

## “If it’s not on the Web, it doesn’t exist at all”: Electronic Information Resources – Myth and Reality

Sarah Stevens-Rayburn<sup>1</sup>

*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

Ellen N. Bouton<sup>2</sup>

*National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA*

**Abstract.** In this paper, we review the current status of astronomical research via electronic means, with an eye towards separating the hype from the hypothetical in hopes of revealing the actual state of affairs. We will review both anecdotal and scholarly work aimed at documenting the state of research using the World Wide Web and demonstrate that although there is enormous potential in electronic research, much of that potential is as yet unrealized. In addition, especially in astronomy, a significant amount of material is not (yet) available electronically and likely will never be. Finally, we will point out the potential danger of a looming paradigm shift in the way astronomers conduct research and the possible consequences thereof.

### 1. Introduction

If we needed a theme, a sort of mantra for our feelings about the Web, the following quote sums it up:

In the great sea of human knowledge, the Web is like a coastal swamp – shallow and chaotic, full of debris and rotting matter, and often stagnant and murky, if nonetheless always crowded and busy with life. Meanwhile the great sea of human knowledge stretches out undiscovered before it. The Web is just a subset of the wonders of intellectual discovery. (White, 1998)

The author of these comments, Jeremy White, writing in the *Australian Personal Computer* magazine, goes on to describe a recent information-seeking incident. As was his usual practice, he went first to the Internet and discovered that his topic was too broad and the calibration of Internet information too

---

<sup>1</sup>ST ScI is operated by AURA, Inc. under contract with NASA

<sup>2</sup>NRAO is operated by AUI, Inc. under cooperative agreement with the NSF

unrefined to produce any useful results. He goes on to say that he, like many other modern day technophiles, has become “desensitised to low-grade and low-quality but high-volume information, desensitised to unstructured search tools and unsystematic research. . .”. We’re pleased to report that he then had the good sense to go to a library that within a matter of minutes was able to provide him with suitable references.

Luckily for us in astronomy, the electronic tools available to us are much better and more refined than is the case in many other disciplines. This is because, although our horizons are infinite, the body of knowledge with which we deal is manageable and relatively compact. This is, however, both its blessing and its curse. Because we have access to so much of it, thanks to services like SIMBAD, NED, and the Astrophysics Data System, we sometimes tend to think we have access to *everything* we need. Unfortunately, this simply isn’t the case. In the next few pages we will review a few of the technological myths currently abounding in our profession and in our discipline, and try to put them in their proper perspective. To help us in this endeavor, we conducted a mini-survey of our own scientific staffs, and we asked the librarians at the South African Astronomical Observatory and at two Indian institutions to ask their scientists to participate as well. As we discuss the myths of the Internet and electronic publishing, we shall be incorporating some of the comments received in response to our questions.

## **2. Myth number 1: Everything (important) is on the Internet; you just need to keep looking until you find it.**

This is actually a two-part myth: (a) that it’s there and (b) you’ll be able to find it, but we will examine both parts together. There was an intriguing study carried out in the Psychology Department of the University of West Florida last year and reported in the *Newsletter on Serials Pricing Issues* no. 188 (Shroyer, 1997). Because of anticipated budget cuts, the library staff conducted an experiment to see what of the current information content of their printed journals might be available electronically. They searched the Web and they also contacted the commercial publishers of psychology journals. They discovered that two-thirds of their titles were represented on the Web, but only 2% had “full-text electronic counterparts to the printed journals.” This correlates exactly with the findings of Lowry and Troll, who reported in 1996 that “less than 2% of the serials published world-wide are currently available in electronic format” (Lowry and Troll, 1996). Of the websites found in the Florida study, over half of them provided tables of contents and/or abstracts (55%), but that meant the other 45% did not.

The situation, of course, is changing rapidly, and astronomy has been much better at making its information content available electronically than have other disciplines. However, of the approximate 40 refereed astronomy titles to which we subscribe, less than ~25% are available in full-text electronically (for a price, of course. See myth 4). This means that in raw numbers, there really isn’t very much of current material that appears simultaneously in paper and on the Web.

