



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Data-centric biology

**Citation for published version:**

Calvert, J 2017, 'Data-centric biology: a philosophical study', *New Genetics and Society*, vol. 37, no. 3, pp. 275-277. <https://doi.org/10.1080/14636778.2017.1389263>

**Digital Object Identifier (DOI):**

[10.1080/14636778.2017.1389263](https://doi.org/10.1080/14636778.2017.1389263)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

New Genetics and Society

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



**Review of Sabina Leonelli (2016) *Data-Centric Biology: A Philosophical Study* Chicago: Chicago University Press**

Jane Calvert  
University of Edinburgh

With the rise of 'big data' we are seeing an increase in data-centrism across many fields of human endeavour, including the life sciences. How should we make sense of these developments? And how should we understand them philosophically? Sabina Leonelli's book *Data-Centric Biology* is the first to grapple seriously with these questions. It challenges us to think in new ways not only about data and biology, but also about philosophy.

Leonelli defines data-centrism as "a particular model of attention within research, within which concerns around data handling take precedence over theoretical questions" (p.178). This requires an answer to the question 'what counts as data?' and Leonelli gives a two-fold one: data are objects that "(1) are treated as potential evidence for one or more claims about phenomena, and (2) are formatted and handled in ways that enable their circulation among individuals or groups for the purpose of analysis" (p.78). It is the second part of this definition that I find particularly striking: part of the definition of data is precisely its portability. In fact, we are told that rather than having "intrinsic representational powers", data "acquire evidential value through mobilisation" (p.198).

Since portability is so important to Leonelli's understanding of data the book focuses much of its attention on making data travel – what she calls 'data journeys'. She shows the importance of both the de-contextualisation of data (so that it is able to circulate) and its re-contextualisation (to make it usable by other researchers). This mobility requires considerable investment of resources and effort particularly from database curators. Leonelli shows how the work of these curators is often not adequately valued or sufficiently funded, despite its centrality to the production of biological knowledge.

Since she argues that in data-centric biology "concerns around data handling take precedence over theoretical questions" (p.178), we might be tempted to draw the conclusion that data-centric research is atheoretical. But this is not Leonelli's position. She maintains that bio-ontologies – "the labels used by curators to classify data for the purposes of dissemination" (p.114) – can be thought of as classificatory theories. Her analysis of bio-ontologies leads her to make the broader point that we should understand theory not only as that which emerges from the attempt to explain phenomena, but also as something that can result from classificatory activities. This belies arguments that were seeing the end of theory in a data-driven world. Instead what she asks us to do is think of theory differently.

This leads me back to my point about how Leonelli challenges philosophy. From her close study of the distinctive nature of data-driven research, she concludes that a philosophy of this topic must give up on generalised abstractions, and instead pay attention to contingencies and specificities. One of the key questions she addresses is whether data-centric biology "is a unique mode of scientific reasoning and practice" (p.193), and she concludes in the negative. She does not think we need a new epistemology of science, precisely because data's "situated nature makes them impossible to associate with one specific form of reasoning or way of carrying out research" (p.198). Instead, it is necessary to pay attention to "the conditions under which data are packaged, disseminated and analysed" (p.181). This emphasis on the situatedness of knowledge resonates strongly with science and technology studies, so by making it central to her analysis Leonelli brings philosophy of science and STS closer together.

And this is something she intends to do. She sees her work as fitting under the heading of “empirical philosophy of science” (p.6), which pays close attention to scientific practices, instruments, institutional settings and social dynamics. Her methods include archival work, interviews and on-line engagements. She has also played an active role in interdisciplinary discussion of, and policy for, data-centric research over the last decade.

Consistent with this multi-pronged approach, she concludes that data-centric research is “conceptually, materially and socially grounded, and inevitably intertwined with broad political, economic and cultural trends and attitudes” (p.198). This sensitivity to social and political factors gives her insights into some of the dangers of data-centric research. For example, she shows how it can inadvertently reinforce current power relations in science, because of the English language dominance of many databases, and because elite established institutions are those that are most likely to be able to fund data production and dissemination.

This is only a brief summary of some of the points that struck me as particularly interesting, but Leonelli’s book has a great deal more to offer those with sociological, historical, philosophical and scientific interests in data-driven research. It opens up new lines of inquiry that will be valuable for much future scholarship. There is no better starting point for someone wishing to reflect on this increasingly important approach to biological research.