The John Marshall Journal of Information Technology & Privacy Law

Volume 31 | Issue 4 Article 9

Spring 2015

Legal Problems in Data Management: IT & Privacy at the Forefront: "Big Data": Ownership, Copyright, and Protection, 31 J. Marshall J. Info. Tech. & Privacy L. 565 (2015)

SESSION FOUR:

"BIG DATA": OWNERSHIP, COPYRIGHT, AND PROTECTION

MODERATOR:

DAVID SORKIN
PROFESSOR, THE JOHN MARSHALL LAW SCHOOL

PANELISTS:

PETER DICOLA
ASSOCIATE PROFESSOR, NORTHWESTERN UNIVERSITY SCHOOL OF LAW

MARCELO HALPERN
PARTNER, PERKINS COIE LLP

PROFESSOR SORKIN: Our next panel on the topic of "Big Data": Ownership, Copyright, and Protection, is Peter DiCola from Northwestern University and Marcelo Halpern of Perkins Coie.

(Applause.)

MR. MARCELO HALPERN: Good afternoon. I know keeping on schedule on a Friday afternoon is important, so we'll talk fast. I'm Marcelo Halpern from Perkins Coie, and I practice in the area of technology transactions law.

What we're talking about here is a little different from what we talked about in the earlier sessions today. We're not talking about data privacy. We're not talking about security. We're not talking about crime.

We're talking about money. What is it about this data that everybody is accumulating and using, why are they doing it, and how do they protect it? Because after all, big data is a big business.

So we are going to go a little through, what it is that we are talking about. Everybody talks about big data. We have had a lot of people mention it today. But what exactly is it, how does it work, and what are the business models around it? Then we are going to talk about the legal frameworks for protecting it.

So as one of the FBI speakers mentioned, they don't do the civil side of this. If there's a hack, if someone steals your data, or what have you, the FBI will try and put them in jail. But what are the damages? How do you protect it from other people that want to do something with it financially? What are your legal protections on it? We'll talk a little bit about how you can do that.

So what is big data? We all talk about it. We all mention it. What are we really talking about here? It's not just this big brother thing. It's not just about privacy and security. It's not just about targeted advertising. It's about ways of making money with data and the growth and the ability to store data.

As our last speaker was saying, the volume of data that's being collected is enormous, and what can you do with it?

So here is how it comes up. Big data includes personal data. That's all the PII (Personally Identifiable Information) stuff and all of the data security things that everybody's worried about.

Trade secrets: The Motorola example that came up, the woman walked away with a bunch of Motorola information.

Sensor data: Iterative things. All the stuff that is being tracked automatically, your FitBit watch bands, your location on your phone, the sensors in your vehicles, are all collecting data, all going into these massive data sets.

Behavioral information: What ad did you click on? When do you

shop? How do you work? How do you get to work? How often are you riding the el? All of that behavioral information is part of big data.

Social media: I have a client that was involved in doing essentially predictive analytics on the basis of people's Internet postings on social media, trying to figure out what that meant in terms of what their voting behavior, based on postings that they would go -- and they were incredibly accurate in painting a picture of people based purely on their voluntary social media postings.

Process capture: So industrial processes. Measuring temperatures, measuring fluids, measuring pressures, all the things that go into process engineering, all being measured, all being collected as part of big data.

Another view of it or a different categories. We're talking about the personally identifiable information. That's where so much of this focus is on, because that's what people are worried about, stealing data and stealing identities, things like that.

But the rest of it is all unidentified stuff. It doesn't have those kinds of protections. It doesn't have those restrictions around it. So demographic information, de-identified data, aggregated data, trend analysis, business records, technical data, all the sensor data, the market data, things like that, none of that has any of those protections around it the way that personally identifiable data does.

Now, leaving aside the question of how unidentified is de-identified data, which is a whole lecture unto itself, is how far do you have to go before you can actually still get back to the individual who generated it, the reality is the vast amount of data is not covered by those kinds of legal protections.

