

Ester Scheck · Jakob Listabarth · Izabela Gołębiowska ·
Pyry Kettunen · Julia Bristow · Katarzyna Słomska-Przech ·
Merve Keskin · Juliane Cron

17 August 2025

Worksheet Round 1

Context and background research of a dataset

Research the background of the dataset and try to find the answers to the questions listed below. Not all of the questions may be applicable to your dataset. Many answers may not be easily found in the metadata; you may need to click through various files and look into supplementary material or other resources related to the topic.

The questions focus more on the context of the dataset than the information saved in the data itself. It is not necessary to analyze the data.

The worksheet was developed based on [Heather Krause's data biography](#). Adaptions and additional questions are set in *italic*.

Biography for the dataset:

Who

- Who collected the data?

- Who owns (*and/or manages*) the data?

- Who provided the data?

- *Who benefits from the way this data is structured?*

- *Who funded the creation of this dataset?*

How

- How was the data collected? What tools and technology were used?

- (How) has the data been cleaned?

Where

- In which locations was the data collected?

- Where is the data stored?

- Where can the data be accessed and by whom?

- *Can you identify any spatial patterns or characteristics?*

Why

- For what purpose(s) was the data collected?

- *What applications has the dataset been used for, or what applications is it intended for?*

When

- When was the data collected?

- *When was the dataset published?*

- *When was the dataset last updated?*

What

- What types of data were collected?

- What do the individual attributes and their values mean?

- *Which categories and terms are used - and are they contested or could they be understood in different ways?*

- *Who or what might be missing from the dataset?*

- *What assumptions does this dataset imply?*

- *How would the dataset be different if it had been authored by someone else?*

- *What biases can you identify in the dataset?*

Your main findings about the dataset: