

# HOLDING MOONBEAMS: THE CHALLENGE OF PRESERVING SCIENTIFIC KNOWLEDGE



*Richard R. Rowe*

*Preserving the scientific information in its new unfamiliar forms is a problem that must be solved and the Biomedical Archives Consortium is offered as a solution. Based on a global network, BMAC would focus on biomedical knowledge but would provide a model for maintaining accessible archives in other subjects. Unfettered access is the aim of this open consortium, with the concept of Copyright Expiration Date (CED) a key objective.*

*Dr Richard R. Rowe  
Chairman, RoweCom Inc,  
15 Southwest Park,  
Westwood, MA 02090  
Tel: (781) 410 3302  
E-mail: rowe@rowe.com*

Much of the recent debate over free scientific information has revolved around access to current information that is central to human welfare such as human genome discoveries, AIDS research, and genetically modified organisms. However, archival services, providing unfettered access to the terabytes of 'old' literature that constitute the foundations of today's knowledge, increasingly present their own challenges.

Many will recall from the musical, *The Sound of Music*, the problems the convent sisters had with their young novice, Maria. She did not fit the traditional mode of a nun; she was very difficult to control. The Reverend Mother puzzled over "How do you solve a problem like Maria? How do you catch a cloud and pin it down?". Finally, almost at a loss, she sings, "...How do you hold a moonbeam in your hand?".

Scientific knowledge today is, in many ways, akin to Maria. It is coming in new, unfamiliar forms, out of the traditional modes that we know how to manage. Managing such knowledge is a problem that we have not yet solved. The emergence of e-publications, for which there are no established archiving traditions and where the formats themselves are unstable and likely to change over time, challenges us. E-resources are inherently heterogeneous, interactive and iterative in nature. Valuable information may slip bypass traditional publishing route through an email or a videoconference. Such information tends to be less fixed in time and form than traditional publications. In addition, current copyright laws continue to limit legal access to archival literature. The increasing volume, velocity and variety of information resources gushing from the digital fire hose is beyond any one institution's ability to capture, sort out and store, let alone make accessible.

Maria was sent off to do childcare with the Trapp family. We

can also choose to send our "Marias" off for someone else to manage. Or we can add new canons to our services and accept the challenges involved in managing these changes ourselves.

The scientific community needs access to quality information in perpetuity and expects it to be filtered, catalogued, stored and made accessible to researchers on demand. Universities and governments today assume the lion's share of maintaining the world's important knowledge in their libraries. Each institution maintains its own collections and finds room, somewhere, to store them, as they age. Most institutions rely upon themselves for the majority of their archival resources. Traditional interlibrary loan services, covered under 'fair use' policies, have limited value since they are slow and expensive. Also, physical possession and ownership generates a sense of institutional security and prestige. However, those institutions can no longer fulfil that mission by themselves. As the volume of information continues to rise, the sheer cost of archiving and providing access to archives has become unsustainable for any one institution.

In order for scientific information to have lasting usefulness, we must address two basic questions: What should we keep? How can we make it accessible? Technical challenges abound regarding the preservation of and access to digital archives that increasingly involve multiple media. But technical solutions necessarily rest upon the professional, economic and public policies that affect the archiving of information. We must respond creatively to the opportunity presented by these policy challenges.

In this fast-moving world, information ages quickly and the aggregate demand for 'old' information is relatively small. However, it is essential for us to preserve such information as may be needed sometime in the future. It therefore seems reasonable to pool resources in order to maintain our growing knowledge archives. To do so, we must modify our assumptions and practices concerning 'old' intellectual property, and allow aging knowledge to devolve into the public domain. Authors and publishers should be able to agree upon a limited period for copyright of scientific research literature, after which scientific information would become a public resource.

There is increasing support for this concept.

