# Promoting ESC Rights of Disadvantaged Groups in Turkiye through Monitoring and Advocacy

## MODULE 2. DATABASES

### Theory

Conceptualising human rights information and building a data model

Funded by the European Union

AMER Association for Monitoring Equal Rights

huridocs human rights information and documentation systems

# Table of contents

# Foreword

Understanding what a database is, why your organisation needs it and how to build one is key to your information architecture journey, to the health and sustainability of your data and to the success towards reaching your final analysis and advocacy goals.

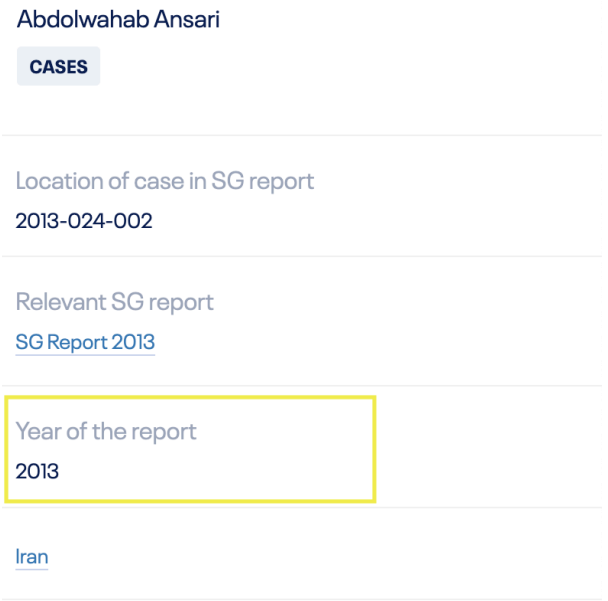
This document presents key information about Uwazi databases, utilised to store and systematise human rights information, and offers guidance throughout a data modelling journey to ensure that a data collection process is more efficient, organised and tailored to the goals of your information management project.

# Glossary of terms

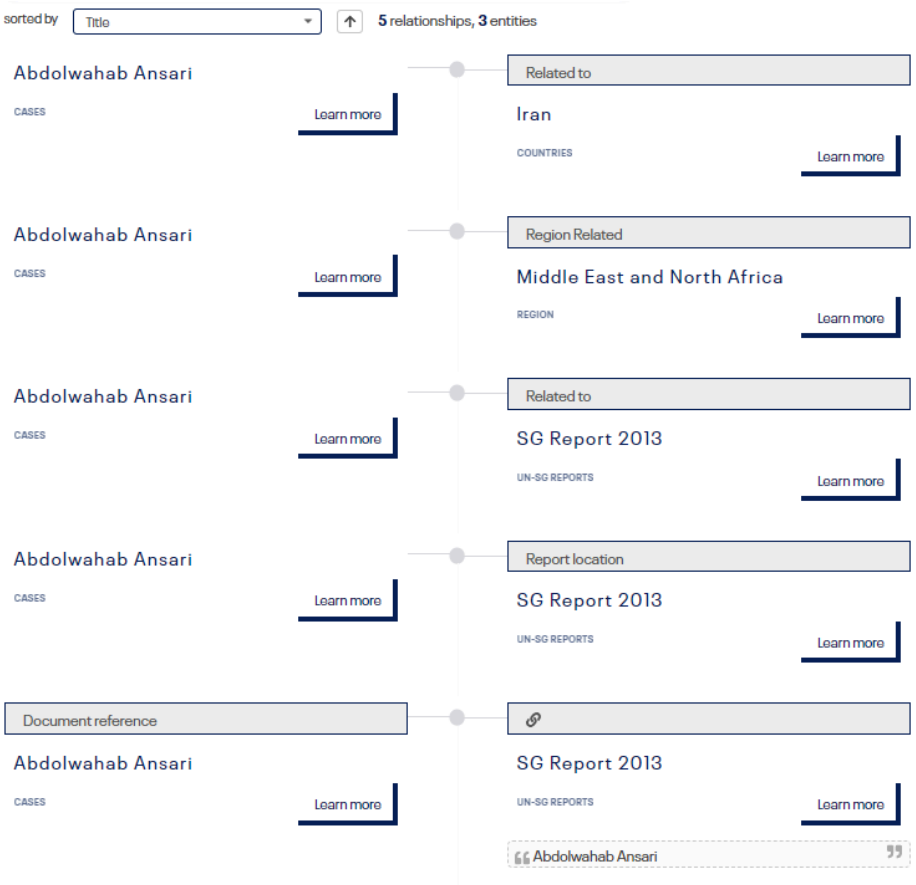
- **Data model** refers to the logical structure of your **collection** or, in other words, how the different pieces of data in your collection are organised and how they relate to one another.
- **Templates** are a starting point for new **entities** in some databases. They can be built and customised with different properties for each type of **entity**. **Templates** can be thought of as **buckets** of information.
- **Entities** are the real-life things such as people, objects, events, etc. that the database contains information about, such as 'Officers', 'Incidents', 'Videos', 'Prisoners', 'Prisons' etc. An **entity type** describes the type of the information that is being recorded such as 'Person', 'Prison', 'Event', etc. In some databases, an entity is a type of information such as court, person, event or otherwise referred to as a **record**. It contains **properties** and can have **relationships** to other **entities**.
    - In Uwazi, an open source database tool developed by HURIDOCS (Human Rights Information and Documentation Systems) entities[1] can look the following way:

[1] These entities are part of ISHR's End Reprisals library accessible at: https://endreprisals.ishr.ch/. Accessed October 2023.


