

See CDP's Digital Preservation Resources, [www.cdpheritage.org/alpha/digital/preservationResources.cfm](http://www.cdpheritage.org/alpha/digital/preservationResources.cfm).

*SOURCE:* Collaborative Digitization Program, [www.cdpheritage.org/digital/projectManagement/](http://www.cdpheritage.org/digital/projectManagement/). This site is updated frequently. Reprinted with permission.



## THE INTERNET

### Basic problems with information on the internet

by Diane K. Kovacs and Kara L. Robinson

**SOME BASIC PROBLEMS WITH INFORMATION** obtained from the internet, or just about anywhere else for that matter, are listed here in the order of their observed frequency on the Web.

**Typos.** The information provided on the internet comes from many sources. Typos are one of the most prevalent problems, because *anyone* can publish information on the internet and often no editors or publishing agencies review the information. The two most likely causes of typos are inaccurate typing because of the informality of the medium and ignorance of the language. English is the lingua franca of the internet, but many varieties or dialects of English exist. Some typos may in fact be spelling variants rather than errors. Terry Ballard and Tina Gunther, in *Typographical Errors in Library Databases* ([faculty.quinnipiac.edu/libraries/tballard/typoscomplete.html](http://faculty.quinnipiac.edu/libraries/tballard/typoscomplete.html)), publish the results of their ongoing analysis of the typos that occur in library catalogs, websites, and other library-related sources.

**Factual errors** (accidental or deliberate). These usually happen because people simply are not checking, or sometimes are just recalling information from confused memories. During an internet searching workshop taught in 1993, the only answer we could find on the internet to the question, "What was the year of the first Thanksgiving?" was 1676. According to Infoplease ([www.infoplease.com](http://www.infoplease.com)), the actual year of the first Thanksgiving is either 1621, 1789, or 1863, depending on whether you mean the first celebration, or the year that it was declared a holiday by George Washington or Abraham Lincoln.

The answer we found in 1993—at a site that no longer exists—was supplied by a 6th grader at a suburban Chicago school. This example is not meant to imply that 6th graders are always a source of inaccurate information. Some 6th graders might publish accurate information if they acquire the facts from an authoritative source (teacher) and/or document their source (encyclopedia, almanac, or website).



**Opinion stated as fact.** Throughout the internet, users can find opinion stated as fact. This problem is very prevalent. Do you question the veracity of something *just* because of who published it? Where the internet is concerned, yes, you must question the

veracity of information based on who said it. You have to ask, "Did the person/doctor/6th grader have training or do research that gives them the authority to provide the information?"

A related issue is the fact that the actual live person who publishes information on the internet can create an online identity that looks good, but has no connection to the reality of the person's real life. This means that checking offline sources to verify authority and credibility is essential.

Can that person provide documentation/proof that what he or she says is accurate? What type of information is provided online to make these determinations? We do the same kinds of evaluation when we work with print resources. Look at the authors of articles and find their sources, research, training, and background before you believe what they say or write.

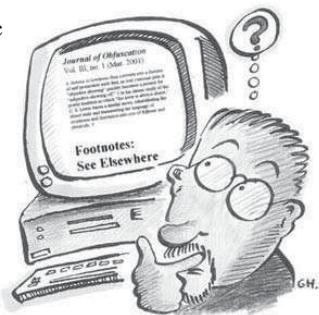


Editors evaluate the veracity of content as well as the authors producing that content at the acquisition stage of publishing, and libraries rely on a publisher's reputation in making their purchasing decisions. During the acquisitions process, librarians rationalize that if a particular publisher accepted and published a book or journal, then it must by association be of good quality. Internet research evaluation is more difficult. It involves more primary research than we are used to doing. One factor is that much information on the internet was originally part of a conversation. Discussion lists, newsgroups, MUDs, and chat transcripts may

be text based, but they are really more akin to speech than to publications. The difference between speech and published information is primarily formality of the language. A three-judge panel that heard the initial arguments in *ACLU v. Reno*, 521 U.S. 844; 117 S. Ct. 2329 (1997) found that on the internet "tens of thousands of users are engaging in conversations on a huge range of subjects. It is no exaggeration to conclude that the content on the internet is as diverse as human thought." Some transcripts of this worldwide conversation are literate and/or authoritative, and others are not.

**Out-of-date information.** Considering how easy it can be to update web pages and other internet information sources, the amount of out-of-date information online is surprising. But people don't always have the time or ability to update information, or to take it offline when it is obsolete. For example, student project websites might remain online long after the project is finished and the student graduates. Another problem is that so much information on the internet is actually archives of discussion lists and newsgroups. It is important to check the dates of the individual postings in such archives, as well as on any other web resource that might be included in your library collection.

**Bias.** Bias is a bigger problem with all sources of information than many people realize. Many internet sites—as well as every other publication medium—provide slanted information to influence how people think about something. An illustrative example is the "Dihydrogen Monoxide Research Division" ([www.dhmo.org](http://www.dhmo.org)). The website uses hyperbole, negative statistics, and words that are meant to scare and alarm people; e.g., dangers, alerts, truth, cancer, "DHMO Kills." Not much documentation supports these claims. Yet



none of the information or facts are false. The bias used in presenting the information gives a skewed sense of the meaning of the information. Only when the user pauses to consider the identity of dihydrogen monoxide does it become clear that this site is intended to illustrate the problems of bias. This site also illustrates the need for selection of resources that take into account educational attainment, reading level, and information needs. The reader needs to have at least some basic chemistry education. Di- (two) hydrogen atoms—H<sub>2</sub>, plus mono- (one) oxygen atom—O, makes H<sub>2</sub>O. The DHMO site is all about water.

### **They're Not Just Using Websites**

**by Paul O. Jenkins**

A fear often expressed by today's academic librarians is that students at their institutions are using websites at the expense of more reputable resources, such as books and journals. In order to study just how true this perception might be at the College of Mount St. Joseph, a private college in Cincinnati with an enrollment of 2,500, I requested faculty to forward to me bibliographies in student papers. Six faculty members, each from a different discipline, responded, leaving me to compile statistics from a total of 116 papers. The results from this sample are discussed below.

#### ***Percentage of citations by resource type***

<i>Discipline</i>	<i>Citations</i>	<i>Articles</i>	<i>Books</i>	<i>Websites</i>
Sociology	260	20%	36%	44%
Physical therapy	234	82%	18%	0%
Religion	149	20%	54%	26%
Humanities	83	11%	42%	47%
Nursing	66	41%	56%	3%
Chemistry	62	48%	27%	24%
Total	<b>854</b>	<b>41%</b>	<b>35%</b>	<b>24%</b>

As the table indicates, the papers yielded 854 citations. A look at the totals reveals that the traditional research sources—articles and books—remain those most often used. Together they made up 76% of the citations, websites accounting for the remaining 24%. Exceptions to this pattern are the sociology and humanities papers, for which websites were used more than any other resource.

No consistent pattern emerged from an examination of course levels for the papers submitted. The only upper-level courses were those from humanities and religion. The physical therapy and nursing courses were midlevel, and those from sociology and chemistry were lower level. Resource requirements outlined by instructors varied from course to course and had a more profound effect.

The physical therapy instructor did not allow the use of any websites; the nursing instructor required that students obtain the permission to use them as sources for their papers. As the numbers indicate, the instructors in the other disciplines adopted a more lenient stance.

Certainly the addition of websites has influenced how students conduct research. If my institution is any true indicator, faculty acceptance of websites as legitimate resources is by no means universal, but seems to have become valid for many. Students have embraced websites and use them with books and articles. For the time being, however, the more traditional resources remain in the ascendancy.

**SOURCE:** Paul O. Jenkins, "They're Not Just Using Web Sites," *College & Research Libraries News* 63 (March 2002): 164.

From Mr. Edwin Worsk  
NO 268 hospital road,  
Box 1287 Port Shepstone  
Chaka -South Africa.

Dear Friend,  
We want to transfer to overseas  
(\$36,000.000.00 USD)Thirty six  
million United States Dollars) from  
a Bank in South Africa. I am  
looking for a reliable and honest  
person who will be capable and fit  
to provide either an existing bank  
account or to set up a new Bank a/c  
immediately to receive this money,  
even an empty a/c can serve, as  
long as you will remain honest to  
me till the end for this important  
business, trusting in you and  
believing in God that you will  
never let me down either now or in  
future.

Election campaign information is biased, almost by definition. For that matter, so is all advertising information. Probably every piece of information reflects bias of some kind, due to the subjectivity of writing. The degree, type of, and reason for bias must be considered in evaluating information.

