

What is a database and do I need one?

Goal of this document: This document provides an explanation about what a database is, when you would need it, and why it's important. Through examples and hands-on activities, the reader will be able to identify how a database could be useful to their work.

Audience: Human rights defenders who are interested in using a database to manage their information about human rights violations.

Contribute: Please feel free to share this resource if you find it helpful or know someone who might have experiences to add.

Status: This is a living document, and everyone is welcome and encouraged to add to it. Ideas that need some expansion are flagged with a sprout. 🌱 Last updated May 6, 2022

Table of contents

How to use this resource	2
What is a database?	2
How have human rights defenders benefited from databases?	3
How do you know when you need a database (and when you don't)?	6
Using spreadsheets to manage human rights information	7
Where it gets complicated to use spreadsheets...	8
What does it take to build a database?	10
Scenario: when to stick with a spreadsheet and when to use a database	12

How to use this resource

Our intention with this resource is to help you think about your current information management system and the ways in which you could improve it so that it is easier to use. In the process of monitoring, documenting, and reporting human rights violations, you will inevitably gather a large amount of information that you will need to sort through and organise. It will also be necessary to ensure that the information is accessible in a way that allows you to communicate the information effectively to others. You already know that your data tells a powerful story. This resource will help you think about how to organise your data so that it can tell that story clearly.

The first section of this resource provides an explanation of what a database is, and the second section will explore some examples of how human rights organisations have used databases strategically. Next, we provide guidance on how you would know if you need a database and when you would need it, and we use spreadsheets as the main alternative information management solution. Finally, we provide a framework that you can use to start to think about the process of developing a database and the costs associated with this process.

What is a database?

A database can be defined very simply as: **compiled and structured data that serves some purpose**. It is a system that allows users to store, maintain, and access information. As a home for all of the information gathered, databases allow us to observe and analyse our information in one central place.

Keeping an organised collection of information can be helpful in that it provides:

1. an accessible and centralised place to store information you want to hold on to, and
2. a way to structure data that will enable the strategic use of information.

Data is considered 'structured' when information is organised according to a predefined format, with set categories for different types of data that make the information easily searchable. Structured data is powerful because it can harness computing power to sort, match, link, calculate, and aggregate information.

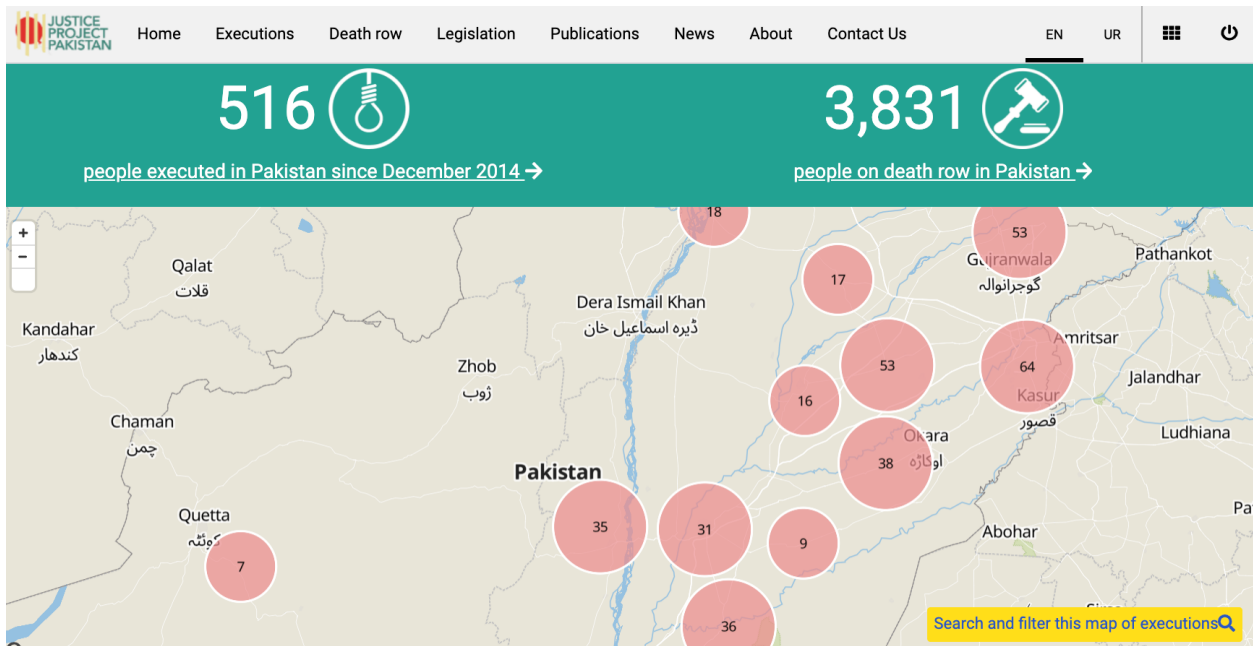
Structured data lets you see your information in different ways. Data is easier to share and allows collaboration when it is structured. Keeping a structured database can also save an organisation a lot of time and money in the long run by reducing repetitive tasks, providing powerful analytical tools, and maintaining long term organisational or community memory. There are many different [types of databases](#), but the most common types of databases that human rights organisations use these days are: relational databases and NoSQL databases.

