

regularly performs the experiment. This is so, even if both labs are recognized experts.

If the question of reproducibility persists once the requisite expertise is established, the answers often reside in subtle differences in methods, cell-line properties or reagents that become apparent only after scrutiny. Two acclaimed cancer researchers required more than a year to harmonize techniques to get similar measures for experiments; success depended on cross-country visits and on reconciling minor differences in how cells were prepared³.

If these steps do not work, we must consider whether the original result was really correct. And we must be prepared to accept the brutal facts, make the required corrections and move on. Great scientists are always willing to embrace the truth with humility and grace — even when it hurts.

ROBUSTNESS

Sometimes we get hung up on particularities of reagents and experiments, but what really matters is what we can credibly infer about how biology works. Moving from experiment to scientific finding requires the third ‘R’ of good science: robustness.

Robust findings become established over time as multiple lines of evidence emerge. Achieving robustness takes rigour and reproducibility, plus patience and judicious attention to the big picture.

As with the other Rs, my own impressions about robustness have been forged through some testing experiences. One involved a large collaborative project that I led

for nearly a decade — the Cancer Cell Line Encyclopedia. Reported inconsistencies with a related project, called the Genomics of Drug Sensitivity in Cancer, prompted intense, clarifying discussions on precisely how measurements are taken and on what sorts of data give the best insights for developing therapies^{4–15}.

A satisfying example of robustness occurred when our team published an unexpected finding, one that points to a vulnerability in certain malignancies: cancer cells harbouring a frequent genetic deletion become more reliant on a protein whose activity is regulated by the very metabolite that becomes elevated when the deletion is present. Two other large groups published similar findings at around the same time, each using distinct approaches^{16–18}. Diverse data that converge on the same observation in aggregate provide robustness, even though any single approach or model system has limitations.

WELL-LIT PATH

Recalling the teenage ‘aha moment’ that kindled my own investigative career decades ago still makes me smile. I did not pursue cancer research because it would offer copious opportunities to coax specific experiments to work reproducibly (such as that vexing chemical reaction). I chose this path because I believe that there are answers out there for

“What really matters is what we can credibly infer about how biology works.”

human disease, and that science holds the key to their discovery and application.

We scientists search tenaciously for information about how nature works through reason and experimentation. Who can deny the magnitude of knowledge we have gleaned, its acceleration over time, and its expanding positive impact on society? Of course, some data and models are fragile, and our understanding remains punctuated by false premises. Holding fast to the three Rs ensures that the path — although tortuous and treacherous at times — remains well lit. ■

Levi Garraway is senior vice-president of Global Oncology at Eli Lilly in Indianapolis, Indiana, USA.

e-mail: garraway_levi@lilly.com

1. *eLife* **6**, e23693 (2017).
2. Horrigan, S. K., Courville, P., Sampey, D., Zhou, F. & Cai, S. *eLife* **6**, e21634 (2017).
3. Hines, W. C. *et al. Cell Rep.* **6**, 779–781 (2014).
4. Safikhani, Z. *et al. F1000Research* **5**, 825 (2016).
5. Havery, P. M. *et al. Nature* **533**, 333–337 (2016).
6. Barretina, J. *et al. Nature* **483**, 603–607 (2012).
7. Garnett, M. J. *et al. Nature* **483**, 570–575 (2012).
8. Haibe-Kains, B. *et al. Nature* **504**, 389–393 (2013).
9. Cancer Cell Line Encyclopedia Consortium, Genomics of Drug Sensitivity in Cancer Consortium. *Nature* **528**, 84–87 (2015).
10. Geleher, P. *et al. Nature* **540**, E1–E2 (2016).
11. Safikhani, Z. *et al. Nature* **540**, E2–E4 (2016).
12. Mpindi, J. P. *et al. Nature* **540**, E5–E6 (2016).
13. Safikhani, Z. *et al. Nature* **540**, E6–E8 (2016).
14. Mehdi Bouhaddou, M. *et al. Nature* **540**, E9–E10 (2016).
15. Safikhani, Z. *et al. Nature* **540**, E11–E12 (2016).
16. Kryukov, G. V. *et al. Science* **351**, 1214–1218 (2016).
17. Mavrakis, K. J. *et al. Science* **351**, 1208–1213 (2016).
18. Marion, K. *et al. Cell Rep.* **15**, 574–587 (2016).

Five ways consortia can catalyse open science

An analysis of more than 50 collaborations shows the secrets of success, write **Joel Cutcher-Gershenfeld** and colleagues from the Stakeholder Alignment Collaborative.

“I am going to my grave with my disk drive in my cold dead hands.” So a senior scientist told a junior researcher, who related the tale at a 2013 US National Science Foundation (NSF) workshop on the reuse of physical samples in the geosciences. Sharing — of data sets, metadata, models, software and other resources — promises to speed discoveries, improve reproducibility and expand economic development. But it requires people to change.

