

# OPEN DATA POLICY COMPLIANCE GUIDE

A Practitioner's Guide to Interpreting and Complying  
with USAID's Open Data Policy



**USAID**  
FROM THE AMERICAN PEOPLE

TOPS  
Small  
Grant



## TABLE OF CONTENTS

INTRODUCTION .....	4
ACKNOWLEDGEMENTS .....	5
WELCOME LETTER .....	6
SECTION I: THE BIG PICTURE PERSPECTIVE .....	8
1. The Importance and Benefits of Quality Data Management.....	9
1.1. What is Open Data? .....	9
1.2. Getting from Here to There .....	9
1.3. The Power of Data Management.....	10
1.4. Who's Responsible for Data Management?.....	10
1.5. What Makes Data Useful?.....	11
1.6. Who Will Use This Open Data?.....	11
1.7. The Goal of This Guide .....	12
2. The USAID Open Data Policy and the Development Data Library: <i>What It is and Why It Matters to Your Organization</i> .....	13
2.1. A USAID Policy Recap.....	13
2.2. Seeing Beyond the Policy .....	14
2.3. What Does "Compliance" Really Mean? .....	15
2.4. Data Is...More than Just Data!.....	16
2.5. Data is Not...Always Sharable!.....	16
2.6. Developing an Internal DM Policy .....	17
SECTION II: QUALITY DATA MANAGEMENT .....	21
3. Improving Your Data Management Practices .....	22
3.1. Developing a Data Management Plan.....	22
3.2. Elements of a Data Management Plan .....	22
3.3. Preparing to Comply.....	28
SECTION III: A STEP-BY-STEP PROCESS: PROPOSAL THROUGH SUBMISSION ....	29
4. Step-by-Step: Dataset Award Negotiations and Communications .....	31
4.1. Responding to a USAID Solicitation .....	32
4.2. Developing Your Proposal.....	34
4.3. Receiving the Award.....	34
5. Step by Step: Dataset Preparation and Submission.....	37
5.1. Preparing and Tracking Datasets.....	38
5.2. Registering and Submitting Data to the DDL.....	40
5.3. What Does Success Look Like? .....	42
SECTION IV: OTHER CONTEXTUAL CONSIDERATIONS .....	43
6. Ethics, Ownership and Working with Partners.....	44
6.1. Securing Informed Consent .....	46
6.2. Exceptions, Embargos & Submissions .....	47

6.3. Data Ownership.....	48
6.4. De-Identification of Data.....	48
6.5. Working with Partners .....	52
7. Staffing and Budgeting .....	53
7.1. Summary of Roles and Responsibilities .....	53
7.2. Budgeting for Your Data Management Plan .....	55
8. Templates, Tools and Resources.....	57
8.1. Master List of Tools .....	57
Tool 1: Policy Discussion Checklist .....	57
Tool 2: Data Life Cycle Infographic .....	57
Tool 3: DCC Data Management Plan Questionnaire.....	57
Tool 4: UMN Metadata Codebook .....	57
Tool 5: Stanford File Naming Handout .....	57
Tool 6: Informed Consent Template .....	57
Tool 7: Ethics and Privacy Talking Points .....	57
Tool 8: Incorrect Example of DDL Submission .....	57
Tool 9: Correct Example of DDL Submission.....	57
Tool 10: DDL Compliance Checklist .....	57
Tool 11: DDL Registration/Submission Form.....	57
Tool 12: Uganda Agric Practices Codebook (Example).....	57
Tool 13: Uganda Agric Practices Data (Example).....	57
Tool 14: Case Study Reference Material.....	57
Tool 15: Requesting Compliance Guidance from Prime (Email Template) .....	57
Tool 16: UK Costing Tool .....	57
Tool 17: Sample DDL Tracker .....	57
Tool 18: Steps and Tools at a Glance .....	57
8.2. Useful Links and Resources .....	57
8.3. Case Study Example .....	59
8.4. Key Terms to Know.....	60
8.5. Dataset Access Levels Defined .....	61
8.6. Exceptions Language (Samples) .....	62
8.7. Acronyms List .....	64

## INTRODUCTION

This resource guide was made possible by a grant from the United States Agency for International Development (USAID) Technical and Operational Performance Support (TOPS) program. The TOPS Micro Grants Program is made possible by the generous support and contribution of the American people through USAID. The contents of the materials produced through the TOPS Micro Grants Program do not necessarily reflect the views of TOPS, USAID or the U.S. Government.

Founded in 1961, Project Concern International (PCI) is an international non-profit organization dedicated to preventing disease, improving community health and promoting sustainable development worldwide. Motivated by our concern for the world's most vulnerable children, families and communities, PCI envisions a world where abundant resources are shared, communities are able to provide for the health and well-being of their members, and children and families can achieve lives of hope, good health and self-sufficiency. PCI is headquartered in San Diego, CA, with U.S. offices in Washington, DC and Seattle, WA. PCI currently operates in 16 countries in Asia, Africa and the Americas. Intervention focus areas include: food and livelihood security; health and nutrition; water and sanitation; and humanitarian assistance and disaster risk management. Cross-cutting areas of focus include community mobilization; gender equity; social and behavioral change; local capacity strengthening; and sustainable impact.

### **Published in 2017 by:**

Project Concern International  
5151 Murphy Canyon Road, Suite 320  
San Diego, CA 92123  
USA

**Author:** Kim Murphy (Consultant), Janine Schooley (PCI), Clara Eder (PCI) and Julia Montijo (PCI)

**Contributors:** Carrie Hessler-Radelet (PCI), Erin Graeber (PCI), and Caroline In (PCI)

**Graphics & Layout:** Jessica Ayala (Consultant)

**Translation:** Valerie Guitton (French) and Erwin Cox (Spanish)

**Cover Image:** Tony Tseng (PCI)

**For more information,** please contact Clara Eder, Vice President, Strategic Information for Impact, at [CEder@pciglobal.org](mailto:CEder@pciglobal.org).

## ACKNOWLEDGEMENTS

We wish to acknowledge the TOPS program for Capacity Building in Best Practices, Assessment and Documentation of Promising Practices for their generous financial support toward the development of this Guide. We also wish to acknowledge USAID, InterAction, Oxfam America, and International Rescue Committee, who together with PCI, created an informal “community of practice” to focus on identifying promising processes, steps, and tools to support compliance with the USAID Open Data Policy. This output informed the development of this resource Guide. Because this Guide is intended to be a living document, we would like to thank – in advance – all of the participants who may attend dissemination events and workshops over the coming months and who may provide useful feedback and ideas to incorporate into future versions.

*This guide was made possible by a grant from The Technical and Operational Performance Support (TOPS) Program. The TOPS Micro Grants Program is made possible by the generous support and contribution of the American people through the U.S. Agency for International Development (USAID). The contents of the materials produced through the Micro Grants do not necessarily reflect the views of TOPS, USAID, or the U.S. Government.*



## WELCOME LETTER

PCI's mission is to empower people to enhance health, end hunger and overcome hardship, and our vision is to enable the most vulnerable people in the world to have the power to lift themselves out of poverty and to create vital, healthy lives for their families and communities. We are honored to partner with USAID in accomplishing this important work.

Our work needs to be 100% transparent, especially as we rely on the generosity of private donors and the U.S. taxpayers, so we understand the philosophical merits of open and shared data. Privacy is also extremely important, and with the support of the Technical and Operational Performance Support (TOPS) program, we developed an Open Data Policy Compliance Guide for practitioners to support NGOs in complying with USAID's Open Data Policy.

This Guide provides data management best practices, policy compliance information, a library of useful tools, and step-by step procedures for submitting data to USAID's Development Data Library (DDL). Designed to be easy to read and follow, the guide will enable food, nutrition and livelihood security practitioners and other stakeholders to strengthen their internal data management systems, maximize ways that data and information can be utilized strategically to improve program quality, and ultimately support USAID's commitment to the value of open and shared data.

We expect this Guide will evolve over time with input from practitioners and peer colleagues like you. We welcome your feedback and invite you to participate in any of the dissemination events and user workshops that will take place during the year.

Many thanks to all who have assisted in making the Guide a tangible and worthwhile tool, and we look forward to working with you as we continue to refine it and tell the story of the effectiveness and importance of our work with real, transparent data.



Carrie Hessler-Radelet  
*President and CEO*  
Project Concern International (PCI)

A young woman with dark hair, wearing a dark blue sweater, is smiling and looking down at a project on a table. She is holding a piece of cardboard. In the background, there are shelves filled with books and other items. The image has a blue tint.

# Section I

## The Big Picture Perspective

### Tools You'll Use...

In this section, you'll be introduced to and use the following tools:

- Tool 1: Policy Discussion Checklist
- Tool 2: Data Life Cycle Infographic

# 1. The Importance and Benefits of Quality Data Management

For entrepreneurs, nonprofits, researchers, scientists and others focused on global health and development, planning for the effective creation, management and sharing of data allows you to get the most out of your programs. Quality data management practices can enable an organization to more effectively use and leverage its data for innovation, quality improvement, impact management and sustainability, and ultimately to achieve critical organizational goals.

## 1.1. What is Open Data?

“Open data” is data that is made available to the public and structured in a way that it is usable. Data must also be able to be found when users are searching for it.<sup>1</sup> Since 2009, the U.S Government has instituted policies designed to make a variety of information and data from its agencies available to the public.<sup>2</sup> But the United States is not the only country focused on such efforts. Open data is part of a larger global trend, with individual efforts underway as well as a consortium of more than 60 countries participating in the Open Government Partnership to make their governments more open, accountable and responsive to citizens.<sup>3</sup>

The U.S. Agency for International Development (USAID) is similarly focused, having drafted comprehensive policies and procedures to guide its programs and operations. To that end, it has created the Development Data Library (DDL), the repository engine which will enable data to be publicly accessible.

## 1.2. Getting from Here to There

It is not a straight or easy path from receiving a grant, creating a program or study, capturing a data point, and making that information searchable, accessible and usable by anyone around the globe. Hundreds of steps and decisions exist between those two milestones.

And getting from here to there involves a host of players. The government plays a pivotal role in crafting and implementing policies that not only address the collection and storage of data, but also mitigate the concerns that consumers and non-governmental organizations (NGOs) have, such as privacy issues. Industry, NGOs and researchers are vital because they are the data visionaries,

---

<sup>1</sup> <https://www.usaid.gov/data/frequently-asked-questions#Q1>

<sup>2</sup> <https://opengovdata.io/2014/us-federal-open-data-policy/>

<sup>3</sup> <https://www.opengovpartnership.org/open-government-declaration>



understanding the importance, meaning and value of every piece of information they collect.

As the McKinsey Group noted in their Open Data Report,<sup>4</sup> success depends upon the cultivation of a vibrant open-data ecosystem. For organizations, that means putting in place the technologies and talent to collect and analyze data. For individuals—as both consumers and citizens—it means being vigilant, savvy providers and users of open data.

### **1.3.The Power of Data Management**

It is important for every organization within this vibrant open-data ecosystem to understand the benefits associated with quality data management. Quality data management promotes transparency and accountability, and can:

- Help save researchers time because data is better organized and easier to find
- Streamline the flow from building blocks to improved outcomes
- Enable an organization to more clearly communicate outcomes to sponsors and peers
- Support better reporting and documentation
- Make the compliance and audit processes easier
- Lead to opportunities for an organization to leverage data for future research efforts
- Elevate the community's awareness of an organization's efforts and accomplishments
- Provide continuity if project staff leave or new researchers join
- Help avoid data duplication
- Ensure publications are maintained and allow for results validation
- Facilitate data sharing with others – industry, government, communities, schools, non-profits and NGOs – to support positive activities and outcomes in other areas
- Utilize shared data to encourage learning and change, be it for better personal habits (eating, self-care health), corporate or community practices
- Ensure that participant and beneficiary identities are protected
- Promote a culture of inquiry and continuous evidence-based learning and improvement

### **1.4.Who's Responsible for Data Management?**

Quality data management is the responsibility of everyone within an organization. That's because steps associated within a comprehensive data management practice

---

<sup>4</sup> <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>

can span a variety of departments and operational roles – from grant proposal writers to program managers and researchers to contract administrators. Chief among those affected of course are researchers, evaluators, and monitoring staff.