How have human rights defenders benefited from databases?

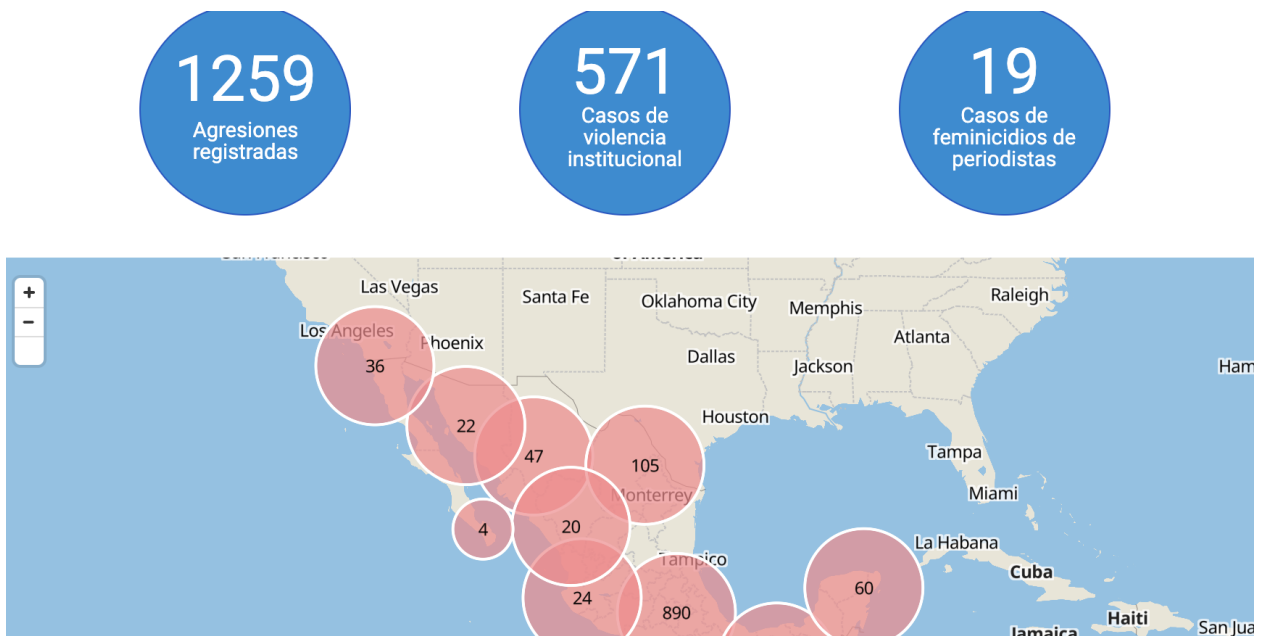
For information about human rights violations to be useful, it needs to be in the hands of activists, advocates, policymakers, and researchers who can leverage it for real world change. From police reports to eyewitness accounts to legal documents, human rights defenders must be able to easily access the information they need in order to bolster and illustrate their claims.

By organising their information so that it is accessible and easy to use, human rights defenders have given their communities the power to expose the truth and advocate for change, preserve memory, and seek justice and accountability.

Below are a few examples of human rights databases that was designed with these goals in mind:

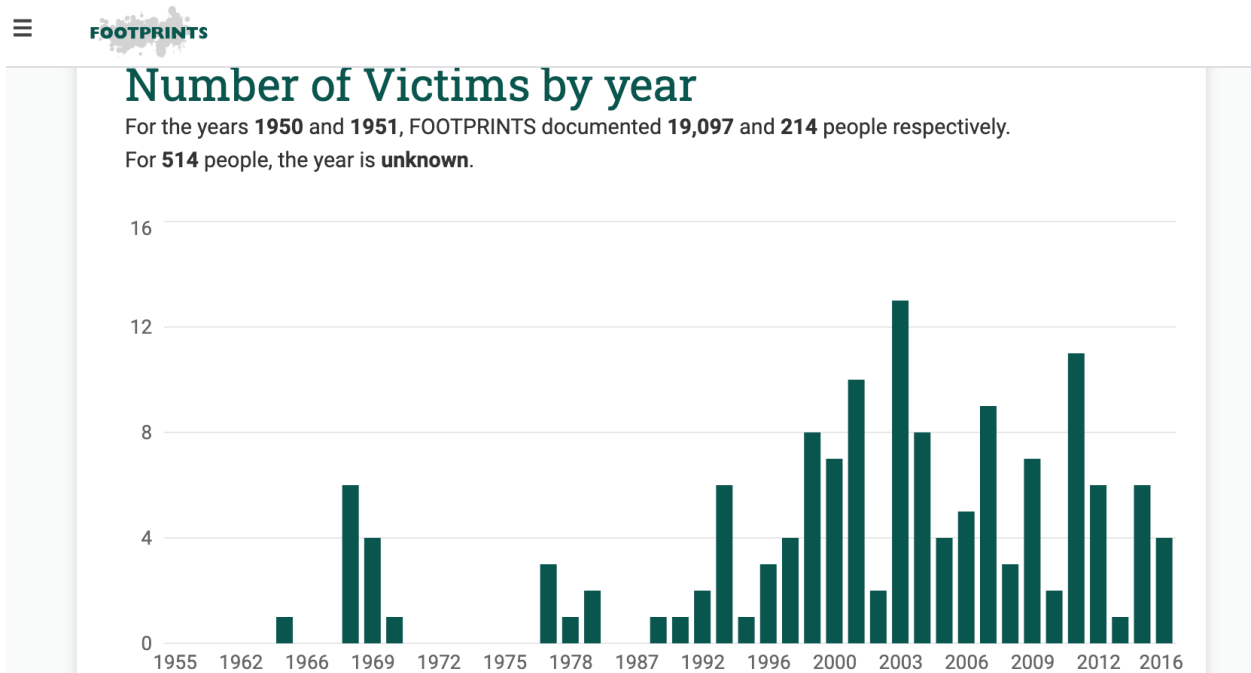


The [Death Penalty Database](#) by Justice Project Pakistan helps defenders **advocate for change** by gathering details about current death row prisoners and information about the 513 prisoners who have been executed in Pakistan since 2014.

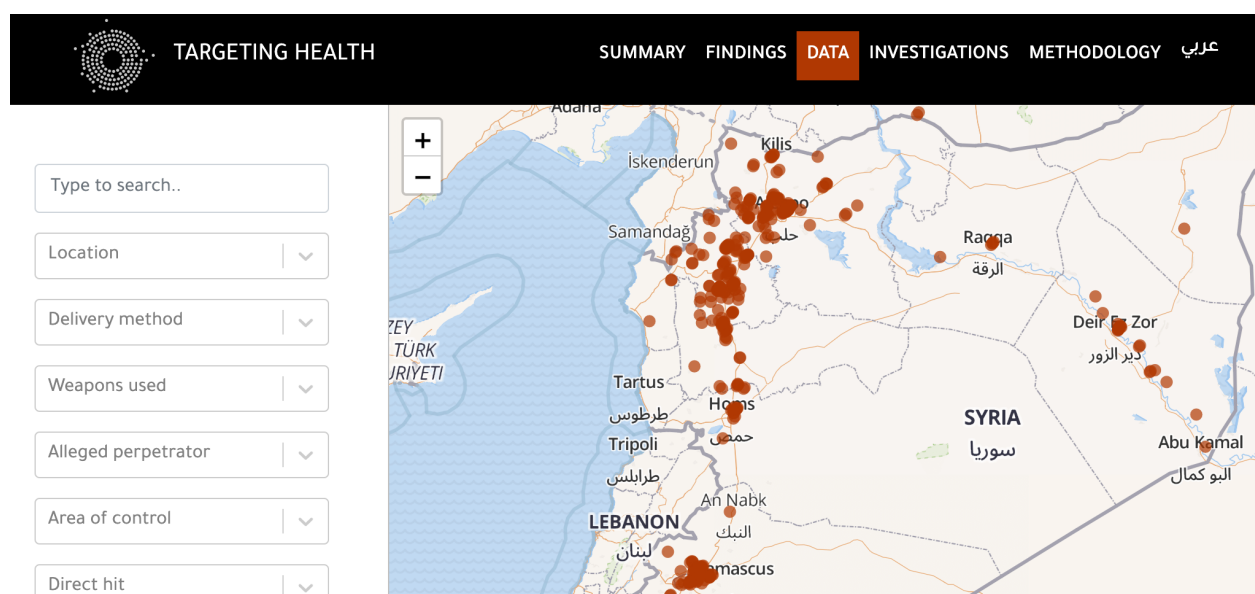


The [Cartography of Attacks Against Women Journalists in Mexico](#) by the Civil Association for Women's Information and Communication (CIMAC) is an online database that offers a comprehensive history of the dangers that women journalists have faced — and continue

to face. This database helps defenders **advocate for change** by monitoring and publishing attacks, and helping journalists avoid this violence.



[Footprints](#) is a database by the Transitional Justice Working Group that contributes to the **preservation of memory** by compiling cases of arbitrary detention, abduction and enforced disappearances committed in and by North Korea, stretching back to just before the outbreak of the Korean War.



The [Syrian Archive](#) is a Syrian-led project that aims to preserve, enhance and memorialise documentation of human rights violations and other crimes committed by all parties to conflict in Syria for use in **advocacy, justice and accountability**.

You can find more examples in the [HURIDOCs Project Spotlight series](#).

How do you know when you need a database (and when you don't)?

	CDCZ	CDDA	CDD B	CDDC	CDD D	CDD E	CDD E	CDD G	CDD H
34968	31,234	BQRS	37	19,24	18	38,7	17,64	27,93	0
34969	39,498	BA-3	28	24,65	30	41,8	34,57	17,47	1
34970	47,893	BRLN	49	27,28	34	0	42,81	28,73	0
34971	28,321	ADAM	65	16,78	17	11,9	39,78	11,7	1
34972	24,78	MADM	33	19,4		12,4	22,4	28,7	1
34973	51,7	DAMA	24	28,49	23,7	17,8	17,89	65,2	0
4974	18,38	ARY2-7	00	22,23	149	1,6	33,74	83,1	1
34975	31,74	BQ47	07	21,98	223	20	42,89	11,7	0
34976	29,7	K3P4	14	22,32	145	17,3	89,73	20,8	1
34977	0	M743	73	20,5	183	19,6	19,49	34,8	1
34978	51,4	MLOK	41	17,97	531	14,1	18,17	78,4	0
34978	53,9	MIK2	61	19,51	724	30,18	51,24	90,3	1
34979	29,78	ZY4Z	88	12,13	553	21,9	32,1	6	1
34980		A/GEN	39	16,59	873	28,7	34,4	12	0
34981	11,73	AAT	23	32,86	781	65,3	24,7	19,7	1
34982	11,	73VR	15	11,23	231	86,4		28,3	0
34983	14,7	WXD	37	20,7	174	79,1		24,7	0
34983	14,	NOT	12	1,79	107	37,5		18,3	0
34984		HORSE	20	24,8	23,5	83,4		19,	1

