

*Manuscript accepted for publication in Bioscience, March 2015*

## **Repertoires: How to Transform a Project into a Research Community**

### *Authors*

Sabina Leonelli, Department of Sociology, Philosophy and Anthropology & Exeter Centre for the Study of the Life Sciences (Egenis), University of Exeter

[s.leonelli@exeter.ac.uk](mailto:s.leonelli@exeter.ac.uk) @sabinaleonelli

Rachel A. Ankeny, School of Humanities, University of Adelaide, Australia

[rachel.ankeney@adelaide.edu.au](mailto:rachel.ankeney@adelaide.edu.au)

### *Abstract*

How effectively communities of scientists come together and co-operate is crucial both to the quality of research outputs and to the extent to which such outputs integrate insights, data and methods from a variety of fields, laboratories and locations around the globe. This essay focuses on the ensemble of material and social conditions that makes it possible for a short-term collaboration, set up to accomplish a specific task, to give rise to relatively stable communities of researchers. We refer to these distinctive features as *repertoires*, and investigate their development and implementation across three examples of collaborative research in the life sciences. We conclude that whether a particular project ends up fostering the emergence of a resilient research community is partly determined by the degree of attention and care devoted by researchers to material and social elements beyond the specific research questions under consideration.

### *Keywords*

Community building; scientific epistemology; data; scientific methods; scientific norms.

### *Acknowledgments*

Funding for this work was provided by the University of Exeter (through support for RAA visiting position at the Exeter Centre for the Study of the Life Sciences) and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 335925 (project "The Epistemology of Data-Intensive Science"). Many thanks to Tim Beardsley, Alan Love and five anonymous referees for helpful comments.

## 1. Introduction

Much work within the history, philosophy and sociology of science has focused on the ways in which scientific collaborations are created and their importance for knowledge development (e.g., Griesemer and Gerson 1993, Hackett 2005, Shrum et al. 2007, Gorman ed. 2010). It is generally acknowledged that, given the international and highly collaborative nature of contemporary biological research, and the interdisciplinary exchanges required to make sense of complex processes (ranging from development to pathogenesis and carcinogenesis), the ways in which communities are built within contemporary life science have a major impact on the quality and type of outputs produced. The ability to create research communities also is crucial to the achievement of integration within biology, whether such integration concerns data, methods, models, insights, disciplines or locations (O'Malley and Soyer 2012, Bechtel 2013, Brigandt 2013, Leonelli 2013a, Plutynski 2013, Vermeulen 2013, Vermeulen et al. 2013). Scholars both within biology and in science studies have pointed to the scale of collaborations as a crucial factor in influencing how research is organised: the sheer number of researchers involved in any one project matters greatly to the ways in which the project is managed, particularly in those cases where researchers have diverse expertise (Lewis and Bartlett 2013) and are based in different geographical locations (Parker et al. 2010, Hilgartner 2013, Davies et al. 2013).

This paper expands on these claims to discuss the material and social conditions under which research communities are not only created, but actually *managed* and *persisting* in the long term. Our focus is therefore on the resilience of scientific collaborations: the conditions under which they can endure and even thrive, despite the high volatility of the research environment in which intellectual priorities, material constraints and funding goals tend to shift very rapidly. In order to explore how communities acquire resilience, and yet retain the flexibility needed to adapt to changing research needs, we examine the evolution of projects that result in active and productive scientific communities. We argue that small, temporary research groups can and sometimes do provide the foundations for building larger, more enduring communities, provided that they develop what we call a 'repertoire': a distinctive and shared ensemble of elements that make it practically possible for individuals to co-operate, including norms for what counts as acceptable behaviours and practices together with infrastructures, procedures and resources that make it possible to implement such norms.

We survey the development and implementation of repertoires in three recent research projects that have played key roles in fostering the emergence of internationally recognised research communities. While we acknowledge that many other factors have contributed to the rise and success of these communities, our account highlights how the researchers involved in the initial projects chose to put the development and maintenance of repertoires at the centre of their project planning and research work from the very outset. We also briefly contrast some related, highly successful short-term research projects that did *not* result in the building of distinct and long-lived research communities, in part because of the absence of the development of repertoires to ground any larger communities. We argue that whether a short-term project results in fostering the emergence of a research community is not only determined by the timeliness and promise of the research questions being asked, or by the technologies utilized, but also by the degree of attention and care devoted by researchers to the material and social infrastructures required to address those questions. We conclude by

reflecting on a key issue concerning current methods, norms and infrastructures in biology: to which extent, and under which conditions, does repertoire-building facilitate or hinder increased integration among biological subdisciplines and approaches?

