

“International Dimensions of Digital Science and Scholarship”
Address to the Association of Research Libraries
and the Canadian Association of Research Libraries
May 18, 2006, Ottawa, Canada

Deanna B. Marcum,
Associate Librarian for Library Services
Library of Congress, Washington, DC

I am delighted to be in Ottawa today, with Canadian as well as American colleagues. Our collaboration in these meetings, so ingeniously begun many years ago, is worthy of celebration. I predict that it will become more so in the future for the simple reason that research libraries in all countries now have more to offer each other—and each other’s patrons—than ever before. That is because digital information technology and Internet communications enable each of us now to contribute beyond our own walls, even beyond national boundaries, to the stock of scholarly resources of the entire world.

Under the theme for this conference—*International Dimensions of Digital Science and Scholarship*—I have been asked to talk about what the Library of Congress in the United States is doing to abet research across national borders. I am delighted to do that because a report has just come to me about collaborative work—work that involves the Library of Congress and indicates what can be done.

1

In July of 2005, the Foundation for Science and Technology, an organization in the United Kingdom, organized a meeting at the Royal Society. The meeting featured a report containing information that a British group had busily been gathering in both the UK and the United States. Members of the group had met in Washington, DC, with

senior staff in a half-dozen of our largest federal research-funding agencies—the National Institutes of Health, the National Aeronautics and Space Administration, the Department of Energy, the National Science Foundation, the Department of Agriculture, and the Department of Homeland Security. These agencies support research-and-development, largely in science and technology, with budgets that in 2005 collectively totaled 56.7 billion US dollars. The British group also met with presidents and provosts of a dozen large US universities represented in our Association of Research Libraries.¹

The British group also met with UK counterparts of the Americans it had visited. The group gathered information from research universities, government departments, research councils, and others in the UK that support research.

What was the group looking for?

I'll get to that, but not before I describe something else it did. The group engaged a British research firm to make a study described as “bibliometric analysis.” That means, first, that the analysts identified research papers that researchers in the US and researchers in other countries, particularly the UK, had authored *jointly*. Then the analysts discovered a couple of things about these collaborative papers. They found that, in the previous five years, collaborations of researchers in the United States with researchers in the United Kingdom had grown faster than US collaborations with researchers in any other G8 country. They also found that publications jointly authored by researchers in the UK and the US had what the report called “a significantly greater impact factor” than did papers produced by researchers in only one country or the other. The measure of “impact” in the study was the average number of citations to each paper by other researchers.²

Well, you will say, that's to be expected, which, in fact, is what somebody in the British group declared. After all, researchers who have the "resources and motivation to overcome difficulties in collaboration over a distance" tend to be the strongest scholars and scientists, so of course their papers will be most often read and cited by others. Moreover, international research projects are likely to be larger than others, and papers about them will attract more citations because of their more prominent profiles.

In response to that, the bibliometricians refined their analysis. They looked at papers in a couple of major scientific journals in which only the most prominent researchers could be expected to publish. And they discovered that papers in those journals that were jointly authored by researchers in the UK *and* the US, received two-to-three times more citations, on the average, than papers in the same journals by UK authors only. "Therefore," they concluded, "UK–USA collaboration does appear to add value, with collaborators combining their talents to achieve benefits they could not have done alone."³

Similar analysis involving papers by researchers in other countries further showed "that UK researchers are not alone in having US co-authors listed on a high proportion of their most highly cited papers." That is because US expenditure on research and development accounts for 37 percent of the R & D expenditure of the entire world. The group reported that "the introductory phase of our study confirmed that the USA, as the world's largest research economy, is the preferred partner for international research partnerships and makes a significant contribution to the leading-edge performance of collaborating nations." US funding helps researchers from abroad take part in top-level projects—and helps the projects attract top-level foreign researchers. Thus everybody

wins. The British group concluded that international collaboration with the US on research is to be encouraged.⁴