Image source:

http://web.archive.org/web/20160606232900im_/https://drawingbynumbers.org/sites/dev.drawingbynumbers.org/files/Numbers2-hires-20120316.jpg

If you're new to information management and databases, it's unlikely that you would intuitively know when it is the right time to switch to a database. In this section, we provide a comparison between a spreadsheet and a database, and what to look out for when you have outgrown your spreadsheet.

Using spreadsheets to manage human rights information

It's very common for human rights defenders to begin recording information about human rights violations in spreadsheets. It makes sense to start this way — a spreadsheet provides a way to organise, filter, search and share the information you have. With some extra effort, you can create dropdown lists in cells to ensure consistency in the terms you use. With a significant amount of effort, you can even [create a database using a spreadsheet](#). But there will likely be a time when the information you want to organise is too complex for a spreadsheet.

Consider this real-life situation from Data & Design How To's, by Tactical Tech:

"...a group we have worked with that documents prisons uses a spreadsheet to track over 40 pieces of data about each prison, including address, capacity and security status. They make numerous updates each day. The source of each piece of data is also recorded, which doubles the number of columns. Over time this increases the amount of time and frustration accrued in managing, analysing and sharing the growing body of information. What started as a simple list, appropriate for a spreadsheet, becomes a major data collection and analysis initiative for which a single spreadsheet soon proves inadequate."

Now, consider this list of redflags found in the Drawing By Numbers guide by Tactical Tech:

Five signs you might have grown out of your spreadsheet¹:

- *You start colour-coding things in the spreadsheet and have created little 'hacks' (like adding 'AAA' or '!!!!' to a row of data to ensure it appears at the top) to find data.*
- *You scroll around a lot to find and edit information or perhaps you have bought a bigger computer monitor so you can see more data on screen.*

¹ Drawing By Numbers guide by Tactical Tech, Data & Design How-to's Note 2: Data basics <http://web.archive.org/web/20160606232900/https://drawingbynumbers.org/data-design-basics/notes-2-data-basics>, Accessed November 2021 to January 2022

- *Different people need to enter data into the spreadsheet so you spend time emailing it around and copy-pasting data into a 'master' spreadsheet.*
- *You regularly have to reformat to fit the needs of different tools to make charts, maps or graphs.*
- *You create multiple spreadsheets to keep count of data in other spreadsheets.*

If you are doing any of the above, it is time to start thinking of a different type of tool.

There isn't a magic formula that will tell you it is time to move from a spreadsheet to a database, unfortunately. Instead, it requires consideration of the pros and cons of each option with the ability to reasonably predict the size and shape of this information management project. Then, it is about determining the right balance between what you need to reach your goals versus what effort and resources you are willing to invest. (We will go more into what it takes to build and maintain a database later on in this resource.)

Where it gets complicated to use spreadsheets...

HURIDOCS and its network developed the [Events Standard Formats](#) methodology (which we now call the *events method*) to provide a way to organise information about human rights violations. The purpose of the *events method* is to capture essential information with regard to *individual acts* of human rights violations to better understand patterns of violence, including "[who did what to whom](#)". It involves gathering information about:

- the facts: what happened, where, and when
- the possible human rights violations that were committed
- the persons involved: which alleged perpetrator did what to which victim, what are the sources of information and which interventions were made.

