erates for prairie voles. They report a positive correlation between how often a male intrudes on a neighboring male's territory and how often his own territory is intruded upon by another wandering male. An EPF male encounters more females than IPF males and thus more opportunities for mating, but so does his partner back home. This trade-off is etched in the genome, with evidence of balancing selection for the above-mentioned *avpr1a* SNPs, but no such evidence at several other locations in the prairie vole genome.

Okhovat *et al.* propose that high population densities favor genetic variants resulting in lower V1aR expression, poorer spatial memory, and more expansive home ranges to capitalize on enhanced possibilities of extra-pair matings. Low population densities would favor the inverse of these traits. In other words, the evolutionary explanation for the persistence of both EPF and IPF males points to the very same cycles of population density that originally motivated Getz's field studies.

The study by Okhovat *et al.* impressively bridges mechanistic and evolutionary analyses to provide a detailed picture of individual differences in social behavior. Future studies should try to integrate the spatial learning and partner preference narratives for both males and females; the joint evolutionary dynamics of male and female traits must be considered to fully understand a mating system (*8*). With the availability of the prairie vole genome, future analyses also will no doubt include efforts to identify other genes that interact with *avpr1a*, in both mechanistic and evolutionary contexts (*9*). Measuring the effects of changes in population density on gene expression throughout the brain will help us better understand how nature and nurture shape social life (*10*). *M. ochrogaster* has come a long way from the traps on the prairie and clearly has much more to teach us. ■

REFERENCES AND NOTES

1. J. E. Hofmann, L. L. Getz, L. Gavish, *Am. Midland Nat.* **112**, 314 (1984).
2. C. S. Carter, L. L. Getz, *Sci. Am.* **268** (6), 100 (1993).
3. Z. R. Donaldson, L. J. Young, *Science* **322**, 900 (2008).
4. M. Okhovat *et al.*, *Science* **350**, 1371 (2015).
5. D. M. Witt *et al.*, *J. Neuroendocrinol.* **3**, 155 (1991).
6. A. G. Ophir, J. O. Wolff, S. M. Phelps, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1249 (2008).
7. S. M. Phelps, A. G. Ophir, in *Cognitive Ecology II*, R. Dukas, J. M. Ratcliffe, Eds. (Univ. of Chicago Press, Chicago, 2009), chap. 9.
8. S. Gavrilets, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9923 (2012).
9. C. C. Rittschof, G. E. Robinson, *Anim. Behav.* **192**, 263 (2014).
10. G. E. Robinson *et al.*, *Science* **322**, 896 (2008).

ACKNOWLEDGMENTS

I thank C. S. Carter, Z. R. Donaldson, L. L. Getz, and T. R. Insel for historical information, C. C. Lutz for comments that improved the manuscript, and M. Lee for expert assistance with the graphics.

10.1126/science.aad8071

DATA ACCESS

Sharing by design: Data and decentralized commons

Overcoming legal and policy obstacles

By **Jorge L. Contreras**[1]* **and**
Jerome H. Reichman[2]

POLICY

Ambitious international data-sharing initiatives have existed for years in fields such as genomics, earth science, and astronomy. But to realize the promise of widespread sharing of scientific data, intellectual property, data privacy, national security, and other legal and policy obstacles must be overcome (*1*). Although these issues have attracted much attention in some circles, they have often taken a back seat to addressing technical challenges. Yet failure to account for legal and policy issues at the outset of a large transborder data-sharing project can lead to undue resource expenditures and data-sharing structures that may offer fewer benefits than hoped. Drawing on our experience with the Belmont Forum, a multinational earth change–research program, we propose a framework to help plan data-sharing arrangements with a focus on early-stage decisions including options for legal interoperability.

> ***"Even if resources do not exist ... technically, there are advantages to fostering legal interoperability among distributed repositories."***

A rich literature beginning with the work of Ostrom (*2*) addresses the organization and governance of common pool resources shared by communities of users in contexts ranging from the global environment to communal living spaces. More recent work has expanded these principles to knowledge commons: collections of intangible resources, such as digital libraries, scholarly publications, and scientific data (*3*). Responding to calls for increased international scientific collaboration, several expert bodies have developed high-level principles for transborder data sharing (*4*–*6*). Although these efforts lay the groundwork for broad data-pooling initiatives, critical design decisions must be made before larger issues of governance and operation.

