CENDI Cost Study

"The Changing R&D Information Economy in the Digital Age"
Report Prepared by Robert Ubell

Making the Transition to a Digital Information Environment

- **Key Judgements**
- **Agency Case Studies Vignettes**

Sponsored by

Department of Commerce *(National Technical Information Service)*
Department of Energy *(Office of Scientific and Technical Information)*
NASA *(Office of Scientific and Technical Information)*
National Library of Medicine
Department of Defense *(Defense Technical Information Center and National Air Intelligence Center)*
Department of Interior *(National Biological Resources Division)*

December 1997

CENDI / 97-3

THE CHANGING R&D INFORMATION ECONOMY IN THE DIGITAL AGE

A Report Prepared for CENDI

By Robert Ubell

President, Robert Ubell Associates

New York, NY

September 1997

Executive Summary

This paper offers the following keyjudgements:

1. Since the Electronic Revolution is in its infancy, experts in the industry cannot predict the long-term effects. With little serious substantiated economic experience or cost studies, there is no current consensus about the consequences of introducing digital technologies. Many recognize that for the immediate future, a dual system—conventional print coupled with new electronic services—may be required. In this environment, costs in the short term are likely to increase.

2. The real payoff from the Electronic Revolution may not be economic. Rather, the investment may be returned in the creation of more effective products designed to generate quality decisions more rapidly. As a corollary, there is broad agreement that as the price per published unit of information declines, costs of delivering more value-added, personalized products to larger populations of end-users will cause costs to rise overall.

Ever since the introduction of digital technologies, conventional wisdom imagined that the combined costs of acquiring, storing, retrieving and disseminating R&D documents would decline when compared with costs associated with delivering conventional print products. The theory held that since a large part of the costs of creating and shipping print on paper—printing, paper, binding, postage—would disappear, organizations would achieve great savings soon after their files were converted.

While the digital age promised to deliver information to end-users in better and faster ways, today managers have discovered that while costs per end-user served appear to have declined[1] and while access has also improved markedly, investments in developing, installing, and maintaining technological infrastructure have escalated. The "cheaper" goal of "better, cheaper, faster" may yet be realized, but only after further investments in technology and human resources are made. If measured in terms of increases in productivity, decision quality and mission effectiveness, however, the return has already been—and will continue to be—enormous.

The Economics of Print Publication

According to the premier guide to journals,[2] there are now more than 165,000 periodicals published worldwide, with approximately 2,000 added annually. Of these, more than 10,000 are peer-reviewed. In the period 1960-1990, prices of science and technology journals increased far in excess of inflation so that the depositories of technical information in the last decades have been exposed to two parallel threats. As the number of research periodicals grows, so have the subscription prices for publications already housed in collections (Figures 1 and 2[3]).

With a river of data overflowing information depositories, several factors accounted for sharp increases in the prices of scholarly material:

1. Balkanization of Science and Technology.[4] In classical fields, such as chemistry and physics, the "twigging effect"—in which research spawns ever-narrower sub-disciplines—stimulated the production of parallel, quite specialized periodicals and other literature (textbooks, monographs, reference works, review series, etc.).

2. Limited Markets. As the literature addresses narrower fields, the marketplace for any one discipline shrinks, so that the support available for any single journal falls to a handful of practitioners and their institutions. While a few highly cited periodicals may boast subscriptions in the tens of thousands, they remain exceptions. Most peer-reviewed journals report subscription totals of between one and two thousand, with a considerable number with less than 500.

---

1 Fulton, Kenneth; Walter Warnick, personal communications.

2 Ulrich's International Periodicals Directory, 34th edition, 1996, R.R. Bowker.

3 Cummings, Anthony M., et al., University Libraries and Scholarly Communication: A Study Prepared for the Andrew W.Mellon Foundation, 1992, Association of Research Libraries.

