2

SCIENTIFIC AGENCY AND SOCIAL SCAFFOLDING IN CONTEMPORARY DATA-INTENSIVE BIOLOGY

SABINA LEONELLI

Philosophers of science are starting to pay attention to the impact of communication technologies, particularly those functioning as means to share results and resources, such as data or materials, on scientific methods and epistemology (Callebaut 2012; Leonelli 2012; O'Malley and Soyer 2012; Ratti 2015). This is especially salient in so-called big data initiatives, where high-throughput means of data production (such as sequencing machines, particle colliders, and space telescopes) are coupled with new technologies for the dissemination, integration, and visualization of the resulting masses of data (such as online databases and software for data analysis). Several commentators have described this phenomenon as an "information turn" in the practices of knowledge production (Castells 1996; Gibbons et al. 1996; Hey, Tansley, and Tolle 2009; Floridi 2013). What philosophers tend to overlook, however, is the significant role of social scaffolds in the development and implementation of these technologies toward generating new research. Social scaffolds include project teams, research networks, scientific institutions, policy bodies, learned societies, governmental committees, and other relevant forms of social engagement and governance. Here, I explore the circumstances under which specific types of social scaffolding facilitate advances in research and the reasons why some forms of sociality are effective in promoting certain kinds of scientific work. I concentrate on cases in which scientists coordinate their efforts with the goal of creating groups responsible for articulating common concerns, making these concerns visible to peers as well as funders and publishers, and developing ways to address them in everyday research practice. As I will show, these groups need to acquire resilience to endure the ever-shifting landscape of short-term funding agreements, fast-moving technologies, and multiple clusters of expertise that support research in any given field. This resilience is necessary, given the challenges and time involved in gaining enough visibility to command the attention of well-established regulatory institutions, such as governmental funders and learned societies. At the same time, these groups of scientists also need to be flexible and responsive enough to retain their usefulness vis-à-vis the shifting needs of relevant scientific communities. I argue that in their attempts to straddle these requirements, scientists tend to rely on well-entrenched social configurations and coordination strategies, some of which political theorists looking at the emergence and establishment of social movements have singled out and examined. Borrowing key ideas from social movement theory, I show how they can help us to understand the evolution of regulatory structures aimed at facilitating scientists' engagement with new technologies to enhance research outputs.

My discussion will be grounded in the examination of two types of organizations that have been heavily involved in developing practices of data dissemination through digital means within the life sciences over the last decade. These are (1) ontology *consortia,* which were created by biologists to promote online tools to classify and disseminate data and have evolved into *de facto* regulatory bodies in bioinformatics and data curation in the United States and Europe, and (2) *steering committees* for model organism communities, whose success in enhancing the cohesion, visibility, and reputation of biological research resulted in their playing significant roles in the governance of research. These are cases in which individual researchers successfully joined forces to build representation and political agency for their scientific concerns that resulted in the creation of organizations with regulatory power over research activities at the national and sometimes even the international level. They are also instances of two broader types of social structures that play a crucial role in the management of virtually every field: consortia and steering committees. Yet these have received little attention from science studies scholars, especially in comparison to "networks" and "laboratories," which have been central units of analysis for social scientific work in this area over the last twenty years.[1]

The chapter is structured as follows. In the first section, I briefly document the emergence of these groups and their successful transformation into scientific institutions with political and epistemic visibility and agency. Next, drawing on ideas from political theory, I argue that viewing these organizations as social movements is a fruitful strategy to make sense of their development from informal groups into well-recognized regulatory bodies. In the third section, I discuss how this process of institutionalization builds on highly entrenched forms of group socialization (*core configurations*) while also fitting the modular and highly dynamic nature of current research networks (Caporael 1997), which typically involve short-term collaborations around individual projects. In conclusion, I reflect on how my analysis could inform studies of the interrelation between institutional and infrastructural scaffolding involved in the evolution of scientific knowledge-making activities.

REGULATING DATA DISSEMINATION IN CONTEMPORARY BIOLOGY

Over the last three decades, scientific societies, governmental bodies, and industry have devoted increasing attention to the opportunities offered by the implementation of new technologies for the production and dissemination of biological research data (Leonelli 2016).[2] The sheer amount of organization, standardization, and infrastructure required to store and disseminate biological data—as well as the bureaucracy, institutional accountabilities, and red tape developed to that end—arguably exceeds anything previously experienced within the life sciences. In the words of prominent scientific commentators: "The introduction in 2005 of so-called next generation sequencing instruments that are capable of producing millions of DNA sequences has not only led to a huge increase in genetic information but has also placed bioinformatics, and life science research in general, at the leading edge of infrastructure development for the storage, movement, analysis, interpretation and visualisation of petabyte-scale datasets" (Southan and Cameron 2009, 119).[3]

The development of efficient data-sharing practices requires insights from the producers and users of data, whose understanding of their quality and significance as research materials is unparalleled. At the same time, individual scientists are not typically in a position to control the considerable resources and man power required to build relevant infrastructures, policies, and standards nor does scientific expertise constitute the only source of insight with regard to the value of research data. Indeed, data management on such a large scale requires a variety of skills, expertise, and insight, which include not only scientific assessment but also social, political, legal, and economic understanding of the circumstances under which data can be stored, maintained, and reused. Biologists interested in data dissemination have long struggled with the complex cluster of expertise and political visibility needed to debate—let alone decide upon—data-management and data-sharing strategies, as demonstrated by the history of data-sharing agreements like the Bermuda Rules (Harvey and McMeekin 2007; Jones, Ankeny, and Cook-Deegan 2018). Two initiatives that groups of biologists have taken in order to organize the public dissemination of research data produced within their field—the Gene Ontology and the Genomic Arabidopsis Resource Network—illustrate how scientists can and do join forces to influence the governance of their research in ways that favor their professional interests and intellectual commitments. Both types of collective action required the development of common standards and practices geared toward the resolution of scientific problems emerging in specific research contexts. At the same time, establishing such standards was intertwined with developing and implementing a regulatory system for scientific research targeted toward addressing the needs and characteristics of the groups involved.

