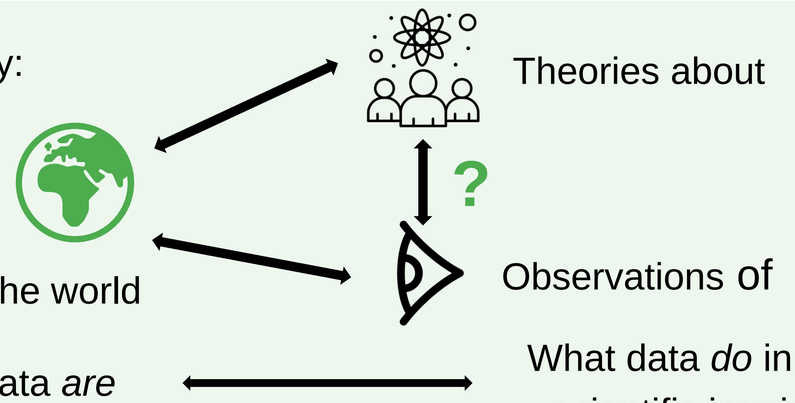
NOT EVERYTHING IS DATA!
A revision of the relational framework of data

Aline Potiron, Institute of Philosophy and Scientific Method, Johannes Kepler University, Linz, Austria.

JOHANNES KEPLER UNIVERSITY LINZ

DATA IN PHILOSOPHY

Dichotomy:



What data *are* ⟷ What data *do* in the scientific inquiry

Relational framework:

- Data are any products of the scientific inquiry that have the dual potential **to be used as evidence to sustain knowledge claims** ① and **to travel between different situations of scientific inquiry** ② (Leonelli 2016, 78).
- It accounts particularly well for the contemporary uses of data such as data-centric and big data practices.

MAIN FINDINGS

Use case: Amplicon Sequencing (AS)

- AS is widely used in microbiology and microbial ecology to "reveal" the underlying characteristics of the microbial community from any environment.
- Principle: DNA sequencing and analysis of a single gene. AS generates millions of DNA sequences.
- This data-centered method (Figure 1) qualifies, in principle, for the relational framework of data.
- Matching the seven steps of AS and their products with the potentials ① and ② of the relational framework (Table 1).