By reviewing steps associated with the data management lifecycle (*Chapter 2*), you can determine which people and teams within your organization are most involved or affected by various tasks and decisions. Depending upon the types and sizes of your programs, you may decide to assign a staff member to serve as a data steward who is responsible for ensuring that your data management practices are executed properly.

## 1.5.What Makes Data Useful?

According to USAID<sup>5</sup>, data is most useful when it is:

- Consistent: Data must be standardized and comparable in order to be used by others.
- Related: An explicit relationship between data is important, where there are unique and consistent IDs to connect the data.
- Scaled: Comparable data across multiple countries can carry more weight.
- Unique: Data that is difficult or impossible to access any other way is inherently valuable.
- Documented: Providing context in terms of how data is collected adds to its meaning.
- Timely: The recency of data matters, especially if it's real-time data.
- Easily accessible: A web-based interface that supports quick and easy downloads makes data more relevant and useful.

## 1.6.Who Will Use This Open Data?

Data becomes particularly valuable – and useful – when combined with other information. Yet data producers and consumers are not necessarily aware of their common interests. For example, the journalism community may not realize that USAID has data that might be relevant to their investigative pieces. Yet, data is critical to stories intended to illuminate issues related to extreme poverty and the impact of development. Similarly, other researchers may not realize the array of data that exists. USAID is therefore committed to helping to create standards for data collection and to establish mutually-beneficial data pipelines.

---

<sup>5</sup> USAID's Open Data Policy, ADS 579 – Development Data, Center for Data Innovation Presentation (PowerPoint), December 9, 2014

## **1.7.The Goal of This Guide**

The goal of this guide is to enable PCI and its peer organizations to make their data useful, unleash the power of their data, by being good data stewards and implementing quality data management practices. The guide is also intended to facilitate compliance with USAID policy and in doing so, ultimately, help to create this vibrant open-data ecosystem.

## 2. The USAID Open Data Policy and the Development Data Library: *What It is and Why It Matters to Your Organization*

### 2.1.A USAID Policy Recap

This guide is not intended to present the Open Data Policy itself, but rather to serve as a roadmap for how to interpret, comply and work within the policy guidelines and submit datasets as directed to the repository, known as the [DDL](#).

Still, a brief recap might be helpful.

In October 2014, USAID released its first ever [open data policy](#), paving the way to provide the global public with regular, systematic access to USAID-funded data for the first time in the organization's history. The Agency modified all of its awards to stipulate that USAID-funded data must be registered in a central repository, and made publicly available, wherever it could do so responsibly, without violating privacy, security and other legal considerations. That same month, USAID established the rules for submitting data to its DDL and defined the process by which that data will be reviewed internally to determine its eligibility for release to the public.

**Dataset** is defined as an organized collection of structured data, including data contained in spreadsheets, whether presented in tabular or non-tabular form. It includes data collected as part of performance monitoring or evaluation, and project data used in reporting to USAID.

*Source: USAID*

#### *Purpose of the DDL*

USAID implementing partners and grant sub-awardees are required to submit datasets that are used (or of sufficient quality) to produce an Intellectual Work that is generated with USAID funding to the DDL in machine-readable, non-proprietary formats.

Unless otherwise directed by the Agreement Officer (AO) or the Contract Officer's Representative (COR) or Agreement Officer's Representative (AOR), the grant recipient (or sub-awardee) must submit the dataset and supporting documentation to the DDL within thirty (30) calendar days after the dataset is first used to produce that Intellectual Work.

When submitted, USAID implementing partners tag the data with labels (metadata) to help consumers better understand, find, and use the data. Datasets are published at [www.usaid.gov/data](http://www.usaid.gov/data), unless legally exempted, after undergoing an internal clearance process. Published datasets also appear automatically at [Data.gov](http://Data.gov).



### The Open Data Mission

The U.S. Agency for International Development is committed to the President's Open Government initiative, upholding the values of transparency, participation, and collaboration in tangible ways that benefit the American people.

- **Transparency.** Government should provide citizens with information about what their government is doing so that government can be held accountable.
- **Participation.** Government should actively solicit expertise from outside Washington so that it makes policies with the benefit of the best information.
- **Collaboration.** Government officials should work together with one another and with citizens as part of doing their job of solving national problems.

Source: <https://www.usaid.gov/open>

## 2.2. Seeing Beyond the Policy

As soon as policies are enacted, there is a rush to discuss issues of compliance. Of course, compliance must be a top priority for any organization that receives funding from USAID. And yet, the Agency wants its partners to take a step back and to recognize the value in submitting and sharing their data – to realize the benefits of becoming an active participant in the open-data ecosystem. Possessing that broader perspective may influence decisions you make regarding data management practices within your organization.

We are all encouraged to keep this in mind as we study the policy's specifics.

### Open Data Drives Crop Selection

In response to the 2010 drought in the Horn of Africa, USAID partners were able to reposition food supplies based on maps and visualizations created from open data. MFarm, an organization in Kenya, [created a mobile application](#) using publicly available crop price data to connect farmers with markets and to help them select crop varieties likely to yield the greatest income.

Source: *USAID Impact Blog*

## 2.3. What Does “Compliance” Really Mean?

If you’ve read the policy, you may have a host of questions swirling around your head. One of which may be: *What do I need to do to comply with this policy?*

Due to the nature of work funded by USAID, the policy is wide-ranging. Compliance will not be achieved by following a list of predetermined steps. You will not simply upload a spreadsheet and mark the compliance checkbox.

Instead, you must take time to consider not only the data you are uploading but also how that data might be understood (and misunderstood), and used by the wider, global population of students, researchers, journalists, peers and government officials who will have access to it. You must be able to communicate your efforts and your vision to end users when they access the DDL, so that they have the benefit of a context-based perspective.

Harkening back to the big picture discussion, you must remember to view your data and its structure from the ecosystem’s prism.

And your decisions will be different from that of peers and colleagues. For example, the question of what constitutes a dataset for your study or program will differ from that of a peer organization. And, how you capture and collect data for a program will likely be in contrast with how a co-worker on another program does. All of these nuances, coupled with program variables such as host country and privacy issues, must be considered.

In short, compliance is a multifaceted and complex matter. It requires a well-thought out plan with the right internal policies and procedures in place to ensure your organization is meeting requirements in a timely manner.

An internal Data Management (DM) policy can be a launch pad for the creation of Data Management Plans, or DMPs, which outline the specific practices and processes for your organization’s individual programs, studies or efforts (see Section II). **Developing an internal policy to drive creation of a DMP is a best practice that will facilitate your efforts to comply with USAID’s requirements and enable you to contribute in a meaningful way to the greater data ecosystem.**

### Compliance vs. Best Practices

To achieve compliance, your organization must complete all of the policy driven tasks and activities required by the Government within the allowed timeframe. *How you comply* – the processes and steps you follow – is up to your organization.

Best Practices are proven methods, strategies, and tools that lead to desired results. Often, Best Practices yield additional benefits, such as improved workflow, reduced errors or greater efficiencies.

This Guide presents a range of Best Practices for your organization to consider.

## 2.4.Data Is...More than Just Data!

The University of Minnesota has compiled an important summary of what data is. . . and is not. This frame of reference will be critical as your organization develops its own internal DM policy.

In the *Reference Model for an Open Archival Information System (OAIS)* ([Wikipedia](#)), data is defined as "[a] reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen."

Types of data include:

- Observational Data
- Laboratory Experimental Data
- Computer Simulation
- Textual Analysis
- Physical Artifacts or Relics

For social science, data is generally numeric files originating from social research methodologies or administrative records, from which statistics are produced. It also includes, however, more data formats such as audio, video, geospatial and other digital content that are germane to social science research.

Digital text is becoming increasingly important, where data may take the form of textual information, semantic elements, and text objects.

## 2.5.Data is Not...Always Sharable!

The federal government<sup>6</sup> provides specific guidance, policies and conditions for recipients of federal funding, which cover what research data is NOT suitable for the purpose of sharing or archiving. These include:

- Preliminary analyses
- Drafts of scientific papers
- Plans for future research
- Peer reviews, or communications with colleagues
- Physical objects (e.g., laboratory samples)
- Trade secrets
- Commercial information

---

<sup>6</sup> 2 CFR 200 – The Uniform Guidance (<https://www.ecfr.gov/cgi-bin/text-idx?node=2:1.1.2.2.1&rgn=div5>)

- Materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law

The federal government further adds personnel and medical information, or other similar data, to this list, “the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.” This is particularly pertinent for those who receive USAID funding.

## 2.6. Developing an Internal DM Policy

To develop a data management policy for your organization, take some time to:

1. Study USAID’s DDL policy and discuss how your organization interprets it;
2. Assess your current data management standards, protocols, practices; a
3. Ask peer organizations if they would be willing to share their policies and practices so that you can learn from others; and
4. Create your own DM policy to drive your compliance efforts.

Each of these items are addressed briefly below.

### *Study the Policy, Discuss Interpretations*

Because the policy has been enacted for a while, you may have already tackled this step. Many organizations have formed internal working groups or have at a bare minimum had program teams assess the impact of the USAID policy on their programs.

To be certain that your interpretations align with the policy, be sure you understand the following key terms from USAID’s perspective (See *Chapter 8, Section 8.4 Key Terms to Know*):

- Dataset
- Data Collection (A group of related datasets, where each dataset can have multiple spreadsheets)
- Intellectual Work

With these definitions in mind, consider the following list of questions. The list is designed to help you to look holistically at the awards and programs across your organization and to identify the raw data elements and their place within a broader context. As you answer these questions, you will be teasing out your team’s interpretations of the policy within the construct of this bigger picture. The list is intended to jumpstart your discussions, interpretations and understanding, but is not intended to be comprehensive of all that you must consider.

At this stage, you should not focus too much on the nitty gritty details. That will happen when you create individual DMPs for each individual award (or program). See *Section II, Chapter 3* for guidance on creating a DMP.



Refer to the [Policy Discussion Checklist, Tool 1](#), included in *Chapter 8*:

- What types of datasets or data collections does your organization typically produce?
  - Can you create lists or categories for these? (Data that is sharable versus data that is not always sharable?)
  - In what type of format are they produced? (i.e., hard copy surveys, Excel spreadsheets, etc.)
  - Are any of these datasets or collections exempt from the policy? If so, why? (Clearly articulate your rationale for exemption based on your understanding of the policy.)
  - In what types of formats are these exempt items usually produced? (i.e., email messages, PDF files, PowerPoint presentations, Word documents, photographs, etc.)
  - Are you able to describe the deeper meaning behind these datasets and data collections? What would you want a stranger to know about your study? What type of context do they need in order to make sense of the raw data? Are there any readily identifiable areas for misunderstandings that can be cleared upfront? Are you capturing this information for your studies? From the beginning of the program, do monitoring and evaluation staff include those explanations?
- What types of Intellectual Works do you typically produce? (program and communications materials, evaluations and assessments, information products, research and technical reports)
  - Are there different types of Intellectual Works that you produce (project evaluations vs annual reports)? If so, how many different types?
  - Does your organization have an intellectual property policy?
- What types of support documentation do you produce? (such as code books, templates and forms, or notes on data quality)
  - Are there types of support documentation that you do not currently produce but that you would like to see your organization begin to generate? If so, what might they be (i.e., definitions or explanations of fields within a dataset)?
- What barriers or challenges do you see with compliance?
  - Do you work with countries that might have policies which limit data sharing? How might you facilitate an agreement with them? What needs to be done upfront or after the fact?
  - Do you have local partners or even patients or participants who are reluctant to have their data shared?
  - Are there privacy, security or ethical issues you will need to address? (See *Section IV, Chapter 6.4* for tips on de-identifying data.)
  - Do you have a variety of different types of awards from different funders, where policy must be refined to accommodate each?