Overcoming personal reluctance is doubly difficult because many aspects of the scientific enterprise undermine sharing.

Right now, most departments, funders and journals presume that data are proprietary from collection to publication. Even when individual scientists and institutional leaders want to do things differently, they face reviewers, colleagues and competitors clinging to conventional models.

As philosopher of science Thomas Kuhn documented more than 50 years ago, the scientific community resists challenges to its orthodoxy. The open sharing of data and other resources is a prime example¹. Conservatism reigns because, as the ‘iron law of oligarchy’ devised by sociologist

Robert Michels in the 1910s predicts, institutions established to achieve a certain goal will prioritize continued existence over the stated objective². And we’ve all experienced ‘path dependency’: that practices are hard to change once established³. For example, despite widespread support, few academic departments have restructured tenure processes to accommodate interdisciplinary work.

Over the past four years, we have studied more than a dozen scientific consortia involved in data sharing, and we’ve mapped the landscape of these ▶

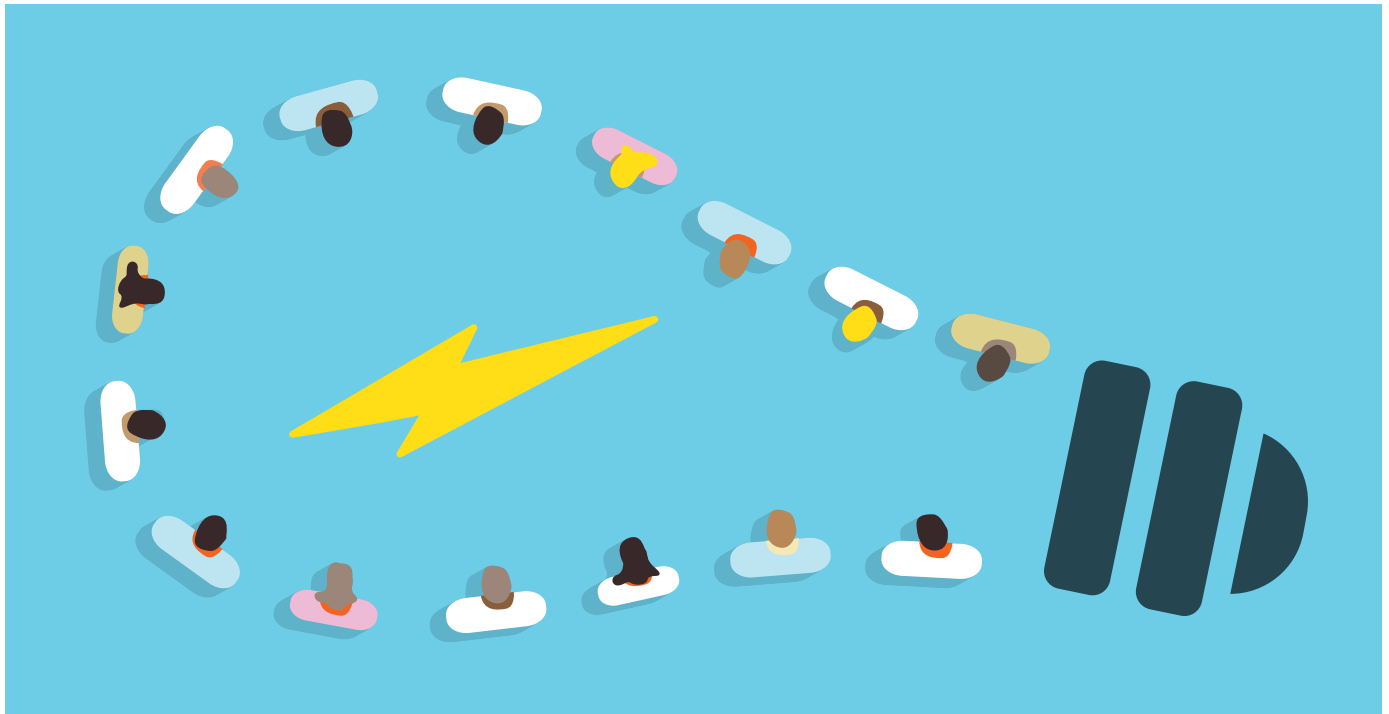


ILLUSTRATION BY OLIVER MUNDAY

▶ and another 44 such initiatives. When they work well, consortia act as catalysts, to accomplish what members cannot do alone^{4,5}. But scientists are seldom taught effective strategies to design and manage such coalitions. Here we distil the lessons from our fieldwork into five ways to foster open science.