Certainly, that ~25% make up the most important and highly respected of our professional journals. We owe a large debt of gratitude to the foresight

of the American Astronomical Society for being among the first to make their information available electronically. But even if it is the most important, it is still not all there is. Peter Boyce has been widely quoted as saying that by the beginning of 1998, “95% of the world’s peer reviewed astronomical literature will be available online.” (Woodward et al., 1996). Considering just the five primary titles in astronomy, and counting only numbers of volumes, we *still* only have about 49% availability online. When we recently spoke with Peter about his assertion, he clarified by saying that he meant current journals only. (If one measures shelf space for the titles in question, we actually have about 68% of the material online, but that’s still a far cry from 95%.) That may be closer to the truth, although we would still argue that 95% is a bit high. Nonetheless, Peter as well as anyone knows that having just the current literature available does not provide all that much. Astronomy has a long history of diversity in its publications. Researchers utilize not just the commercially available journals, but also important works produced by observatories and astronomy departments which have been distributed freely to others. This diversity has meant the community of astronomers was more open and accessible internationally, without regard to the economic viability of the parent country or institution. There have always existed both a “right to publish” and a “right to have one’s work accessible.” As for back volumes, we would estimate that about 10% or less of our collections are available in full-text via the Internet and the 90% not available online is enormously important for astronomical research. When supernova 1987A went off, astronomers didn’t go rushing for the current journals. They needed to know what had been going on in 30 Doradus years earlier.

Our astronomers commented on this in the survey, saying things like “I regularly go to the paper versions for those journals which are not available on the Internet.” On the other hand, 46% admitted that there was a possibility that they would omit from their research papers that were not available electronically.

The ADS has done a superior job in making the primary journals for the past 20 or so years available electronically, and as Guenther Eichhorn shows in a poster for this conference, they are well on their way to making earlier volumes available as well. But these journals are but a dozen titles out of the several hundred that we buy and maintain for the use of our scientific staffs. Does that mean we can dump the other journals? Heaven knows we could use the shelf space, but we’re not quite sure we’re ready for that. And what about the 50% of our collections that are monographic works or conference proceedings? There are a few of these available electronically, but in an incredibly piecemeal fashion and subject to the whims of crashing disks and non-astronomer systems people who can easily decide that this great meeting proceedings that hasn’t been accessed in two years is no longer a good use of the storage medium.

### **3. Myth number 2: The journal is on the Web.**

What exactly does this mean? Is it *all* on the Web? What about letters to the editor, instructions to authors, obituaries, announcements of prizes and awards? A recent discussion on the liblicense listserv (<http://www.library.yale.edu/~llicense/index.shtml>) centered on the topic of “when is a journal not a journal?” Producers of “aggregator databases,” those

wherein a third party offers “full-text” of a multitude of titles from many different publishers, may actually be re-keying the text from the original, thus introducing typographical errors and spurious corrections. The posting to make this point used the amusing tale of a paper on Mormonism, wherein some over-eager spellchecker had changed all the instances of the word Mormon to moron.

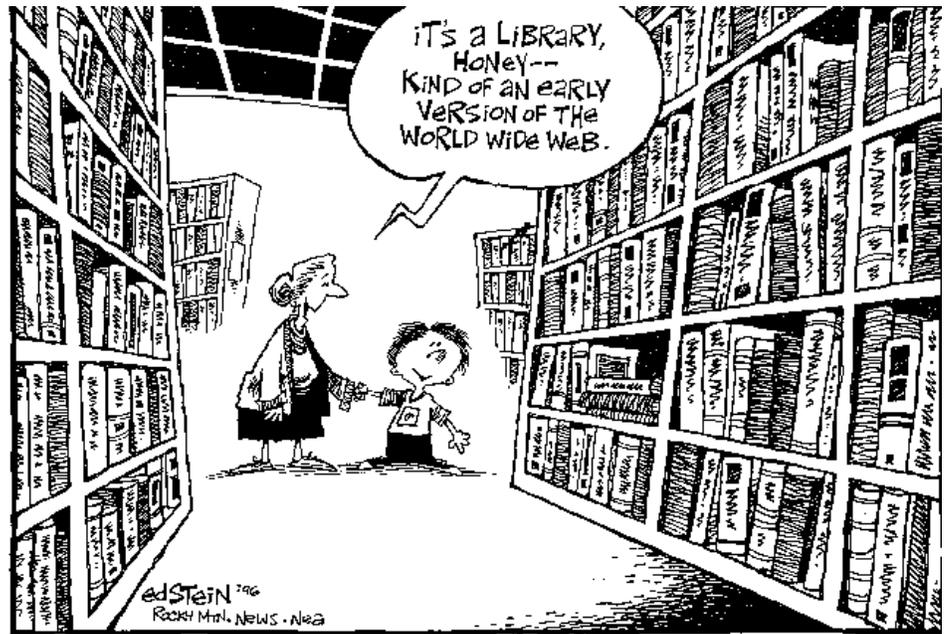
From a more positive angle, consider those materials now available on the Web that we could not possibly hold in paper: collections of materials that do not lend themselves readily to print distribution; e.g., data sets from NASA missions, extensive tabular data or collections of images. The online version of the January 1998 *Astronomical Journal* has a movie embedded in one of the articles. Looking at the paper version, one sees only the first frame of the movie and an exciting description of what the reader is missing by reading the paper edition instead of viewing the online one.