## **2. Repertoires in Contemporary Life Science**

It is necessary to begin by clarifying why we want to use the notion of 'repertoire' to describe particular types of research norms and infrastructures. We understand a repertoire to be a stock of skills, behaviors, methods, materials, resources and infrastructures that a group habitually uses to conduct research and train newcomers who want to join the group. Indeed, the development of a repertoire is strongly tied to the establishment of a group identity in the first place, and parallels can be drawn between our discussion of repertoires and the sociological literature on the role of communities in field and discipline-formation (e.g., Ben-David and Collins 1966, Griffith and Mullins 1972, Mullins 1972, Parker and Hackett 2012, Gerson 2013). In particular, we are using the notion of community to identify a group of individuals brought together by repeated interactions around one or more goals, which can range from the pursuit of a given interest to the production of a tool, the development of a procedure and/or the use of a common space (whether physical or intellectual). We use 'repertoire' in a different sense than the treatment proposed by G. Nigel Gilbert and Michael Mulkey (1984). They identify two major interpretative repertoires (also termed 'linguistic registers') which occur frequently in scientific discourse and analyze how these are employed to account for error and belief. By contrast, we are not primarily or solely focused on scientific discourse.

An example of a repertoire in our sense is the set of resources, institutions and expertise that have come to define research work in contemporary systems biology, which include mathematical skills, knowledge of molecular biology, centers and funding dedicated to systems biology, and a social commitment to openness in research, expressed for instance in developing and contributing to 'omics' databases (e.g., KEGG, Kanehisa et al. 2012) and building opportunities to share and debate models (e.g., the model repository BioModels; Chelliah et al. 2013). This repertoire is now widely recognized as characteristic of systems biology work, and is often used to identify membership in what is otherwise an extremely varied community with no obvious shared theoretical commitments (Calvert 2010). For instance, a molecular biologist who has an interest in systems, but has no mathematical skills and does not collaborate with people who do, is not generally regarded as someone who is directly engaged in systems biology. A modeler who does not cooperate with peers and share data, models and expertise, and who does not work with molecular data, also would be viewed as an outsider.

Those working in established scientific communities have fairly settled and shared repertoires that arise out of their scientific training, institutional memberships, funding sources and experiences in labs or other work sites. Much like the idea of 'paradigm' put forward by Thomas Kuhn (1962), repertoires thus include material, social, and conceptual components, such as targeted venues and data infrastructures, shared theoretical commitments and a common pool of instruments and materials. Repertoires are particularly close to Kuhn's notion of 'exemplar', which he characterised as the knowledge, methods and assumptions used to address questions within any given research paradigm (Kuhn 1962). Also, similarly to Kuhn and Knorr-Cetina, we reject the

characterisation of research communities as focused largely on shared theories as constitutive of a discipline or field (e.g., Toulmin 1972, Darden and Maull 1977, Shapere 1977). While theoretical insights and disagreements often have important roles to play, they are by no means the only rallying points when a community is being developed. In fact, some communities (such as those in systems biology and the model organism case discussed below) are created in the absence of common theories, which enables groups of researchers to exploit the same instruments, resources, and infrastructures to explore a wide variety of perspectives and ideas.

At the same time, repertoires include elements that Kuhn (and Mulkey and Gilbert 1984) did not explicitly consider, namely the social and institutional resources and infrastructures that are critical to contemporary biology, such as databases, scientific committees, learned societies, modes of funding (and related commitments), and activities such as sequencing and phenotyping that are simultaneously conceptual, performative and material. In this way, our approach follows Karen Knorr-Cetina's suggestion that "the social is not merely 'also there' in science [...] it is capitalized upon and upgraded to become an instrument of scientific work" (Knorr-Cetina 1999, 29). Also in contrast to paradigms, repertoires include procedures and norms specifically aimed at stimulating institutional and financial support, such as promissory discourse and marketing strategies designed to increase the funding appeal of specific projects; they are permeable and mutable entities, which are constantly adapted to the broader research and funding environment (indeed, they owe much of their resilience to this flexibility); and, much like the 'thought collectives' discussed by Ludwig Fleck (1979 [1935]), they do not preclude their users from taking advantage of several repertoires at once. As we illustrate below, scientists typically need familiarity and engagement with more than one repertoire, in order to maximise their chances of funding as well as to enhance the visibility and impact of their research. Finally, repertoires are clearly performative, and thus we have selected this terminology because of its resonance with its usage in non-scientific fields such as music, where the notion of a repertoire is well-established (e.g., Faulkner and Becker 2009).