Consortia and the Case of the Gene Ontology

The term *consortium* has recently acquired popularity within the life sciences as a way to refer to scientific collectives brought together by a common set of concerns. These span from an interest in specific phenomena (e.g., the Beta Cell Biology Consortium, devoted to pancreatic islet development and function, http://www.betacell.org/) to solving a common technical problem (e.g., the Flowers Consortium in the United Kingdom, aimed at creating a common infrastructure for synthetic biology, http://www.synbiuk.org/) or promoting a specific standard or technique (e.g., the Molecular Biology Consortium [MBC], founded to further high-throughput analysis of biomolecular and subcellular structures via a superbend X-ray beamline at the advanced light source, http://www.mbc-als.org/). The members of a consortium, which can be individuals as well as groups, labs, and institutes, do not need to be located in the same geographic site or belong to the same discipline. Indeed, the term is typically used to designate groups of scientists based in different institutions around the world and from a variety of disciplinary backgrounds. Consortia are sometimes fueled by dedicated funding, most often provided by governmental bodies interested in supporting a specific area of scientific work. In other cases, financial support is achieved by bringing together a variety of public and private resources. One example is the

Gene Ontology Consortium, which was created to develop and promote a particular tool for online data dissemination: the Gene Ontology (GO).

GO was created in 1999 as an alternative to the classification systems for genomic data proposed within medical informatics. The group of curators involved in the GO Consortium started their involvement as scientists discontented with how data were organized in databases at that time. They set out to create a resource that would do a better job of representing biologists' needs. In 1998, the group consisted of only five representatives from the yeast, mice, and fly communities, who saw themselves as fighting for a biology-driven bioinformatics. Their involvement with GO stemmed from their dissatisfaction with the ways in which medical informatics, as a field, was handling the setup of data-sharing tools in biomedicine, particularly model organism biology. They felt that the voices of biologists actually producing and working with these data were not being heard and endeavored to produce a set of tools grounded in biological know-how and geared toward the expectations and needs of biology users (for more historical detail, see Leonelli 2009, 2010). In 2000, funding for their efforts started to trickle in, and they found themselves in a position to recruit more like-minded researchers from other model organism communities. Following the explosion of data-intensive methods and related data infrastructures, these efforts came to be more widely recognized as crucial to the future development of biological research as a whole. The GO group expanded to include a head office based at the European Bioinformatics Institute (EBI) in the United Kingdom, counting up to ten researchers at any one time and at least twenty affiliated data curators spread around the world. These curators come together as a collective in regular meetings, online discussions, and funding applications. While many of the curators shift periodically, depending on project funding and local institutional arrangements, some have persisted as a long-term core group of affiliated scholars since the start of the project. GO has been increasingly institutionalized, both as part of the EBI and through strong links with the National Centre for Biomedical Ontology in the United States. Still, it continues to rely on voluntary contributions of participants, both financially and in terms of man power and data donation. For example, representatives from FlyBase, the database devoted to the dissemination of data on the fruit fly *Drosophila melanogaster,* contribute as much as they can justify under the remit of their project funding. Many others involved with organism databases do the same (e.g., the Arabidopsis Information Resource and WormBase, for the nematode *Caenorhabditis elegans*).

In previous work (Leonelli 2009, 2010), I have discussed the function of the GO Consortium as a powerful force within biology and beyond. The consortium has successfully developed procedures and technologies through which users can interact and upload, retrieve, and analyze data. It has also strongly influenced what counts as professional training for data curators in model organism databases—most notably, by helping to establish the International Society for Biocuration, which largely defined best practices for this field and strengthened its professional standing. Moreover, it has contributed to promoting values such as open access to data, intercommunity cooperation, and diversity in epistemic practices across biology, as well as fostering the pursuit of common goals, including specific kinds of cross-species integrative biology. All these activities involve networking with both the biological communities interested in the data being disseminated and the funding bodies and learned societies involved in supporting the relevant biological fields. The successes of GO signal the impressive increase in regulatory power, international visibility, and political resonance that this group has enjoyed since its origin. The GO Consortium has played an important role as an agent of change within the biological community.