4 Cummings, Anthony M., et al., page 94.

3. Continual Generation of Research Reports As fields grow, not only does the number of periodicals published in that discipline tend to increase—so that in certain dynamic areas, there may be as many as two dozen or more scholarly journals competing for papers—but existing publications also tend to expand to meet demand. A journal launched, say, twenty years ago as a quarterly with 400 pages a year in a "hot" field may have expanded over the years to a monthly publishing today more than 2,000 pages annually.[5]

4. Uniqueness of Papers in Science and Technology.[6] Unlike the wide availability of competitive commodities in supermarkets where consumers select one or two brands of breakfast cereal off the shelves, articles published in the technical literature, for the most part, are not interchangeable. Information managers cannot simply choose one journal over another and hope to satisfy their research staff. The New England Journal of Medicine, for example, cannot be substituted for the Journal of the American Medical Association. Both are essential.

5. Escalating Publishing Costs Despite the introduction of various new technologies in the printing processes,[7] so-called "first-copy" costs have escalated. Peer-reviewing[8] alone is among the most labor-intensive operations, with dozens of pieces of correspondence, telephone calls, faxes and e-mails transmitted merely to accept (or reject) one article—much of it at the expense of the publisher who must also support salaries, fringe benefits and overheads of editorial teams at the publisher's site and often at satellite offices at distant research institutions. Once accepted for publication, a paper must be shepherded through copyediting, typesetting, proofreading, indexing and abstracting, among dozens of other major and minor operations before it is actually printed. Add to these the mounting cost of postage, not only to deliver the publication to subscribers, but also to market it to its intended audience.

6. Publishers' Economic Dilemma Now consider the high cost of printing slimmer print runs, owing to smaller markets for finer slices of research, to photocopying, and to declines in funding. As prices climb, acquisition budgets have historically not kept pace, forcing institutions to cancel low impact journals. With slipping circulation, publishers announce a spiral of price increases.[9] (Certain European publishers have exploited currency fluctuation to further boost prices and profits.) Consequently, over the last several years, publishers report that unit sales for periodicals have declined on average between three to five percent a year.

The result is an information crisis.[10] The investment needed to acquire and manage the fire-hose gushing from the knowledge industry has overwhelmed the collection and management of essential R&D information. Maintaining the paper status quo offers the unappetizing prospect of failing to control costs.

The Digital Information Economy

Faced with a paper crisis, information managers welcomed the digital revolution, expecting that computerization would help control price inflation by eliminating some of the more expensive aspects of publishing and managing published information. The result would contribute to a "digital dividend," after saving the cost of printing, paper, postage, storage, maintenance, and other aspects of both the publishing and delivering economies. Miles of shelves would be vacated; entire buildings would be emptied as solid, bound volumes were transformed in electronic impulses. What's more, computerization would offer far more information to far greater numbers at greater speed and with sharper precision.

But as a recent study on the economic impact of networked information revealed,[11] a number of obstacles undermined the coming of the electronic utopia:

---

5 Cummings, Anthony M., et al., page 95.
6 Cummings, Anthony M., et al., page 95-96.
7 Cummings, Anthony M., et al., page 128.
8 Crawford, Susan Y., et al., page 101.
9 Cummings, Anthony M., et al., page 97-98.
10 Cummings, Anthony M., et al., page 11.

1. Investments in "First-copy" Costs Increased[12] Commercially available digital subscription costs to scholarly materials increase because much of the editorial activity required to generate print continues in electronic publishing[13] and, in many cases, these costs rise, owing to additional labor needed to "markup" text for electronic dissemination, especially on the World Wide Web. Many publishers report that "markup" costs have increased editorial expenses between 15 and 25 percent.

2. Highly Trained Technical Staff at Higher Salaries Increase Human Resources Budgets Conventional print publication and technical information management has been traditionally shepherded through many of its various stages—from initiation to end-user delivery—by a handful key management staff, supported by a cadre of entry-level employees without advanced training. The digital information economy, however, depends critically on unprecedented numbers of highly qualified programmers, software developers, Web masters and others armed with technical degrees, commanding salaries far greater than the average wage for print support staff.

3. Rapid Obsolescence of Hardware and Software Results in a Treadmill of Technology Costs By now it is widely acknowledged that much of the installed infrastructure and accompanying software being used today by publishers, librarians, database producers, end-users and others in the information chain will be obsolete 24 to 36 months after implementation. Investment required to upgrade systems has introduced a new treadmill of technology costs.

