

Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology

Sabina Leonelli

Bulletin of Science Technology & Society published online 20 August 2013

DOI: 10.1177/0270467613496768

The online version of this article can be found at:

<http://bst.sagepub.com/content/early/2013/10/08/0270467613496768>

Published by:



<http://www.sagepublications.com>

On behalf of:

[National Association for Science, Technology & Society](#)

Additional services and information for *Bulletin of Science, Technology & Society* can be found at:

Open Access: Immediate free access via SAGE Choice

Email Alerts: <http://bst.sagepub.com/cgi/alerts>

Subscriptions: <http://bst.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [OnlineFirst Version of Record](#) - Oct 8, 2013

[OnlineFirst Version of Record](#) - Aug 20, 2013

[What is This?](#)

Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology

Bulletin of Science, Technology & Society
XX(X) 1–6
© 2013 SAGE Publications
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0270467613496768
bsts.sagepub.com


Sabina Leonelli¹

Abstract

The collection and dissemination of data on human and nonhuman organisms has become a central feature of 21st-century biology and has been endorsed by funding agencies in the United States and Europe as crucial to translating biological research into therapeutic and agricultural innovation. Large molecular data sets, often referred to as “big data,” are increasingly incorporated into digital databases, many of which are freely accessible online. These data have come to be seen as resources that play a key role in mediating global market exchange, thus achieving a prominent social and economic status well beyond science itself. At the same time, calls to make all such data publicly and freely available have garnered strength and visibility, most prominently in the form of the Open Data movement. I discuss these developments by considering the conditions under which data journey across the communities and institutions implicated in globalized biology and biomedicine, and what this indicates about how Internet-based communication and the use of online databases affect scientific research and its role within contemporary society.

Keywords

data, open science, biology, biomedicine, scientific knowledge, databases, political economy

Scientific data can be defined as material artifacts that are collected and used as empirical evidence for the plausibility of claims about the nature of reality (“The earth revolves around the sun”) and/or the efficacy of specific interventions (“500 milligrams of paracetamol help relieve headache”). Their role as evidence for scientific claims makes them into public objects, which, at least in principle, can and should be widely scrutinized to assess the validity of the inferences drawn from them. And yet the vast majority of scientific data generated in the second half of the 20th century have only been accessed by small groups of experts; and very few of those data, selected in relation to the inferences made by the scientists who analyzed them, have been made publicly available through publication in scientific journals. This management of data dissemination is tied to a view of scientific knowledge production as an esoteric and technical process, where even trained researchers become so specialized as to be unable to assess data produced by fields other than their own. Within this view, scientists invest time and effort in scrutinizing data produced by colleagues only when they have reason to doubt their interpretation or suspect foul play; and concerns with data production and interpretation, including issues associated to the emergence of “big data” in the biological and biomedical sciences, remain remote from global civil society.

Since the start of the new millennium, the Open Data movement has challenged this technocratic way of conceptualizing practices of data sharing and their political, social, and economic significance. The movement brings together scientists, policy makers, publishers, industry representatives, and members of civil society around the globe who believe that data produced by scientific research should be made publicly accessible online and freely usable to anyone. The Internet provides a platform for scientists to exchange data, materials, and opinions in real time, no matter where they are geographically located. Participants in the Open Data movement embrace this opportunity. They typically advocate that data can and should travel beyond the specific setting in which they are generated, because free and wide data dissemination enhances the chance that people who have not been involved in their production (whether they are scientists or not) will contribute to their interpretation. Accordingly, the production of scientific

¹University of Exeter, Exeter, UK

Corresponding Author:

Sabina Leonelli, Department of Sociology, Philosophy and Anthropology, University of Exeter, Byrne House, St Germans Road, Exeter, EX4 4PJ, UK.

