

*Data
Protection is
People
Protection*



DATA PROTECTION & PRIVACY GUIDES



The following content is designed to help Mercy Corps staff better understand and implement responsible data practices. It brings several existing Mercy Corps policies and guidance documents together with some simple tutorials and links to other resources. The information can be taken all together as a comprehensive guidebook, or each section can be used as a standalone guide to a particular topic.

While the guide is primarily intended for Mercy Corps staff, and is designed to accompany the [Responsible Data Toolkit](#), we are releasing it under open license so that it may benefit partners, peer-organizations, and others in need of example policies, templates, and instructions for implementing responsible data practices. You can download all of our guide content via our Github page at <https://github.com/mercy Corps/DPP-guides>.

Table of Contents

Understanding Sensitive Data	1
<i>Defines what sensitive data is and provides guidance for its collection and use.</i>	
Privacy Impact Assessments	3
<i>Provides details regarding PIAs and the assessment templates required by Mercy Corps' Responsible Data policy.</i>	
File Sharing Best Practices	6
<i>Presents an overview of best practices using G Suite at Mercy Corps and a brief tutorial.</i>	
Deidentifying Data	13
<i>Provides a brief overview of encryption and provides an example of one way to encrypt a file using software available within Mercy Corps.</i>	
Encrypting a File	21
<i>Contains an overview of de-identification—including anonymization and pseudonymization—and an example of one way to deidentify a dataset using spreadsheet software.</i>	

Citations and acknowledgments

The content provided in this guide is all part of good data management practices and the larger data life cycle, or the overall activities for individual data management as part of a program or response. The following resources are excellent places to start for a more complete understanding of managing your data responsibly. We have used or cited portions of all of these resources throughout the Guide and refer to specific chapters or sections that are most pertinent to a specific topic.

- › The Cash Learning Partnership’s *Data Responsibility Toolkit* is designed for cash and voucher practitioners specifically, but is a gold standard in guidance for responsible data. The Toolkit is available in [English](#), [Arabic](#), [French](#), and [Spanish](#).
- › The [Electronic Cash Transfer Learning Action Network’s Data Starter Kit for Humanitarian Field Staff](#) provides a series of data tip sheets for understanding various aspects of good data management and protection practices..
- › The [International Committee of the Red Cross’ Handbook on Data Protection in Humanitarian Action](#) is a detailed guide to almost every aspect of humanitarian data.
- › The [International Federation of Red Cross and Red Crescent Societies’ Data Playbook](#) is an excellent resource of exercises, session plans, checklists, and other materials to help you organize conversations and activities with your team to develop responsible data activities.
- › The [Humanitarian Data Centre’s online tutorial for conducting a disclosure risk assessment](#) is a very specific, technical resource, but one that is indispensable for truly reducing the risk of data being used to identify individuals.
- › The [Engine Room’s Handbook of the Modern Development Specialist](#) is a good overview of data in the context of international development activities.

License

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).



Understanding Sensitive Data

Understanding the different classifications of data can be difficult, but they are an important part of working with humanitarian data. For example, what is the difference between personal data and sensitive data? Particular types of data may require special care in accordance with regional or national laws or organizational policies, and can present different types of risk to both program participants and organizations. Sensitive data is a sub-category of personal data and this section provides detailed guidance for processing and understanding it.

☆ Importance

A majority of Mercy Corps programs and other activities collect some kind of personal information about individuals. In many cases programs also collect information about an individual's cultural profile, sexual orientation, health, or biometrics and genetics. These types of information are considered sensitive data and if they are disclosed, accessed or shared improperly, could result in:

- › harm to a person, such as sanctions, discrimination and security threats;
- › a negative impact on Mercy Corps' ability to carry out activities and reduced trust or public perception.

It is vital to take the necessary steps to protect these data.

Guidance

This section contains two documents:

- › “Guidance for processing of sensitive data” introduces you to sensitive data, key terms, and things to consider when planning its collection, storage, analysis, and sharing.
 - Mercy Corps staff can access the Guidance in [Mercy Corps' internal Digital Library](#).
 - Anyone can download the template as a Microsoft Word document in [English](#), [Arabic](#), [Spanish](#), [French](#), and [Russian](#).
- › “Sensitive Information Assessment (SIA) Template” can be used with a Privacy Impact Assessment to document all additional safeguards being employed for sensitive data. This document also sets out the different legal bases that can be used to justify the collection and use of sensitive data.
 - Mercy Corps staff can access the SIA template in [Mercy Corps' internal Digital Library](#).
 - Anyone can download the template as a Microsoft Word document in [English](#), [Arabic](#), [Spanish](#), [French](#), and [Russian](#).

Further Assistance

Planning for the collection or use of sensitive data should be part of a larger strategy for the data life cycle of a program or activity. There are several resources that can help you.

- › The International Federation of Red Cross and Red Crescent Societies (IFRC) *Data Playbook* is an excellent resource of exercises, session plans, checklists, and other materials to help you organize conversations and activities with your team. In particular, [Module 4 - Responsible Data](#) is a good place to start.
- › The International Committee of the Red Cross' *Handbook on Data Protection in Humanitarian Action* is a detailed guide to almost every aspect of humanitarian data. Chapter 3 specifically deals with the legal basis for personal data processing.
- › The Cash Learning Partnership's *Data Responsibility Toolkit* is designed for cash and voucher practitioners specifically, but is a gold standard in guidance for responsible data. Specifically, *Tip Sheet #2, "Design and Plan"* discusses lawful basis for sensitive data. The Toolkit is available in [English](#), [Arabic](#), [French](#), and [Spanish](#).

Privacy Impact Assessments

This guide will help you understand a Privacy Impact Assessment (PIA) and contains the PIA guidance and template used at Mercy Corps. The PIA template contains a series of questions that create a framework for identifying the potential privacy risks related to data collection and management that are part of implementing a new program or technology. A PIA is also important when the context of a program changes significantly and new risks or scenarios need to be considered.

A PIA is required anytime a new program, project, or technology involves the collection or use of personal or sensitive data.

☆ Importance

A PIA allows you to analyze how a particular project or new technology will affect the privacy of the individuals involved. A PIA also helps to document mitigation strategies that protect participant's privacy and strengthen public confidence in our work. A PIA ensures that potential problems are identified early on, when addressing them will be simpler, less costly, and will not risk harming program participants or staff.

Principles

The principles behind a PIA are similar to those for any secure use of personal data. Below are some key principles that have been adapted from the [Cash Learning Partnership \(CaLP\)](#):

- › Identify the privacy risks to individuals.
- › Identify the privacy and data protection compliance liabilities for your organization.
- › Demonstrate accountability and compliance with the policies that protect program participants, partners, and staff.
- › Ensure that the organization is promoting the right to privacy in its humanitarian activities and acting as an ethical data steward.

Guidance

Mercy Corps' staff can find the [PIA Guidance in the Digital Library](#). The document contains answers to frequently asked questions related to PIAs and links to the internal Mercy Corps archive of completed PIAs for comparison. Mercy Corps' PIA Guidance is available to anyone in [English](#), [Arabic](#), [Spanish](#), [French](#), and [Russian](#).

Remember that:

- › A PIA is a *process* used to identify and minimize the privacy risks. Completing a PIA form is not the end of the process! Revisit the PIA again after your project starts to make sure there are no new changes that introduce new risks. If there are, document the changes and new mitigation strategies needed to minimize any new risk.

› Conducting a PIA involves working with people at Mercy Corps, and sometimes with partner organizations and others to identify and reduce privacy risks. For example, if you are using a new technology, you may need to research whether the company you are working with has a privacy policy and what technological safeguards they use to ensure data is protected. You may also need to educate yourself about relevant privacy regulations in your country of operation. Three websites that you can use to monitor national-level data and privacy laws are:

- [Data Protection Laws of the World](#);
- The [U.N.Conference on Trade and Development database of Data Protection and Privacy Legislation Worldwide](#); and
- [One Trust Data Guidance database of Global Privacy Laws](#).

› It can be helpful to compare PIAs of similar programs. You can conduct this research on your own or reach out to the Data Protection and Privacy team for assistance.

Templates

Mercy Corps' staff can find the long-form [PIA Template in the Digital Library](#). The long-form PIA Template is available to anyone in [English](#), [Arabic](#), [Spanish](#), [French](#), and [Russian](#).

Each long-form PIA template includes five use-cases, which are explained below. Clicking the links below will take you to a screen where anyone can download English versions of the individual use cases in the **.odt** format (compatible with Microsoft Word and open source applications like OpenOffice and LibreOffice) by clicking **View raw** or the **Download** button.

› a new [Policy](#)

› a new [Process or Procedure](#)

› a new [Software or Technology System](#)

- This is primarily for implementation of new global, country-wide or team-specific systems.
- If you are selecting or using a new system as part of a larger project or program, use the project or program option instead.

› a new [Vendor or Partner](#)

- This is primarily meant for validation of a vendor, partner, or third party's activities as part of a unique or one-time activity.
- If you are selecting or using a new vendor, partner or third party as part of a larger project or program, use the project or program option instead.

› a new [Project or Program](#)

- This can be for any phase or aspect of a project or program.
- *This is the most comprehensive PIA option*, and includes language for also selecting new software or tech systems, and/or a new vendor or partner.

Further Assistance

- › The Electronic Cash Transfer Learning Action Network’s [Data Starter Kit](#) provides a tip sheet for PIAs (**see Tip sheet #1**).
- › The Information Commissioner’s Office of the UK provides a [detailed code of practice for conducting privacy impact assessments](#).
- › The International Committee of the Red Cross’ *Handbook on Data Protection in Humanitarian Action* is a detailed guide to almost every aspect of humanitarian data. Chapter 5 specifically deals with privacy impact assessments.