What is important to know about the *events method* at this point is that the smallest unit of analysis is a single act. For the sake of this resource, this simply means that for any single act, there may be multiple people involved. From our experience at HURIDOCS, we find that some form of this *events method* approach is helpful in organising information about human rights violations.

However, using this methodology requires something more than a simple spreadsheet. To explain why a spreadsheet falls short, please listen to the part of this presentation starting at 7:53 and ending at 12:50. It gives a clear explanation of the pitfalls someone would run into when trying to record information about a specific act.

[Embed this video starting at 7:53

 Community Discussion: The events method for documenting human rights violations]

To summarise, the example shared in the video demonstrates the following common pitfalls in trying to record events in a spreadsheet:

1. **Capturing more than one violation can be complicated.**
Often, the violation perceived as the most serious is the one that is included in the record. But so much valuable information is left out if this is the approach. Or, only a few violations are able to be captured in one record, also limiting what can be understood about the situation.
2. **Capturing more than one perpetrator can be complicated.**
There are many times when we need to capture more than one perpetrator in relation to a victim. But this is hard to do if you only have one row in a spreadsheet to work with.
3. **Connecting the appropriate perpetrator to the violation can be complicated.**
When all the information is organised according to one entity type (in this case, the victim entity), then it is difficult to capture nuanced information about other information surrounding the victims, such as the relationship between the perpetrator and the violation.
4. **You may end up duplicating the victim record.**
In order to capture all of this information, you may duplicate the victim record, which would lead to inaccurate counts.

Using a database will help you avoid many of these pitfalls related to data structure because you can use multiple entities to organise your information. Here's why this approach is helpful:

1. You can have one entity to represent information about the victim, another entity to represent information about the act/event that happened, and another entity to represent the perpetrator.
2. Using entities, you can create a record for the victim without having to create records for the event or perpetrator first. Furthermore, you can delete an incident record without losing data about the victim because the information lives in two different entities.
3. When updating data in entities, there is less chance of inconsistencies and errors. For example, if updating the address for a prison entity that is already related to multiple prisoner entities, you only need to do it once (in the prison entity).

- Using entities, you can associate multiple victims with a single record (e.g. incident record) without creating sorting and counting problems.

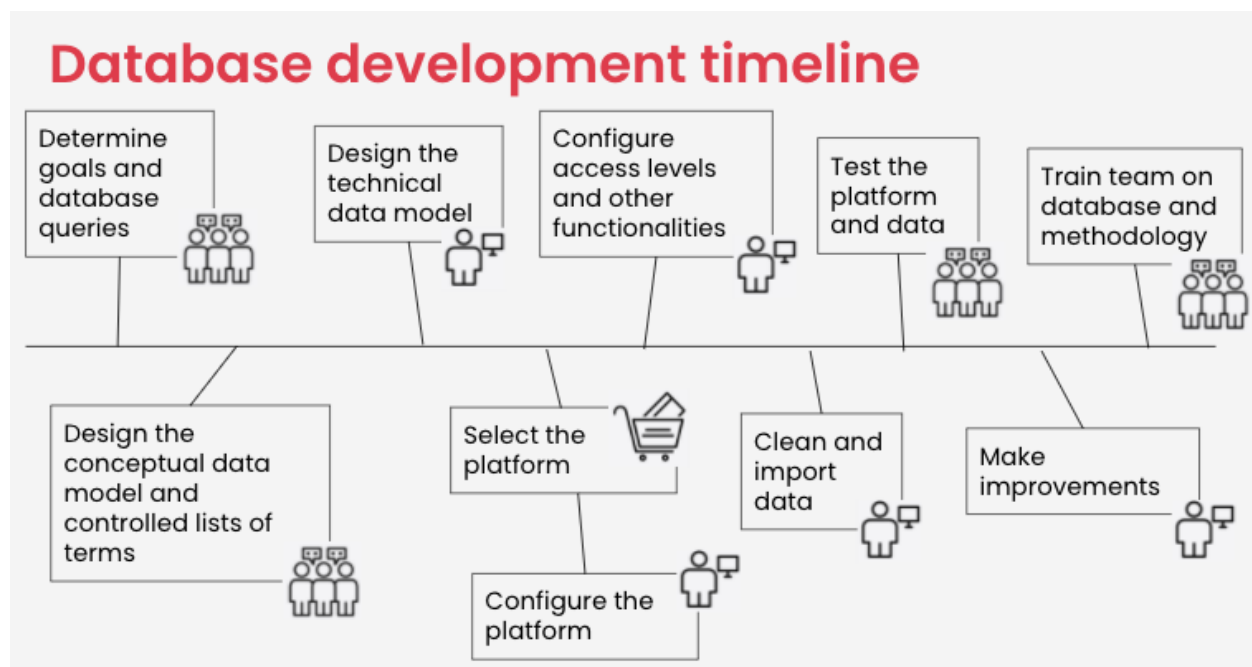
You will learn more about how to design your data model in the section titled

