POLICY PREREQUISITES AND TECHNOLOGY LIMITATIONS

As you've seen, copyright policy interacts with technology by tending to limit technological development—but also by using technology to undermine aspects of public policy. Technologies introduced into the library without active consideration of a policy framework may conflict with and undermine library policies.

There's another, perhaps more frustrating interaction of policy and library technology. Some policies *require* technology to make them effective—and sometimes, the technology just isn't up to the task. When policymakers will not (or cannot) accept the limitations of technology, you have a constant source of friction and difficulty.

This chapter considers two examples in very different areas, although both may involve online searching. First, a long discussion of filtering policy and technology; then, a brief discussion of library metasearch systems.

Filtering

While the name "censorware" more accurately describes what Internet filtering software actually does, this discussion uses the neutral term "filtering." But don't be fooled: Except for whitelists (programs that only allow people to visit *approved* sites), filtering software works by censoring the presumed bad and significant chunks of stuff that isn't bad, not by acting as a filter that lets all the good stuff, and only the good stuff, through it.

The Mar/Apr 2004 *LTR* (40, n. 2), "Filtering and Filtering Software," offers guidelines and pointers on individual filters and how to choose among them.[1] For libraries that don't receive funding that requires them to install filters, Nancy Kalikow Maxwell offers an aging but still useful discussion of "Alternatives to Filters" in the Mar/Apr 2001 issue of *LTR* (38, no. 2).[2]

All policies related to Internet filtering (except for library policies *not* to filter) have one thing in common: They depend on effective technology to carry them out. That poses a problem.

Policy Issues

Various conflicting policies come into play with Internet filtering. That's true for all uses of filter software, not just filtering within libraries.[3] Just within libraries, these policies may be involved:

Federal law and case law requiring filters. For libraries receiving certain kinds of Federal support, the Children's Internet Protection Act (CIPA) mandates that filters be present on all computers capable of accessing the Internet. As upheld by the Supreme Court, the CIPA mandate is actually quite narrow as it applies to adults.

State and local law and policy requiring filters. Some states and localities require filtering for some or all library computers. Those laws and ordinances are likely to define broader filtering coverage than CIPA's

narrow coverage and unlikely to provide such immediate recourse as the case law for CIPA. In such cases, the local policy probably conflicts with federal policy. Some localities have adopted policies that library computers should *not* be filtered—which now means doing without some federal funding and may conflict with more restrictive state laws.

Federal, state, and local policies regarding free speech, privacy, and confidentiality. Guarantees of free speech and press appear at the highest level of federal policy—the Constitution itself. While privacy is not explicitly guaranteed in the United States Constitution, it is an explicit Constitutional right in several states, such as California. Various federal, state, and local policies—and vast quantities of case law— provide varying levels of protection for confidentiality. To the extent that freedom of speech and the press implies freedom to read, these policies conflict with policies requiring filters, leading to patterns of court fights and uneasy compromise decisions.

National and local library policies. The American Library Association (ALA) recommends policies that do not allow for age-based discrimination in the provision of library services. ALA also takes an absolute official stance against the use of filters. Many, perhaps most, local libraries have formal or informal policies somewhat more nuanced than ALA's recommendations. The mix of policies involved includes collection development policy, age-related policies, appropriate-use and time-and-place policies, confidentiality and privacy, and policies regarding offensive behavior. Some of these policies will be written; some exist only in the minds and habits of librarians. The whole mix makes filter decisions more complicated.

Today's Most Important Policy: CIPA

To date, only one federal law relating to filtering has been upheld by the Supreme Court of the United States. It's worth looking at CIPA, as modified by the Supreme Court's decision, in more detail. CIPA is not the first federal attempt to impose restrictions on Internet content. CIPA is, however, the first to survive court battles.

First, a few quotes from the decision itself (selected from a lengthy ruling):

> The Government has broad discretion to make content-based judgments in deciding what private speech to make available to the public . . . Internet terminals are not acquired by a library in order to create a public forum for Web publishers to express themselves. Rather, a library provides such access for the same reasons it offers other library resources: to facilitate research, learning, and recreational pursuits by furnishing materials of requisite and appropriate quality.

> Because public libraries' use of Internet filtering software does not violate their patrons' First Amendment rights, CIPA does not induce libraries to violate the Constitution. . . . To fulfill their traditional missions of facilitating learning and cultural enrichment, public libraries must have broad discretion to decide what material to provide to their patrons . . .

> The decisions by most libraries to exclude pornography from their print collections are not subjected to heightened scrutiny; it would make little sense to treat libraries' judgments to block pornography any differently.