Repertoires within research communities are relatively stable and often complex, ensuring the basis for longer-term collaborations within groups. The extent to which these repertoires create opportunities for collaborations beyond the community that adopts and develops them is less clear. It is certainly the case that repertoires can be mobilized and redeployed by individuals or groups in a variety of ways to serve numerous purposes. The flexibility of a repertoire and its adaptability to new research questions and circumstances is critical for its usefulness, particularly for the purposes of invention and discovery.

To better understand the features and significance of repertoires, consider the role of repertoires in short-term projects. These projects form the basis for the vast majority of funding allocated by governmental agencies to scientific research, and thus constitute a large proportion of research work carried out in the biosciences. Their length varies between one to five years, and they are typically geared to the exploration of a specific research hypothesis or at achieving a specific and delimited scientific milestone by a team of researchers. The team usually includes individuals trained in a variety of fields, whose joint expertise is viewed to be ideally suited to tackling the question or goal at hand, but who may not have previously worked together. Indeed, short-term projects are sometimes used as a way to forge new interdisciplinary links and bring new

methodological or conceptual tools to the study of a given problem. Thus short-term projects typically involve efforts geared to making individuals with different backgrounds and interests work harmoniously towards the same goal. What interests us here is the fact that these efforts *infrequently* give rise to a new repertoire. Rather, participants in these sorts of projects often fall back into using existing and familiar repertoires (sometimes hybridizing several in a somewhat inconsistent manner), or succeed in achieving their goals without establishing the grounds for a novel, resilient and shared repertoire that can support or promote an ongoing research community.

We now reflect on the characteristics of scientific projects that have succeeded not only in fulfilling their research goals, but also in establishing a stable research community (and even, sometimes, a new field). In each case, we observe that a key move in this process was the development of a resilient repertoire. Thus these case studies highlight some of the conditions under which a repertoire can serve as the basis for the establishment and ongoing productivity of a research community.

### **3 – Building Repertoires**

#### *Case 1: From Sequencing Projects to Biocuration*

Bio-ontologies are an achievement of the bioinformatic efforts directed at an efficient organization and distribution of data produced by genomic research. They provide a framework through which heterogeneous sets of biological data can be classified, stored and retrieved through freely available, online databases (Rubin et al. 2008). For the purposes of this case study, we focus on the bio-ontologies collected by the Open Biomedical Ontologies Consortium (<http://www.obofoundry.org>), an organization founded to facilitate communication and coherence among bio-ontologies with broadly similar characteristics (Ashburner et al. 2003), and particularly the Gene Ontology (GO), which is widely regarded as the most successful case of bio-ontology construction to date and used as a template for several other prominent bio-ontologies (Ashburner et al. 2003, Brazma et al. 2006).

Bio-ontology terms behave in similar ways to other classificatory categories: they stabilize objects of knowledge in ways that enable, but at the same time constrain, future research. The knowledge captured by bio-ontologies is bound to change with further research, and they manifest themselves differently in each research context. Resolving the tension between stability and flexibility of classificatory categories is crucial to the success of bio-ontologies and is a core responsibility of curators, who engage in adapting and updating bio-ontologies so that they mirror the research practices and knowledge of their users. Thus, curators mediate between the diverse assumptions and practices characterizing the work of bio-ontology users and the need for bio-ontologies to conform to universal requirements such as consistency, computability, ease of use and wide intelligibility. Curators' interventions are crucial to the effective functioning of bio-ontologies, and ideally need to be informed by a wide range of expertise, including IT and programming skills, training in more than one biological discipline (allowing them to bridge between different scientific contexts) and familiarity with experimentation at the bench (so that they understand observational statements made in the context of specific experimental settings, as well as anticipating the expectations of the users of the bio-ontologies).