The British group then went to work on recommendations about how to make the UK's international research partnership with the US grow. The group summarized its recommendations in a report drafted just last March. Promoting transatlantic research, said the group, does not require "expensive artificial incentives to collaborate." Instead, it could be achieved with a modest degree of help to overcome "the natural obstacle [of] distance," in the report's words, and such "artificial ones," as "lack of information, different funding systems," and "outdated perceptions."⁵

In fact, in some areas, the report noted, significant cooperation between scientists in the UK and the US already exists: "Special agreements are in place to secure strong cooperation" in research for defense, homeland security, and other "sensitive and strategic fields of research." However, the group found that researchers in transatlantic projects have difficulty getting funding agencies in their respective countries to provide bilateral support. Projects dependent on funding from both sides are in danger of what the group called "double jeopardy." Consequently, the group recommended the establishment of protocols between UK research councils and US funding agencies to help researchers of both countries who want to collaborate "in areas of shared priority."⁶

Such an agreement was signed last November between the US National Science Foundation and the UK's Economic and Social Science Research Council. The foundation also has announced an agreement with a German funder for cooperative work in chemistry between US and German researchers. To support more international

collaboratives, the foundation has instituted a program called “Partnerships for International Research and Education.”⁷

Also, the Engineering and Physical Science Research Council in the UK and its Department of Trade and Industry are collaborating with the US Department of Energy and our Sandia National Laboratory in an exchange program for postgraduate scholars engaged in research on hydrogen technologies. The Economic and Social Sciences Research Council in the UK is collaborating with the Social Science Research Council in New York to fund scholarly exchanges between the two countries. And approximately 100 researchers from the UK are receiving postdoctoral fellowships to work at laboratories at the National Institutes of Health in the US. NIH also funds a Health Science Research Scholars program, which enables postgraduate students from the US to do biomedical research at Cambridge and Oxford universities. The British group’s report calls for more programs “to assist the inward and outward mobility of new and young researchers.”⁸

Also of great importance, the British study recognized that technologies for digital reproduction and Internet access can give a boost to research collaborations. In the words of the report, “The extensive digitization programmes currently underway, together with the development of efficient search engines, etc., will considerably enhance the potential for more research collaboration . . . between academics in the UK and the USA,” and in fields besides engineering and science. Research links created between Cambridge University and the Massachusetts Institute of Technology are cited by the report as examples of how, in its words, “large infrastructure projects and others in the arts, humanities, and social sciences are being shared across the network to good effect.”⁹

Conversely, however, the study also found that international collaboration is now “hindered by the lack of research materials available in digital form.”¹⁰ That may seem odd to those of us familiar with all the digitized books and other electronic resources available on the Web from the Gutenberg and Perseus projects, the “Making of America” project of Cornell University and the University of Michigan, the “American Memory” project to which many of you have contributed at the Library of Congress, the JSTOR aggregation of digitized scholarly journals, and the digitization projects at the University of Virginia, the University of Toronto, and most if not all of the rest of the Canadian and American libraries represented in this room, along with many large libraries elsewhere in the world. But you also know that none of us has digitized more than a fraction of the totality of our collections.

The British report calls for more digitization of both primary and secondary sources to “underpin collaborative research.” The need would be not just to build internationally usable collections but also “to build effective and sustainable virtual research communities.” The British study group believes the following:

While a variety of digitization projects are underway in both the UK and the USA, there has hitherto been little coordination between them. There is an opportunity to build a truly colossal and cross-searchable transatlantic database that would open up many exciting new avenues of collaborative and comparative research.¹¹

Perhaps that is what we will eventually get from such mass digitization projects as the Google Book Search library and the Open Source Initiative, in both of which several of you are involved. These projects will digitize at least parts of libraries in the UK and Europe as well as in Canada and the US. And work to facilitate scholarly use of digital aggregations is going forward in such organizations as the Digital Library Federation,

which now includes two British institutions in its primarily American membership. However, the report was talking of projects to digitize materials needed to support specific international projects of research.