So this is a track -- I love this website. It's called Ghostery, that tracks from a single website. What are all the connections that that website is going to and what are they doing with the information? So this was just from their website, going to Tmobile.com and all of these little dots represent different kinds of data collection points from a single visit to T Mobile. Some of these are publishers. Some of these are advertisers. Some of these are unknowns. That's a great category, right? We don't know what they're doing with it, but they're connected to this website, and if you do this on a couple of different websites, you start realizing just how many touch points there are for everything you see, and every one of them is collecting some data.

Well, there is a reason for it. Why are they doing all of this? Well, it's a data-driven ecosystem. Sharing information is important in iterative things. It makes things work together. Data is primarily passively collected and generated. So most of this data is not coming from people actively doing something, it's just in the air, it's being collected as it goes along.

This diagram came from the FTC's Internet of Things workshop

and shows a little bit about where some of this data is coming from and how is it getting shared along the way. All of this comes from individuals, comes from devices, comes from work spaces, public spaces and so forth, and not all of it is explicitly grant -- has permission explicitly granted for using that information; it's part of the collection issue of how this all comes about.

So why do we care about all this? We got lots of data. We've already said a huge amount of it doesn't pertain to an individual. It's not individually traceable, so why do we care? Well, it's big business. So what do you do with this stuff?

Obviously, the information that's knowingly provided is important, and that's the part that people are mostly worried about; personal data, behavioral information. It comes from registrations, subscriptions, click-through behavior and things like that, and it's used largely for -- in legitimate uses, leaving aside the illegal uses -- targeted advertising, offers that are potentially of value to the individuals, as well as the offers that you don't really care about, consumer behavior, tracking trends, what is the bestselling model of X these days? What are features people are looking for in the next device or the next article of clothing, what have you, trend information that is derived from individual behavior.

Stuff that's unknowingly provided, the sensor data, the social media data, things that are automatically collected. They are used for testing performance, defined equipment. It's used for traffic flow, used for identifying opinions, positions, postures, what are people doing with it.

Then, of course, process improvement. You want to improve your product. You track how the product is going along the way. All that feeds into the data sets.

So the value proposition. Ultimately, data analytics can influence behavior. A great example of this is the -- obviously, the targeted advertising is the one everybody looks at. But how are they doing it? So there is the example that people looked at recently was Target that developed a model based on behavior tracking that helped to identify when a woman was pregnant, based on what purchases they were making. They didn't know this person was pregnant, but based on the purchasing habits, well, it came to pass that they applied this model. It was quite good, or they thought, and they got a very irate phone call from a father who said, my 16-year-old daughter is just getting advertising from you for baby products. She is not pregnant. Why are you targeting her for baby products? Of course, this runs up the chain at Target. Somebody high up at Target calls the father back to apologize. The father says, she's pregnant. She was buying unscented shampoos. She was buying certain types of foods and vitamins. There were changes in behavior that Target had figured out meant that this was probably someone who was pregnant. It's really effective. So targeted advertising, obviously.

Data has a network effect. The more data you have, the more you can do with it, and the more information you can get out of it. And it's not just more of the same kind of data. It's the compilation of data coming from multiple sets, which gets more complicated.

So look at something like the weather app on your phone. The weather app on your phone gets the information from the National Weather Service. It is government information, but they combine it with information that they have. They combine it with algorithms they have to come up with their own unique weather forecast. That's different from the forecast that some app might have.

The power of combining the multiple data sets gives them that additional ability. The big change in the data world is that in big data, volume trumps precision. What I mean by that, in the old days, if you will, when people had to collect data manually, there was a limitation of what you could do because you couldn't collect as much data. So it was very sample driven. How do we get a nice randomized sample of people? How do we get enough information from that? Then how do we derive projections from that with margins of error and things like that factored in? High statistical issues of how many data points do you need before you can identify something?

With the volume of data that you have available now, you don't need that. You're no longer looking for -- to answer a question, you are looking for an answer and then deciding what the question is, because you have the sample from everybody. Correlation becomes more valuable than causation. When you're looking to make money, you don't necessarily care why someone's buying this. You just want to know that they're buying it. This is a great cartoon on that. You really don't care. It's correlation. It just means that it works, and you can leave it up to the social scientists to discover why.