The group associated with what is now known as the GO Consortium began as a group of outsiders, motivated by their unhappiness with how data were organized in databases, and determined 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, fighting to establish a biology-driven bioinformatics. In 2000, funding for their efforts started to trickle in, and they found themselves in a position to recruit more like-minded biologists and bioinformaticians from other model organism communities such as the plant community formed around *Arabidopsis thaliana*. The group is now substantially larger, including a head office based at the European Bioinformatics Institute in the UK and at least 30 affiliated bioinformaticians spread in model organism communities around the world, which arguably constitute a scientific community. The Consortium has become a model for how biological data infrastructure should work and what it should look like, and it has been increasingly institutionalized, for instance as part of the National Centre for Biomedical Ontology in the United States. Nevertheless it continues to rely on the funding provided by each participating model organism community, derived from short-term governmental grants whose renewal depends on performance, and on work done by participants who are committed to the usefulness and importance of the Gene Ontology as a biological resource.

The GO Consortium serves as a powerful centralizing force within model organism biology (Leonelli 2009): it regulates what counts as professional training for curators; it enforces common values such as open access to data, inter-community co-operation and diversity in epistemic practices across biology; it fosters common goals, such as a desire to pursue comparative and integrative biology; it channels and reinforces the support of specific funding sources, which in turn strengthens the commitments of all participants to a fair and equitable contribution; and it establishes the 'rules of the game' by establishing common procedures and technologies through which users can interact among each other and upload, retrieve and analyze data. It was initially founded with the intent that it would function in the long-term, which marks a difference from the other cases explored here. Nevertheless, all of the attributes described above contributed to the establishment of a shared repertoire and the building and persistence of a research community over the past fifteen years, and has led to the GO Consortium being viewed as an agent of change within the biological community (cf. Hine 2006).

### *Case 2: From Simple Organisms to Model Organism Research*

Model organisms are relatively low-cost, low-maintenance research materials that are easy to control and on which a substantial body of knowledge can rapidly be accumulated, since repeated use of and reference to the same organism provides a great opportunity for sharing knowledge across a vast constellation of biological disciplines, groups and research schools (Ankeny and Leonelli 2011). Indeed, some organisms have become important platforms for interdisciplinary collaboration across research programs in fields as diverse as molecular biology, physiology, development, reproduction (Friese and Clarke 2012) and even ecology (Bevan and Walsh 2004). Classic examples include the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, the zebrafish *Danio rerio*, the budding yeast *Saccharomyces cerevisiae*, the weed *Arabidopsis thaliana* and the house mouse *Mus musculus* (NIH website 2010).

What are now known as 'model organisms' began simply as experimental organisms that came to be utilized for research within genetics and developmental biology. Some had long histories within a variety of branches of the life sciences, while others were specifically chosen because of their potential for pursuit of multiple levels of organization within the organism. One key example of an organism that fits the latter model, and which also illustrates how research projects can evolve if a repertoire comes to be shared, is the nematode *C. elegans*. The 'worm,' as it is commonly known, became the focus of investigation in the late 1960s by a research group at Cambridge (de Chadarevian 1998, Ankeny 2001). Although working under the auspices of the Laboratory of Molecular Biology (LMB), the scientists involved in this project came from a range of training backgrounds, including developmental biology, genetics, biochemistry, information technology, medical research and neurobiology, to name a few of the fields. The project initially focused on producing complete developmental lineage maps as well as a catalogue of genetic mutations, but also came to serve as the basis of a growing community focused on this organism with a set of shared goals and understandings about preferred methods for doing biological work. This background, together with what came to be a well-established community, laid the groundwork for efficient use of this organism as one of the first foci of the massive mapping and sequencing projects within the Human Genome Project (HGP).

This community is a clear case of the building of a repertoire that allowed a research community to persist beyond the completion of a specific project (in the first instance the LMB-initiated project, and later the sequencing efforts), and without wedding it to a particular subfield within biology. This repertoire included the very concept of a 'model organism'; the know-how, expertise, protocols, instrumentation and data accumulated by participating scientists; long-term, blue-skies funding support particularly from the US and UK governments, which attracted participants to the community and enabled its development in relatively well-resourced conditions; and an ethos of sharing data and techniques prior to publication, all of which contributed to the continuity of the research efforts and their abilities to build over time. In addition, the production, use and dissemination of the actual specimens of these organisms was increasingly standardized and centralized through the establishment of stock centers. Finally, the establishment of a range of infrastructures including databases to gather both published and unpublished data in a standardized manner has provided essential contributions to the community that has resulted.

