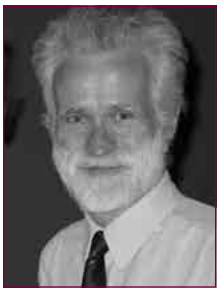
Open Access to Professional Information

This IMS Presidential Address was given by Jim Pitman at JSM in Salt Lake City.



In the Science–Technology–Medicine publishing world there has been rapid growth over the last 20 years, with a market growth rate of about 8 per cent per year in the 1990s, and astonishingly high profit margins: over 30 per cent. Worldwide, the Science–Technology market is worth around US $4 billion per year. There is a strong trend towards consolidation, with aggressive entry into the market of private equity houses. For example, in 2003, Candover and Cinven acquired Kluwer Academic for 600 million Euros, and BertelsmannSpringer for 1.1 billion Euros. The resulting consolidation, now called "Springer", headed by Derk Haank (former Elsevier CEO), publishes over 1,000 journals and 5,000 book titles, and the revenue approaches a billion Euros per year. Earlier this year, Wiley acquired Blackwell for $1.08 billion.

So a few large publishers (Springer, Wiley-Blackwell, Elsevier, Taylor & Francis) now take more than half the total market revenues. The rest is divided between a large number of smaller publishers: societies, university presses, and so on.

The transition from paper to electronic format has been a windfall for established publishers. Authors now submit in close to production format, there are decreased production and distribution costs, more sales to libraries rather than individuals—and yet no price reduction.

There has also been a shift in responsibilities, as publishers take over archiving responsibility from libraries: subscribers don't get copies but pay for licenses to view; publishers control access and linking systems, which tend to trap the user in the publisher's website.

There is a trend towards databases, with the journal becoming less important as a unit than when it was a physical volume. These databases include publisher silos like ScienceDirect and SpringerLink; full-text database aggregators (EBSCO, ProQuest, etc.); portals for abstracting and indexing (ISI Web of Science, MathSciNet, ACM Guide, CIS); and content hosts (Ingenta/Vista, Project Euclid).

Scholarly Communication Crisis

The term "crisis" became popular with librarians a few years ago. Better terms would be "struggle" or "war": we can expect this to continue for decades. But what does this struggle involve?

- Loss of access to the scholarly research literature, as the rising costs of journal subscriptions exceed institutional library budgets.
- Loss of library funds for books due to explosive increase in the journal and database sectors.
- "Big Deals" (SpringerLink, ScienceDirect, etc.), which are good for big publishers and big library administrators, and a force behind mergers in the publication industry, but bad for small publishers, small libraries and the academic community, because they reduce academic control over journal selection, amplify inequalities between institutions/countries, and shut out the broader community.

Most librarians think the current system is dysfunctional and unsustainable.

To quote Carol Kaesuk Yoon ('Soaring Prices Spur a Revolt in Scientific Publishing', *New York Times,* December 8, 1998): "In fact, researchers say, academia is a paradise for publishers. First the public pays for most scientific research through, for example, the National Science Foundation. Then universities pay the salaries of scientists who do virtually all the writing, reviewing and editing. Universities sometimes even provide free office space to journals. Finally, authors typically sign over their copyright to publishers, who can sometimes bring in many millions of dollars a year in subscriptions for a single high-priced journal—subscriptions paid by university libraries supported by tax dollars and tuition."

Encouraging signs

There are some encouraging signs, however. There are free and open source software and licenses (e.g. the GNU Project, LaTeX, R Project for Statistical Computing, Creative Commons, and many others); StatLib open data and software; free electronic journals (including, since 1996, the journals *EJP* and *ECP* now supported by IMS/Bernoulli); PubMed; Public Library of Science; open content collaboration such as Wikipedia; and the Google Scholar index.

Of particular interest to IMS members is arXiv.org: created in 1991 by Paul Ginsparg, and now at Cornell, arXiv is a successful eprint service in physics, mathematics, non-linear science, computer science, quantitative biology, and now, statistics (arXiv/stat opened April 2007, supported by IMS and Bernoulli Society). ArXiv demonstrates the separation in electronic media of two functions of "publication" (making public): first, communication and archiving, and second, peer review and certification. The 'Physics culture' is now permeating to other fields, with authors depositing on arXiv at time of journal submission and with no discernable loss of journal subscriptions on account of arXiv. Almost all publishers now tolerate arXiving; authors can insist on it. Currently,

arXiv offers alerts, search, and indexing; long-term, IMS plans to promote arXiv/stat as a community-supported repository with associated services (integration with CIS, author name authority, departmental listings, etc.)

Business models

There are now two competing business models: the traditional Gated Access model, supported by library subscriptions; and the Open Access model, supported by some combination of author fees, sponsorship and advertising. IMS has adopted a mixed model: by assisting authors to post their work on arXiv while still retaining journal subscriptions, by allowing all journal content to become open after a suitable delay, and by supporting a mixed portfolio of gated and open journals.