Now, have the team ask itself these types of questions:

- What is the spirit of the policy as it relates to your organization's type of work? Are there disconnects or conflicts within your interpretations?
- Are you clear on your submission timeline?
  - Can you determine when the 30 day timeframe will fall after you've completed an Intellectual Work? Do you need to address the timeline with the COR/AOR?
- Are there any grey areas that may affect your organization and its programs?
- Can you clearly articulate those grey areas and formulate a position or stance of what your program teams must do?
- What is the bare minimum in terms of compliance that your organization must do? If you wanted to set the gold standard in terms of compliance, what would your organization choose to do?
- What types of resources (e.g., human, financial) will you need in order to achieve compliance? (See **Tool 16** for guidance on budgeting and costing.)

Depending upon the size and focus of your organization, you may opt to have an in-house compliance professional to spearhead and oversee efforts related to the execution of a DMP, or you may rely on individual program teams to use the DMP as a basis for all of the tasks associated with compliance.

## ***Assess Your Current Practices***

Before you are able to create an internal DM policy and subsequent DMPs, analyze and address your data management standards, protocols, practices.

### ***The Data Management Life Cycle***

In your current data management practice, do you follow a formal data management life cycle? As a best practice, proven data management practices follow a common life cycle that starts as early as the planning and proposal submission process. Review these life cycle steps and determine whether you have a similar process in place. If not, would these steps work for your organization? Could they be tailored for your needs? Decide whether you will adopt this life cycle or document your own.

## **Steps in the Data Life Cycle<sup>7</sup>**

### **1. Proposal Planning & Writing**

- Review of existing data sources, determine if project will produce new data or combine existing data
- Investigate archiving challenges, costs, consent and confidentiality

---

<sup>7</sup> Source: The University of Virginia Library, Research Data Services & Sciences, "Steps in the Data Management Lifecycle." <http://data.library.virginia.edu/data-management/lifecycle/>

- Identify potential users of your data
- 2. Project Start Up**
  - Create a data management plan
  - Make decisions about documentation forms and content
  - Conduct pretest of collection materials and methods
- 3. Data Collection**
  - Organize files, backups and storage, frequently asked questions (FAQs) for data collection
  - Think about access, control and security
- 4. Data Analysis**
  - Document analysis and file manipulations
  - Manage file versions
- 5. Data Sharing**
  - Determine file formats
  - Further document and clean data
- 6. End of Project**
  - Deposit data in data archive (repository)

You may also refer to the [Data Life Cycle Infographic, Tool 2](#), included in *Chapter 8*:

### *Create Your Own Policy*

With a solid understanding and interpretation of the policy, your organization can create an overarching internal data management policy. Your DM policy should include documentation of your interpretations.

This policy can serve as a starting point as you develop individual DMPs to address the specifics of each program award. For example, the policy can apply across your organization as general guidelines. Then, for individual awards or groups of awards, you may create tailored DMPs to address specifics of those programs.

If you'd like a bit more assistance in creating your policy, use the policy wizard available through Digital Impact (<https://digitalimpact.io/policies/policy-wizard/>). It includes a series of questions that will prompt generation of boilerplate content that you can customize for your organization and programs.



# Section II

---

## Quality Data Management

### Tools You'll Use...

In this section, you'll be introduced to and use the following tools:

- Tool 3: Digital Curation Centre (DDC) Data Management Plan Questionnaire
- Tool 4: UMN Metadata Codebook
- Tool 5: Stanford File Naming Handout
- Tool 6 Informed Consent Template
- Tool 7: Ethics and Privacy Talking Points
- Tool 8: Incorrect Example of DDL Submission
- Tool 9: Correct Example of DDL Submission



### 3. Improving Your Data Management Practices

Creating a dataset for public release requires greater care in formatting and documenting than generally given to datasets that remain in the hands of the individuals who collect and analyze the data. One tool that can be used to guide the processing of data for public release is a Data Management Plan, or DMP. This Chapter includes guidance, best practice information and tools from a variety of sources, including USAID, colleges and universities, and global research organizations. It describes an array of components that should be considered when developing your DMP.

Most development data generated with USAID funds are generated by USAID's implementing partners. However, many USAID operating units (missions, bureaus, and independent offices) also manage data, especially the data submitted from partners on indicators selected by those missions for reporting in their annual Performance Plan and Report (PPR). Although the guide is written with an emphasis on the work of USAID's implementing partners, it will prove useful for USAID operating units as well.

#### 3.1. Developing a Data Management Plan

Many USAID awards include data collection activities from the start. Award recipients with planned data collection activities should develop a DMP well in advance of the start of these activities. Recognizing that any number of elements may change, the DMP should be viewed as a living document to be updated as often as necessary. It is generally easier to develop documentation of a data collection and analysis activity before the work proceeds rather than to create quality documentation after-the-fact.

In any case, data management planning should be done early. Waiting until a data oriented activity is complete can mean extra work; however, by planning appropriately, data management can make the process of collecting, analyzing and using data much more organized and efficient.

(To understand where a DMP fits within the data submission process, see **Section III.**)

#### 3.2. Elements of a Data Management Plan

The context behind data is very important. To begin, analyze each individual award in terms of your organizational policies on intellectual work and the commitments you have made to produce intellectual work. A DMP should capture and present this context, so that the likelihood of misinterpretation of the data can be reduced. The elements of a DMP presented here help to capture that important context as well as

guide how the data will be treated. This guidance builds on the elements described by the University of Minnesota as well as the University of Arizona (based on the elements developed by the Inter-University Consortium for Political and Social Research as part of their [Framework for Creating a Data Management Plan](#)).

You may also consult the [Data Management Plan Questionnaire, Tool 3](#), found in *Chapter 8*. It is a tool developed by the Digital Curation Centre<sup>8</sup> which is also available online: <http://www.dcc.ac.uk/resources/data-management-plans>.

### ***Data Description***

- What type of data will be produced?
- What data will be collected?
- What is the subject matter scope and geographical scale of the data?
- Who is the primary audience for the data?
- What other audiences might have an interest in the data?
- Are there other existing data that are relevant to what you are collecting?

### ***Access and Sharing***

- How are you planning on archiving, backing up, and sharing your data?
- Why did you choose this method?
- What terms of use do you have, if any?

### ***Data Collection Methodology***

- Who is collecting the data?
- What access or permissions are necessary for that data once it is submitted?
- If enumerators are hired for this task, what confidentiality agreements do they need to sign?
- How will you ensure they hand over all the data that is collected?
- If data will be collected over several days, what steps will you take to secure that data at the end of each day?
- Are there any personal safety concerns to address for your team of enumerators? How will they safeguard technology devices and prevent theft?
- How will the data be collected (survey, extraction from routine data collection systems, direct observation of events, etc.)?
- What technology will be used to collect the data (paper, telephone, laptop, remote sensors, etc.)? To determine which collection methodology might work best, consider these questions:
  - Why are you considering the specific method or technology you have in mind?

---

<sup>8</sup> DCC. (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/data-management-plans>

- How will its data collection approach fit with the workflow you envision for this effort?
- What specific benefits do you see of one approach over the others?
- What is the setting (research site, clinical setting, field) of the interviews?
- If an electronic device will be used, what is the connectivity that will be required during data collection? Can the location support that type of connectivity?
- If paper is being considered, how will the information eventually be stored as electronic information?
- How will data security maintained?
- Who will actually document the data (research coordinator or subject)?
- How will the data be integrated with other electronic research data?
- What are the roles and responsibilities of the data collection team? (See [Tool 3, Data Management Plan Questionnaire](#) for a table of roles and responsibilities.)
- Who will be responsible for the quality control during data collection?
- What is the time period allotted for data collection?

### *Data Analysis Methodology*

- What statistical software (or other analysis software) will be used to manipulate and analyze the data?
- How will data analysis affect de-identification efforts?
- Are there tools or software needed to create/process/visualize the data?
- Will you use pre-existing data? From where?
- How will the data be merged with other existing data relevant to the analysis?
- Who will document the transformation of the data during analysis so that the analysis could be replicated by others?

Frequently, to facilitate the analysis of a dataset, USAID implementing partners may use proprietary software packages to manipulate, analyze and store data; for example, a statistical package such as SPSS or SAS. The requirement to submit data in a non-proprietary format calls for the maintenance of *two* distinct versions of the data set or the conversion of the dataset to a non-proprietary format just prior to submission.

In addition, some USAID implementing partners are concerned that potential users of a dataset may lack familiarity with its structure and, as a result, make errors when converting non-proprietary data to a software package for their own analysis. In instances where a potential user chooses to use the same software as the partner, there is no need for concern because the DDL will make public both the non-proprietary format and the version in the statistical package.

To facilitate use of the data by others as well as to guard against inadvertent error by others handling the data, each dataset should be accompanied by complete documentation including a codebook defining the variables, a data dictionary describing the relationships among the files within a dataset if more than one, and a document describing the methodology used to collect the data and any other relevant contextual information to help others interpret the data correctly. (For an example of a codebook, see [Tool 13, Uganda\\_Agric\\_Practices\\_Codebook\\_wValues](#). Or, [click on the link](#) to view a list of data dictionary examples.)

## ***Metadata***

Metadata captures the details surrounding the data set contents and properties. This description should supplement the title and enable potential users to determine the relevance of the dataset in relation to their own particular interests and help facilitate their appropriate use of the data.

- What types of metadata will be produced to support the data?
- Have you considered how certain types of metadata (such as GPS coordinates) may affect de-identification?
- What descriptors of the data will be most helpful to allow others to understand the nature of the data and to locate in a data archive?
- What standards will be used for documentation and metadata?
  - Consult the Project Open Data Metadata Schema 1.1 (<https://project-open-data.cio.gov/v1.1/schema/>)
  - Or, use the UMN readme.txt template to document your data. See [UMN Metadata Codebook, Tool 4](#) in *Chapter 8*, or view the readme file online: <http://z.umn.edu/readme>
- Is there a good project and data documentation format/standard in existence that you can adopt or customize?
- What directory and file naming convention will be used? For guidance, you may reference the [Stanford File Naming Handout, Tool 5](#), in *Chapter 8*, as well as the Project Open Data Metadata Schema 1.1, which should prevail if there is a conflict between the two sources (<https://project-open-data.cio.gov/v1.1/schema/>)
- What project and data identifiers will be assigned?
- Is there a community standard for metadata sharing/integration?

## ***Intellectual Property Rights***

- Who will own the rights to the data and other information produced by the project (for example, does a partner in the country of origin of the data claim ownership that might inhibit later distribution)?
  - Which projects apply?
  - Do any involve a beneficiary database?
  - Have any evaluations been conducted on the award?

- How many commitments, special studies, annual awards, etc.?
- How will permission be obtained to use and disseminate the data if a question as to ownership exists?
- Will these rights be transferred to another organization for distribution and archiving?
- What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?
  - Do you have a de-identification plan?
  - How will you address a situation where a participant may request their data be returned? (After data has been de-identified and separated?)
- Does your data have any access concerns? Describe the process someone would take to access your data.
- Any special privacy or security requirements (e.g., personal data, high-security data)?
- Any embargo periods to uphold?

### Research Embargo

With regards to research, USAID may embargo, or temporarily withhold from public release for a reasonable period (e.g. 12 months), a dataset that is subject of a pending publication or pending patent application. You must still submit the Dataset to USAID, and with agreement of the COR/AOR, it can be held as non-public until the conclusion of the embargo period.

