



## The Responsibility to Share: Sharing the Responsibility

Martin Yuille, Bernhard Korn, Troy Moore, et al.

*Genome Res.* 2004 14: 2015-2019

Access the most recent version at doi:[10.1101/gr.2677304](https://doi.org/10.1101/gr.2677304)

---

**References** This article cites 7 articles, 3 of which can be accessed free at:  
<http://genome.cshlp.org/content/14/10b/2015.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# The Responsibility to Share: Sharing the Responsibility

Martin Yuille,<sup>1,8</sup> Bernhard Korn,<sup>2</sup> Troy Moore,<sup>3</sup> Andrew A. Farmer,<sup>4</sup> John Carrino,<sup>5</sup> Christa Prange,<sup>6</sup> and Yoshihide Hayashizaki<sup>7</sup>

<sup>1</sup>Medical Research Council geneservice, Genome Campus, Hinxton, Cambridge CB10 1SB United Kingdom; <sup>2</sup>Deutsches Ressourcenzentrum fuer Genomforschung, Im Neuenheimer Feld 506, Heidelberg D-69120 Germany; <sup>3</sup>Open Biosystems Inc., Huntsville, Alabama 35806, USA; <sup>4</sup>BD Biosciences Clontech, Palo Alto, California 94303, USA; <sup>5</sup>Invitrogen Corporation, Carlsbad, California 92008, USA; <sup>6</sup>The I.M.A.G.E. Consortium, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, California 94550, USA; <sup>7</sup>RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045 Japan

## A New Breed of Clones

The transcriptome, promoterome, and phenome clone sets described in this special issue of *Genome Research* are among the first of a new breed of clones. Their novel features affect how distributors on the path from originator to end user handle such resources. These new features are individual design, collective use, large scale, and flexibility. Whereas pre-genome clones bore random lengths of genomic DNA or cDNA from random locations in the genome or transcriptome, each of these post-genome clones bears a specialized vector harboring a tailored sequence that the originator selected and/or designed down to the last base pair.

These clone sets are intended for collective use in two senses. Firstly, an entire set may be used in one experiment with the aim of identifying the subset that elicits some detectable phenotype in vivo. Secondly, an entire community of scientists (e.g., the worm community) may use the same clone set to identify further subsets of clones that each elicit other phenotypes. When they share their data, its combined scientific value increases disproportionately.

The scale of these experiments is orders of magnitude greater than those studying one clone at a time. Since the 1970s, constructs have been arduously created by restriction and ligation. This technology is now being displaced by less-arduous in vitro recombination-based methods that permit systematic approaches to the study of defined sequences. Recombination-based cloning makes it far easier to derive subclones and swap specific sequences (representing, for example, orthologous protein

domains or epitope tags). This flexibility means that initial clone sets will become the basis for many rounds of subcloning to permit finer experimental definition of function. The simplicity of these methods and the growth of bioinformatics and robotics facilitate production of large-scale clone sets whose experimental value increases geometrically, but only if distribution is timely, accurate, and efficient.

Therefore, whereas the genome project used large numbers of clones for just one main purpose, post-genome molecular biology requires large clone sets for many inter-related purposes. This then creates opportunities—and problems—not only for investigative biologists, but also for infrastructure biologists (defined as those who created the machinery to facilitate the genome project that now needs adaptation to post-genomic biology).

This Commentary explores how the distribution infrastructure may respond to this new breed of clones in the current policy framework. Distribution is undertaken by a number of specialist organizations globally. The authors of this Commentary are drawn from public sector organizations that may be working under not-for-profit policies and from the private sector. In some ways, we compete with each other, but our joint authorship—a first, as far as we are aware—reflects perhaps the most important feature of the response required by post-genome biology, the requirement for greater coordination.

## Policy Framework

Two recent documents have laid out principles affecting distribution of materials and related data sets. The principles impact on authors who originate new resources and their funders, on editors and their reviewers, and on distributors and end-users.