> Especially because public libraries have traditionally excluded pornographic material from their other collections, Congress could reasonably impose a parallel limitation on its Internet assistance programs. As the use of filtering software helps to carry out these programs, it is a permissible condition under Rust.

Concerns over filtering software's tendency to erroneously "overblock" access to constitutionally protected speech that falls outside the categories software users intend to block are dispelled by the ease with which patrons may have the filtering software disabled.

If some libraries do not have the capacity to unblock specific Web sites or to disable the filter or if it is shown that an adult user's election to view constitutionally protected Internet material is burdened in some other substantial way, that would be the subject for an as-applied challenge.[4]

The first four paragraphs support CIPA as a policy, justifying it based on other national and local policies. The last two paragraphs relate directly to technology limitations and, as expanded, constitute critical differences between the original law and the law as it was upheld.

Following are some comments on those notes, if only to show how complex CIPA is as a policy matter:

Because "most" public libraries don't collect "pornography" (the casual term used to signify "harmful to minors" material as in CIPA)—a conclusion based on the unfortunate testimony of two so-called expert witnesses— therefore *all* libraries can reasonably be required to reject online "pornography."

After you say "most" once, you can just change that to "libraries" as a whole—after all, why worry about the minority?

Rust (*Rust v. Sullivan*) was another questionable Supreme Court decision, one that upheld Congress' right to require that family planning services receiving federal funding eliminate abortion counseling. Bad law is used to support bad law.

The majority decision happily quotes 1930 guidelines for material selection and Donald Davis's more recent claim that universal access would be detrimental to users. A footnote seems to say that it's really America's libraries that are desperate to prevent patrons from viewing "pornography"—Congress is just helping out the cause. Thus, the seven percent of libraries with universal filtering morph into most, then all, public libraries.

Where do we get the absolute assertion that public libraries do not install Internet terminals to provide a forum for Web publishers to express themselves, but rather to provide patrons with online material of requisite and appropriate quality? The decision states it at least twice and possibly three times. How is it that open access or even "filtered" access could be justified as "providing patrons with online material of requisite and appropriate quality?" That's a mystery that's never been solved.

Clarifying the Policy

CIPA never called for blocking of text, only images—a point that's more important after the decision. Images to be blocked fall into two categories for adults, three for anyone under seventeen—a rather broad definition of "children." The two general categories are child pornography and obscenity, both categories that are already illegal in the United States. Presumably, Web sites offering images of child pornography or obscenity reside outside the United States; otherwise, the government should be shutting them down entirely, not merely limiting library access.

The third category relates to those younger than seventeen and is the most troublesome: "Harmful to minors." As defined in CIPA, these are images that fail the three-part test for obscenity but "with respect to minors" added. The

problem here is that the three-part test for obscenity is so difficult. Briefly, the image must *simultaneously*:

1. appeal to a prurient interest in nudity, sex, or excretion;

2. depict, describe, or represent in a *patently offensive way* an actual or simulated sexual act, sexual contact, or lewd exhibition of the genitals; and

3. lack serious literary, artistic, political, or scientific value.

Think about those elements, particularly the first two. In essence, the image must simultaneously turn you on and offend you—an odd combination at least for most healthy minds. It must also be trash (the third clause), but that goes without saying in the United States.

CIPA does not call for filtering of pornography as such (most of which is Constitutionally protected) and does not call for filtering of *textual* pornography at all, even for children.

A library with strong local support and enough backbone could make a strong case for taking a truly minimalist approach to CIPA. That approach would include installing an open-source blocking program configured to block illegal sites (child pornography and obscenity) on all computers. It also would be configured to block images only based on URLs for sites adjudged by government authorities to be harmful to minors and to block those images on computers in use by adults with explicit warning of the blocking—a one-button "unblock this" function, and a similar one-button "don't filter this session" function—for the search session as a whole.

The truly minimalist approach: The URL blocking lists would be populated when governmental agencies provide lists of illegal sites (which are all presumably outside the United States, or they'd be prosecuted and shut down) and sites adjudged to be harmful to minors. Such lists don't exist? Then don't block anything. The burden of proof should be on those wanting to prevent access, not on librarians or online publishers.

Few libraries are ready to take such an assertive and minimalist approach. If the local community has that sort of support, the library won't be using federal funds and won't be using censorware on adult computers anyway.