Example of that. Home Depot decided to dive into the data of taking their purchasing patterns and comparing it to weather patterns, and what they were looking for is, what do people buy when a hurricane is projected? So if a hurricane is predicted in Florida, you know, people are going to go -- and you figure out, okay we can all guess, they're going to buy batteries. They are going to buy shelf-stable foods. They are going to buy candles, generators, plywood to board up the windows, right?

So they looked at all the stuff, and they said, what else can we do with this? What else are people buying? They're buying Pop-Tarts, specifically strawberry Pop-Tarts.

(Laughter.)

They discovered that no one had predicted and no one would necessarily have guessed that in a hurricane people buy inordinate numbers

of Pop-Tarts, and they tested this proposition.

(Laughter.)

So what they did, the next time there was a hurricane predicted, they put the Pop-Tarts next to the batteries, and they sold more Pop-Tarts than you could possibly image.

(Laughter.)

Well, you could now backtrack from that and say, well, why do people buy Pop-Tarts? Oh, it's comfort food. Oh, it doesn't need to be cooked and things like that. But without that just volume -- vast amount of data and correlating it with external data sets, they would never have figured that out. They made a lot of money on Pop-Tarts.

So where is all this going? For science fiction fans, the Foundation Trilogy talks about a world where, based on the information that's available, you can predict people's behaviors with a fair degree of accuracy, and what's interesting about the concept here is, we're not talking about predicting individual people's behaviors. So we're not saying, you are to buy the Pop-Tarts, but we're saying that we know enough about patterns that there is a large enough population of people that will buy Pop-Tarts that it's worth putting that in the stores.

You can make broad prediction on large populations which, after all, is where the money is. I don't care which one of you buys as long some of you buy and enough of you buy to make money.

So the question is, how do we protect ourselves, and how do we protect our clients from these data sets? We all know we now have these incredibly valuable stores of information, and what can we do to protect ourselves? So we're concerned because data is inherently hard to protect. It's being gathered in a lot of ways, leaving aside, again, the criminal element of hackers and things like that. It's all over the place. It's not necessarily well organized.

The legal protections around it, which is what we are going to get to in a minute, is difficult. Copyright protection of databases is thin. Facts are not generally protectable. Copyrightable data inside of a database retains some of these protections, but the database itself under U.S. law, not easy to protect. Trade secret is a great mechanism if you can keep it a secret, and it could apply to internal corporate data but very hard when you're now taking external data sets, which, of course, gives you that network effect and that leveraging effect to be able to create a value from it. So trade secret, some value there, but not a ton.

We're also concerned, because just because you have the data doesn't mean you can use it. It doesn't mean that you can share it and that's where all these data privacy rules start coming in to play. I got

lots and lots of data on my customers. I can't necessarily share it with you, but it's of huge value. It's huge value to you, which means that you're willing to pay me to give you some of that information. How can I get access to that and monetize that? Such information, all this stuff, of course, has its own protection and elements around it.

Then there are contractual restrictions. When I do choose to share my data set with you, I am going to put some contractual restrictions around what you can do with it and handle the data sharing arrangements that way.

So with that, we are going to turn to what are the legal protections around this.

MR. PETER DiCOLA: So, hi, I'm Peter DiCola. I teach copyright law at Northwestern. Copyright, we are going to -- Marcelo talked briefly about trade secrets. That might be the place to start, like with the data you keep secret, you think about is that worth it. Of course, anything that can be reversed engineered, trade secret doesn't do any good for you.

So some of these things, like someone learning what Home Depots knows about Pop-Tarts, well, that can be reverse engineered now because they're putting the batteries next to the Pop-Tarts, so there must be -- there might be a reason, right? So trade secret isn't going to help you protect the value of that.

When we talk about aggregation of data and these collections of things, then we get into some interesting areas of copyright. One of the cases I am going to talk about every IP student will know. Some of the other ones, the IP students might not. Those of you who aren't familiar with copyright at all, you might not be aware of these cases. They are kind of interesting.

Okay. So to start with copyright protection, what is -- copyright protection, ever since January 1, 1978, is automatic in anything that is an original work of authorship that's fixed in a tangible medium of expression. So lots of people still think you have to, for example, mail a letter to yourself or think you have to register to have a copyright. There is all this folk wisdom about it. Hopefully, the lawyers in the room know that's not true, that you have copyright right away. You do have to register to bring a lawsuit and to enjoy certain, you know, procedural benefits and statutory damages, and things like that.