Nonetheless, “the journal is on the Web” may not be a particularly helpful piece of information. *Where* on the Web is often the more relevant question. About five years ago, we mounted our first Library Web pages, with links to all of the important astronomical sites we figured our staffs would need. Care to guess how many of those links would still be accessible today? Even the very carefully managed AAS journals ended up moving from a link at the AAS to one at the University of Chicago Press within a relatively short time after they opened for business. But the address did not change fast enough to stop the original address from being embedded in the Library of Congress record for the electronic ApJ, and it sits there uncorrected to this day. This brings us to:

#### **4. Related myth number 3: That information is on a Web page.**

In point of fact, a reference that was cited yesterday may have vanished tomorrow. It may have moved, or it may disappear entirely. Or the server may be down. Will it *always* be available on the Web? Barring loss, theft, fire, natural disaster, or paper disintegration, the paper copies will still be in the library years from now. Will the commercial publishers, for whom astronomy publications are certainly not a big money-maker, see fit 10 or 20 years from now to continue to archive and make available what they may consider to be a low-use journal? The AAS sets aside money for refreshing the archive and moving/converting it to future formats, but are the commercial publishers likely to do this? Another example: NRAO is on the email distribution list for the VSOP News (VLBI Space Observatory Programme), which is distributed bi-weekly and contains ongoing information on HALCA (Highly Advanced Laboratory for Communications and Astronomy), of great interest to NRAO staff. In addition to being distributed by email, VSOP News is available on the Web. However, the NRAO Library prints and archives it because a staff member, who is part of the project, has said that after the mission has ended the material will almost certainly be taken off the Web. Similar concerns abound for all of the newsletters currently available electronically. Perhaps these have no possibility of historic importance, but we simply don’t know that and they change their addresses more often than politicians change their opinions. Several observatories have decided to no longer produce paper versions of their observatory newsletters. That means that the history of these institutions could well be lost to future generations.

Another concern that cannot be ignored is the difficulty of keeping track of versioning of Web publications. Versions of a Web page may change, so that what one found before may not be there later, even if the link is good. Recently, the ApJ had to deal with the interesting question of what to do about an author's name being misspelled. It's awfully easy to correct the spelling in the online edition, but doing that introduces two related problems. First, the online edition then no longer is a bibliographic match of the paper edition. As paper versions fade from view, this may not be a problem, but as long as both exist, publishers must be extremely careful about what they do, be it ever so minor. The second problem a correction introduces is: how does the user know which version is being referenced? With the change of the spelling of an author's name, it would be easy to find out, but what if it's correcting a plus sign to a minus in an equation? If there is any change in the author's name spelling or of a plus to a minus sign, then the electronic version is correct and the paper is not. Is everyone going to be certain in the future which is the permanent archival copy? If the electronic version becomes the archival copy, will people know with what year/volume the paper ceased to be that?



© 1996 Ed Stein, reprinted with permission

### 5. Myth number 4: Information on the Web is free.

The travelling companion to this myth is that e-access to journals will lower library costs. In truth, access to the electronic versions of journals costs money. Sometimes it is a comparatively reasonable amount of money, for instance the *Astrophysical Journal*, but sometimes prices are so high that access to the electronic version is out of reach for any but the large university libraries. Sometimes access (always for a fee) is only available to consortia members, thus effectively cutting off the majority of astronomers who are at the independent research institutions (NRAO, ST ScI, ESO, etc.). Roy Tennant, writing in *Library Journal*, says “We do not create digital libraries to save money. We create them to greatly expand access, increase usability and effectiveness, and establish entirely new ways for individuals to interact with information. Rather than being cheaper to create, maintain, and preserve than print collections, the evidence so far seems to indicate the opposite. Meanwhile, we are retaining and expanding our print collections” (Tennant, 1997). Confirming Tennant’s remarks are those of Lowry and Troll (1996), who looked at the costs of locally maintaining digital archives and concluded that “Based on a ten-year replacement cycle, digital storage and access will cost academic libraries 16 times as much as print to store locally.” Since their study (and probably in some cases because of it), the pattern seems to be moving away from local archives for electronic journals, but that simply shifts the costs back to the publisher, who will certainly be forced to pass them on to the libraries anyway.

