

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

Participants

William Arms, Cornell University

Keith Kupferschmid, Software &
Information Industry Association

Scott Teissler, Turner
Broadcasting Systems, Inc.

Sarah Wiant, American
Association of Law Libraries

Roy Kaufman, John Wiley & Sons,
Inc.

Rebecca Pressman

Howard Besser,
InterPARES/consortium to
preserve public television in
digital form/New York University

Kenneth Crews, Copyright
Management Center

Edward Lee Lamoureux, Bradley
University

Ken Frazier, Association of
Research Libraries and American
Library Association

Victor Perlman, American Society
of Media Photographers, Inc.

Janice Pilch, University of
Illinois at Urbana-Champaign

Carl Johnson, Brigham Young
University

Allan Adler, Association of
American Publishers

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MR. RUDICK: We do have one new person, I think. Pronounce your name and tell us where you're from.

MR. TEISSLER: Scott Teissler, I'm the chief technology officer and the chief information officer and the executive vice-president for technical operations for Turner Broadcasting and for CNN.

MR. RUDICK: Preservation of websites -- we in the Study Group recognized that websites, however we may want to define those, present special issues. I remember that two years ago, when the National Digital Strategy Advisory Group in the Library of Congress had its first meeting, this was one of the first things we talked about.

Many people said, my god, we've got to do something about this. Well, some people have been doing something about it.

Now, we're going to talk about how and in what way our law should change to officially recognize something that people have already recognized in practical terms.

We've talked about ephemeral in the context -- or in a more technical way, but now we're talking about ephemeral in a different nature, which is, we're talking about an area of activity which is perhaps not intended by the people who created it, to be for posterity, but this is part of our culture, maybe not the most elegant part, but an important part.

Historians, archivists, librarians, care about this as much as they do about other things that they preserve for posterity. There are a couple of unique characteristics.

Websites are generally intended by people who put them up -- maybe less worried than the typical producer of copyrighted content, about whether there are copies. They may want them copied.

They don't have much of an incentive to keep and preserve things. In some cases, having embarrassed themselves, they may hope that what they did was destroyed.

There was a reference to the websites of political parties a moment ago, and that's a good example. They should be embarrassed; I'm embarrassed, and that's a nonpartisan remark, actually.

There are a number of specific issues and we'll come to the fact that it's difficult to capture websites, except through harvesting processes that are somewhat automated. Even though the answers to some of these questions may be implicit and we may think we know what they are, we need to discuss them and we need to discuss the appropriate details.

So, given what I've discussed -- sorry --

MS. RASENBERGER: One clarification: That is, when we refer to websites, it's a shorthand way of talking about online content.

MS. GASAWAY: No, no, no.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MS. RASENBERGER: Well, yes, but Internet content, which is what we have in the Federal Register Notice.

MS. GASAWAY: But we're not talking about licensed journals.

MS. RASENBERGER: We're not limiting it to the worldwide web. There may be additional technologies.

MS. GASAWAY: Okay.

MR. RUDICK: We're talking about stuff that's put out, wherever it's put out, without restrictions.

MS. RASENBERGER: Yes, unrestricted, publicly-available content. That's one of the questions.

MR. RUDICK: Given the general characteristics that we have just discussed, or that I have just expressed, I should say, I'm going to read all the questions, then we'll go back to the first one:

Should a special exception be created to permit the online capture and preservation by libraries and archives, of certain website or online content?

Second, if so, what types of limits should be imposed? Examples could be a class such as something that's noncommercial, something that's not an object of commerce. You could refer to whether access is restricted by passwords or registration; access controls or the absence of them.

Then going on from there, we're going to ask you whether the copyright owner should be permitted to opt out, and, if so, how? Should robots.txt, or other no-trespassing commands be available; should there be some way of notifying or notification that the owner wants the content to be captured and preserved.

Finally, the last question is, should there be restrictions on public access or not, to content so captured, and, if so, of what nature?

So, the first question: Should there be a special exception, and, if so, what limits should be imposed?

MS. WIANT: I would like to see libraries allowed to preserve online content, in particular, speaking from the perspective of the law, there are many web pages that are created sometimes by state and local government, actually, sometimes even by the Federal Government, that, when changed, in parties, the information just totally disappears.

Sometimes it's not just the parties; sometimes information that's gathered by the Government, may be up for a particular window of time, and then just totally disappear, and that may be the only record.

For instance, federal committees sometimes build web pages and have very valuable information on them, but at the end of the term of Congress, those committee pages completely disappear and that information is just totally lost.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

I would like to see, at least under some circumstances, that libraries be allowed to enter those pages.

MS. RASENBERGER: Keith?

MR. KUPFERSCHMID: I hate to sound like a broken record, but I will. As I said this morning, I think this is also a question that raises a whole bunch of other questions, and it's hard to answer the question in a vacuum, not knowing who this exception would apply to, if the exception would include definitions of library and archives or whomever is permitted to take advantage of this exception; what we mean by "other online content." As a matter of fact, there was a little back-and-forth about that.

Perhaps most importantly, from the perspective of SIIA and our members, I hate to bring up the acronym, DMCA, but I will. How does a DMCA apply to these websites that are archived?

What happens to these archived websites that include, for instance, pirated material? That material has been taken down, and that's why you can't find it, pursuant to a notice and takedown under the DMCA.

What is the entity that captures and preserves that archive? What are their responsibilities here?

Let me give a real live example that happened to us, gosh, about a year and a half ago. We had some content -- content, not software -- that was stolen and posted on a couple of websites. It soon blossomed into about 20 websites.