[\[Resource section\] Designing your conceptual data model \(living document\)](#) 🌱.

What does it take to build a database?

It's important to know up-front that building a database requires time, effort, and financial resources. For many groups, just building the database requires more than they have to invest. But it's really helpful to think this through and plan it out so that your team isn't shocked to learn: how long it will take, who needs to be involved, and how much it will cost.

To explore what it takes, we have put together this general database development timeline with the following stages:



You can think through each stage and what it would require in your context. Here are some general notes and considerations for each stage:

Stage	Effort and time, from whom	Financial
-------	----------------------------	-----------

		considerations
1. Determine goals and database queries	This is a collaborative process that will involve your primary audience (those that will be <i>using</i> your database), and would require a facilitator	Cost of salaried staff
2. Design the conceptual data model and controlled lists of terms	This is a collaborative process that will involve your primary audience, and would require a facilitator	Cost of salaried staff
3. Design the technical data model	This will likely require support from someone with experience developing data models.	Cost of consultant (if needed)
4. Select the platform	This will require some research, asking peers for advice, understanding your team's capacity for hosting and maintaining a platform.	Cost of purchasing the platform software. Cost of hosting the platform.
5. Configure platform (data model, access levels, etc)	This might require support from someone with experience configuring this platform.	Cost of consultant (if needed)
6. Clean and import data	This might require support from someone with experience cleaning and importing data. Gathering all your existing data might take time.	Cost of consultant (if needed)
7. Test the platform and data	This is a collaborative process that will involve your primary audience, and would require a facilitator	Cost of salaried staff
8. Make improvements	This might require support from someone with experience configuring this platform.	Cost of consultant (if needed)
9. Train team on database and methodology	This is a collaborative process that will involve your primary audience, and would require a facilitator	Cost of salaried staff to do the training, and to be trained

Scenario: when to stick with a spreadsheet and when to use a database

Let's say an organisation is recording the names of prisoners on death row in their country.

→ A spreadsheet would likely be a fine solution for organising this information if this organisation is only collecting simple information about this person, such as:

- ◆ Name
- ◆ Birthdate
- ◆ Gender
- ◆ Prison
- ◆ Crime allegedly committed

→ But, a database would be more appropriate than a spreadsheet if the organisation wanted to record information about any of these attributes. For example, it might be helpful to record information about the prison (e.g. location). Or, it might be helpful to capture information about the legal process that this person was involved in (e.g. courts, judges).

A spreadsheet is very simple (can be as simple as a table), accessible (available on most computers), and affordable (Google sheets are free). But the downside is that it has some significant limitations related to collaboration and data analysis (as explained above). While the downside of building and running a database is the amount of effort and resources it takes, the benefit is that, if done well, it can protect and preserve your hard-won information in a way that enables you to understand your data in the most useful way.



Next resource section:



[Resource section] Determining the goals and research questions (living document)