- **Property** (or **attribute**) refers to aspects of the entities and/or relationships. Attributes will be displayed as fields on each entity in a database. For example, for a 'Person' entity, attributes could include: 'First name', 'Last name', 'Gender', or 'Ethnicity'.
    - In Uwazi, properties are contained within entities and have different types, as seen below:



- **Relationships** indicate how different entities are associated with each other. This can be real-life relationships between different things in your data model. It is often expressed using the formula noun-verb-noun (prisoner *detained at* prison). In Uwazi, a **relationship** allows **entities** to be connected to each other.
    - In Uwazi, relationships between entities can be viewed as connections, as seen below:




- **Thesauri** (or **Thesaurus)** refers to lists of terms that are referenced in **select** and **multiselect** properties.

# What is a database?

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

Keeping an organised collection of information can be helpful as 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 filter, sort, match, link, calculate, and aggregate information.

In addition to making data more accessible, there are other benefits to structuring data:

- Structured data lets you view your information in different ways.
- Data is easier to share and allows collaboration when it is structured and clean.
- Keeping a structured database can also save an organisation's future human and financial resources by reducing repetitive tasks, providing powerful analytical tools, and maintaining long-term organisational or community memory.

## Why do I need a database instead of a spreadsheet?

Deciding to move from a spreadsheet to a database requires you to consider 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, what information you actually have and what effort and resources you are willing to invest.

Let's say your organisation is recording the names of individuals whose social rights have been violated. A spreadsheet would likely be a fine solution for organising this information, if your organisation is only collecting simple information about each person, such as:

| |
|---|
| ◆ Name |
| ◆ Birthdate |
| ◆ Gender |
| ◆ Violated social right |
| ◆ Date when a violation occurred |

A **spreadsheet** is very simple (it can be as simple as a table), accessible (it is available on most computers), and affordable (e.g. Google Sheets are free). On the other hand, there are significant limitations related to collaboration, data connection, and analysis. Building a database can harness, protect and preserve your hard-won information in a way that enables you and your users to understand your data in the most useful way.

There are 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 can be captured in one record, also limiting a better understanding of complex situations.
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. Other

information related to the victim, such as the relationship between the perpetrator and the violation, can become lost.

4. **You may end up duplicating the victim record.**
   In order to capture all of this information, you may need to duplicate the victim record, which would lead to inaccurate counts and duplicity of facts.

A **database** would be more appropriate than a spreadsheet if your organisation wishes 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).

If the information you want to organise is too complex for a spreadsheet, consider this list of red flags found in the Drawing By Numbers guide[2] by Tactical Tech:

*Five signs you might have grown out of your spreadsheet:*

| |
|---|
| • *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.* |
| • *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.* |

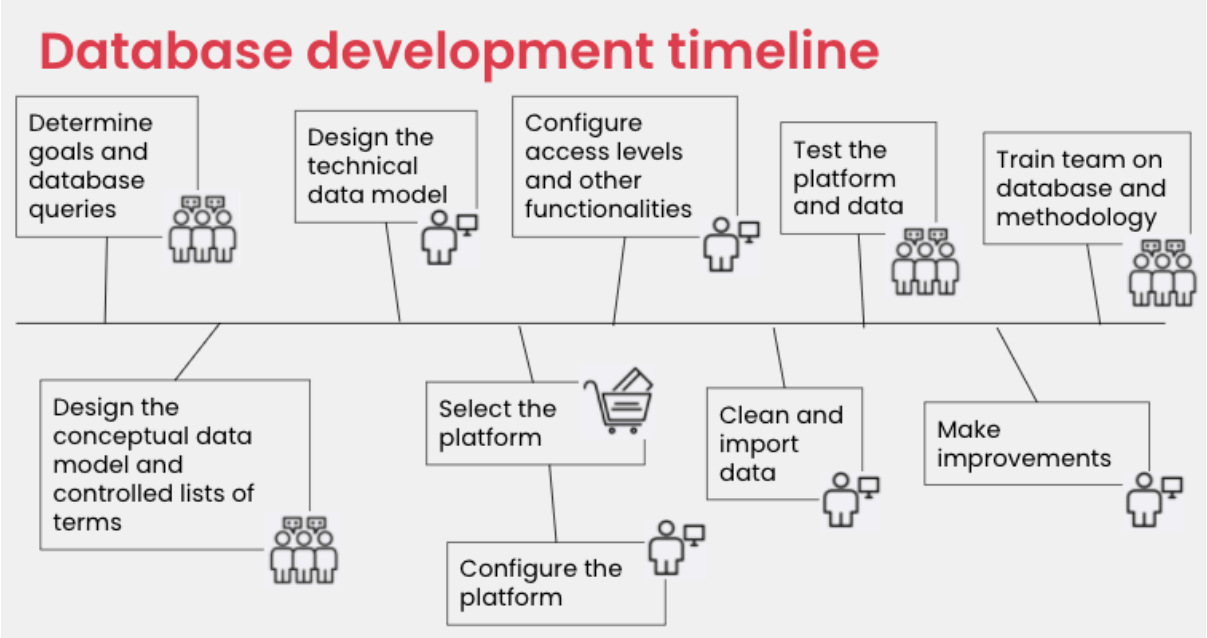
[2] Tactical Tech. "Data & Design How-to's Note 2: Data basics." Drawing by Numbers. Accessed November 2021 to January 2022. https://web.archive.org/web/20160606232900/https://drawingbynumbers.org/data-design-basics/note-2-data-basics.