We sent notice of takedowns, got the materials taken down, but the material was cast on an archival site and the people who had it on the website and took it down pursuant to a notice of takedown, were very happy to provide links to the material up on this cached archive version of it.

We had to take it down, but you can still find it here. We contacted the entity. I won't name names here, but we contacted the entity and asked them to take it down, and they did eventually take it down.

But it did take quite a long time, and we were very thankful to actually be aware of that to be able to contact them.

They were somewhat aware of the DMCA, actually, but that's a problem. We have an exception along the lines of allowing online capture and preservation of websites.

It's not just going to be the one entity that's doing it; there are going to be hundreds, if not more, doing that, that will require the copyright holders to start tracking those down to get it taken down and off the original websites.

Also, there are related concerns in addition to the DMCA one that I mentioned. They have to do with the impact that an exception here might have on those who operate the website, instead of being captured, and those whose content is contained on those websites.

It's really essential, as I pointed out before, that any sort of capture and preservation activities, have no negative impacts on those who maintain or originate the website or its content.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

For example, just because the website's content is no longer available for free, or at all, does not mean the user should be able to get that content from the person who captured the content.

I'll give you another example, then I'll close up here, so I won't go past my three minutes. We had another case of piracy, this one taking place on Yahoo. This is a real, really scary example, because what happened was, this user group was pirating medical texts.

At least in one instance of pirating this medical text, they had copied -- I mentioned dosage tables earlier -- they copied dosage tables, and those who were copying the dosage tables, screwed up and they were just moved up one, so there was actually a fatal amount in one of the dosage tables, and you had better believe that we hoped no one ever sees it again, because that could be a real problem.

And if there are people archiving these websites, who just don't comply, that type of situation can be a real problem.

Just to close the loop on that, I should mention that there are several people who are being prosecuted for this. The ringleader is someone named Adam Pariah. I should mention that he will get up to two years in prison. He was a pediatrician, and he was also convicted on child pornography charges. You get an idea of the type of individual we're dealing with here.

MR. FRAZIER: Adam Pariah?

(Laughter.)

MR. KUPFERSCHMID: There are a lot of questions that need to be answered before we can move down this path to consider an exception in this context. We're happy to talk about this, but, realize there are real concerns. We've had them before when there was no such exception.

If an exception were to be crafted, I fear that we would have a lot more problems just like the two I identified.

MR. RUDICK: Thank you. Howard?

MR. BESSER: When someone goes to court and tries to prove that someone had ripped off their content on the web, what do they bring as evidence? They go to a library or archive or they go to the Internet Archive, in order to get that to bring in as evidence.

In fact, there's a major social value in preserving these things that helps the content providers. The Internet Archive identified a large number of requests from -- a large proportion of their requests for information come from people who want to see what their web pages looked like a few years ago, because they don't have them anymore.

I'd just like to harken back to about 35 years ago when CBS sued the Vanderbilt Television News Archive, because they said this is depriving them of their rights under copyright law, that the recording of the Nightly News was a great harm to them.

In fact, here I have one of my student's Ph.D. dissertations on this. He's been doing research for the last year and a half on the history of the Vanderbilt issue.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

What happened then, before it was settled in court, Congress came in and crafted the 108 exception for audiovisual news. What we've seen in the last 35 years is no significant market harm to the TV industry, a huge social benefit in terms of preservation of our heritage, teaching and research into the past.

What started out the Vanderbilt Archive, was holding public servants and holding the media accountable for things that they claimed. So, one of my questions is, do we want to have this contention of a lawsuit and a new act of Congress hanging over us, every time a new technology or a new genre comes along, that, in fact --

And I look at Vanderbilt as a really good example of something that was opposed by rights-holders, but, in fact, ended up being to their benefit.

MR. RUDICK: Scott is next. I'll just review the queue to see if anybody is missing. After Scott, I've got Ken Frazier and Bill Arms and Edward.

MR. TEISSLER: This is a bit of a macro answer. You know, a thing that's different here is that the scale and the scope and the complexity and diversity and the rate of change and the rate of evolution of the universe of website or website-like content or things websites might become, is vast beyond any of these examples we're comparing it to.

Our view is that without some help, which I'll describe in a minute, the community of archives and libraries is completely over-matched, facing this array of content going forward.

You've got kind of a choice here, choosing to win a battle in this microscopic part of the total space, and that microscopic part of the total space, with respect to preserving it, without better tools, without better cognizance on the part of content creators, who are going to have to make provisions for their content to be archive-able, without a set of emergency standards and practices to facilitate archiving, you're going to wind up archiving a negligible fraction of the available universe.

And, opting out, as opposed to opting in, isn't going to help. In the case of opting in, you presumably established a coincident intention to make it a work tool to establish the basis for common practice, to establish the basis for an industry of tool providers to emerge to help.

The thing I encourage here is, you watch out for the case of the specific example, consider the size of the problem confronting the current size of the solution, and ask ourselves what we can do to make sure that the available means to a solution grow.

MS. GASAWAY: I just wanted to remind you that the opt-in and opt-out comes up on another --

MR. TEISSLER: I'll repeat myself later.

MS. GASAWAY: Because that's going to be the next question. Bill? Oh, I'm sorry, Ken.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MR. FRAZIER: The examples cited about pirated content, seem to me to prove just the opposite point. Content was taken down. We're talking about freely-available websites. Here we have an example where you not only have the benefit of being able to support a charge of privacy, but the content does come down.

We're also dealing with content that is freely available, that is out there on the Internet. One could say, to borrow a phrase that I heard on the break, that there's kind of an implied license that it be available and kept preserved.

