive Information

The data collected by many development programs are considered sensitive, requiring a host of additional considerations for privacy, informed consent, and use. Sensitive data are any data points that could cause harm if they are improperly disclosed. This is especially true in development programs and when working with vulnerable populations. These kinds of data could include personally-identifiable information (PII), such as names and addresses, or information about personal attributes, such as ethnicity, religion, political views, language, sexual orientation, or HIV status. Identifiers of personal information can lead to the individual (direct identifiers) while others can be used in conjunction with other data to identify the person (indirect identifiers). It is important to note that indirect identifiers can be more difficult to classify because it may not be immediately clear how the data can lead to identification; potential indirect identifiers can sometimes require additional consideration to properly address any possible security or privacy concerns. Not all sensitive data are personally-identifiable. Organizations such as the Humanitarian Data Exchange also identify categories of Demographically-Identifiable Information (DII).[14]

Data that could be used to infer demographic characteristics of a person, even if their individual identity is concealed, would be considered DII. Examples of DII might include native language, birthplace, or religion. Sometimes seemingly innocuous data can be used to reconstruct DII, as in a recent study that claimed to identify Muslim taxi drivers based on their inactivity at designated prayer times.[15]

A good approach to sensitive information collection is to apply "Lean Data" principles.[16] The key features of Lean Data are an emphasis on using data for value creation (rather than merely reporting) and an embrace of data collection methods and technologies that favor efficiency and speed. In general, you should collect the minimum possible amount of sensitive information, limit the extent to which these data are copied or moved, and delete them once you no longer need them. Lean Data principles also help minimize the burden on data subjects. People in need of assistance may find it intrusive to fill out lengthy forms or surveys asking highly-personal questions. We should not assume that people forfeit their rights to and desire for privacy whenever they receive assistance from one of our programs. To apply Lean Data principles, you should consider ways to avoid or minimize the collection of sensitive and/or identifiable information—for example, collecting only the year of birth instead of full birth dates. Considering the outcomes of data collection from similar projects may also be of assistance here. If a specific type of data proved fruitless in the past, then maybe this time you can skip collecting that data. Similarly, if previously available data can serve as a proxy for newly collected data, then its use could reduce the collection burden for both the program and participants.

13 Chima, Raman Jit Singh (2016). Access Now Position Paper: Understanding the "Right to be Forgotten" Globally
14 https://data.humdata.org/about/terms
15 Berlee, Anna (2015). Using NYC Taxi Data to Identify Muslim Taxi Drivers.
16 The Lean Data methodology was developed by Acumen, see: https://acumen.org/lean-data/. For a more extensive description, see Dichter, Sasha, et al. (2016). The Power of Lean Data. Stanford Social Innovation Review.

Key Consideration: Sensitive Data

How are you handling your sensitive data and PII? Consider:

- Limiting Collection
- Encryption
- Tracking of PII
- Staff Training

See USAID Privacy Policy Program (ADS 508)

Note that this is only applicable to U.S. data subjects but can be used as a reference.

See USAID Privacy Basics

While Lean Data principles sound straightforward in theory, they can be more complex in practice. For example, if data are being collected in order to build a predictive model, you may not know in advance which attributes will be useful in predicting your outcome of interest. If you "under-collect," your data might not be rich enough to support your project's goals. The use of repurposed third-party data can also be problematic, especially if you don't know the circumstances of its collection or what circumstances might make some attributes sensitive. Third-party data should, when possible, be reviewed for PII, sensitive information, and indirect identifiers.

In addition to applying Lean Data principles, applying disclosure limitation methods or controls can sometimes mitigate the risk associated with sensitive data. Data managers can mitigate risk in a variety of ways, including de-identification, pseudonymization, and/or aggregation. In many cases, however, data from which PII have been removed can be "re-identified" by combining them with other data sources. In one famous example,[17] Netflix released hundreds of thousands of "anonymized" user ratings and offered a USD $1 million prize to researchers who could significantly improve their movie recommendation algorithm. Names and other identifying information had been removed to protect the privacy of Netflix users. Within two weeks, however, researchers had cross-referenced the Netflix data with non-anonymous reviews on IMDB.com. Although PII had been removed from the records released by Netflix, a person's pattern of reviewing obscure movies functioned as a "fingerprint" that was unique enough to match their records in the two databases. While no harm was done in most cases, disclosing a person's movie-viewing habits could lead others to infer political views or lifestyle preferences they would rather keep private.