The mouse presents a clear contrast case: although undeniably a vital contributor to the sequencing projects and to a range of biomedical efforts over the 20th century (Rader 2004, Lewis et al. 2013), those who work on the mouse as an experimental organism come from a wide range of disciplinary backgrounds, interests and overarching goals, and sources and modes of funding. For instance, the vast majority of research on mice takes place in private rather than public facilities, with accountabilities both to specific companies and to the production of knowledge for use by society at large. There are very diverse values associated with the goals underlying research such as those related to purer biological research versus those that underlie medical research particularly in conjunction with pharmaceutical testing and other more commercialized endeavors (Davies 2013). As a result (and despite attempts in that direction), there are no centralized stock centers for mice strains. Specimens are not always shared across laboratories, and when they are, the transaction is typically costly, thus limiting access to those who have the financial resources to pay for them. Although many scientists

working on the mouse worked together during the mouse genomic sequencing that was part of the HGP and developed some shared databases and other resources in this process, the results were directly related to the project at hand and did not generate a broader community that continued to work together in any large numbers after the conclusion of the sequencing projects. We claim that this outcome is related to the lack of generation of a repertoire: there did not come to be a series of shared practices or aims among the diverse groups that worked on the mouse genome, nor concepts, protocols, institutions, shared financial resources or other components of a repertoire that could serve to unify these disparate groups. Various groups went back to their previous methods for doing scientific work, for instance in relation to studies of alcoholism (see Ankeny et al. 2014).

### *Case 3: From metagenomic sequencing to the microbiomes*

As soon as the costs and labor involved in genome sequencing started to drop in the early 2000s, meta-genomic sampling became a popular source of projects across the life sciences. The opportunity to sequence many organisms within a short timeframe made it possible to sample and investigate microbial life forms, which in turn created opportunities for shifts in the very conceptualization of organisms (Dupré and O'Malley 2009, O'Malley 2014). Among the neologisms associated to the practice of metagenomics, perhaps the most prominent is the idea of 'microbiome,' which emerged in the early 2000s in association with the Human Microbiome Project and similarly human-directed initiatives (such as the Gut Microbiome Project). Despite its multiple interpretations (Huss 2014), the idea of the microbiome was eagerly adopted and used as a banner by a vast variety of biological initiatives, all eager to tap into the increasing funding allocated to such efforts and the research opportunities afforded by the associated technologies. Examples of such projects are the Earth Microbiome Project, investigating variation of ecosystem niche structures at biogeochemical scales; the American Gut project, which uses crowdsourcing as a means of collecting data about the microbes populating the guts of American citizens; the Soil Microbiome, examining the microbial diversity of pre-agricultural prairie soils in the USA; the Home Microbiome Study, looking at the association between microbes of families and their homes; and the Hospital Microbiome, looking at hospital environments during construction and after opening.

Projects such as these are typically associated with the following set of features: they are all funded by large governmental grants from the National Science Foundation, the National Institutes of Health and other agencies in the US and Europe; they engage (some more successfully than others) in international standardization efforts for the types of data, technologies and software that they use, such as the Minimal Information Standards which attempt to regulate the format of data files produced to facilitate cross-project integration and comparison (e.g., the .biome data file); they re-purpose widely used technologies such as sequencing towards new intellectual goals, taking particular advantage of the increasing speed and decreasing costs with which sequencing data can be obtained; they operate on a very large scale, relying on vast samples of data acquired via metagenomic investigations of several microbial populations, taken at different times over the same or comparable areas (thus generating so-called 'big data'); they take an ecological approach via by conceptualizing organisms (such as humans) as well as eco-systems as multi-species environments with unique 'microbial footprints'; and they make extensive use of new social media and crowdsourcing opportunities, such as



those offered by Twitter and websites, to enhance their public profile and attract volunteers in order to collect samples and help analyze results.

Given this success and their relatively cohesive features, we propose that microbiome projects have come to constitute yet another example in contemporary biology of a well-functioning repertoire which is successfully redeployed in a multiplicity of different domains. A key motive in this story is money, particularly its sources and how these shape the repertoire (as contrasted simply to the funding of a project or similar). Inertia is created by the public relations trappings which in turn allow the repertoire's application on a mass scale and its redeployment. Repertoires also have life cycles that can vary widely: the microbiome and bio-ontologies examples illustrate the power of a repertoire over a relatively short span of time, while the model organism example instantiates a particularly durable and resilient repertoire.

