



COPYRIGHT 2.0 – RIGHTS IMPLICATIONS AND USER-GENERATED CONTENT

COLLERAN: Thanks. Hope you enjoyed that. That was meant to be a shameless promotion (inaudible). That part was supposed (inaudible). I think that really sets the tone for today's session, just how much material – we all know there's a lot of material out there, and it's being used in a lot of different ways. But what about copyright – because it's becoming a bigger and bigger issue every day. And Web 2.0 and the prevalence of user-generated content has really – blown really sky high. And we talked about the original creative downstream creator. How to manage the content, how to manage the (inaudible). (inaudible) said it best when she said the consumer and the creator are blurring. But what also is blurring is the concept (inaudible) rules around the IP – rules around IP and what you need to think about when we're providing tools and technology to our users that allow all our content to be put out there in a lot of different ways.

(inaudible) very interactive discussion. I hope that it's a pretty passionate discussion. We were just in a conference room talking about this, and we'd probably still be in there (inaudible) yank these guys out (inaudible) it's such heavy stuff. So we're going to hear from three experts in the field.

The first, as you can see will – the first actually will be Greg Merkle, who's the vice president and creative director of Dow Jones. Greg is in a unique position where he works for a traditional publisher of content, but they are also delivering this content in very innovative ways. And he's a thought-leader in this area, so he's – you'll see he's kind of in the middle there a little bit on the subject.

Dan Gisolfi who's the IT software architect in the Emerging Internet Technologies Group at IBM – and by the way, if you're going to have a panel, (inaudible) have somebody from IBM on your panel. He's going to talk about, and your going to see some great and interesting stuff, because they have some technology that has enabled (inaudible) mashups. And I think a couple people in this room especially will be interested in what they see up there.

And last but not least, the voice of reason, the lawyer. We have Robert Lands, who is the partner of Finers Stephens & Innocent out of London, and he's going to come

and talk to us a little bit about what's going on in the international stage in terms of copyright, and how the law is putting major constraints on this, and maybe the law needs to change, maybe it needs to be reinterpreted. Let's see. So enough of me. Without further ado, Greg.

MERKLE: All right, good afternoon everyone. It's a real pleasure to be here. Again, my name is Greg Merkle. I have a really unique role in the Enterprise Media Group, where I have the ability to envision what the end user experience is going to be around information. The creative director is not necessarily a marketing job function. I do a lot of engineering (inaudible) how users will interact with information (inaudible) any of your software.

I thought I'd make sure – we're just talking about the ecosystem of the user, and I wanted to just use this as an example. In the corporate enterprises we serve, and actually within the entire 2.0 (inaudible) sharing environment, you have essentially three classifications – someone who creates content, you have people who participate, and then you have a reader. And (inaudible) – there's actually another one that's called the couch potato, who sit on the extreme periphery of the overall information ecosystem.

What's interesting about this overall ecosystem, or this user base, is that the creators are publishers, traditionally. They were journalists and editors. And now creators are bloggers, could be (inaudible) providing output of the wiki. The participants could have been individuals historically who wrote in, part of the op-ed or the editorial component of the newspaper. And now that participatory margin is so rich (inaudible). And then readers who just consume information.

So just want to set the stage by creating and envisioning a (inaudible) view, where I (inaudible) start to pull information together from multiple sources. In the world that I'm now investigating in corporate enterprise, where they are trying to build very (inaudible) information (inaudible) by creating very high performance workflow applications, including information in the context of the user's workflow. We're starting to see the emergence of consumerization, the trends of the personalized home, moving into the enterprise. So I want to just envision building that (inaudible). So I have an RSS feed. Before I get into this I'm going to poll the audience. Who uses an RSS feed (inaudible) as the primary means of reading the news? (inaudible). So three of you – four of you. (inaudible) this morning. I'm a (inaudible), so I use a reader. I use traditionally as well (inaudible) information via RSS, I think I have about 250 feeds that I (inaudible) read. And it's impossible to get through that using DAT (inaudible) technology.