Steering Committees and the Case of GARNet

A similar case study demonstrates the ways in which model organism communities have organized and coordinated themselves, resulting in an affirmation of their identity as key actors within the scientific landscape. Such organization is provided largely by steering committees: groups of representatives from the community who meet regularly to discuss future directions for the community as a whole (typically, some of the most active principal investigators [PIs], either elected by the community or sometimes self-appointed). One of these steering committees is GARNet, the Genomic Arabidopsis Resource Network. GARNet consists of plant scientists working on the model organism *Arabidopsis thaliana*. Most committee members are elected for a three-year term by UK researchers who self-identify as having an interest in *Arabidopsis* research, with efforts made at every election to ensure a fair representation in terms of research interests, gender, and geographical spread. Coordination and long-term memory is provided by two GARNet coordinator posts, one that has been in place since the committee's birth and another consisting of different individuals over the years; the committee chairs and PIs of the GARNet grant, who have shifted over the years but continue to maintain a close affiliation with the group even after the end of their mandates; and two ex officio committee members (the director of the European Arabidopsis Stock Centre, who has been part of the committee since its birth, and myself as an *Arabidopsis* historian and plant data expert since 2009). GARNet was created in 2000 as part of the gene function initiative funded by the Biotechnology and Biological Sciences Research Council (BBSRC) in the United Kingdom. While its initial remit was to ensure the availability of functional genomic technologies across UK plant science labs (Beale et al. 2002), GARNet has succeeded in obtaining two further rounds of funding from the BBSRC and has established itself as one of the most important organizations for the coordination, steering, and representation of basic plant research in the United Kingdom and internationally. This has happened through several initiatives, including (1) establishing a website and regular newsletter, which constitute unique information sources for new resources and initiatives in the field (principally concerning data but also embracing experimental techniques and instruments, as well as new funding opportunities); (2) organizing annual meetings attracting *Arabidopsis* scientists but also, increasingly, other plant scientists interested in updates on opportunities, techniques, and technologies for cross-species research; (3) coordinating dialogue among key stakeholders in the field, including learned societies like the Society of Biology, key funders such as the BBSRC, and the publishing industry responsible for the leading journals in plant science; (4) setting up surveys across the plant community, with the objective of articulating scientists' perception of what constitutes interesting new research directions and communicating it to funders (e.g., a survey commissioned by the BBSRC on the status of system biology in plant research); and (5) monitoring the number of resources funding bodies allocate to plant science vis-à-vis other parts of biology and lobbying for more resources and attention to be allocated to plant scientists.

As a result of these activities, GARNet now plays a central role in mediating the transition of the UK plant science community from a focus on functional genomics to system/synthetic plant science and translational research. Indeed, GARNet played a key role in integrating research conducted on *Arabidopsis* (traditionally funded by the BBSRC and viewed as fundamental research with no immediate applicability) with research carried out on crops such as barley, maize, and wheat (traditionally funded by the Department for Environment, Food and Rural Affairs and viewed as applied biotechnology). The rapprochement of these two communities was needed and overdue: *Arabidopsis* research has advanced to yield precious insights for agriculture (e.g., how to increase plant yield) and emerging biofuels (e.g., how to increase cell metabolism to make plants produce more butanol). Additionally, crop science is realizing that *Arabidopsis* research provides excellent comparative tools for research across plant species. GARNet has taken the lead in coordinating meetings among investigators in both communities, resulting in the founding of the UK Plant Science Federation (Leonelli et al. 2012). GARNet has also strongly affected the provision of bioinformatic services to plant scientists and biologists interested in *Arabidopsis* data. In 2009, the National Science Foundation decided to dramatically cut funding to a key database, the Arabidopsis Information Resource (TAIR), due largely to a lack of long-term sustainability for such an infrastructure. GARNet organized two international workshops that gathered powerful PIs, information technology (IT) experts, and funders to discuss models for the long-term maintenance and development of databases in plant science, helped find an agreement for how TAIR was to survive and develop in the future, and provided guidance on how similar databases could be made more resilient and useful to researchers.

SELF-REGULATORY EFFORTS AS SOCIAL MOVEMENTS

Consortia and steering committees, exemplified in the previous cases, share a number of features. They are self-organized collectives, whose joint activities begin without a great deal of support from well-established institutions or even from the communities in which they operate. Individuals propose themselves as representative champions for their communities, with the duty to voice scientists' existing concerns and facilitate solutions to those problems. These collectives support a wider spectrum of values and ideals than the specific issues they emerged to tackle, such as fostering initiatives requiring broad changes in the governance of the social system within which they are working. Initially, these organized efforts were devised as provisional responses to a localized issue in data management and dissemination. They persist with minimal dedicated funding thanks to the voluntary support and contributions of members of the communities they represent. Despite a precarious status in the early stages of their operation, these self-organized collectives have garnered visibility and political power, building their credibility by strongly connecting to their communities and attempting to articulate scientists' concerns in a way that bridges communication gaps with relevant peers and other stakeholders. It is not a coincidence that the biological communities that managed to organize themselves in this way are among the largest and most successful today. As a consequence, the model organisms these groups have championed are currently recognized as the most important in experimental biology (Ankeny and Leonelli 2011), indeed exemplifying a specific mode of doing research that has come to define much of the field (Ankeny and Leonelli 2016). All this has happened within a relatively short period of time: both the GO Consortium and the GARNet steering committee have gone from outsider status to participating in the primary regulation of biological research within the space of ten years.

