Practical guide: tips for producing reliable data in the human rights field

Introduction

How do we know if certain behaviors or facts should be labeled under specific categories in our database? How do we decide whether labelling a violation as a physical aggression or as torture? International standards and laws are useful instruments but sometimes they are not enough, as reality is complex and ambiguous and makes us to doubt.

One of the main sources of inconsistencies in databases is the lack of homogeneous criteria for recording data. In this document you will find advice on how to mitigate these inconsistencies and give your database more coherence, uniformity and reliability.

Communicate in a clear and explicit manner

If you want your data to be reliable and trustworthy, you need to explain how you collected and recorded them in your database. It is not about revealing or exposing your sources (victims, witnesses, etc.), but about detailing how the study was made and how did you build the indicators. This process, as we will see later on this document, is complex and determines the results we will get. If we want to be reliable we need to reference the standards or laws that we are using and explain where do our specific indicators come from.

Avoid ambiguity

Polysemic, synonymic and ambiguous terms create inconsistencies when working with data. There might be words or sentences with various meanings (polysemy) or, by contrary, several words or phrases to name the same thing (synonymy).

For example, we often say indistinctly "right to asylum" and "right to refuge" to talk about the protection and assistance that the state must guarantee for everyone forced to flee his or her country because of persecution, war or violence. We use two different names to refer to the same thing or very similar things, as in some contexts there might be differences between them. If we store data related to this right in our database but we do not select the preferred term, it is very likely that inconsistencies will appear. Similarly, if we use polysemic words we should make clear which meaning we are referring to.

Also, there are some terms frequently used in our daily life or in media that do not have an empirical correlation. Selecting the most adequate terminology in a database is a complex process. For instance, the use of the term “political prisoner” in the media might have different meanings that do not always match the specific way in which cases are registered in the database.

Choosing the appropriate terminology for a database is a complex process. In the HR field, the glossaries created by specialised organisations constitute a good reference that we can take as a starting point.

The HURIDOCS thesauri list can help you a lot in this process. It includes several lists of standardized terms which serve to categorize phenomena in the HR field.

Make sure your concepts are measurable

In social sciences, the term operationalization refers to the process involved in measuring a phenomenon that is not directly accessible by defining observable variables that we can measure.

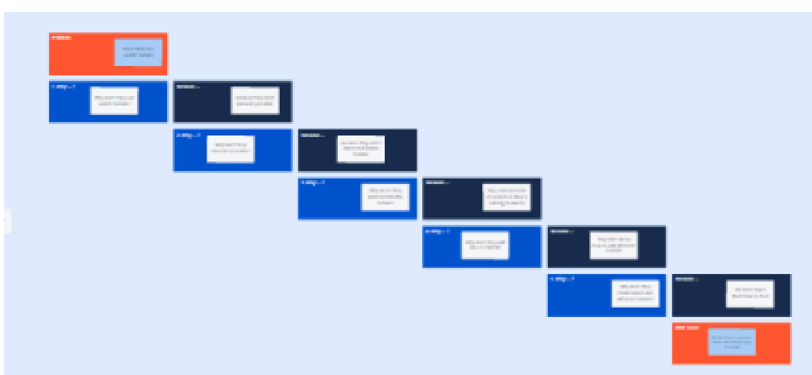
The operationalization entails translating abstract concepts into observable and measurable indicators and variables. Every time we label a given fact as a certain type of human rights violation, we are actually going through an operationalization exercise. For instance, if you see severe marks of a physical aggression committed by a state agent, you can interpret that as sign of mistreatment or torture, two different categories that might refer to distinct observable phenomena depending on the context.

To successfully achieve the operationalization process you need to define the concepts you will use, the variables of analysis, and be clear, systemic and consistent on how exactly you will measure what you are studying. This entails congruence between the conceptual definition, the operational definition and the observational indicators. This is a key component on your data system as it will determine the analysis, so It's extremely important to consider how you will operationalize the variables that you want to measure before collecting data.

