Vinton G. Cerf

DOI:10.1145/2436256.2436258

Open Access

OPEN ACCESS CONTENT” is a term that holds different meanings depending on the perspective and context. For some, it denotes “free of charge”; for others, it may mean “downloadable.” In principle, most definitions revolve around the notion that content is easily found and freely available. It is a topic that has been widely discussed and, with the advent of the Internet and the Web along with “digital publishing,” it has become an important touch-point for the research community.

For many, discussions about open access ultimately lead to the economics of publishing, which has changed dramatically in recent years, as most content has moved online and print has increasingly become a secondary medium. With print, there is a significant cost to produce physical copies, to distribute, archive, and provide access to them. In the digital world, printing and distribution costs may ultimately disappear, but the creation, distribution, and archiving of digital content rely on more complex technology than print and costs for such technology are not insignificant. Access costs in the digital world are related to digital communication network access while in the physical world they revolve around postage and transportation. And the costs of composing and making digital content discoverable will likely never go away completely. In the digital world, the traditional print functions of a librarian are still needed, since most would agree that full-text search is not a substitute for a skilled research librarian who adds value through context, experience, and personal interaction with a searcher.

It is easy to fall into the trap of comparing apples and oranges when discussing the economics of open access. For example, one might compare the cost of operating a real-world library

with its shelves of books, magazines, CDs, to the cost of a high-density, high-capacity disk-drive system. We must think more broadly about the function of curating, cataloging, and indexing content, and maintaining storage systems (whether print-based or digital) to illuminate the more relevant considerations. There is, however, another trap I find myself falling into occasionally: comparing digital content that is *merely* the equivalent of print content; that is, static papers, text, imagery, and so on. That this is an overly narrow notion is readily understood as one begins to think about interactive presentations of research results, archiving of research data, *and* software needed to interpret, analyze, or present research results.

As our capacity to store digital information grows, it is predictable that the need and appetite for storing research data and analytical software will also grow. There are consequences to this direction. For one, we need to assure the data formats, the analytic software, and other metadata can be preserved *and understood* for hundreds if not thousands of years. Curating the collection of digital reports, interactive applications, and large-scale databases is a substantial challenge. I have written on the topic of *bit rot* in the past and the term applies here as well. Not only do we need to archive the raw bits of reports, data, metadata, and application software, we must also maintain a system context in which all of this material remains accessible and usable.

For some, this leads to the conclusion that we may need software emulators of older hardware, older compilers for past source code languages, OS copies, and instantiations of applications relevant to published research reports and data. One can readily reach the conclusion that maintaining such a diverse archive is challenging and costly.

Another aspect of high-value archiving is completeness. While it is arguable that many unaffiliated parties may maintain some content independently, there is great value in knowing that *all* content of a particular class can be found in a readily accessible archive. The ACM Digital Library is a prime example of an archive valued for its comprehensive character. Indeed, the value and important contribution of some of its content is recognized years after publication. Therefore maintaining a comprehensive collection contributes to scholarship in a critical way.

As the research community moves toward digital publication of content, algorithms, analytical software, data, and metadata, it seems inescapable that business models will be needed to assure the longevity, utility, and comprehensive nature of archival information. There appear to be many ways in which these costs may be defrayed. Research grants may cover some or all of these costs, subscriber fees may provide another path, historical “page charges” in the print world may be replaced by their equivalent for maintaining a sort of *digital vellum* in which content and its surrounding ecosystem can be made eternal and enduring.

This topic is the subject of ongoing discussion throughout ACM. Practical steps toward resolving alternatives are being taken, as detailed by ACM Publications Board co-chairs Ronald Boisvert and Jack Davidson in the February 2013 issue of *Communications* (p. 5).

The recent steps taken by ACM are important but not the end of the story. I look forward to exploring these and other ideas with members of ACM, the research community, and the organizations that sponsor research in all fields, including our own.

Vinton G. Cerf, ACM PRESIDENT