So back to this world. So an RSS feed from Factiva, which is a user-defined (inaudible), so I want (inaudible) content that's around a specific subject. I can get RSS feed directly from a publisher, and I can also activate (inaudible). If there's a wiki internally, I can get the RSS feed of the wiki, the conversation (inaudible). If there are internal (inaudible), I can get the RSS from (inaudible). And then there's

something called Yahoo types – (inaudible) internal blog which I may not have any control over. RSS from (inaudible). So I just wanted to show that there's eight varieties of RSS on one single screen.

And just to go a little bit deeper, to talk about Dapper. (inaudible) some of these applications. Dan is going to cover this in probably much greater detail. But this morning as we were talking, Dan created, and I actually captured – Dapper actually enables you to go get data from a website that doesn't provide syndication, RSS, or XML, (inaudible) library, and you can create a theme, and you can save it as XML or RSS. Swivel (sp?) is actually a data repository of statistical information (inaudible). And there's census data there, government (inaudible) data, private information – really interesting – but this is just a screen grab of estimated price by (inaudible) who are Web 2.0 companies. And there's a list of about 100, I just moved it to 10. And the stats are available in graphical form, which I can share with (inaudible), etc. It's really, really quite powerful.

Yahoo Pipes – this was actually an interesting feed from CNet. It's a green technology (inaudible), so it looks like a simple list of headlines. It's actually aggregated by using a visual development tool called Pipes. And Dan will probably go into much greater detail with this. But it can essentially go to URLs and builds some rules – (inaudible) blogs like CNet.com and looking for keyword energy, blogs at CNet.com looking for a keyword green. URLs creating some filters. Outputting them, combining them, de-duping them, sorting them, filtering them by date, and then outputting them into RSS. This tool exists today. So just to show you some of the rich development (inaudible) that are emerge – and there are two other development (inaudible). So going back to my little (inaudible), imagine all these mechanisms powering this environment, and after really (inaudible) having development (inaudible).

Another area that we'll talk about in our conversation are emerging standards. (inaudible) access protocol and XRML and (inaudible), so if anyone in the audience is interested in that or anyone has any other information to share, that'd be great to talk about that as well.

I'm going to hand it over to Dan.

GISOLFI: OK, so Greg gave you a very nice introduction about – from a top-down view of how to create a mashup or a (inaudible) as you've described. (inaudible) don't have to be just RSS feeds, they can be very interactive widgets and gadgets, and so there's a whole space, a whole ecosystem out there led by this group of gadgets. Microsoft's got (inaudible) and so on.

What's happening across the Web is that you have the Web becoming the palette of software building blocks. And so at IBM, and what I've been doing for the past couple years, is a mashup enabler. It's going off and being able to take content – work with content providers like Greg's Dow Jones – and be able to enable their

clients to be able to (inaudible) building blocks, and be able to (inaudible) creation of new situational applications or mashups. And when you think about that, those situational applications are going to be very much a combination of – fabricated from a variety of different sources. So as a content provider, as a publisher in the audience, we don't know the long tail of Web 2.0 where your content's going to be. If you're afraid of that, that's part of the discussion we're going to be having here today. If you're going to embrace the long tail, well then how – at what point do you want your content to be out there and available, accessible, and then what limits do you want to put on it?

So what Greg, in his introduction, actually gave you is an introduction to some of the tools that are out there that, like Dapper, like Swivel. And what I want – I'm just putting Dapper (inaudible) – there's no relationship with Dapper. It's actually a company in Israel. But what Dapper and Swivel and some of these companies out there on the Web today, introduce to you today – every one of you in the audience, as non-programmers, can walk up to these tools, create yourself new content from existing content, maybe even your own company site, and then publish that new content as a new conduit, a new widget, to share or use in situational applications, again so that a knowledge worker or situational application developer, who I like to call a mashup assembler, can go and create.