Open Access

The international Open Access (OA) movement now supports the availability of electronic content free of charges and restrictions. Supporting organizations include arXiv, PubMed, PLoS (Public Library of Science), and Creative Commons, which promotes the use of licenses suitable for OA publishing. Such licenses are compatible with copyright, peer review, revenue, print, preservation, prestige, career advancement, and indexing.

Open access increases the impact of scholarly work by making it accessible to the widest possible audience; it facilitates knowledge transfer between different educational levels, or countries, or subjects; it takes full advantage of the internet for search, access, navigation and organization; and it allows small and medium-sized publishers to compete with multinational corporations.

There is significant institutional support for OA from major libraries, NIH, the British Government, Wellcome Trust, Max Planck Society and others; but also significant institutional opposition from, for example, the Association of American Publishers and the UK Royal Society.

Open Access attiudes

Attitudes towards OA vary widely. Harold Varmus (PLoS) believes OA is so important that scholarly societies should not support their activities by gated access. In contrast, John Ewing (AMS) holds that OA is a fad which distracts us from the real issues: Big Deals by avaricious publishers, hoarding of the historical archive, faulty application of usage statistics, and version control for electronic publications. While I agree these are important issues, I regard OA as an ideal, like freedom, which I hope can rally the academic community to protect its knowledge base from commercial control.

I regard Open Access as an ideal, like freedom, which I hope can rally the academic community to protect its knowledge base from commercial control

In more practical terms, OA achieves society missions with minimal overheads, is attractive to most authors, and encourages cooperation between societies. A typical scholarly society must balance two imperatives: to achieve its mission of broad dissemination of professional knowledge, and to stay financially viable. Following the success of commercial publishers, many societies have overbalanced towards restriction/commercialization of their knowledge base.

Raym Crow, writing for the Scholarly Publishing and Academic Resources Coalition, has recognized structural constraints which limit societies' performance and explain why commercial publishers largely own the market: a low tolerance of risk, lack of business expertise, insufficient market leverage, and undercapitalization. Commercial publishers have exploited these constraints by the creation of new journals and the acquisition of society titles (*JRSS*, *Scand. J. Stat*, etc.).

Coalitions of scholarly societies and publishing cooperatives offer ways to work around these constraints. Examples are Project Euclid, JSTOR, and some recent IMS collaborations.

Many authors prefer their articles to be made available with open access. But some choose otherwise for various reasons, such as the prestige of restricted outlets, or a desire to get full credit for work while partially withholding it to maintain some advantage in research competition. This attitude was encapsulated for me in a fragment of conversation I heard last year passing by a Berkeley coffee shop. One academic, confidingly, to another: "You know, I really don't like sharing my best ideas with the broader community". I rather prefer Thomas Jefferson's attitude to ideas: "He who receives ideas from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me."

Many of you may know the story about Richard Stallman and the printer program, which led to the GNU Project and copyleft software licenses. (The story can be found in Wikipedia, thanks to the GNU Free Documentation License.) A similar incident led to my own commitment to create open access electronic outlets for expository material in probability and statistics.

To describe the incident I should first

Continued on Page 14

Open Access *continued*

provide a little background. By age 20, I had learned the asymptotic distribution of the Kolmogorov-Smirnov statistic from reading my father's monograph on statistical inference. Since then I have worked on analytic and combinatorial models whose asymptotics are described by functional limit theorems featuring a Brownian bridge in the limit, and on the distribution of various functionals of the Brownian bridge. A few years ago I was invited by editors of Wiley's *Encyclopedia of Statistical Sciences* to revise and expand the entry on Brownian bridge. As the intention was to create an online version of the encyclopedia, I inquired what, if any, rights I might retain to publicly post an electronic version of the article. The answer was none, not even the right to post the article on arXiv or on my own homepage.

I was sufficiently shocked by this that I emailed about 70 senior probabilists and statisticians, asking whether they supported this kind of commercial control of professional information, and if not whether they would support the creation of an open access outlet for expository articles in probability and statistics. I got about 50 replies by return email indicating strong support for an open access alternative. IMS and the Bernoulli Society then quickly supported creation of the pair of open access journals *Probability Surveys* and *Statistics Surveys*.

Much work remains to be done to create an open access equivalent of Wiley's Encyclopedia. But suitable open linking infrastructure for online encyclopedias can already be seen in place in Wikipedia, MathWorld and PlanetMath, and it seems only a matter of time before such functionality is available with all entries validated by scholarly societies.

IMS's role in promoting OA

In addition to the five IMS copyedited subscription journals (*Annals of Statistics, Annals of Probability, Annals of Applied Statistics, Annals of Applied Probability*, and *Statistical Science*), we have five Open Access journals, jointly with the Bernoulli Society: *Electronic Journal of Probability, Electronic Communications in Probability, Probability Surveys, Electronic Journal of Statistics*, and *Statistics Surveys*. We also have four more affiliated and co-sponsored publications: *CIS*, *JCGS*, *Bernoulli*, and *ALEA*, and of course, this *Bulletin* is available OA electronically.