4. Acquiring Digital Information Introduces New Roles In the print economy, subscription agencies and library book wholesalers occupy an intermediary space between institutions and publishers—consolidating orders, servicing claims, among other labor-intensive tasks—freeing information managers to perform more satisfying intellectual services for their clients. With the introduction of digital products, highly trained and more costly specialists must now evaluate new technologies, integrate systems and negotiate terms, either directly with providers or with third-party integrators.[14] The information manger must now become a licensing agent, since much digital information is now leased rather than purchased. Most information managers must now often seek legal advice about contract terms—an entirely new expense.

5. Costly New Technical Support Services Are Now Required As new technologies are released—and as end-users are given direct access to information over networks—institutions, software and hardware vendors, telecommunications systems, information providers and database producers must now provide experienced online or toll-free technical help and customer support services to greater numbers of users, expanding the scope, sophistication, and cost of end-user services.

6. Technology Generates Greater Demand at Greater Expense[15] With computer screens on every desktop, with instant access to vast quantities of information, with alerting services that identify up-to-the-minute information, users have come to expect the delivery of high-impact data never before as easily accessed. As speeds improve, as databases become more easily accessible and as more information is mounted on networks, end-users are coming to demand more expensive hardware, software, and telecommunications networks.

7. Electronic Publishing Adds to Costs by Adding New Products and Services Digital technologies make it possible to provide individuals and their supporting agencies with products and services never before possible with print. Today, penetrating search engines, graphical interfaces, and other new technologies have revolutionized the way in which information is searched and assimilated. Billion-dollar digital information companies—commercial and nonprofit—such as Knight-Ridder, Lexis-Nexis, Chemical Abstracts, ISI, and others have introduced online products that have become essential information-gathering tools, adding to the total cost of acquiring information, rather than offering less-expensive alternatives.[16] What's more, institutions—many in government—that produce databases for scholars and the public have been forced by their mission not only to

---

11 Ubell, Robert, Cost Centers and Measures in the Networked Information Value Chain, April 1997, Coalition for Networked Information.

12 Cummings, Anthony M., et al., page 95.

13 Crawford, Susan Y., et al., page 106.

14 Malone, Thomas W., et al., page 488.

15 Molholm, Kurt N., personal communication.

16 Malone, Thomas W., et al., page 496.

acquire expensive commercial databases, but must continuously update technologies to identify, gather, categorize and disseminate great volumes of data to satisfy consumer demand.

8. Intellectual Property Management Introduces New Investments[17] Because electronic data is easily accessible and widely available, the proprietary rights of publishers and authors may be compromised. Consequently, all of the players in the digital arena must now introduce safeguards to prevent as nearly as possible the wholesale exploitation of information by unauthorized users.

9. Data Conversion and Digital Archiving Can Be Prohibitively Expensive[18] Converting legacy paper files to digital media is a mammoth and very expensive task. While certain institutions and providers have committed themselves to mounting historical documents on networks, it is not certain whether this effort will be universal, given the great investment needed.

10. Paper Documents Must Be Maintained in Parallel with Electronic Files at Great Cost Recognizing the extraordinary costs involved in converting paper to digital media, most collections must continue to maintain their historical documents, forcing institutions to continue to support the storage and personnel infrastructure at—or near—previous levels. What's more, of the 10,000-plus peer-reviewed journals available, less than 1,000 are currently online. While publishers are accelerating the mounting of periodicals on networks, it is unlikely that more than a couple of thousand will be accessible electronically by the turn of the millennium.[19] What's more, many documents are inappropriate for electronic dissemination (at least not at present). One manager reports that an average document used by his staff is approximately 110 pages—far larger than is convenient to access on computer screens.[20]

11. "Repurposing" Adds Large Unanticipated Costs As technology evolves, files mounted previously on more primitive digital networks can become obsolete when new platforms, search engines, and interfaces fail to accept earlier records. Unanticipated investments then become necessary to convert existing files to conform to the new technology or to acquire entirely new files that essentially replicate the old but cannot be accessed by updated software. This dilemma faces all of the players in the information industry—R&D laboratories, archives, publishers, database producers, among others.

