

Post-publication sharing of data and tools

Despite existing guidelines on access to data and bioresources, good practice is not widespread. A meeting of mouse researchers in Rome proposes ways to promote a culture of sharing.

Sharing scientific data through publication has long underpinned the cycle of discovery and is the dominant means by which scientists earn credit for their work. More recently, technologies generating very large data sets and novel biological materials have given rise to principles under which communities share data and materials (pre- and post-publication), and to a new sharing infrastructure — large public databases and repositories. Although much attention has been given to practical and ethical guidelines for prepublication data release from large-scale ‘community resource projects’, summarized in the Bermuda Principles¹ and the Fort Lauderdale report², sharing of data and resources from hypothesis-driven research has largely been addressed piecemeal by individual communities, journals and funding agencies.

We report here the efforts of one such community to address issues of particular relevance to the free sharing of data and resources for mouse biology, genetics and functional genomics. Our community has had more than six decades experience with strategies for sharing mice, and more recently for cell lines. When it comes to resource sharing, the two greatest impediments to fully exploiting global research using the mouse as a model organism are the barriers created by material transfer agreements and the underutilization of public mouse repositories.

Community discussion

At a meeting in Rome in May organized by the CASIMIR consortium, a European project examining mouse research infrastructure, participants attempted to establish an agenda for community discussion. This meeting was attended not just by mouse investigators, but by representatives of funding agencies and journals, intellectual-property specialists and sociologists. The resulting Rome Agenda was designed to assist the stakeholders in developing a coordinated and directed approach to the main factors inhibiting free sharing of the fruits of publicly funded mouse research.

Two of the most important shared resources and research outputs in the field are mice and embryonic stem cells. The imperative to share such resources was probably first articulated by the US National Institutes of Health (NIH) in

March 1984. Yet even today, numerous unique mouse strains are not made available to the research community despite the existence of publicly funded mouse repositories provided for this purpose (see International Mouse Strain Resource (IMSR), www.findmice.org). Comparison of the number of knockout mice recorded by the international Mouse Genome Informatics (MGI) database (<http://www.informatics.jax.org/>) with those deposited in IMSR repositories suggests that currently only 35% are available in this way. This is an encouraging doubling of the percentage available since last assessed in a 2006 NIH survey. To further improve this figure, however, it is important that the sharing ethos is consistently observed by the mouse community and investment in repositories continues to keep pace with the generation of new strains.

Experiences shared at the meeting indicated that enforcement of existing policies regarding data and resource deposition is variable, and that despite increased emphasis on the importance of sharing by journals and funding organizations in recent years, there is evidence that geneticists and genomic researchers are withholding data and research materials with increasing frequency³. It is one thing

"Enforcement of existing policies regarding data and resource deposition is variable."

to encourage data deposition and resource sharing through guidelines and policy statements, and quite another to ensure that it happens in practice, as a recent informal survey of proteomics data deposition has revealed⁴.

Consequently, although many of the issues discussed in Rome are of specific concern to mouse biology and functional genomics, several have relevance to the wider biological sciences. For example, the issues surrounding licensing and patenting of genetically manipulated mice and embryonic stem cells could apply to many research tools that are generated through hypothesis-driven research. We hope that our experiences and recommendations can inform and stimulate broad discussion in the community as a whole and we ask readers to participate in an online forum to that end (see <http://tinyurl.com/mo4gh8>).

A strong message from Rome was that funding organizations, journals and researchers need to develop coordinated policies and actions on sharing issues. The Rome Agenda described

and summarized here (see ‘The Rome Agenda, overleaf’), represents a challenge to stakeholders to coordinate their efforts to facilitate the ready exchange of data and resources and to share good practices already implemented by some organizations and journals.

Access to publication-associated data

Prepublication data release is comprehensively discussed in an accompanying paper from the Toronto group⁵, whose conclusions were broadly supported in Rome. For publication-associated data, the meeting strongly endorsed the recommendations of the National Academy of Sciences UPSIDE report⁶, which lays out detailed guidelines for data sharing, not least the principle that data on which publications are based should be made available immediately on publication.