The scientists engaged in these efforts demonstrate an acute awareness of the deep ties between power and standardization and of the ways these ties affect day-to-day research practices. They have effectively created systems of governance via a complex web of activities (including sophisticated marketing strategies and enrollment techniques) within which the standards and norms they propose may help to address issues emerging from scientific work. How should we characterize these groups of scientists and their activities? What kind of collective agency is in operation, and how does it achieve both power and impact? One way to consider these questions is in light of discussions about the emergence and status of so-called new social and scientific/intellectual movements. Drawing from this literature is not a new idea, and I will refer to authors who have advanced similar views with respect to scientific agency. However, I believe this to be a powerful lens with which to analyze the development of contemporary biological knowledge, particularly the creation and implementation of standards and infrastructures to disseminate data. From this corpus of literature, I have extracted four characteristic features of social movements that can be readily observed in both case studies. I propose that we view these scientific consortia and steering committees as social movements because they exhibit these four characteristic features:

1. They emerge in response to changing research needs and landscapes.
2. They establish new practices.
3. They create a vision for how research should be conducted in the future.
4. They become political actors with the power to engender social, scientific, legal, and political shifts (e.g., data-sharing policies, rules for database access, publication strategies, or shifts to the credit system in science).

Movements as Reactions

Della Porta and Diani (1999, 6) define new social movements as

1. informal networks, based on
2. shared beliefs and solidarity, which mobilize about
3. *conflictual issues,* through
4. the frequent use of various forms of *protest.*

The emphasis within this definition is on the role of movements as *reactions* to the existing status quo. This is an important and suggestive intuition; the collective action characterizing consortia and steering committees is driven by the desire to resolve existing problems. For GO and GARNet, these problems emerge from scientific practice. To this end, a high level of epistemic and political agreement is required and must be targeted to specific issues. Consortia and steering committees are committed to using a rational, knowledge-based approach to reach such consensus; these are expert movements for an expert community and usage. This often means antagonizing the establishment, as in the case of many nonscientific social movements.

> A movement is a social/intellectual movement by our definition only if, at the time of its emergence, it significantly challenges received wisdom or dominant ways of approaching some problem or issue and thus encounters resistance. (Frickel and Gross 2005, 207)

For Frickel and Gross, the notion of "resistance," interpreted as opposition to a discriminating majority, is central. I agree that for cases of consortia and steering committees a degree of resistance and challenge to previous practices and normative demands that characterize a field or domain is involved. But although this provides a key motivation for collective action, there is another noteworthy goal central to the collective agency that initiates consortia. This is to draw attention to issues that have not been the focus of funding agencies or of the scientific community and yet have caused trouble for research (or are likely to do so in the future). In these cases a regulatory need is going unrecognized by regulatory bodies; thus, there is an opportunity to delegate decision-making power (and annexed responsibility) to a new form of agency or actor. If successful, some people or institutions are willing (or forced) to absorb the regulatory need, either because they are identified as likely candidates or because they are created for that purpose. Additionally, other scientists are happy to delegate responsibility to these new movements; they willingly give up their decisional power over the issues. A similar dynamic is currently seen in the rise of organizations such as the Research Data Alliance (2016), which started as a group of open science advocates and lobbyists in 2010 and within five years became a reference point for governments and funding agencies looking for guidance on how to collect and mobilize research data across all areas of society.

Movements as Collective Creation

Another significant feature of social movements is that they aim to create something new: "Temporary public spaces, movements of collective creation that provide societies with ideas, identities, and even ideals" (Eyerman and Jamison 1991, 4). GO is a good example of this kind of consortium, which is primarily geared toward the development of new tools and knowledge. GO managed to channel the creative energies of a number of prominent biologists and bioinformaticians into the development of a unique and highly popular database. At the same time, building the momentum and opportunity for such an endeavor is itself an imaginative and laborious act. Social movements have been defined as "luxury goods" because they need support in order to take off on the scale required for collective action to be effective. Thus, they are typically organized around "hot issues" most likely to attract the attention of funders and peers. (This is the case with both data infrastructure and synthetic and translational plant biology.) It is also critical to note the importance of the collective experience of unity through action as a means to form a social identity. The formation of a social nucleus with a distinct identity and sense of membership happens simultaneously with the focus on a common set of issues. Notably, the social unity or cohesion of the group is more important than agreement or consensus on the specifics of the issue itself; what matters is the individuals in the group's sense of agreement and belonging and their willingness to invest resources toward the same normative vision. Indeed, both GO and GARNet have contributed greatly to forming a well-defined research community bound together by similar worries and obligations.[4] Unavoidably, this has also involved conflicts over boundaries, the exclusion of individuals or groups for financial, geographical, or personal reasons (no matter how inclusive both the GO and the

GARNet groups strive to be), and the formation of other communities striving to counter or emulate their increasing visibility and resources.

Movements as Signs of Change

Melucci (1996, 1) proposes yet another definition of social movements:

> Movements are a sign; they are not merely an outcome of the crisis, the last throes of a passing society. They signal a deep transformation in the logic and processes that guide complex societies. Like the prophets, movements "speak before": they *announce what is taking shape* even before its direction and content has become clear.

Thus, according to Melucci, social movements have the key function of voicing a normative vision—in this sense they are "signs of change." This function is visible in both case studies in which the collectives in question have developed specific visions of what counts as good science (e.g., norms regulating standardization and data curation in databases; a commitment to enhancing research efficiency through collaboration and coordination in plant science). These visions play a key role in forming social identities (see section 2.2), but they also contribute to wider debates about the appropriateness of specific goals, norms, and methods in research at large and the changes that new technological and social developments foster. Another example can be seen in the ideas of Science 2.0 and Open Science, which the European Commission has used over the last decade to capture a perceived ongoing shift in the practice and results of science. This feature parallels the study of the formation of *communities of promise* with a common imagination, such as can be observed in the case of epistemic networks formed around stem cell research (Martin, Brown, and Kraft 2008), and, more generally, the study of the development and function of scientific *imaginaries* (Jasanoff and Kim 2015).