## Cech Committee Report

The decision by *Science* to publish Celera's report on the human genome sequence—despite incomplete access to the data—provoked wide debate (Marshall 2000; Powledge 2001). Among its effects was a decision by the U.S. National Research Council to evaluate the responsibilities of authors for sharing data and materials and to recommend standards for sharing. The NRC Committee on Responsibilities of Authorship in the Biological Sciences under the chair of Tom Cech made its report last year (Committee on Authorship in the Biological Sciences). In response to the question, "Shouldn't there be exceptions to the general responsibility to share?" the Cech committee admitted that there might, on occasion, be legal or other considerations (e.g., forbidding export of certain biological materials), but the committee determined that there was a basic principle to uphold. This was propounded as the Uniform Principle for Sharing Integral Data and materials Expeditiously (UPSIDE), which was to be implemented through 10 main recommendations (see Box 1). The committee stressed the beneficial consequences of UPSIDE by noting that universal adherence without exception will promote cooperation and prevent divisiveness in the scientific community, maintain the value and prestige of publication, and promote the progress of science.

## OECD Report

The Paris-based Organization for Economic Co-operation and Development (OECD) aims to help governments adopt strategic orientations by deciphering emerging issues and by identifying policies that work. An OECD task force of scientists from 19 countries chaired by Hideaki Sugawara (DNA Data Bank of Japan) produced a report in 2001 on Biological Resource Centers (Working Party on Biotechnology). The report

### <sup>8</sup>Corresponding author.

E-MAIL [myuille@geneservice.mrc.ac.uk](mailto:myuille@geneservice.mrc.ac.uk);  
FAX 44-1223-494512.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.2677304>.

**Box 1. NRC Committee on Responsibilities of Authorship in the Biological Sciences**

## UPSIDE principles

In the committee's view, the principle has five corollary principles:

1. Papers should include all data or algorithms needed to support their major claims.
2. If this information cannot be included (e.g., data set is too large), authors should make it available to all researchers at no charge (e.g., online).
3. By the time of publication, authors should deposit data in publicly accessible data repositories if these are in general use.
4. Authors should identify which materials others might request and should state how to obtain those materials.
5. Authors should make patented materials available under license.

## UPSIDE recommendations

Building on these principles, Cech makes ten recommendations:

1. Scientists should continue to contribute to framing database protection laws.
2. Reviewers for journals should advise authors on the what and how of making materials available.
3. Authors should not apply conditions to others' use of unpatented new materials (e.g., requiring collaboration).
4. We should move toward one standard Material Transfer Agreement with a streamlined process.
5. Researchers should receive their requested materials expeditiously (<60 days).
6. Journals should have a clear policy on sharing data and materials, with a policy on authors' nonadherence to the policy.
7. Research funding policy should include a material and data sharing policy for PIs.
8. Researchers should report an author who delays despatch—first to the journal (>60 days), then to the authors' employer or funder (>90 days).
9. Funders should pay PIs to cover costs of dissemination of data and materials.
10. Researchers receiving data and materials should acknowledge this appropriately.

challenged governments to address the question, "How do we move from technologies based on mineral resources (metals, coal, oil, etc.) and on physics, chemistry, and engineering to technologies increasingly based on biological resources, and, more particularly, on something that is essentially invisible—the living cell and its genes?" Its proposals to address this question recommended countries to establish accredited Biological Resource Centers with agreed international standards and data links (summarized in Box 2).

**Assuring Distribution**

In our individual capacities, and as far as they affect our organizations' activities, we support these UPSIDE principles and OECD recommendations and propose

that all of us involved in biological research—grant applicants, grant reviewers, funders, authors, publishers, editors, and reviewers, as well as infrastructure scientists—need to review how to act to ensure that methods for sharing resources will meet the needs of post-genome biology.