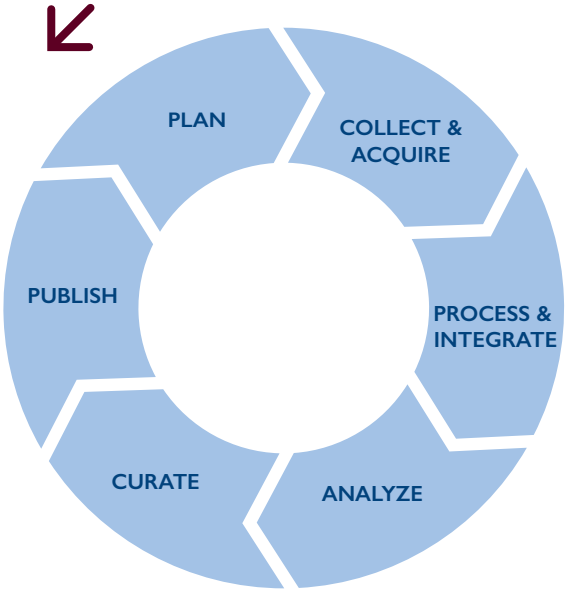
For the vulnerable populations served by USAID and implementing partner programs, the stakes of having their personal data re-identified could be significantly higher. When thinking through how to best protect PII in interventions, it is necessary to think about how to prevent re-identification. As illustrated by the example above, the impact of available qualitative data or third-party data on re-identification needs to be taken into consideration. Best practice is to identify which types of data could act as indirect identifiers and train staff on how to minimize their unnecessary collection.

17 Lubarsky, Boris (2017). Re-Identification of 'Anonymized' Data. *Georgetown Law Technology Review 202.*

TOOL

4 Tracking and Protecting Sensitive Information

Data managers use a data life cycle to help design for data and think about the flow of data throughout interventions. To protect sensitive information, it is necessary to catalog and track all copies of sensitive data throughout the data life cycle. This worksheet is meant to help track copies of sensitive data. Example questions to consider at each stage of the data life cycle are given as a starting point.



		Notes
Plan	Have you planned for how to handle sensitive data during key events and intervention closeout? (See Tool 2, Key Events Planning Table, page 10).	
Collect & Acquire	Which devices will be used to collect sensitive information? How will these devices be secured from theft or unauthorized access? Devices can include personal equipment, laptops, flash drives, etc.	
Process & Integrate	If staff are using personal equipment for data collection, how are data removed post-integration?	
Analyze	How will those analyzing data store copies of the data? How will devices used be secured?	
Curate	What protections will be used to de-identify data?	
Publish & Share	When sharing data, how will data be transferred? Examples can include shared drives, flash drives, etc. What processes are in place to mitigate risk during transfer?	

Key Consideration: Data Aggregation

Data aggregation is a common way to protect privacy and security. Rather than reporting results for each individual, aggregated data report summary statistics, such as values averaged over a population. Aggregated data are less detailed, and may be harder to analyze or reuse. Aggregation typically cannot be undone. When people refer to "disaggregated data," they are typically describing incomplete aggregation. In an education program, one might report individual test scores for each student (no aggregation), or a single average score for the entire school (complete aggregation). One could also report "disaggregated" results by calculating averages for categories such as gender, ethnic subgroup, or class year. While such data can be more analytically useful, they may expose students and graduates to privacy risks. As an extreme example, if only one female student took a particular class, reporting gender-aggregated scores would reveal her personal score while preserving her male classmates' anonymity.

When thinking about aggregation, some points to discuss are:

- Does aggregation allow for sufficient protection of PII (or other sensitive information, i.e. DII)?
- Does aggregation of the data skew the data and potentially impact transparency/accountability? For example, if survey data are aggregated by taking a sample mean, the presence of outlier responses may impact the results.
- Will data disaggregation required for data analysis create new privacy risks? How can they be mitigated? For example, could disaggregating data based on ethnic subgroup lead to perpetuating inequalities?

Afandi Djauhari/NetHope

Key Consideration: Protecting Vulnerable and Marginalized Populations

Imagine you are working on an intervention that focuses on preventing mother-to-child transmission of HIV. At a press conference to announce the positive results of your intervention, photos of the HIV-positive mothers were taken. These images are ultimately shown on national TV—exposing the identity of the HIV-positive mothers. Women revealing their HIV-positive status has resulted in abandonment by spouses, denial of access to medication, and increased incidence of intimate partner violence, and could be a risk for the women who participated in your press conference.

This example illustrates how important it is to consider the sensitive information of vulnerable or marginalized populations. To help protect the sensitive information of your beneficiaries, discuss the following:

- Identify any vulnerable or marginalized groups involved in your intervention. This could include: mentally disabled individuals, geographically isolated groups, those engaged in stigmatized and/or illegal activities, etc.
- Which attributes or pieces of data might be considered sensitive in the context of your intervention?
- Are there any risks of inadvertently "outing" vulnerable populations? This could happen from public notification of the intervention, participant selection processes, collection methods, or other events associated with your intervention. If this is a concern, what are you doing to mitigate risks?