- *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**.

If you decide to move from a spreadsheet to a database, you need to consider both the **pros** and **cons** of each option as well as predict the size and shape of your information management project. The next step is to determine the right balance between what you need to reach **your goals** versus what **effort** and **resources** you are willing to invest.

It's important to know up-front that building a database requires time, effort, and human resources. To explore what it entails, we have put together this general database development timeline with the following **nine 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 |
|---|---|
| **1. Determine goals and database queries** | This is a collaborative process that will involve your primary audience (those that will be using your database), and would require a facilitator. |
| **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. |
| **3. Design the technical data model** | This will likely require support from someone with experience in developing data models. |
| **4. Select the platform** | This will require some research, asking peers for advice, understanding your team's capacity for hosting and maintaining a platform. |
| **5. Configure the platform (data model, access levels, other functionalities, etc)** | This will require some research such as consulting software manuals or online forums. It might also require support from someone with experience in configuring the selected platform. |
| **6. Clean and import data** | This might require support from someone with experience in cleaning and importing data. Gathering and migrating all your existing data might take time. |
| **7. Test the platform and data** | This is a collaborative process that will involve your primary audience, and would require a facilitator. |
| **8. Make improvements** | This might require support from someone with experience in configuring the selected platform. |
| **9. Train team on database and methodology** | This is a collaborative process that will involve your primary audience, and would require a facilitator. |

Planning resources and producing a project timeline to build a database is usually followed by the *data modelling* phase, encompassing a few exercises, helping create a database architecture.

# What is a data model?

---

Human rights work relies heavily on data such as police records, witness testimonies, videos of incidents, media reports, etc. This data needs to be translated into information that can be understood by those who have the power to make change – citizens, journalists, policy makers, lawyers or judges. This is especially important for human rights defenders because they often need to show that violations are systematic. To make data more understandable, it is often helpful to have some kind of database that can help you organise and analyse the information that you have. In order to create a database, you need a plan for how to structure your data within this system – this plan is called a **data model**:

- A data model establishes how data/information is structured/organised for the purpose of designing an information system.
- Your data model will lay the groundwork for how your data/information is analysed.
- You need a data model in order to design a database that will allow you to carry out the required analysis.

Video [to embed into the pdf]
Example:

# How do you build a data model?

## Step 1: Group your information

How would you group the information you have collected? What are the main categories of information? (e.g. incident, victim, court, law, perpetrator)

It is important not to get too hung up on finding the perfect words to use to categorise your information at this stage. Here is an example from a data model meeting with partners:

> After a few hours of back and forth of deliberation and discussion about what are called the 'types of information in the database', we found it helpful to simply agree that we are talking about three main buckets of information, referred to as 'templates': Thing 1 [template 1], Thing 2 [template 2], Thing 3 [template 3]. The group knew that these were important buckets because the information within them shared the same properties.

For example, you might already know that one of your main buckets of information is an **incident**, because each example of an incident has a date, location, people involved, description, etc. Furthermore, you may have many research questions related to incidents, so it makes sense for it to be a main bucket of information in your data model.