**Deliberate fraud** is a rapidly growing problem, given the ubiquity of the Web. Medical fraud on the Web has increased. Business or consumer frauds are also common. The best defense is to know where to check to see if an offer or claim really is too good to be true. Medical claims made on websites might be checked out using the Federal Trade Commission's Operation Cure-All site ([www.ftc.gov/bcp/conline/edcams/cureall/](http://www.ftc.gov/bcp/conline/edcams/cureall/)) on which the FTC reports ongoing health fraud investigations and warnings, or the Quackwatch website ([www.quackwatch.com](http://www.quackwatch.com)). Other valuable sources of information are Quatloos!—Cyber-Museum of Scams and Frauds ([www.quatloos.com](http://www.quatloos.com)), Scamorama ([www.scamorama.com](http://www.scamorama.com)), and Scambusters ([www.scambusters.org](http://www.scambusters.org)).

*SOURCE:* Diane K. Kovacs and Kara L. Robinson, *The Kovacs Guide to Electronic Library Collection Development* (New York: Neal-Schuman, 2004), pp. 22–24. Reprinted with permission.

## Libraries and internet filtering, 2005

by Lori Bowen Ayre

**LIBRARIES BEGAN USING INTERNET FILTERS** in the late 1990s due to community pressure and the Children's Internet Protection Act (CIPA). CIPA is a federal law that requires all computers in a public library to be filtered if that library accepts any federal funds for computers that access the internet or the costs associated with a connection to the internet. It took effect on July 1, 2004. According to the National Conference of State Legislatures, 21 states have filtering laws that apply to schools or libraries. While most of these laws require publicly funded institutions to adopt internet use policies, some mandate filters. Legislators are convinced that filters effectively protect minors from harmful, web-borne internet content. To the extent that filters are expensive and may pose a threat to free speech or open access, legislators (and much of the public) have decided that the protections for children outweigh any such concerns.

The use of filters in public libraries has increased steadily. Norman Oder reported in the January 2002 and January 2005 issues of *Library Journal* that the percentage of public libraries filtering increased from 25% in 2000 to 65% in 2005, yet many librarians argue that filters have no place in a library.

The American Library Association fought CIPA in the courts and took the position that "the use of filtering software by libraries to block access to constitutionally protected speech violates the Library Bill of Rights." Some libraries refused to install filters and gave up federal funding instead. Filters were seen as antithetical to the mission of the library.



While some libraries were developing internet policies explaining their reasons for not using filters, other libraries were quietly installing them. The libraries installing the filters soon found that filters alleviated many thorny problems they'd been grappling with. It turned out that filters did prevent children from bumping into unexpected and unwanted websites and advertisements. Filters served as a deterrent for public porn browsers. Libraries found that with some effort, they could implement and enforce their internet-use policy. For the first time, libraries had a way to control how their public computers were being used. The filters didn't do the job perfectly, but the fact was, there were no more complaints from patrons after the filters were installed. The filters were *good enough*.

## How filters work

Filters operate on a system of categories. Websites, or sometimes individual web pages, are categorized by filter companies. The library's filter administrator utilizes the categories to build filter profiles. For example, the adult filter profile might allow all categories of content to pass through except items categorized as "sexually explicit." The children's filter profile would undoubtedly block the "sexually explicit" content as well as other categories deemed inappropriate for children such as "hate," "firearms," and "violence."

The filter company decides how each site will be categorized. Filter companies fiercely protect their process for categorizing websites and equally fiercely protect the websites identified within each category. Part of the value of the filter is in the number of websites categorized, because sites that have yet not been categorized will not necessarily be blocked.

Ironically, librarians—professionals trained to catalog and evaluate content—subcontract their cataloging job to software companies when they install a filter. Unlike librarians, the subcontractors are not information professionals, and they typically use automated methods to classify the 3 billion web pages on the internet.

The features available in state-of-the-art filters are too numerous to recount here. For a thorough summary of filter features and to compare filters, feature-by-feature, visit [libraryfiltering.org](http://libraryfiltering.org). However, certain features are particularly important for libraries, such as the ability to control what is blocked, how to override blocked pages, how granular the blocking is (site, page, domain, IP address), and what information is presented to end-users when they encounter a blocked page.

Most library filters leave the choice of what to block in the hands of the system administrator, who sets up filter profiles and selects the categories to

block. Filters that don't allow for different filter profiles might be suitable for home use but are not appropriate in a library setting.

Most filters provide some mechanism for overriding blocked pages either on the fly using an administrator password, or by adding sites to an "always allow" list that supersedes the block on a page caused by its categorization.