There are three obvious checkpoints where resource sharing can be implemented. The first is at the time of submission of a proposal for funding. Although some funders do have a general policy on sharing data and materials (e.g., see [http://www.mrc.ac.uk/index/strategy-strategy/strategy-science\\_strategy/strategy-strategy\\_implementation/strategy-other\\_initiatives/strategy-data\\_sharing.htm](http://www.mrc.ac.uk/index/strategy-strategy/strategy-science_strategy/strategy-strategy_implementation/strategy-other_initiatives/strategy-data_sharing.htm) and <http://www.wellcome.ac.uk/en/1/awtpubrepdat.html>), they could make the intention to

distribute a new resource an explicit condition of support when appropriate. The budget pages in grant applications could specify a budget item to cover costs of resource sharing. If the applicant names a distributor who has agreed to acquire the planned resource, this would strengthen the case for funding, as it would provide evidence that the resource is likely to be of wide interest to the community. Distributors do not want to fill their freezers with rarely requested resources.

*Science* recently reported a survey showing that >25% of U.S. geneticists say they cannot replicate published findings because other investigators will not give them the data and materials they need (Stokstad 2002). Therefore, a sec-

**Box 2. OECD Report: Biological Resource Centers Underpinning the Future of Life Sciences and Biotechnology**

## Summary of recommendations

1. Establish national BRCs.
 

Selectively seek to strengthen existing ex situ collections of biological data and materials and, when needed, create new collections, including in non-OECD countries, and raise those collections to the quality required for accreditation as national BRCs.
2. Develop an accreditation system for BRCs based on international criteria.
 

Support the development of an accreditation system for BRCs based upon scientifically acceptable objective international criteria for quality, expertise and financial stability.
3. Create international linkages among BRCs.
 

Facilitate international coordination among national BRCs. This should be based upon modern informatics systems that link biological data to biological materials across national BRCs and upon common technological frameworks. Biological Resource Centers 50
4. Coordinate standards, rules, and regulations taking BRCs into account.
 

Take into account the objectives and functioning of BRCs when establishing and harmonizing national or international rules and regulations. Develop policies to harmonize the operational parameters under which BRCs function, including those governing access to biological resources, as well as their exchange and distribution, taking into account relevant national and international laws and agreements.
5. Establish a global BRC network.
 

Support the establishment of a global BRC network that would enhance access to BRCs and foster international cooperation and economic development. A global BRC network would greatly improve the conditions under which biological materials and information are preserved and exchanged. How this challenge is met may affect the future of life sciences and biotechnology for many years to come. It is a challenge that calls for the full support of governments, the scientific community, and the private sector. BRCs should be encouraged to coordinate their activities so as to best serve their essential functions in response to the needs of sectors that depend on their biological resources.

ond checkpoint arises when a manuscript describing the new resource is submitted for publication. Only half of the leading journals have a policy in this area (see Table 1). This includes *Genome Research*, which instructs authors as follows:

“Researchers who submit papers to this journal are prepared to make available to researchers all materials needed to duplicate their work. Material from a publication must be easily available to the broader community in publicly held databases and repositories when available, and at the *Genome Research* Web site, and if desired at the author’s Web site, when they are not. *Genome Research* will NOT consider manuscripts where data used in the paper is not freely available on either a publicly held Web site, or in the absence of such a Web site, on the *Genome Research* Web site. There are NO exceptions.”

These instructions go some way toward addressing the issue of resource distribution and are in-line with the Cech Committee recommendations. However, we believe that journal policies should be stronger. They might inform authors that, before accepting a manuscript for publication, authors will confirm that the named distributor(s) has agreed to distribute. Currently it takes six months, or more, from the time of initial contact between author and distributor and the time of release of the resource. This is because their negotiations on issues such as a Material Transfer Agreement can be lengthy.

We encourage scientists to use a recognized distributor (rather than trying to do it themselves), because this helps to control quality and to ensure the long-term availability of the resource. An author’s scientific interests may change over time, as may his/her

funding situation, leading to difficulties in distribution.