File Sharing Best Practices

This guide covers file sharing best practices for Mercy Corps team members using G Suite apps, particularly Google My Drive. To simplify this guide, we will be discussing sharing a single *file*, such as a spreadsheet. However, the same options are available when sharing a My Drive *folder*.

Note: Mercy Corps is transitioning to Microsoft 365 for file sharing. Once best practices for that platform have been established, a similar document or resource will be created.

☆ Importance

There are several reasons why it is better to share a file by hosting it in Google Drive and sending a link rather than as an attachment in an email.

- › **Security:** you can easily change who is permitted to access or edit your file. You can also make the link time sensitive by only allowing access to the file for a given period of time.
- › **Version control:** when sharing a file that is hosted online, many people can access it at once and all changes and comments will stay in one file. Sending a file as an attachment often results in multiple versions of the same document with different file names, edits, comments, etc. The document owner will spend a lot of time trying to compile all of this into a single file! Using a link also guarantees that the recipients always have access to the most updated version of the document.
- › **File sizes:** Some IT departments impose restrictions on the size of attachment that are allowed. Sending a link allows you to share files of any size.
- › **Easy editing:** Files shared as Google documents in Drive, Microsoft Word in OneDrive, or similar formats allow the recipient to open and interact with the document using a web browser: they do not have to have the most recent version of a particular type of software.

Principles

Whenever you share documents, you should consider the following:

- › **Carefully consider who will create the file or folder, who will own/administer it, and who will access it. If short-term contractors create and administer files there is a risk of the files or the access leaving with the contractors when they leave!**
 - Only give access to those who need the file.
 - Confidential, private or personally identifiable information (PII) content must *always* have restricted access. If you are unable to determine if or how the content should be restricted, please seek assistance from Legal or the Data Protection and Privacy team.

) Use the appropriate level of permissions.

- Consider an example where a new project charter needs to be created. Most likely, only the team responsible for charter creation should have full access. When it's time to get feedback from others, grant additional permissions that only allow viewing or commenting.
- Use extreme caution in granting access requests to team members who may accidentally be using their personal email accounts. Instead of granting access to a personal email account, grant access to their Mercy Corps account, and ask the team member to log in with those credentials.

) Permissions change over time.

- If you are only working with someone for a short period of time or with people outside Mercy Corps, consider giving temporary permissions. If you forget to remove the access later, an expiration date will ensure their access is suspended at the right time.
- Periodically audit the list of who has access to your files, folders or shared drives to ensure you remove access for team members who have changed roles, or are no longer with Mercy Corps.

) Risky content requires extra steps.

- Personally identifiable information (PII), demographically identifiable information (DII), or other types of personal data are protected under multiple data protection laws. Before you share personal data, verify the legal requirements for sharing that information with others. Inappropriate personal data sharing can put program participants, donors, partners and Mercy Corps team members at risk. If you have questions about personal data or data protection laws, please email the Data Protection and Privacy team at dataprotection@mercy Corps.org.
- If the information is considered confidential or proprietary for business purposes, share it only with required parties and consider granting temporary access.
- If the person receiving the file works in an insecure location, or if the contents include personal data, consider encrypting the file or protecting it with password protection. See the Encryption and de-identification sections for examples of how to do this.

) Never move files without the owner's permission.

- Moving files may alter who has access and make it impossible for others to find the file! Always check with the document owner before moving a shared file to a new location.

File Sharing: GDrive

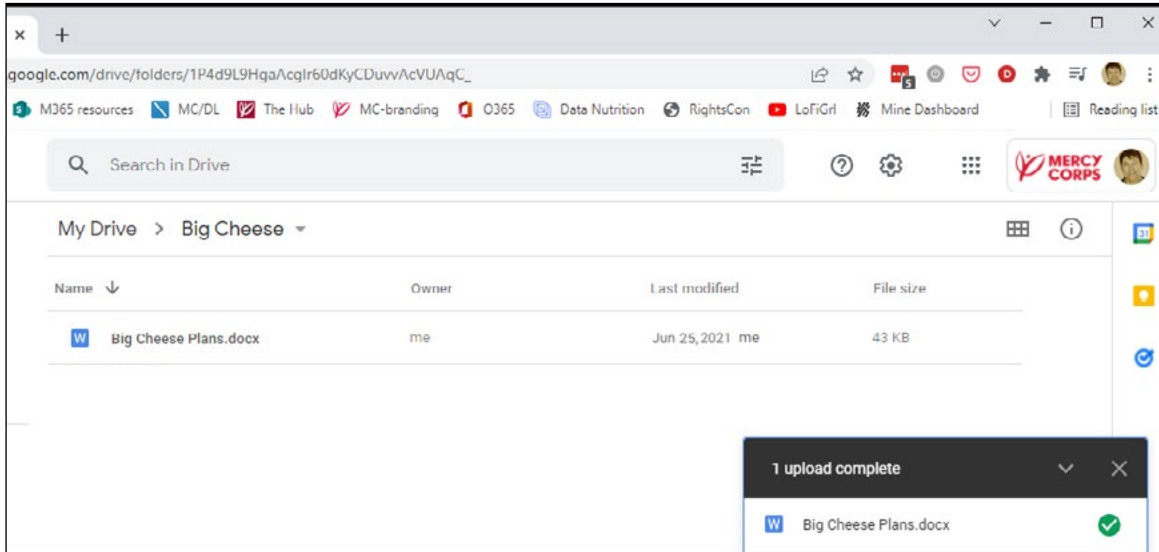
Here is a sample use-case for sharing a file using best practices in Google Drive. Imagine that the year is 2020 and Mercy Corps is working with a consultant (PNW Rocks), to create materials for a major new initiative at Mercy Corps, that is code-named Big Cheese. The Big Cheese project will not be announced publicly until July 2021 so it's important to limit who has access to the file. To get started, we'll want to collaborate on next steps for the project in a file called "Big Cheese Plans."



Instructions

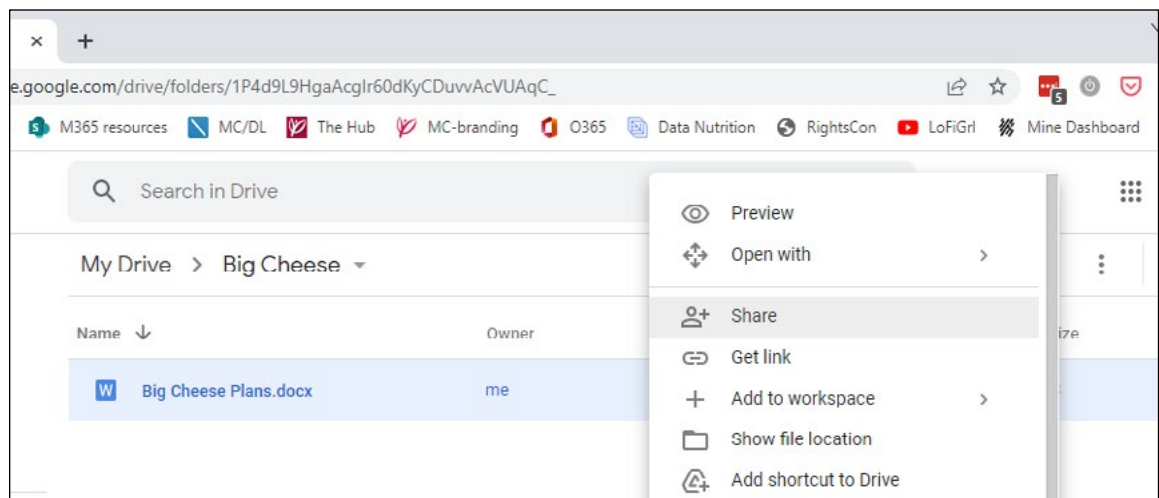
Step 1 - Upload

Upload the file to My Drive.



Step 2 - Share

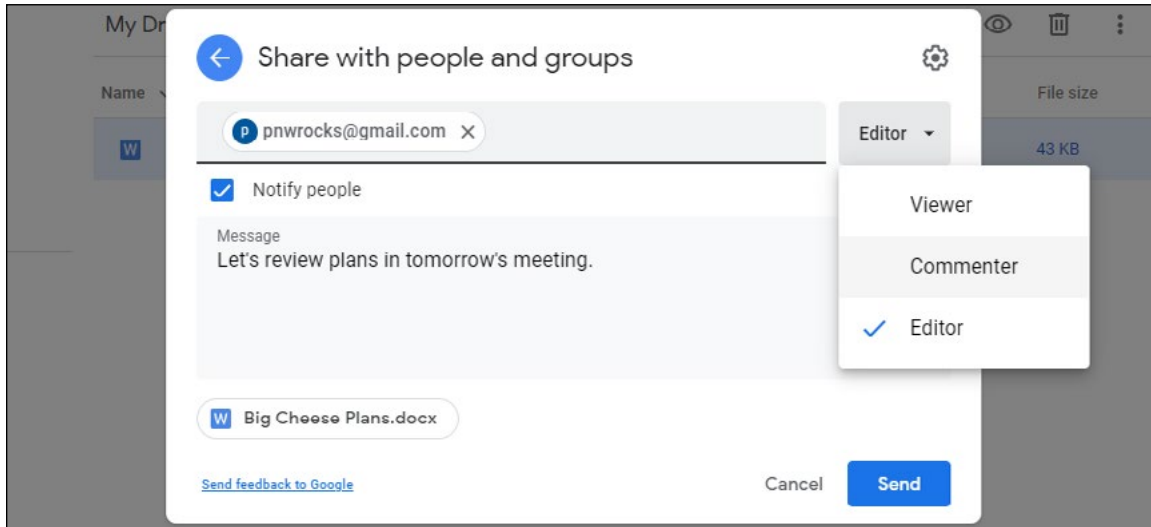
To share the file, right-click the file and then click **Share**.