Email: s.leonelli@exeter.ac.uk

knowledge is portrayed as involving the centralized collection and “mining” of data sets gathered by different research communities across the globe. Pooling together results, it is argued, maximizes the chances of spotting significant patterns in the data that are collected, and thus of transforming data into knowledge (OECD, 2009; Royal Society, 2012). This in turn may improve the quality, accessibility, and transparency of research and speed up the rate of scientific discovery. Some scientists even refer to this shift as a new, “data-driven” paradigm of research (Hey, Tansley, & Tolle, 2009).

Whether research is actually being driven by data, rather than theories, hypotheses, models, or policy challenges, remains disputable (Leonelli, 2012a, 2012b). What is clear is that data are increasingly conceptualized as inherently valuable products of scientific research, rather than as components of the research process that have no value in themselves. This involves viewing data as open to several possible interpretations, whose validity and usefulness depend on the questions, interests, and materials characterizing the specific context in which data are adopted (Leonelli, 2009b, 2013). It also involves viewing data as research outputs that can be published and cited without necessarily having been used as evidence for a specific claim (as required within traditional journal publications).

This methodological, conceptual, and institutional shift in the scientific status of data is particularly visible within biology and biomedicine, where the emergence of big data in the form of large molecular data sets (generated through new technologies such as sequencing machines and microarray experiments) has developed hand in hand with an ethos of data sharing. Indeed, the widespread sharing advocated by the Open Data movement has also long been supported by biologists working on key nonhuman organisms for molecular research, such as the fruit fly *Drosophila melanogaster* (Kelt, 2012; Kohler, 1994; Leonelli & Ankeny, 2012). Furthermore, prominent biomedical researchers such as Francis Collins (who used to direct the Human Genome Project and now heads the National Institute of Health) have forcefully argued against the privatization of human genomic data, for instance, by underwriting international agreements to release sequence data as soon as they are produced.¹

Over the past decade, funding agencies such as the National Institutes of Health, National Science Foundation, European Research Council, and Research Councils UK have strongly endorsed this innovative perception of how data should be managed. They are actively promoting Open Data as key to the advancement of basic research and its translation into applications with immediate social impact, such as therapeutic or agricultural innovations.² In practice, this means that they are pressuring their grantees to release data to public databases—a move that in turn affects how scientists set up their research and how they measure and develop their outputs. Many researchers now invest considerable time and resources into donating data to public

repositories; and they regard the consultation of online databases as a first and crucial step toward the development of new lines of inquiry.

An important question to raise vis-à-vis these scientific developments, which also helps decipher their significance to global civil society at large, concerns the reasons why the Open Data movement has acquired such prominence in contemporary scientific and public discourse. Given the enormous achievements of 20th-century science, where data sharing was confined to small sections of the (predominantly Western) scientific community, why are funding bodies insisting on Open Data as crucial to 21st-century research? This is the central question that I wish to address in this article. A standard answer to this question, often endorsed by journalistic accounts of the power and promise of big data, points to Open Data as a crucial way for scientists to exploit the emergence of new technologies, such as genome sequencing and Internet-based social media. It is certainly true that the availability and widespread uptake of new information and communication technologies, as well as the introduction of new methods of data generation, play a crucial role in making it possible to produce and share information on the scale advocated by the Open Data movement. And yet I believe that the emergence and political impact of the Open Data movement are not a mere consequence of technological advances in data production and communication, nor are its implications restricted solely to science.

In what follows, I argue that the scientific concerns underlying the Open Data movement need to be evaluated in relation to at least four other sets of factors. First, Open Data provides a common platform for scientists, scientific institutions, and funders (in both the private and the public sphere) to discuss and tackle the practical difficulties involved in making data travel and be reused. Second, it feeds into concerns with transparency, legitimacy, and return on investment on the part of science policy and funding bodies. Third, it aligns with the challenges posed by the globalization of biomedical research to new parts of the world, and the resulting infrastructural fragmentation and geographical dispersion of research processes. And fourth, it exemplifies the embedding of scientific research in market logics and contexts. To make it at all feasible for data to travel, market structures and political institutions need to assess not only their scientific value but also their value as political, financial, and social objects: The increased mobility of data is unavoidably tied to their commodification. Thus, not only does the widespread dissemination of biomedical data challenge established ways to produce, control and use scientific knowledge, it also plays a key role in mediating global market exchange and international politics, thus achieving a prominent social and economic significance well beyond science itself.