Walt Crawford, writing in *Online* notes that if every text of any length is printed each time it is used, there are enormous disadvantages to the all-digital library. He estimates that a typical public library would spend more on printing and licenses than its current total budget and would use at least 50 times as much paper. Further, if publishers can enforce the pay-per-view model that the Association of American Publishers favors, then the download fees must be added to the print costs. The resulting total may well be more than it would cost to buy the print version, and the print version may be used over and over by many different users (Crawford, 1998). In our survey of staff use of electronic journals, 91% of tenure/tenure-track astronomers printed the paper out over half the time and two-thirds of them printed over 90% of the time. 44% of this same group also reported that they are more likely to print out the electronic article than they are to photocopy from the paper journal. We have shifted the paper costs from the publisher to the reader or the reader’s institution.

### 6. Myth number 5: Problems with technology (e.g. slow download time, bandwidth problems, access problems) will be solved in the next year or two, and communications won’t cost us anything.

This is what Crawford calls “the great technological handwave.” It is “the futurist’s response to any shortcomings in technology, any unmet needs, anything that’s lacking. . . . The great technological handwave turns ‘ifs’ into ‘whens’ and ‘whens’ into ‘just a couple more years.’ [It] rejects budgetary arguments, since as we all know, technology just keeps getting cheaper and cheaper until it’s essentially free” (Crawford, 1998; see also myth 4 above: information on the

Web is free. As an example of myth 3 above, “That information is on a web page”, we note that the link to our original paper copy of Crawford’s article printed from the Web page in February 1998 had changed by 1 April).

Several of our survey respondents commented on the frustration that often accompanies electronic access: “I find that I’ll often try to find a paper on-line before going to the library. But slow access, poor search engines or search engine failures, inconsistent interfaces, and papers where no E-version is available (or where the data tables aren’t included), make doing things electronically really annoying sometimes.”

We in astronomy are quite lucky in that our sources of information are relatively few and the places where we look for them are fairly well-known, but we are still at the mercy of the Internet vagaries when we step even slightly out of our narrow discipline.

In considering the consequences of the technological handwave, we in the United States often lose sight of what is going on elsewhere. We get annoyed when an occasional summer thunderstorm knocks out our access for an hour or so, but we forget that there are those for whom there may routinely be two hours a day when there is no power because of a systematic rotation of outages in order to conserve the limited power available.

Interestingly, the connectivity problems in other countries really do affect us more than we might like to admit. We had several respondents who commented that they were less likely to use *Astronomy and Astrophysics* because “their website was so slow that I gave up.” This leads us into our last section, the looming paradigm shift in astronomical research.

## **7. Paradigm shift in astronomical research?**

One of our greatest concerns has been whether the availability of electronic journals, accessed either directly or through ADS, influences the papers people cite in their research in a negative way. Do people omit from their citations those papers for which full text is not available? What about those papers not included in ADS indexing? In our survey, we asked “In what ways, if any, has the availability of electronic journals changed the way you do research?” and “Does the availability of online resources influence which papers you cite in your own research; e.g., are you more likely to omit papers you can’t get easily?”

Person after person praised the speed of doing literature searches using ADS, the ability to quickly find complete and correct references to partial citations or vaguely remembered articles. One person noted that he was much more likely to look up a reference right away rather than making a note and checking it later in the library. However, when talking about changes, most people told us about the ways the abstracting and indexing of ADS (and SIMBAD and NED) affected them, not ways the electronic journals did. A commonly expressed thought was, “So far, the electronic journals haven’t made much difference. However, the online references to data and literature (SIMBAD, NED, and the ADS) are essential. . .”. One person said he reads the current journals more frequently since he can access them in spare moments from his office. However, several people noted that they “no longer browse the journals as vigorously as before

and so [are] less likely to get ‘distracted’ and read articles outside ... current research interests.”