So in IBM, one of the tools we've been creating for this space is something called QEDwiki. And QEDwiki is a mashup maker. Now, I'm not going to spend time – I want to focus on copyrights – on the issue of copyright here today. So I'm not just going to focus on the tool (inaudible) anybody who'd want to see it. But what I do want to show is to point out (inaudible) example, some of the problems. So after arriving here last night and getting the pamphlet of the folks who were attending, I just went and looked at a couple of the sites from attendees. And I looked at British Library and the Berkshire Publishing sites. Now, these are two sites didn't have RSS feed sites (inaudible) pick up quickly, but I was able to create my own RSS feeds, and I used Dapper to do that. And what we're going to talking about, or my role here today, is to introduce you to one of the issues or problems, some of the things you should think about as publishers.

So the first thing that you should think about in this mashup – or what I like to call it mash board – this is very personal. This is Dan's research paper. I created it. At no point have I told you I'm sharing it. I could. I could share it with my colleagues. But you don't know where Dan's scope is with this page that I've created. The second thing is you can say that Dan put his name on it. The other thing that you can see here is that I told you that I went to British Library, but I called it Web 2.0 Library. I also went to Berkshire Publishing, but I called it My American News search engine. And I'm just going to interact with these (inaudible) Dapper (inaudible) go off and do an RSS query on that. And while that's going on (inaudible) – (inaudible).

M: (inaudible).

GISOLFI: (inaudible). And these are running two dynamic searches against the British Library and Berkshire Publishing. I'm also going to do a search here on China, and that's going to go off and hit the Financial Times. A couple things are happening here. And issues you should think about with mashup. Those publishers in the room, does Berkshire and British Library, are they comfortable with having their content on the same page as Financial Times? Is the British Library comfortable with the fact that Dan put the British Library as title there? He put Web 2.0 Library. Same thing for Berkshire. Conversely, isn't it cool that their content is being (inaudible) at this low level from (inaudible) out there, working with this content. Now, if I wanted to drill into this content, you can see that I really can't get far here in the British Library. I'm probably just going to get a summary of information that was given to me off the news feed, right? But if I wanted to (inaudible) for information get that. If I went to Berkshire, same thing, I'm just going to get – going to lead into the information.

Couple of things to point out about these types of widgets. These two widgets that I (inaudible) from Dapper, we've classified these as standalone widgets. As you noticed, I had two things wrapped independently with each one of those. They're just scripts, Web scripts. Whereas when I click in this section here, in this form, I put (inaudible) for example, and through our mashup maker, we were able to wire that interaction to an RSS feed search against Financial Times. So as you think about ways your content can be published, the use of mashup makers, you want to be able to make sure that any user is able to interact with your content as well as other content on the same page. And that's where mashup makers and some of the technology (inaudible) Web 2.0 become interesting.

So with that, I hope I've paid some of the issues and interesting comments or discussion for us to (inaudible).

LANDS: (inaudible). So I expected maybe some jaws to drop – I'm looking at Karen – and when I had the privilege of spending a day at the IBM research labs where I first saw this from Dan, I guess my face turned white. Because this goes beyond just links. This goes into cutting and pasting and sending those things out (inaudible) your content. So that's really what we want to talk about today. And I think that's a great lead in, so what about copyright law? Where does that come in and did something just get violated there? I think so. There was no –

F: (inaudible).

LANDS: Just to expand on that question which was what about moral rights? In case people don't really know where that question was coming from, in Europe we have a system of moral rights that goes alongside copyright. So the creator of the copyright work has certain rights that can't be transferred, generally. Rights like the right to be identified as the author, the right to object to derogatory treatment of their work.

F: (inaudible).

LANDS: Right. Yeah, and there's been a lot of talk about attribution being the key issue. But it's more than attribution because I, as an author of one of those things in Dan's presentation, I may want my name to be credited as the author. But I may not if you've combined my stuff with other people's stuff and it may look bad for me. So moral rights is quite a complex issue, in Europe anyway. In the States I believe it's not really part of copyright law.