It seems to me that the issue also that the information might be wrong, is so ubiquitously true across all kinds of information, that citing that as a reason why it should be preserved, strikes me as yet another way to get information.

MR. ARMS: Actually, Ken made the central point I wanted to make. I think we all have tremendous sympathy for your examples, but I don't think any exemption of the types proposed, would make any difference to it.

The soft-spot piracy you're talking about, is being put out on unrestricted sites that any web crawler can grab. How to propagate notice of takedown of a sort of complex system, I think, actually, a well-managed archival system might help. It certainly won't make that problem any more difficult.

MR. RUDICK: Ed?

MR. LAMOUREUX: I wanted to address the issue of restricted public access archives as an end product. Is it time?

MR. RUDICK: No.

MR. LAMOUREUX: Do you want to wait?

MR. RUDICK: Yes.

(Laughter.)

MR. RUDICK: We appreciate it. Roy, I think you were next.

MR. KAUFMAN: One quick observation and then one thing that we just need to watch out for: With respect to Government websites, Federal Government websites, public domain, I don't think there's anyone here who would object to a rule enabling freely caching local government and state government websites or question the validity of that.

That's only a small part of the phenomenon, but we can probably move past that one pretty easily, because I think there's a huge public interest in being able to maintain those websites.

To the extent that the local governments may have some arguable copyright protections, that's a good place for this.

The other thing is, I just sort of want to respond to something that Scott said, which is that this is very hard. You guys are in the

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

trenches and we think that we are, actually, as publishers, in the archival and preservation trenches, as well.

Note how hard this is, note how little of the content we'll really be able to save, at the end of the day. With all the things that exist, whatever we do, we should make sure that we don't create exemptions that discourage private initiatives by undercutting any financial incentive or ability to create financial incentives for archival projects.

I don't know what the answer to that question is, but there's got to be a lot of different answers to these problems.

With exceptions, you need to be very careful that you don't disincentivize entities who might have a commercial interest in trying to preserve their own material. Make sure that you don't undercut that, so that it doesn't get done by anyone, or gets done inefficiently.

MR. RUDICK: Okay, anyone else? Ken?

MR. CREWS: Just a reminder footnote: When we do get to opt-in/opt-out, help us remember what somebody said in Los Angeles about the restrictions on the White House website. You might be surprised, Roy.

MR. KAUFMAN: Hey, it's public domain.

MS. GASAWAY: Some of them have a .txt file.

MR. CREWS: But my point is, seeing the kind of curve that this discussion in the last several minutes has taken, I wonder what we're trying to accomplish here.

If we're doing this in the shadow of the Internet Archive and a few such large-scale initiatives, downloading massive amounts of material, and, as we heard from Mr. Kahle in the previous roundtable, not getting a lot of complaints about what he's doing, and being able to respond to complaints quite efficiently, to remove material at the request of rights-holders, if that's working in the marketplace, and, in effect, being litigated, which it is, we're going to have the benefit of sitting tight here and watching what the courts think about this.

I wonder if we can do anything better than what's currently going on here. Let's let the marketplace experiment for awhile; let's let the fair use experiment for awhile; let's hear from the courts, and let's see how this evolves in the not-too-distant future.

I think any actions to try to be specific in the statute today, may be premature.

MR. RUDICK: Lolly wants to ask a follow-up, and I do, too.

MS. GASAWAY: I don't know if it's a follow-up or just a reminder. As you mentioned, you have the Internet Archive, but we have also received information from other libraries that want to do a much more targeted thing, just the websites of political parties in the last election or just something like that.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

What the Internet Archive is doing, does not necessarily answer that, because there's a curation function that goes into the other one. I just didn't want us to lose that, that there might be other sorts of structures or other entities that want to do much smaller portions.

MR. RUDICK: Kenny, in response to your point about let things shake out under 107, I know the Library of Congress has told us in the 108 Group, for example, how urgent some of these things are. I think that after Katrina, you were asked to assemble.

MS. RASENBERGER: There are people in this room who could talk about what the Library's is doing, much better than I could, but the Library of Congress does have website preservation, generally speaking.

It's been event-based. You mentioned Katrina, which we've done in partnerships; the elections, the war in Iraq. We have largely gone on a permissions-basis, which, needless to say, is almost unworkable, I think.

On one collection, I know -- I don't know if you want to pop up and say anything about it, but there's something like 3,600 hours on permissions, with a 24-percent response rate.

MS. GASAWAY: And I think it should have been public domain. It was political websites for the election.

MR. CREWS: If I may follow up on this one with just one or two sentences, why didn't you just use fair use?

MS. RASENBERGER: The Copyright Office sits in the Library of Congress, and we don't believe that capturing on a massive basis like that, is necessarily fair use.

MR. RUDICK: Let me get you off the hook. (Laughter.)

MR. RUDICK: We have ten minutes more on this question. What I would like to encourage us to do, is to focus on -- while there are bullets in my copy, I don't know that you could hear that they were bullets when I described them, but let's do a hypothetical.

Let's say we had an exemption -- forget opt-in and -out. That's coming later.

Should it be limited to a class of websites? What should the standards be? Noncommercial? Object of commerce? I don't know what words you want to use, but if you were to say this type of website shouldn't be included, somebody may think every website should be included.

But let's say you were going to try to put some limits around this. What would those limits be? Would it relate to the use of passwords, access controls? If you think there should be limits, what should they be? Bill?

MS. GASAWAY: Scott came in first.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MR. RUDICK: My eyesight and hearing are limited.

MR. TEISSLER: I say good luck in deciding what a website is, what it includes and what it doesn't include. Good luck now, good luck five years from now, good luck ten years from now, because it's a recipe for investing a lot of energy in a futile mission, at least generating enough cases of the polymorphous perverse to undo what good it may do for the bulk of average websites.