A SPECTRUM OF CENTRALIZATION. Although little empirical research exists on commons structures for data sharing and related costs, we have observed four basic structural models for scientific data pools along a continuum ranging from the most to the least centralized (see the table).

(i) ***fully centralized***: all data are aggregated in a single, centrally managed repository;

(ii) *intermediate distributed*: repositories are distributed and separately maintained, but may be interconnected by a central access portal, share technical service components, and utilize a common data-exchange format [sometimes called a federated database system (*7*)];

(iii) *fully distributed*: repositories are maintained locally and are not technically integrated, but share a common legal and policy framework that allows access on uniform terms and conditions (legal interoperability);

(iv) *noncommons*: repositories are largely disaggregated and lack technical and legal interoperability and, at most, may share a common index.

Centralized repositories with curation, analytics, and quality control can enhance the value of the data they contain [e.g., the GenBank repository of DNA and RNA sequence data (*8*)]. Centralized structures, however, come at a cost and may be impractical in many transborder collaborations because of political, legal, and organizational issues. But the alternative to a fully centralized commons need not be a noncommons. The shortfalls of noncommons models include incompatible data formats, inability to search across data sets, underutilization of data resources, individualized and inefficient access requirements, and difficulties moving data across national boundaries. Distributed commons structures, however, offer a meaningful subset of benefits with lower cost and resource commitments than fully centralized models.

For example, an online portal through which researchers can access multiple inde-

---

[1] *S. J. Quinney College of Law and Department of Human Genetics, University of Utah, Salt Lake City, UT, USA.*
[2] *School of Law, Duke University, Durham, NC, USA.*
**E-mail: jorge.contreras@law.utah.edu*

pendent repositories may feel like a centralized commons to users, but avoids the cost and governance overhead of a centralized repository [e.g., the Global Earth Observation System of Systems (GEOSS)]. Portal-based structures may make it easier for a central administrator to provide users with value-added services and aggregated statistics [e.g., the World Data Center for Microorganisms (*9*)], and allow users to more easily query, combine, and analyze multiple data sources (*7*).

Even if resources do not exist to link repositories technically, there are advantages to fostering legal interoperability among distributed repositories (*10*). To achieve this across jurisdictions, rules for data access and usage must be compatible with each other, must comply with laws and regulations of relevant jurisdictions, and must address rights of ownership and control granted to data generators (*11*). Legal interoperability can enable researchers to access and use data across multiple repositories without seeking authorization on a case-by-case basis, which increases the likelihood that more data will be put to productive use.

Perhaps the most straightforward path to legal interoperability is simply to contribute data to the public domain and waive all future rights to control it (*11*). This approach has been advocated by more than 250 organizations that have endorsed the 2010 Panton Principles for open data in science (*12*). Alternatively, researchers who wish to receive attribution credit for their contributions, but are otherwise willing to relinquish control over them, have released data under standardized Creative Commons licenses that have been widely used for other online content, including open-source code software, music, and photographs.

Despite the simplicity and appeal of these approaches, they are not always feasible. Data will often remain subject to legal regulation that, for instance, explicitly or implicitly reveal personally identifiable information, were obtained from human research subjects, relate to sensitive technologies, or disclose infrastructural details. Wilbanks and others, recognizing these requirements, have called for new models of informed consent and privacy protection to facilitate broad, socially beneficial sharing of at least some categories of such data (*13*).