A critical assessment of the significance of the Open Data movement for contemporary society at large needs to take account of all these factors, which foreground the indissoluble

ties of scientific research to global political economy. I shall now briefly discuss each set of factors with reference to recent developments in biological and biomedical research.

Let us start from the importance of *bringing together a variety of stakeholders to tackle the difficulties involved in making data travel*. The fact that data are produced by a variety of groups, for different purposes, in different parts of the world, poses immense scientific, logistical, ethical, and structural challenges to their open dissemination. Whether scientific data are shared, among whom, and to which effect depend on the existence of appropriate regulatory, social, and material infrastructures, such as (a) workable databases, guidelines on data donation, and servers located in safe locations where data storage can be guaranteed in the long term; as well as (b) well-coordinated networks of individuals, scientific groups, companies, and institutions that take responsibility for developing, financing, and enforcing those infrastructures and the related instruments, computers, and software (Bowker, 2001; Edwards, 2010; Hilgartner, 1995; Hine, 2006; Leonelli, 2010a, 2010b; Star & Ruhleder, 1996).

Virtually every area of biological and biomedical investigation is in need of dedicated databases, which are specifically targeted to collecting and visualizing the varieties of data involved in each field. For instance, cancer research requires the integration of genetic, physiological, and clinical data, as well as data coming from nonhuman organisms such as mice. Providing common access to such diverse data sets is a complex requirement involving the coordination of several stakeholders in cancer research, ranging from individual research laboratories to pharmaceutical industries, governmental agencies, hospitals, and public health services. Even well-funded attempts such as the Cancer Bioinformatics Grid have struggled to coordinate such an extensive and varied group.

In response to the technical demands involved in these efforts, companies like IBM and leading universities such as Harvard and MIT are training a whole new cluster of “data scientists” specializing in the construction of data infrastructure. Data scientists need to combine knowledge of specific areas of the life sciences (e.g., genetics or cell biology) with skills in bioinformatics and computer science—a new set of skills that makes the curation of data, hitherto seen as a technician’s job, into a branch of scientific research in its own right (Hilgartner, 1995; Hine, 2006; Lenoir, 1999; Leonelli, 2009a, 2010a, 2012c).

However, this approach requires institutional and financial backing, in order to guarantee that data scientists have the status and power needed to develop and implement their contributions. This is hardly surprising when looking at previous examples of such efforts in the history of science: There are well-documented cases of data infrastructures that were unable, in the longer term, to coordinate the diversity of interests, values, terminologies, and constraints involved in the relevant research areas, as well as to muster enough financial and human resources to support themselves.³

The resources and skills required to achieve such coordination are clearly not only technical but also social (Leonelli, 2009a, 2010a, in press). Thus, many scientific institutions are incorporating mechanisms to recognize and reward the contributions of data scientists. Industries are also increasingly supporting bioinformaticians as key contributors to R&D; and the publishing industry is trialing new types of journal devoted to documenting data dissemination strategies (e.g., GigaScience, started in 2012).

These new infrastructures are in turn embedded within wider political and economic contexts, which brings me to the second set of factors underlying the current push toward Open Data: the emphasis placed by public institutions responsible for science funding, often under pressure from national and international policy, on *fostering public trust in science as a source of reliable knowledge and thus as a legitimate source of information*.