IT Security

When we assure data subjects that we will keep their data private, we are committing ourselves to robust information security. This includes guarding IT systems against outside attackers, but also against misuse by project staff or by those who might inherit a database after project close-out. Security professionals sometimes refer to "successor attacks"—a scenario in which hypothetical future users of a system are corrupt or unscrupulous. Even if you trust your partners and staff, you should make it difficult for future users to misbehave.

Be sure that your IT security practices conform to USAID operational policies (as described in the "Legal and Policy Issues" section of this document on page 4). IT security encompasses both digital access controls and the physical security of facilities where computers and storage media are kept. While risk can never be eliminated completely, the goal of your IT security efforts should be to mitigate risk at a level that is appropriate to the needs of your intervention.

John O'Bryan/USAID

Consideration	Resource to Address Consideration
Given the nature of the data, what level of security for IT systems and training for staff are needed?	See USAID Information System Security (ADS 545). USAID Security Training Policy, Standards, Guidelines, and Plan. To obtain a copy of this document, email ato@usaid.gov. *Note these are strictly applicable to systems owned by, or for the use of, USAID. At the same time, they may serve as a helpful reference for others.*[18]
Have you established a process for privacy incidents, including data breaches, especially if it falls under international or U.S. law? Who needs to be notified? Do you have the resources to track and manage the incident/breach?	See USAID Breach notification policy and plan (ADS 508mai), AAPD 16-02 Media and Information Handling and Protection, Privacy and Security IT Incident Reporting and GDPR: Notification of Data Breach.[19]

When you build systems that are intended to be taken over by host-country governments or other institutions, there may also be a security risk from under-resourced successors. If it is unaffordable to maintain and update services such as antivirus software or firewalls, the system's new owners may begin to cut corners. Partners who have not been adequately trained in security practices may choose weak passwords or leave server rooms unlocked. We should ensure that secure systems are sustainable within our partners' human and financial resources. Although not a complete checklist, this document provides highlights of good IT practices (Tool 5, IT Security Highlights Checklist, page 21).

18 To obtain a copy of this plan, email ato@usaid.gov.
19 USAID and the USG have not yet established an official position regarding the GDPR. You should consult with GC regarding if and how to factor GDPR into data planning and management.

5 TOOL
IT Security Highlights Checklist

The physical and technological protections put in place to safeguard data are the cornerstone of IT security. IT security is a large and complex field and will not be fully covered in this document. Nonetheless, below is a high-level checklist to get you started in establishing and maintaining IT security.

✓	
	Examine the flow of data from collection to storage, analysis, and preservation to identify points of vulnerability. Consider conducting a risk assessment.
	Establish processes to identify lost, corrupted, or tampered data. Determine the potential risks from these data and potential ways to mitigate these risks.
	Establish a procedure to deal with privacy incidents including data breaches (see page 20).
	Create backup and archival systems necessary to avoid inaccessibility and loss of data. Have the proper resources to meet these needs.
	Determine who has access to data throughout the intervention and how this will be tracked/monitored. Limit access where possible. Determine who has remote access and establish how this will be done securely.
	If using a cloud-based server, understand how your provider is preventing unauthorized access and ensure they meet any security requirements.
	Provide staff training on secure management, sharing, and transmission of data. This should include preparing staff on how to address a situation where someone powerful asks for inappropriate data access.
	Determine which devices contain sensitive information and the rules regarding retirement or off-premise use of these devices. If IT support/repair services are provided by vendors outside of the intervention team, determine data protection methods to protect sensitive data on machines. (See Tool 2, Key Events Planning Table, page 10).
	Establish oversight mechanisms to ensure that secure data storage guidelines are followed. Outline consequences for not following protocols.

Putting Data to Work

In this section, we will explore data use and the processes related to sharing and curating data for re-use. Throughout USAID's Program Cycle, data and information from monitoring, evaluation, and other analyses are used for accountability and learning and should inform adaptive management so that activities and projects achieve their objectives. Additionally, the broader development community can utilize data that has been shared and curated from past interventions both for general use and to improve future interventions. To create usable datasets as well as encourage re-use of data, it is essential to consider issues surrounding data quality, data sharing, and curation. Achieving high quality, curated datasets allows you and others to put your data to work.