The third checkpoint for implementing the sharing of resources is provided by distributors themselves. Identification and acquisition of new resources is their *raison d’être*. To do this efficiently, they need to make sure the rest of the scientific community understands their increasingly important role and their problems. One major problem is that space is running out. After acquisition and release of a new resource, initial interest eventually wanes; the distributor has changed roles and become an archivist. The cost of archiving old clones is borne by increasing the price of new clones. Today, there are tens of millions of samples held in freezers.

Possible solutions to counter the effects of rising clone prices include destroying resources in which interest has waned, reducing storage costs through other means, or separating out archiving. However, interest in older resources can wax again when a new application emerges. We have seen this with Bird’s 1994 CpG island libraries (Cross et al. 1994) that now find new applications in microarrays. All researchers know that clearing out the freezer never really works, as that tube at the back is the one we may need tomorrow. Distributors may alleviate this by offering an old resource back to the originator. However, the chances are that the originator’s interests will have shifted and that it will be someone else who invents the new application. Cutting costs can be achieved by reducing security (reducing the number of copies of each resource) or by implementing new and cheaper technologies to replace the conventional storage system of glycerol stocks held at  $-80^{\circ}\text{C}$ . An attractive proposal is the use of ambient temperature, paper-based

technology, of which one example is the RIKEN DNA Book (Kawai and Hayashizaki 2003). This would cut storage costs, but work is needed to ensure that acquisition and retrieval costs would not increase substantially.

The OECD has suggested that archiving is as important for tomorrow’s biotechnology as is husbanding of mineral resources for conventional industrial technologies. This raises the issue of whether distributors make good archivists. Although they possess relevant expertise, do they have sufficient long-term security? Public sector distributors receive public support or underwriting, but this is usually time-limited. Private-sector distributors are part of the high-risk biotechnology industry. A halfway measure that each distributor could take would be to pledge that in the event of the distributor ending its activity, it would permit another distributor to acquire its resources for continued distribution. We are aware of at least one former distributor that has demonstrated its commitment to sharing its responsibilities by doing this on a no-fee basis.

### Individual Design

The individual design of clones in post-genomic clone sets is one factor of two that obliges distributors to implement tighter quality controls. The first resources distributed by the first specialized genomics resource center—the Medical Research Council’s UK Human Genome Mapping Project Resource Centre (now superseded by MRC geneservice)—were small numbers of manually constructed genomic fragments of unknown sequence that probed polymorphisms to establish their map position. Quality controls were restricted to

**Table 1.** Policies on Sharing Materials and Data of 56 Most Frequently Cited Journals by Publisher and by Content

	Percentage of journals				
	All journals (N = 56)	Society or Association publishers (N = 37)	Commercial publishers (N = 19)	Life sciences journals (N = 38)	Clinical-medical journals (N = 18)
Journals with a policy	55	51	58	68	28
Policy specifies:					
Sharing materials	39	30	58	47	22
Sharing software	2	0	5	3	0
Depositing data	41	35	58	53	17
Statement of consequences	2	3	0	0	6
Whom to contact	4	3	5	3	6

Adapted from the Committee on Responsibilities of Authorship in the Biological Sciences, reprinted with permission from The National Academy of Sciences, courtesy of The National Academies Press © 2003.

checking for phage contamination. This was quite reasonable, given that the clones were themselves the object of investigation, and thus there was, in effect, a collaboration between the resource center and end-users who were expected to return the mapping data that they had acquired to the resource center. This data was intended for integration with existing mapping data so as to refine the human genetic map.