Recently, interest has focussed on what is an appropriate period for copyright protection of scientific knowledge. The Public Library of Science ([www.publiclibraryofscience.org](http://www.publiclibraryofscience.org)) has suggested six months. Some publishers have argued that one fixed time period is not appropriate, since the "half life" of scientific articles varies substantially from one field to another. The issue is primarily an economic one. Historically, virtually all of a publisher's revenue has come from the original sale of its journals. The argument that many institutions would simply wait six months to get access to a scientific article rather than continuing to purchase journals as they do now is questionable. Timeliness is far too important a variable these days. It would be useful to analyze the use patterns over time of digitally published articles, by discipline. Our common goals should be to make scientific information readily accessible to all those who need it, within a reasonable period after it has been published, and to do so in a manner that does not seriously undermine the financial base of scientific publishing.

No matter what the length of the original copyright may be, we need to begin using the Copyright Expiration Date (CED), as a standard eight-digit XML tag attached to each scientific article, to enable us to know when any particular article is unfettered by copyright. Authors could attach a simple "copyright expiration provision" as a condition of their assignment of copyright to publishers, in order to ensure that their articles would be in the public domain after the agreed-upon period.

### **Traditional archival models**

There are several models designed to address the problem of what scientific information should be stored and how to make it economically accessible:

*Decentralized institution-based archives.* This is currently the most prevalent model but, as the volume of print and electronic-based knowledge increases exponentially over time, the sheer cost of this model will make it unworkable for even the largest of institutions, let alone the smaller ones and those on the wrong side of the digital divide.

*Publisher-based solutions.* Historically, most publishers have not archived their material in an

accessible manner – that has been the role of libraries. The additional cost to publishers of archiving their publications would constitute a significant financial burden that would have to be paid by someone and, finally, there is no guarantee that any publisher will exist in perpetuity. In at least one case a publisher has asked a university to maintain its archives. However, if one accepts the proposition that knowledge should be organized and preserved by discipline rather than by publisher, a maximally useful archive will have content from multiple publishers. Thus a publisher-based solution has limited value.

*Mega centres.* There is much discussion about building large centralized digital archives. This is a logical extension of the 'national library' concept. The International Congress of Scientific and Technical Information has explored digital archiving issues ([www.icsti.org/icsti.forum/33](http://www.icsti.org/icsti.forum/33)) and ([www.icsti.org/icsti.forum/35](http://www.icsti.org/icsti.forum/35)) giving particular attention to this model. JStor, an ambitious and popular non-profit journal archiving initiative, is another effort to create a "mega centre." The U.S. National Library of Medicine and the Canada Institute for Scientific and Technical Information (CISTI) are in many respects de facto biomedical mega centres. This model is often deeply dependent upon government or foundation funding. The model has significant economies of scale, especially for low-demand information.

However, since the value of information is almost purely "contextual," this model has serious limitations as a truly comprehensive archive for any particular scientific discipline, particularly in the digital world. What is important for one field of inquiry may be irrelevant to another and vice versa. The decision to 'keep' or 'throw away' should be done within the framework of a particular body of knowledge. Furthermore, the absolute quantity of information that today is being generated on a global basis soon will be far beyond the capacity of any one organization to maintain effectively. We must learn how to share the responsibility of maintaining and providing access to the world's knowledge.

*Centres of excellence.* This model relies upon libraries, which from their beginning have maintained print archives, extending their

archiving functions to digital resources and focusing upon a particular area of expertise. It calls for libraries to share the task of maintaining accessible archives of important knowledge in a collaborative global network. Library centres of excellence throughout the world currently have in-depth archives in their fields of specialization. In many cases, they could readily provide these archival services globally. There could be multiple, redundant library centres for the most important areas of scientific knowledge, so that we would not be dependent upon only one such centre. These services could be available through the internet on a paid subscription basis, thus enabling the archival centres to offset the costs of maintaining high quality services on behalf of the world as a whole. Such services would provide access to both the print and the digital resources in a particular discipline.