Data Quality

High-quality decisions require high-quality data. Biased, inaccurate, or incomplete data can harm data subjects, their communities, and development organizations. If we are misinformed about what is happening, we may misdirect resources, pursue ineffective interventions, or deepen existing inequalities. Per ADS 201, all USAID data that is reported externally must go through a data quality assessment (DQA).

One useful tool is the USAID-recommended Data Quality Assessment Checklist.[20] This checklist contains a series of questions about data validity, reliability, timeliness, precision, and integrity. In most cases, these quality standards present trade-offs with each other. In addition, activity managers and designers need to balance various aspects of data quality with cost. Rather than being used only for an after-the-fact assessment of collected data, data quality guidance can be used to proactively design processes and safeguards that support successful activities.

In addition to your own program data, it is important to assess the quality of any third-party data that will be used in project implementation or evaluation. It may be better to use less data than to rely on data you cannot trust. Also consider and document the limitations of the data and what they can and should be used for.

20 See ADS 201sae and USAID's Data Quality Standards & Conducting a DQA for more information.

Morgana Wingard/USAID

TOOL

6 Using Responsible Data Practices to Meet Data Quality Standards

This chart is meant to help you think about how you can best meet the data quality standards detailed in the USAID DQA. Example questions are given to help you incorporate responsible data into your DQA and improve your results.

Standard	Definition	Example Considerations	Additional considerations/notes
Validity	Data should clearly and adequately represent the intended result.	What staff training processes are in place to facilitate accurate and unbiased data collection?	
Integrity	Data collected should have safeguards to minimize the risk of transcription error or data manipulation.	Who can edit data, at which point, and for what purposes? Are rights to edit as restricted as possible, both in terms of who can edit the data and during which intervention phase(s) they access it?	
Precision	Data should have a sufficient level of detail to permit management decision-making.	Do any of your privacy protections (i.e., data aggregation or lean data) impede data use for decision making?	
Reliability	Data should reflect stable and consistent data collection processes and analysis methods over time.	Are there standard protocols in place to promote responsible data handling—such as protocols/approaches for data aggregation to promote reliable interpretation?	
Timeliness	Data should be available at a useful frequency, should be current, and should be timely enough to influence management decision-making.	When planning the frequency of collection, do you consider burden to the beneficiary?	

Freddy Feruzi

Data Retention

The data curation and retention stages of the data lifecycle ensure that data are well-managed over the long term and continue to maintain and grow their value through time. Curation and retention involve archiving, preservation, and other activities to maintain data usability. These stages typically require specialized expertise and IT environments.

For most producers of development data, data curation should focus on submission of data and related documentation to a trustworthy digital repository. USAID's Development Data Policy (ADS 579) outlines USAID's curation approach for Agency-funded data and establishes the Development Data Library (DDL) as the Agency's central digital repository for detailed program and activity data.[21] The Development Data Policy guides the submission of USAID-funded data to the DDL (See Digital Repositories and the DDL on page 25).

Curation and retention activities often surface risks that were not well identified or managed at earlier stages in a program or activity. As an example, the process of preparing to submit data to a repository might reveal detailed information about individuals that were not sufficiently de-identified. It might uncover data columns (i.e., variables such as survey questions) that draw attention to individual data points or outliers. Curation often exposes risks and legal issues that result from inadequate documentation.

Preparation to submit to a repository could also reveal problems with informed consent documentation or the absence of necessary agreements. They can also discover risks related to privacy incidents as well as future combinations and use of data. It is important to identify and address these risks when planning to submit data to a digital repository.

Curation and retention activities also introduce possible new risks. In particular, they introduce risks that result from the ability to combine datasets and use them in unanticipated ways in the future. For more on this, see the "Sensitive Information" section of this document on page 15. To avoid these risks, you should work closely and carefully with repositories to submit appropriate data and necessary documentation as well as assess how they protect the data you submit. You can consult with repository experts early in a program to understand the responsibilities for submitting different types of data and documentation and also understand how you can best document risk and risk mitigation associated with your data submissions. Finally, you can ensure that you submit all necessary documentation, particularly any risk assessments and legal agreements, along with data submissions.

21 See ADS 579 for more information.

Panos

Digital Repositories and the DDL

Digital repositories are an excellent tool for data curation. USAID's DDL serves as a digital repository that provides a curation environment. The DDL facilitates long-term preservation, long-term access and re-use, well-governed, responsible management for the long-term, and responsible dissemination via publication and sharing. As part of USAID's commitment to open data and data-driven decision-making, all USAID awards as of October 2014 are now required to register and make available raw data to the DDL. These data should be related to the development of an "intellectual work" (if data are intended to support an analysis that will be submitted to USAID as a deliverable, then the data are probably an intellectual work). The DDL holds data with several different access levels:

- Public
- Restricted access
- Non-public

When submitting, the user provides a proposed access level along with a justification (which includes a detailed risk assessment of the data). The justification for withholding public access should reference one of the six criteria for the exemptions to openness (i.e., threats to national security interests, threats to personal safety of U.S. personnel or recipients of U.S. resources, etc. See ADS 579.3.2.3 for full list).