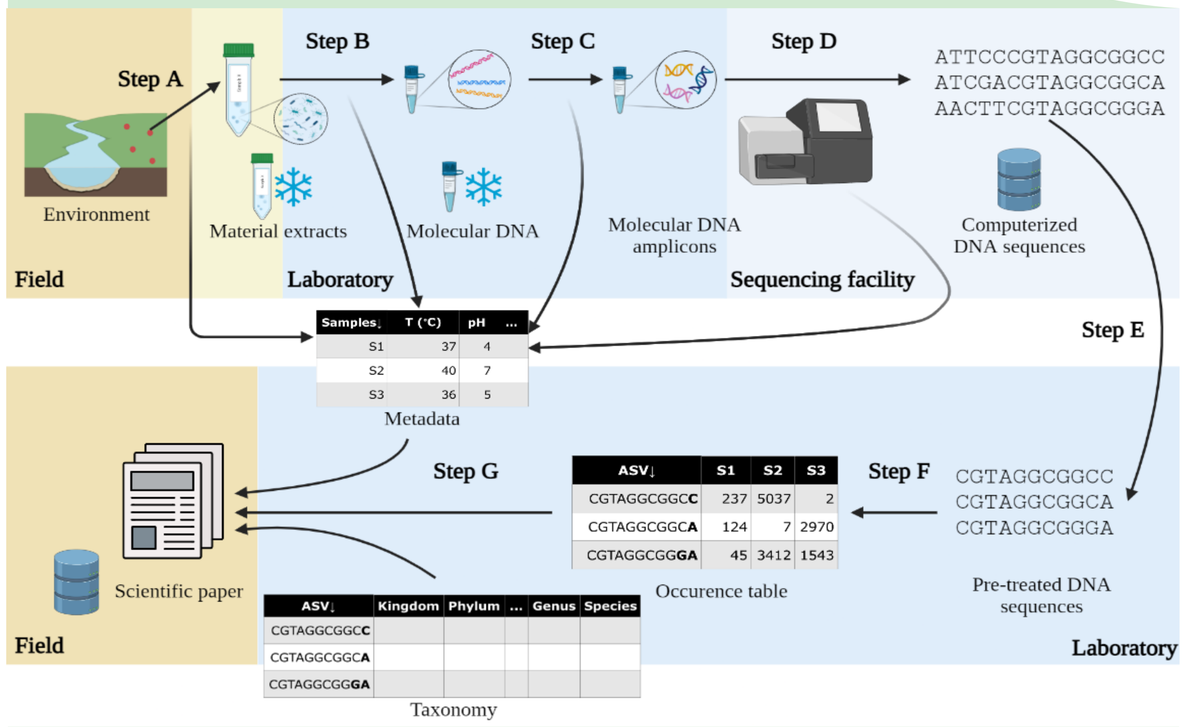


Figure 1: Schematic representation of the seven steps of AS method.

Table 1: Classification of the seven steps of AS method along ① and ②.

Empirical results	**Step.** Products of the inquiry	Potential ①	Potential ②
Samples	**A.** Material extracts	Y	Y
	B. Molecular DNA	Y	Y
	C. Molecular DNA amplicons	Y	Y
Primary data	**D.** Computerized DNA sequences	Y	Y
Secondary data	**E.** Pre-treated DNA sequences	Y	Y
	F. Occurrence tables	Y	Y
Knowledge claims	**G.** Scientific papers	N*	Y

*I take them to contain the evidence either explicitly or implicitly.

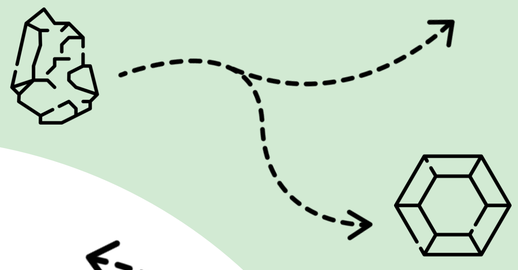
Problems

- Scientists do distinguish among these products i.e. *prima facie* problem for practitioners of practice-oriented philosophy of science.
- One may ask philosophically what make data special in scientific inquiry if all products of that inquiry are data.

The relational framework of data revised

1st criterion: the realization of the potentials ① and ②
2nd criterion: the distinction between travelling and being reused.

- **Travelling**: Literally changing position between at least two things (possibly several) such as location, time but also discipline, and, more globally, situation.



- **Reused**: Transformation – physical change, production of a new product – of one product in at least one other, different situation of scientific inquiry that the one where it comes from

Samples

Any products of the scientific inquiry that do not travel between different situations of scientific inquiry and that are used to construct the evidence for knowledge claims

Primary data

Any products of scientific inquiry that travel between different situations of scientific inquiry and are reused in different situations of scientific inquiry.

Secondary data

Any products of the scientific inquiry that travel and serve as evidence for knowledge claims They are not reused in different situations of scientific inquiry.

CONCLUSIONS & PROSPECTS

Conclusions

- The framework is closer to the actual scientific practice.
- It makes salient two distinctions underestimated in philosophy: samples vs. data and among two types of data.

Prospects

- Define models along these lines to test the robustness of these distinctions.
- Study other use cases in biology to test the generalization of these enhancements.

REFERENCE & ACKNOWLEDGMENTS

Reference

Leonelli, S. (2016). Data-centric biology: a philosophical study. The University of Chicago Press
https://doi.org/10.1080/14636778.2017.1389263

Acknowledgments

Julian Reiss and Emanuele Ratti for their careful proof readings of different versions of this work. An earlier version of this work was presented at the ISHPSSB in 2021 where it received many useful comments. The figure was created with Biorender ®