See some examples:

## [TEMPLATE 1] Person

What kinds of information do you want to capture for each **[Person]**, and why?

| Type of information (properties) | What is the format? | Why is this important to collect? |
|---|---|---|
| **Date of birth** | Date field | Capturing dates of birth will let you examine age patterns within your datasets. |
| **Gender (or Sex)** | Select field | Documenting gender (or sex) of individuals can help you understand if gender (or sex) plays a role in a documented human rights situation. |
| **Occupation** | Select field | Compiling this information can be relevant in cases when there is a reason to suspect that a crime or a human rights violation was committed by a perpetrator due to the victim's professional affiliation (i.e. murders of journalists or reprisals against human rights defenders and activists). |
| **Ethnicity** | Select field | Understanding what ethnic identities are claimed by individuals can be important, especially when investigating cases of discrimination that are potentially based on ethnic or racial grounds. |
| **Related individuals** | Relationship | Capturing how individuals (victims, witnesses and sources of information) are interrelated will help one see patterns and reveal connections that did not seem obvious before. |

## [TEMPLATE 2] Perpetrator

What kinds of information do you want to capture for each **[Perpetrator]** and why?

| Type of information (properties) | What is the format? | Why is this important to collect? |
|---|---|---|
| **Name and Surname** | Text field | Name and surname pertain to basic information about a person, which, nonetheless, is crucial when using information gathered from public sources. Many human rights defenders use open source intelligence (OSINT) methods to find perpetrators' profiles on social media platforms or to search for information about them online. |
| **Age (or Date of Birth)** | Numeric field (or Date field) | In the same way, learning and registering perpetrators' age will give you an opportunity to use OSINT techniques in order to identify a perpetrator |
| **Affiliation** | Relationship (to Institutions) | It is important to capture any potential relations between perpetrators and law enforcement authorities. It can help you prove that a certain number of perpetrators are abusing the power delegated to them by authorities. |
| **Position and rank** | Select field | As a number of human rights violations are committed by state employees, it is important to keep track of changes in the ranks. |
| **Distinctive marks** | Rich text field | Registration of distinctive marks or features may help investigators with the identification of perpetrators in situations when the identity is unknown. |

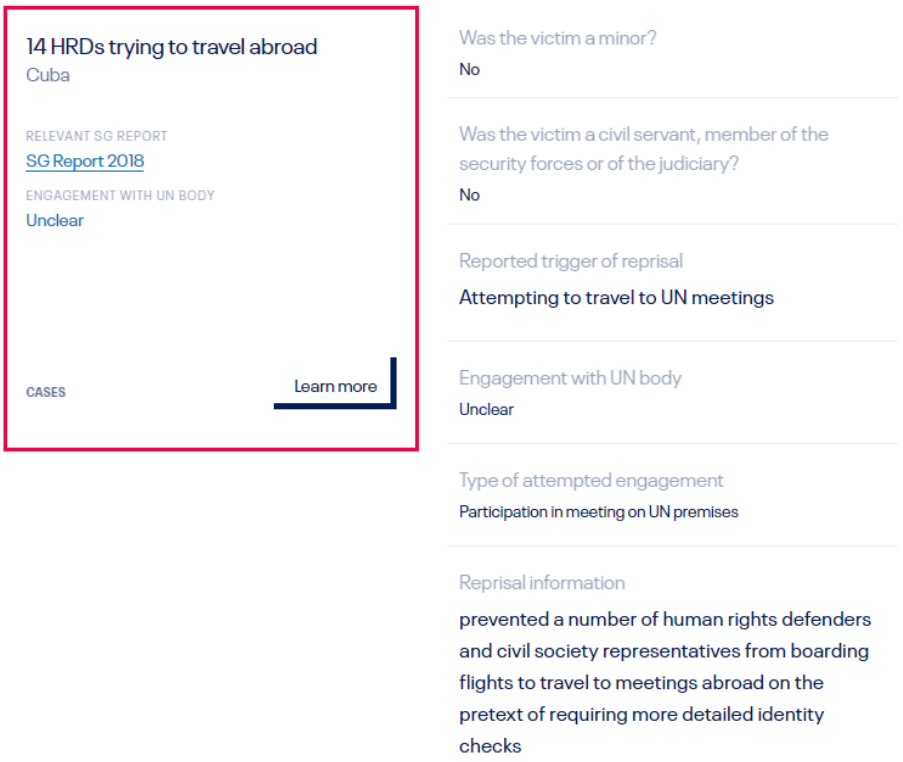
## [TEMPLATE 3] Case

What kinds of information do you want to capture for each **[Case]**, and why?

| Type of information (properties) | What is the format? | Why is this important to collect? |
|---|---|---|
| **Description of case** | Rich text field | The description of case field will allow for providing a context to a documented event in a narrative form and prove helpful if something cannot be conceptualised through standard thesauri or quantitative indicators |
| **Issues** | Rich text field | This field can be used for explaining main issues raised in the case and highlighting why the case is significant. |
| **Date** | Date field | Capturing dates throughout your documentation efforts will help you use temporal concepts more efficiently that can aid in building a stronger case and timeline of events. |

A **case entity** can look the following way in an Uwazi database[3]:

[3] These entities are part of ISHR's End Reprisals library, accessible at: https://endreprisals.ishr.ch/.
Accessed October 2023.




## Step 2: Identify relationships between templates

How are **templates**[4] and their properties of information related to each other? (For instance, a victim is involved in an incident, a court receives a case, a court makes a decision, a judge relies on a law.)