Discuss the following questions when considering the use of a digital repository:

- Have you contacted any digital repositories that may be suitable for your data to discuss requirements?
- Have you consulted ADS 579 to review submission requirements for the DDL?

Key Consideration: Planning for Data Retention

The processes of archiving, retention, and/or disposal are key for properly curating data.[22] Make sure your implementing partners consider the following for data that remain in their possession post-intervention:

- What are their archival, retention, or disposal methods for protecting personal or sensitive data?
- How will project staff be trained on proper archival, retention, and disposal practices?
- How are implementing partners monitoring and tracking these processes?

22 USAID staff must follow ADS 502: The USAID Records Management Program for records management. Under terms of contract, implementing partners submit detailed data to the DDL and documents to the DEC for long term storage.

Data Sharing

USAID's Development Data policy, ADS 579, describes data as "assets for USAID, its partners, the academic and scientific communities, and the public at large," and outlines an approach to making USAID's data open and machine-readable by default. When possible, we aim to create data that can serve as a public good and share them as widely as possible.

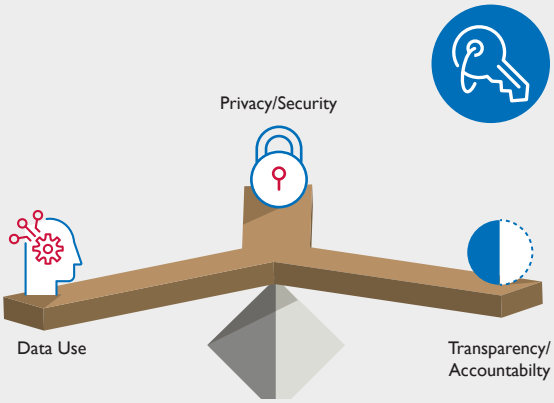
This openness needs to be coupled with respect for the privacy of data subjects, law enforcement or national security concerns, and host-country laws. Especially when we partner with host-country governments through Memoranda of Understanding (MOUs), non-disclosure agreements and contracts, it is important to clarify responsibilities and expectations (see the "Legal and Policy Issues" section of this document on page 4). There is often a tension between ensuring personal privacy and reaping the benefits of data sharing.

There may be cases where you decide to share data with specific partners. In these cases, legal documents such as MOUs or non-disclosure agreements can be a good way to ensure that data-receiving partners are committed to the same standards of data protection and ethical use that you put into place with your program and staff.

Key Consideration: Data Sharing

Data sharing is key to transparency and broader data use. Use of existing data via sharing or open data also provides opportunities to stretch resources. However, sharing data can create new privacy and security concerns.

When thinking about data sharing, some important points to discuss are:



- What processes are in place for data sharing?
  - Who has authority to publish, distribute, or release data; to whom; and when?
  - Does documentation of internal and external data sharing include who shared the data, why, when, and with whom?
- Will data be de-identified or aggregated prior to release? If so, how does the risk of re-identification compare to the sensitivity of the data? (See the "Sensitive Information" section of this document on page 15).
- If sharing data with the data subjects or the communities that contributed data, when and in which format will data be shared? What level of data protection is required?

Key Consideration: Third-Party Data

Imagine that an international organization is interested in creating a database of bacterial pathogens that cause foodborne illness. The database will contain whole-genome sequences of bacterial strains collected from foods at various points in the supply chain from farmer to fork. The speed of the technology along with a centralized database of samples from around the globe will allow for rapid detection of foodborne pathogens and their source, bolstering food security and preventing illness. To create this database, the international organization is obtaining data from third-party, in-country labs in government, academia, and the private sector. There are concerns from data collectors about data misuse and privacy given the sensitivity of data. For example, sequence data related to virulence or antibiotic resistance could pose national security risks that must be balanced with public health benefits.[23]

The above example highlights the need to discuss responsible data practices in third-party data use and sharing. To help you think through the considerations that are relevant to third-party data, discuss the following:

- Have you spoken with providers of third-party data about responsible data practices? Have third parties applied appropriate, responsible data practices in collecting, storing, and sharing data?
- What are the terms of use when sharing data obtained from a third-party? Are there any issues with informed consent?
- What are the data restrictions on and reporting requirements for third-party data and how does this impact data ownership of derivative works?