Example:

Concept	Variable	Indicator	Categories
Right to water	Amount of water	Litres per person per day	Sufficient: >50 litres Insufficient: <50 litres
	Accessibility	Distance to the water source Collection time	Accessible: <1km Inaccessible: >1km Accessible: <30 min Inaccessible: >30 min
	Water quality	Physical and chemical characteristics	...
	Affordability	Price related to the household income	Affordable: <3% of the household income Unaffordable: >3% of the household income

You need to take baby steps to go through from abstract ideas to empirical facts, from vague to concrete terms, from complex to simple. To do so, it is always useful to graphically represent the stages of this process. This [Miró template] (or any other visualization option) will help you verify from top to bottom and vice versa that concepts and indicators are properly linked.



The definitions and indicators must be coherent, so variables and categories as well. Although it looks as a clear and rational process, we often see categories that show misunderstanding of definitions.

Work examples from the HR field

Example 1: measuring economic indicators.

¿How do we determine what are low, middle and high incomes?

Low and middle income: income > $3,20 per day

This amount marks the poverty threshold.

High income: income > 50 % median income

The income amount that divides a population into two equal groups: the ones who are above and below the amount. The median is a statistical concept commonly used together with the mean and the mode, although each one is calculated differently.

Source: Human Rights Measurement Initiative. Methodology Handbook (2019). https://humanrightsmeasurement.org/wp-content/uploads/2019/06/HRMI-Methodology-Guide-2019-version-2019.06.06.pdf

Example 2: indicators for arbitrary detentions

We can talk about arbitrary detention if one of these circumstances occur:

1) *Detention without warrant*
2) *Detention period and conditions are clearly illegal and unproportioned*
3) *The detention was made by armed civilians not identified as police officers*
4) *The detention was made according to the testimony of undercover agents.*

Source: "Ruta de salida legal Para la liberación definitiva de personas presas políticas", de Abogadas y abogados defensores de personas presas políticas.

Example 3: indicators to measure the respect of children's rights in the justice system. Quantitative indicators developed by UNICEF to determine if the justice system respects children's rights:

1. *Number of children arrested during a 12 month period per 100,000 child population*
2. *Number of children in detention per 100,000 child population*
3. *Number of children in pre-sentence detention per 100,000 child population*
4. *Time spent in detention by children before sentencing*
5. *Time spent in detention by children after sentencing*
6. *Number of child deaths in detention during a 12 month period, per 1,000 children detained*
7. *Percentage of children in detention not wholly separated from adult*
8. *Percentage of children in detention who have been visited by, or visited, parents, guardian or an adult family member in the last 3 months*
9. *Percentage of children sentenced receiving a custodial sentence*
10. *Percentage of children diverted or sentenced who enter a pre-sentence diversion scheme*
11. *Percentage of children released from detention receiving aftercare*

The results of these measurements are translated into different scenarios that allow us to conclude to what extent does the judicial system respect the children's rights.

Source: Manual for the Measurement of juvenile justice indicators:
https://cdn.penalreform.org/wp-content/uploads/2013/05/JJI_Indicators_man_en_2006.pdf

Naming variables in a precise and objective manner, with references to observable phenomena, as we have shown in the examples, is crucial. When it comes to name the variables, you should reflect on how you are going to measure: by degree, intensity, quantity, etc. The examples mentioned above use formulas such as “percentage of” or “number of”, which are independent from subjective criteria.

Now that you have seen concrete examples, what you think about this?

To what extent do you think the government respects physical integrity?

a) *A lot*
b) *Enough*
c) *A little*
d) *Nothing*

The responses of the example will be based on the respondent' perception. The term "a lot" does not have the same meaning for everyone.

We need to formulate questions in such a way that answers can have a concrete reference, avoiding subjective answers, except if what we want to measure are, precisely, perceptions.

Concordance between variables and categories

The term variable refers to the characteristics or attributes of the unit of analysis. Although in databases it is common to define items for search based on a list whose components might not have relation or coherence between them, in social science methodology there are ordering criteria which might be of help in case we want use those items to make reliable and valid analysis.

In social science methodology and statistics, a variable is defined as the characteristics or attributes of a given unit of analysis. These attributes are classified in categories (or thesauri). The categories describing a variable must be exhaustive and exclusive.

1. Exhaustive: available categories must reflect the whole range of possible options. In case we don't know that range in advance, we can include the option "other".
2. Mutually exclusive: the contents or possible answers inside every category must be clear and not overlapping between them. This way we will gain clarity and we will avoid confusion when it comes to assign categories.

Although assigning categories looks like an easy task, in practice we can see many inconsistencies. Most of the time the categories we assign show how definitions were not understood properly.