IMS plans to collaborate to develop further OA journals, and to provide OA to biographies, bibliographies, reviews, images, celebrations and memorials, collected works, glossary and encyclopedia entries, interactive graphics, and teaching materials

Recent IMS Initiatives

IMS has created cooperative publishing agreements with various partners — the Bernoulli Society, StatLib, l'Institut Henri Poincaré, APS-INFORMS. IMS seeks further such cooperation through a new framework for Affiliated and Supported Journals and Societies. For details, see the IMS website, or contact the IMS Exec.

IMS–Bernoulli cooperation: This includes our joint membership agreement; next year's 7th World Congress in Probability and Statistics (Singapore, July 14–19, 2008); the production and marketing by IMS of the journal *Bernoulli* (editorial control remains with the Bernoulli Society); the co-publication of the five OA journals listed above, including the newly-launched *Electronic Journal of Statistics* and *Statistics Surveys*.

IMS–StatLib cooperation: By agreement with the Carnegie Mellon Statistics Department, IMS will use StatLib as a repository of OA supplementary materials associated with IMS journals, such as data, graphics, software, documentation, or additional proofs or appendices. This will be subject to technical data standards and review by editors and referees. The electronic medium allows authors the option to present much more than is possible in the traditional journal article format. The first supplementary material is associated with the launch of a new IMS Journal, *The Annals of Applied Statistics* (*AOAS*).

IMS–Institut Henri Poincaré cooperation: The journal *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* is currently produced by contract with Elsevier, but is switching to IMS production starting in 2008. This is a fine example of cooperation between professional organizations to reduce the influence of commercial publishers.

IMS–APS–INFORMS cooperation: APS–INFORMS is the Applied Probability Society of INFORMS, the Institute for Operations Research and the Management Sciences. IMS–APS–INFORMS have a joint membership agreement, and a new agreement to develop an OA electronic journal in the area of applied probability and operations research, whose focus will be on serving the interests of APS members and the broader applied probability community. This is an exciting development: a small society seeking recognition and community support through OA publishing, with assistance from IMS.

IMS's Future Plans

In future, we plan to collaborate to develop

further OA journals, and to provide OA to more publication types, such as biographies, bibliographies, reviews, images, celebrations and memorials, collected works, glossary and encyclopedia entries, interactive graphics, and teaching materials.

We also plan to organize and index such content to create and maintain a high quality OA web of professional information in statistical science. We hope to engage society members as contributors and curators, and distribute the maintenance problem over numerous societies with a common web infrastructure.

To achieve this, we need technical and administrative resources, for which we have launched the **IMS Open Access Fund**. This is a new IMS fund to which individuals and organizations can donate, dedicated to the support of existing and new OA ventures which further the IMS mission. We also need human resources: volunteers to help guide development, especially those with expertise in software development, and volunteers for editorial service.

Conclusion

The academic community is engaged in a struggle to create and protect an open environment for scholarly communication.

As researchers and administrators — by your choice of publication outlets, by your choice of data formats and software systems (open or proprietary), by how you support the editorial system of peer review, by how you direct your professional societies — your actions affect the system of access to professional information in our field.

So I encourage you to consider the effect of these actions as you make them, and I invite you to join IMS in building a coalition of professional organizations committed to open access to professional information in statistical science. ■

How can *you* promote OA?

Readers:

Join professional societies like IMS which work to provide you and others with open access to high quality scholarly information, and provide further services as membership benefits.

Encourage your librarians to subscribe to journals published by societies supporting OA.

Encourage your departments and universities to support OA publication.

Acknowledge the value of high quality OA publications in promotion cases

Authors:

Preferably submit your articles to society-run journals with copyright agreements which allow self-archiving of final versions on arXiv; open access to publisher version, at least after some delay; and re-use of content in derivative works

Wherever you submit…

post a copy on arXiv at the time of submission,

don't sign restrictive copyright agreements: amend them to retain the right to post the final version of your work on arXiv or other open access repository,

maintain your publication list on your own website (preferably machine readable e.g. bibtex),

provide links to full-text whenever possible,

digitize and post your old work on the web, and look for more stable repositories than your own website

Editors and Referees:

Refuse to work for journals with overly restrictive copyright policies.

Work for society-run journals which promote OA publication

Work to raise the standard of OA journals to be more attractive to authors than commercial journals.

Create alternatives (like the Editorial Board of Elsevier's *Topology,* which resigned en masse in December 2006; the first issue of the London Math Society's *Journal of Topology* is scheduled for January 2008.

Librarians:

Purchase society-run journal titles and resist Big Deals

Index OA sources, support OA digital repositories

Department Chairs and Deans:

Recognize the value of high quality OA publications in personnel review cases

Increase the visibility of your departments and schools by encouraging OA posting of eprints, and encouraging faculty to list their publications online with links to full text

Higher University Administrators:

Support creation and maintenance of OA repositories

Encourage posting of all research in OA repositories

Install systems for indexing and displaying of your school's research output, such as the Duke Faculty Database System

Society Administrators

Encourage your members to use OA repositories

Develop open indexes and aggregations to showcase your members' work

Find ways to maintain fiscal viability without undue access restrictions to journal content, as IMS has done.

Seek efficiencies of scale by cooperating with other societies on OA publication ventures

Do get in touch if you would like to help!