Conclusion

The goal of this document is to start a conversation among USAID staff and partners about how we can responsibly handle our data—maximizing the benefits of data use while identifying and mitigating risk. This will allow us to be responsible to our beneficiaries, ourselves, and the broader development community. These conversations should lead to countries and partners building their technological and data use capacity, in support of USAID's Journey to Self Reliance. This document is not an official statement of policy or legal requirements; for that you should turn to the ADS, M/CIO and your Resident Legal Officer (in Missions) or the Office of General Counsel (in Washington).

While some of us do not think about data every day, responsible data practices are not a detour or a distraction from development work. Instead, responsible use of data is grounded in good development practices of consultative stakeholder engagement, evidence-driven programming, and respect for the people we serve.

23 FAO (2016). Applications of Whole Genome Sequencing in food safety management.

Endnotes

For more on how other development organizations approach data responsibility, consider policies and recommendations from the following organizations (Please note these resources were gathered in 2017, no additional resources have been gathered since then):

Document and Link	Sector
International Organization for Migration (IOM) Data Protection Manual (2010) https://publications.iom.int/books/iom-data-protection-manual	Humanitarian Assistance (Migration)
FrontlineSMS (FLSMS) Data Integrity Guide (2011) http://static1.squarespace.com/ static/56e1a99907eaa0941d037b0a/56e1aaff06dcb7bbf42a7ba7/56e1aab506dc b7bbf42a750f/1457629877708/ frontlinesms_userguide.pdf?format=original	Technology (Mobile)
GSMA Association (GSMA) Privacy Design Guidelines for Mobile Application Development (2012) https://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/ gsmaprivacydesignguidelinesformobileapplicationdevelopmentv1.pdf	Technology (Mobile)
International Committee of the Red Cross (ICRC) Professional Standards for Protection Work (2013) https://www.icrc.org/eng/assets/files/other/icrc-002-0999.pdf	Humanitarian Assistance
Cash Learning Partnership (CaLP) Protecting Beneficiary Privacy (2013) http://www.cashlearning.org/downloads/calp-beneficiary-privacy-web.pdf	Finance
Principles for Digital Development (Working Group) (PDD) Digital Data Principles (2014) http://digitalprinciples.org/	Technology (ICT4D)
The Engine Room – Responsible Data Forum (RDF) Shooting Our Hard Drive into Space and Other Ways to Practice Responsible Development Data (2014) https://responsibledata.io/wp-content/uploads/2014/10/responsible-development-data-book.pdf	Cross-Cutting
Oxfam Responsible Program Data Policy (2015) http://policy-practice.oxfam.org.uk/publications/oxfam-responsible-program-data-policy-575950	Cross-Cutting
Girl Effect (GE) Girl Safeguarding Policy: Digital Privacy, Security and Safety Principles and Guidelines (2016) https://www.ictworks.org/wp-content/uploads/2016/05/GE-Girl-Digital-Privacy-Security-Safety-v-May-2016.pdf	Gender
The Engine Room – Responsible Data Forum (RDF) The Handbook of the Modern Development Specialist (2016) https://responsibledata.io/resources/handbook/	Technology (ICT4D)

Document and Link	Sector
UN Global Pulse (UNGP) Privacy and Data Protection Principles (2016) http://www.unglobalpulse.org/privacy-and-data-protection-principles	Technology (Big Data/ICT4D)
Electronic Cash Transfer Learning Action Network (ELAN) A Data Starter Kit for Humanitarian Field Staff (2016) http://elan.cashlearning.org/	Finance
International Committee of the Red Cross (ICRC) Rules on Personal Data Protection (2016) https://www.icrc.org/en/publication/4261-icrc-rules-on-personal-data-protection	Humanitarian Assistance
World Food Program (WFP) Guide to Personal Data Protection and Privacy (2016) https://docs.wfp.org/api/documents/e8d24e70cc11448383495caca154cb97/download/	Humanitarian Assistance
Oxfam Responsible Data Management Training Pack (2017) http://policy-practice.oxfam.org.uk/publications/responsible-data-management-training-pack-620235	Cross-Cutting
The Harvard Humanitarian Institute (HHI) Signal Code: A Human Rights Approach to Information During Crisis (2017) https://hhi.harvard.edu/publications/signal-code-human-rights-approach-information-during-crisis	Humanitarian Assistance
National Institute of Standards and Technology (NIST) Privacy Risk Management for Federal Information Systems, National Institute of Standards and Technology Internal Report (NISTIR) 8062 (2017) http://csrc.nist.gov/publications/drafts/nistir-8062/nistir_8062_draft.pdf	Technology (Information Systems)
Catholic Relief Services (CRS) Responsible Data Principles and Guidelines (2017, pending publication)	Cross-Cutting
United Nations Development Group (UNDG) Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda (2017) https://undg.org/document/data-privacy-ethics-and-protection-guidance-note-on-big-data-for-achievement-of-the-2030-agenda/	Technology (Big Data/ICT4D)

Annex 1: Key Words

Adaptive management: A management approach used in uncertain, complex, or changing contexts. Adaptive management incorporates experimentation and learning to increase knowledge and select improved courses of action—as opposed to rational planning or crisis management.