12. Database Producers Must Accommodate Costly "Mixed Economy."Today, vast quantities of bibliographic and other data generated by commercial, government and nonprofit institutions consists largely of conventional print documents. In order to acquire, catalog and disseminate print information, database producers—in and out of government—must continue to rely on traditional methods or expensive digitizing equipment. As the digital economy emerges, computerized information must be added to existing files, sometimes requiring parallel technologies, staff, and other infrastructure in order to accommodate both print and online documents. What's more, because the digital economy is in its infancy, few common standards have been accepted that a wide variety of non-compatible electronic files must be integrated at great cost.

---

17 Scott, R.L., "Transitioning to a World of Electronic Scientific and Technical Information Exchange," paper delivered at Inform '97, May 7, 1997, pages 9-10.

18 Arms, William Y., "Scholarly Publishing on the National Networks," Scholarly Publishing, Vol 23, April 1992, pages 158-169.

19 Ubell, Robert N., Scholarly and Professional Journal Publishing Industry, Robert Ubell Associates, 1994, pages 1-2.

20 Molholm, Kurt N., personal communication.

Conclusion

It is impossible to predict when annual expenditures to support the digital environment will fall below what is required to maintain a paper economy.[21] While the "digital dividend" has not yet made information budgets shrink, it has certainly had large and lasting effects. Among these are:

- deeper accessibility to the world literature[22]
- broader availability to larger populations[23]
- increased R&D productivity[24]
- sophisticated manipulation of data[25]
- pin-pointed selection of critical information[26]
- simplification of research methods[27]
- easy navigation through complex databases[28]

In order to achieve these obvious benefits, continued investment in the digital economy is required now.

References

Arms, William Y., "Scholarly Publishing on the National Networks," Scholarly Publishing, Vol 23, April 1992

Cummings, Anthony M., et al., University Libraries and Scholarly Communication: A Study Prepared for The Andrew W. Mellon Foundation, 1992, Association of Research Libraries

Crawford, Susan Y., Julie M. Hurd and Ann C. Weller, From Print to Electronic: The Transformation of Scientific Communication, 1996, ASIS

Drabenstott, K.M., Analytical Review of the Library of the Future, 1994, Council on Library Resources

King, D.W. and J. Griffiths, "Economic Issues Concerning Electronic Publishing and Distribution of Scholarly Articles," Library Trends, Vol 43, No 4, Spring 1995

Lynch, Clifford., "The Transformation of Scholarly Communication and the Role of the Library in the Age of Networked Communication," Serials Librarian, Vol 23, No 3, 1993

Malone, Thomas W., Joanne Yates and Robert I. Benjamin, "Electronic Markets and Electronic Hierarchies," Communications of the ACM, Vol 30, No 6, June 1987

Report on the Task Force on a National Strategy for Managing Scientific and Technical Information, 1994, Association of American Universities

---

21 Fulton, Kenneth, personal communication.
22 Cummings, Anthony M., et al., page 104.
23 Scott, R.L., page 6.
24 Malone, Thomas W., et al, page 484, 488.
25 Scott, R.L., page 13.
26 Scott, R.L., page 6.
27 Fulton, Kenneth, personal communication.
28 Malone, Thomas W., et al. Page 494.

Scott, R.L., "Transitioning to a World of Electronic Scientific and Technical Information Exchange," paper delivered at Inform '97, Oak Ridge, TN, May 7, 1997

Siler, Sara, "Dictionary Dynasty Fights to Survive in Electronic World," The New York Times online "CyberTimes," June 30, 1997

Strassman, Paul, "People: The Untold Part of the Cost Story," ComputerWorld, April 14, 1997

Ubell, Robert, Cost Centers and Measures in the Networked Information Value Chain, April 1997, Coalition for Networked Information

Ubell, Robert, US and Canadian Survey of Electronic Products and Services (1993-1995), 1995, Robert Ubell Associates

Ulrich's International Periodical Directory, 34th Edition, 1996, R.R. Bowker

Acknowledgments

This report was based, in part, on telephone conversations with the following experts whose participation is gratefully acknowledged:

Hal Espo, former Senior Director, Strategic Alliances, Knight-Ridder Information

Kenneth Fulton, Executive Director, National Academy of Sciences

Karen Hunter, Assistant to the Chairman, Elsevier Science Publishers

Robert J. Massie, Director, Chemical Abstracts Service

A.W. Kenneth Metzner, Vice President, Director of Electronic Publishing, Academic Press

Kurt N. Molholm, Administrator, Defense Technical Information Center, Department of Defense

Ann Okerson, Associate University Librarian, Yale University

Walter Warnick, Director, Office of Scientific and Technical Information, Department of Energy

Robert Ubell

Robert Ubell is President of the scientific and technical publishing consulting firm, Robert Ubell Associates. Over the past 15 years, the company has generated a number of key studies on the development of electronic publishing, including, most recently, Cost Centers and Measures in the Networked Information Value Chain for the Coalition for Networked Information. In the commercial sector, clients have included some of the more prominent publishers, including Thomson, McGraw-Hill, and Harcourt. The firm has participated equally in the nonprofit arena, with such notable clients as the American Institute of Physics, the American Chemical Society, and the American Cancer Society.

The author of more than 50 publications in such periodicals as Nature, The New England Journal of Medicine, and New Scientist, Mr. Ubell has also written or edited some 17 books for such distinguished publishers as Oxford University Press, Cambridge University Press, and Henry Holt, among others. Mr. Ubell lectures widely on scholarly publishing and has taught at Columbia College of Physicians & Surgeons and MIT.

Prior to founding Robert Ubell Associates, Mr. Ubell was American Publisher of the premier science weekly Nature, and founder of the monthly, Nature Biotechnology. Earlier, he was Editor of The New York Academy of Sciences

award-winning magazine, The Sciences. He began his career at Plenum Publishing Corporation, where he became Vice President and Editor-in-Chief. Last year, Mr. Ubell served as President of the Internet biological information site, BioMedNet.

A member of numerous boards and committees, he has served in various capacities for the National Academy of Sciences, the National Research Council, the National Academy of Engineering, among many other scholarly and government agencies and associations. He is on the Board of Directors of Marcel Dekker, Inc., the scientific and technical publisher in New York. On September 2, he joins Dekker as Executive Vice President for Electronic Publishing.

MAKING THE TRANSITION TO A DIGITAL INFORMATION ENVIRONMENT

KEY JUDGEMENTS

1. Fully Electronic Information Management is in its infancy and the full life cycle cost is not yet understood

2. The real payoff of the digital world may not be simply economic, but, instead the ability to use better information for better decision making

3. There is often a shift in where costs are incurred and savings are gained. The information management function may increase costs but offsets to users provide an overall life cycle return on investment for the agency or nation.

CASE STUDIES

Medicine

Research is needed to make information fully electronic and this adds to the development costs that allows future payoffs. For example, in the visible human project (with goals in teaching and research about the human body) methods need to be developed to link image data to symbolic text-based data, which is comprised of name hierarchies, principles and theories. Standards do not currently exist for such linkages. Information research and development money is needed to accomplish this. The Visible Human Project will allow:

- Non-invasive colon cancer screening
- Simplified plastic surgery
- Prostate cancer surgical rehearsal
- Surgical simulation
- Revolutionizing the study of anatomy in high school, college, and professional school
- Radiation absorption modeling
- Crash testing simulations

Another medical example: In Medicine, effective mapping of the human genome may help to find the causes for some disease and help in prevention, treatment or cures. A key component of the human gene map project is the management of information. Because of the volume and the dynamic nature of the data and the need for broad access and dissemination of the data, electronic versions are essential.

Defense

The Department of Defense DTIC is making DoD directives and instructions (D&I) electronically available whereas previously they were only available in print and microfiche form. In an analysis of economic benefits, DTIC estimated that over the course of a 9 month period, about $200,000 could have been saved by users if electronic D&I were available. Other non-quantifiable user benefits(savings) include timeliness , avoidance of storage and on-site file maintenance of material and currency of information. However, DTIC had to invest in certain life cycle processes to make this possible. They needed to scan, OCR, index, store, develop search and retrieval software, and maintain transmission capability. Based on experience with the prototype of full text with technical reports, it is known that DTIC must process the electronic files prior to producing an acceptable output format. This was the production step previously done by the publications offices including layout, editing, format conversion which costs are now at DTIC. Development of more efficient processes for carrying out these functions is an investment that must be made to realize the thousands of dollars of user benefits.