And I haven't got any slides, and if I did they'd probably be very boring compared to these guys. But I thought maybe if we – this is very sophisticated stuff. If we take a step back and take a look at what's going on now in the courts, because that's the future, we might be able to see what the problems might be, and possible solutions. You probably realize that it's every IP lawyer's favorite sport to sue either Google or YouTube. Lots of that going on.

Google has been sued in Belgium, and also by a French news agency in America over its Google News service. And that does – although at an embryonic state compared to this – a similar thing, where it will take headlines and a sentence or two from the start of a news report and put it together on a page. In Belgium, the Belgian court said that that was a copyright infringement and made Google block the – stop linking to those Belgium news services that it was taking content from. And Google actually got a little revenge in there by blocking the sites completely from Google, so you couldn't even find the newspaper in the web search. But they could've interpreted the ruling a little too literally, deliberately. But that case shows that copyright is not the same the world over. In the UK I don't think Google's news service would be an infringement actually, because taking a headline is not a copyright infringement, because a headline is too short to be able to reword. And taking the first line of the story probably isn't taking a substantial part of that story. So it's not an infringement. But the law may be different in Belgium. And that's one of the issues with copyright using these new Web services, because copyright is not the same in all the world.

YouTube is getting sued because a) it has money, and so people think it's worth suing them. Nobody bothered suing them until (inaudible). And two, because it's publishing copyright material, quite often they've been submitted by users. And in doing so it's doing what these services would like to do – it's making a reproduction of that work, and that's an unlawful activity. It's also under English law infringing the right of communication to the public. This is the right to make content available to people to download as and when they want to see it. It's the law's reaction to technology. Because in Europe we had a rationalization of copyright a few years ago where they said, look, previous law just said you've got a right to broadcast, you've got a right to make a cable transmission. But now technology's moving so quickly that we can't fit things into those small boxes anymore. So they changed the terminology and they said, instead of that, we'll

make it a communication to the public right. So, by any means. And that's what it would be alleged that YouTube is doing. In America there's a similar things – look at my notes, I'm not a U.S. lawyer, so if I say anything that's complete rubbish under American law or any other law then tell me, because I'm an English lawyer. And in America they have a thing called contributory infringement, where YouTube didn't create the content, it didn't record a television show and publish it – one of its users did – but it could be liable contributory infringement because it's then publishing that on. There's a defense – if they didn't know about it. But the argument is – that's currently being tested – they should've known about it because that's what everybody does with YouTube.

So those issues are playing out right now. And I'd be happy to look at those and more in relation to these new Web 2.0 services which are going far beyond that.

COLLERAN: And what did you think about when you were thinking of copyright law? When you saw – when we had our discussion earlier on today, and what you just saw, what did you think?

LANDS: Blind panic, really. (laughter) There's stuff that is clearly fine. Got to say that. There's stuff there that is on a traditional basis not being a problem at all. And some of it is really just linking. And linking has never really been tested in the courts. There was a case in 1997 that involved one newspaper (inaudible) another newspaper's content, and they kind of tried to say, well, it could be a trespass. They tried to look at old law to fit it to this new situation. But it was never really resolved, and it still isn't, but everybody links, it's been going for years. So that kind of stuff is fine. Also, taking an RSS feed is generally going to be OK, if you're just taking it and not repurposing it and modifying it, because the site providing the RSS feed is giving you an implied license to do so. It becomes much more complicated when you're not just taking the content and publishing it, but you're taking the content, mashing it up with other stuff, and then publishing it. Then it really could be a problem.

COLLERAN: Greg and Dan, you took two RSS feeds and put them together and renamed them. How do you feel about yourself now?