It may be possible to tailor a CIPA-compliant mechanism that does little to damage First Amendment rights of those seventen and older. It's possible those age ten and younger should either use computers that access only whitelisted sites, or should use filtered computers—although that possibility isn't backed by convincing factual evidence.[5] That leaves tweens and teens. CIPA treats tweens and teens—people aged eleven to sixteen—the same as six-year-olds. It seems unlikely that teens stand to be harmed by sexual images found on free, legal sites. Tweens are a little tougher case, but even for that age group "save the children" is a misleading cry.

But those are policy arguments and CIPA is the dominant policy at the moment.

ALA's FAQ

ALA offered its own answers to frequently asked questions relating to CIPA, which also help illustrate the complexity of the interrelated policy issues. Some of that FAQ discourse includes:

> In cases like this, where no single opinion has the support of a majority of the Justices, the narrower concurring opinions typically govern future interpretations.

It appears that, under the Supreme Court's decision and the government's interpretation of the statute, libraries must turn off the filter upon request by an adult, without inquiring into the adult's "purpose" for disabling the software. In fact, both concurring opinions made clear that any library that burdens patrons' rights through an improper or restrictive application of CIPA's disabling provision could face a future lawsuit.

Minors undoubtedly have Constitutional rights to receive information, but the Court did not address those rights at length in its decision. It is nonetheless clear that CIPA permits minors to request that a library unblock specific websites.

The Supreme Court's various decisions in the CIPA case certainly suggest that a library that imposes filtering requirements without disabling faces a risk of litigation if adult or minor patrons cannot access Constitutionally protected speech.

There is no obligation to use any particular filter in the library. . . . Because the inherent flaws of blocking software make it impossible to ensure that [covered materials] are filtered, a library will be deemed CIPA-compliant as long as it makes a "good faith" effort to block these categories of online materials.[6]

Paraphrasing from other portions of the FAQ, here's what ALA recommends:

Inform the public: Post signs in hard copy and/or on the computer screens informing patrons that:

Because the library receives federal funds, federal law requires blocking software.

The blocking software is inherently imprecise and flawed—it will block access to a "vast array of Constitutionally protected material" and is also "incapable of protecting against access to Internet material that is obscene, child pornography, or harmful to minors."

The library can unblock individual sites that have been blocked erroneously and "will disable the entire filter for adult patrons 17 and over upon request. The requesting patron will not have to explain why he or she is asking that the site be unblocked or that the entire filter be turned off. *The library encourages patrons to request that the filter be disabled."* [Emphasis added]

Facilitate disabling of the filter: In addition to the signs, suggestions include:

Segregating computers for unfiltered access by adults, who would sign a form or display identification showing they're 17 or over and seek unfiltered access "for lawful purposes."

Adopt a smart card system—and have the computers "offer adult patrons the option of Internet access with the filter enabled or disabled," at the welcome screen with a "click to declare you will use the Internet for lawful purposes" button.

Amend Internet use policies to reflect changes or responses to CIPA.

Technology Limitations

Those are the policies and a tiny fraction of the issues. They require effective technology. CIPA requires an Internet filter that only blocks images, that blocks only those images that are obscene, child pornography, or "harmful to children" (and the latter only for computers being used by people younger than seventeen), that can have the filter disabled for a site or an entire session immediately upon any adult user's request, and that can unblock an image (or those images on a site) for someone younger than seventeen who asks, if the library finds that the block is incorrect.

Can any product meet those requirements? Probably not. Depending on the approach taken, it can fail in ways that do more or less harm to Internet

access, but any approach appears doomed to failure—not only now, but permanently (short of universal and universally applied content labeling, which seems highly unlikely).

Filtering software in general doesn't work. It fails to stop a substantial amount of "bad stuff" and blocks a significant amount of "good stuff" in the process. Products that do better on underblocking (letting the bad stuff through) tend to be much worse on overblocking (blocking the good stuff). Much of the material that's overblocked does not fall in the gray areas of legal pornography. Rather, it's accidentally trapped through filtering methodology.

Advances in so-called artificial intelligence are unlikely to solve that problem either for text or, particularly, for images. It's hard enough for software to even recognize the nature of an image (it may not be possible on a general basis).

How would a computer distinguish between an image of sexual activity between eighteen-year-olds (legal pornography, *possibly* filterable for children) and an image of such activity between two sixteen-year-olds (child pornography and illegal in all cases)?

How would a computer distinguish between an artwork involving sexual contact or depiction of sexual organs—there is no shortage of classical statues and paintings involving both, vividly and accurately depicted—and worthless trash involving the same elements?