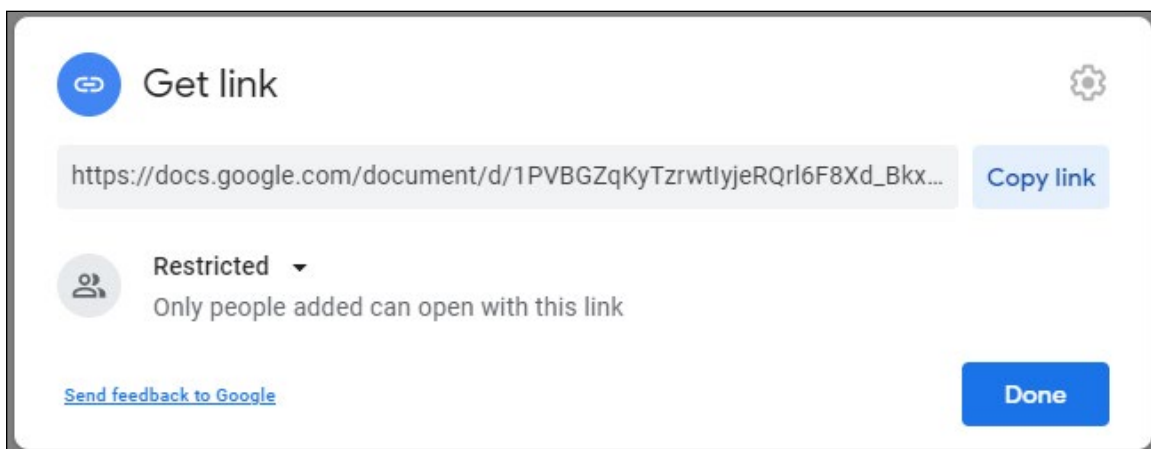
- › Only give access to those who need the file. When you share a file in My Drive, the default setting is **Restricted** (to persons or groups), which is the best practice. Remember, confidential, private or personally identifiable information (PII) content should *always* have restricted access!
- › The **Anyone with the link** option should only be used for files that hold no confidential information and are open to the public. If the **Anyone with the link** setting is used on a file with confidential private, personal or financial data, it could easily and accidentally be shared and put Mercy Corps at immediate legal risk and make it easy for bad actors to use the information for criminal purposes.

Step 3 - Select level of access

Enter the email address for the person you want to share with, then choose the level of access. Google defaults to **Editor** access, which is only appropriate for team members who need full access to the document. When you are engaging stakeholders for feedback, choose **Viewer** or **Commenter**. Best practice is to notify the person, and add a message, explaining why you've shared the file. To notify, leave that **Notify people** box checked. When done with changes, click **Send**.



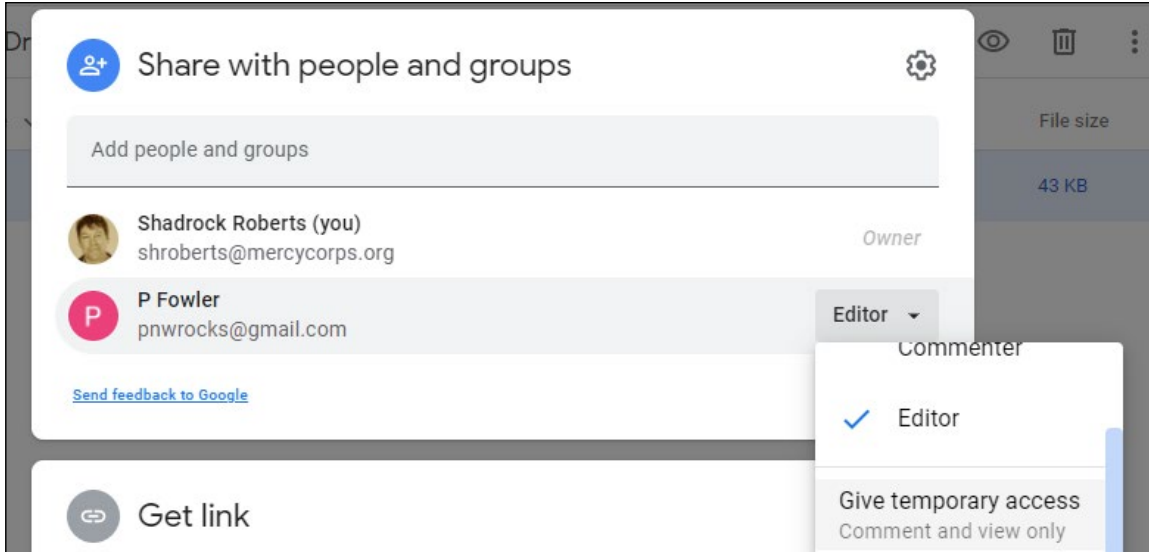
Another option is to send a separate email with a link to the file. To notify separately via email, uncheck the **Notify people** box. After clicking **Done**, right click the file and choose **Get Link**. In the new popup window, click the **Copy link** button, and then paste it into your email.



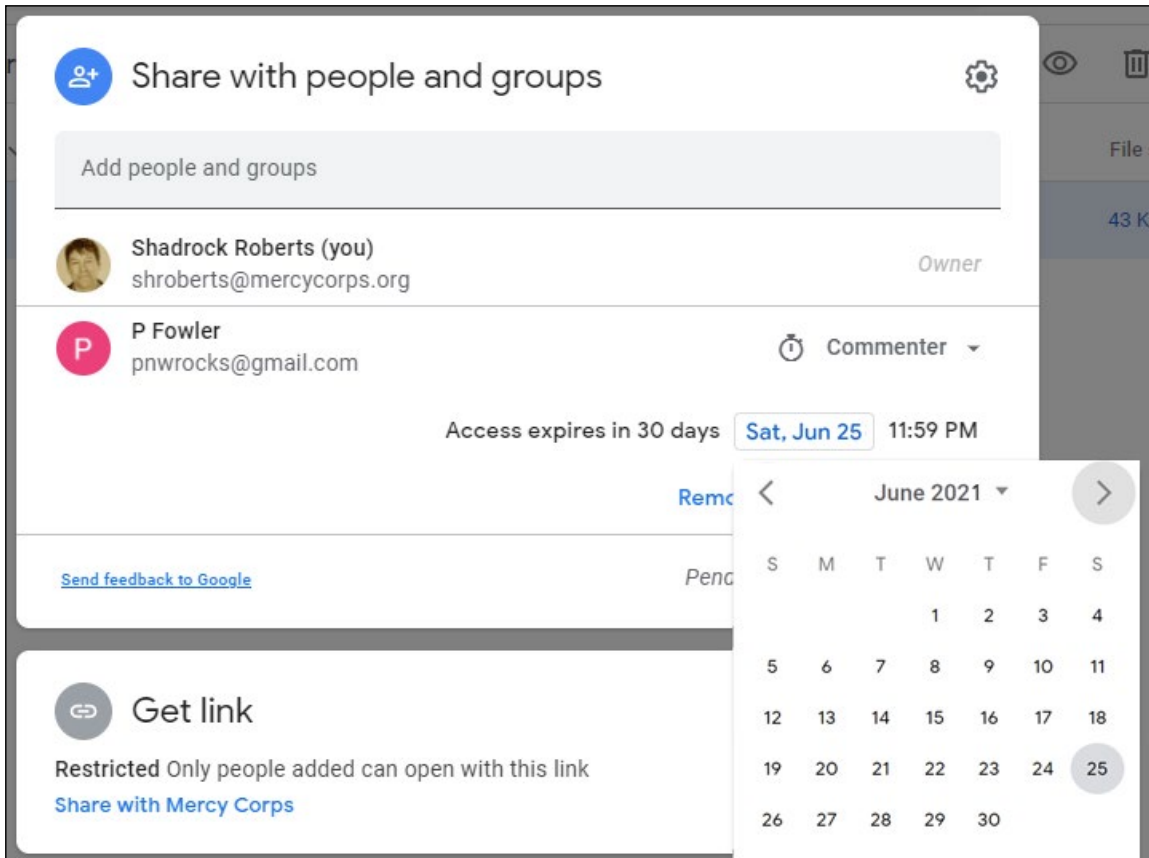
- › To learn more about access levels, visit [Share and collaborate in My Drive](#).
- › If you restrict a file, and someone with access shares the link with another person, that person will not automatically have access to that content in Google Drive. Instead, they'll see a web page with the opportunity to request access. The request for access will go to the file owner. Access requests should be reviewed, and those receiving access requests should not simply grant permission to any and all requests, without reviewing the above notes and considerations.

Step 4 - Temporary Access

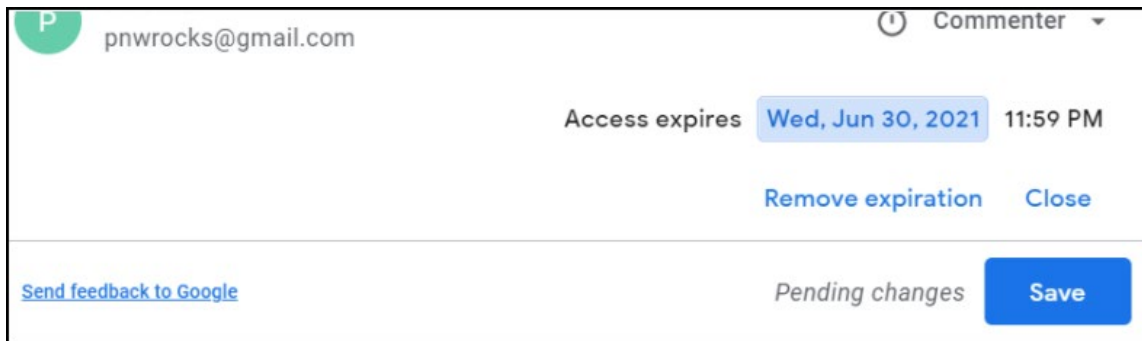
Once permission has been granted, the sharing time period can be shortened. To limit access, right click the file again, and click **Share**. Find the email address you just shared with and right click the access level. You will see new options appear; select **Give temporary access**.