Currently, funding bodies rarely require investigators to deposit their mice in public repositories, although many encourage it, with the consequence that mutant lines may be lost or not fully exploited. The meeting strongly recommended that, at least on publication, journals should insist that mice and embryonic stem cells be deposited in a public repository within a specified time frame, the length of which still requires community consensus. Additionally, funders should be willing explicitly to cover the costs of deposition of mice arising from projects into public repositories.

We recommend that it becomes mandatory for scientific papers to explain where and how to access data and resources generated as part of the investigation. We are aware that some journals already have strong policy positions in this area, insisting that large data sets must be deposited in public databases, and that all reasonable requests for materials from other researchers must be fulfilled. There is however, heterogeneity with both policy and enforcement; surprisingly, many journals have no written policy on the availability of either bioresources or primary data.

In addition, papers should acknowledge any other data or materials used and the originating sources. This might be facilitated by the addition of metadata tags linking to data and bioresources⁴. A mechanism, such as a digital object identifier for resources in public repositories, would allow ready searching of the literature for specific bioresources, which is currently extremely difficult. It

would also add incentives for complying with data release and deposition policies by attributing credit to researchers who do share.

When it comes to compliance, journals and funding agencies have the most important role in enforcement and should clearly state their distribution and data-deposition policies, the consequences of non-compliance, and consistently enforce their policy. The costs of proactive 'policing' (explicit review at the end of grants or following publication) may be disproportionate, but a consistently implemented reactive policy, in a culture in which sharing is the ethical norm would, we believe, suffice.

Where they don't yet exist, clear criteria should be developed for reviewers of grants to help them assess data and material-sharing plans submitted as part of a funding proposal. There are already examples of good practice in this regard from the NIH⁷, the Howard Hughes Medical Institute, and several UK funding organizations such as the Wellcome Trust and the Medical Research Council⁸⁻¹⁰. Data-sharing plans are required in proposals, efforts are made to facilitate sharing, such as putting investigators in touch with repositories and, for some organizations, compliance is an important consideration in funding renewal.

Deposition of data and resources into public repositories is important for the validation of published results, as well as facilitating reuse. Although it is usual practice for major public databases to make data freely available to access and use, any restrictions on use should be strongly resisted and we endorse explicit

encouragement of open sharing, for example under the newly available CC0 public domain waiver of Creative Commons¹¹.

Licensing, patenting and material transfer agreements

Recent experience from technology-transfer programmes in the public sector discussed at the Rome meeting reflects a growing consensus among technology-transfer professionals that the patenting of mouse resources and genes is expensive and a poor return on investment. (Not least because most research tools are available under non-exclusive licences, whether patented or not.) This is reflected in a 1999 NIH policy that discourages filing of patents on mice as research tools generated from work done in its intramural research programmes. We recommend patenting research tools and methods only under exceptional circumstances, although patents may still be appropriate for research methods that are broadly applicable to multiple research fields.

Regardless of whether mouse resources or research methods are patented, licensing terms should be as broad as possible, acknowledging that academic institutions are both developer-providers and recipient-users of new mouse models, so there is little benefit in imposing obstacles on the availability and use of mice in the form of patents, licences and material transfer agreements (MTAs). Moreover, researchers

should be free to breed these mice for internal research purposes and to cross-breed them to develop innovative new mouse models.

With commercial use, any licensing of mice or methods to the private sector should include a broad reservation of rights on behalf of academic and not-for-profit institutions to use the mouse or method for non-commercial research purposes. In accordance with the sharing policies of some funding institutions, such as the NIH, it would be inappropriate to include licensing terms requiring royalty

"We recommend that materials and data be shared under the least restrictive terms possible."

reach-through or product reach-through on subsequent inventions, and institutional policies on intellectual property, technology transfer and licensing should reflect these principles. Equally, repositories should be able to distribute mouse resources to industry

under reasonable terms and conditions.

Within the academic community, processing of MTAs has become a major impediment to the open and timely dissemination of mouse resources and associated data¹². Onerous terms and conditions in many MTAs have increased transactional costs for institutions and have become a major cause of delay in negotiations and the sharing of resources. We recommend that materials and data be shared under the least restrictive terms possible. If documentation is necessary for any reason, then the minimum NIH sharing policy should be applied¹³. This 1999 policy states that materials developed

The Rome Agenda

Access to data and materials

- The data on which publications are based should be made available immediately through public databases on publication. Journals should insist that mice or embryonic stem cells are deposited in a public repository within a specified time frame.
- It should become mandatory for publications to explain where and how to access data and materials generated during the investigation. Publications should acknowledge any other data or materials used, the originating sources and availability.
- Grant reviewers should be provided with clear guidelines to assess data- and material-sharing plans, whether these have been met in the application, and whether the mechanism of sharing proposed would meet appropriate

goals if the work was to be funded or ultimately published.