In a recent interview, Clifford Lynch, developer of the University of California’s Melvyl system, notes that “there’s a sense in which the journal articles prior to the inception of that electronic abstracting and indexing database may as well not exist, because they are so difficult to find. Now that we are starting to see, in libraries, full-text showing up online, I think we are very shortly going to cross a sort of critical mass boundary where those publications that are not instantly available in full-text will become kind of second-rate in a sense, not because their quality is low, but just because people will prefer the accessibility of things they can get right away.” (Educom Review Staff, 1997)

When asked in our survey if the availability of online resources influenced which papers they cited in their own research, 64 people answered “no,” but 55 said “yes.” One person, answering yes, added “a warning that in this new electronic age, if it isn’t on-line, for many purposes it might as well not exist.” Several people did specifically say the online resources *broadened* their citations by helping them to find papers they would not otherwise have known about. However, others said that although they would cite all relevant papers, they would very likely be unaware of papers that are not available or abstracted online. One person said, “Literature searches are now at least 2-3 orders of magnitude faster than before the advent of ADS. In those days the only way was to wade through *Astronomy and Astrophysics Abstracts* and it could take several days of solid work just to get all the references for one topic. Now it takes a few minutes.” While it is certainly correct that ADS is orders of magnitude faster than A&AA, we are concerned by the implication in this statement that ADS and A&AA are equals and one would therefore retrieve from ADS all the references gleaned from a thorough search of A&AA. Another person recognized the danger of electronic searches: “The serious problem I see is that electronic searches give the ‘illusion’ [of being] 100% complete just because they come ‘automatically’ (but if you use an inadequate key for your search, the result may be HIGHLY incomplete!)”

It may also be completely overwhelming, by providing entirely too many references. One of the most common complaints we hear from people is that in using online searches (SIMBAD, NED, ADS) it is very difficult to narrow the search terms enough to prevent a flood of unrelated and unwanted hits. Unnecessary information is counterproductive, and wastes the time of the person who must look through it to determine its relevance.

Some people said they read fewer papers of general interest than they used to: “Paper journals allow me to skim results from individuals, institutions, or oddball interests that appeal to my curiosity and fill in niches of information that were of subconscious interest to me, sometimes of great usefulness on a topic which at the moment is not the focus of my immediate search.” And another said, “I spend less time sifting potentially relevant sources and tend rather to focus quickly on a few (i.e. I think it has reduced the breadth of reading).”

In his book, *The Gutenberg Elegies*, Sven Birkerts writes about “the gradual displacement of the vertical by the horizontal, the sacrifice of depth to lateral range.” He asks, “how is one to assess the relative benefits and liabilities of these intrinsically different situations? How do we square the pluses and minuses of

horizontal and vertical awareness? .... Inundated by perspectives, by lateral vistas of information that stretch endlessly in every direction, we no longer accept the possibility of assembling a complete picture. . . . The computer, our high-speed, accessing, storing, and sorting tool, appears as a godsend. It increasingly determines what kind of information we are willing to traffic in; if something cannot be written in code and transmitted, it cannot be important" (Birkerts, 1994: 72-74).

The computer is a tool that has radically changed the ways in which both astronomers and librarians work, and it will continue to do so. But if we are to be responsible users of technology, we must constantly be aware of both its myths and its realities, the ways it enhances what we do and the ways it limits us. If we are blinded by its dazzle, we can no longer assemble a complete picture.

### References

- Birkerts, S. 1994, *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. (Boston: Faber and Faber)
- Crawford, W. 1998, *Online*, 22,  
<http://www.onlineinc.com/onlinemag/OL1998/crawford1.html>
- Educom Review Staff. 1997, *Educom Review*, 32(6),  
<http://www.educom.edu/web/pubs/review/reviewArticles/32642.html>
- Lowry, C. B., & Troll, D. A. 1996, *Serials Librarian*, 28, 143
- Shroyer, A. 1997, *Newsletter on Serials Pricing Issues*, No. 188,  
<http://www.lib.unc.edu/prices/>
- Tennant, R. 1997, *Library Journal*, 122(20), 31
- White, J. 1998, *Australian Personal Computer*, 19(4), 6
- Woodward, H., Rowland, F., McKnight, C., Meadows, J., & Pritchett, C.  
<http://telecaster.lboro.ac.uk/dawson9.html>