A calendar will appear. Navigate to the month when access should expire, and click the corresponding date.

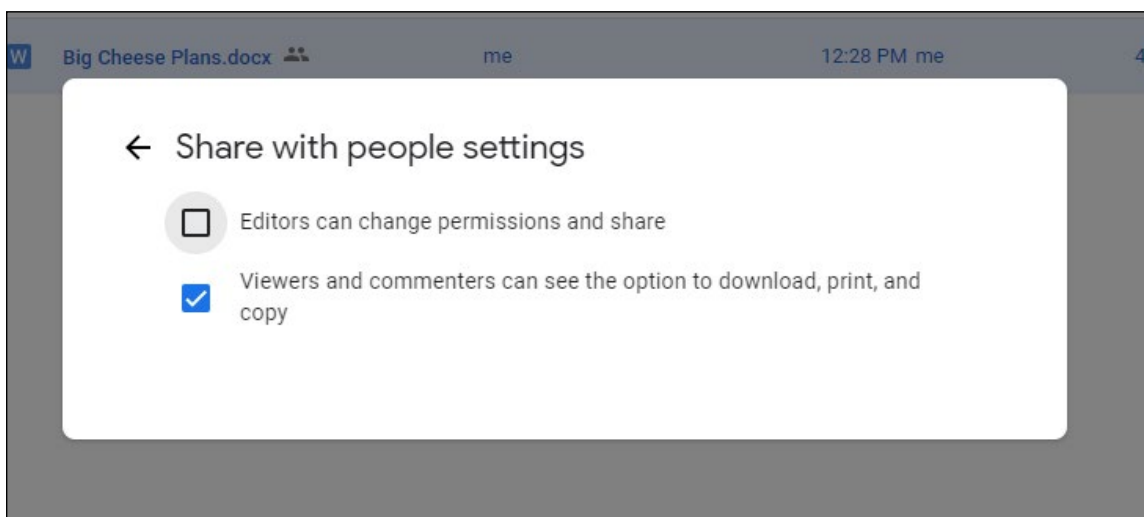


The calendar will disappear, and the display date will change. Once you can see that access will expire on the correct date. Click **Save**.



Step 5 - Additional Options

If you have concerns about others taking inappropriate actions with the content, there are additional options available in the settings screen. File settings can be accessed from the gear icon in the upper right of the sharing window. Click the gear and you'll see options to restrict further sharing, or disabling the option to download, print or copy. For any feature you want to disable, simply uncheck the box. The file will quickly update, saving the new setting.



› To learn more, visit [Restrict sharing options on Drive](#).

Additional considerations

- › Do not place confidential files inside folders that are shared widely. Shared folders permissions trickle down to each file and subfolder, so your confidential file could be accessed by anyone with permission to the parent folder! Instead, move your confidential file to a new location, outside the larger shared folder structure.
- › Once a file has been shared, you may have the option to **Make a Copy** and/or **Move** the file to another location. Never move a file without the owner's permission!
- › If you need to see the file in your My Drive or in a secondary location, the best practice is to use the [Add shortcut to Drive](#) option.
- › If a file copy is made, or the file is moved, be aware it will not have the same permissions as the original file unless you explicitly set those permissions.

This guide does not cover file sharing in Google's shared drives, nor external file sharing platforms. To learn more about these, visit Google's [Best practices for shared drives](#) or [managing shared drives page](#). Each of these pages is available in a variety of languages: scroll to the bottom of the page to select your language.

The best way to control access to files using Drive is to create a Google group and assign permissions to those in the group. Google groups are not just for sending emails; groups are powerful and convenient tools for managing folder and file permissions. [Find out more about Google Groups](#).

If your project requires the use of applications outside of G Suite, encourage your team to download [Google Drive for desktop](#). This program allows you to view any documents in My Drive or Shared drives as if they were on your laptop, even when offline, and without having to download the item or convert it to a Google format.

Further Assistance

How and with whom you will share data should be part of a larger strategy for the data life cycle of a program or activity. There are several resources that can help you.

- › The International Federation of Red Cross and Red Crescent Societies (IFRC) [Data Playbook](#) is an excellent resource of exercises, session plans, checklists, and other materials to help you organize conversations and activities with your team. In particular, [Module 7 - Data Sharing](#) is a good place to start.
- › The International Committee of the Red Cross' [Handbook on Data Protection in Humanitarian Action](#) is a detailed guide to almost every aspect of humanitarian data. Chapter 2 specifically deals with data sharing.
- › The Cash Learning Partnership's [Data Responsibility Toolkit](#) is designed for cash and voucher practitioners specifically, but is a "gold standard" in guidance for responsible data. See especially, [Tip Sheet #6, "Data Sharing"](#). The Toolkit is available in [English](#), [Arabic](#), [French](#), and [Spanish](#).

Deidentifying Data

This guide provides an example of removing personally identifiable information (PII) from a dataset. There are several ways to “de-identify” data, which refers to the processing activities or methods that work to prevent a Data Subject’s identity from being revealed. Two common types of de-identification are “Anonymization” and “Pseudonymization.”

Anonymization is the process by which Personal Data is rendered anonymous so that an individual (or “data subject”) is no longer identifiable: it is a permanent change to the data. Common methods include removing personally identifiable information or scrambling values across certain sets of PII.

Example: imagine an organization has survey data that contains fields for name, national ID number, village name, ethnic affiliation, age, education level, and health indicators. In this case, removing name and national ID number would be the first step in making the data anonymous since these “direct attributes” are personal data that directly identify an individual. The “indirect attributes” of village name, ethnic affiliation, age, education level, and health indicators would remain.

However, even though some attributes seem “anonymous” they may not be. If the survey was collected in a very small village where only two residents identify as a particular ethnic affiliation, and they are each of different ages, then using those two indirect attributes could allow for those individuals to be identified! The process by which all attributes are examined to reduce the risk of re-identifying a data subject is called Statistical disclosure control. The first step in this process is a disclosure risk assessment and the Humanitarian Data Centre has an [online tutorial for conducting a disclosure risk assessment](#).

Pseudonymization, on the other hand, describes the processing of personal data in a way that personal data can no longer be attributed to a specific data subject without the use of additional information, such as a key code.

Example: imagine a survey contains your name, email address, age, nationality, and workplace. Pseudonymization takes the data that’s identifiable about you specifically (your name, email address, age) and makes it inaccessible and separate from non-identifying data, like your nationality. Pseudonymous data can be put back together at some point so that all information can be linked back to a specific source or person. This is why pseudonymization requires that the additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to the data subject.

Should you choose Anonymization or Pseudonymization?

Anonymizing will generally be safer and reduce the risk of exposing PII. However, this can sometimes make the data too general, which may not make it useful for programs such as cash voucher assistance. In the case of health programs that involve vaccinations or other treatments, it may be important to contact individuals for follow-up treatment. In both of these cases pseudonymization would be the best choice since you can always put the data back together to identify an individual when needed.

There is no single right answer about when to choose one method over another and it is important to understand why the data were collected, the potential risks associated with holding that data, and the needs of the program, before choosing how to deidentify your data.

It's also important to understand that the techniques used to both anonymize data and to hack data are becoming ever more sophisticated and that **even de-identified data isn't always one hundred percent secure**. When in doubt, contact your data or IT team for assistance.

☆ Importance

Recent **data breaches at the international Committee of the Red Cross, email hacks at the U.S. Agency for International Development, and improper data sharing by the U.N. High Commissioner for Refugees** all show several ways in which humanitarian data are at risk. Data from household surveys, needs assessments and other forms of microdata make up an increasingly significant volume of data in the humanitarian sector. These types of data are critical to determining the needs and perspectives of program participants and the communities we work in, but these data also present risks. Understanding how to assess and manage the sensitivity of these data is essential to ensuring that they are used in a safe, ethical and effective manner in different response contexts.

Some advantages of using anonymized data over personal data include:

- › protecting against inappropriate disclosure of personal data;
- › fewer legal restrictions apply to anonymized data; and
- › allowing organizations to create open or publicly accessible data while still complying with their data protection obligations.

Principles

De-identifying data is part of data processing, and personal data processing undertaken by humanitarian organizations should comply with the following principles.

- › **Fairness and lawfulness of processing:** methods must comply with regional, national, or local legislation or policies that may limit what data can be de-identified and how certain technologies are used. Any processing of personal data should be transparent for the data subjects involved.
- › **Purpose limitation:** humanitarian organizations should determine and set out the specific purposes for which data are processed. These purposes should be explicit and legitimate.
- › **Proportionality:** ensure each particular activity related to the processing of personal data is appropriate for the stated goal. For example: is only the minimum required amount of data being collected? Are appropriate technical and organizational measures in place to reduce the risks associated with data processing?
- › **Technology changes:** new datasets and new tools for analyzing them change and advance rapidly, and so do the means by which data are hacked or stolen. It is important to understand new and emerging risks to your data and continue to adjust your methods and practices accordingly.

Pseudonymization

This is an example of one way to de-identify data in a spreadsheet. There are a wide variety of ways to perform de-identification and this example uses a “key code” to remove personally identifiable information found in direct identifiers and keep it in a separate file. Personally Identifiable Information (PII) is information that can be used to identify an individual. Common examples are name, address, phone number, date of birth, and social security or national ID number.

Instructions

You can follow these [Pseudonymization-instructions](#) to walk through a basic example of pseudonymizing a data set. The exercise uses [a sample data set found in the data folder of the online guide](#).