FIVE PATHS TO OPENNESS

Build out from the middle. In 2013, the US Office of Science and Technology Policy stated that papers and data sets created with federal funds should be made broadly available to “accelerate scientific breakthroughs and innovation, promote entrepreneurship, and enhance economic growth”. But top-down initiatives on their own do not change behaviour.

Professional societies, funders, publishers and academic departments can operate in the middle, mediating between directives from on high and bottom-up actions. Unfortunately, traditional institutions are often both slow to change and unable to create the infrastructure needed to bridge visionary initiatives and daily practice.

Lack of integration across research sites and teams, spanning diverse fields and pursuing separate projects, can be fatal. For example, the US National Institutes of Health launched the National Children’s Study more than a decade ago, with the intention of coordinating 40 research sites in tracking 100,000 children from birth. In 2014, after US\$1.2 billion had been spent, the study was cancelled.

A more modest effort had greater success. In 2014, the NSF convened a three-day

meeting of around 30 facilities that curate, share and preserve scientific data across the geosciences. The formation of a Council of Data Facilities (CDF) was on the agenda, but when facilitators asked whether everyone was ready to draft a charter, two-thirds of attendees were not. This was a surprise. NSF-funded facilities wanted to go ahead. But those funded by other US agencies, such as the Department of Energy and NASA, were mostly there to safeguard their funding or operations — the iron law of oligarchy in action.

What the group did agree was to let the NSF data facilities draft a ‘strawperson’ charter. This accommodated the interests of opponents and was unanimously adopted. The CDF, launched later in 2014, has fostered initiatives to credit authors for the sharing and reuse of their data and for advancing common standards — tasks that no individual facility could have accomplished.

Consortia can cut across groups to cause fundamental shifts. All the successful ones we have studied help researchers to interact beyond established disciplinary or institutional silos. Influence rather than authority rules. Consortia advance interdependence while recognizing members’ independence.

Forge a shared vision. Nearly all models of how communities and organizations change emphasize the need to establish a shared vision. This is not a task that groups of scientists typically undertake. The lack of hierarchy in science means that an imposed, top-down vision is unlikely to succeed, and may even be impossible to organize.

The EarthCube initiative offers a model for how consortium can reach a unified vision. EarthCube was launched in 2011 to facilitate sharing in the Earth and space sciences. Its goal is to develop cyberinfrastructure that supports an estimated 200,000 geoscientists. Initially, more than 200 thought leaders participated in road-mapping exercises that revealed the absence of any single, all-encompassing cyberinfrastructure. Further outreach to more than 1,500 potential users provided the ‘voice of the customer’ in 27 discipline-specific workshops. Scientists realized that many concerns that they believed were unique to their own domains were actually widely shared.

Cyberinfrastructure experts worked with representatives across many disciplines to craft the programme. In 2015, the EarthCube shifted from outreach and planning to building tools and resources. While that work continues, a need for more outreach and engagement is emerging. The work of maintaining a shared vision is never complete. An NSF leader told a journalist last year that technology develops rapidly, whereas the social aspects of sharing data take a while to develop⁶.

Accommodate diverse, changing interests. A stakeholder can be a research team, an academic department, a professional society, a funder, a publisher or another entity. These generally have both competing and common interests. Successful consortia align stakeholders to recognize and promote mutual benefits and to appreciate separate characteristics. This requires building consensus

and resolving conflicts.

A wealth of specialisms can complicate data sharing: fields and disciplines have their own technical tics, data structure, classification systems and more. For example, we began surveying geoscientists as part of the NSF EarthCube initiative in 2013. Our 1,500 respondents identified their primary and secondary research fields, as well as more granular areas of expertise. They listed more than 700 unique areas of expertise such as basalt geochemistry, geomicrobiology and so on.

A scientist using sonar to map the seafloor will use different instruments and speak a different language from one studying riverine carbon cycles. But sharing across fields is where significant value is realized. For instance, a limnologist using sensor data on river flooding can benefit from collaborating with a data scientist who is curating satellite images of the same river. EarthCube supports data tools, so it stays mindful of what is needed to address differing terminology, technical needs, methods, norms and other matters in a systematic, rather than ad hoc manner.

Understanding exactly what various groups hope to get out of a project is important. In 2014, leaders from US supercomputing centres, along with university scholars, government agencies, publishers and others, formed the National Data Service (NDS) to promote middleware (software bridging systems and applications) and software services needed for data sharing.

Over a period of six months, they forged a shared vision, but then discovered that the vision meant different things to different stakeholders. For example, scientists just wanted tools and methods that would enhance their workflow. Software developers were motivated by the chance to build popular tools and methods. Cyberinfrastructure providers needed common tools, rather than customized solutions, to serve an increasingly diverse set of clients. Each group had to recognize the others' distinct reasons to participate for the consortium as a whole to make progress.