Anyway, we start with this: The idea, the thresholds are actually not about registration, not about any administrative tasks, but are about whether the work is original. The Code doesn't define work. The Code doesn't define original. So work might be something really small, like a piece of data, like one number. The Code doesn't say that it's not, so we're going to see cases where you took one number from me. That's copyright infringement.

So that's the thing that happens in the world. You may not have been aware of that. We are going to see a recent case where that happened. They weren't successful, just to spoil it. In case you were worried, oh, my god, I've used numbers for things. No, it's not that bad, but it's just the logic, the internal crazy logic of copyright law can get you to a place where you're actually filing that complaint and litigating it in federal court.

So, anyway, we don't have a definition of work. It can be something really, really small. Also, originality is a low bar. It doesn't mean original, like, wow, that's original, what an original thought. It just means that it is independently created, that you didn't take it from somewhere else.

So then there's a big limitation on this, which is that copyright only applies to specific expression, not abstract ideas, and drawing that line between abstract ideas and protectable expression has been real -- it has been a really interesting journey in copyright, as we look at software over the last 30 -- 30 plus years now.

Software is really functional, and so in some sense it's abstract. It seems like it could be a process, a system of methods. There are other aspects of software that are patentable, and yet there are other aspects of software the courts have deemed to be copyrightable.

Similarly with data. You might say -- you might look at data and say, well, that's a fact or that's a discovery, that's a thing about the world. That's not protectable. It's an abstract idea. In another sense, it might be someone's expression or someone's opinion about what number best characterizes some feature of the world, right? So this is one limitation, but it's going to be a contested boundary.

Then there's this interesting, really long-standing regulation that the Copyright Office has. It says, okay, these are examples of words not subject to copyright. Sometimes courts wrestle with this language and sometimes they don't. The courts that have really thought about it have given deference to the Copyright Office's interpretation of this, but one way to implement the distinction between abstract ideas that are not protectable under copyright and protectable expression, is there are just some shortcuts here.

So words and short phrases is the one -- I am not going to go through all of these, but words and short phrases are not supposed to be copyrightable. So one example, your title isn't copyrightable, if you think about it that makes sense. There are lots of things that have the same title, if you think about it. There isn't one person who's locked up the title for the life of the author plus 70 years. That isn't what we've done.

So words and short phrases are not copyrightable. So how do we get to -- does that apply to the case where someone has used a database and taken one number or five numbers or ten numbers from it? Is a single

number treated like a word or short phrase, or is it not? That's one of the questions for us to look at.

These are some others I grew up with. In the interest of time, I will just hustle through it.

So the next big concept in copyright is that we have to be ready to think in layers. So even if you take a bunch of facts -- let's say you have, I don't know, an example close to home for all of us in law schools, facts that -- what U.S. News collects about law schools. They collect a lot of facts about law schools, employment rates where you're located, how many students they have, things like that. Then they arrange them, right? They select which facts they care about. They may care about the size of a library; they might not. They may care about whether you have got a copyright class; they may not. I don't think they ran schools based on whether they have copyright. They should, but they don't.

So, then, that selection arrangement of facts, that layer on top of it, that organization, that in and of itself might be creative. It might be protectable as an original work of authorship, and the Code specifically says that that is protectable. So compilation is a category of work, and this gets a little metaphysical, right? But you could think of a book as a compilation of words, right? So it's not that you have your own protection on each individual word in your book, but if have written an entire novel, you have arranged a lot of words in a very particular way, and so, in some sense, what you've got is a compilation copyright. So you could think of almost anything that's created as a compilation.

But for our purposes today, it's a little simpler. We are just going to think about how did you arrange the data in a database? What were the variables you kept track of and for which people did you keep track of those variables? The scope of those compilations, it doesn't mean that you get protection of the individual entries necessarily, but it does mean that you get this protection in the organization.

So when would that be important? That would be important if someone took your entire database. If you collected -- if someone copied U.S. News and World Reports' rankings of all the law schools with all their data in the columns and the way that they are arranged and just republished it as their own, that would be copyright infringement, for sure. Not just because of the images or the formatting of the page, but because the compilation is protectable by the way that they selected and arranged that data.