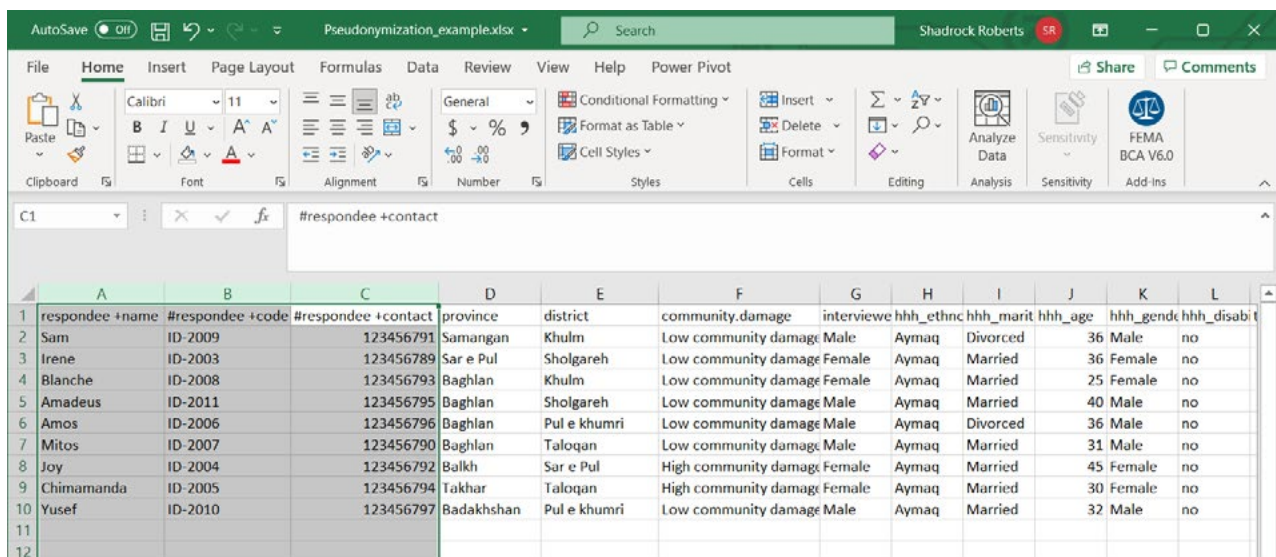
Once you have pseudonymized the sample data, you can continue with the Humanitarian Data Centre’s tutorial [for conducting a disclosure risk assessment](#).

Step 1 - Identify PII

Start by identifying PII in the data. Ideally, you will have metadata—data or a document that defines your data—to help you understand which fields contain PII. In the sample data, there are three columns that contain potential PII:

- › **#respondee +name** appears to contain a name.
- › **#respondee +code** likely contains an identification number of some kind.
- › **#respondee +contact** possibly contains a mobile phone number

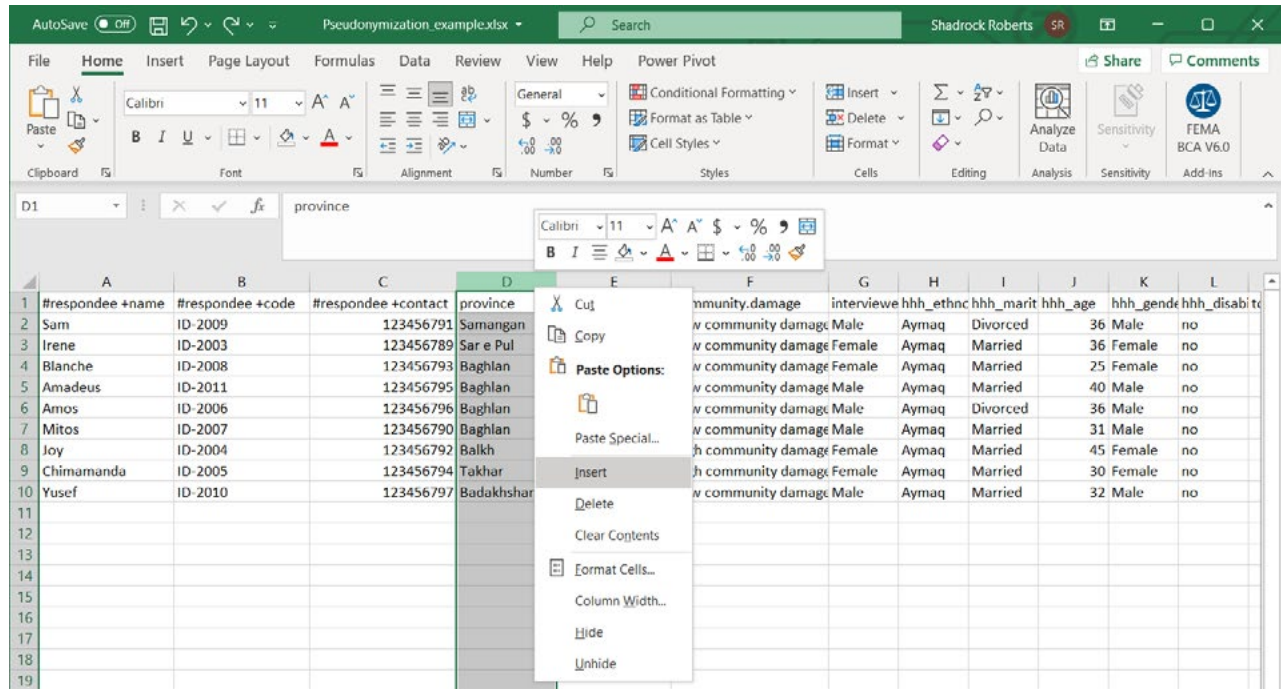
Each of these direct identifiers uses the [Humanitarian Exchange Language for tagging data](#).



	A	B	C	D	E	F	G	H	I	J	K	L
1	respondee +name	#respondee +code	#respondee +contact	province	district	community.damage	interviewee	hhh_ethnc	hhh_marit	hhh_age	hhh_gendr	hhh_disabi
2	Sam	ID-2009	123456791	Samangan	Khulm	Low community damage	Male	Aymaq	Divorced	36	Male	no
3	Irene	ID-2003	123456789	Sar e Pul	Sholgareh	Low community damage	Female	Aymaq	Married	36	Female	no
4	Blanche	ID-2008	123456793	Baghlan	Khulm	Low community damage	Female	Aymaq	Married	25	Female	no
5	Amadeus	ID-2011	123456795	Baghlan	Sholgareh	Low community damage	Male	Aymaq	Married	40	Male	no
6	Amos	ID-2006	123456796	Baghlan	Pul e khumri	Low community damage	Male	Aymaq	Divorced	36	Male	no
7	Mitos	ID-2007	123456790	Baghlan	Taloqan	Low community damage	Male	Aymaq	Married	31	Male	no
8	Joy	ID-2004	123456792	Balkh	Sar e Pul	High community damage	Female	Aymaq	Married	45	Female	no
9	Chimamanda	ID-2005	123456794	Takhar	Taloqan	High community damage	Female	Aymaq	Married	30	Female	no
10	Yusef	ID-2010	123456797	Badakhshan	Pul e khumri	Low community damage	Male	Aymaq	Married	32	Male	no
11												
12												

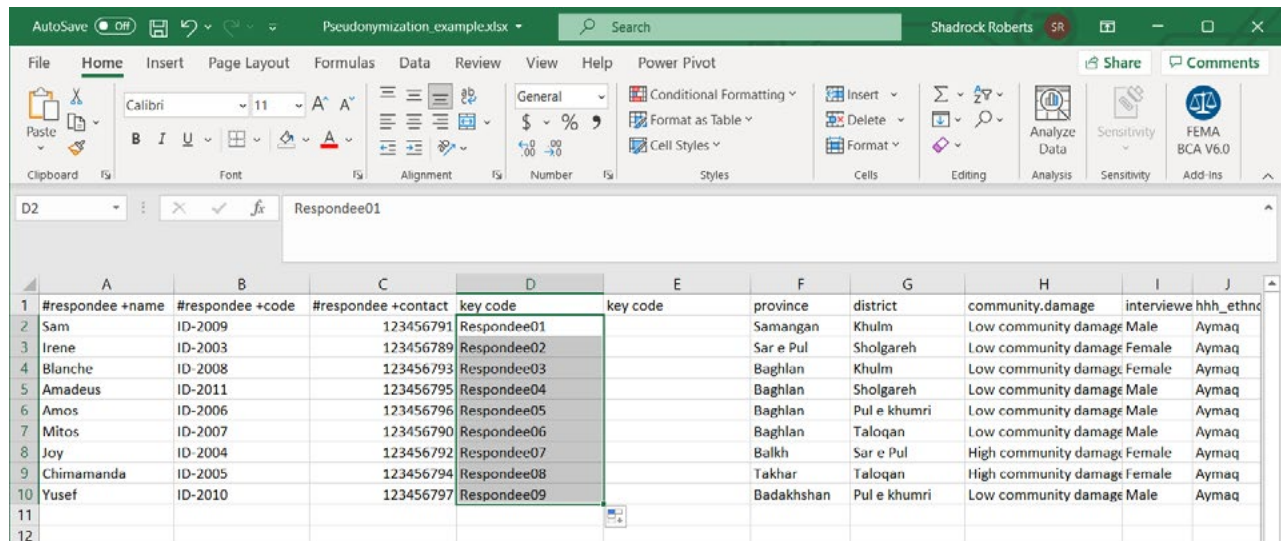
Step 2 - Create New Columns for the Key Code

We will use a key code, a value that we generate, to break out the PII. Since the direct identifiers are all grouped together, we'll create two new columns between columns C, **#respondee +contact** and column D, **province**. In Excel, we do this by highlighting a column to the right of where we want to insert new columns, right-click on the column and select **Insert**. Repeat this process again to create another empty column.