#### **4 - Conclusions: When Are Repertoires Useful?**

A repertoire clearly differs from mere methods or technologies (such as sequencing, which is very widely utilized within different scientific contexts, and hence does not constitute a repertoire in its own right) and fields defined more narrowly for instance through institutions or theoretical commitments. The idea of a repertoire captures what happens when specific projects become 'blueprints' for the way in which whole communities should do science, including the complex procedures developed to ensure the long-term maintenance of material infrastructures and the continuation of the required financial and institutional support. In most cases, short-term collaborative projects do not result in repertoires; for instance they may reveal fundamentally different commitments, incompatible work practices or fail to secure longer-term resources. This process is normal and even necessary, as participation in projects does not always result in substantial shifts in researchers' habits and collaborations: first, such shifts are not necessary for the production of significant scientific contributions; and second, even in cases where such shifts could be helpful, a tenured principal investigator typically must be working on several different projects at any one time, and cannot devote the same amounts of time and attention to all of them. Thus, we are not advocating that all projects *should* result in resilient repertoires and communities, or that those projects that do are of higher quality or in any sense better than those that do not; instead our focus is in outlining characteristics that seem to be shared by those that *do* evolve in this manner.

We have briefly outlined a series of mini-examples where the building of repertoires has allowed short-term and smaller-scale projects to transform into ongoing, productive, and resilient research communities. It is clear that the type of community involved and its history, and the way it is run and coordinated, have significant influences on research practices and outcomes. The building of repertoires is an iterative process that proceeds in parallel manner to the development of communities that are committed to using them. This process warrants a longer analysis in order to articulate the key features that contribute to the type of repertoire that results in a successful community, as well as to identify and contrast different types of repertoires. Yet even our brief discussion shows that one obvious benefit of repertoires is their flexibility: they can be used across multiple branches of biology, and often are deliberately constructed so as to avoid committing to any specific subfield and hence are structured to exploit interdisciplinarity.

At the same time, the adoption of repertoires unavoidably creates strong commitments to particular techniques, assumptions, values, institutions, funding sources and methods, which although initially productive can sometimes act as constraints to future integration and innovation. The use of microarrays produced through Affimetrix technology provides an excellent example of these tendencies. In the late 1990s, Affimetrix became the main provider of DNA microarrays, though the patenting and commercialisation of their GeneChips tools. These tools arguably became part of the established repertoire for genome-wide studies, as they enabled researchers to rapidly produce results in standard formats, thus guaranteeing comparability and the implementation of community guidelines for data annotation such as the Minimal Information About Microarray Experiments (Brazma et al. 2001, Rogers and Cambrosio 2007). This dependence became problematic when other companies started to produce competitive and arguably better arrays through different technological platforms, and the research community had to negotiate a transition from the accepted standard to a wider variety of approaches (for an historical example of a similar process, see Anorova et al. 2010 on failures of big data biology in 1960s).

The development of infrastructures and related community norms, such as databases and guidelines on data sharing, often includes attempts to be versatile vis-à-vis existing repertoires, because these structures need to be utilized by a variety of epistemic cultures in order to be used efficiently and successfully (Leonelli 2013b). However, the fact that mass-produced instruments for data production (such as mass spectrometers and microarray chips) are engineered and implemented on a wide scale channels research in a particular direction. It tends to canalize research towards the production and dissemination of very specific data types, especially in cases where data are generated primarily because researchers have the right instruments to do the work quickly and cheaply, rather than generating data to answer specific questions. In these types of cases, the resulting data can flood the research landscape in a disproportionate manner and sometimes without quality checks, and hence have considerable negative consequences. This in turn creates incentives to keep exploring these types of data rather than creating data in more deliberate ways in response to specific projects, which might be seen as conservative strategy and ultimately problematic for scientific discovery.