Moreover, the report points to a major omission in all this: The national libraries in the United States and the United Kingdom—that is, the Library of Congress and the British Library—though engaged with each other informally in such international ventures as the Internet Preservation Consortium, have “no formal links” for collaboratively facilitating transatlantic scholarship.¹²

What exactly does this mean? It means, explains the report, that much more could be done to develop scholarly exchange programs in the two national libraries, primarily by “developing a critical mass of resident fellows,” and by “enhancing creative interaction between visiting scholars and research-active curatorial staff.” Also, the report says, both the British Library and the Library of Congress are “well positioned through their national roles and professional skills to link into other research libraries, archives, and data centers in their respective countries.”¹³

The report also notes, however, that new collaborative work is beginning. With encouragement from funding organizations in both the UK, and the US, including our National Endowment for the Humanities, several things are happening:¹⁴

- The Arts and Humanities Research Council and the Economic and Social Sciences Research Council, both agencies of the UK, have agreed to provide funds for as many as twenty British scholars each year to spend up to nine months doing research in the Library of Congress. The program began in April with six existing award holders and will become fully operative next year.

- The Library of Congress will absorb space and curatorial costs for the visiting British scholars, who will be encouraged to interact with other scholars from abroad in the Library's Kluge Center, which accommodates a program for enabling senior scholars and postgraduate fellows to use the library's resources.
- The British Library has agreed to develop its own center for visiting scholars, and to organize workshops for and exchanges of scholars on research topics about which the two national libraries have major collections.
- The Gatesby Charitable Foundation in the UK will provide a substantial grant to refurbish two floors in one wing of the Jefferson Building of the Library of Congress as a center for scholars from the UK.
- Through the Joint Information Systems Committee in the UK, funding is being provided for a five-year collaboration by the Library of Congress, the British Library, and other organizations to digitize resources for scholars, beginning with newspapers, sound recordings, official records and publications, and other resources of a documentary nature.

The Arts and Humanities Research Council in the UK intends to support an academic program both to inform and to encourage the digitization work. Bruce Cole, the chairman of the U. S. National Endowment for the Humanities will lead the US side of a transatlantic steering group for the overall initiative, and Clive Field of the British Library will chair the joint digitization activity.¹⁵

The British Report shows how digital technology and Internet access are helping to open new possibilities for the facilitation of scholarship by libraries. But to take full advantage of the technologies, libraries must consider changes in major areas of traditional activity. Let me touch on two such areas: collection development and bibliographic control.

In the area of collecting, most if not all of your libraries are investing increasing proportions of their budgets in digitizing parts of their own collections, and in acquiring access rights to electronic databases, digitized journals, and other virtual resources. In addition, numerous academic libraries are developing digital repositories for preserving and providing access to scholarly products created electronically by their own faculty members in research and teaching—repositories of the kind pioneered by MIT in the DSpace model, which several other institutions also are using. At the federal level in the United States, another kind of digital-collection repository is in development that may soon get a new legislative boost.

On the second day of this month, May 2, Senator John Cornyn, a Texas Republican, and Senator Joseph Lieberman, a Connecticut Democrat, introduced Senate bill 2695, entitled “The Federal Research Public Access Act of 2006.” The bill applies to research funded in whole or in part by US government agencies with research budgets of more than 100 million US dollars. The bill is aimed particularly at agencies that make grants for non-classified research, performed not within the agencies but within universities, healthcare services, and other outside groups. Eleven federal agencies are likely to be involved, including the departments of Agriculture, Commerce, Defense, Education, Energy, Transportation, Homeland Security, and Health and Human Services,

and the National Science Foundation, the Environmental Protection Agency, and the National Aeronautics and Space Administration.¹⁶

The bill would require that if a researcher receives financial assistance from such a federal agency, and produces a paper accepted for publication in a peer-reviewed journal, the researcher must provide an electronic copy of the paper (or the published version if the publisher agrees) to the agency that helped fund the research. The bill would require that the agency place the paper in a digital repository designed for long-term preservation, whether maintained by the funding agency or by some other organization. And the bill would require that the repository provide the public with free, online access to the paper as soon as possible and no later than six months after publication in a journal.¹⁷

Our federal government already has been conducting a research-repository experiment. In May 2005, our National Institutes of Health (NIH) announced what was called a “Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research.” Under the policy, NIH asked researchers working with its financial support to submit resulting manuscripts to the National Library of Medicine’s PubMed Central, a digital repository that preserves and provides public access to peer-reviewed results of biomedical, behavioral, and clinical research.¹⁸ However, the NIH made its policy voluntary, and only 4 percent of eligible research made it into PubMed Central, which led some to consider the policy “a failure.”¹⁹ An Open Access Working Group at the National Library of Medicine making it mandatory to deposit research supported by NIH in PubMed Central. All this was the background for the new, more comprehensive Senate bill.