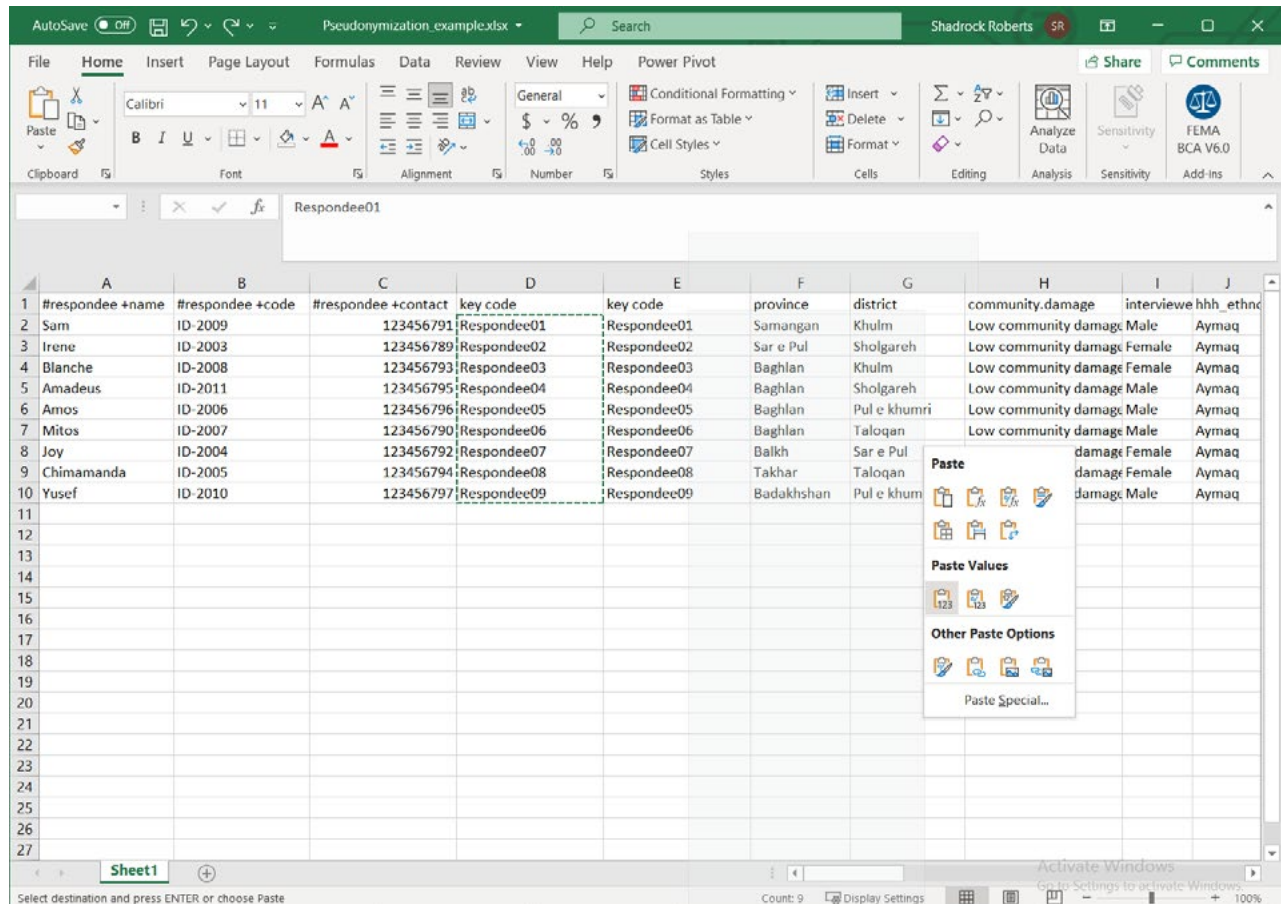
Step 3 - Create the Key Code

Start by naming your new columns. We'll use "key code" in each of them: each column will hold the same values. This would be a good time to update any metadata about this dataset to explain what **key code** means! Next, we'll use **Excel's Auto Fill feature** to create a simple code. Type **Respondee01** in the first cell. Next, highlight that cell, click on the drag handle in the lower right corner of the cell, and drag down to the end of the data set. This will automatically fill in the final number of each record so that each respondee now has a new code.



Step 4 - Duplicate the Key Code and Remove Formulas

Now we will copy the key code and paste it into the adjacent column. You can do this using basic keyboard commands such as **ctrl + C** or highlight the cells you want to copy, right click on them, and select **Copy**. In the adjacent column highlight the cells you want to paste the new key code into, right-click, and choose **Paste**. I've chosen to specifically paste only values. If you have used a formula to create a new code, then it will be important to retain *only the values* for use as a key code!



The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J
1	#respondee +name	#respondee +code	#respondee +contact	key code	key code	province	district	community.damage	interviewe	hhh_ethnc
2	Sam	ID-2009	123456791	Respondee01	Respondee01	Samangan	Khulm	Low community damage	Male	Aymaq
3	Irene	ID-2003	123456789	Respondee02	Respondee02	Sar e Pul	Sholgareh	Low community damage	Female	Aymaq
4	Blanche	ID-2008	123456793	Respondee03	Respondee03	Baghlan	Khulm	Low community damage	Female	Aymaq
5	Amadeus	ID-2011	123456795	Respondee04	Respondee04	Baghlan	Sholgareh	Low community damage	Male	Aymaq
6	Amos	ID-2006	123456796	Respondee05	Respondee05	Baghlan	Pul e khumri	Low community damage	Male	Aymaq
7	Mitos	ID-2007	123456790	Respondee06	Respondee06	Baghlan	Taloqan	Low community damage	Male	Aymaq
8	Joy	ID-2004	123456792	Respondee07	Respondee07	Balkh	Sar e Pul	damage	Female	Aymaq
9	Chimamanda	ID-2005	123456794	Respondee08	Respondee08	Takhar	Taloqan	damage	Female	Aymaq
10	Yusef	ID-2010	123456797	Respondee09	Respondee09	Badakhshan	Pul e khumri	damage	Male	Aymaq

The 'Paste' context menu is open over cell E10, showing options: Paste, Paste Values, and Other Paste Options. The 'Paste Values' option is highlighted.

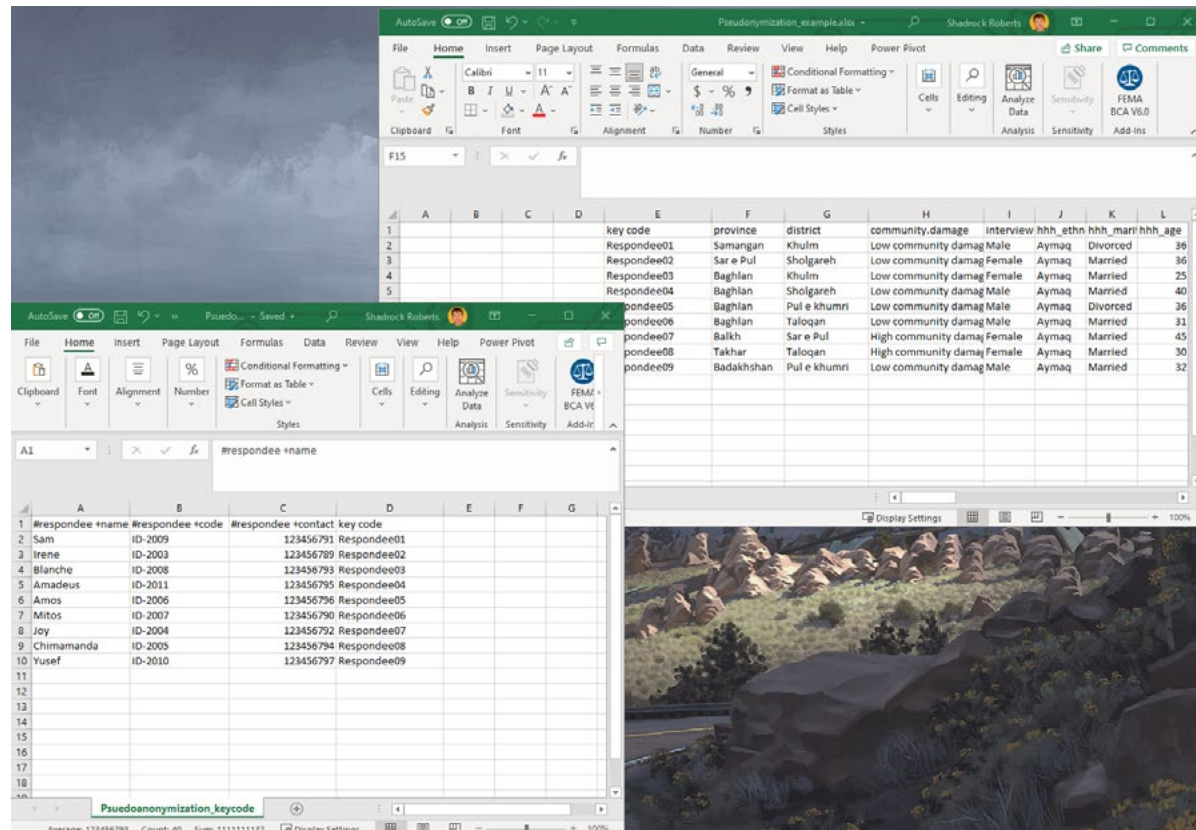
Step 5 - Separate Direct and Indirect Identifiers

Highlight the columns that contain the direct identifiers with PII along with one of the key code columns. In this example, we are highlighting columns A-D. Right-click on them and select **Cut**.