Okay. So we will move ahead.

So the first case, this is a case -- a couple of the cases we are going to do about phone books. It's hard, because they are getting more and more dated as we go. I have to explain to my students what a phone book is.

(Laughter.)

This is a case -- Feist is a case that's important to remember. This is a case about the white pages, not the yellow pages, and that distinction I do have to explain to my students, right? White pages are about residential addresses. The yellow pages are about businesses and organized by category and might be based on ads and purchases by the companies.

So this first case is about the white pages, and a rival phone book publishing company grabbed -- just copied all of -- so Feist was trying to publish a rival phone book. They took the whole listing of the white pages, and the argument was that this was a copyrightable compilation. The way they would do this, how would you detect copying? Well, you have some fake people in your phone book, right? That means you have some made-up names. There would be telltale signs. So it was clear that the phone book had been copied. That was not the interesting debate.

But this case ended up going up to the Supreme Court to decide what is actually protectable as a compilation. The first step is to say, well, people's names and their addresses, those are facts, so those aren't copyrightable; the individual entries in the database, not copyrightable. There's no argument about that.

What Rural Telephone was arguing, is that the way that those names were arranged was arranged in an original way, that they had expressed -- it was an expressive work to have arranged the names as they did. What was their creative way of expressing all the names in this area? They put them in alphabetical order.

(Laughter.)

You didn't know that. A lot of us didn't know that. 1991, late '80s, alphabetical order was invented.

(Laughter.)

Not any earlier. You would have thought, but before it was weird, you could put "R" in front of "K." It was very strange.

(Laughter.)

So Rural Telephone had sorted that out for us, and they claimed the copyright. No. So, anyways, just as their copyrights and opinions say, this is ridiculous. It's not -- yes, ways of selecting and arranging data could be original, but this one isn't. Even though it's a really low bar, this isn't above that bar. The way you selected the people for the phone book, well, those are the people who live in this town. And the way you arranged them was alphabetical order. None of those things count.

Okay. So we start with a low bar. The white pages don't make it through. So far so good. --This is just the way they articulated the originality standard.

So now we are in the post-Feist world. The Supreme Court doesn't do a lot of copyright cases, and so Feist is their last word on originality and compilation. So we're still sort of -- all the cases we're talking about are working with the language in Feist.

This is a case about maps. This is someone who in the late '60s -- Mason, in the late '60s, created maps where he took the geological -- USGS surveys and then overlaid information about property lines and surveys and things like that, so he created these really detailed maps.

This isn't it. This is just a map of Montgomery County, which is where he was. I can't use his map, because it's copyrighted. It's not on the web.

So Mason ends up winning. He said because he took a map -- sure, you might think the map, the features of the land, those are facts, but he added on to it and selectedwhich things are going to go on my map. So his map was about property lines and property values, in his opinion, of how to overlay the surveys on to the geographical features. So that was protectable. So when Montgomery Data grabbed his maps and updated them and amended them, that was deemed to be copyright infringement. Okay. So that was a protectable compilation, to add data to the map.

Then we get the case about the yellow pages where -- and this is very close in time to Feist, and the thought was, well, okay, maybe the white pages aren't copyrightable, but the way we've arranged the business listings, those are. So, you know, we chose to use the word "restaurants" instead of dining establishments. We chose to say "doctors" instead of "physicians." We chose these words and how to arrange people, and the Eleventh Circuit doesn't buy it, but there's a really strong dissent in this case saying, actually, those were --that was over the low bar. So there was a little bit of debate. You can imagine that maybe some aspect of yellow pages would be copyrightable, but in this case, not what Donnelley took.

Okay. Now we get into the really interesting cases for databases. This is about the RedBook for car values. So Maclean-Hunter publishes the RedBook. They have the listings of what is a car worth, and to do that, they argued that this is a compilation of their views of the car values.

Now, it's pretty clear that their decision about what cars to review and what variables are important about cars, those are a protectable selection and arrangement. On the compilation level, that upper layer that we talked about, this is protectable.