MR. RUDICK: But we're ambitious and can use all the luck we can get.

MR. TEISSLER: This is a good way to make the problem even harder.

MR. RUDICK: Bill?

MR. ARMS: Having said that, let me try a strawman. I think the only sensible division is material that has been put up on the networks with no restrictions on access by the general public.

That is the only sensible dividing line. If there is a password or any other sort of restriction, it's out of scope.

MS. GASAWAY: You say, being put up for the general public, with no access restrictions?

MR. ARMS: Being put with no restrictions on access by the general public.

MR. LAMOUREUX: I assume, following up on that, it means captured with the same restrictions.

MR. RUDICK: Say that again.

MR. LAMOUREUX: If there are restrictions to the material, the capture would capture the gateway. If I needed a password before, I'd still need a password again. I can't circumvent the restriction by looking at the archive.

I kind of wanted to speak to that general principle. The notion that I would promote through archival means, limiting access to perfectly accessible web pages, as a function of archiving, is, to me, counterintuitive to the very notion of cyberspace.

I just cannot imagine it, that the law would encourage locking up, that which is freely available, because it was only freely available yesterday, and, because time has passed, now the procedures are so different that if we go to look at that, in what ways would we restrict public access to that which was, yesterday's worldwide web, other than as those restrictions had been in place on that day?

MR. RUDICK: In the queue, I have Janice, Bill, Keith, Ken Frazier. Okay, Janice?

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MS. PILCH: I have a similar thing to say. As people put material up on the web for the public to see for free, that's the kind of material that we're talking about -- unrestricted free access to material intended for the public's viewing and use.

And, you know, this almost gets into a similar thing of, you know, restricting a digital copy to four walls. But we're not going to go there right now.

It's artificial then to try to limit it. Such websites should be preserved exactly as they are created in the first place.

The second thing to say is that -- responding to what Laura said a minute ago about libraries wanting to create limited captures of certain websites, if the threat and the danger of capturing unlawful copies of material that some people are concerned about, is eliminated, because libraries will be copying only materials they know not to be causing problems, why should we not allow this exception?

It is, again, causing no harm, for preservation purposes.

MR. RUDICK: Bill Arms.

MR. ARMS: I jumped the queue. Someone else with my name contributed that.

(Laughter.)

MR. RUDICK: Keith?

MR. KUPFERSCHMID: I agree with what Bill said before about restricting this to what is publicly available, as long as you're not hacking through passwords or firewalls or anything of that nature, and you're just talking about publicly- available, accessible websites.

I think the focus here should be websites, rather than the content. By focusing on the website, the nature of the website really tells you more about the intentions and missions of the owner of that website, rather than if you focus on the specific items of content themselves.

I think that also, you've got to consider different business models out there. There are websites, existing websites or the ones that come most often to mind, that will provide content, today's content for free, and maybe even yesterday's and the day before.

But, after awhile, they start restricting that. Maybe they restrict last week's content. You actually have to register, you have to pay some kind of subscription.

When you get to last year's content, you actually have to pay for versions of that. That kind of business model has to be accounted for.

Just because you were giving it away for free today, doesn't mean that it's free tomorrow. In addition, I know a couple of companies, offhand, that their information is freely available on the web maybe for today and tomorrow, but eventually, what they do is, they restrict access to that information, then they use that information and they repackage it in other ways, a lot of time as historical data or historical information. That also needs to be accounted for.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

Are we talking about robot exclusions later?

MS. GASAWAY: In about three minutes.

MR. KUPFERSCHMID: All right.

MR. RUDICK: We have three in the queue. If you're brief, we can accommodate all of you. Ken Frazier, Scott, then Alan.

MR. FRAZIER: Very briefly, one of the most common forms of restriction is registration, which isn't intended to be a restriction at all. It's a way of being able to identify user communities.

So, if you view registration as a restriction, you could be failing to preserve a lot of content that ought to be preserved.

With respect to the business models that have content available freely now, and then for sale later on, it seems to me that that business model is completely compatible as an archiving strategy for the open Internet. You could do both things and still have the opportunity to archive open, free websites.

MR. RUDICK: Scott?

MR. TEISLER: I just wanted to point out a couple of things about being publicly available. cnn.com is publicly available. There's some content on there that's sort of our content, which we're not particularly concerned about being copied, because it's in use.

But there's a lot of content in there that we, in turn, get under license, and which we're allowed to put up for a finite amount of time under such licenses, even though there are no barriers between the public and that content.

A lot of DRM is transparent, by the way. There may not be a password or some protection scheme; it may simply be there to allow somebody to see something, but not to copy it.

So, again, without tools, further help and cooperation between the sites and the archives, there aren't really good ways to solve these problems.

MR. RUDICK: Mary, with a brief question.

MS. RASENBERGER: Sure. In hearing what some of you are saying, I'm wondering if it could be addressed through a concept such as whether the content being called up, is, in itself, an object of commerce. How do you distinguish between those kinds of sites where the site owner might have an issue with it being preserved and called, and those that are not, which is probably the majority of websites out there, which is -- their purely informational content isn't what they are selling; it's more like sort of the leaflets, the pamphlets of the past.

I just would like to get your reaction on how that notion sticks with you, if the restriction would be, if the content is, in and of itself, an object of commerce.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MR. RUDICK: We have in the queue, Alan, Ed, maybe Howard, and then we have to cut it off and move on.

MR. ADLER: I was going to add to some of the examples Keith had mentioned before. This comes into Mary's comments, too.