Promoting the transparency and accessibility of science is particularly important in the face of its technical nature, which makes it hard to comprehend for the vast majority of citizens, and the disillusionment surrounding technoscientific achievement in the West. Cases in point are the controversies raging over the safety of genetically modified foods in Europe and the recent criminalization of earthquake specialists in Italy, accused of failing to predict the 2009 devastation of the city of L’Aquila. Perhaps the most blatant recent case of public mistrust in science is the controversy following the public release of e-mails exchanged by researchers at the Climatic Research Unit of the University of East Anglia in 2010 (an episode often referred to as ClimateGate). This was a case where a perceived lack of transparency in how climate data were handled fuelled social mistrust in the scientific consensus on global warming, which in turn affected public support for the implementation of international measures against climate change. Many national governments and international organizations like the European Research Council support the free circulation of data in the hope that it will increase the transparency and accountability of scientific research—and thus, potentially, its trustworthiness and social legitimacy. Similarly, the Royal Society has pointed to Open Data as an opportunity to prevent scientific fraud and disclose the evidence base for scientific pronouncements to the general public, so as to avoid the kind of miscommunication and misunderstanding underlying ClimateGate.

The increasing reliance on and support for online data dissemination is also intimately related to *the globalization of science beyond traditional centers of Euro-American power* (my third point). Open Data are implicated in transforming the geographies of science and its relation to local economies, as illustrated by the rise of centers of biomedical research excellence in the Global South. Centers such as the Beijing Genomics Institute, one of the powerhouses of contemporary data production in biomedicine, interact with researchers across the globe largely through digital means and do not see themselves as requiring the support of

extensive local or even national research infrastructure and traditions. Thanks to widespread data dissemination over the Internet, they can quickly learn from results produced elsewhere and contribute their own share of data to international databases and research projects, thus gaining visibility and competing with established programs in the United States, Japan, and Europe. Nations that have not figured as prominent producers of scientific knowledge throughout the 20th century, such as China and Singapore, are thus devoting increasing financial support to research, in the hope to attract a highly skilled workforce and boost their industrial productivity and economic prospects.

The conditions of possibility for data sharing, and the diverse motivations, stakes, and incentives underlying data donation to online databases, illustrate how widespread data dissemination has created new forms of inclusion and exclusion. One might think that laboratories in poor or underfunded regions would strongly support data sharing, for it makes data produced with expensive technology accessible to them, thus raising their chance to produce cutting-edge science, and that rich laboratories, which regard the possession of such technologies as providing them with a competitive edge, would be reluctant to donate data—particularly since donation requires additional labor. However, taking account of the considerable resources and diverse expertise needed to transform data into new knowledge helps acquire a more realistic view on the benefits and costs of data sharing. Underfunded laboratories actually struggle to access online resources, appropriate bandwidth, adequate expertise, and computers powerful enough to analyze data found online and are coming to terms with the difficulties involved in developing resources and standards for data donation. By contrast, many rich laboratories have found that data donation offers the opportunity to participate in international networks and receive help with data analysis, thus accruing their own prestige, visibility, and productivity. Even major pharmaceutical companies like GlaxoSmithKline and Syngenta are contributing to the development of public databases, in the hope of outsourcing their R&D efforts, improving their public image, and gaining from the availability of data produced through public funding.

This brings me to the last set of factors that I wish to discuss as relevant to explaining the prominence of Open Data as a contemporary social and scientific movement. This is the extent to which data dissemination practices are embedded into global political economies. *The very idea of scientific data as artifacts that can be traded, circulated across the globe, and reused to create new forms of value is indissolubly tied to market logics, with data figuring as objects of market exchange.* National governments and industries that have invested heavily on data production—through the financing of clinical trials or genome sequencing projects—are keen to see results. This requirement to maximize returns from past investments, and the urgency typically attached to it, fuels the emphasis on data needing to travel widely and

fast to create knowledge that would positively impact human health.