Larger, better-characterized clone collections were developed as the genome project strategy moved from mapping to sequencing. However, quality controls were unchanged. Sydney Brenner's proposed shortcut to acquiring the purportedly most interesting sequences in the genome by sequencing cDNA clones prompted collaborations like the IMAGE (Integrated Molecular Analysis of Genomes and their Expression) Consortium in 1996 (Lennon et al. 1996). There are now in excess of 5 million IMAGE clones from human sources as well as model organisms. In practice, their value was mainly in transcriptional, rather than genomic analysis. However, the IMAGE Consortium increased the scientific value of its libraries by deriving and resequencing subsets like the full-length cDNA Mammalian Gene Collection (Strausberg et al. 1999). The RIKEN murine full-length cDNA collection (Okazaki et al. 2002) is comparable (although is not widely available). Constructing large clone sets and subsets required the use of robotics. This has led to the second factor, which obliges distributors to implement tighter quality controls. In research labs, robots are used to construct and pick libraries. Distributors use robots to copy and rearray libraries. These manipulations cannot be performed with standard microbiological aseptic techniques, thereby increasing risks of bacterial or phage contamination, well-to-well cross-contamination, and empty wells (no inoculum). These problems are intermittent, and may only be detected as complaints by end-users some time after distribution or after numerous rounds of resource replication. Current solutions require close cooperation with the originating lab and are expensive, because they are labor intensive. However, if the originator makes early contact with the distributor, some problems can be resolved with ease. For example, originators may make pools of >1 clone/well for reasons of economy and with a view to limited use of the resource. However, after extensive rounds of resource replication, pooling leads to underrepresentation of slower-growing clones in each pool. Pooling in

this way is therefore a false economy from the distributors' point of view.

### Collective Use, Large Scale, and Flexibility

The development of transcriptome, promoterome, and similar resources may require refinements of the resource. These refinements may be both experimental and bioinformatic in nature. Hence, more than one version of the resource will exist, and each version will have annotations that will change over time. It is important to note that collective use is inherent in creating the completed resource, as multiple labs will contribute to the refinement. Distribution of the resource to these labs is an essential part of the project to create the completed resource.

When distribution becomes part of a research project, distributors need to consider their role with some care. Should distributors be responsible for ensuring that the catalog of versions of the resource is kept up to date? Is it rather the responsibility of the resource originator, or of each lab that is refining the resource; or is it the responsibility of the project's funders? Which of these has the greatest interest in assuring the biological and the bioinformatics quality of the resource in the short and long term? Can updating the catalog await formal publication of results in one journal or another?

Similar problems exist for pregenomic resources with postgenomic applications (in both transcriptome and proteome analysis), such as the IMAGE collections. Thus, IMAGE Consortium member Greg Lennon has noted how the resource would benefit from an expression database (None 1999). This point also touches on collective use, as multiple labs gather data on expression and its effects in numerous assays. But, who should be responsible? An investigator may say "Why should I update catalogs with data that I didn't accrue?" A funder may say "Why should we update catalogs on experiments we didn't support?" An editor may say "Why should I update catalogs with data from papers we didn't accept?" A distributor may say "My job is to distribute clones, not data."

An important lesson from the early days of the genome project is that updating cannot be left simply to the good will of the community. In those days, clones that detect DNA polymorphisms were distributed to end-users on the understanding that new data would be returned to the distribution center. Some 90% of end-users never returned their data.

For IMAGE clones used in spotted arrays, a community solution is emerging built on the MIAME (Minimal Information About a Microarray Experiment) principles under the auspices of the international Microarray Gene Expression Data Society (<http://www.mged.org/index.html>). Data on methods and results is warehoused by organizations such as the European Bioinformatics Institute.

A comparable solution for dealing with post-genomic clone and data sets does not exist. However, the need for such a solution is growing. It will require the cooperation of clone originators, distributors, end-users, and the numerous journals in which they report their results. An important first step is to put in place standard formats that can cope with the flexibility and complex functionality of post-genomic clone sets. Simply a link to a genome sequence database is inadequate. Hilson and colleagues have proposed a useful format (B. De Meyer, C. Lurin, I. Small, and P. Hilson, unpubl., [http://www.orfeome.org/orfweb/modules/UpDownload/store\\_folder/Hilson/De\\_Meyer\\_et\\_al.pdf](http://www.orfeome.org/orfweb/modules/UpDownload/store_folder/Hilson/De_Meyer_et_al.pdf)). Their proposal arose from an EU-funded project, of which some distributors are members. The proposal is that there is some Minimum Information About an ORF (MIAO) that should be held on each clone and its derivatives to assist investigators in discovering the originator's intentions in designing the clone. Adopting standard nomenclature, such as MIAO, would allow methods and results data sets to be attached to each clone, thereby facilitating curation, the design of further experiments, and meta-analysis. Again, the question is, who should be responsible?