- Funding organizations should be willing explicitly to cover the costs of deposition of materials arising from projects as part of the project budget.

Licensing and patenting

- The public sector should patent mice as research tools only under exceptional circumstances.
- Licensing terms for mouse resources or research methods should promote the establishment of a mouse 'research commons'.
- Materials and data should be shared under the least restrictive terms possible. Material transfer agreements for sharing materials between academic and not-for-profit institutions should be avoided or simplified.
- Researchers should be free to

breed shared mice for internal research purposes and to cross-breed to develop new mouse models.

- Licensing of mice or methods for commercial use should include a broad reservation of rights for academic and not-for-profit institutions.
- Licensing terms should not include inappropriate royalty reach-through or product reach-through on subsequent inventions, and institutional policy should reflect this.

Data and resource-sharing infrastructure

- Further dedicated sustainable investment in public databases and repositories should be encouraged.
- Funding agencies should provide researchers with clear direction on

expectations for data/resource/publication sharing, and should ensure appropriate data-sharing plans at the outset of projects and facilitate sharing as data and resources are generated.

Standards and tool development

- Data structure and semantics need standardizing and adopting.
- Metadata should be consistently attached.
- Investment is needed in computational tools to make use of standards and interoperability for data sharing and reuse.

Attribution and reward

- Attribution of data or resources should be enforced by journals and databases.
- A system for measuring attribution is needed to provide rewards for data sharing.

using NIH Federal funding should be freely transferred between researchers using "... either no formal agreement, a cover letter, the Simple Letter Agreement of the Uniform Biological Materials Transfer Agreement (UBMTA), or the UBMTA itself".

The Jackson Laboratory in Bar Harbor, Maine, an example of good practice, has applied these principles for many years. The laboratory provides mice to academic and not-for-profit researchers with the simple notification that the mice are to be used solely for research purposes and are not to be sold or transferred to third parties without permission.

Data and resource-sharing infrastructure

The view of meeting participants was that the largest part of the data underlying publications is archived on journals' 'supplemental information' sites or authors' own sites. These data are often formatted in a non-standard way, not readily searchable, and in the long term not guaranteed to persist. In a 2006 survey of major journals, Anderson *et al.*¹⁴ found that on average only 83% of supplementary data were still accessible a year after publication (for one journal this was as low as 33%) and that it seemed that approximately 10% of all data that was supposed to be available through a supplementary website was never available at all. It is clear, therefore, that the issue of long-term sustainable public repositories needs to be addressed by funding agencies, publishers and the community.

Many of the major public data repositories have no stable underlying funding and there are data types, particularly new ones, without appropriate public data repositories. We encourage further investment and recommend that public database coverage and stability be looked at in a coordinated way by funding organizations and the community with increased urgency. A good model is provided by the UK Biotechnology and Biological Sciences Research Council's Bioinformatics and Biological Resources Fund, which provides dedicated funding for development and sustainability of public resources and informatics tools.

Standards and tool development

Shared data are useful only if they are searchable and usable. For both attributes data must be formatted in a standard way, conform to standard structure and semantics and have appropriate metadata attached. It is clear that the community is still a long way from achieving these standards; further support and community discussion is needed. The full utility of standards such as MIBBI (Minimum Information for Biological and

Research commons

A research commons is a set of resources available to all scientists, either as part of the public domain or on standard terms and conditions that facilitate scientific collaboration, efficient reuse of materials and data, and dissemination of knowledge.

Biomedical Investigations) will be attained only by developing tools for data retrieval, mining and computation. The Gene Ontology bioinformatics initiatives provide a good example of how parallel development of tools and standards generates added value. Dedicated funding is needed to develop key elements of database infrastructure, including interoperability and data integration.