GISOLFI: I'm not doing anything wrong. One of the things we talked about earlier, and I want to point out to the crowd, if you think this is just something that you just saw and it's really not happening a lot, I think you're drastically wrong. This is very pervasive on the Web today. And when we were discussing this in the conference earlier I argued that this is no different from a blogger going off and relinking back to other blogs or other news articles or other websites. The difference here is that, in the case of like Yahoo Pipes, I can actually take content from two different RSS feeds and derive an entirely new RSS feed. Now, as an author of that RSS feed, I have two issues, or two comments. One is that I feel that I can repurpose this content. I'm not making money off it, that's a different story. But I also feel that I want attribution for the RSS feed, just like that guys who I'm borrowing from want

attribution. So this gets into – and, Greg, you mentioned this earlier – about standards. We really – RSS and (inaudible) syndication technologies out there today does not deal with copyright and attribution. And I think from the panelists here, I think we're in favor of that.

MERKLE: Yeah, I think one of the wonderful things about standards is that they standardize. That means that opens up exchange, the interchange of ideas, the interchange of information. RSS was considered to be a very disruptive technology early on, and a syndication technology. So its primary goal is to get content into other environments and drive usage back to websites. It's evolved into something much, much richer. And in fact the RSS specification is evolving to facilitate that, out of the adaptation of the original format. And I mentioned standards earlier, that the standards are emerging and are just starting to take into consideration these downstream use cases of how information is actually being consumed, and where it's being consumed. And as Dan mentioned, it's pervasive, and you're not going to stop it. And when we talked earlier about there are two paths you can take – you can either build thicker and higher law around your content, or you can actually take a very close look at the social phenomena of being able to repurpose and remix information, and understand how rights can be infused into that environment.

I work for Dow Jones. We're publishers, so we publish the *Wall Street Journal*, we have the Wall Street Journal Online, and many other online consumer websites. Out of those websites we do produce RSS feeds. They're available to registered subscribers. The experience is very contained. The user clicks on a link, the link could be available anywhere, and the user is then driven back to the site, drives advertising impressions – essentially it's a linkage model. The problem gets a little more interesting when the RSS feeds are synthesized out of multiple data sets, as (inaudible) Dan put together.

We serve corporate enterprises as a different line of business for the B2B market, where we serve corporate enterprise with (inaudible) like Factiva. And we then have – we ask that the enterprise sign a global license for all their users. And essentially we are then managing the royalties, the rights to distribution of that information on behalf of the IPs through our product. It's a very expensive proposition to architect that kind of authentication mechanism for the massive data sets we have.

And we're now meeting with clients who are using – who use our content, and they're taking some of the practices that are emerging on the outside, the Web 2.0 practices of mashing up, and creating something called an enterprise mashup, where they're actually creating APIs and almost Dapper-like and Pipe-like interfaces into their own data. Imagine Yahoo Pipes into their CRM, because it's so hard to get information out of it, and the remixing it. That's one that's an incredibly gray area, with regard to the reuse of that information.

COLLERAN: I don't what the publishers in the room think, but I think Greg is in a unique position because he works for a semi-traditional publisher, but obviously a forward-looking publisher, but also is in the new technology area. But I don't know if that is a gray area (inaudible). Because you just talked about a lot of content being put out into their own RSS feeds, and I just assume they're distributing them, and they're not just sitting and reading them themselves. So I don't know – Kate, do you want to comment on that?

F: (inaudible).

LANDS: There's protection for the author of the blogs, and there should be protection for you as a publisher, because you are going to liable if there is something that's defamatory in those blogs, or if – an example that I gave you earlier – if there's medical information, for example, and the dosage is incorrect. So there's that issue as well as the copyright issue. But from the copyright standpoint, the problem I think I have as a lawyer, which is the problem that a lot of people have with Google Books and that kind of thing, is that what's going on now is your taking the content and saying I can take your content and repurpose it, unless you tell me not to, which goes against what the law actually says. Which is that you've got a right to prevent people doing anything unless they ask your permission – apart from certain (inaudible). So that's the problem with it.