We suggest that the underlying principle that can help determine responsibilities is that of core competencies. The core competence of an investigative lab is in the formulation of hypotheses in a particular area of biology, in the analysis of data from experiments that explore those hypotheses, and occasionally, in inventing new methods. Infrastructure labs or organizations have competencies in the handling of biological materials or related data. Advances in technology are accentuating these distinctions and tending to create new divisions of labor—new competencies. Established competencies tend to migrate to the infrastructure or to service sectors (a clear example is the migration of oligonucleotide synthesis from the bench to the supplier). Turning these truisms into changing practices is hard for us all, but it seems

that the time for choices regarding the distribution of postgenomic resources has arrived.

### End Note

The process of preparing this commentary has revealed a high degree of cooperativeness between the authors. Therefore, we have decided that it would be appropriate to meet to work out how we should actually achieve the things to which we are committing ourselves—ensuring that new resources and related data are available to the research community, embracing new technologies appropriate to our competencies, supporting the UPSIDE and OECD principles, and keeping down costs. The key will be to act in the integrated, yet quasi-autonomous, way that seems to be characteristic of the new postgenomic biology.

### REFERENCES

Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council. 2003. *Sharing publication-related data and materials:*

- Responsibilities of authorship in the life sciences.* The National Academy of Sciences, The National Academies Press, Washington, D.C.  
<http://books.nap.edu/catalog/10613.html>
- Cross, S.H., Charlton, J.A., Nan, X., and Bird, A.P. 1994. Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* **6**: 236–244.
- Kawai, J. and Hayashizaki, Y. 2003. DNA book. *Genome Res.* **13**: 1488–1495.
- Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M.B. 1996. The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33**: 151–152.
- Marshall, E. 2000. Human genome: Storm erupts over terms for publishing Celera's sequence. *Science* **290**: 2042–2043.
- None. 1999. Hot papers in genomics. *The Scientist* **13**: 17–18.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. (The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team) 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Powledge, T.M. 2001. Changing the rules? The agreement between Celera and *Science* magazine concerning Celera's publication of its human genome sequence is upsetting many researchers in bioinformatics. *EMBO Rep.* **2**: 171–172.
- Stokstad, E. 2002. Data hoarding blocks progress in genetics. *Science* **295**: 599.
- Strausberg, R.L., Feingold, E.A., Klausner, R.D., and Collins, F.S. 1999. The mammalian gene collection. *Science* **286**: 455–457.
- Working Party on Biotechnology. 2001. Organization for economic co-operation and development. Biological resource centres underpinning the future of life sciences and biotechnology. SourceOECD Science & Information Technology. No. 7, pp. 1–68.

### WEB SITE REFERENCES

- <http://www.mged.org/index.html>; Web site of the Microarray Expression Data Society.
- [http://www.mrc.ac.uk/index/strategy-strategy/science\\_strategy/strategy-strategy\\_implementation/strategy-other\\_initiatives/strategy-data\\_sharing.htm](http://www.mrc.ac.uk/index/strategy-strategy/science_strategy/strategy-strategy_implementation/strategy-other_initiatives/strategy-data_sharing.htm); Describes the policy of the Medical Research Council on sharing data and materials.
- <http://www.wellcome.ac.uk/en/1/awtpubrepdat.html>; Describes the policy of the Wellcome Trust on sharing data.
- [http://www.orfome.org/orfweb/modules/UpDownload/store\\_folder/Hilson/De\\_Meyer\\_et\\_al.pdf](http://www.orfome.org/orfweb/modules/UpDownload/store_folder/Hilson/De_Meyer_et_al.pdf); MIAO, the minimum information about an ORF.