Common agenda

Despite oft-repeated statements of good intentions, stakeholders do not always share common interests. Within academia, a fear of 'helping the opposition' runs alongside concerns about the ethical or responsible use of freely shared data. A culture of sharing and open access is made more difficult by policies promoting the commercialization of research¹⁵, ineffective sharing infrastructure and inadequate data standards. Combined with unrealistic expectations from institutions of the value of exclusive licensing to the highest bidder, these factors can slow the progress of discovery and translation.

As an antidote to these concerns, the Rome meeting strongly encouraged sharing behaviours that promote a 'research commons' (see box, above). The heart of a research commons is one in which academic research is not impeded by restrictions on use and access to data and materials, in line with the principles of the Creative Commons¹¹. Adoption of a set of 'mouse research commons' principles would increase the effective use and economic value of publicly funded research by avoiding duplication of effort, unnecessary creation and use of live animal models, and facilitating reuse of data.

We know from the Jackson Laboratory's experience with its repository that developers of new mouse resources are willing to comply with an unrestrictive distribution policy as a condition for acceptance of their resources, so we believe the mouse research commons is not just a utopian dream. Rather it should create a paradigm shift to establish this as a norm for the research community. ■

Paul N. Schofield is in the Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, CB2 3EG, UK, **Tania Bubela** is in the Department of Public Health Sciences, University of Alberta, Edmonton Alberta, T6G 2V2, Canada, **Thomas Weaver** is

at the MRC Mary Lyon Centre, Harwell, Didcot, Oxfordshire, OX11 0RD, UK, **Lili Portilla** is at the National Center for Research Resources, Bethesda, Maryland 20892-4874, USA, **Stephen D. Brown** and **John M. Hancock** are at MRC Harwell, Mammalian Genetics Unit, Harwell Science and Innovation Campus, OX11 0RD, UK, **David Einhorn** is at the Jackson Laboratory, Bar Harbor, Maine 04609, USA, **Glaucio Tocchini-Valentini** is at the Istituto di Biologia Cellulare, 00015 Monterotondo Scalo, Rome, Italy, **Martin Hrabe de Angelis** is at the Institute of Experimental Genetics, Munich, Germany, **Nadia Rosenthal** is at the EMBL Monterotondo, 00015 Monterotondo, Rome, Italy, for the **CASIMIR Rome Meeting participants**. A complete list of the authors and their affiliations accompanies this article online.

e-mail: PS@mole.bio.cam.ac.uk

1. Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing Bermuda, 25-28 February 1996 (HUGO, 1996); available at www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml
2. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility (Wellcome Trust, 2003); available at www.wellcome.ac.uk/Stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf
3. Cohen, W. M. & Walsh, J. P. *Innov. Policy Econ.* **8**, 1-30 (2008).
4. *Nature Biotechnol.* **27**, 579 (2009).
5. Toronto International Data Release Workshop Authors *Nature* **461**, 168-169 (2009).
6. Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences (National Academy of Sciences, 2003).
7. Final NIH Statement on Sharing Research Data, NOT-OD-03-032 (National Institutes of Health, 2003); available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
8. Medical Research Council Policy on Data Sharing and Preservation; available at www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/DataSharingInitiative/Policy/index.htm
9. Wellcome Trust Policy on Data Management and Sharing; available at www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm
10. Biotechnology and Biological Sciences Research Council Data Sharing Policy; available at www.bbsrc.ac.uk/publications/policy/data_sharing_policy.html
11. Science Commons Database Protocol <http://sciencecommons.org/resources/faq/database-protocol>
12. Walsh, J. P., Cohen, W. M. & Cho, C. *Res. Policy* **36**, 1184-1203. (2007).
13. Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources Federal Register **64**, 72090-72096 (1999); available at http://grants.nih.gov/grants/intel-property_64FR72090.pdf
14. Anderson, N., Tarczy-Hornoch, P. & Bumgarner, R. E. *BMC Bioinformatics* **7**, 260 (2006).
15. Nelson, R. R. *Res. Policy* **33**, 455-471 (2004).

The views expressed in this paper represent the consensus of the meeting participants, and do not necessarily reflect the current policy of their respective Institutions.

Join the discussion at <http://tinyurl.com/mo4gh8>

See online special at <http://tinyurl.com/dataspecial>