

<http://somatosphere.net/?p=9752>

Open access, open data, open science...what does “openness” mean in the first place?

2015-02-11 09:17:30

By Nadine Levin

Recently, the research community has been flooded with encouragement to make things “open,” meaning: freely and easily accessible, in a variety of ways, and to a great variety of audiences. This impetus to be open has taken the form of debates over “[Open Source](#)” software licensing, “[Open Access](#)” to the published results of research[i], “[Open Innovation](#)” across organizations, and “[Open Data](#)” in research and government contexts. We are in the era of the “[New Scientific Revolution](#)”, where the fruits of research will (supposedly...) flow out of the academic ivory tower to transform health, society, and beyond. (Most academics will probably be familiar with these terms by now, even fleetingly, through publicity from the [AAA](#) or [The White House](#).)

Such [encouragement, commitment, and raw faith in the power of “open”](#) is—as [Sabina Leonelli](#) and I have been exploring for the past year and a half—an articulation of what is broadly known as the “Open Science” movement. This movement aims to make not only publications and data, but also research materials and methods open, in an effort to [foster equality, widen participation, and increase innovation and productivity](#). Indeed, Open Science has been promoted globally as a key component of modern society by organizations such as The [Wellcome Trust](#), [Royal Society](#), [National Institutes of Health](#), [Center for Open Science](#), Open Knowledge Foundation, and many others. It has become increasingly central to science policy in both the [United Kingdom](#) and the [United States](#).

But, as Sabina and I have been documenting, there remains a remarkable lack of clarity as to how the implementation, practice, and enforcement of Open Science—or to rephrase it, “openness” in science—should occur. [Policies have different terms and requirements, educational institutions have different infrastructures](#), scholarly communities have different commitments, and individual researchers have different priorities and circumstances[iii]. Consequently, the project has been exploring what exactly “openness” means in modern society: how, for whom, when, and where does openness occur, and what forms of scientific labor and value does the norm of openness highlight or obscure?

In everyday research, Open Science takes many forms. It can be researchers putting their data into online databases such as [GenBank](#), [figshare](#), or into [journal repositories](#). Or, it can be researchers developing standards for data format and quality, led by institutions like the [European Bioinformatics Institute \(EBI\)](#) and the [National Center for Biotechnology Information \(NCBI\)](#). More commonly, it might involve researchers publishing in open access journals like the [Public Library of Science \(PLOS\)](#), or publishing open access articles in highly-regarded journals like Nature, Science, and Proceedings of the National Academy of Sciences. The range of these activities is staggering—involving multiple things, groups of people, and geographies—and to describe it would be beyond the scope of this blog post.

Open Science, or “openness” in science, brings unquestionable benefits. The biomedical researchers we have interviewed, for example, emphasize that it avoids the duplication of work, and ensures the transparency of peer review. Likewise, anthropologists emphasize that it enables the public—and research subjects/informants—to access and learn from the academic research that purports to be about and for the public.

But it also raises critical—and significantly less discussed—[questions about how such openness is being done, and is playing out, in practice](#). In other words, *how*—in terms of making things not only available, but also usable and useful—should openness be done? *By whom*—early career researchers or established professors—should openness be practiced, and *to whom*—people in Universities, companies, or developing nations—should the presumed benefits of openness be directed? *When* in research—at the very beginning, before publication—should openness occur? These questions, which are lurking beneath the surface of the various “open” movements, [pose serious challenges](#), as researchers attempt to negotiating the uneven, non-uniform terrain of the Open Science wilderness.

To explore some of these questions and challenges, as a starting point, it helps to remember that such encouragement to be “open” is not new, and dates back to the mid-20th century (a la Robert Merton) and even earlier^[iii]. Its contours are not even or uniform, as it primarily relates to societies that have the digital infrastructures to share and generate information (for example, in Africa and South America). Understanding “openness” in contemporary society is all about context, [something that anthropologists and social scientists are well attuned to notice and study](#).

Through interviews with a variety of UK-based advocates of “Open Science,” this is exactly what we have learned. Despite the overwhelming emphasis and value placed on being “open,” there are a plethora of cases in which researchers struggle to be open because of contextual,

situational considerations. These considerations are about a range of things: credit and career structures, the pressures of competition and industry collaboration, modes of intellectual property, and compliance with University and Government policies. These might be classified as the social, economic, and political contexts that, as many scholars in anthropology and science and technology studies have observed, govern scientific norms and practices. But the challenge of investigating how researchers go about being “open” lies in understanding how—and if—the institutions or communities that set particular agendas for openness reinforce particular norms and values, which at times conflict with researchers practices and experiences.

To anchor this discussion a real-life example, let me turn to a moment to a case study that emerged in our interviews with one of the 22 researchers we interviewed in the United Kingdom as part of this project. This particular principal investigator (PI) was involved in systems biology research to develop models of plant biology, which used complex equations to describe cellular processes. These models used large volumes of molecular data, and therefore relied on the existence of databases containing the data from various research groups and labs.

The PI has experienced developing such a database, emphasizing that in the data-intensive world of systems biology, [the data contained in databases not only needed to be available, but also useful](#). In other words, [just having access to raw data would not help](#) build models: instead, researchers needed access to curated data, which had been carefully cleaned up, edited, and annotated with metadata describing the conditions under which the data had been collected.