Within my industry, for example, a website is part of the product. In the educational field, for example, I sell textbooks that come together with a DVD and the Mac sleeve, but also come with an access password to a website.

The website is designed to work specifically in conjunction with the textbook and the DVD. Some people might say, so what? It's a website, it's up there, you're not taking the textbook, you're not taking the DVD.

Nevertheless, it's part of the package product. In that respect, I think that probably copyright owners would object that that should be treated as any other website that may be there for information purposes, with our without the restrictions.

MR. RUDICK: Ed?

MR. LAMOUREUX: I can't speak directly to the technology of it, but it strikes me that a non- archive metatag here, is appropriate and would probably solve about 99.99 percent of these problems.

If you're a newspaper or CNN and you want to pull your stuff back or charge for it, you've got one metatag that says this ain't going into an archive, because that's how we do it, and then we just pass it on.

But Ken's stuff, that teaches the world about copyright, and he only charges -- what was that? \$220,000 for the PowerPoints?

MR. CREWS: At least.

(Laughter.)

MR. LAMOUREUX: That one is free and available. It's a fairly easy, I think, technological solution.

MS. RASENBERGER: Which gets us to the opt-out question.

MR. RUDICK: Before we do that, Howard, you're the anchor man, except for Jule.

MR. BESSER: Let me try to address part of what Mary's question, in a way to try to get at some of the industry reaction against some of this.

I think if we're thinking in really long periods of time, if we are not thinking that we're going to archive this website and we're going to make it available next week, next year, ten years from now, if we're thinking in terms of 50 years, hundred years, a thousand years, if we're thinking in those types of terms, I think most of the rights-

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

holders are not very concerned about how that's going to break their business model.

And, you know, where my concern is, is that these things that represent our culture, our history of this time period, be accessible in the future, and I don't -- history has told us that we cannot trust the publishers, the studios or whatever, to preserve this, and we shouldn't; that's not their job; that's the job of libraries.

But it's a job that becomes more and more difficult. If we're trying to look in that long run, then we're not necessarily imposing any kind of problem on top of business models.

You know, I don't think that anything that I've heard from you folks here, really -- you're really seriously concerned about being able to get something 50 or 100 years from now, and that that disturbs the business model.

Related to that is also the evidentiary kind of thing. In fact, the Vanderbilt Archive began from a very conservative man who was upset with the news that he thought was too liberal for the time period, and he wanted to be able to have proof to show that these things were this way at that time.

Again, that's a kind of function that maybe in some ways is an archival function of evidence; in some ways -- this is what I brought earlier -- it's a legal function for you to go after someone who has taken the website, and you can prove and get the Internet Archive to show that, in fact, they copied your website.

MR. RUDICK: Thank you. If any content owner wants to respond, you can submit written comments. Jule, unless you have a quick one --

MR. SIGALL: Just a quick one to make the observation that it seems like most people are talking about this problem in terms of trying to determine the intent of the person who created the website, as to whether it should be archived or not.

One of the problems seems to be that we're using proxies for that intent, in some respects, in the way that they generalize or generally allow or disallow access to a particular portion of the website.

The problem is, sometimes that access may have nothing to do with archiving the content. It may be for business reasons; it may be for other reasons, to find out who the user community is.

So that's merely a proxy. It sometimes can be a wrong proxy. We'll talk, I assume, in the next section about robot.txt. That's one way you can ascertain the intent with respect to archiving.

Then there are other circumstances where you may want to permit archiving, even in defiance of that intent, for certain types of material on the Web.

So, to me, the question is, what kinds of mechanisms can be put into the statute to accurately assess the intent of the parties and also what circumstances do we want to allow archiving, even in disregard of that intent? That gets into harder questions, which is really what opt out and opt in are about.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

On whom should the burden fall to ascertain or assert the intent with respect to archiving? That is, I think, the question on opt out and opt in.

MS. GASAWAY: Great.

MS. RASENBERGER: We've got two more questions. We've got less than half an hour.

MS. GASAWAY: We've got -- we'll do 15 and 15, then. We're going to repeat them.

The first one is the opt-in/opt-out issue; the second one is restrictions on public access to captured content. So we might have an embargo period or what have you.

On the opt-in/opt-out issue, if the Internet works as an opt-out system, should robot.txt or other similar no-trespassing commands be honored by anyone capturing it? And would that give content owners sufficient confidence that their material was not being archived, if they didn't want it to be?

That's the question. Then we can talk about opt in, but let's do opt out first. That's sort of the easier one, I think. Bill?

MR. ARMS: I've made some notes here, and Jule's three points were almost identical to what I said or what I was writing here, with a few subtle differences.

First of all, I think we need a technique by which websites can indicate their policies and preferences, and we need something, and maybe the robot.txt model is a good one, one of the key things being that it's simple and it's automated.

We've got so much data, these things have got to be automated. So these questions, is there access now or do you intend to make it closed access in the future, is the sort of thing that can be indicated by the site.

The second is that the default should be that there are no restrictions, that if there's no indication made, the fact that this has been put up with open access, should be taken as a default that you can go ahead and preserve it and make it available to others.

The interesting one is, are there some occasions in which the public interest makes it important to preserve, even overriding some of the preferences? Let me give you two examples, which I think are really important:

First of all, there are people and organizations who wish to obliterate history. And your example of state governments taking things down when they change, is a good example.

The websites of prominent figures such as politicians and other examples, I think there's a public interest in these cases that should be preserved. The other, and perhaps I think the answer to Kenny's question of why can't we just leave the present situation as it is, to me, is the newspapers.