This is particularly evident when considering the extraordinary expectations linked to the Human Genome Project and its potential to contribute to medical advances. Indeed, it could be argued that the allure of big data lies precisely in the impossibility to predict and quantify their potential as evidence in advance. This makes it possible to inflate or deflate expectations to suit the dynamics of venture capitalism, while at the same time respecting the methods and ultimate goals of the scientific effort at hand. If we were able to predict exactly how a specific data set could be used in the future, and thus which data should or not be widely disseminated, we would not need Open Data in the first place: The point of free and widespread data dissemination is that one never knows who might be able to view which data and see something new in them, or indeed whether such fruitful use of data is at all possible.⁴ As exemplified by Mike Fortun's 2008 analysis of the deCODE case, the opportunity to circulate and reinterpret data is unavoidably couched in vastly promissory terms, and it is hard to differentiate a priori between a fruitful data sharing initiative and one that is unlikely to yield scientific insight—which makes financial investments in this area both risky and potentially rewarding.

Furthermore, the open dissemination and reuse of data do not challenge only notions of competition and property within established scientific communities but also notions of property, privacy, and effective communication in industry, government, and civil society.

Data acquired from patients in clinical trials or participants in personalized genomics, for instance, have clear economic value, and many pharmaceutical companies welcome the opportunity to access personal information unwittingly circulated by citizens who are not aware of its value as “data” for medical research—a move widely disputed by legal scholars, advocacy groups, and medical associations as an infringement of privacy (Tutton & Prainsack, 2011). The dissemination of data of relevance to innovation in food security or bioenergy, such as molecular data on plants and plant pathogens, is similarly plagued by uncertainties about intellectual property, particularly in cases of public–private partnerships between governmental agencies and companies such as Monsanto or Shell. It is yet unclear how fundamentally data sharing will transform industrial practices and relations to intellectual property, but it is notable that such an option is being actively considered.

Kaushik Sunder Rajan (2006) and Chris Kelty (2008) have shown how free data access has greatly helped maximize exchange and downstream capital flows. At the same time, data mobility is not free in the sense of being devoid of financial and social costs. Data sharing requires human resources and capital: Even the most successful initiatives are confronted with the exponential costs involved in maintaining and expanding data infrastructures in the long term

and are struggling to produce sustainable business plans for their activities (Bastow & Leonelli, 2010). Indeed, the European Union has denounced the costs associated with funding the current plurality of online databases in biomedicine as unsustainable in the long term, and is pushing for the centralization of facilities for data sharing as a possible solution (most prominently through ELIXIR, a gigantic effort currently underway at the European Bioinformatics Institute to coordinate and eventually integrate data sharing initiatives in biology and medicine; see the ELIXIR website: <http://www.elixir-europe.org/>). The National Science Foundation, which has funded many successful data-sharing initiatives at the turn of the millennium, is also attempting to rationalize its investments in this area and is now asking database curators to provide self-sustaining business models (e.g., International Arabidopsis Informatics Consortium, 2010).

To conclude: The emergence of technologies that facilitate the production and dissemination of biological data on a large scale is certainly a key reason for the visibility and political support garnered by the Open Data movement in recent years. Furthermore, adequate expertise and infrastructures are needed to assess the reliability and potential significance of data, thus ensuring that data journeys can contribute to the development of knowledge perceived to be relevant and useful within a variety of domains. In turn, the development of technologies and expertise for the care of data, not to mention their production and use to create new biomedical knowledge and interventions, is made possible by the availability of institutions that help define the financial value of data as commodities and the conditions under which data can be made to travel around the globe.