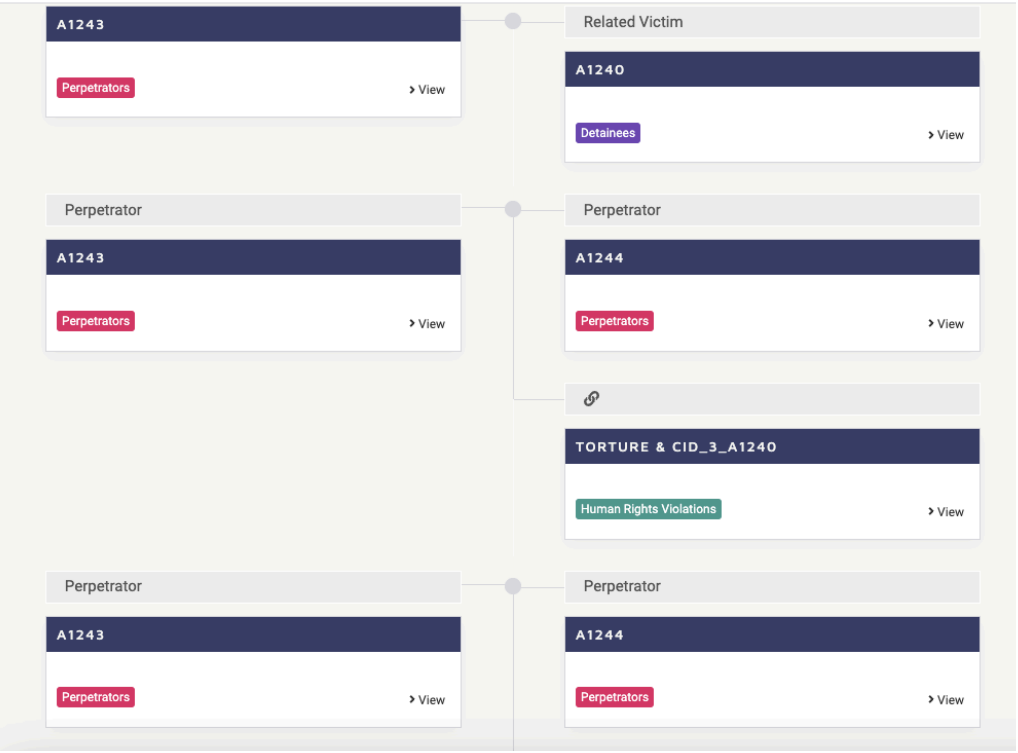
In the image below, you can see how **relationships** between various records can be illustrated in the dedicated panel[5] in Uwazi. There are 3 different relationships that you can view: relationships between a perpetrator and victim, between two (or more) perpetrators, and between perpetrators involved in a human rights violation.

[4] **Templates** are a starting point for new **entities** in some databases. They can be built and customised with different properties for each type of **entity**.
[5] These entities are part of the North Korean Prisoners Database, accessible at: https://nkpd.io/. Accessed October 2023.

Relationships help to establish how each entity interacts with each other.
In some cases, Relationships make it possible to investigate and analyse “who did what to whom” (and when, and where, etc).

One of the ways to show entities' interaction follows this formula: noun-verb-noun. For instance: Perpetrator (noun) committed (verb) an alleged crime against a victim (noun). Moreover, it could be a case related to the location entity or a law that is violated: CASE No 3 contains a violation of the Article 2 of a CERTAIN DOCUMENT.

Other examples: as a documentalist, I want to create connections/relationships between different reports so that we can see areas of intersection and overlap.
In March 2021, VICTIM A presented a testimony before JUDGE J in relation to EVENT X .