Similar issues emerge in the case of ontologies used to order and retrieve data within databases, where the need to produce standards of wide usability is hard to accommodate given the wide diversity and dynamism characterising the research projects in which data are produced and re-used. More exploration is needed of cases where repertoires create opportunities for wider collaboration, or in fact constrain such collaborations. Biology (and perhaps most, if not all, scientific research) has always been characterised by tensions between standardization and innovation, conservatism and novelty, and consensus and dissent around scientific norms. The establishment of repertoires *may be* one *keyway* to cope with these tensions, as they allow assembly of a set of tools and methods on which a community can build further research (until of course these tools eventually become obsolete or inadequate). The important lesson to be drawn by considering the development and role of repertoires in the contemporary biosciences relates to the importance of scientific methods and infrastructures dedicated to community building, which nevertheless are mutable and evolve over time. Integration within biology cannot happen unless at least some researchers invest considerable time and effort in building resources and settings in which they can be

deployed by a wide variety of participants, including those whose contributions could not have been anticipated. Indeed, we have shown that while repertoires initially set up to serve short-term projects may well end up supporting a large community of scientists over a long period of time, when the researchers involved choose to put the development and maintenance of repertoires at the centre of biological discussions from the very outset of a new project. Whether a single project ends up fostering the emergence of a research community (and eventually a repertoire) is partly determined by the degree of attention and care devoted by researchers to material and social elements beyond the specific research questions under consideration.

We propose that this way of analyzing the practice of science opens up a new methodological approach for the doing of philosophy of science, inasmuch as the terminology of repertoires allows us to better understand the relation between individual contributions and collective practices and norms, and also to consider the research practices and behaviors related to policy, finance, ethics, norms, public relations and/marketing and institutions, thus facilitating a more comprehensive and thus accurate view of the drivers of scientific change. The political economy of science becomes central to this story – rather than viewing it as a mere ‘externality’, we return it to its critical place as something strongly relevant to the epistemology of science and what actually works and serve as role models in science, in terms of questions to be asked, methods to be used, norms to be adopted and communities to be supported.

The development of a repertoire is an important moment in the growth of a scientific community, in which key goals and values come to be explicitly articulated and efforts are aimed at making it feasible to achieve these goals, often through the inclusion of new groups and approaches. The material and social structures implemented through such efforts undoubtedly create constraints for future research, but perhaps more importantly, they also constitute a major platform for integrative research, as long as the researchers involved remain aware of the need to continuously reflect on and revise their practices, and to recognize and welcome challenges.

## References

Ankeny RA. 2001. The natural history of *C. elegans* research. *Nature Reviews Genetics* 2: 474–478.

Ankeny RA, Leonelli S. 2011. What’s so special about model organisms? *Studies in History and Philosophy of Science* 42: 313–323.

Ankeny RA, Leonelli S, Nelson NC, Ramsden E. 2014. Making organisms model humans: Situated models in alcohol research. *Science in Context* 47: 27: 485–509.

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 Program, 1957–Present. *Historical Studies in the Natural Sciences* 40: 183–224.

Ashburner M, Mungall CJ, Lewis SE. 2003. Ontologies for biologists: A community model for the annotation of genomic data. *Cold Spring Harbor Symposia on Quantitative Biology* 68: 227–236.

- Bechtel W. 2013. From molecules to behavior and the clinic: Integration in chronobiology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 493–502.
- Bevan M, Walsh S. 2004. Positioning Arabidopsis in plant biology: A key step toward unification of plant research. *Plant Physiology* 135: 602–606.
- Brazma A, Krestyaninova M, Sarkans U. 2006. Standards for systems biology. *Nature Reviews Genetics* 7: 593–605.
- Brazma A et al. 2001. Minimum information about microarray experiments (MIAME)–towards standards for microarray data. *Nature Genetics* 29: 365–371.
- Brigandt I. 2013. Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 477–492.
- Calvert J. 2010. Systems biology, interdisciplinarity and disciplinary identity. In Parker J, Vermeulen N, Penders B (eds). *Collaboration in the New Life Sciences*. Farnham: Ashgate, pp. 201–218.
- de Chadarevian S. 1998. Of worms and programmes: *Caenorhabditis elegans* and the study of development. *Studies in History and Philosophy of Biological and Biomedical Sciences* 29: 81–105.
- Chelliah V, Laibe C, Le Novère N. 2013. BioModels Database: A repository of mathematical models of biological processes. *Methods in Molecular Biology* 1021: 189–199.
- Darden L, Maull N. 1977. Interfield theories. *Philosophy of Science* 44: 43–64.
- Davies G, Frow E, Leonelli S. 2013. Bigger, faster, better? Rhetorics and practices of large-scale research in contemporary bioscience. *BioSocieties* 8: 386–396.
- Davies G. 2013. Arguably big biology: Sociology, spatiality and the knockout mouse project. *Biosocieties* 8: 417–431.
- Dupré J, O'Malley MA. 2009. Varieties of living things: Life at the intersection of lineage and metabolism. *Philosophy & Theory in Biology* 1.
- Faulkner RR, Becker HS. 2009. *Do You Know—?: The Jazz Repertoire in Action*. Chicago: University of Chicago Press.
- Fleck L. 1979 [1935]. *The Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.
- Friese C, Clarke AE. 2012. Transposing bodies of knowledge and technique: Animal models at work in the reproductive sciences. *Social Studies of Science* 42: 31–52.
- Gerson EM. 2013. Integration of specialties: An institutional and organizational view. *Studies in History & Philosophy of Biological and Biomedical Sciences* 44: 515–524.
- Gilbert GN, Mulkay M. 1984. *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse*. Cambridge: Cambridge University Press.

