



## Unlimited Access—Limitless Success

Laurie Goodman

*Genome Res.* 2001 11: 637-638

Access the most recent version at doi:[10.1101/gr.11.5.637](https://doi.org/10.1101/gr.11.5.637)

---

**References** This article cites 2 articles, 1 of which can be accessed free at:  
<http://genome.cshlp.org/content/11/5/637.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

# Unlimited Access—Limitless Success

“Knowledge is of two kinds: we know a subject ourselves, or we know where we can find information upon it.”  
—Samuel Johnson

There is a long-standing tradition in the area of scientific publication that material presented in a publication is made available to interested parties within the community. Sharing this material serves a twofold purpose: First, anything presented in the literature can be duplicated; second, and more importantly, others can add to the information base using these materials, thus more rapidly increasing our understanding of a field, disease, biological process, etc.. Although this has been a general tenet of the scientific community, there are and will continue to be cases where individuals prefer to maintain research information in a more proprietary way and for a variety of reasons, many of which are equally useful to society.

There are many reasons why individuals wish to publish their work, including the wish to be recognized as the person who accomplished a particular goal or made a discovery, or as a means to advertise a product. The desire to “make your mark” in print is a major part of all publications, especially when careers can and do rely on the number and type of publications an individual has to his or her credit. Regardless, the main point of publishing in the scientific literature is to educate and to increase the community’s ability to further work in that area; the availability of data and tools from such publications is an essential part of the process.

The publication of the Human Genome Draft Sequence by the Public Consortium and by the private company Celera Genomics (Human Genome Sequencing Consortium 2001; Celera Genomics 2001) brought to the fore the vagueness of policies that journals have set with regard to data availability upon publication. *Genome Research’s* own written policy on the sharing of material and information upon publication was as follows: “It is also understood that researchers who submit papers to this journal are prepared to make available to researchers materials needed to duplicate their work. Authors of accepted manuscripts must submit mapping and sequence data to the appropriate data bank and provide an accession number for these data at the page proof stage.” One can easily concede that there is a great deal of room for determining what is actually the letter of such agreements. In light of recent discussions over the availability of sequence data from human genome papers in *Science* and *Nature*, the Editors of *Genome Research* wish to clarify our policy on data and material availability from papers published in *Genome Research*.

Upon publication in *Genome Research*, all related

sequence information must be deposited in one of the public databases; at this stage, this means depositing the sequence data in EMBL (<http://www.ebi.ac.uk/embl/Submission/index.html>), GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>), or DDBJ (<http://www.sakura.ddbj.nig.ac.jp/>). We do agree that there is some advantage to the community that material from private companies, such as the human genome sequence from Celera, is publicly available in some way. We feel, however, that the adoption of a policy by which sequence data can be provided publicly—but not necessarily through one of the public databases—would create a slippery slope that may make sequence data accessibility increasingly difficult and less useful.

The current public sequence databases, although administered separately, cross-compare and upload information from each other; thus, these separate databases effectively provide a single source for sequence information. If we were to allow one group to maintain their sequence data on their own site, sequence data would become more fragmented. Also under those circumstances, each time we publish a paper containing sequence information, each additional group would then have the right to request that they maintain their sequence data on their own site as well. Such a policy can only lead to further fragmentation of sequence data that is inherently most useful when it can be directly combined and compared with related material. A further compounding issue is that these multiple sites may not be maintained in perpetuity because individuals, laboratories, and companies cannot guarantee that such sites will continue to be maintained in an appropriate fashion.

Data availability does not, however, only mean sequence data. On publication in *Genome Research*, any data that has a public submission site must be deposited in its appropriate public database. This includes, for example, expression data (array data and SAGE data can both be deposited in GEO (Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>); and only array data can be submitted to ArrayExpress [<http://www.ebi.ac.uk/arrayexpress>]). SNP data should be submitted to SNPdb (<http://www.ncbi.nlm.nih.gov/SNP/>); note also that several databases for other organisms, such as Flybase (<http://www.flybase.org>), maintain sites for public deposition of SNP data as well. Protein 3-D structure data should go to PDB (<http://www.rcsb.org/pdb/>); this site contains mostly only experimentally-determined structures.

There are other data in papers for which there are

no public databases accepting submissions, but rather there are curated databases where the curators keep track of the literature and maintain and update the database with new information. For example, information on protein domains is available from a number of databases such as PFAM (<http://pfam.wustl.edu/>), ProSite (<http://www.expasy.ch/prosite/>), and PRINTS (<http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINTS.html>). We encourage authors to make these databases aware of newly published material, if possible, but such data should be made available as described below.

In cases where there are no public databases available, the *Genome Research* Web site will maintain flat files of such datasets. The authors can make this material available on their Web sites as well. Papers that present novel computer software must have the source code freely available to everyone, enabling individuals to reproduce the results reported in the paper and also to advance research in related areas. We recognize that there are reasons that individuals would wish to keep such information available only to academicians; however, at this stage the separation of academia and business is no longer clearcut. Instead, we fully expect individuals who wish to publish their work, and thereby make the information available to the community, to legally protect this material via copyright or patents. Data on pedigrees should also be exchanged, once published, but we recognize that the rights of the families need to be appropriately protected in this process. Nevertheless, researchers should be able to design a way to distribute this information while still maintaining con-

fidentiality, and we encourage those who are involved with research using pedigree data to come together to find ways to better utilize these resources as a group. Thus, the families who are involved in this research can expect to reap the benefits of such work sooner.

Authors should also be prepared to exchange resources such as clones, animal stock, cell cultures, etc., but our readers must clearly understand that resources such as these do have limitations with regard to ease of exchange: There may be extensive time required for preparation, high cost, and, quite often, limited availability of the original source. Authors should strive, when possible, to send their material to the public repositories that are available to handle some of these types of resources.

In short, as much as is reasonably possible, material from a publication must be easily available to the broader community—in public databases and repositories when available, and at the *Genome Research* and author's Web site when they are not. By pursuing publication, the author's goal is to educate, enlighten, and enrich the scientific community to generally further the pursuits of the community at large; to do so, he or she needs to provide all the related resources from that publication to that community.

—Laurie Goodman

## REFERENCES

- International Human Genome Sequencing Consortium. 2001. *Nature* **409**: 860–921.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. *Science* **291**: 1304–1351.