

The Role of Bio-Ontologies in Data-Driven Research: A Philosophical Perspective

Sabina Leonelli

s.leonelli@exeter.ac.uk

What role does 'theory' play in data dissemination and re-use? This question fuels debate over so-called 'data-driven research': is such an approach possible, how can it be characterised and how does it challenge or complement existing research practices?

This project explores these questions by focusing on the epistemic roles of bio-ontologies. I argue that bio-ontologies (1) express the background knowledge used to pursue biological research and (2) guide such research towards future discoveries. In this ways, they play the role traditionally assigned to theory in scientific discovery.

BACKGROUND Data-Driven Research

What does it mean for research to be based on empirical evidence? This question, one of the oldest within the philosophy of science, is being reformulated and reconsidered within contemporary biological and biomedical science. Technological innovation and shifting ideas about what counts as evidence have transformed current practices of data collection. Data gathering has become increasingly automated and technology-driven, thus acquiring relative independence from hypothesis-testing, theorisation and explanation. Several biologists have argued that the extraction of knowledge from automatically generated data may constitute a new, 'data-driven' approach to scientific research. This project examines the characteristics and significance of data-driven research.

METHODS Philosophy, History and Social Studies of Biology

This research employs empirical methods extracted from history and sociology (e.g. archival research, interviews, participative observation) to ground a philosophical analysis of conceptual issues in contemporary biology. Since 2004, I have interviewed several groups of scientists involved in the development and use of bio-ontologies; attended relevant scientific conferences; and presented results to philosophers, historians, sociologists and biologists.

ANALYSIS The Epistemic Roles of Bio-Ontologies

- (1) REPRESENTATIONAL -- *Bio-ontologies constitute representational maps of the biological knowledge underlying data-driven research.*
Thanks to their precisely defined terms, bio-ontologies explicitly formulate knowledge about entities and processes that is widely assumed across research communities, yet is usually dispersed across publications and groups. They can be constantly updated to reflect research developments: bio-ontologies need not be true or universal, as long as they capture the assumptions and practices underlying data sharing and re-use.
- (2) HEURISTIC -- *Bio-ontologies constitute a network of theoretical hypotheses guiding data-driven research towards future discoveries.*
The adoption of specific terms and definitions has a strong impact on how data are used in new research contexts. They inform their users' understanding of how phenomena are defined beyond their own field. They also define the evidential scope of the datasets classified and distributed through databases.

CONCLUSIONS

Moving from Dissemination to Discovery

The epistemic power of bio-ontologies lies in their ability to facilitate a move from the *dissemination* of data to the *use* of data towards new discoveries.

Bio-Ontologies as Theories

The definitions used to disambiguate bio-ontology terms play the role traditionally assigned to scientific hypotheses. They are descriptions of phenomena relied upon when planning new research; at the same time, they are open to testing and modifications resulting from empirical findings. Bio-ontologies can thus be fruitfully characterised as theories in Claude Bernard's sense: 'intellectual tools necessary to the evolution of science and suitable for the discovery of new facts'.

References

- > Leonelli S (2008) Bio-ontologies as tools for integration in biology. *Biological Theory* 3, 1: 8-11.
- > Leonelli S (2009) Centralising labels to distribute data: The regulatory role of genomic consortia. In Atkinson P, Glasner P and Lock M (eds.) *The Handbook for Genetics and Society: Mapping the New Genomic Era*. London: Routledge.
- > Leonelli S (2009) On the locality of data and claims about phenomena. *Philosophy of Science* 76, 5.
- > Leonelli S (forthcoming 2010) Packaging data for re-use: Databases in model organism biology. In Howlett, P and Morgan, MS (eds) *How Well Do 'Facts' Travel*. Cambridge, MA: CUP.
- > Leonelli, S (forthcoming 2010) Documenting the emergence of bio-ontologies. *History and Philosophy of the Life Sciences*.

Acknowledgments

This research is funded by the UK Economic and Social Research Council. Special thanks to my colleagues at Egenis, Mary Morgan and the Gene Ontology team for helpful discussions.