The newspaper collections of our major libraries are one of the most important collections that they have, and, at present, there is no

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

way they're being collected. I personally believe that the Library of Congress should not be allowed to collect major newspapers.

I actually believe that they should be required to collect the major newspapers, because I think they are so important. I know the Library of Congress well enough to know that they're very sensitive.

They are the Library of Congress, and they respond to the same pressures as Congress responds to. But I think there are some things like that, which, under carefully controlled situations, need to be archived and preserved, even if the owners are reluctant to have that happen.

The next question is restrictions on access. Clearly, this has got to be combined with controlled access, so I think Jule had it right.

MS. GASAWAY: I have Roy and then Ed.

MR. KAUFMAN: Some of what Bill said, tying in what I think Kenny said, gets back to a lot of the issues here. I'm not going to go back over what we discussed before, but there are so many things here that we're trying to write an exemption for, that really covers all of these things -- opt- in/opt-out, but no opt-out, if there is compelling public interest.

You sort of do come back with, well, once we write this, this becomes, did you do this, did you not do this? Sometimes, fair use, for all its ambiguity, has a beauty to it, which is to say this is political stuff, it's most likely to be protected by fair use, regardless of whether the RNC or someone else says it should or shouldn't be copied or searched.

Every time we try to go down the drafting thing, you become somewhat limited in what you are -- from a user perspective, I'm as worried as I am from the publisher perspective. I want important stuff to be kept, but if we have to spell everything out, we won't get everything, so if we do our best to spell out, we'll miss stuff, the stuff that falls out, but is enumerated under the legal theory that I learned the first week of law school, which is, basically, if you didn't do this, you can't do it and you're not going to be able to look to fair use. Maybe this is an area or maybe in some certain areas here where it's difficult, we should just accept fair use as a value and not just try to write around all of this.

MS. GASAWAY: Anyone else? Keith and Kenny?

MR. CREWS: Being a great fan of fair use, Roy, I appreciate the added insights and support on that.

Fair use has the additional ability of allowing us to respond to something that I emphasized earlier in the afternoon, and I'm seeing it emerge in this discussion, as well.

I see the value for one standard for being able to make the copy at all, then the possibility of a different standard for what you can do with that copy, once you have it.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

Fair use allows that flexibility. If your question is, can I make that copy, fair use may give, relatively speaking, a broad, optimistic answer.

If the question then becomes, can I put that copy out for the rest of the world to access off of my server, fair use can give an answer to that, too. Surely, I think most of us would agree that it's not as broad, whatever that means, as the first answer.

Fair use offers flexibility to respond to different needs, different circumstances, in the big, unpredictable future that lies ahead.

MR. KUPFERSCHMID: That's a step in the right direction, because that gives a good idea of what the intent of the website operator is. And, certainly, with regard to any exception in this area, an archive not able to demonstrate a pattern of practice of abiding by that rule, they shouldn't be able to take advantage of this exception.

Once again, it's very tough to look at these issues in a vacuum. Howard mentioned previously that these websites aren't going to be made available for 50 or maybe a hundred years.

None of this is written down, none of this is agreed to. All of this can certainly change in perspective, as opposed whether it's going to be available next week or a month from now.

Obviously, you can imagine that the issues are a lot different to publishers, depending on exactly if and when this is going to be made publicly available.

The last point I'll make is a comment within this group of questions about opt-out provisions, which is the Federal Register notice related to the software.

I was very confused by this question. I'm perhaps the only one around the table that represents the software industry. Most, if not all of the software that's needed to read websites, is now own- able for free.

You've got Macromedia Flash and Adobe Acrobat Reader, and there is other software out there that is available for free. So I'm a little confused about the question about needing to download copies of software.

If you're talking about operating software such as Acrobat Reader and Acrobat Writer, that's a very different question. I know my software companies that I represent, would be very concerned, because they have archival software.

That would be database software that would be a significant issue.

From my standpoint, I would have to find out a lot more about this to be able to give a more knowledgeable response.

MS. GASAWAY: Howard, Scott, then Ken. Just in order to point out, so that we make it clear, Howard commented on the 50 to a hundred years before it would be available.

I don't think the Working Group has necessarily said that's where we would be. We don't know.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

I guess the things from the last election might be interesting to somebody before 50 years, but I may be wrong. I just wanted to make sure that you knew that that was just coming from another participant at the roundtable, that the Study Group has not set a 50-year embargo at all.

MS. RASENBERGER: In LA, in fact, I think six months is what happened.

MR. KUPFERSCHMID: Six months, a hundred years.

(Laughter.)

MR. BESSER: What I was trying to suggest with that, is that there may be some common ground we can find for different purposes, and, again, you'd have to come back to what the purpose is. The kind I was saying - - my purpose, one of the highest priorities for me, is that we won't lose our record of this time period. And for that kind of thing, that could be embargoed for quite a while.

There are other purposes, as well, but maybe we can reach some common ground right now, so let me address a couple other things. Keith raised the issue about software.

I think the real question is where we get a quick time player. We can't download that. You can't get -- that's the question.

We're not talking about Acrobat Writer; we're talking about a reader, and we're talking about an old version that is no longer distributed.

That's the kind of issue that we're worried about. Now, as we try to start kind of parsing what might be the things that we would want soon, versus long-run, in terms of websites to preserve, I think one of the things to keep in mind is that particularly for the long run, there are certain things that one would want to be able to overcome: Signals from the website that say don't crawl this just for historical purposes.

The kind of thing that I'm thinking of, mixed-content websites may contain particular essays, writings, or authored pieces, and those are the things that I really trust the publisher will take care of for a long time.

But one of the things that we're interested in, is the context around that; how did someone find that, what are the parent pages; what is the stuff around it? What is the context?