Notably, elaborating such a vision does not necessarily involve an explicit contrast between it and preexisting views. Making visions identifiable as new entities (i.e., as signs of change) is as important as building some continuity with the intellectual traditions characterizing the epistemic communities to which the vision is directed. A vision needs to be anchored somewhere in order to be understood. The language used to express the vision, the practices it involves, and the problems it is supposed to solve all need to be situated in specific contexts that coevolve with the vision itself. If spokespersons for a new vision cannot latch on to (and influence) one or more preexisting intellectual traditions, they will find it difficult, if not impossible, to enroll new participants in their movements (Frickel and Gross 2005, 221). This is on display in both GARNet and the GO Consortium because participants stress that these organizations championed existing understandings of good practice and reliable data sharing within model organism biology, particularly in the face of other ways of handling data preferred by other communities.

Movements as Rising Power

One final characteristic of social movements concerns the role of power dynamics in the emergence and operation of consortia and steering committees, including the importance of long-term influences on the environment.

> When backed by dense social networks and galvanised by culturally resonant, action-oriented symbols, contentious politics leads to *sustained* interaction with opponents. The result is a social movement. (Tarrow 1998, 2)

The actions of GO and GARNet (as well as other types of scientific organizations) result in the acquisition of political representation and agency on national and global agendas, even though their immediate target is primarily needs arising from day-to-day research practice. This large-scale political representation and agency often goes well beyond the resolution of the initial problems and can be referred to as these movements' *rising power.* Although often mentioned in sociological and anthropological studies of emerging fields, the ways in which such power is developed in and through scientific practice deserves much more research. For example, by what diverse paths does a movement quickly develop an internal hierarchy and administration in order to function, which in some cases transform into a semiofficial agency? The National Institute for Health and Care Excellence in the United Kingdom, which started as a grassroots movement of doctors trying to monitor the safety of guidelines provided by the National Health Service, is now a major evaluation agency with tremendous clout over government and patient organizations. A key element in this type of development is "access to key resources" (Frickel and Gross 2005, 214). These resources include (1) *organizational structures,* such as channels for information flow (e.g., conference venues and publications), frequently linked to epistemic cultures; (2) *intellectual power,* grown in parallel with the reputation and personal credibility of the movement's leaders and with the assessment of their vision and actions developed by peers over time; and (3) *long-term employment* within academia for at least some of the movement's leaders, which provides the stability and continuity necessary to the blossoming of collective agency on a large scale. Another important element is the ability to raise bottom-up support or *micromobilization* (Frickel and Gross 2005, 220). All of these display parallels to the situation outlined by Kaushik Sunder Rajan (2006, 52) in relation to what he calls "new corporate activism": corporations' political strategies for influencing the outcome of issues affecting their organizations.

ENTRENCHED CONFIGURATIONS AS SOURCES OF SOCIAL ROBUSTNESS

I have described four characteristics that social movements seem to have in common with scientists' attempts to regulate their own activities. Both ontology consortia and steering committees are instances of collective self-regulation stemming from perceived needs in a scientific field (e.g., conflict or lack of resources). They formulate creative solutions to such problems, which are developed and implemented by groups of individuals with the expertise to recognize the problems and to present them so that others within their field will recognize them as well. Additionally, these groups exhibit an entrepreneurial ability to devise ways in which risky, collective efforts can contribute to solving those problems. Focusing on these characteristics thus helps to explain how groups such as the GO Consortium and GARNet managed to evolve into well-recognized regulatory bodies.

The process of institutionalization at work in these groups relies heavily on widely entrenched forms of group socialization, which these organizations exploit in order to achieve two crucial and yet potentially contrasting goals. First, they maintain an enduring identity and some stability, which enables them to keep growing in scale, ambition, and visibility. Second, they retain the flexibility needed to fit the highly mutable and volatile nature of current research networks. The capacity to adapt to changes is crucial in the contemporary landscape of scientific funding, where intense competition for relatively small pots of money makes the majority of biological research dependent on collaborations around short-term projects. Collaborators, as well as the topics of interest, can and often do change radically from project to project. Scientists need to manage this environment in order to make interesting new links to people, fields, and topics, as well as maintain and develop existing interests and collaborations. Stability arises out of constant renewal; the necessity to enhance the robustness of social scaffolds in the face of environmental perturbations is one of the most fascinating aspects of these scientific initiatives.

A prime example showing how these groups depend on these forms of socialization is their reliance on charismatic individuals as group leaders who carry authority as well as recognition within the main communities of interest. In the case of GO, for instance, it is notable that the initial impetus toward the development of bio-ontologies was provided by key figures in model organism biology, such as Michael Ashburner, Suzanne Lewis, and Judith Blake, whose scientific authority among their peers was already established and well recognized. Building on existing credibility and reputation, these figures were able to attract the attention of their peers and the trust of funders, thereby creating a tidal wave of interest that culminated in the formation of a thriving community of developers and users of bio-ontologies. In the case of steering committees such as GARNet, we find similar dynamics; highly visible scientific figures in plant science and systems biology, such as Andrew Millar, loaned their credibility to the committee as it was being formed.