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	#respondee +name	#respondee +code	#respondee +contact	province		community.damage	interviewe	hhh_ethnc	hhh_marit	hhh_age	hhh_gende	hhh_disabi
2	Sam	ID-2009	123456791	Samangan		w community.damage	Male	Aymaq	Divorced	36	Male	no
3	Irene	ID-2003	123456789	Sar e Pul		w community.damage	Female	Aymaq	Married	36	Female	no
4	Blanche	ID-2008	123456793	Baghlan		w community.damage	Female	Aymaq	Married	25	Female	no
5	Amadeus	ID-2011	123456795	Baghlan		w community.damage	Male	Aymaq	Married	40	Male	no
6	Amos	ID-2006	123456796	Baghlan		w community.damage	Male	Aymaq	Divorced	36	Male	no
7	Mitos	ID-2007	123456790	Baghlan		w community.damage	Male	Aymaq	Married	31	Male	no
8	Joy	ID-2004	123456792	Balkh		w community.damage	Female	Aymaq	Married	45	Female	no
9	Chimamanda	ID-2005	123456794	Takhar		w community.damage	Female	Aymaq	Married	30	Female	no
10	Yusef	ID-2010	123456797	Badakhshan		w community.damage	Male	Aymaq	Married	32	Male	no

Next, open a new spreadsheet and paste these values using the keyboard shortcut **ctrl + V**, or other method. Save the new spreadsheet. You now have two spreadsheets: one of them contains indirect identifiers while the new sheet contains the direct identifiers with PII. Both datasets contain a key code for each record in the data so that all data can be recombined when necessary.



Next Steps

Both files contain a key code that will allow them to be put back together. One way to do this in Excel, is to use the **VLOOKUP** function to automatically populate cells based on the value of other cells. In this case, you could populate empty cells in the original file with the missing PII based on the **keycode** value.

Because the new file contains the direct identifiers containing PII, it must be stored securely. One excellent way to do this is to encrypt the file and to use cloud storage to limit who has access to the file (**see the Encryption and File Sharing Best Practices guides**).

Remember: while the original spreadsheet has been deidentified by removing the direct identifiers that contain obvious PII, the other indirect identifiers have the potential to be combined with other data or analyzed in such a way as to allow for an individual to be identified.

For this reason, both files should still be stored securely. If you wanted to share the original, non-PII, file more widely it would be critical to perform a *disclosure risk assessment* to ensure the minimum amount of risk that the data could be re-identified. The Humanitarian Data Centre has an [online tutorial for conducting](#)

a [disclosure risk assessment](#) using the [open source statistical software “R”](#). Additionally, [Poverty Action Lab’s De-identification for data publication](#) web page provides an excellent discussion of data de-identification and includes sample code for the [statistical software Stata](#). For Mercy Corps staff, [Draft Guidance from T4D](#) is available internally and provides additional Excel formulas.

Finally, all of these steps together help mitigate risk or exposing PII, so they should be listed in the PIA ([see the Privacy Impact Assessment guide](#)) so that others understand how these data are being protected.

Further Assistance

Deidentifying data is part of good data management practices and the larger data life cycle, which is the overall activities for individual data collection as part of a program or response. The following resources are excellent places to start for a more complete understanding of managing your data responsibly.

- › The Cash Learning Partnership’s Data Responsibility Toolkit is designed for cash and voucher practitioners specifically, but is a gold standard in guidance for responsible data. The Toolkit is available in [English](#), [Arabic](#), [French](#), and [Spanish](#).
- › The Electronic Cash Transfer Learning Action Network’s [Data Starter Kit for Humanitarian Field Staff](#) provides a series of data tip sheets for understanding various aspects of good data management and protection practices.
- › The International Committee of the Red Cross’ [Handbook on Data Protection in Humanitarian Action](#) is a detailed guide to almost every aspect of humanitarian data. Chapter 2 specifically deals with deidentifying data.
- › The Engine Room’s [Handbook of the Modern Development Specialist](#) is a good overview of data in the context of international development activities. The section on [Sharing Data](#) specifically deals with de-identification.

Encrypting a File

This section addresses a basic example of encrypting a file using a Microsoft feature available on Mercy Corps computers. There are a range of factors to consider when encrypting a file, but here we focus on using a password and encryption of a single file. See the links below for resources that explore encryption in more depth. For this guide, however, it is useful to understand the subtle difference between “password protection” and “encryption”.

Think of password protection as a box with a lock on it. When you “password protect” your document, you are putting it in an electronic box and locking it with a password: only those with the password can open the box. However, if the password you choose is not very strong, or if it is shared with the wrong person, someone can easily get into the box and view your document! By contrast, encryption uses complex algorithms to encode information, which requires having a key to decrypt that information. Think of this as taking your document and running it through a paper shredder that assigns a special key to put the document back together again.

When you combine password protection and encryption, you are effectively doubling your protection. If someone successfully breaks the password to the electronic box, they will only be able to see the bits of shredded paper without also having the proper key. All Mercy Corps’ laptops are encrypted using Microsoft BitLocker. This prevents a Mercy Corps laptop’s hard disk from being removed and accessed on another computer.

☆ Importance

Encryption is critical because it helps ensure the privacy and security of information. Without encryption, data can be intercepted and read by anyone who has access to it. When considering whether or not to encrypt data, ask yourself, “What is the risk to Mercy Corps’ program participants, staff, and partners if these data were lost or stolen?” A good rule of thumb is to encrypt anything containing personally identifiable or sensitive information.

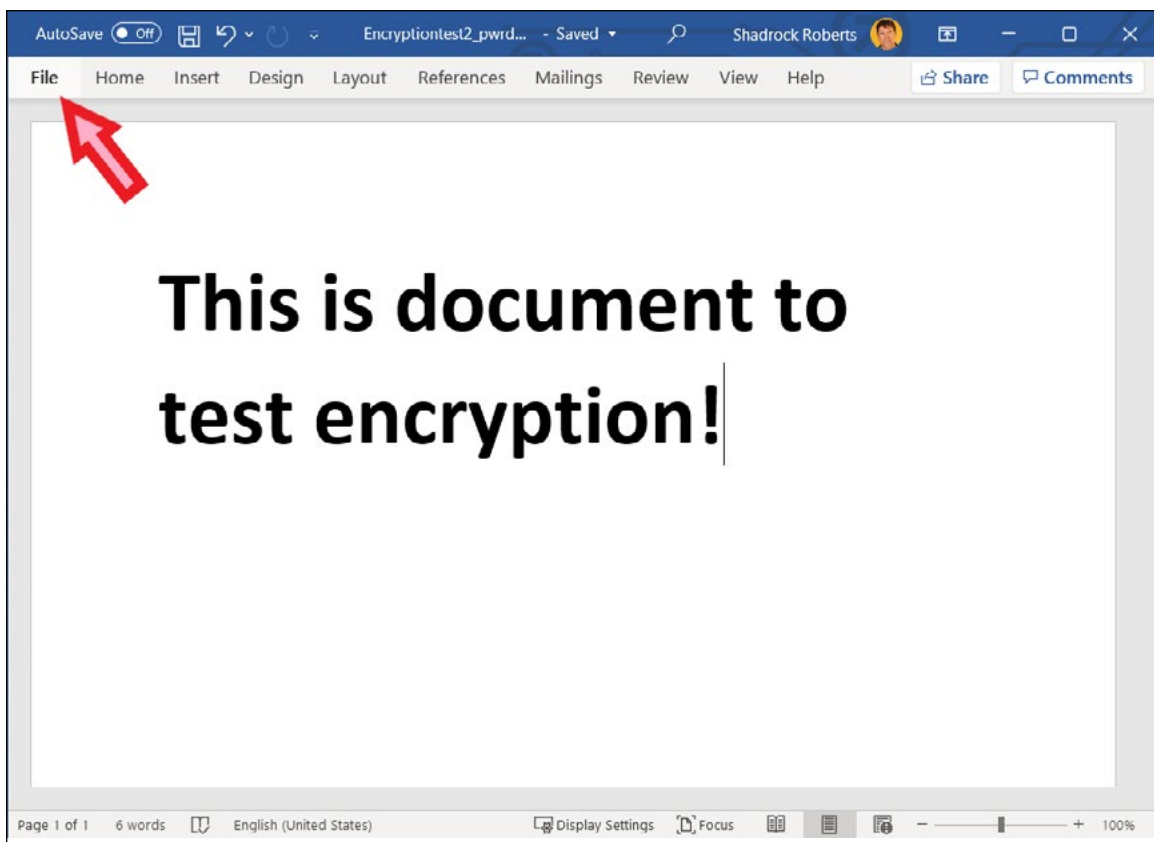
Principles

- › Use approved Mercy Corps systems for encrypted data transfer and storage (e.g. Microsoft SharePoint or Google Drive). When in doubt, ask your local IT team for advice.
- › Encrypt sensitive data at all stages of its collection, use, transmission, and storage.
- › Use strong passwords and do not reuse passwords. Lists of passwords circulate online and make it easier for someone with one of your passwords to access more than one of your accounts or files! You may want to use a password manager, such as Lastpass. However, password managers may be vulnerable to cyber attack by fake apps, so it is critical that password managers are used as part of a broader approach to securing data.