Aerospace

In the aviation and aeronautics sector, aviation products are the largest positive industrial contributor to the U.S. balance of trade. This edge is eroding, especially to the French. France is known for engaging in highly effective, active information capture, with on tenth of our research budget. They now have 40-50% of the aircraft and space launch business. Japan, another key competitor has STI organizations track the career of researchers in 60 countries and employs 5,000 experts specifically to monitor 8500 selected journals from 50 nations. The U.S. has increased its budget in science, technology, and education, but the STI budget in NASA has declined over 50% in recent years. Competent mechanisms to disseminate and access results to those intended to benefit are possible and critical in an electronic environment if we are to have effective return on our R&D investment dollar.

Taxes

For the taxpayer, the National Technical Information Service (NTIS) has made access to over 700 tax forms with instructions, publications and other information available on the WWW 24 hours a day, 365 days per year. The IRS needed on online solution that protected IRS internal systems and that could be rapidly expanded during the peak tax season. During the 1997 tax filing period, the web site received over 117 million hits and over 6 million file downloads. The IRS has been able to significantly reduce the overall expense and response time of delivering tax information as well as enabling more constituents to reach the IRS.

Government Contracting

The Davis-Bacon Act requires wage rates established by the U.S. Labor Department to be paid on federally funded construction projects. Rates are established by surveying wages for each occupation in four types of construction for each county of the 50 states, District of Columbia, and the territories. There are about 3000 documents.

Prior to online availability of this data, it was only available to the public as a 7000-page document with weekly updates of as many as 1500 pages. Ten day currency is required in contract bidding. Print delivery was not always timely. The customer of the print service paid $2000 per year for the full country. With the current online availability through NTIS's FedWorld, the data are updated every Friday. The cost is $600 per year for unlimited access.

The advantages of NTIS's database system over paper are:

- Cost Savings
- Timely Delivery
- Powerful and effective search engine
- Rapid access to specific data
- No labor for paper maintenance
- Electronic documents can be incorporated into solicitations, proposals, and contracts without rekeying. This also eliminates keying errors.

Federal Records

As the federal government continues to shift from paper to electronic record keeping, the National Archives and Records Administration (NARA), has major responsibilities to provide guidance on how agencies should handle this electronic environment. In addition, the hardware and software needed to access the growing electronic legacy collections must be provided in order to continue to make these collections available to researchers and the public.

Shifts in planning, training, and technology that must take place in order to continue to perform NARA's mission in an electronic environment. In the strategic plan of NARA it states, "We have to figure out how to cope with a paper explosion that threatens to overwhelm us, how to get on top of accelerating technological change, how to take advantage of opportunities that technologies offer, and how, in a time when the federal government is cutting back, to make the most of our resources." In addition to paper NARA is trying to cope with growing quantities of computer-generated records. In the Federal Government, these range from millions of e-mail messages to vast scientific databases, all of which require new methods for appraisal, preservation, and public access. Clearly, electronic records problems are costly to overcome and require more staff with technical expertise, a resource we can never have enough of. Moreover, the current shrinking of government only makes the job larger because as agencies streamline, programs end, and military bases close, more information is cleaned out and accumulated records need to be dealt with.

NARA is implementing an Electronic Access Project (EAP) to manage electronic collections and to provide better access, but basic descriptive work (i.e., cataloging to provide accessible content) must be performed first in order to create the basis for centralized access. EAP had significant infrastructure investments (http://www.nara.gov/nara/vision/ eap/eaprec.html.). Congress has appropriated $4.5 million for this project. The operating budget for NARA also acknowledges the need for continued improvements, beyond the initial development of EAP, and the increased need for NARA to provide guidance in the electronic environment. In times of shrinking Federal resources, NARA's FY98 budget increased 4% over FY97 to accommodate the transition to a digital information environment.