DESIGN CONSIDERATIONS. If a collaborative research project has sufficient resources to create a centralized data repository with accompanying infrastructure and staffing (potentially millions of dollars up-front and thereafter for fully staffed and curated repositories), important benefits can be achieved. In most cases, however, this level of funding will not be available and a distributed data commons could be a desirable alternative. We found, in our experience with the Belmont Forum, that the project's leadership gave substantial weight to early aspirational statements regarding broad data sharing. Sufficient consideration may not have been given to potentially useful distributed data structures. When, at the conclusion of a lengthy planning stage, it became apparent that a centralized commons was beyond budgetary constraints, the decision was made to settle for no commons at all and rely on lofty but nonspecific data-sharing principles to motivate researchers to share data on their own (*14*). To help avoid such dilemmas in the future, we offer the following actionable framework for evaluating distributed data commons early in the project-planning phase:

How many data repositories are under consideration? If the number is small, then fully distributed, unlinked repositories (i.e., no commons) may suffice. Researchers may easily access each repository, and the cost of a commons structure can be avoided.

Are there resources to develop a common data portal? As the number of repositories increases, some form of commons structure will likely facilitate data sharing and usage. Although the cost is not trivial, a common portal can enhance the value and usability of the data. If funding for a data portal is not available, planners may wish to consider a fully distributed commons with legal interoperability.

Structural models for scientific data pools

Data-sharing options

BENEFITS AND COSTS	CENTRALIZED	INTERMEDIATE DISTRIBUTED	FULLY DISTRIBUTED	NONCOMMONS
Incremental research benefits				
Data access	Access to all data in unified manner	Access to multiple repositories through central portal	Access to each repository separately, but under a common usage/access policy and single approval	Ad hoc coordination with other repositories only
Data analytics	Most powerful search, analysis, quality assurance of aggregated data	Cross-repository searching and analytics; Metadata and aggregate statistics can be developed by central authority	Index/catalog only	Index/catalog only
Costs				
Up-front costs	Structure and build centralized repository; Develop data interoperability mechanisms; Develop common usage policy	Develop data interoperability mechanisms; Develop common usage policy	Develop common usage policy	Few up-front costs
Ongoing centralized costs	Operating and maintaining central repository; administering policies	Operating and maintaining portal; administering policies	Administering policies	No central costs
Ongoing distributed costs	Few distributed costs	Operating and maintaining repositories	Operating and maintaining repositories	Operating and maintaining repositories
Governance overhead	Central repository	Central portal/services, each distributed repository, and interrelationships	Each distributed repository and interrelationships	Each distributed repository with minimal coordination

Are data regulated in the relevant jurisdictions? This question is relevant no matter which commons structure is selected. If data are not regulated or subject to human-subject, privacy, health, or similar legal regimes, consider releasing data to the public domain or licensing under a common-use license. If data are regulated in one or more relevant jurisdictions, planners should consider engaging legal experts to develop a common data access and use policy that complies with regulations in each jurisdiction. For example, if data include human genetic information, both genetic nondiscrimination laws and data privacy regulations should be considered. Legal interoperability, and the ability for users to access and use all data on consistent terms via a single authorization, will be achieved only if the most stringent jurisdiction's regulations are observed in each case or are otherwise addressed (*13*).

Although the Belmont Forum will doubtless produce a wealth of valuable earth science data, initial appreciation of data-sharing options might have facilitated decision-making and planning among its many national participants and might have resulted in a more robust data-sharing structure. Addressing these design choices early—while acknowledging budgetary, legal, and political constraints—can save planning and implementation costs later.

REFERENCES AND NOTES

1. J. H. Reichman, P. F. Uhlir, *Law Contemp. Probl.* **66**, 315 (2003).
2. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge Univ. Press, Cambridge, 1990).
3. B. M. Frischmann, M. J. Madison, K. J. Strandburg, *Governing Knowledge Commons* (Oxford Univ. Press, New York, 2014), chap. 1.
4. Policy RECommendations for Open Access to Research Data in Europe (RECODE), http://recodeproject.eu.
5. International Council for Science, World Data System Strategic Plans 2014–2018 (ICSU, Paris, 2011).
6. B. M. Knoppers *et al., Genome Med.* **3**, 46 (2011).
7. Institute of Medicine, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* (National Academies Press, Washington, DC, 2015), chap. 6.
8. J. L. Contreras, *Science* **329**, 393 (2010).
9. J. H. Reichman, P. F. Uhlir, T. Dedeurwaerdere, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge Univ. Press, New York, forthcoming 2016), chap. 8.
10. J. Palfrey, U. Gasser, *Interop: The Promise and Perils of Highly Interconnected Systems* (Basic Books, New York, 2012).
11. Group on Earth Observations (GEOSS), "Mechanisms to share data as part of the GEOSS data-core" (White paper, GEOSS, Geneva, 2015); http://bit.ly/GEOSSdata-sharing.
12. P. Murray-Rust, C. Neylon, R. Pollock, J. Wilbanks, "Panton Principles: Principles for open data in science" (2010); http://pantonprinciples.org.
13. J. Wilbanks, in *Privacy, Big Data, and the Pubic Good* (Cambridge Univ. Press, Cambridge, 2014), chap. 11.
14. Belmont Forum Steering Committee, "A place to stand: e-Infrastructures and data management for global change research" (Belmont Forum Secretariat, Paris, 2015).