## Ethics and Privacy

All *human subjects research* (i.e., research on people *not* operational research), should be submitted to the Institutional Review Board (IRB). NGOs are NOT exempt from IRB review and approval. Routine monitoring and evaluation (M&E) activities, however, may not need to be approved by the IRB since they are not defined as research (even though some methodologies are similar). When a program involves human subjects research, informed consent which addresses ethics and privacy should be a priority.

- How is informed consent being handled?
- Do you have an informed consent process in place?
  - If not, do you need to create one? See **Tool 6, Informed Consent Template** in *Chapter 8*.
- How will privacy be protected during the analysis phase and beyond?
- Will the data collection activity be subjected to a review by an Institutional Review Board (IRB) and, if so, which IRB?
- Have you addressed any local/national/international issues related to your award? For example, are you challenged by local implementing partners who may resist sharing data? Do you expect any resistance from patients or participants? Will the local or national government be wary? See the **Ethics and Privacy Talking Points, Tool 7** in *Chapter 8* for guidance on how to work with the mission, the local government and partners.



### ***Format***

- What format(s) will you use for the submission, distribution, and preservation of the data? To note, USAID will eventually move to an online system for submissions. As a best practice for open data, spreadsheets should not be formatted (although additional versions that are formatted may be submitted as supplemental materials). They must also be saved in a csv, non-proprietary format. Samples of both formatted and unformatted spreadsheets ([Tools 8 and 9](#)) appear in *Chapter 8*.
- At least one format should be platform-independent and non-proprietary as required by the policy and to assure that the data will be reusable in the future.
- To facilitate analysis by others likely to be familiar with any proprietary software used during the collection or analysis phase, the data may be submitted for publication in more than one format.

### ***Archiving and Preservation***

- What procedures will you use to ensure long-term archiving and preservation of your data?
- How will the relevance of the data change over time (will the relevancy of the data to the subject diminish over time or disappear completely)?
- What are the budget costs of preparing data and supporting documentation?
- How will the data be archived for preservation and long-term access?
- How long should it be retained (e.g., 3-5 years, 10-20 years, permanently)?
- What file formats? Are they long-lived?
  - Are there data archives that are appropriate for your data (subject-based or institutional)?
- Who will maintain the data for the long-term?
- Who will ensure that any updates to the data are subsequently uploaded in the DDL?
- Do data codes need to be destroyed to prevent identification? If so, what process and procedures will you use?

### ***Data Sharing and Reuse***

- Because you will upload your data to the DDL for others to reuse, how will the data be discovered? And shared?
  - Who is the audience for reuse? Who will use it now? Who will use it later?
- When will you publish it and where?
- Any tools or software needed to work with the data?

### ***Storage and Backup***

- Will the data be reproducible? What would happen if it got lost or became unusable later?
- How much data will it be, and at what growth rate? How often will it change?
- Where and how will you store your data to ensure their safety (several copies are recommended)?
- How will data be managed during the project (Include information about version control and file-naming conventions)?

### **3.3.Preparing to Comply**

You now have an overarching Data Management Policy for your organization, as well as a Data Management Plan for at least one award. Armed with this information, you can prepare your organization for compliance. In essence, you need to decide what your organization and your team must do in order to follow through on all of the steps and activities required by your internal policies and your DMP.

For example, if you will use a proprietary software program to collect and organize data, or a commercially available program such as SAS, then you must decide how and when you will create duplicate data spreadsheets for submission to the DDL. Or, if you need to address sensitive informed consent issues, how and when will you tackle that? As you can see, preparing to comply will require some thought, time and resources.



# Section III

---

## A Step-by-Step Process: Proposal Through Submission

### Tools You'll Use...

In this section, you'll be introduced to and use the following tools:

- Tool 10: DDL Compliance Checklist
- Tool 11: DDL Registration/Submission Form

This section includes a variety of useful information, tools and best practices. It is designed to help you structure an efficient workflow to support Open Data Policy compliance. This guidance is intended to serve as a foundation that can be customized to the unique needs and activities of your organization. Ultimately, we hope it provides a solid starting point for your efforts and simplifies the overall process.

Ideally, you should read this entire section before you implement any plans so that you gain an appreciation for the entire process – the continuum of compliance. It will also enable you to make decisions with greater foresight.

### ***A Birdseye View of the Process***

This Section (*Chapters 4 and 5*) presents the activities associated with these steps:

1. Respond to a USAID solicitation (See *Chapter 4*)
2. Develop your proposal (See *Chapter 4*)
3. Receive the award (See *Chapter 4*)
4. Prepare and track datasets (See *Chapter 5*)
5. Register and submit data to the DDL (See *Chapter 5*)

## 4. Step-by-Step: Dataset Award Negotiations and Communications

Before you even begin, there are several key terms and definitions to understand as you tackle the open data compliance process. They were touched upon in earlier chapters but are defined at a high level below. More detailed descriptions can be found in *Chapter 8*.

- **Dataset** is defined as an organized collection of structured data, including data contained in spreadsheets, whether presented in tabular or non-tabular form. Data collected as part of surveys, performance monitoring or evaluation, and any project data used in reporting to USAID. Dataset in this context *does not* include qualitative data, aggregated performance reporting data, and unstructured data (email messages, PDF files, PowerPoint presentations, word processing documents, photos and graphic images, audio files, collaboration software, and instant messages).
- **A Dataset Collection** is a group of related dataset spreadsheets. USAID's ultimate vision is that an end user could conduct a search across a group of affiliated collections.
- **Intellectual Work** is defined as works that document the implementation, monitoring, evaluation, and results of international development assistance activities developed or acquired under a federal award. This may include program and communications materials, evaluations and assessments, information products, research and technical reports, and articles and papers, whether published or not.
- **Supporting Documentation** is defined as information describing the dataset, such as code books; data dictionaries; data gathering tools, templates and forms; data gathering methodologies; notes on data quality; and explanations of redactions<sup>9</sup>. Supporting documentation defines the fields within the dataset and any categories or labels within the dataset that may require explanation to an individual not familiar with the data.

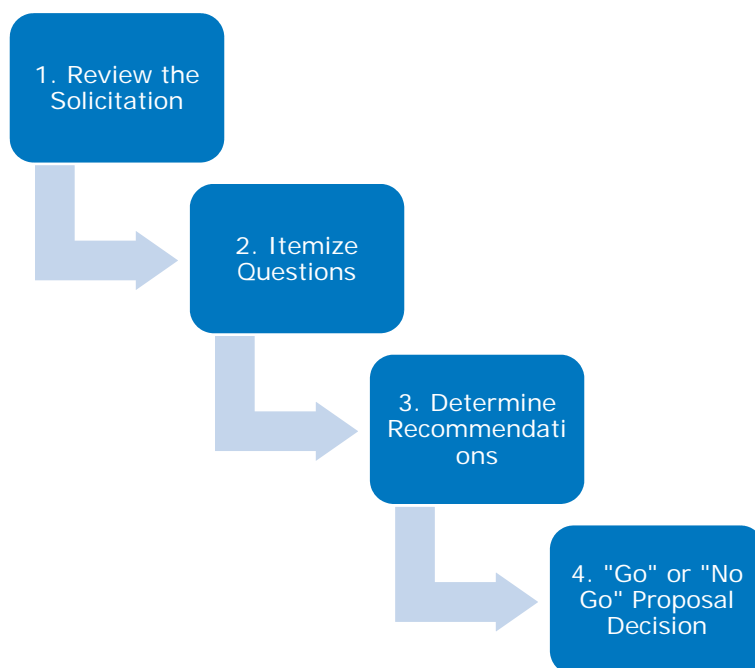
---

<sup>9</sup> Redaction means the censoring or obscuring of part of a text for legal or security purposes.



## 4.1.Responding to a USAID Solicitation

Once USAID puts a call out for grant proposals, you should begin to think about the issue of open data compliance. This proactive approach will save you time and money and allow you to be more efficient and organized throughout the process.



1. Review the Solicitation.
  - Consider creating a proposal team that includes a Proposal Manager (or Grants Coordinator), a Principal Investigator or Research Advisor, and a Technical Advisor. (See *Chapter 7, Section 7.1, Summary of Roles and Responsibilities*.)
  - Identify likely types of data collection.
  - Confirm the date when questions are due to USAID.
2. Itemize any questions related to the Solicitation.
  - Submit your questions prior to the deadline.
3. Determine whether or not your proposal team will make recommendations for:
  - A specific access level ("restricted public" or "non-public") to be applied to select or all datasets;
  - An exemption from dataset transmission;
  - An appropriate timeframe for submission; or
  - An embargo to temporarily withhold data from the public.

See subsection below for Dataset Access Level Definitions as well as *Chapter 6, Section 6.2, Exceptions, Embargos & Submissions* and *Chapter 8, Section 8.6. Exceptions Language* for samples to support each of these scenarios. This language can be included in your set of solicitation questions.

**To note**, if you are unsure of whether or not your data will meet ethical standards for release, do not hesitate to choose the most restrictive level of access when submitting and be sure to include the rationale for that restrictive access decision.

4. Go or No Go Decision – Based on the USAID responses to questions or if USAID does not accept your organization's recommendations, your team will need to decide whether or not to submit a proposal.

### *Dataset Access Level Definitions*

Datasets must be documented with one of these three access levels as mandated by the [Open Data Policy \(Section III.4\)](#), in consultation with [Project Open Data](#):

- **Public:** Dataset is or could be made publicly available to all without restrictions. For the "Public" access level, your team may choose to document technical or resource barriers to increasing access to this Dataset.
- **Restricted Public:** Dataset is available under certain use restrictions. One example, among many, is a dataset that can only be made available to select researchers under certain conditions, because the dataset contains sufficient granularity or linkages that make it possible to re-identify individuals, even though the dataset is stripped of Personally Identifiable Information (PII). Another example would be a dataset that contains PII and is made available to select researchers under strong legal protections. This category includes some but not all datasets designated as Controlled Unclassified Information (CUI), consistent with Executive Order 13556. For the "Restricted Public" access level, you must document the conditions under which select individuals may obtain access to this dataset.
- **Non-Public:** Dataset is not available to members of the public. This category includes datasets that are only available for internal use by the Federal Government, such as by a single program, single agency, or across multiple agencies. This category might include some but not all datasets designated as CUI, consistent with Executive Order 13556. Some non-public datasets may still potentially be available to other intra-agency departments or other government agencies, as discussed in OMB Memorandum M-11- 02: Sharing Data While Protecting Privacy. For the "Non-Public" access level, you must document your rationale for why the data cannot be made public.

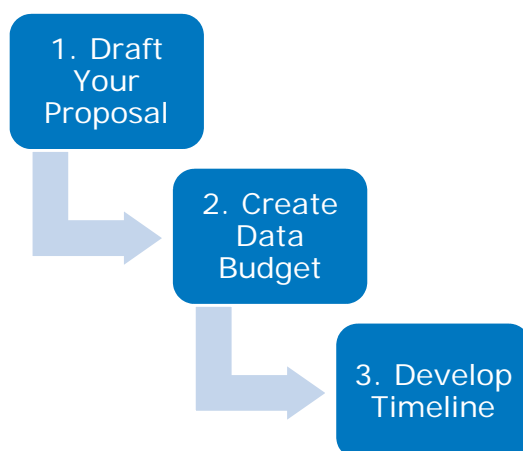
You will propose an access level for each dataset submission and include a rationale for proposing either a restricted public or non-public designation. The Operating Unit will weigh in on that proposal, with the legal officer making the final determination after consideration of all factors.

Your team will want to take time to understand all access angles – how best to protect and to share your data – as you craft your rationale.

For more details, also see *Chapter 8, Section 8.5. Dataset Access Levels Defined*.

### 4.2. Developing Your Proposal

In addition to drafting your proposal narrative, you should consider all of the resources (costs and human capital), activities and timeframes associated with compliance.