An archivist is concerned about context, primarily concerned about context, whereas the publisher might be trusted to keep the actual piece, written piece, alive for quite a bit of time, the picture of how people found that and how it was contextualized in this time period, is one that I don't see any publisher showing any concern about trying to preserve, and rightly so.

That's what libraries should be involved in.

MS. GASAWAY: Scott?

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MR. TEISLER: To be responsive to Keith's specification, there's an expression called the invisible web. When web crawlers go around the web, they only find a fraction of the content out there.

There are lots of sites that are published dynamically, so there's never static pages. In fact, if you want to use a site like that, you need the publishing software.

A prominent content member of the Internet Archive is cnn.com. Of course, there's about five years of cnn.com in there. It gets a lot of use.

If you take a look at what you can sort of recover from the 2001 issue of cnn.com, what you'll be able to recover from 2006 cnn.com, there's a big difference. That site is not assembled for you until you visit it.

It's not, by the way, just the big people like Time Warner that are going to do this. Most websites aren't draft creations of HTML programs.

They are put up by services who facilitate their creation of websites for people. It used to be GeoCities, now it's myspace.com, where the same kind of dynamic assembly takes place.

If archiving is going to work, opt-in/opt-out or whatever, most of it may be measured by the massive content. The content on the web is going to have to contain instructions on how to compartmentalize this stuff. Otherwise, it's the QuickTime Four problem times a million.

MS. GASAWAY: Bill may have later data than I do, but the last study I saw about what percent of web pages were commercial, was that only 30 percent of all web pages were commercial. I want to know if there is a later study, so a whole lot of them are noncommercial to begin with.

MR. TEISLER: Same problem with counting. You can't see it, you can't count it, so these commercial pages aren't being counted.

MS. GASAWAY: They may not be there. Some of the noncommercial may not be there, either.

MR. ARMS: I just want to agree completely with Scott on this point. One of the interesting things about the Internet Archive is what a small proportion of the web you can nowadays get, just by access-crawling, and it's getting less all the time.

What is interesting, though, is that many, many sites -- and you may be one of them -- put your material up in special places, so that the search engines can have access to them, and those are potential places for archivists to go.

MS. GASAWAY: We've got three people in the queue, to stop on opt-in/opt-out.

MR. KUPFERSCHMID: This is the first I'm hearing the concern. I would be more than happy if somebody represents 800-plus members. This is the first time we've heard this concern, and I'm sure we can get our

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

software companies to the table and find a solution to this, to make sure you guys have the field versions and software to access this.

MS. GASAWAY: Ken Frazier, back to Keith, then Alan, and we're done.

MR. FRAZIER: The library community is especially concerned about opt-out with respect to government information, where we are seeing a growing tendency on the part of government, to issue public information, intending it to be public, and then deciding to withdraw that information from the public.

It's not just the Federal Government; it can also be state governments. Here, we have a compelling public interest in the preservation of that content over time.

Very often today, that content is available only in a website form. There is no analog, there is no other option.

The research library community is not only in a position of having to contemplate the preservation of content, in defiance of the copyright owner, but actually preservation of the content, in defiance of the law, because the standard is so high to preserve that content, that it's a measurable, visible problem.

MR. RUDICK: Alan, one minute, then we've got to wrap up.

MR. ADLER: There's always a compelling interest which tends to drive things forward. Usually, to those who share the interest, it's to the exclusion of other competing interests. I would suggest here that while creating an historical record is certainly an important interest, remember that there's an entire generation that was very willing to trade e-mail for personal correspondence, and with that trade-in for grammar and syntax, and viable information, for that as well.

The historical record, people say, has suffered, because of that. Nevertheless, it's a product of society adapting new technology and moving onward. My point here is just that simply there are no immutable compelling interests, I think, even in terms of establishing an historical record here, that necessarily trumps some of the other interests, competing interests that have been served.

I'd point out that if we really are concerned about being comprehensive in gathering all these websites, you're going to have a tremendous compilation of pornography, not to mention infantile material that's put up by children who have learned better than most adults today, to create their own websites, personal websites that they use.

So there's going to have to be some, I think, recognition of the fact that there needs to be targeting of the activity in order to be able to justify it in terms of the compelling interest that is driving it. Again, we would remind people that it is an exemption from the general rules.

MR. RUDICK: The last question, which we'll barely squeeze in, is, let's suppose -- it's a hypothetical, okay -- it doesn't mean there's a decision to do this, but if there is a special exemption for websites,

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

if there's a special provision for preserving websites, whatever provisions there might be for opt-outs and opt-ins, what restrictions should there be on public access to content?

Some of you, Bill, for example, have already, I think, commented on this. Are there any other comments on this subject?

MR. TEISLER: On the question you just asked? Opt-in solves a lot of this. For opt-in to be useful, for all the reasons I have been saying, you have information about the material at your site. ; There have been ways invented to do this -- generic categories or rights' expression language.

In this way, you'd at least have explicit guidance about how to solve the subsequent access problem. The thing we also want to keep in mind, is that what a library or an archive is, especially a virtual library or archive, is five or ten years, 15 or 20 years or 50 years from now.

It's interesting to speculate about, but I think our general belief is that it will tend to de-emphasize physical premises. Most access will be remote access.

We'll have other problems that we have today with rights management and remote access. For me to imagine the statute anticipating all the variations on the theme to get this done, is a little hard.

The lesson, over and over again about how the net works, is that you devise scaleable means of ways of expressing instructions and preferences, and exceptions across that solve business problems and mitigates intellectual property problems, and it's a different approach than trying to solve it in the statute.

