Describe your data collection choices

Every data collection campaign, even the most comprehensive CASTEMO annotation, necessarily **makes choices**, and is **selective**.

In collaboration between several users, and also as time goes by, it becomes increasingly tricky to remember what data collection guidelines you used, what you included, what you skipped... while this is, obviously crucial for the **interpretation of results** based on the data collected. Therefore:

> You should always describe your data collection choices.

Furthermore:

> It is generally better to keep the description of data collection choices very close to the data themselves.

In CASTEMO knowledge graphs, we recommend to append the description of the collected data as part of the **metadata of the Territory** which holds this CASTEMO data. There is a **pre-defined section Protocol** under each Territory which allows you to capture the basic ones. Beyond that, you might want to describe your choices in a more narrative and comprehensive way in a **document** (webpage, Google Doc), which can be linked from the Protocol.

Revision #2
Created 27 April 2023 15:17:05 by Prof. Dr. David Zbíral, Ph.D.
Updated 23 June 2025 10:22:10 by Prof. Dr. David Zbíral, Ph.D.