1. Draft your proposal. As you draft your proposal response, it's not too early to think about the data management practices you'll put into place if you win the award!
2. Create a Data Budget to address the costs associated with preparing project data for submission, according to the approved recommendations. (See *Chapter 7, Section 7.2, Budgeting for Your Data Management Plan* for guidance on preparing a data management budget.)
3. Develop a realistic timeframe for dataset submission.
  - Be sure to draft a rationale for any unique exceptions that may affect your timeline (i.e., exemption status, embargo period, or submission timeframe) and include it in your proposal.

### 4.3. Receiving the Award



Congratulations! If your organization is selected to receive the award, you can roll up your sleeves and apply quality data management practices to your program efforts.

1. Develop a DMP – At the start of the award, develop a DMP to address the specific dataset submission requirements (required levels of restriction, controls set up to ensure the validity of the data, research publication, etc.). (See *Section II, Chapter 3* for guidance on how to develop a DMP.)
  - As you develop your DMP, be sure your team discusses critical issues such as working with local or national governments, de-identifying data, securing informed consent, and structuring the data correctly for proper submission.
2. Negotiate and communicate data specifics. This involves reviewing the DMP with the COR/AOR and providing specifics regarding the type of datasets you will submit.
  - Since your proposal included much of this information, this may simply be a confirmation or validation of the specifics you presented. Or, it may involve some negotiations between your team lead(s) and COR/AOR until you reach agreement.
  - If changes are necessary, you should update your DMP. Remember, it is a living document that is updated and adapted throughout the Award in collaboration with the COR/AOR.
  - Review the **DDL Compliance Checklist, Tool 10** in *Chapter 8* to be sure that you have considered all types of data and intellectual works.
3. Develop a risk mitigation plan which details how you will de-identify data. Before any data collection activities commence, outline how personal information will be safeguarded and how data will be made anonymous. (See *Chapter 6, Section 6.3 De-Identification of Data* for tips.)
  - Prior to submission to the DDL, datasets should be stripped of PII, such as names or birthdates.
  - Specify your process for de-identifying data. Will you hire an outside vendor to assist your team?
4. Secure Informed Consent. If your research involves human subjects, you must obtain “informed consent” from all participants, or from the subject’s legally authorized representation, prior to any data collection activities.
  - The IRB must review and approve all research that involves human subjects and the informed consent process before research can begin.
5. Inform your partners, including award sub-recipients, about the need for all parties to comply with USAID’s open data policy.

### Tip! Informed Consent

During the informed consent process, research participants should be informed about the confidentiality of the information they share with researchers. It is the responsibility of the research lead to ensure information about how the data will be used, and who will have access to it, is include in consent forms. Please see your IRB Administrator for questions regarding informed consent.

Also see *Chapter 6, Section 6.1 Securing Informed Consent* and **Tool 6, Informed Consent Template**.

- Share your DMP with them.
- Clarify your expectations, as well as roles and responsibilities of each team member.



## 5. Step by Step: Dataset Preparation and Submission

After you have reached agreement with the COR/AOR, your team can begin to utilize your award and focus on your project. As you do, it is critical to be intentional in the ways that you collect and prepare data. It will be far less time-intensive – and much less hassle – to have your data organized from the onset in a way that will facilitate compliance, rather than having to organize it after the fact.

Submissions to the DDL are done by a two-step process, which includes “registering” the data first, and then “submitting” the data.

The open data requirements from the White House include a notation that data must be in a non-proprietary format. That means, spreadsheets should be saved as a “csv” rather than an “xlsx” file version. One unfortunate consequence of this requirement is that spreadsheets **should not** have any tabs (multiple sheets within one spreadsheet file). To view an example of an unacceptable, formatted spreadsheet see [Tool 8](#); to view an example of an acceptable spreadsheet see [Tool 9](#).

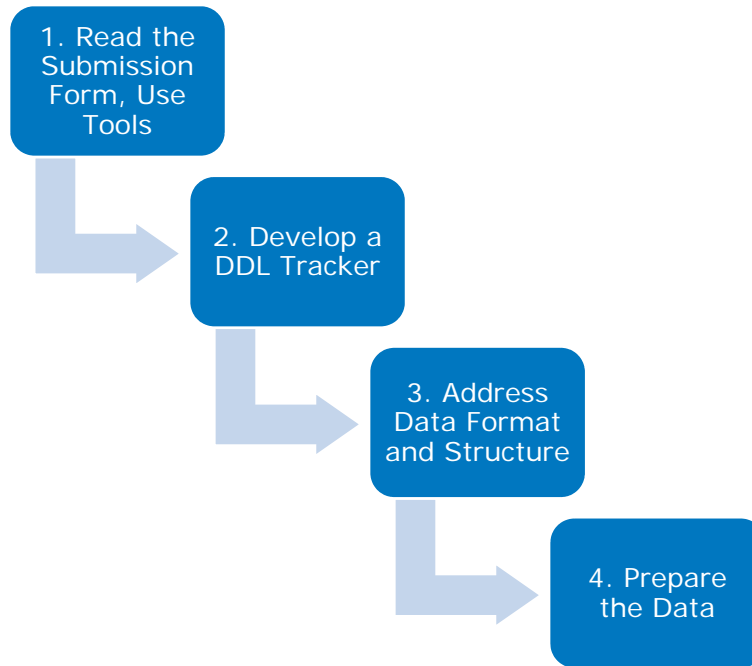
USAID encourages partners to submit more than one format for a dataset when appropriate. For example, according to USAID, when the analysis is done using a statistical program such as SPSS, the partner should convert that SPSS file to a csv, ingest the csv into the platform and attach the SPSS version as an external attachment.

### Tip!

Take some time to become familiar with the DDL Registration/Submission Form. Understanding the information that is being requested will help you streamline your efforts. (Reference [Tool 11](#), [DDL Registration/Submission Form](#), in *Chapter 8*.)

## 5.1.Preparing and Tracking Datasets

Preparing and tracking your datasets will be some of the most important activities your team handles in relation to open data compliance.



1. Use the **DDL Compliance Checklist (Tool 10)** and read the **DDL Registration/Submission Form (Tool 11)** to guide each step of the data capture and preparation process.
  - The best way to become prepared and to understand what you will need to submit is to review the submission form. The most current version can be located online: <https://www.usaid.gov/data/ddlsubmissions>
  - Be sure to communicate any changes to evaluation design, special studies, or any other important details that may have already been communicated to the COR/AOR and approved.

### Who Does What?

If your organization has not done so already, determine who will be responsible for data ownership. This person will spearhead all of the registration and submission activities for your program. They should also be responsible for ensuring that you are complying with USAID policy and with your own organization's DMP. If discrepancies are found, they would work to resolve them appropriately.

You may reference the table in *Chapter 7, Section 7.1* on Roles and Responsibilities.

- Additionally, use the DDL Compliance Checklist, which further guides your efforts to collect and organize all of the information you'll need to comply.
- 2. Develop a DDL Tracking System
  - Develop a DDL Tracking System (See the [Tool 10 DDL Compliance Checklist](#), 1.1 item and also [Tool 17 Sample DDL Tracker](#)).
  - Consider a tracker that's both project based and which is integrated across your entire organization.
- 3. Address data format and structure.
  - Create contextual descriptions and metadata, as well as code books and dictionaries.
  - Document any other details that you think a user would need to know to understand your data and not misuse it.
  - Write a "Read Me" file specifying "*What you should know. . .*" that a user could reference before viewing or downloading your data.
  - Develop meaningful spreadsheet column headers.
  - Use an intuitive file naming structure for your spreadsheets.
- 4. Prepare your data. Preparing your data involves making sure that you have all of the materials you will submit – such as spreadsheets, code books, and reports – and that those items meet the requirements specified by USAID.
  - Address/confirm you have codes to the data and that the identifiable data has been (or will be) stripped. (See *Chapter 6, Section 6.4* for tips on de-identifying data.)
  - Develop a plan for data entry and cleaning it.
  - If you are unsure as to whether your data has been sufficiently cleaned, there is a question on the registration form (Does this meet [USAID quality standards](#)? See page 16 of the ADS Chapter 579 pdf file.) In that field, you can include your concerns and explanation.
  - **To note**, many surveys are not useful if they are stripped of identifiers. In those cases, the data can be registered to the DDL with the recommendation that it not be submitted. USAID will then make a determination.

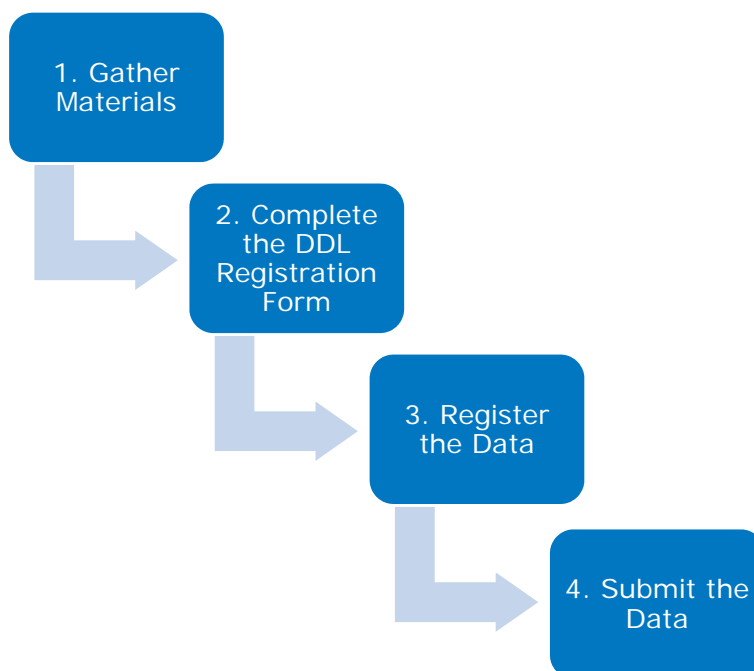
### Tip!

Many users may potentially access the spreadsheets you upload to the DDL. If your project utilizes a meaningless file naming convention – such as *11057-LOT.csv* – it will be difficult for users, and perhaps even for your own team members, to navigate the repository and find information. Instead, consider contextual filenames, such as *Premie\_Wheat\_Allergy\_Sept17Survey*. Similarly, choose column header titles that will make sense for the layperson or create a detailed codebook to help users decipher materials.

## 5.2. Registering and Submitting Data to the DDL

Uploading data to the DDL is a two-step process. It begins with “registering” your data with USAID using an online form that captures pertinent information about the data you plan to submit. The Registration step alerts the DDL team that a dataset exists (although no data will be attached.) Upon receiving the alert, USAID staff activates a secure file transfer application and responds to the submitter with instructions for transmitting the data.

That does not mean that the data is publically available quite yet. It first undergoes a clearance process and review by several entities, including USAID’s Privacy, FOIA and Legal Counsel offices. Once returned, it is formatted for public access (this can take 6 weeks to complete).



1. Create and gather supporting documentation and materials
  - During award negotiation or once you realize there will be data collection activities, provide the COR/AOR specifics regarding the type

### Raw Data vs. Aggregated Data

USAID does not want items submitted to the DDL that are submitted to the Mission. Often partners submit already aggregated data, rather than the raw data. The Mission is reporting on an aggregate across all of its partners. If each partner is doing a primary data collection that underlies the indicator being reported to the Mission, that underlying indicator should likely be reported to the DDL.

of datasets you intend to collect and submit. Then, before you complete and submit your registration form, email the COR/AOR to notify them about the list of datasets you will be registering and request confirmation that they received your email. Once the COR/AOR has acknowledged your request, you may complete the DDL Registration Form.

- Use the **DDL Registration/Submission Form (Tool 11)** as a guide for gathering all of the information you will need. As previously noted, materials will include data dictionary, code book, user manuals, syntax files, the documentation of data cleaning process, etc. (Also see **DDL Compliance Checklist, Tool 10, and the Sample DDL Tracker, Tool 17**)

- Assign Access Levels. You should have specified access levels during the grant award agreement phases. If you have data that has since been deemed sensitive, you must indicate this during the Registration process and provide a rationale for the change. You may also include any concerns you have regarding access levels in the comments section of the [Data Registration Form](#) (e.g., informed consent concerns or data quality issues). (See *Chapter 8, Sections 8.5 and 8.6* for more details on Dataset Access Levels). Similarly, if you believe there is now a compelling reason to not transmit data to USAID, you should note this in the “Access Level Comment” field when initially completing the DDL Registration Form. This information will then be routed to the COR/AOR, who will make the determination.

### 2. Register the data.

- As previously mentioned, Data Registration and Submission is a two-step process; with Registration being the first step.
- When you are ready to register, complete all of the fields on the form referencing the materials you’ve gathered.

#### Registration vs. Submission

Registration involves entering ALL of your information – including information about your datasets, such as metadata.

The registration process looks very similar to what you might consider a submission! In fact, even the first link you click on is titled: DDL Submission Form.

Think of it as the “alert” phase when you are letting USAID know that you are readying your project to comply with their Open Data Policy.

#### Tips!

After you register, USAID will send you a link with an expiration date. **You must respond by submitting your data within the allotted timeframe (30 days)**; again, having all your materials on hand when you do.



- Once you register the data, USAID will review and respond within 24 hours or so by sending you a link. You can then follow their instructions and submit your data based on their response.
  - The link they send will expire if you do not respond within the allotted period of time. So again, be sure to have all of your data ready for submission.
2. Submit the data
- Once you receive the approval from USAID, you must submit your datasets.
  - In some instances, USAID may have questions or additional requests which you will need to address before final submission.
3. Address award closure data requirements
- Within thirty (30) calendar days after award completion, you must register/submit all datasets and supporting documentation that have not previously been submitted to the DDL, along with an index of all datasets and intellectual works created or obtained under the award and an itemized list of any and all DDL submissions. You should also provide the itemized list of DDL submissions to your Compliance Unit.
  - In accordance with the mandatory standard provision, your Compliance Unit will submit the itemized list to the COR/AOR.

## 5.3.What Does Success Look Like?

In *Chapter 8, Section 8.3*, you will find an example of study data successfully uploaded to the DDL. Titled, "Baseline Study of Food for Peace Title II Development Food Assistance Program in Uganda," the snapshot shows information that is available to the public through the DDL, including the study's title, its description, metadata, uploaded data ([Dataset Spreadsheet and Data Dictionary, Tools 12 and 13](#)) and the [Program Reference Report \(Tool 14\)](#).



# Section IV

---

## Other Contextual Considerations

### Tools You'll Use...

In this section, you'll be introduced to and use the following tools:

- Tool 12: Uganda Agric Practices Codebook (Example)
- Tool 13: Uganda Agric Practices Data (Example)
- Tool 14: Case Study Reference Material
- Tool 15: Requesting Compliance Guidance from Prime (Email Template)
- Tool 16: UK Costing Tool
- Tool 17: Sample DDL Tracker

## 6. Ethics, Ownership and Working with Partners

While it's likely your organization is fully committed to data transparency and openness, you may have concerns that particular provisions of the USAID Open Data Policy may compromise the ethical integrity of your research conducted with human subjects or jeopardize your research partnerships with host countries.<sup>10</sup>

Here are some key concerns:

- *Confidentiality and Beneficiary Protection* – Often, participants provide informed consent to have their responses used for specific purposes and by a specific group of people (e.g., the research team). However, under USAID's Open Data Policy, participant information may now be accessible to the general public, which raises concerns about how to ensure respondents are truly informed about the ways in which their information may be used. In addition, such a disclosure may impact data quality and a respondent's willingness to engage in data collection activities.
- *De-identification and Security Issues* – In some instances, de-identified datasets (those from which personally identifiable information has been redacted) may still contain information that would allow participants to be (re)identified. Ensuring adequate security of data storage should also be a consideration.
- *Partnerships* – When work is partially funded by another donor, or data is owned by a host country, USAID officials have indicated that it will be your responsibility to inform such parties of the obligation to submit and share data collected with USAID funds. If, however, you believe this could potentially threaten a partnership, you are encouraged to reach out to the

### Ethical Considerations

According to the UK Data Archive, the key principles of research ethics that have a bearing on sharing or archiving confidential research data are:

- A duty of confidentiality towards informants and participants
- A duty to protect participants from harm, by not disclosing sensitive information
- A duty to treat participants as intelligent beings, able to make their own decisions on how the information they provide can be used, shared and made public (through informed consent)
- A duty to inform participants how information and data obtained will be used, processed, shared, disposed of, prior to obtaining consent
- A duty to wider society to make available resources produced by researchers with public funds (data sharing required by research funders)

<sup>10</sup> The International Rescue Committee provided input to help shape this list of concerns.

COR/AOR and consult with USAID since compliance requires that at a minimum datasets be registered.

## 6.1.Securing Informed Consent

The United Kingdom's Data Archive is that country's version of USAID's DDL. The UK Data Archive website contains a variety of important reference information (see *Chapter 8, Section 8.2* for these useful links). According to the UK Data Archive, written consent documentation typically includes an information sheet and a consent form signed by the participant. This allows background information to be as detailed as necessary, while keeping the signature form short and concise.<sup>11</sup> The UK Data Archive's guidance is presented below.

An **information sheet** should cover the following topics:

- Purpose of the research
- What is involved in participating
- Benefits and risks
- Terms for withdrawal:
  - Participants have a right to withdraw at any time without prejudice and without providing a reason
  - Thought should be given to what will happen to existing, already provided, data in the event of withdrawal
- Usage of the data:
  - During research
  - Dissemination
  - Storage, archiving, sharing and re-use of data
- Strategies for assuring ethical use of the data:
  - Procedures for maintaining confidentiality
  - De-identifying data where necessary, especially in relation to data archiving
- Details of the research:
  - Funding source
  - Sponsoring institution
  - Name of project
  - Contact details for researchers
  - How to file a complaint

The **consent form** should be written in plain language free from jargon. It should allow the participant to clearly respond to each of these points:

- The participant has read and understood information about the project
- The participant has been given the opportunity to ask questions
- The participant voluntarily agrees to participate in the project

---

<sup>11</sup> UK Data Archive (1967-2017): <http://data-archive.ac.uk/create-manage/consent-ethics/consent?index=3>

- The participant understands that he or she can withdraw at any time without giving reasons and without penalty
- Procedures regarding confidentiality are explained (use of names, pseudonyms, de-identifying data, etc.)
- Separate terms of consent for interviews, audio, video or other forms of data collection
- Use of the data in research, publications, sharing and archiving are explained and whether it will be made available online or on your organization's website, for example.
- Signatures and dates of signing for the participant and the researcher

The participant should be given a copy of the form and the researcher should retain the signed original. (You may reference **Tool 6, Informed Consent Template**, in Chapter 8 to help you create your own consent form.)

You may also review the templates and resources that the USAID Privacy Office uses to assess potential privacy issues, which can be found on their website: <https://www.usaid.gov/who-we-are/agency-policy/series-500/references-chapter>.

## 6.2.Exceptions, Embargos & Submissions

Some scenarios may point to “exceptions, embargos and submissions.” These include situations where your organization may decide that you want to hold your data back from submission. To do so, you must communicate with USAID as to your reasons and your rationale. (See *Chapter 8, Section 8.6 Exception Language Samples*.)

If you believe that your data collection activities for a country program fall under one of the exception categories (e.g., privacy concerns, timeframe, access level issues, embargo scenarios), your project lead should:

1. Meet with the respective COR/AOR to discuss the case;
2. Request (in writing) that the COR/AOR grant an exception, citing one (or more) of the exception criteria (See the policy documentation itself, ADS 579, Section 3.2.3). You must present a concrete justification.

Additionally, you may submit a request to the COR/AOR to temporarily withhold research data from public release for a reasonable period of time (e.g., 12 months), citing section 3.3.3 of ADS 579, Embargos on Data Publication.

For monitoring data that is continually updated (quarterly or more frequently), USAID has recommended that you speak with the COR/AOR to determine an appropriate timeframe within the DMP to submit data. USAID has recognized that asking awardees to submit such data 30 days after submitting quarterly progress reports is not reasonable. As a result, this data should be considered incomplete and not ready for public access or use.



If at any time you are unclear regarding the policy requirements, it is advised that you seek clarification from the COR/AOR. Any exceptions, embargos, contract amendments or other agreements between the COR/AOR and your organization should be documented in writing and stored within the program files.

## **6.3.Data Ownership**

Under the standard terms and conditions in USAID awards, the implementing partner owns the data. However, the federal government has the right to obtain and use the data. The Bayh-Dole Act of 1980 changed the way the USG addressed intellectual property developed under awards. In the interest of spurring private sector innovation, the Bayh-Dole Act required that the USG provide intellectual property rights in data developed with USG funding to small businesses and non-profits, with the USG reserving a license. Bayh-Dole was later extended to all USG partners. The standard clauses in the Federal Acquisition Regulations, as well as provisions for Assistance, stipulate that partners have ownership rights, with a broad license provided to the USG. Under both acquisition and assistance, USAID is generally provided a license to obtain and use the data produced under an award.

## **6.4.De-Identification of Data**

This section presents the UK's Data Archive's guidance for de-identification of data.<sup>12</sup>

Before data obtained from research with people can be shared with other researchers or archived, you may need to de-identify them so that individuals, organizations or businesses cannot be identified. Here we provide guidance on de-identifying quantitative data appropriately in order to retain as much meaningful information as possible.

Re-users of data have the same legal and ethical obligation to NOT disclose confidential information as primary users. De-identification may be needed for ethical reasons to protect people's identities in research, for legal reasons to not disclose personal data, or for commercial reasons.

Personal data should never be disclosed from research information, unless a respondent has given specific consent to do so, ideally in writing.

In some forms of research, for example where oral histories are recorded or in anthropological research, it is customary to publish and share the names of people studied, for which they have given their consent.

---

<sup>12</sup> UK Data Archive (1967-2017): <http://data-archive.ac.uk/create-manage/consent-ethics/anonymisation>

Procedures to de-identify data should always be considered alongside obtaining informed consent for data sharing or imposing access restrictions.

A person's identity can be disclosed from:

- **Direct identifiers** such as names, addresses, postcode information, telephone numbers or pictures
- **Indirect identifiers** which, when linked with other publicly available information sources, could identify someone, e.g. information on workplace, occupation or exceptional values of characteristics like salary or age

Direct identifiers are often collected as part of the research administration process but are usually not essential research information and can therefore easily be removed from the data.

De-identifying research data can be time consuming and therefore costly. Early planning can help reduce the costs.

De-identifying techniques for quantitative data may involve removing or aggregating variables or reducing the precision or detailed textual meaning of a variable. Special attention may be needed for relational data, where connections between variables in related datasets can disclose identities, and for geo-referenced data, where identifying spatial references also have a geographical value.

*Qualitative data does not need to be submitted to the DDL and, therefore, does not need to be de-identified for these purposes.*

## ***UK Data Archive Quantitative Data De-Identification Examples***

**Remove direct identifiers from a dataset** such as detailed personal information. Such identifiers are often not necessary for secondary research.

***Example:** Remove respondents' names or replace with a code. Remove addresses, postcode information, institution and telephone numbers.*

**Aggregate or reduce the precision of a variable** such as the respondent's age and place of residence. As a general rule, report the lowest level of geo-referencing that will not potentially breach respondent confidentiality. The exact scale depends on the type of data collected, but very detailed geo-references like full postcodes, wards or names of small towns or villages are likely to be problematic. Coded or categorical variables which may be potentially revealing can be aggregated into broader codes.