What has propelled data into becoming protagonists of contemporary biomedicine is their ambiguous status as at once local and global, free commodities and strategic investments, common goods and grounds for competition, potential evidence and meaningless information. Openness, defined through the opportunities for dissemination associated with the internet, is a defining characteristic of big data science, policy, and infrastructure. The vision underlying the Open Data movement is that data risk remaining meaningless if they are prevented from travelling far and wide and that travel endows data with multiple forms of scientific as well as financial, social, and political value. This flexibility to multiple uses and future scenarios, as well as to the diverging interests of potential users, is crucial to a contemporary biomedicine that is concerned with movement and that sees the constant rearticulations of knowledge to and through value.⁵

Acknowledgments

I am deeply thankful to Kaushik Sunder Rajan for pushing me to develop these ideas and for his insightful comments on several versions of this article. The Knowledge/Value network, which Kaushik created and continues to lead, provided me with an excellent opportunity to explore the interrelations between epistemic, political, and economic dimensions of data sharing, for which I

am grateful. For suggestions of relevance to this piece, I am particularly thankful to Gail Davies and other participants to the workshop “Knowledge/Value-3: Information, Databases and Archives” (held in September 2012 at the University of Chicago Beijing Center) and to Staffan Mueller-Wille, Barbara Prainsack, and Susan Kelly.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: My research was funded by the U.K. Economic and Social Research Council (ESRC), as part of the ESRC Centre for Genomics in Society and Grant ES/F028180/1.

Notes

1. One well-known example of such agreements are the so-called Bermuda Rules (Harvey & McMeekin, 2007; see also ongoing research by Rachel Ankeny and Robert Cook-Deegan).
2. These shifts in science policy are evidenced by the recommendations of the National Working Group on Expanding Access to Published Research Findings in the UK, also known as the Finch Report, and the subsequent statement by Research Councils UK, both from 2012.
3. For a fascinating case of a data infrastructure whose origins date back to the 1950s and whose success has been marked by moments of failure and profound transformation, see the Long-Term Ecological Research community (Anorova, Baker, & Oreskes, 2010; Baker & Millerand, 2010).
4. Let us not forget that the vast majority of genomic data made available online over the last ten years have yet to be used in ways that are scientifically meaningful; and that some scientists view much of the data being put online today as pointless information, which will never be exploited despite its wide accessibility.
5. These themes are expanded on in Sunder Rajan and Leonelli (2013).

References

- Anorova, E., Baker, K., & Oreskes, N. (2010). Big science and big data in biology: From the international geophysical year through the international biological program to the Long-Term Ecological Research (LTER) Network, 1957–Present. *Historical Studies in the Natural Sciences*, 40, 183–224.
- Baker, K. S., & Millerand, F. (2010). Infrastructuring ecology: challenges in achieving data sharing. In J. N. Parker, N. Vermeulen, & B. Penders (Eds.), *Collaboration in the New Life Sciences*. Aldershot (UK): Ashgate.
- Bastow, R., & Leonelli, S. (2010). Sustainable digital infrastructure. *EMBO Reports*, 11, 730–735.
- Bowker, G. C. (2001). Biodiversity datadiversity. *Social Studies of Science*, 30, 643–683.
- Edwards, P. D. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge: MIT Press.