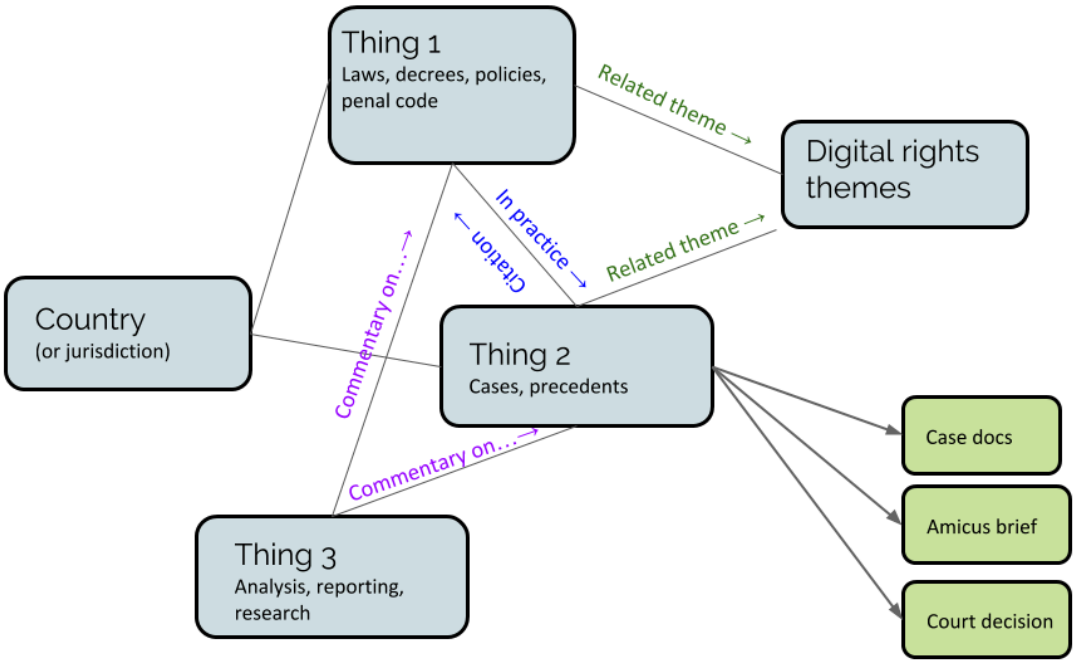


## Step 3: Identify properties for each template

**Properties** (or attributes) refer to aspects of the entities and/or relationships.[6] What information do you need to collect about each category in order to answer the questions you prioritised in Step 2? (For instance,  names, locations, dates, categories.) What are the lists of category items your team will use to describe/categorise this information?

---

Once you have grouped your information into templates, identified your properties and established the relationships among the different pieces of information, your data model might look something like this:




[6] Properties or attributes will be displayed as fields on each entity in a database. For example, for a 'Person' entity, attributes could include: 'First name', 'Last name', 'Gender', or 'Ethnicity'.

## Step 4: Create your list of terms also known as 'Thesauri'

A **thesauri,**[7] the plural form of thesaurus, refers to a list or terminology used for categorising different aspects of human rights violations in a consistent manner.

### [LIST/THESAURUS] Violations of ESC rights (example)

What are the items to be included in this list/thesaurus?

| Items/terms | Definition |
|---|---|
| **Forced eviction** | Forced eviction occurs where a person, group or community is involuntarily removed from their home and/or land they occupy (permanently or temporarily), without appropriate legal and procedural protections[8]. |
| **Discriminatory exclusion from access to adequate food** | This type of discrimination is a distinction, exclusion or restriction which has the intention or effect of nullifying or impairing access to food or the means for its procurement, based on prohibited grounds.<br>It is a violation of international human rights law if a State:<br>• denies individuals or groups access to food or productive resources, based on prohibited grounds of discrimination; or<br>• fails to redress laws, policies or practices which have a discriminatory impact, even if they may appear neutral at face value and regardless of intent; or<br>• fails to take measures to address discrimination by private actors, for example if public authorities fail to take measures to change attitudes in the family |

[7] **Thesauri** (or **thesaurus)** refers to lists of terms that are referenced in **select** and **multiselect** properties.
[8] OHCHR. "Examples of Definitions of Human Rights Violations." Inspira Learning Management System (LMS). Accessed September 2023. https://elearning.un.org/CONT/GEN/CS/UNHR_V3/Module_01/story_content/external_files/Examples%20of%20definitions%20of%20human%20rights%20violations.pdf.

| | |
|---|---|
| | whereby boys and men are given privileged access to scarce food resources at the expense of women and girls[9]. |
| **Lack of access to medical services, facilities and treatment** | It is a violation of international human rights law if States:<br>• Fail to ensure the equitable distribution of medical services and facilities, such as between rural and urban areas, or between wealthier and less wealthy communities; or<br>• Fail to ensure non-discriminatory access to health care, for example for people living with HIV/AIDS or for displaced populations; or<br>• Directly impede access to medical services, facilities or treatment, for example by limiting access to health services as a punitive measure, e.g. during armed conflicts; or<br>• Do not take adequate measures to protect persons against actions of third parties which threaten access to health care, for example by failing to take adequate measures to address cultural attitudes preventing women from accessing pre- and postnatal health care; or<br>• Fail to take whatever steps are necessary, using the maximum of available resources, to achieve progressively the full realisation of the right of access to medical services, facilities or treatment for all, with special attention to under-served areas and populations, and an emphasis on the development of primary health care; or<br>• Adopt deliberate retrogressive measures without compelling justification.[10] |
| **Failure to address domestic violence** | Domestic violence is violence committed in the home or private sphere, generally between persons belonging to the same family or household, and often between intimate partners. It is a violation of international human rights law if the State fails to:<br>• take effective measures to prevent domestic violence, among other things; or |

[9] Ibid.
[10] Ibid.