Nearly all of the filters available at the time CIPA was upheld had more serious technological limitations, some of them based on company policy. Most blocked text as well as images, typically blocking entire pages or sites.

None had categories defined as narrowly as those in CIPA. "Sexual activity" and even "pornography" provide for much broader censorship than CIPA's three areas. Nearly all of them kept lists of blocked sites in encrypted form and considered those lists proprietary, making it impossible for librarians to review the site lists.

Worse, most filters blocked sites not only on a case-by-case basis, but also by algorithms based on text or other aspects of sites that had never been reviewed. Realistically, *most* blocking (including most sites in encrypted lists) is necessarily done by algorithm, not by human review: Otherwise, given the size and growth of the Internet, filters would block a tiny percentage of questionable sites.

Since CIPA was upheld, there's some evidence that one or two commercial providers and open-source solutions come closer to CIPA's requirements: Images only, narrowly defined blocking, immediate disabling of the filter on qualified request—and, to meet library policy needs, site lists that can be reviewed and modified. Can any library actually review a list of millions of sites, many of them with names so nasty that only the bravest librarian would ever go there? That's another question.

Questionable Policy with Even More Questionable Technology

That's where things stand. It's reasonable to assume that most public library directors are loath to have unfiltered Internet computers in the children's room. Ideally, the computers in that room—which should be less prominent than the physical playthings and the physical collection—would be limited to a substantial set of Web sites selected by librarians and other trusted agencies.

Children in this context should mean sub-teenagers, those in elementary school. Cautious libraries might go along with a "parental permission" scheme for terminals outside the children's room and for all underage patrons, but such schemes aren't ideal. A parental-permission scheme uses borrower-ID flags to turn on or off filtering—but any filtering scheme blocks thousands of legitimate site and fails to block thousands of sites that parents don't want their kids to see.

Filters don't work, and don't work in ways that fundamentally violate not only the First Amendment but also library principles. That's fairly well documented. Since the government's witnesses in the CIPA case essentially agreed that filters both overcensor and underfilter, it's hard to make a counterargument other than the absurd, "Some day we'll get it right, just trust us."

Filters designed to meet CIPA policy may be appallingly loose by some community standards. They're likely to underblock substantially in order to avoid severe overblocking. That means there *will* be "offensive" images on library computers at some point, even in children's areas (if those computers don't use whitelist filtering). That may be the best you can do, and maybe it's the best for which libraries *should* aim.

Metasearch

Metasearch goes by many names: Federated searching, distributed search, broadcast search, cross-database searching, sometimes even portal.

There's nothing new about cross-database search and retrieval: Dialog and other services were doing it long before the Web. In some cases, metasearch is just a fancy word for cross-database search and retrieval. But it's really more than that, partly because the databases searched aren't all from the same supplier, partly because it can involve so many databases.

You can see metasearch at work in the Web environment at clusty, ez2www, www.info.com, and a number of other sites. These sites send a search out to some number of search engines (for example, Google, MSN, Yahoo!), bring back a portion of the results from each engine, remove duplicates, and display a combined result set (possibly including the rank in each search engine). The metasearch system may also add value, for example, by suggesting topical clusters within the result.

Web metasearch is relatively straightforward. Since every Web search engine returns sites in "relevance" order, it's plausible to do a combined relevance ranking. Since every Web search result includes a unique URL, it's trivially easy to combine duplicate results from different sites.

Library metasearch (and campus metasearch) may be more ambitious: Hundreds, sometimes thousands, of databases may be involved, as opposed to the handful of Web search engines. It's also more difficult because those databases lack the consistency of Web search engines and because results don't have convenient unique labels.

Policies and Desires

Library metasearch attempts to satisfy a set of library policies and desires. Briefly, the policy (stated or unstated) is to make the library's resources readily accessible to users. More broadly, that includes the desire to make

searching easy and consistent and to expose resources that might otherwise be hidden and go unused.

Libraries also have somewhat conflicting policies. Specialized resources are acquired to serve specialized needs and may have specialized interfaces to serve those needs. Libraries need to serve expert users in narrow fields as well as novice users with broad needs. That means supporting native interfaces for databases as well as the single metasearch interface. Typically, it also means making movement from a metasearch result to the databases represented in that result as simple as possible.

A metasearch interface may ask the user to select databases or groups of databases first, or it may suggest appropriate databases based on a search. The metasearch interface may present a combined set of results or list the result size for each database that produced results, with links to get those records. The metasearch interface may encourage the user to switch to native interfaces for preferred databases or make it easier to stay within the one-stop search shop.