Real example: we see this categorization within the "gender" variable:

a) Male, b) Female, c) LGTB

In the example, two variables are mixed: gender and sexual orientation. Thinking that gender identity and sexual orientation are connected It's a common misconception, but it is false.

Even if we don't know that much about these concepts, we can discover that something is wrong when we see that the existing categories are not mutually exclusive. A man or a woman can consider themselves LGTB or heterosexual, among other orientations. The *error comes from assuming that women and men are heterosexual, which is false. This miscategorization is the result of our heteronormative bias.*

To avoid these issues is important to pay attention when we define the variables or the terms we will use for data collection and registration.

Tips to use HR standards for variables definition

★ Be careful on using international human rights standards as an unique or exclusive source of indicators definitions for data registration purposes. Using universal standards as a source is useful and necessary but not enough.

ACNUDH, said: “The enunciation of human rights regulations (...) can be very general; furthermore, many human rights seem to overlap. This means that the provisions of the rights treaties are not particularly useful for defining appropriate indicators”. (See the full statement in the footnote)[1].

★ Select definitions consciously and strategically. Remember that the definitions you choose will affect the type of data you include and exclude in your analysis. The definitions have ulterior consequences and reflect our position regarding the topic to be studied.

★ Even in the human rights field, the conceptual definitions we establish for research or data collection purposes could be based on broader perspectives than the judicial or legal standards.

---

[1] “The enumeration of human right standards in treaties and their further elaboration by the treaty monitoring bodies and other human rights mechanisms and instruments may remain quite general and many human rights appear to overlap. So human rights treaty provisions are not particularly helpful in the identification of appropriate indicator(s). As a starting point, it is therefore important that the narrative on the legal standard of a human right is transcribed into a limited number of characteristics or attributes of that right. By identifying the attributes of a right, the process of selecting and developing suitable indicators or clusters of indicators is facilitated as one arrives at a categorization that is clear, concrete and, perhaps, more “tangible”. Indeed, the notion of attributes of a right helps in making the content of a right concrete and makes explicit the link between identified indicators of a right on the one hand and the normative standards of that right on the other”.

★ Definitions for data collection and registration are the result of a theoretical framework, and empirical observational items in a given concrete context.

★ We suggest reviewing different sources to verify that the concepts and terminology for data collection are accurate and valid. Test your definitions and assumptions with case law and research studies.

★ In order to enrich and validate our definitions and terminology, we must review the bibliography, talk to specialists from different disciplines, and critically analyze it, considering the potential implications for measurement purposes.

★ Attend the particularities of your context and the diversity of voices.

Culture, gender or age configure our point of view in a differential way. It is essential to accommodate various standpoints to mitigate biases and achieve a comprehensive perspective in the human rights field. For example, when discussing collective rights, how does the indigenous community define property?

When using legal and judiciary sources:

★ Remember that the interpretation of the universal human rights standards and law relies on specific context, so it can differ according to the specificities of a case, vary over location, time or the juries involved.

You can search judicial responses of international bodies in Latin America using SUMMA. This database was created by the Center for Justice and International Law (CEJIL) and grants free access to every active or closed case opened in the last decades by the Commission and the Interamerican Court of Human Rights, the bodies of the Interamerican Human Rights System.

★ Local law could contain systemic bias and does not always reflect a comprehensive human rights perspective.

Sometimes using legal definitions as the main source is the best option. For example, using the penal code as a source to define arbitrary detention. However, in other cases, law could reflect systemic biases, based on racism or gender discrimination. For example, in the past, law and judicial resolutions excluded women from civil and political rights, as the right to vote; or until recently in many judicial cases sexual abuse was not considered a rape crime if the perpetrator was a husband; or for instance, until the seventies in Spain women needed a licence to postulate for jobs or have a property.

These examples reflect how legal sources could sometimes be restrictive from a human rights perspective and we should use them based on critical judgment to define the variables for our database.

Train your team

Your team should be trained to understand all the work processes related to data management. Without standardization and common criteria, your data won't be comparable or useful to measure phenomena, as it will depend on the subjectiveness of the one who collected or recorded it in the system.

Everyone in the organization needs to know that the ability to harness data and, thus, to achieve our goals, depends on the homogeneity of the criteria: the definitions of the things we are going to measure and how we are going to measure them.