The PI acknowledged that transforming data for a “raw” into a “useful” state required the effort and time of the systems biology community, often in the form of “data donation,” where people might not be recognized according to traditional publication-based metrics for their time and effort^[iv]. And yet, he believed that such data donation would not only benefit his own research, but would ultimately benefit the systems biology community as a whole, by allowing researchers to do novel experiments or validate existing models.

So, to give people an “incentive” to donate high-quality data to the database, the PI gave people who donated data access to a suite of online data analysis tools, which let them run experiments on the large collection of data within the database. This, he said, provided an “immediate return” to people for their efforts, and was a sort of reciprocal “gift” to the data donors.

But then, the PI explained that this presented him with a “dilemma of

openness,” a challenge which he did not know how to solve. He acknowledged that he relied on the complete openness of the database users, but then he acknowledged that, in return, he was not able to be completely open with the data analysis tools he gave users access to. Although he gave the data donors the ability to use these software tools online, he did not give them the source code: in other words, he did not enable the data donors to use the software on their own computers, nor did he give them access to the internal “logic” of the data analysis tools. His reasons for doing so were, he claimed, to ensure that people did not install the data analysis tools on their own computers, and subsequently stop contributing to the database. He felt that [if he did not incentivize people to carry out data donation and sharing, they would cease to do it.](#)

The point we are trying to make with this case study is that there are a variety of different types of openness, all of which hinge on [making some things open, and some things closed](#). Openness is not a binary, but rather indicates a particular strategy for engaging in a type of openness, in order to achieve a specific end result. In the end, this PI’s struggle typified what many researchers expressed as an attempt to strike a balance between providing benefits to the individual researcher and to the community. The PI’s engagement with a particular type of openness indicated an attempt to encourage people to contribute to the overall systems biology community through data donation, but also to give people the opportunity to derive personal benefit by using data analysis tools on a large pool of data.

However, this case study also raises broader questions about the types of norms and values that are embedded within Open Science initiatives, and how these affect practicing researchers. As [Javier Lezaun and Catherine Montgomery have written in their excellent paper on Open Innovation in pharmaceutical research](#), notions of “sharing” and “openness” are predicated on the labor of researchers and institutions, who feel the imperative to put research materials into circulation. Within Open Science, researchers are encouraged to make labor-intensive and value-laden outputs open. And yet, it is precisely these outputs that often remain unacknowledged or under-valued, and which researchers resist making, or do not want to make, open.

Overall, it helps to remember that the materials and objects of research do not have value in and of themselves, and instead require work—through policies, norms, economies, and infrastructures—to have or be denied value. Sabina and I, in a paper we are currently developing, argue that Open Science generates notions of value ([which some argue is part of the neoliberal economy of higher education](#)), which become embedded in the objects that researchers are encouraged to share and circulate. While the more “traditional” outputs of scientific research—data, papers, intellectual

property—are valued by the Open Science movement, they also contain less tangible aspects of value to researchers—skill, labor, knowhow, attribution, credit. This presents a very real tension for researchers, like the PI whose story I recounted above, as they try to negotiate openness in practice.

[Nadine Levin](#) is a Postdoctoral Fellow at UCLA in the Institute for Society and Genetics, where she is working on an NSF-funded project on “What is metabolism after big data?” and what consequences this has for biomedicine. Previously, she has done work at the University of Exeter on how Open Access and Open Data policies affect the practice of post-genomic research, and also on how intellectual property regimes in the arts and humanities are affected by “the digital”. She completed her DPhil in 2013 in Anthropology at Oxford University, with a dissertation that explored how researchers in the field of metabolomics create, analyze, and use data to make claims about metabolism and health.

Notes

[i] You can read more about how this affects Arts & Humanities researchers in [another blog post I've written](#)

[ii] See the recent Nuffield Council on Bioethics report on “[Biological and health data](#)”

[iii] See [Chris Kelty's work on model organism newsletters](#)

[iv] See [Sabina Leonelli and Rachel Ankeny's forthcoming paper on data donation and credit attribution in the edited volume “Postgenomics”](#)

AMA citation

Levin N. Open access, open data, open science...what does “openness” mean in the first place?. *Somatosphere*. 2015. Available at: <http://somatosphere.net/?p=9752>. Accessed February 11, 2015.

APA citation

Levin, Nadine. (2015). *Open access, open data, open science...what does “openness” mean in the first place?*. Retrieved February 11, 2015, from Somatosphere Web site: <http://somatosphere.net/?p=9752>

Chicago citation

Levin, Nadine. 2015. Open access, open data, open science...what does "openness" mean in the first place?. Somatosphere.
<http://somatosphere.net/?p=9752> (accessed February 11, 2015).

Harvard citation

Levin, N 2015, *Open access, open data, open science...what does "openness" mean in the first place?*, Somatosphere. Retrieved February 11, 2015, from <<http://somatosphere.net/?p=9752>>

MLA citation

Levin, Nadine. "Open access, open data, open science...what does "openness" mean in the first place?." 11 Feb. 2015. Somatosphere. Accessed 11 Feb. 2015.<<http://somatosphere.net/?p=9752>>