The biggest advantage of metasearch is that users enter a search once, usually in a Google-like single box, and (in some cases) don't have to puzzle over which of several hundred databases will suit their needs. The metasearch engine may offer other advantages, but that's the primary benefit.

Well-implemented metasearch has real advantages for many libraries, for many users, for many situations. Individual databases and online catalogs with their native search interfaces also have substantial advantages over metasearch for many libraries, for many users, for many situations.

One solution is rarely right for all circumstances. One user interface or one search technique is unlikely to suit all circumstances. "The user" is even less meaningful as a term than "the library." Faculty and post-doctoral students aren't the same as freshmen, and students taking honors courses in their majors have different needs than students rushing out a survey course paper.

Current Technology Limitations

Library metasearch systems can be problematic due to several technological issues, including the following:

- Metasearch almost always means lowest-common-denominator searching, using those few search techniques that appear comparably implemented among a range of databases.

- Response time may vary wildly among a set of databases, and the best results may arrive after the user already has decided to use early arriving results.

- Databases vary in size enormously, from specialized databases with thousands or tens of thousands of records to bibliographic databases with tens of *millions* of records. Mixing results from such disparate databases may tend to make the most relevant results difficult to find among less relevant results from much larger databases.

- Databases vary in type: Full-text aggregations with minimal metadata, catalogs and pure indexes with rich metadata but no full text, databases that aren't textual at all, and crawled Web pages that aren't databases. These varied types make relevance ranking and common presentation formats difficult.

- Many library-related databases don't return results in "relevance" ranking, and there is no common definition of relevance among those that do. Some databases don't sort results at all, particularly for large results. When the first one hundred records from one "relevance-ranked" database are combined with the first one hundred from a database that returns results in alphabetic order by article title and with another that returns them in reverse chronological order (most recent first), any attempt to provide a common-ranked listing will be misleading at best.

- Essentially, it is impossible (under current conditions) to retrieve *all* of the records in a metasearch result, which means it's also impossible for a metasearch engine to apply its own relevance ranking in a fair and robust manner.

- The most common search-and-retrieve protocol, Z39.50, is far from universal among library databases and rare outside the library field. It's also implemented differently in different databases.

- Meaningful de-duplication may be difficult or impossible across a range of resources, given differences in metadata and the lack of universal standards for metadata outside of traditional cataloging.

Do these technological limitations mean that metasearch is a bad idea or should be abandoned? Not at all; but they do mean it needs to be approached realistically and in view of its limitations.

The heading of this section says "Current Technology Limitations": That's because these limitations may be overcome in the future, at least to a considerable degree. A computer may never be able to determine whether a given image is or is not obscene or harmful to children—but advances in standards and in cheap computing and telecommunications may make metasearch limitations less meaningful over time.

The National Information Standards Organization (NISO) has established a Metasearch Initiative to address some of the issues noted previously. That initiative won't solve all of the problems, but it will almost certainly alleviate some of them.

Conclusion

It's comforting to believe that technology can always solve the problems caused by technology, and that any technological limitation will disappear in a few years. Neither is true nor seems likely to become true. Policymakers may assume more of technology than technology can deliver. Compromises need to be made—sometimes in policy as well as technology.

Notes

[1] Lori Bowen Ayre, "Filtering and Filter Software," *Library Technology Reports* 40, no. 2 (Mar./Apr. 2004).

[2] Nancy Kalikow Maxwell, "Alternatives to Filters," *Library Technology Reports* 37, no. 2 (Mar./Apr. 2001).

[3] See *A Starting Point: Legal Implications of Internet Filtering*, September 2004, The OpenNet Initiative, for a discussion of policy issues and conflicts in filter use in general, including corporate and government filtering.

www.openinitative.org

www.fepproject.org/whitePapers/NRC/whitePapers.html

www.ala.org/washoff/WOissues/civilliberties/cipaweb/adviceresources/questionsanswers.htm

[4] You'll find more extensive selections from the decision and discussions of that decision in *Cites & Insights: Crawford at Large* 3, no. 9 (Midsummer 2003), a twenty-page issue entirely devoted to the CIPA decision. Much of the commentary in this section is rewritten from that issue.

[5] Marjorie Heins makes a strong case that there's no scientific or objective proof that material deemed "harmful to minors" *is* harmful to minors, in "Identifying What Is Harmful or Inappropriate for Minors," a March 5, 2001, white paper for the Free Expression Policy Project.

[6] CIPA FAQ on the ALA Web site, from a version dated July 8, 2003.