Data:[24] Recorded information, regardless of form or the media on which it may be recorded. The term includes technical data and computer software. The term does not include information incidental to contract administration, such as financial, administrative, cost or pricing, or management information.

Data Aggregation: A common form of anonymization that is performed by aggregating records on individuals to create a new dataset which summarizes totals by classification (e.g., taking individual patient files and summarizing numbers of patients by disease, gender, and age). Data aggregation can also refer to the combination of multiple datasets from different sources into one master dataset.

Data Curation: The submission of data and all related documentation to a repository for long-term preservation.

Data De-identification: A form of data cleaning that removes identifiers and other sensitive information from a dataset.

Data Life Cycle: The stages through which data move. These include: 1) Plan, 2) Collect and Acquire, 3) Process and Integrate, 4) Analyze, 5) Curate, and 6) Publish and Share.

Data Management Plan: A planning tool to help project/activity managers ensure proper management, protections, curation, and resource planning for an activity's data throughout its life cycle.

Data Ownership: A term referring to who "owns" the data—more explicitly who has final, legal authority over access and use of the data. USAID policy states that data is owned by the organization who collects the data; however, the U.S. Government retains an unrestricted right to access and usage of the data.

Data Pseudonymization: A form of data de-identification, pseudonymization means the replacement of identifying information with codes or pseudonyms. It is possible to re-identify individuals by obtaining (or inferring) matched lists of codes and real identifiers.

24 Taken from Glossary of ADS Terms (last updated May 23, 2018).

Data Sovereignty: The principle that national governments should have exclusive authority and control over public digital assets.[25] A nation's data sovereignty is compromised if data required for government operations are controlled by outside entities.

Demographically Identifiable Data (DII): Demographically identifiable data is defined as, "data points that enable the identification, classification, and tracking of individuals, groups, or multiple groups of individuals by demographically defining factors. These may include ethnicity, gender, age, occupation, and religion,"[26] which may then be automatically considered sensitive.

Direct Identifier: A direct identifier is information that directly identifies an individual, including formal name, national identification number, genetic code, fingerprints or other biometrics.

Evaluation:[27] Systematic collection and analysis of information about the characteristics and outcomes of strategies, projects, and activities, conducted as a basis for judgments to improve effectiveness and timed to inform decisions about current and future programming. Evaluation is distinct from assessment or an informal review of projects.

Indirect Identifier: An indirect identifier is information that, when combined with other information, such as a combination of age, location, gender, education, and employment, can facilitate the identification of a data subject. For example, in a small village of 150 households, the description "a Christian male aged 35 with three children under age five who works in a local factory," may allow his community to identify the individual easily; in a large city, it may be less likely to reveal a precise identity, unless Christians or male factory workers are particularly rare.

Informed Consent:[28] Informed consent, as defined by the U.S. Government, involves three features: 1) disclosing sufficient information about direct risks and benefits to a participant so he or she can make an informed decision on whether to participate, 2) making sure the participant truly understands this information, and 3) making sure the decision to participate is truly voluntary. In general it is best practice to seek consent whenever an activity involves individuals; however, it is federally required when it involves human subjects research.

Information Technology (IT):[29] Any services or equipment, or interconnected system(s) or subsystem(s) of equipment, that are used in the automatic acquisition, storage, analysis, evaluation, manipulation, management, movement, control, display, switching, interchange, transmission, or reception of data or information. The term "information technology" includes computers, ancillary equipment (including imaging peripherals, input, output, and storage devices necessary for security and surveillance), peripheral equipment designed to be controlled by the central processing unit of a computer, software, firmware and similar procedures, services (including provisioned services such as cloud computing and support services that support any point of the lifecycle of the equipment or service), and related resources.

Lean Data Principles: A set of principles that can be applied to data collection and used to bolster efficiency and create value with stakeholders.[30] One of the core principles is to limit data collection only to data that is needed.