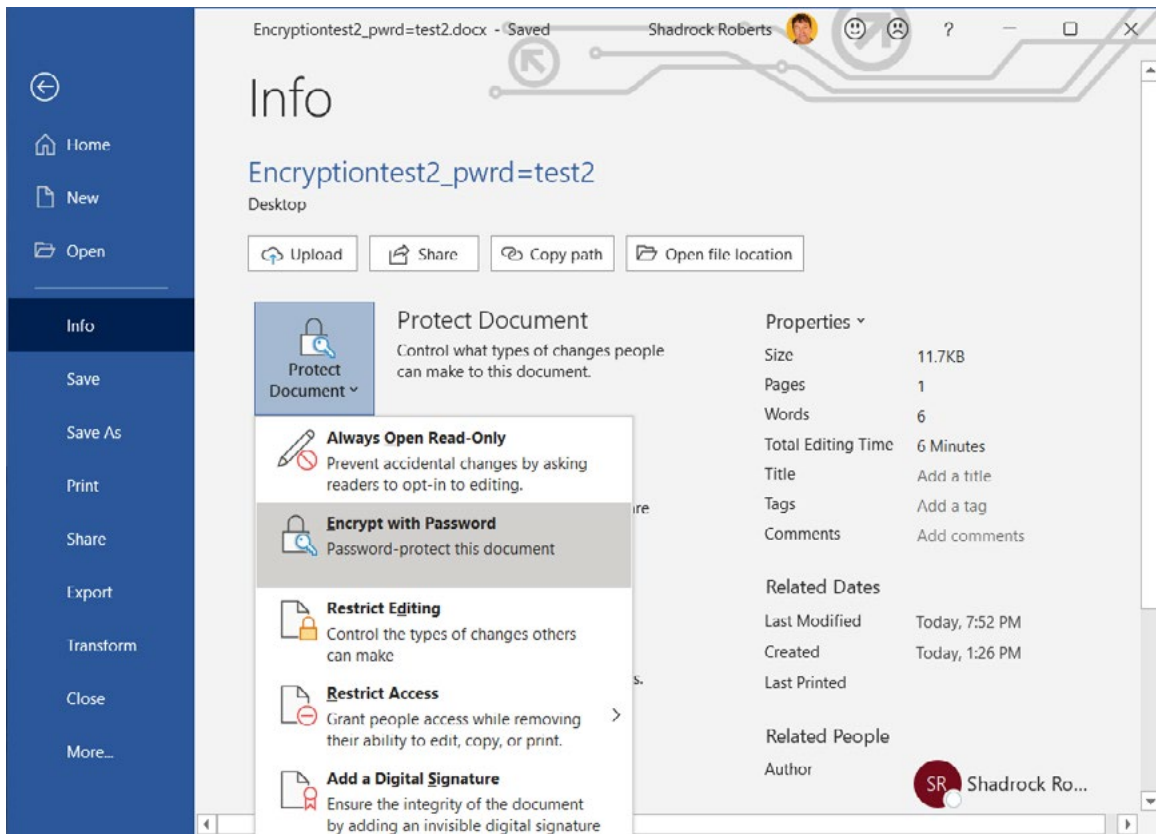
- › In a team environment, encryption is only as good as the weakest link. If even one person fails to use encryption, your program data is at risk. It is extremely important to communicate this to your team: encryption is not just an issue of technology, but of behavior change as well.
- › Understand laws that govern encryption in your country. Local laws in a number of countries (such as Sudan, Yemen and Pakistan) place limits on encryption software. When in doubt, ask your local IT team for advice: in general they will work with you to ensure that your computers' hard drive is appropriately encrypted using Intune.

Instructions

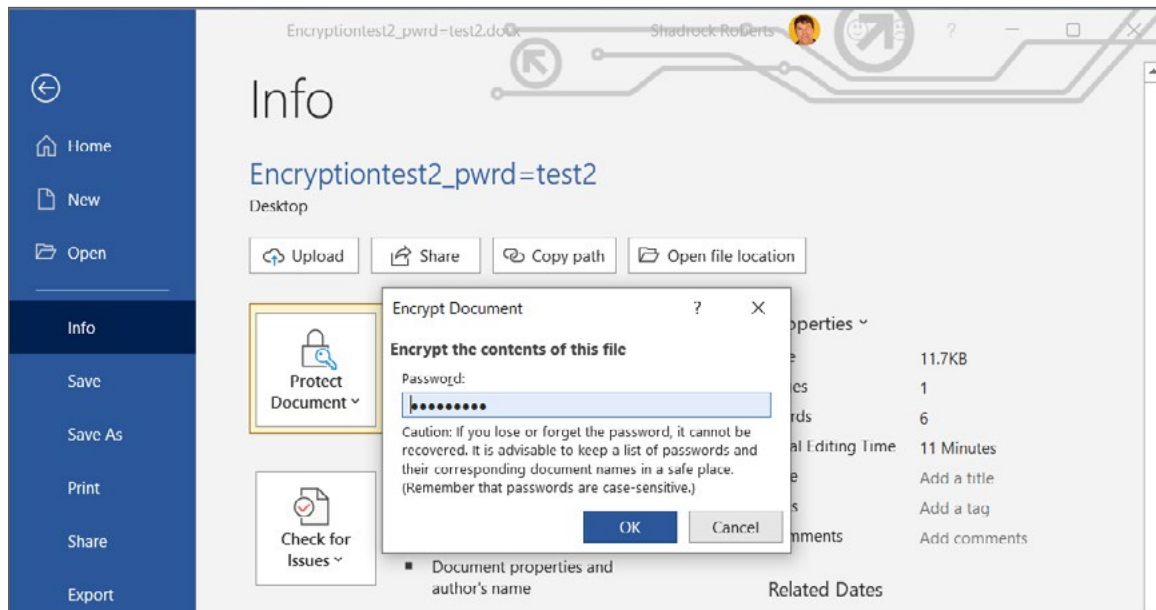
- 1 Open the Word, Excel, or PowerPoint file you want to encrypt and select the **File** menu.



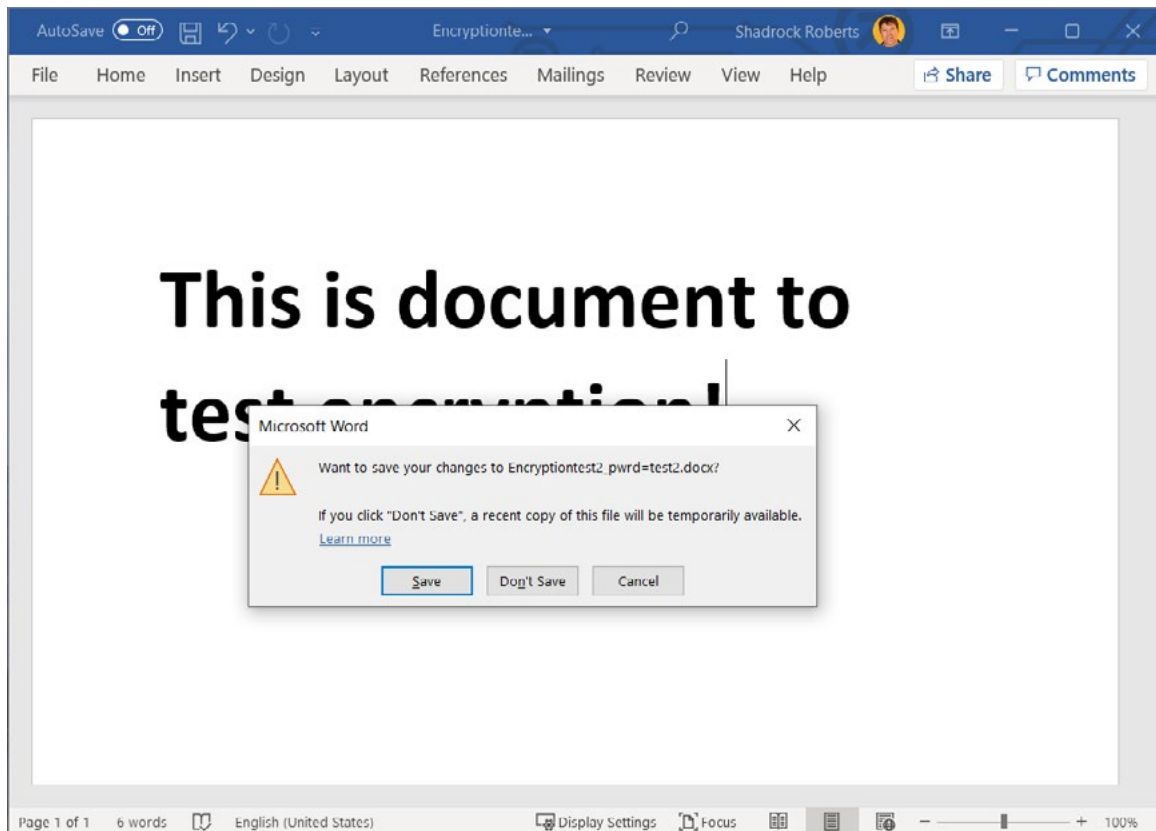
2 Navigate to **Info > Protect Document > Encrypt with Password**.



3 Type a password, click **OK** then type it again to confirm it.



4 Save the file to make sure the password takes effect.



Now you can share the file and the password with those who require access. It is best practice to place the file on a Mercy Corps approved cloud service such as G Suite or SharePoint. Remember to send the file link and password links separately. For example, you could share the file using Google Drive (**see the File Sharing section**) and generate a notice that the file has been shared via Google, then share the password via an email to a colleague.

Further Assistance

- › The Electronic Cash Transfer Learning Action Network's [Data Starter Kit](#) provides a tip sheet for encryption (see Tip sheet #5).
- › The Electronic Frontier Foundation provides a more [detailed look at various forms of encryption](#).
- › The Engine Room's [Hand-Book of the Modern Development Specialist](#) includes a section on Managing Data that provides additional, high-level, thoughts on encryption.

CONTACT

HEATHER LOVE

Director, Global Data Protection and Privacy | IT
hlove@mercycorps.org

SHADROCK ROBERTS

Data Protection Specialist | IT
shroberts@mercycorps.org

About Mercy Corps

Mercy Corps is a leading global organization powered by the belief that a better world is possible. In disaster, in hardship, in more than 40 countries around the world, we partner to put bold solutions into action—helping people triumph over adversity and build stronger communities from within. Now, and for the future.



Global Headquarters

45 SW Ankeny Street
Portland, Oregon 97204
888.842.0842
mercycorps.org

European Headquarters

40 Sciences
Edinburgh EH9 1NJ
Scotland, UK
+44.131.662.5160
mercycorps.org.uk