***Example:** Record the year of birth rather than the day, month and year; record postcode sectors (first 3 or 4 digits) rather than full postcodes; aggregate detailed 'unit group' standard occupational classification employment codes up to 'minor group' codes by removing the last digit.*

**Generalize the meaning of a detailed text variable** by replacing potentially disclosive free-text responses with more general text.

***Example:** Detailed areas of medical expertise could indirectly identify a doctor. The expertise variable could be replaced by more general text or be coded into generic responses such as 'one area of medical specialty', 'two or more areas of medical specialty', etc.*

**Restrict the upper or lower ranges of a continuous variable to hide outliers** if the values for certain individuals are unusual or atypical within the wider group researched. In such circumstances the unusually large or small values might be collapsed into a single code, even if the other responses are kept as actual quantities, or one might code all responses.

***Example:** Annual salary could be 'top-coded' to avoid identifying highly paid individuals, even if lower incomes are not coded into groups.*

**De-identify relational data** where relations between variables in related or linked datasets or in combination with other publicly available outputs may disclose identities.

***Example:** In confidential interviews on farms the names of farmers have been replaced with codes and other confidential information on the nature of the farm businesses and their locations have been disguised to de-identify the data.*

*However, if related biodiversity data collected on the same farms, using the same farmer codes, contain detailed locations for biodiversity data alone the location would not be confidential. Farmers could be identified by combining the two datasets.*

*The link between farmer codes and biodiversity location data should be removed, for example by using separate codes for farmer interviews and for farm locations.*

## **UK Data Archive Qualitative Data De-Identification Examples**

*Again, qualitative data does not need to be submitted to the DDL. However, the UK's Data Archive offers useful guidance for de-identifying text-based information.*

Often researchers presume that participants want their data destroyed or kept confidential and inaccessible. However, they might be quite willing to have their data shared with other researchers if appropriate pseudonyms and other protections are provided.

If you need to de-identify textual data, information should not simply be removed or blanked-out. Pseudonyms, replacement terms, vaguer descriptors or systems of

coding should be used to retain maximum content. Best practices for de-identifying text are:

- Do not collect disclosive data unless this is necessary, e.g. do not ask for full names if they cannot be used in the data
- Plan de-identification at the time of transcription or initial write up, except for longitudinal studies where relationships between materials may require special attention during editing
- Use pseudonyms or replacements that are consistent within the research team and throughout the project, e.g. use the same pseudonyms in publications or follow-up research
- Use 'search and replace' techniques carefully so that unintended changes are not made, and misspelled words are not missed
- Identify replacements in text clearly, e.g. with [brackets] or using XML tags such as <seg>word to be de-identified</seg>
- Retain unedited versions of data for use within the research team and for preservation
- Create a de-identification log of all replacements, aggregations or removals made - store such a log separately from the de-identified data files

## Example De-Identification Log

Interview / Page #	Original	Changed to
<b>Int1</b>		
p1	Age 27	Age range 20-30
p1	Spain	European country
p3	Manchester	Northern metropolitan city or English provincial city
p2	20th June	June
p2	Amy (real name)	Moirra (pseudonym)
<b>Int2</b>		
p1	Francis	my friend
p8	Station Road primary school	a primary school
p10	Head Buyer, Produce, Sainsburys	Senior Executive with leading supermarket chain

## Example

In an interview transcript, a person's name is replaced with a pseudonym or with a tag that typifies the person [farmer Bob, paternal grandmother, council employee]. This is also done when reference is made to other identifiable people. An exact geographical location may be replaced with a meaningful descriptive term that

typifies the location [southern part of town, near the local river, a moorland farm, his native village].

(Chapter 8, Section 8.2 of this Guide contains online links to more information on both qualitative and quantitative anonymization.)

## **6.5. Working with Partners**

Working closely with partners, whether they are the prime award recipient or a sub-awardee, is critical to strong data management.

If you are the prime award recipient, use your DMP and this Guide to provide your partner with clear direction on how to comply. If you are a sub-award recipient, you may use [Tool 15](#) as an email template to [Request Compliance Guidance from the Prime](#).

## 7. Staffing and Budgeting

### 7.1. Summary of Roles and Responsibilities

The International Rescue Committee (IRC) developed a list of common roles and responsibilities associated with data management activities, and matched those responsibilities with the typical data management life cycle. (This information is merely an example that you can build upon. It has been edited for use within this guide and appears in the table below.)

Positions include:

- Proposal Lead / Grants Coordinator
- Regional Program Officer
- Research Advisor
- Country Program Staff
- Technical / Regional Director
- Regional Measurement Action Coordinator

POSITION TITLE	RESPONSIBILITIES
<b>Proposal Lead / Grants Coordinator</b>	<p><i>When USAID call for proposals is issued:</i></p> <ul style="list-style-type: none"> <li>• Send solicitation to assigned proposal team members.</li> <li>• Submit questions to donor.</li> </ul> <p><i>During proposal development:</i></p> <ul style="list-style-type: none"> <li>• Ensure recommendations on restricted access, embargo periods, submission timeframes, and exemptions are included in the proposal.</li> </ul> <p><i>At agreement review:</i></p> <ul style="list-style-type: none"> <li>• Confirm Office of Development (ODP) requirements in agreements reflect your organization's assessment of the proper access level, exemption status, embargo period, and submission timeframe for the data; flag issues.</li> <li>• Send a notice to all partners, including sub-grantees, about the need to comply with USAID's Open Data Policy.</li> </ul>
<b>Regional Program Officer</b>	<p><i>At agreement review:</i></p> <ul style="list-style-type: none"> <li>• Confirm ODP requirements in agreements reflect your organization's assessment of the proper access level,</li> </ul>



POSITION TITLE	RESPONSIBILITIES
	<p>exemption status, embargo period, and submission timeframe for the data; flag issues.</p> <ul style="list-style-type: none"> <li>• Manage requests for agreement changes.</li> </ul>
<b>Technical Advisor</b>	<p><i>When USAID call is issued:</i></p> <ul style="list-style-type: none"> <li>• Review solicitation (or concept note) and make recommendations about access level, exemptions, submission timeframes, and embargos.</li> </ul> <p><i>During proposal development:</i></p> <ul style="list-style-type: none"> <li>• Advise on activity/budget inputs required to prepare the project data for submission and draft relevant language to be included in proposal.</li> </ul> <p><i>After award granted:</i></p> <ul style="list-style-type: none"> <li>• Develop DMPs in collaboration with country program staff.</li> </ul>
<b>Research Advisor</b>	<p><i>When USAID (research) call is issued:</i></p> <ul style="list-style-type: none"> <li>• Review solicitation (or concept note) and make recommendations about access level, exemptions, submission timeframes, and embargos.</li> </ul> <p><i>During (research) proposal development:</i></p> <ul style="list-style-type: none"> <li>• Advise on activity/budget inputs required to prepare the project data for submission and draft relevant language to be included in proposal.</li> </ul> <p><i>After (research) award granted:</i></p> <ul style="list-style-type: none"> <li>• Provide technical support, as needed, on DMPs and de-identifying data (for research).</li> <li>• Ensure information about how data will be used, and who will have access, is include in consent forms.</li> </ul>
<b>Country Program Staff</b>	<p><i>After award granted:</i></p> <ul style="list-style-type: none"> <li>• Develop DMPs in collaboration with team members.</li> </ul>

POSITION TITLE	RESPONSIBILITIES
<b>Technical / Regional Director</b>	<p><i>When USAID call is issued:</i></p> <ul style="list-style-type: none"> <li>Determine whether to pursue a proposal if USAID does not grant recommendations around access level, exemptions, submission timeframes, or embargos.</li> </ul>
<b>Regional Measurement Action Coordinator</b>	<p><i>After award granted:</i></p> <ul style="list-style-type: none"> <li>Provide technical support, as needed, on developing DMPs and de-identifying data (for measurement).</li> </ul>

## 7.2. Budgeting for Your Data Management Plan

Senior USAID officials have indicated that if significant additional resources are needed to comply with the DDL policy (e.g., to de-identify data), award recipients should include the appropriate funding in the grant requirements. This section presents an edited version of Best Practices from the Data Observation Network Earth (DataONE) for budgeting the costs associated with data management.<sup>13</sup>

As a best practice, a partner must first acknowledge that the process of managing data will incur costs. Partners should plan to address these costs and the allocation of resources in the early planning phases of the project – as early as the proposal process. This best practice focuses on data management costs during the life cycle of the project, and does not aim to address costs of data beyond the end of the project.

Budgeting and costing for your project is dependent upon institutional resources, services, and policies. As a result, you should first consult with your sponsored project office, your office of research, tech transfer resources, and other appropriate entities to understand resources available to you.

There are a variety of approaches to budgeting for data management costs. All approaches should address the following costs in each phase:

- Short-term costs
- Long-term costs
- Internal/external costs

<sup>13</sup> Data Observation Network Earth, DataONE, <https://www.dataone.org/best-practices/provide-budget-information-your-data-management-plan>

- Equipment/services costs (i.e., compute cycles, storage, software, and hardware)
- Overhead costs
- Time costs
- Human resource costs

Then, address outsourced versus internal costs. Internal (or insourced) costs are items that are managed directly within the program group. Outsourced costs are items that are contracted or managed outside of the program group.

### *Phases of the DataONE Data Life Cycle*

DataONE aligns its costing activities with phases of the data management life cycle, and coordinates with its centralized IT services department. Your organization may not have a centralized IT group, but may still need to consider these types of expenses or even the possibility of using external service providers. Use this list to consider a variety of costing options:

- **Collect** - Likely both insourced and outsourced costs. Coordinate with central IT services or community storage resources to ensure appropriate data storage environment and associated costs during this phase or throughout the life of the project.
- **Assure** - Likely insourced costs. This phase is primarily focused on quality assurance/control, and costs will primarily be incurred around time and personnel.
- **Describe** - Likely insourced costs. This phase includes initial and ongoing documentation as well as continuous development of metadata. Documentation captures the entire structure of the project, all configurations/parameters, as well as all processes during the course of the entire project.
- **Deposit** - Likely both insourced and outsourced costs.
- **Preserve** - Likely both insourced and outsourced costs. Coordinate with central IT services or community repository environments that are equipped to provide preservation services. This phase will be tied closely to the costs of the collection phase.
- **Discover** - Likely insourced costs. Coordinate with librarians, IT service providers, or repository providers to identify and access data sources.
- **Integrate** - Likely insourced costs. Coordinate with IT service providers or other service groups to merge and prepare data sources for analysis phase.
- **Analyze** - Likely insourced costs. Coordinate with central IT services or other workspace providers to connect data sources with appropriate analysis and visualization software.

You may also reference [Tool 16, UK Costing Template](#) and [Tool 18: Steps and Tools at a Glance](#).

## 8. Templates, Tools and Resources

In this final chapter, we've compiled the many resources used and referenced throughout this guide.

### 8.1.Master List of Tools

- Tool 1: Policy Discussion Checklist
- Tool 2: Data Life Cycle Infographic
- Tool 3: DCC Data Management Plan Questionnaire
- Tool 4: UMN Metadata Codebook
- Tool 5: Stanford File Naming Handout
- Tool 6: Informed Consent Template
- Tool 7: Ethics and Privacy Talking Points
- Tool 8: Incorrect Example of DDL Submission
- Tool 9: Correct Example of DDL Submission
- Tool 10: DDL Compliance Checklist
- Tool 11: DDL Registration/Submission Form
- Tool 12: Uganda Agric Practices Codebook (Example)
- Tool 13: Uganda Agric Practices Data (Example)
- Tool 14: Case Study Reference Material
- Tool 15: Requesting Compliance Guidance from Prime (Email Template)
- Tool 16: UK Costing Tool
- Tool 17: Sample DDL Tracker
- Tool 18: Steps and Tools at a Glance

### 8.2.Useful Links and Resources

The following online resources were consulted during development of this guide. Each offers more in-depth information that you may find useful.