### The BMAC proposal

We propose to create a BioMedical Archives Consortium (BMAC) dedicated to ensuring unfettered access to biomedical archival knowledge. BMAC would create a global network that combines the benefits of both the mega-centres and the centres of excellence. Although we are proposing to begin with a focus on biomedical knowledge, the model could readily be applied to other scientific and scholarly fields.

Ten to twelve mega-centres would assume the responsibility of maintaining archives for the core biomedical literature that is no longer current. These mega-centres would maintain "mirrored" copies of the archives and provide subsidized access to those publications for any institutional or individual member of the consortium.

In addition, BMAC would promote the development of a global network of specialized biomedical library centres of excellence. Such centres would have full access to the mega-centres for the core archival literature and would augment that core literature with value-added information. This would include such resources as grey literature, papers from meetings and conferences, and active databases related to that centre's area of expertise. We envision there would be hundreds of such specialized centres networked throughout the world.

BMAC itself would be a standard-setting

co-ordinating body. It would have several functions:

- establish technical and organizational standards for biomedical archives;
- identify, certify and de-certify mega-centres and centres of excellence;
- ensure 'optimal redundancy' in the global network of biomedical archives;
- maintain a meta-catalogue of global biomedical knowledge resources.

The BMAC would be an open consortium supportive of the numerous existing and planned initiatives for providing widespread online access to biomedical archives. It should be global in scope and could be an independent body, or could work under the auspices of an existing international body such as the World Health Organization.

The consortium would provide access not only to digital archives but would also maintain comprehensive catalogs of the full range of biomedical knowledge, including paper-based materials that could be requested on demand.

With BMAC in place, institutions could provide their members with unfettered access to extensive bodies of biomedical knowledge. They would have access not only to the traditional core literature but also to the more esoteric and idiosyncratic resources available from the specialized centres of excellence. Medical schools, research centres and hospitals would no longer need to use valuable space and time storing old publications, at significant cost, 'just in case' at some point they might need access to them.

The BMAC model would rely upon annual subscription revenues from institutions and individuals. Annual fees would be based on the size and nature of the institution. Supplemental fees would be charged for non-standard services. This economic model would enable the centres to improve their services, including using the most advanced technologies to provide hyperlinks to related information. In some cases, public subsidies would be needed in order to increase the quality and quantity of the service and to provide unfettered access to users. However this

model would, to a significant degree, reflect the supply-and-demand characteristics of biomedical research and practice.

The "moonbeams" that flow through today's digital knowledge networks will not be easy to capture and preserve for another day. Providing access to such knowledge over time constitutes a new challenge, requiring new ways of thinking about our roles as experts in managing knowledge. As Mother Superior, we need to change our canons. We must learn how to demand less direct control over content and how to facilitate more open and interactive knowledge networks that behave, like Maria, in unexpected and exciting new ways, adding value to the quality of our work and our lives.

An initial meeting to explore this collaborative model for biomedical archives, with representatives of professional societies, publishers and libraries participating, was held on 19 August 2001 at the MIT Press in Cambridge Massachusetts, coinciding with the International Federation of Library Associations meetings that week in Boston. At that meeting, it was agreed that the proposed BMAC model was worth further exploration with an expanded group that should include medical researchers and practitioners, as well as librarians and publishers.

A second BMAC planning meeting is scheduled for January 20, in New Orleans from 2 p.m. to 5 p.m. in conjunction with the American Library Association's mid-winter conference. All persons interested in exploring and developing the BMAC proposal further are invited. For more information on the agenda and location of the meeting, go to: [www.biomedarchives.org](http://www.biomedarchives.org).

#### Links:

1. BioMedical Archives Consortium  
<http://www.biomedarchives.org>
2. International Congress of Scientific and Technical Information <http://www.icsti.org>
3. JStor <http://www.jstor.org>
4. Public Library of Science  
<http://www.publiblibraryofscience.org>
5. RoweCom, Inc. <http://www.rowe.com>