ACKNOWLEDGMENTS

J.H.R. has received support from the National Human Genome Research Institute (award no. P50HG003391). J.L.C. and J.H.R. served as members of the U.S. delegation to the Belmont Forum organized by NSF.

10.1126/science.aaa7485

PHOTO: PEOPLEIMAGES/ISTOCK PHOTO

SIGNAL PROCESSING

Matched filtering of ultrashort pulses

Optical signal processing can help reduce noise in detection of ultrafast electrical pulses

By **Michael Vasilyev**

The need to detect a small signal obscured by a large amount of noise is like the problem Prince Charming faced in his search for Cinderella. Fortunately, he had the magic glass slipper that perfectly fit only her. In signal detection and estimation theory, such a magic glass slipper that perfectly fits only the signal of interest is called a "matched filter." Because every change in the signal shape requires a change of the matched filter, it would be highly advantageous for the filter to be dynamically reconfigured according to the expected signal. For ultrashort (<100 ps) electrical signals that are near the limits of today's electronics, such reconfigurable matched filters become extremely challenging to make. On page 1343 of this issue, Ataie *et al.* (*1*) show how optical signal processing could help to achieve this goal. They demonstrate the detection of a single 80-ps pulse in the presence of a large amount of noise by converting electrical signals to optical signals and subsequently approximating the matched filter with a potentially reconfigurable optical scheme.

The matched-filter concept arises in the context of optimal detection of a signal $x(t)$ (see the figure, panels A and B) that is possibly scaled in magnitude, shifted in time, and degraded by additive noise (panel C). The optimal receiver correlates its input with the matched filter for a time-shift τ $[x(t-\tau)^*]$ producing the correlation function (panel D) with a tall and narrow peak at $\tau = 0$, thereby permitting easy detection. In the frequency domain, this is equivalent to multiplying the received spectrum by the matched filter in the shape of the complex conjugate of the expected signal spectrum. For received signals shifted in time, the matched time-domain filter is shifted accordingly, or the frequency-domain filter is multiplied by a linear phase response. It is well known (*2*) that for a signal degraded by white noise, the matched filter maximizes the signal-to-noise ratio (SNR) in the optimal receiver. Optical receivers with matched filtering have been demonstrated to operate within a fraction of a decibel from the quantum sensitivity limit (*3*) and are being considered for use in future near-Earth and interplanetary laser communication links.

Physically, the use of the matched filter integrates all of the signal's temporal or spectral components in such a way that they add up coherently (they are in phase) to create the peak of the correlation function, whereas the noise's components add incoherently (they have random phases). Thus, the resulting SNR is improved relative to the SNR of the individual components. The matched filter not only recovers the arrival time of the signal, but also compensates for the spectral phase distortions caused by signal propagation (dispersion) in microwave transmission lines and optical fibers, which, if left uncompensated, lead to unwanted pulse broadening (see the figure, panel B) as well as chirping and ringing effects in time domain that further drown the signal in the noise (panel C).

The signal domain can be straightforwardly expanded to two-dimensional images. Here, the linear phase response in

---

Department of Electrical Engineering, University of Texas, Arlington, TX 76019, USA. E-mail: vasilyev@uta.edu