

What are data?

Current debates on big data and data-centric research have reignited social and cultural interest in what counts as data, and under which conditions data are transformed into knowledge. I propose a philosophical perspective on these issues that builds on long-standing discussions of inferential processes within philosophy, and on the history of scientific data production and use. I consider data as tools for communication, whose main function is to enable intellectual and material exchanges across individuals, collectives, cultures, governments, and – in the case of biology – species, and whose mobility across these groups is a hard-won achievement.

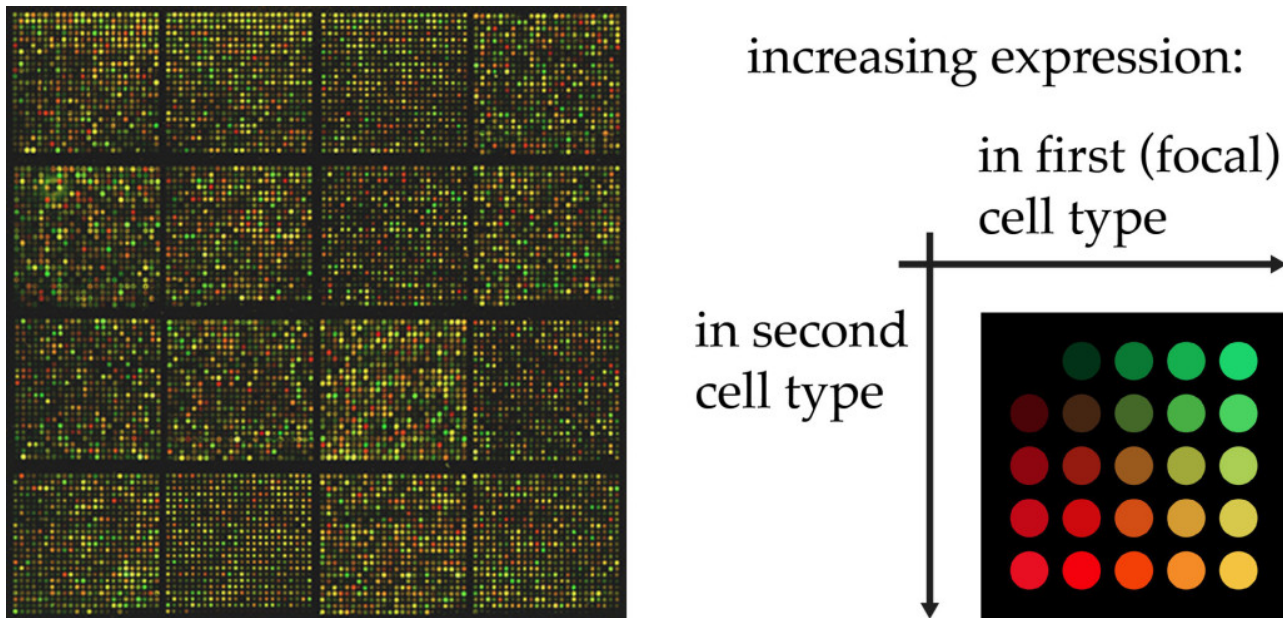


Fig. 1. Biological data on gene expressions, which scientists call 'microarray data'.

When conceptualizing data as *tools* that people use to understand the world and communicate with each other, one stresses their role as sources of knowledge, rather than knowledge in themselves. An object becomes a datum when it is treated as potential or actual evidence for one or more prospective knowledge claims. For instance, the observation “this tree has 46 brown leaves” can be used as evidence (or not) depending on the circumstances under which it is made as well as the circumstances under which it is interpreted. If a small child utters this sentence as a result of counting leaves in his garden, and he and nobody else uses it for any other purpose, the observation does not function as a datum, since it is not being used as evidence for any claim. But suppose that the observation is heard by a climate scientist interested in the rate of leaf shedding in deciduous trees: such a person may use it as evidence for claims about the relation between shedding and climate change in a specific region. A plant scientists interested in one of the many pathogenic infections currently threatening forests in the Northern Hemisphere, which typically

make the leaves of affected trees turn brown, may instead interpret the observation as evidence that the tree is infected by a pathogenic fungus. One can imagine many more situations under which this same observation may function as evidence for other claims, depending on who is evaluating it, with which background knowledge and for which purposes. It follows that the evidential value of a datum such as this can vary. Furthermore, I argue that the scope and variety of claims which data may be used to corroborate is impossible to fully predict, since it is so steeped in specific and changeable factors. The extent to which the evidential value of a given dataset can vary is impossible to determine a priori.

In this view, 'data' is a relational category applied to research outputs that are taken, at specific moments of inquiry, to provide evidence for knowledge claims of interest to the researchers involved. Data consist of a specific way of expressing and presenting information, which is produced and/or incorporated in research practices so as to be available as a source of evidence, and whose scientific significance depend on the context in which it is used. They do not have truth-value in and of themselves, nor are they straightforward representations of given phenomena. Rather, data are fungible objects, which are defined by their *portability* across different contexts of use and their *prospective usefulness as evidence*. The same set of objects may or may not be functioning as data, depending on whether it can be passed around a community of researchers and on which role it is made to play in scientific inquiry; and the ways in which data are organised and visualized for dissemination has a significant impact on the ways in which data are re-used across contexts.

Sabina Leonelli

*Exeter Centre for the Study of the Life Sciences (Egenis) & Department of Sociology,
Philosophy and Anthropology, University of Exeter, Exeter, UK*

Publication

[What Counts as Scientific Data? A Relational Framework.](#)

Leonelli S.

Philos Sci. 2015 Dec 1