

The Open Triangle

*Michael J. Ackerman, Ph.D.**

During the past year, there has been more than the usual discussion about open software. The discussion has expanded into an area being referred to as “open science.” Open science has three components: open literature, open data, and open software. Proponents of open science represent this concept by the following formula:

open literature + open data + open software = open science

What is fueling the “open” movement, and what will it mean to clinical medicine?

The potential negative that you must understand is that “open” does not always mean “free.” It is important that you keep this in mind in order not to be misled by the rhetoric. Open may be associated with both direct and indirect costs that are borne by the customer of the open item. With this thought in mind, let’s explore the brave, new open world.

At the risk of antagonizing journal publishers, I will begin with one of the components—open literature. Several years ago, when Dr. Harold Varmus¹ was the director of the National Institutes of Health (NIH) he pushed the idea that if the NIH, and therefore the American people, had paid for the research, the findings of that research should be made freely available. This is a problem in that academic publishing through journals usually requires the author to transfer copyright of the information to the journal publisher. If the findings are released on the Internet, for example, by the author before the copyright exchange, the findings are usually not accepted for publication. If the author releases the findings after the exchange of copyright, there is a potential legal problem with the journal publisher. In that sense, access to the results of scientific findings is through controlled-access literature not open-access literature. Whether or not anyone is being denied access to the information he or she needs for any reason is not what is at issue. What is at issue is the control of intellectual property that was paid for in advance by a third party that wants to exercise some control or influence over the disposition of that intellectual property.

This is not a new problem; just the medium has changed. Early on, a person either subscribed to a journal or went to a library to read a journal that was subscribed

to by the library. Authors purchased reprints from the journal so that they could fulfill requests for copies of their publications. Then the copy machine was invented. Instead of purchasing the intellectual property from the copyright holder, people went to the library or to a colleague who subscribed to the journal and made a copy of the needed article. Most people do not realize that this is theft of intellectual property through breach of copyright. A case ended up in Federal court and ultimately resulted in the fair use doctrine in the U.S. Copyright Act.² Equilibrium was restored.

Then the World Wide Web came into being. The Web is just like the copy machine, only faster, cheaper, and more easily accessible. If an article is posted on the Web, it is accessible by anyone in the world who has access to the Internet, not just people who have a colleague or access to a library that has a subscription. Yes, multiple copies can be made by a copy machine and distributed far and wide (in violation of U.S. copyright), but an electronic copy can be sent many times faster and cheaper by e-mail. It is the ubiquity and low cost of worldwide instant access that have prompted the patrons of the research community to rethink their position on the entire enterprise.

The publishing industry is changing and finding a new equilibrium. There are many new economic models including open access and electronic-only access. The traditional paper-and-postage model appears to still be viable. But one thing is clear: open is not free. The current assumption is that if readers of the journals do not pay a subscription fee, either directly or indirectly, then the authors of the journal articles will have to. Perhaps the funding sources will prefer this latter model as it allows the cost of publication to be included in the cost of doing the research. But such author fees will raise the cost of doing research. The evolving experiment continues.

Having raised the ire of the publishing community, I will now take on the research and clinical communities. Besides money, the basic commodity of research and clinical medicine is data. Research experiments are done in order to capture data in an effort to prove a new theory. Clinical medicine data are obtained from patient laboratory tests in order to be more certain about a diagnosis. What is done with the data after the initial purpose is served? The information is put in the file, either paper or electronic, for possible future reference. Is there not a second life for data after the initial purpose has been served?

This is what “open data,” another component of the open science model, is all about. The argument parallels

*Assistant Director for High Performance Computing and Communications, National Library of Medicine, Building 38A, Room B1n-30, 8600 Rockville Pike, Bethesda, MD 20894; e-mail: ackerman@nlm.nih.gov.

This article was written by the author in his private capacity. No official support or endorsement by the National Library of Medicine is intended or should be inferred.

Copyright © 2006 by Greenbranch Publishing LLC.

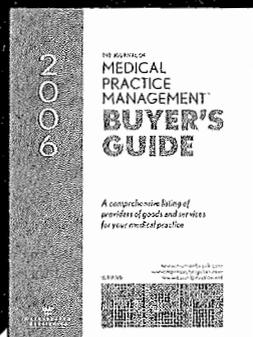
the argument for open literature. The data are not really owned by the researcher or clinician. Data are owned by the person who paid for the data collection, i.e., the granting agency or the patient. If the granting agency requires or the patient allows, depersonalized or aggregate data could be shared and used for other purposes. Meta-analysis studies based on published literature are a distant attempt at the reuse of data. Its often-fatal flaw is that it is really a reuse of synthesized information not original data. Open data would allow these researchers to re-analyze the original sources in order to uncover more universal findings. Without these data, findings are inferred leading to controversy and the requirement for additional

experiments to resolve the controversy. Aggregated patient data might lead to new insights in the current state of the public health:

But remember, open is not free, and both clinical and research data have monetary value to their holders. Just as in the case of open literature, open data gets back to economics. More on this and the third component, open software, next time. ■

REFERENCES

1. Shreeve J. Free radical. *Wired Magazine*. June 2006;14:136-143.
2. U.S. Copyright Law, Section 107, Fair Use; www.copyright.gov/fls/fl102.html.



Order extra copies of the MPM Buyer's Guide

- 1 copy \$49.95**
- 2-3 copies \$44.95 each**
- 4-5 copies \$39.95 each**
- 6 or more copies \$35.95 each**

Yes! Send _____ copies

Total amount due: \$ _____
(price includes shipping/handling)

Payment Enclosed
(checks payable to *Greenbranch Publishing*)

Charge my:
 Visa MasterCard American Express

Account # _____ Exp. Date _____

Name

Company Name

Address (no PO boxes, please!)

City/State/Zip Code

Order by Mail:
 MPM Buyer's Guide
 Greenbranch Publishing
 PO Box 208
 Phoenix, MD 21131

Order by Fax: (410) 329-1510

Order by Phone: (800) 933-3711