GISOLFI: So one thought there is – in fact it was in the keynote this morning, about (inaudible) the long tail. I think all publishers in the audience have to come up with some sort of threshold for their business. At what point do they embrace Web 2.0 and this long tail? And when they do so, what is this tactical or long-term strategy for addressing what content can be made available and to what extent? And by whom, and for what purposes? Is the British Library going to go after Dan for putting (inaudible) up there? It may be a little costly to hire a lawyer to do that.

LANDS: I'll do it for free.

(laughter)

GISOLFI: You have to really think about what is worth, when is it justified, going after (inaudible)? And then, what we we're talking about earlier is, in a sense, Dapper and our tool and a lot of these mashups are just conduits for the information. The end user is the one who's using the information. And at that point, what does that end user do with it? And maybe that's where copyright comes into play. So I just like the fact that there's a lot of questions being asked here. There's a lot of activity going on on the Web. This is being done all over, bloggers do it, mashup assemblers do it. And if content publishers want their information to get out there to places where they never thought they would be accessible, and they want to embrace the long tail, they're going to have to bring down some of that wall, as Greg pointed out earlier.

F: (inaudible).

COLLERAN: Yeah, I think that's why we talked about that earlier, that awareness is a big (inaudible). (inaudible) around awareness, and around the awareness of the users that are using these tool, and what they can and can't do. And at the same point, like these guys were talking about, do you lower the law a little bit. Because copyright law, at least in the United States, has not changed. The law is the law. So if, Kate, you're pasting an article in a blog, and I don't think you are, but if you are passing and sending it out, that's – current copyright law tells you that you can't do that. And I think that what – struggling (inaudible) make aware to you all, is these guys are building – Dan and his team and people are building tools that make this so easy. So it's such a conundrum. And I think Dan even has a little bit of a different opinion on it is violating some of copyright or not.

GISOLFI: As I mentioned before, and we discussed, I really feel that RSS and (inaudible) those specifications need to be updated to have metadata about attribution in there. So I'm all for that. And I think the tool inventors out there have no problem (inaudible) information, (inaudible). Right now it's set in stone. And we also – we need to make sure it's simple. Because, as a blogger, I'm sure you go through this every time you go to link to another site, it stops (inaudible), right? And so when I'm going on as a mashup (inaudible) bring a page like Greg laid out for you, I'm trying to think about all the data intersections (inaudible) –

F: (inaudible).

GISOLFI: (inaudible) attribution, but you also want – I have a job (inaudible), right? And from an IBM perspective, our focus here is very – one side of the application spectrum. We're focusing on situational applications with situational needs of the end users or knowledge workers. And when you're in the situation, you don't want to bogged down by certain things. And so I also think – I also think the point of copyright also comes into at what point does it get published? If I created that application on my desktop, do I need to deal with attribution? How many of you have Excel documents on your desktop that you cut and paste stuff into? Would you give attribution? Would you need to? It's sitting on your desktop. It's sitting on your desktop, you never e-mail it to anywhere. So my point is the point of publishing the mashup or situational app, or sharing it, I think that attribution becomes important. So I know I'm saying a lot of things that people aren't happy with –

F: (inaudible).

MERKLE: No, I think there are – the emerging standards for RSS today is not a very rich container for the information. There's another version called ATOM – (inaudible) familiar with it. Essentially it has more descriptive name value pairs in the XML. (inaudible) actually takes into consideration the author, the rights to the – and potentially even the sub rights to the content, where the content may be – have a

bibliographic reference, or there may be media embedded into that article, which has a different set of rights than the actual container. So it's a very thick problem when it comes to rights, and there are gross rights and net rights, just like a film would have. Standards are beginning – just beginning – to be taken in consideration, the downstream consumption. So I think the question is how can the technology do that for the user that's actually – I don't think it can effectively. I think that it's incumbent on the publisher to actually create, with a standard –

END OF FILE