- Visit the [Open Data Research Network](#) and the [Open Data Institute](#) for information regarding **the impact of open data in developing countries**.
- The **Data Management Policy Wizard** is available online through Digital Impact (<https://digitalimpact.io/policies/policy-wizard/>).
- The University of Minnesota and the University of Arizona, and the Inter-University Consortium for Political and Social Research, have posted their [Framework for Creating a Data Management Plan](#).

- The Digital Curation Centre has a host of online tools, including a **Checklist for a Data Management Plan**: <http://www.dcc.ac.uk/resources/data-management-plans>
- Project Open Data Metadata Schema 1.1 (<https://project-open-data.cio.gov/v1.1/schema/>)
- The **UNM Metadata Codebook Text (Tool 4)** readme file is online at: <http://z.umn.edu/readme>
- The University of Michigan's **Informed Consent tools** can be found at: <http://research-compliance.umich.edu/informed-consent-guidelines>
- The most current version of the **USAID DDL Registration/Submission Form** is at: <https://www.usaid.gov/data/ddlsubmissions>
  - To view a list of data dictionary examples, [click on the link](#)
- USAID is using the following template to **assess potential privacy issues**: <https://www.usaid.gov/sites/default/files/documents/1868/508mah.pdf>
- The USAID has posted detailed answers to frequently asked questions: <https://www.usaid.gov/data/frequently-asked-questions>
- UK Data Archive Resources:
  - Information on **de-identification**: <http://data-archive.ac.uk/create-manage/consent-ethics/anonymisation>
  - **Quantitative De-identification**: [Read more about quantitative anonymisation](#).
  - **Qualitative De-identification**: [Read more about qualitative anonymisation](#).
  - Additional **consent form templates**: <http://data-archive.ac.uk/create-manage/consent-ethics/consent?index=3>
- Data Observation Network Earth, DataONE, **DMP Best Practices**: <https://www.dataone.org/best-practices/provide-budget-information-your-data-management-plan>
- A glossary of terms associated with Open Data is posted at: <http://opendatahandbook.org/glossary/en/>
  - The USG's glossary is posted here: <https://www.data.gov/glossary>

### 8.3.Case Study Example

There have been several studies uploaded to the DDL. Below is just one example, chosen at random, to illustrate program description, objectives, metadata, and uploaded data spreadsheets.

#### *Baseline Study of Food for Peace Title II Development Food Assistance Program in Uganda*

In fiscal year 2012, USAID's Office of Food for Peace (FFP) awarded funding to private voluntary organizations (PVOs) to design and implement a multi-year Title II development food assistance program in Uganda. The main purpose of the Title II program is to improve long-term food security of chronically food insecure population in the target regions. FFP contracted a firm, ICF International to conduct a baseline study in targeted areas of the country prior to the start of the new program. The purpose of the study was to assess the current status of key indicators, have a better understanding of prevailing conditions and perceptions of the population in the implementation areas, and serve as a point of comparison for future final evaluations. Results would also be used to further refine program targeting and, where possible, to understand the relationship between variables to inform program design. The study was conducted in 2013, while FFP expects to conduct final evaluations as close as possible to the end of the program five years later.

The specific objectives of the baseline study were:

- Establish baseline values of key impact and outcome indicators prior to implementation of the Title II programs and set targets;
- Assist the PVOs in establishing target levels for improvements in these indicators over the five-year Title II program cycle;
- Inform PVOs about the current food security situation so they can refine their program design and implementation strategies and improve efficiency by targeting the areas and subgroups that will benefit most; and
- Provide FFP baseline indicator values that can be compared across countries through meta-analyses of indicator results. The baseline study consisted of two components: a representative population-based household survey focused on the collection of data for required impact and outcome indicators, and a qualitative component to add depth, richness, and context and serve to triangulate information from survey findings and analysis.

*Currently there are seven (7) different data sets, each with a csv file and a data dictionary.*

#### **SAMPLE ASSOCIATED DDL FILES**

- **Uganda Agricultural Practices Data** ([csv file](#), [Tool 12](#))



- **Associated Data Dictionary (Tool 13)**

### References

(Tool 14) [http://pdf.usaid.gov/pdf\\_docs/pnaed238.pdf](http://pdf.usaid.gov/pdf_docs/pnaed238.pdf)

### Additional Metadata

Access level	<a href="#">Public</a>
Bureau code	184:15
Contact	<a href="#">Jill Moss</a>
Keywords	baseline survey, ffp, food for peace, food security, karamoja, title ii development program, uganda
Landing page	<a href="#">Visit page</a>
Language	en-US
License	<a href="#">See this page for license information.</a>
Modified	2013-10-15
Programs	Agriculture, Maternal and Child Health, Monitoring and Evaluation, Nutrition, Water Supply and Sanitation
Publisher	USAID
References	<a href="http://pdf.usaid.gov/pdf_docs/pnaed238.pdf">http://pdf.usaid.gov/pdf_docs/pnaed238.pdf</a>
Rights	Under the terms of an agreement with USAID, a partner owns data it collects. Under the authority of the license that partners grant to USAID, USAID posts the data with a CC-BY license providing attribution to the partner.
Spatial	Uganda
Unique identifier	3c32504d-ffba-46c9-8c7b-30449def4143

## 8.4.Key Terms to Know

- **Dataset** is defined as an organized collection of structured data, including data contained in spreadsheets, whether presented in tabular or non-tabular form. Data collected as part of performance monitoring or evaluation, and project data used in reporting to USAID. (For example, lot quality assurance, cluster samples, facility surveys, school attendance, crop and weather monitoring, sector-related sampling) and quantitative data underpinning evaluation findings (for example, baseline, midterm, final surveys, and other

related datasets). Datasets may also include other survey data that inform project design (for example, population-based and integrated household surveys, economic assessments, organizational capacity assessments).

“Dataset” in this context *does not include qualitative data*, aggregated performance reporting data, unstructured data (email messages, PDF files, PowerPoint presentations, word processing documents, photos and graphic images, audio files, collaboration software, and instant messages).

- A **Dataset Collection** is a group of related datasets, where each dataset can have multiple spreadsheets.
- **Intellectual Work** is defined as works that document the implementation, monitoring, evaluation, and results of international development assistance activities developed or acquired under a federal award, which may include program and communications materials, evaluations and assessments, information products, research and technical reports, progress and performance reports, and other reports, articles and papers prepared under an award, whether published or not. PCI, for example, generally produces three types of intellectual work for any given project, including project evaluations (baseline, midterm, endline reports), special and/or research studies, and annual reports.
- **Supporting Documentation** is defined as information describing the dataset, such as code books, data dictionaries, data gathering tools, templates and forms, data gathering methodologies, notes on data quality, and explanations of redactions<sup>14</sup>. Supporting documentation defines the fields within the dataset and any categories or labels within the dataset that may require explanation to an individual not familiar with the data. When available, scopes and methodologies, such as survey protocols and instruments used to collect and analyze the data must also be submitted to the DDL, along with annotations to inform the general public of any known data quality issues. Datasets must be accompanied by metadata as required by Project Open Data and other metadata assigned by USAID<sup>15</sup>.

## 8.5.Dataset Access Levels Defined

Access levels must be designated for each dataset, with documentation of the rationale for the designation in the DDL. You will recommend access levels during the proposal and award acceptance phases, which will be approved by your COR/AOR.

Datasets must be documented with one of these three access levels:

<sup>14</sup> Redaction means the censoring or obscuring of part of a text for legal or security purposes.

<sup>15</sup> This may be accomplished by following the instructions at [www.usaid.gov/data](http://www.usaid.gov/data).

- **Public:** Dataset is or could be made publicly available to all without restrictions. For the “Public” access level, Operating Units may also document technical or resource barriers to increasing access to this Dataset.
- **Restricted Public:** Dataset is available under certain use restrictions. One example, among many, is a [dataset] that can only be made available to select researchers under certain conditions, because the dataset contains sufficient granularity or linkages that make it possible to re-identify individuals, even though the dataset is stripped of PII. Another example would be a Dataset that contains PII and is made available to select researchers under strong legal protections. This category includes some but not all datasets designated as Controlled Unclassified Information (CUI), consistent with Executive Order 13556. For the “Restricted Public” access level, Washington Operating Units, in consultation with GC, or overseas Operating Units, in consultation with the appropriate RLO, must document the conditions under which select individuals may obtain access to this Dataset.
- **Non-Public:** Dataset is not available to members of the public. This category includes datasets that are only available for internal use by the federal government, such as by a single program, single agency, or across multiple agencies. This category might include some but not all datasets designated as CUI, consistent with Executive Order 13556. Some non-public datasets may still potentially be available to other intra-agency Operating Units and/or other government agencies, as discussed in OMB Memorandum M-11- 02: Sharing Data While Protecting Privacy. For the “Non-Public” access level, Washington Operating Units, in consultation with GC, and overseas Operating Units, in consultation with their RLO, must document the rationale for specifying that the data cannot be made public.

### 8.6.Exceptions Language (Samples)

Limited exceptions to making submitted data publicly available may be granted in the following instances:

1. When public disclosure threatens national security interests;
2. When public disclosure is likely to jeopardize the personal safety of U.S. personnel or recipients of U.S. resources (which should include beneficiaries, to be confirmed by USAID);
3. When public disclosure would interfere with the agency's ability to effectively discharge its ongoing responsibilities in foreign assistance activities;
4. When there are legal constraints on the disclosure of business or proprietary information of non-governmental organizations, contractors, or private sector clients;

5. When the laws or regulations of a recipient country apply to a bilateral agreement and restrict access to information;
6. When data reveal private information about individuals that must be kept confidential consistent with ethical guidelines and federal regulations; or
7. When examining information that falls within these categories, USG agencies have sufficient flexibility to protect sensitive information from disclosure, *on a case-by-case basis*.

If the TA has determined any of the above situations apply, the Grants Coordinator should submit this as a question to the donor. Sample language for each request is included below:

- *Restricted or non-public access level:* "We expect that some of the data subject to the Open Data Policy will be particularly sensitive and that simply de-identifying the data will not ensure confidentiality for beneficiaries. Would USAID consider applying access level "X" to data from this project for the following reasons: RATIONALE [cite one or more of the acceptable reasons per section 3.2.3 of ADS 579]?"
- *Exemption from data transmission:* "Would USAID consider an exemption from dataset transmission to the DDL for the following reasons: RATIONALE?"
- *Timeframe of submission:* "We expect that some of the data subject to the Open Data Policy will need to be continually updated (quarterly or more frequently). As such, would USAID consider X [e.g., once per annum] as an appropriate timeframe for submitting these data, given that such data would be incomplete if submitted within 30 days of each progress report and therefore not of use to the public?"
- *Research data embargo:* "Would USAID consider applying an embargo to temporarily withhold from public release research data that are subject to pending publications for X [a reasonable period of time, e.g., 12 months] to allow researchers adequate time to analyze and publish works based on the data?"

## 8.7.Acronyms List

AO	Agreement Officer
AOR	Agreement Officer's Representative
COR	Contract Officer's Representative
CUI	Controlled Unclassified Information
DM	Data Management
DMP	Data Management Plans
DDL	Development Data Library
DCC	Digital Curation Centre
FAQ	Frequently Asked Questions
FFP	Office of Food for Peace
IRB	Institutional Review Board
M&E	Monitoring and Evaluation
NGO	Non-Governmental Organization
OAIS	Open Archival Information System
PII	Personally Identifiable Information
PPR	Performance Plan and Report
PVO	Private Voluntary Organizations
PCI	Project Concern International
TOPS	Technical and Operational Performance Support
USAID	United States Agency for International Development