MR. RUDICK: It's a thoughtful comment, but if there were a statutory provision, let me ask the question very specifically: Should online access -- it's been preserved, it's somewhere -- should online access, ever to be permitted in this context?

MS. GASAWAY: It's a good -- it's a weird thing. We're discussing a website and you can't get online access.

MR. RUDICK: And should a certain period of time have to lapse before making the content available? If you've spoken previously on this, it's been recorded and you don't have to say it again. Would anybody else like to comment on either of those two specific questions? Yes?

MS. PILCH: I think that as long as opt-out provisions are honored, any sites that people don't want distributed or copied, will not be. Yes, I think the sites should be -- online access should be allowed, and, again, I can't see any reason why you wouldn't have that, and I don't see any reason to restrict it to premises, either, if you know you're dealing with material that no one has objected to being displayed.

MR. RUDICK: Howard, Keith, Roy, Ed.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MR. BESSER: I'd like to remind people of something I said in the previous session, that, in fact, all material eventually is supposed to go into the public domain, and that, certainly at that point, everything should be available, including things that people opted out on.

MR. RUDICK: One-minute rule, guys, and you can always submit written comments. Keith, Roy, Ed, and that's it.

MR. KUPFERSCHMID: In the background paper, there's a statement, a possible restriction being that you have to wait till the information is quote/unquote, stale, because it has no value at that point.

I found that comment very interesting, because I'm somebody who is front and center on this debate on database protection legislation, and we were trying to push that issue for a long time and trying to get some kind of federal law to protect from misappropriation and talk about that information actually does have information, and at some point, that value is going to be less.

But, at least during that period of time, it ought to be protected. I was getting stonewalled, we were getting nowhere with the library community, so I think that certainly if we're going to discuss this issue, I'll be more than happy to discuss these issues about when the information actually does have value and making sure it's protected during that period of time, if that's talked about at the same time.

MR. ADLER: We're going to attach the database legislation.

(Laughter.)

MR. RUDICK: Roy?

MR. KAUFMAN: This is all balance. Can I imagine a point at which the archival copy of something online is available online? Yes, of course.

What's the time period? What are the ; conditions? Is this opt-in or opt-out? If it's opt- in, can you set your own time conditions by a piece of code? If it's opt-out, can you set limitations on a code?

There are so many things here that, yes, I can envision it. I think it's a good idea. Ultimately, it will happen, certainly when it's in the public domain, and hopefully even before then, but are we opting in? We can agree to a much shorter time period, if we're opting in.

Are we opting out? Well, it's different. Yes, it's pushing it.

MR. RUDICK: Ed, you clean us up.

MR. LAMOUREUX: As you know, the Vanderbilt Archive follows procedures for allowing people to use televised materials. One of the procedures is, once we send the material to you, you can't show it on television. That's the very purpose.

Any archive and archived online materials, would have to, by its very nature, I think, definitionally, make those materials available through responsible means.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MS. GASAWAY: I have one sort of short follow-up question to Bill and Scott. Something you said this morning, Bill, and some of the blogging software, isn't there something that does what Scott was talking about, that sort of sets those parameters that allows the blog to be archived?

MR. TEISSLER: The present set of generally-recognized instructions are aimed at crawling and indexing and abstracting, rather than the archiving. The notion here is that a more general set of instructions capable of conveying nuance --

MS. GASAWAY: But there is something, because someone had mentioned that about blogging, and I just wanted to clear that up.

MR. ARMS: Technically, it's straightforward.

MR. RUDICK: Mary, one last one.

MS. RASENBERGER: I don't want to get into a conversation about this now, but it's something I want you to think about if you're submitting written comments: If we were to have such an exception, whether it should be limited to certain institutions or whether some of your problems with this kind of exception would be improved with limiting it to certain institutions -- I know, Bill, you spoke about the Library of Congress. For instance, could something like this be limited to the Library of Congress or other types of institutions?

MS. GASAWAY: All right, we want to thank you all for your participation today. You've given us much to think about. Your comments will be considered and reacted to and reflected upon and used by the Study Group to develop our recommendations.

MR. RUDICK: Written comments: I know we mentioned this a couple of times, but we'd love to have them. I know how frustrating it's been sometimes to have to give an incomplete answer and not respond to what a colleague has said.

Here's your chance. We promise that someone will read all of them, and all of us may read some of them.

MR. LAMOUREUX: Just a quick question: Can we make an application with the written comments, or not? In my application, I made arguments. Is that in, or do I need to repeat it?

MS. RASENBERGER: You do not need to repeat it. Once you submit it, it's been read and it's on the record. If you have additional comments, though, please submit them; we'll be accepting them. The procedures are in the Federal Register notice, and we'll receive them through April 17th.

Transcription
Section 108 Study Group, Public Roundtable #2
March 16, 2006, Rayburn House Office Building, Washington, D.C.

Topic 4: New Website Preservation Exception

MS. GASAWAY: Some people really did just apply by saying I'd like to apply and here are the two topics I want to address, and they didn't write a statement.

MS. RASENBERGER: And some people did have arguments. Those are on the record; we have those.

MR. LAMOUREUX: Thank you.

MS. RASENBERGER: I just wanted to note that the transcripts of this roundtable discussion, and also the one in Los Angeles, will be up on our website within probably a couple of weeks, give or take a little bit.

I wanted to thank Dick and Lolly very much for all their hard work, and also Joe Keeley who was here for a little while, but I guess he's gone again. Thank you again for arranging for us to use the room, and I thank all of you very much for your participation.

(Whereupon, at 4:30 p.m., the roundtable discussions were concluded.)