Successful consortia recognize that stakeholders and interests are dynamic. The development of unique digital identifiers for physical samples exemplifies this principle. Junior scientists, who tend to be more digitally adept, have been early adopters; some senior scientists prefer to stick to marking samples with felt-tip pens, which limits digital sharing. Over time, this ratio will change. Data-sharing initiatives must span such differences and adapt as participants, needs and technology evolve.

To be effective, consortia should re-map stakeholders and their interests periodically. To do this, some of us launched a new firm, WayMark Analytics, through an NSF

programme. Initiatives should be assessed for their likely effects, and evaluated after the fact for their actual effects.

Multiply impacts. Coalitions breed broader levels of cooperation. Consider the work of the CDF with science publishers. Many publishers insist that submissions include data, but few see data curation, sharing or storage as part of their business. Thus, publishers routinely accept difficult-to-use data packages, such as flat PDF files with insufficient metadata. Before the CDF, individual data facilities had to work out separate agreements with each publisher.

The CDF multiplied its impacts by collaborating with leading publishers to form the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS). Now, data submitted with articles is increasingly matched with the best facility for curation and reuse. This reduces effort, improves curation and enables access for others.

Similarly, the Biomarkers Consortium created new ways for stakeholders to work together. It defined a pre-competitive space for drug manufacturers, biotechnology

“Understanding exactly what various groups hope to get out of a project is important.”

companies, regulators, public research agencies, academics, patient advocacy groups and trade organizations.

This meant devising and formalizing ground rules, even just to allow meetings to take place. Competing companies needed assurance that they would not run afoul of anti-trust laws. The US Food and Drug Administration needed to participate in discussions about tools for drug development without compromising its regulatory authority.

The consortium has now completed more than a dozen projects, resulting in reliable biological readouts that are accelerating and easing drug development to speed effective treatments.

Co-evolve. Consortia enable science and the infrastructure for sharing data to co-evolve. The iPlant Collaborative was funded by the NSF in 2008 to build a platform linking high-performance computing centres to plant scientists. Initial participation was disappointing: they built it, but people didn't come. As the science changed, high-performance computing was needed for genomic data, and usage increased dramatically. In fact, the resources proved useful beyond botanical data. So, in 2015, iPlant expanded its focus to become Cyverse, which provides infrastructure for very large data sets and complex analyses across the life sciences. This was a

deliberate shift in step with changing science, and broadened the collaborative's impact.

COMING TOGETHER

Successful consortia avoid duplication of efforts, identify gaps and accommodate widely different rates of change. The NDS was launched in the United States two years after funders in the United States, Europe and Australia established the Research Data Alliance (RDA). Both services promote data sharing. Initially, the RDA was concerned that the NDS would duplicate its efforts. After six months of dialogue, it became clear that the NDS focused on advancing technology, whereas the RDA focused on social systems (community-generated use cases, identification of needed standards).

There is overlap, but there are many more ways in which these two consortia are distinct and complementary. In the early stages, the iron law of oligarchy and path dependency threatened to pull apart the two initiatives. Instead, the groups drafted a memorandum of understanding, aligning interests and defining complementarity. Each organization and their stakeholders benefit from their coming together.

Science is increasingly a collaborative enterprise, but its infrastructure, social and technical, lags behind. A transition to open science cannot depend solely on policy statements, voluntary action or academic departments. Multistakeholder consortia can serve as essential catalysts by following these principles. ■

Joel Cutcher-Gershenfeld is professor in the Heller School for Social Policy and Management, Brandeis University, Waltham, Massachusetts, USA, and a member of the Stakeholder Alignment Collaborative.

Karen S. Baker, Nicholas Berente, Courtney Flint, Gabriel Gershenfeld, Brandon Grant, Michael Haberman, John Leslie King, Christine Kirkpatrick, Barbara Lawrence, Spenser Lewis, W. Christopher Lenhardt, Mathew Mayernik, Charles McElroy, Barbara Mittleman, Namchul Shin, Shelley Stall, Susan Winter and Ilya Zaslavsky are members of the Stakeholder Alignment Collaborative. e-mail: joelcg@brandeis.edu

1. Kuhn, T. S. *The Structure of Scientific Revolutions* 31–65 (1962).
2. Michels, R. *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy* (Hearst's International Library Company, 1915).
3. David, P. A. *Struct. Change Econ. Dynam.* **5**, 205–220 (1994).
4. Cutcher-Gershenfeld, J. et al. *Data Sci. J.* **15**, 8 (2016).
5. Knight, E., Cutcher-Gershenfeld, J. & Mittleman, B. *MIT Sloan Mgmt Rev.* **57**, 16 (2015).
6. Witze, A. *Nature* **538**, 303 (2016).

Full author affiliations accompany this article online. Some authors declare competing financial interests: see go.nature.com/2nnnrs9 for details.