Given the responsibilities already weighing on the shoulders of these leading figures, individuals not well known for their scientific contributions, who nevertheless possessed the right set of competencies and skills, performed much of the actual legwork and coordination work. These individuals were willing to sacrifice time and resources toward making the enterprise successful at a time when resources allocated to the group were very scarce. In the case of both GO and GARNet, these turned out to be junior academics with broad-ranging scientific interests who were intrigued by the social organization of their communities. They had a strong drive toward promoting cooperative behaviors in science and often talked about the importance of "serving" the community of researchers by setting up useful data infrastructures. In some cases, family commitments made it difficult for these individuals to pursue a full-time career in research. Perhaps unsurprisingly, the majority of these individuals were women. Thus, to some extent, this embodied the well-established social configuration of womanhood as nurturing and service-oriented, providing colleagues and peers with trustworthy resources and highly skilled labor that did not fit the formal structures for scientific credit and measures of excellence.

Another form of socialization that features heavily in the history of these organizations (and also in the history of many social movements) is that of personal friendship. In both cases, the regulatory power of collective action was reinforced through informal networking, including late-night discussions, joint trips, and workshops involving a regular set of core attendees and the formation of strong personal bonds among some of them. This included a willingness to bring other friends and collaborators on board. These informal bonds became particularly important during times of trouble, when problems with the organization forced its members to regroup and rethink their strategies and general approach. GARNet faced such a moment at the end of its first ten years of funding, when it became apparent that its continuation would depend on its ability to (1) demonstrate the levels of support and appreciation for GARnet's work to the BBSRC and (2) formulate a vision for future work that embraced the whole of plant science, rather than only the *Arabidopsis* community, which tracked recent trends toward cross-species research (of the type that GARNet itself fostered, for instance, through helping to set up the UK Plant Science Federation). GARNet members appealed to prominent individuals in plant science with whom they had collaborated in the past and who were happy to testify to the usefulness of the organization and help articulate its vision for the next funding cycle.

These forms of socialization play the role of *core configurations,* which Linnda Caporael and colleagues (2014) have characterized as "subgroups of face-to-face interactions that are posited to recur in daily life, ontogeny, history, and plausibly, as part of human evolutionary history" (58). They can be identified and singled out on the basis of the specific functions they accomplish; indeed, their success in achieving a given purpose is what "explains their continued replication" (Caporael 1997, 282). Caporeal has focused on the size of groupings—the number of individuals involved—as a fundamental feature of core configurations, and my analysis of specific cases of collective agency in biology confirms her emphasis on groupings of a relatively small size, which enables strong personal relations and the ability to quickly reorganize in response to external challenges. Additionally, I have highlighted the distribution of the social roles required to spur a social movement to grow and become established, especially a scientific organization with these characteristics. Core configurations like personal friendship, by virtue of their proven track record in bringing and keeping individuals together, have become entrenched forms of socialization, which individuals fall back upon when attempting to achieve conceptual and institutional changes.[5] As such, these configurations provide stability and visibility to fledgling organizations, such as consortia and steering committees, while also enhancing their flexibility to changes in the environment. The result is robust social entities.

THE DISSEMINATION of scientific data relies on a great variety of material and social scaffolds, ranging from well-established institutions that determine data-sharing policies and related credit systems (funding agencies, policy bodies, academies, learned societies) to venues through which data can travel (annual conferences, data journals, repositories) and other types of organizations involved in the production and reuse of data (universities, networks). In this chapter I have considered two ways in which scientists have coordinated their actions and agendas to shape science governance and policy related to the means of data dissemination in biology. Both consortia and steering committees have played—and continue to play—crucial roles in supporting and structuring data curation practices, as well as making them visible and recognized by longstanding scientific institutions. In so doing, they have themselves acquired an institutional role and acted as key social scaffolds for the development and implementation of data-intensive biology. Looking at these organizations as social movements helps to identify some of the core strategies or configurations that helped to develop the ideas, values, and priorities of a few individual scientists on a large scale, thus shaping knowledge-making practices at the international level.

This analysis resonates with Wiebe Bijker's invitation to recognize that specific patterns of agency by groups of scientists play an important role in large technological systems (Bijker, Hughes, and Pinch 1987). It also shows why attention to social and institutional dynamics is critical to understanding scientific practices. Activities such as data sharing, data interpretation, publication patterns, the choice of topics for future research, and scientists' commitment to specific norms need to be analyzed with reference to their broad institutional and social contexts, especially in cases where scientists themselves play a key part in developing and shaping those contexts. In turn, social structures such as formal and informal committees and groups, often brought together by a common concern or goal, function as scaffolds for the development of new institutions (Gerson 2013; Wimsatt 2013; Caporeal et al. 2014). As illustrated by the speed with which both GARNet and GO have developed from a small group of scientists into large, influential organizations, an evaluation of the cultural role and impact of specific groups and associated norms and behaviors should take into account the highly dynamic context in which they operate. Different characteristics of social scaffolding help at different moments in the development of such institutions. For example, while imposing strong leadership may prove fatal when a feeling of community participation and engagement is required for social cohesion, it may well help when dynamics change, and social coordination is more effectively centered on the activities of a charismatic individual or subgroup. The same can be said for the extent to which norms of engagement are codified (e.g., participation in GARNet networks was voluntary but subject to specific rules of engagement—dictated by the broader funding structure through which it was supported—from the start), the choice to rely on given technologies versus attempting to develop new ones (GARNet drew its visibility from the former, while GO acquired social and political influence by virtue of the latter), and the choice to highlight existing "gaps" in governance versus trying to build new areas of influence (GO, notably, started with the former and ended up pursuing the latter).