The US Association of Research Libraries has endorsed the new legislation, along with the Association of College and Research Libraries, the Association of Academic Health Sciences Libraries, and SPARC, the Scholarly Publishing and Academic Resources Coalition.²⁰ But the Association of American Publishers, fearing loss of journal sales, has declared opposition, and Congressional approval is far from assured.²¹ If the bill does become law in the United States, our National Library of Medicine and our National Agricultural Library, among others, may become more deeply involved in developing research repositories—libraries, in effect, that use digital technology to expand access to scientific research discoveries worldwide.

I am not as familiar with the state of open-access activities in Canada. But I know that CARL as well as ARL was a signatory in 2001 to the Budapest Open Access Initiative, sponsored by the Open Society Institute,²² and that the Canadian Association of Research Libraries has urged “Canadian granting councils and universities to consider adopting self-archiving policies that would mandate researchers to publish their work in open-access journals or deposit their works into open-access repositories.”²³ And I understand that the Canadian Institutes of Health Research are in the process of developing a new policy on “Access to Products of Research.” In fact, I believe that last Monday, May 15, was the deadline for recommendations from the public about what that policy should contain.²⁴ I look forward to hearing from my Canadian colleagues what the outcome turns out to be.

An additional area in which change is coming to facilitate scientific research and general scholarship is the area of bibliographic control. If national libraries, under the impetus of new policies and new legislation, do, indeed, assume responsibility for

building repositories of scientific research and other scholarship produced with public tax dollars, we will be compelled to devise new approaches to providing and guiding access to the contents. We will need a new access structure, focused on delivery, not of catalog information, but of content, itself.

Contrary to popular belief, libraries developed cataloging initially as a means of fostering access. We have taken pride in our catalogs, which once were a great innovation. In the pre-Internet era, cataloging records became our most effective means of letting scholars know what books and journals and other collections were in our holdings. Over time, we automated these records, and developed networks of bibliographic records that made them widely available to the world.

Now, in the 21st century, the creation of digital information resources accessible over the Internet has accelerated, and so has the development of search services designed to help information seekers find what they want. Users increasingly discover that it is easier to pursue a subject by typing it into a Google search box than to track down references obtained from even an online library catalog. Users prefer the convenience of search engines even when searches produce overwhelmingly long lists of possible resources, without much guidance about which ones contain reliable information or most correspond to a user's individual needs. For a long time, scientists, in particular, have made little use of our catalogs, which have always been more suited to study in the humanities and social sciences.

The Library of Congress was instrumental in putting the current structure of bibliographic controls in place. The library has continued to play an active role in the organizations that govern international cataloging policy. And on a daily basis, we have

provided bibliographic information to libraries around the world. New technology now gives us new opportunities to expand access again, and national libraries such as ours must again take responsibility for developing structures that will facilitate research.

Shortly after I joined the Library of Congress as associate librarian for library services, I spoke at a meeting of the US Association of Research Libraries about the need to rethink our bibliographic infrastructure. I also asked all directors of services within my purview at the library to begin redesigning our services and products by focusing on needs of users. And I asked our directorate for Acquisitions and Bibliographic Access to streamline our processes to make the library's resources more quickly and conveniently accessible.

Our catalogers actually had started reconceptualizing their role before I arrived. In fact in the year 2000, during the library's bicentennial celebration, our cataloging managers convened an international symposium to consider bibliographic changes for the digital era. Subsequently our Acquisitions and Bibliographic Access staff members have been working on a new strategic plan and laying groundwork for changes we will need. We realize that millions of items among our special collections are hardly accessible even now because we can provide the public no more than cursory bibliographic descriptions of them. They will be further out of public sight as researchers turn more to Internet search engines than to our catalog. And now that we can provide content electronically, we must give priority not to describing but to delivering it.