Gorman ME (ed). 2010. Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration. London: MIT Press.

Griesemer J, Gerson E. 1993. Collaboration in the Museum of Vertebrate Zoology. *Journal of the History of Biology* 26: 185–203.

Griffith B, Mullins NC. 1972. Coherent groups in scientific change: 'Invisible colleges' may be consistent throughout science. *Science* 177: 959–964.

Hackett EJ. 2005. Introduction to the special guest-edited issue on scientific collaboration. *Social Studies of Science* 35: 667–672.

Hilgartner S. 2013. Constituting large-scale biology: Building a regime of governance in the early years of the Human Genome Project. *Biosocieties* 8: 397–416.

Hine C. 2006. Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science* 36: 269–298.

Huss J. 2014. Methodology and ontology in microbiome research. *Acta Biotheoretica* 9: 392–400.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Research* 40: D109–D114.

Knorr-Cetina K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge: Harvard University Press.

Kuhn TS. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

Leonelli S. 2009. Centralising labels to distribute data: The regulatory role of genomic consortia. In Atkinson P, Glasner P, and Lock M (eds), *The Handbook for Genetics and Society: Mapping the New Genomic Era*. London: Routledge, pp. 469–485.

Leonelli S. 2013a. Integrating data to acquire new knowledge: Three modes of integration in plant science. *Studies in History and Philosophy of Science Part C* 4: 503–514.

Leonelli S. 2013b. Global data for local science. *Biosocieties* 8: 449–465.

Lewis J, Atkinson P, Harrington J, Featherstone K. 2013. Representation and practical accomplishment in the laboratory: When is an animal model *good-enough*? *Sociology* 47: 776–792.

Lewis J, Bartlett A (2013) Inscribing a discipline: tensions in the field of bioinformatics. *New Genetics and Society* 32(3): 243–263.

Mullins N. 1972. The development of a scientific specialty: Phage group and the origins of molecular biology. *Minerva* 10: 51–82.

National Institutes of Health (NIH) 2010. Model organisms for biomedical research. (Accessed 20 March 2015; <http://www.nih.gov/science/models/>)

O'Malley MA, Soyer OS. 2012. The roles of integration in molecular systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 58–68.

O'Malley M. 2014. *Philosophy of Microbiology*. Cambridge University Press.

Parker JN, Vermeulen N, Penders B (eds). 2010. *Collaboration in the New Life Sciences*. Surrey: Ashgate.

Plutynski A. 2013. Cancer and the goals of integration. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 466–476.

Rader KA. 2004. *Making Mice: Standardizing Animals for American Biomedical Research, 1900–1955*. Princeton: Princeton University Press.

Rogers S, Cambrosio A. 2007. Making a new technology work: The standardization and regulation of microarrays. *Yale Journal of Biology and Medicine* 80: 165–178.

Rubin DL, Shah NH, Noy NF. 2008. Biomedical ontologies: A functional perspective. *Briefings in Bioinformatics* 9: 75–90.

Shrum W, Genuth J, Chompalov I. 2007. *Structures of Scientific Collaboration*. Cambridge: MIT Press.

Shapere D. 1977. Scientific Theories and Their Domains. In Suppe F (ed). *The Structure of Scientific Theories*. Urbana: University of Illinois Press, pp. 518–565.

Toulmin S. 1972. *Human Understanding: The Collective Use and Evolution of Concepts*. Princeton: Princeton University Press.

Vermeulen N. 2013. From Darwin to the census of marine life: Marine biology as big science. *PLoS One* 8: e54284.

Vermeulen N, Parker JN, Penders B. 2013. Understanding life together: A history of collaboration in biology. *Endeavour* 37: 162–171.