In closing, it is critical to stress again that social scaffolds affect the production and transmission of knowledge through their tight interrelation with the development of material and infrastructural scaffolds. Indeed, the existence of organizations such as GARNet and GO has been strongly correlated with the development of computing facilities and data-extraction methods in molecular biology. The effective alignment of these material and social structures has made a significant difference to the methods and strategies for data production and interpretation currently in use within biology.[6] Philosophical research focused on the status of data in contemporary science, as well as the ways in which inferences are drawn and corroborated, needs to look beyond specific instances of data use and examine why certain configurations of norms, instruments, and methods become established and their implications for the development of knowledge-making practices.[7] The analysis herein points to an important direction for future work in the philosophy of science: the need to challenge the minimalist and asocial conceptualizations of scientific agency pervading much of contemporary philosophy. This type of work will help philosophers understand the material, social, conceptual, and institutional conditions for knowledge production as a necessarily interconnected and historically situated whole.

NOTES

Warm thanks to colleagues in GARNet and GO, Rachel Ankeny, Barry Barnes, Dario Castiglione, Elihu Gerson, James Griesemer, Alan Love, Bice

Maiguashca, Raffaele Marchetti, Hans Radder, Bill Wimsatt, and my colleagues at Egenis for helpful discussions on this topic. This research was funded by the European Research Council grant award 335925.

1. See the overviews of science and technology studies (STS) work on the social organization of scientific research provided in Bijker, Hughes, and Pinch (1987), Hackett et al. (2008), and Atkinson, Glasner, and Lock (2009).

2. For example, there are many STS analyses of standardization procedures, the role of standards as "coordination devices" for complex networks of actors (Latour 1987; Bowker and Star 1999), the relation between biomedical regulation and the production of "objective" knowledge (Cambrosio et al. 2009), and the way in which standards foster accountability and trust by facilitating the enactment of "rituals of verification" (Power 1997). The specific case of bioinformatic standards has also been subject to several studies (e.g., Hilgartner 1995, Bowker 2000, Hine 2006, Chow-White and Garcia-Sanchos 2012, Mackenzie 2012, and Lewis and Bartlett 2013, as well as my own work on the subject).

3. For analyses of the notions of scale at play in "big biology," see Davies, Frow, and Leonelli (2013).

4. For more detail on the ethos of model organism communities and the importance of repertoires in shaping research fields, see Ankeny and Leonelli (2011, 2016).

5. I am here thinking of simple entrenchment: "An evolving adaptive system with a recurring developmental trajectory, and differential entrenchment generating different degrees of evolutionary conservation" (Wimsatt 2013, 83).

6. For an expanded version of this argument, see Leonelli (2016).

7. To this aim, Rachel Ankeny and I have proposed to adopt the notion of *repertoires* as units of analysis for scientific organization and change over time (Ankeny and Leonelli 2016).

REFERENCES

Ankeny, Rachel A., and Sabina Leonelli. 2011. "What's So Special about Model Organisms?" *Studies in the History and the Philosophy of Science: Part A* 42 (2): 313–23.

Ankeny, Rachel A., and Sabina Leonelli. 2016. "Repertoires: A Post-Kuhnian Perspective on Scientific Change and Collaborative Research." *Studies in the History and the Philosophy of Science: Part A* 60:18–28.

Atkinson, Paul, Peter Glasner, and Margaret Lock, eds. 2009. *The Handbook for Genetics and Society.* Abingdon, UK: Routledge.

Beale, Michael, Paul Dupree, Kathryn Lilley, Jim Beynon, Martin Trick, Jonathan Clarke, Michael Bevan, Ian Bancroft, Jonathan Jones, Sean May, Karin van de Sande, and Ottoline Leyser. 2002. "GARNet, the Genomic Arabidopsis Resource Network." *Trends in Plant Science* 7 (4): 145–47.

Bijker, Wiebe E., Thomas P. Hughes, and Trevor Pinch, eds. 1987. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology.* Cambridge, Mass.: MIT Press.

Bowker, Geoffrey C. 2000. "Biodiversity Datadiversity." *Social Studies of Science* 30 (5): 643–83.

Bowker, Geoffrey C., and Susan L. Star. 1999. *Sorting Things Out: Classification and Its Consequences.* Cambridge, Mass.: MIT Press.

Callebaut, W. 2012. "Scientific Perspectivism: A Philosopher of Science's Response to the Challenge of Big Data Biology." *Studies in the History and the Philosophy of the Biological and Biomedical Sciences* 43 (1): 69–80.

Cambrosio, Alberto, Peter Keating, Thomas Schlich, and George Weisz. 2009. "Biomedical Conventions and Regulatory Objectivity: A Few Introductory Remarks." *Social Studies of Science* 39 (5): 651–64.

Caporael, Linnda. 1997. "The Evolution of Truly Social Cognition: The Core Configurations Model." *Personality and Social Psychology Review* 1 (4): 276–98.

Caporael, Linnda, J. R. Griesemer, and W. C. Wimsatt, eds. 2014. *Developing Scaffolds in Evolution, Culture, and Cognition.* Cambridge, Mass.: MIT Press.

Castells, Manuel. 1996. *The Rise of the Network Society.* Oxford: Blackwell.

Chow-White, P., and Miguel Garcia-Sanchos. 2012. "Bi-directional Shaping and Spaces of Convergence: Interactions between Biology and Computing from the First DNA Sequencers to Global Genome Databases." *Science, Technology, and Human Values* 37 (1): 124–64.