25 Irion, Kristina (2013). "Government Cloud Computing and National Data Sovereignty." Policy and Internet, 4(3-4), 40-71
26 See https://signalcode.org/code-intro/glossary/
27 Taken from Glossary of ADS Terms (last updated May 23, 2018).
28 Adapted from U.S. Department of Health and Human Services.
29 Adapted from Glossary of ADS Terms (last updated May 23, 2018).
30 Dichter, Sasha; et al., (2016) "The Power of Lean Data." *Stanford Social Innovation Review.*

Learning:[31] In the development context, learning refers to a continuous process of analyzing a wide variety of information sources and knowledge—including evaluation findings, monitoring data, innovations and new information that bring to light new best practices or call into question received wisdom and collected observations and tacit knowledge from those who have particularly deep or unique insight in a given area—leading to iterative adaptation of strategies and project design and/or implementation, in order to sustain the most effective and efficient path to achieving development objectives.

Metadata: Data about data. For example, if a dataset consists of data subject names, addresses, and birthdates, the metadata might describe naming conventions (e.g., given names followed by family name), address structure (e.g., house number, street name, city, province and, postal code), and date format (e.g., MM/DD/YYYY). Metadata can help people who are not part of your team interpret data without uncertainty.

Monitoring:[32] The ongoing and systematic tracking of data or information relevant to USAID strategies, projects, and activities. Relevant data and informational needs are identified during planning and design, and may include output and outcome measures that are directly attributable to or affected by USAID interventions, as well as measures of the operating context and programmatic assumptions.

Open Data:[33] Data that "anyone can freely access, use, modify, and share for any purpose."[34] For this to happen, data must be legally open (in the public domain or under liberal terms of use with minimal restrictions) and technically open (machine readable and preferably in a non-proprietary electronic format). That said, open data is not private or personal data.

Personally Identifiable Information (PII):[35] Information that directly identifies an individual. PII examples include name, address, social security number, or other identifying number or code, telephone number, and email address. PII can also consist of a combination of indirect data elements such as gender, race, birth date, geographic indicator (e.g., zip code), and other descriptors used to identify specific individuals. Same as "information in an identifiable form."

Privacy Incident: An incident which has the potential for unauthorized disclosure, access, changes, or removal of personal or sensitive data. Please note that an incident only needs to raise or expose the potential for harm, rather than having explicitly resulted in harm.

Sensitive Data: An additional subcategory of "personally identifiable information" that includes sensitive personal information. Sensitivity is context-specific. There is no consensus definition about what characteristics are included in sensitive data. Examples of sensitive personal information could include: racial or ethnic origin, sexual orientation, physical or mental health information, political opinions or affiliations, criminal records, biometric records, genetic information, membership in a union or other similar organization, ex-combatant status, or refugee displacement status.

Stakeholders: All parties involved or impacted in an activity, including USAID staff, other U.S. partners, prime and subcontractors/agreement holders, local partners, government partners, advocacy organizations, local community, and individual beneficiaries.

31 Taken from: Glossary of ADS Terms (last updated May 23, 2018).
32 Taken from: Glossary of ADS Terms (last updated May 23, 2018).
33 Adapted from: The World Bank: Open Data Toolkit (copyrighted 2017).
34 See https://opendefinition.org/.
35 Taken from: Glossary of ADS Terms (last updated May 23, 2018).

Annex 2: How This Document Was Created

This document was developed by USAID's Global Development Lab based on research by FHI 360 and Sonjara, Inc. The considerations herein are anchored in best practices that were identified based on a literature review, a stakeholder analysis, and case studies. The research and development of the document had four phases:

- **Phase 1:** Conducting research into responsible data practice, including a literature review of international and humanitarian best practices, legal landscape analysis, and key informant interviews with stakeholders at USAID and implementing partners (March-July 2017).
- **Phase 2:** Developing draft responsible data considerations based on the literature review (June 2017).
- **Phase 3:** Testing the applicability of the guidelines in real-world digital development activities in Nigeria and Kenya (June-July 2017).
- **Phase 4:** Refining the document based on the country case study findings, stakeholder analysis, and feedback (July 2017-February 2019).

Annex 3: How This Document Aligns with the USAID Program Cycle

While the structure of this document does not mirror the USAID Program Cycle, many aspects of responsible data use are important at multiple phases of the Program Cycle. For example, it is important to plan for IT security before implementation begins, during implementation, and after the activity closes to ensure that data remain secure. Also note, this document does not serve as an authoritative guide to responsible data practice or data management under the USAID Program Cycle. While some of the resources and approaches shared here may be helpful, additional efforts are ongoing with USAID's Bureau for Policy, Planning, and Learning (PPL) and Office of the M/CIO to provide detailed guidance necessary to operationalize responsible data practices at a formal program level.

USAID
FROM THE AMERICAN PEOPLE

fhi360
THE SCIENCE OF IMPROVING LIVES

mSTAR