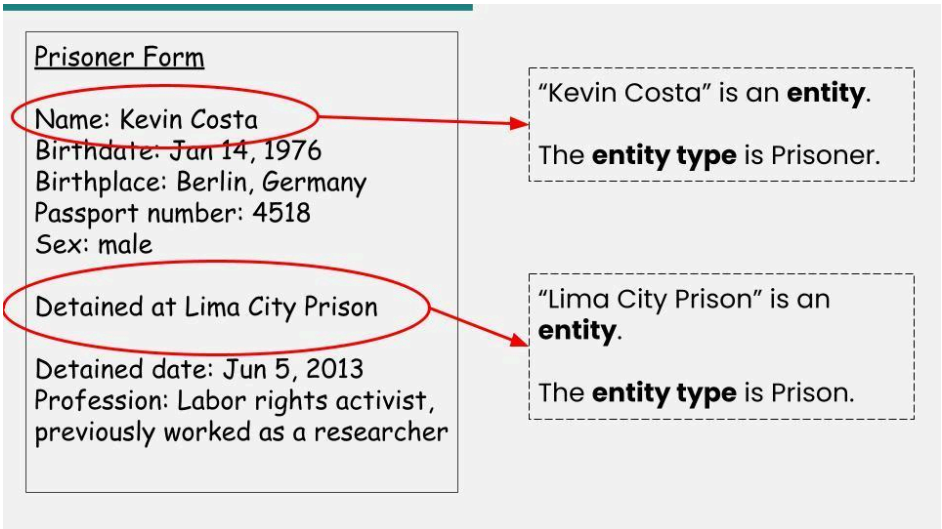
| | |
|---|---|
| | • establish effective systems to protect persons from real and immediate risks to life or integrity, where public officials know, or ought to know, that acts of violence are[11]. |
| **Child labour** | Work that deprives children of their childhood, their potential and their dignity, and that is harmful to physical and mental development. It refers to work that:<br>• is mentally, physically, socially or morally dangerous and harmful to children; and/or<br>• interferes with their schooling by: depriving them of the opportunity to attend school; obliging them to leave school prematurely; or requiring them to attempt to combine school attendance with excessively long and heavy work[12]. |

# Data model examples

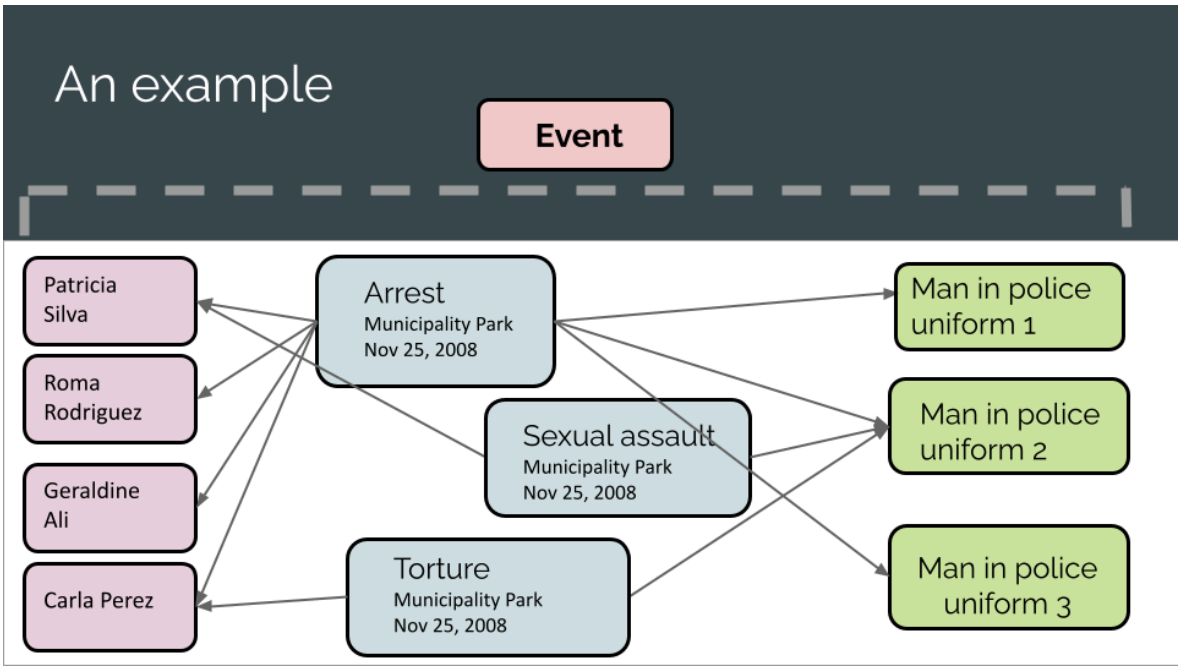
Once drawn, your data model will illustrate how your **records** (otherwise referred to as **entities**), are going to be related between each other, potentially letting you see what else is missing from the constellation.

Take a look at the slide presented below. The **Prisoner form** contains information about Kevin Costa, detained at Lima City Prison. It also suggests that we detect two entity types based on this information, specifically, **Prison** and **Prisoner**, and implies that a future data model would show that **Prison** and **Prisoner** are related.

[11] Ibid.
[12] International Labour Organization. "What is child labour." Accessed 31 August 2023. www.ilo.org/ipec/facts/lang--en/index.html.