Davies, Gail, Emma Frow, and Sabina Leonelli. 2013. "Bigger, Faster, Better? Rhetorics and Practices of Large-Scale Research in Contemporary Bioscience." *BioSocieties* 8 (4): 386–96.

Della Porta, Donatella, and Mario Diani. 1999. *Social Movements: An Introduction.* Oxford: Blackwell.

Eyerman, Ron, and Andrew Jamison. 1991. *Social Movements: A Cognitive Approach.* University Park, Pa.: Penn State University Press.

Floridi, Luciano. 2013. *The Philosophy of Information.* Oxford: Oxford University Press.

Frickel, Scott, and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70 (2): 204–32.

Gerson, Elihu. 2013. "Integration of Specialties: An Institutional and Organizational View." *Studies in the History and the Philosophy of the Biological and Biomedical Sciences* 44:515–24.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1996. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Society.* London: Sage.

Hackett, Edward J., Olga Amsterdamska, Michael Lynch, and Judy Wajcman, eds. 2008. *The Handbook of Science and Technology Studies.* 3rd ed. Cambridge, Mass.: MIT Press.

Harvey, Mark, and Andrew McMeekin. 2007. *Public or Private Economies of Knowledge? Turbulence in the Biological Sciences.* Cheltenham, UK: Edward Edgar.

Hey, Tony, Stewart Tansley, and Kristine Tolle, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Redmond, Wash.: Microsoft Research.

Hilgartner, Stephen. 1995. "Biomolecular Databases: New Communication Regimes for Biology?" *Science Communication* 17 (2): 240–63.

Hine, Christine. 2006. "Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work." *Social Studies of Science* 36 (2): 269–98.

Jasanoff, Sheila, and Sang-Hyun Kim, eds. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power.* Chicago: University of Chicago Press.

Jones, Katherine Maxon, Rachel A. Ankeny, and Robert Cook-Deegan. 2018. "The Bermuda Triangle: The Pragmatics, Policies, and Principles for Data Sharing in the History of the Human Genome Project." *Journal of the History of Biology*, online first.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society.* Cambridge, Mass.: Harvard University Press.

Leonelli, Sabina. 2009. "Centralising Labels to Distribute Data: The Regulatory Role of Genomicconsortia." In *The Handbook of Genetics and Society: Mapping the New Genomic Era,* edited by Paul Atkinson, Peter Glasner, and Margaret Lock, 469–85. Abingdon, UK: Routledge.

Leonelli, Sabina. 2010. "Documenting the Emergence of Bio-ontologies: Or, Why Researching Bioinformatics Requires HPSSB." *History and Philosophy of the Life Sciences* 32 (1): 105–26.

Leonelli, Sabina. 2012. "Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies." *International Studies in the Philosophy of Science* 26 (1): 47–65.

Leonelli, Sabina. 2016. *Data-centric Biology: A Philosophical Study.* Chicago: Chicago University Press.

Leonelli, Sabina, Berris Charnley, Alex R. Webb, and Ruth Bastow. 2012. "Under One Leaf: An Historical Perspective on the UK Plant Science Federation." *New Phytologist* 195 (1): 10–13.

Lewis, Jamie, and Andrew Bartlett. 2013. "Inscribing a Discipline: Tensions in the Field of Bioinformatics." *New Genetics and Society* 32 (3): 243–63.

Mackenzie, Adrian. 2012. "More Parts than Elements: How Databases Multiply." *Environment and Planning D: Society and Space* 30 (2): 335–50.

Martin, Paul, Nik Brown, and Alison Kraft. 2008. "From Bedside to Bench? Communities of Promise, Translational Research, and the Making of Blood Stem Cells." *Science as Culture* 17 (1): 29–41.

Melucci, Alberto. 1996. *Challenging Codes: Collective Action in the Information Age.* Cambridge: Cambridge University Press.

O'Malley, Maureen A., and Orkun S. Soyer. 2012. "The Roles of Integration in Molecular Systems Biology." *Studies in the History and the Philosophy of the Biological and Biomedical Sciences* 43 (1): 58–68.

Power, Michael. 1997. *The Audit Society: Rituals of Verification.* Oxford: Oxford University Press.

Ratti, Emanuele. 2015. "Big Data Biology: Between Eliminative Inferences and Exploratory Experiments." *Philosophy of Science* 82 (2): 198–218.

Research Data Alliance. 2016. "About Us." Accessed November 30, 2016. https://www.rd-alliance.org/about-rda.

Southan, Christopher, and Graham Cameron. 2009. "Beyond the Tsunami: Developing the Infrastructure to Deal with Life Sciences Data." In *The Fourth Paradigm: Data-Intensive Scientific Discovery,* edited by Tony Hey, Stewart Tansley, and Kristin M. Tolle, 117–23. Redmond, Wash.: Microsoft Research.

Sunder Rajan, Kaushik. 2006. *Biocapital: The Constitution of Postgenomic Life.* Durham, N.C.: Duke University Press.

Tarrow, Sidney G. 1998. *Power in Movement: Social Movements and Contentious Politics.* Cambridge: Cambridge University Press.

Wimsatt, William C. 2013. "Articulating Babel: An Approach to Cultural Evolution." *Studies in the History and the Philosophy of the Biological and Biomedical Sciences* 44:563–71.