- Fortun, M. A. (2008). *Promising genomics: Iceland and deCODE genetics in a world of speculation*. Berkeley: University of California Press.
- Harvey, M., & McMeekin, A. (2007). *Public or private economics of knowledge? Turbulence in the biological sciences*. Cheltenham, England: Edward Elgar.
- Hey, T., Tansley, S. & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.
- Hilgartner, S. (1995). Biomolecular databases: New communication regimes for biology? *Science Communication*, 17, 240-263.
- Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science*, 36, 269-298.
- International Arabidopsis Informatics Consortium. (2010). An international bioinformatics infrastructure to underpin the Arabidopsis community. *Plant Cell*, 22, 2530-2536.
- Kelty, C. M. (2008). *Two bits: The cultural significance of free software*. Durham, NC: Duke University Press.
- Kelty, C. M. (2012). This is not an article: Model organism newsletters and the question of "open science." *Biosocieties*, 7, 140-168.
- Kohler, R. E. (1994). *Lords of the fly: Drosophila genetics and the experimental life*. Chicago, IL: University of Chicago Press.
- Lenoir, T. (1999). Shaping biomedicine as an information science. In M. E. Bowden, T. B. Hahn, & R. V. Williams (Eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems (ASIS Monograph Series; pp. 27-45)*. Medford, NJ: Information Today.
- Leonelli, S. (2009a). Centralising labels to distribute data: The regulatory role of genomic consortia. In P. Atkinson, P. Glasner, & M. Lock (Eds.), *The handbook for genetics and society: Mapping the new genomic era* (pp. 469-485). London, England: Routledge.
- Leonelli, S. (2009b). On the locality of data and claims about phenomena. *Philosophy of Science*, 76, 737-749.
- Leonelli, S. (2010a). Documenting the emergence of bio-ontologies: Or, why researching bioinformatics requires HPSSB. *History and Philosophy of the Life Sciences*, 32, 105-126.
- Leonelli, S. (2010b). Packaging small facts for re-use: Databases in model organism biology. In P. Howlett & M. S. Morgan (Eds.), *How well do facts travel? The dissemination of reliable knowledge* (pp. 325-348). Cambridge, England: Cambridge University Press.
- Leonelli, S. (2012a). Classificatory theory in data-intensive science: The case of open biomedical ontologies. *International Studies in the Philosophy of Science*, 26, 47-65.
- Leonelli, S. (2012b). Making sense of data-driven research in the biological and the biomedical sciences. *Studies in History and Philosophy of the Biological and Biomedical Sciences*, 43, 1-3.
- Leonelli, S. (2012c). When humans are the exception: Cross-species databases at the interface of clinical and biological research. *Social Studies of Science*, 42, 214-236.
- Leonelli, S. (2013). Integrating data to produce new knowledge: Three modes of integration in plant science. *Studies in History and Philosophy of the Biological and Biomedical Sciences*. Advance online publication. doi:10.1016/j.shpsc.2013.03.020
- Leonelli, S. (in press). Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties*.
- Leonelli, S., & Ankeny, R. A. (2012). Re-thinking organisms: The epistemic impact of databases on model organism biology. *Studies in History and Philosophy of the Biological and Biomedical Sciences*, 43, 29-36.
- National Working Group on Expanding Access to Published Research Findings. (2012). *Accessibility, sustainability, excellence: How to expand access to research publications* (Finch Report). Retrieved from <http://www.researchinfonet.org/publish/finch/>
- OECD. (2009). *OECD guidelines for human biobanks and genetic research databases (HBGRDs)*. Retrieved from www.oecd.org/sti/biotechnology/hbgrd
- Research Councils UK. (2012). *Research Councils UK policy on access to research outputs*. Retrieved from http://www.rcuk.ac.uk/documents/documents/RCUK%20_Policy_on_Access_to_Research_Outputs.pdf
- Royal Society. (2012). *Science as an open enterprise. Report*. Retrieved from <http://royalsociety.org/policy/projects/science-public-enterprise/report/>
- Star, S. L., & Ruhleder, K. (1996). Steps towards an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7, 63-92.
- Sunder Rajan, K. (2006). *Biocapital: The constitution of post-genomic life*. Durham, NC: Duke University Press.
- Sunder Rajan, K., & Leonelli, S. (2013). Introduction: Biomedical trans-actions, postgenomics and knowledge/value. *Public Culture*, 25, 3.
- Tutton, R., & Prainsack, B. (2011). Enterprising or altruistic selves? Making up research subjects in genetics research. *Sociology of Health & Illness*, 33, 1081-1095.

Author Biography

Sabina Leonelli is a senior lecturer in the Department of Sociology, Philosophy and Anthropology at the University of Exeter, United Kingdom. She uses historical and sociological research to foster philosophical understandings of knowledge-making processes, focusing particularly on data-intensive science, model organism biology, and plant science. She is the coeditor of *Scientific Understanding: Philosophical Perspectives* (2009) and the associate editor of the journal *History and Philosophy of the Life Sciences*.