Let's take a look at an example of a data model below:.



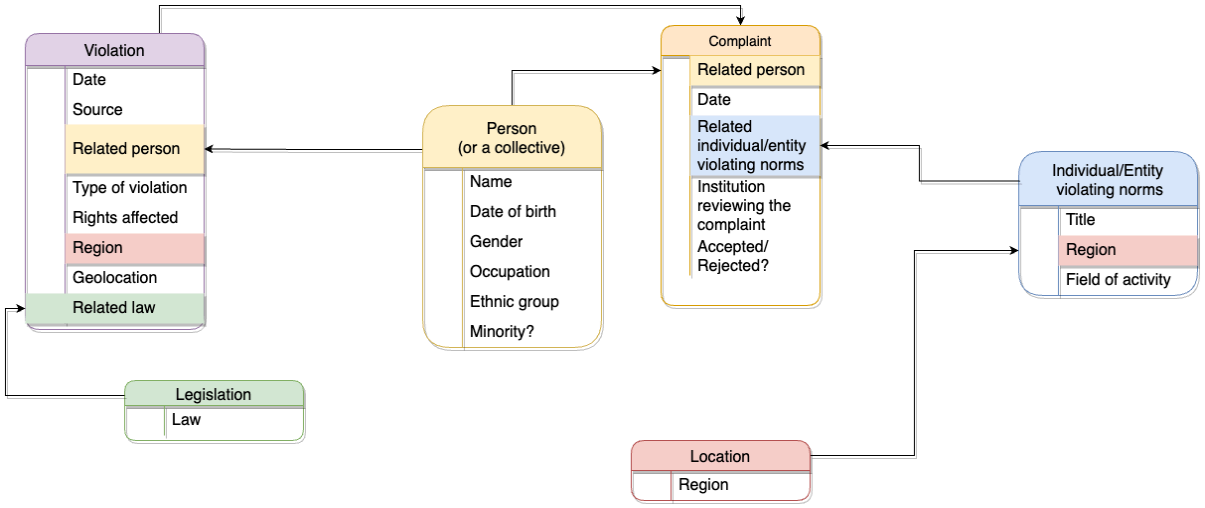

This data model depicts how templates (**Victims, Perpetrators**) are related to each other in a chain of acts, while at the same time illustrating the natural complexity of documentation of human rights violations.

Investigators often reveal intertwined connections between perpetrators, committing crimes against multiple victims as well as victims who are victimised by multiple perpetrators. The documentation of situations in which individuals play multiple roles has even more pitfalls: in some situations, victims give testimony about other crimes that they witnessed which, therefore, makes them primary sources. There is sometimes a pathway from victim to perpetrator and the other way around while crimes against offenders occur too. The roles of individuals are intrinsically fluid, and data models function as the exact mediums that transmit and reflect this natural state of things.

When documenting human rights violations in countries experiencing a breakdown of the justice systems (we can, for instance, think of authoritarianism, in which courts do not act independently and make decisions, empowering authoritarian rule), its institutions can also be viewed as violating people's fundamental rights. When building a data model for documentation in such contexts, you should deliberate carefully how to capture it and create an information architecture that will equip you to perform an accurate analysis.

When it comes to monitoring ESC rights, one can also think of a variety of approaches that can be used. Adopting a victim-centred approach can guide you toward focusing the entire workflow on examining a victim's profile by ensuring a granular corroboration of the evidence associated with the case.

Meanwhile, documentation can have other priorities, say, encompassing an entirety of facts about an incident with numerous related elements. In this case, a sample data model for monitoring ESC rights can look the following way.




Bear in mind that any data model can be further refined and adjusted to better suit your documentation purposes.

## Final considerations

Now that you understand how to create a data model, define goals, and research questions for your documentation project, you can move forward by selecting a database solution that fits your needs. When comparing tools and applications, consider doing the priority-setting exercise once again: what do you deem crucial in this selection?

- Finding the most user-friendly platform?
- Choosing the one offering the largest number of possibilities?
- Using a tool developed based on open source principles?
- Selecting a tool that meets the highest data security standards?
- Lastly, a very important question to ask is whether or not a tool will allow you to incorporate the data model you built.

Acknowledging that all these elements and components are crucial and should not be assessed as one against another, try to prioritise them depending on your goals and bring your considerations to your team.

If your team needs assistance with a database configuration, give thought to contacting human rights organisations such as HURIDOCS (Human Rights Information and Documentation Systems). Alongside developing an open source secure database application *Uwazi*, we also offer broad methodological and technical support services to organisations and human rights defenders to work together toward configuring powerful databases, tailored to their unique documentation needs.

# Promoting ESC Rights of Disadvantaged Groups in Turkiye through Monitoring and Advocacy

**MODULE 2. DATABASES**

**Theory**

Conceptualising human rights information and building a data model

Funded by the European Union

AMER Association for Monitoring Equal Rights

huridocs human rights information and documentation systems