I have found that making even small changes in the bibliographic system at the Library of Congress can produce tremors across the country. And I have no desire to throw out quickly a system that we, ourselves, have done so much to put in place.

National libraries around the world, working with colleague institutions in each of their countries, will need to collaborate on developing a workable new system based on access to content rather than access to description. Difficult work lies ahead—work that must be done thoughtfully and carefully.

The questions we collectively face go far beyond the two areas of change that I have described. In the broader framework of library transition, I have been thinking about such additional challenges as the following:

- How can a library such as mine—traditionally viewed as a “library of last resort”—translate that collections concept for the digital environment in which everyone is a publisher?
- How important is it to preserve and provide access to the popular oral-history collections that every group, at least in the US, seems to be generating?
- How do we allocate resources between needs for traditional preservation and demands for digital preservation?
- As institutions that have been deeply committed to the humanities, social sciences, and cultural heritage, what should be the role of our libraries in supporting research in the sciences in the 21st century? For example, how do we work with research centers and scientific laboratories to see that large data sets and other resources for science are maintained appropriately?
- Recognizing that managing digital preservation for the long term will take a collaborative effort, how do we develop the models of governance and funding required to make such collaboration successful?

- Finally, but far from least, how do we re-tool our staff members, many of whom are quite senior and expert in their fields, to work in the digital information environment?

On all of these fronts, we must work together to find the next steps. Service to scholarship, in the sciences as well as other areas, internationally as well as domestically, will be increasingly required of us as the future unfolds. If we can adequately adapt our services and programs, that future will be wonderfully productive.

Thank you.

¹ Professor Sir Gareth Roberts, “International Partnerships of Research Excellence, UK-USA Academic Collaboration,” draft manuscript, March, 2006, p. 2.

² Ibid., p. 3.

³ Ibid., p. 4.

⁴ Ibid., p. 4-5.

⁵ Ibid., p 5.

⁶ Ibid., p. 6-7.

⁷ Ibid., p. 7-8. The areas of mutual concern are described as “Cyber Infrastructure/e-Social Science and Human and Social Dynamics.” The German funder is the Deutsche Forschungsgemeinschaft (DFG).

⁸ Ibid., p. 12-13.

⁹ Ibid., p. 16.

¹⁰ Ibid., p. 9.

¹¹ Ibid., p. 9-10.

¹² Ibid., p. 9.

¹³ Ibid., p. 9.

¹⁴ Ibid., p. 10.

¹⁵ Ibid., p. 11.

¹⁶ Janice Anderson, "Federal Research Public Access Act of 2006," e-mail message to the Legal Information Preservation Alliance, May 8, 2006. Also, SPARC, "The Federal Research Public Access Act of 2006, Frequently Asked Questions for University Administrators & Faculty," <http://www.arl.org/sparc/resources/frpaa.html>, accessed May 12, 2006.

¹⁷ Ibid.

¹⁸ National Institutes of Health, "NIH Public Access Working Group of the NLM Board of Regents," http://www.nlm.nih.gov/od/bor/workgroup_charge.html, accessed May 12, 2006.

¹⁹ Anderson, "Federal Research."

²⁰ SPARC, "The Federal Research."

²¹ Anderson, "Federal Research."

²² Open Access Institute, "Budapest Open Access Initiative," <http://www.soros.org/openaccess/>, accessed May 15, 2006.

²³ Canadian Association of Research Libraries, "Open Access, the Canadian Context," http://www.carl-abrc.ca/projects/open_access/oa_can_context-e.html, accessed May 15, 2006.

²⁴ Canadian Institutes of Health Research, "Invitation to Provide Advice and Suggestions for CIHR's Policy in Development—Access to Products of Research," <http://www.cihr-irsc.gc.ca/e/30948.html>, accessed May 13, 2006.