But there's dicta in the case that says that maybe even an individual car price is protectable, because that's the result of a statistical re-

gression that spits out a number of, here's our estimated valuation of this car. So maybe one number, because it compiles other numbers from a statistical formula, maybe that's protectable.

Now, that wasn't crucial, the holding in this case, but the Ninth Circuit picks up on it in a case about coin prices and said, yep, we're going with that, because in Maclean-Hunter, the whole -- CCC had taken the whole RedBook. They had just taken the whole database, and so all they needed was to say, well, you took our whole selection and arrangement.

But here, only individual coin prices were taken from Kapes' newsletter. So here the Ninth Circuit actually had to rely on the idea that an individual coin price was protectable under copyright. They held that it was Okay. So this is the high watermark for protection in really small things.

Not all the cases go this way. Obviously, I will skip through the overviews.

We have some other ones where it goes the other way. People -- like in ads, as in the phone book case, we see a few more. In more recent years, our courts are resistant to this kind of thinking, that the individual data.

So in a case that should be close to the law student's heart. This is a case between two law school publishers, Matthew Bender and West. These the two that merged? I forget, maybe someone else.

This was a case about whether the star pagination in West's Reporters of cases is copyrightable. So Matthew Bender made a CD-ROM, taking the cases, making it searchable, and allowing you to go through the West Reporters and search for cases. Obviously, the cases are in the public domain. Those aren't copyrightable. They're public. They're government works. However, West's argument was that you could re-create their way of arranging the cases, because you would know that their original compilation was, hey, we figured out what pages these things should go on. So the little indicators of, --the star, this was page 636; star, this is page 637 that was copyrightable. The court rejects that, again, over dissent, but a divided panel. The Second Circuit says that those aren't copyrightable.

Then we get a case about load ratings on ball bearings. Here the District of Connecticut says we're not buying that a single load rating reflecting opinion or an expressive thing, we are going to say that that is not copyrightable, that one number that is summarized.

Then the Southern District in a case about interest rates says that doing this over here, quoting Bancorp's national average of interest rates, to put in an advertisement, that our interest rate is above the national average; that one number, 3.95 percent, not copyrightable. Why? Well, because it's a simple average of government published -- a five-government published interest rate. So they said that method -- first of

all, that method of compilation doesn't seem very original. You're given five numbers, take an average. Humans have done that before. Sometime before 2013 it may have occurred to someone to compile them that way.

So, that was part of the reasons, but also the court expressed doubts on the number of lines basic to these lines between abstract ideas and expression, that this was not copyrightable expression.

So we see some cases going the other way on this issue of the single entries of the databases which can end up being important, because a lot of what happens in these disputes is people taking some subset of database and saying, okay, I'd like to use these particular instances.

So just to sum up. Apart from the copyright cases, the individual element is probably unprotected. Those are most likely viewed as sensor data, things like that, probably characterized as facts. Curated database might be protected as a compilation, but it might not. It depends on how much of it someone takes. It depends on the circumstances. Outputs of algorithmic data analysis might be protectable but might not. It depends on whether they go with the coin price case or whether they go more with the interest rate averaging case and the sense of those opinions. So you can -- you can see that the case law goes another direction.

What this leaves you with is that the terms of use and the contract agreement are really going to be important. So, of course, one of the most famous cases about terms of use from now almost 20 years ago, Pro CD v. Zeidenberg,[1] Mr. Zeidenberg paid a lower price for a version of Pro CD software. He wanted to resell it for a higher price, wanted to take advantage of the differential pricing, kind of exploit the idea of the difference between individual pricing and business-to-business pricing.

The Seventh Circuit says that the shrink wrap contract is enforceable. It doesn't matter that copyright law, as a default, would have a first sale doctrine to allow you to resell the CDs.

So the question in these kinds of cases about how contracting and copyright overlap is the question in a way of, if copyright says that this thing is left out of protection, can you, nonetheless, protect it by contract? In particular, can you ask your users that are agreeing to your terms of use, to say, oh, yeah, I won't make fair uses of the data? I won't take advantage of the statutory exceptions that I have to make backup copies of software.

In Pro CD versus Zeidenberg, the Seventh Circuit takes a strong position that, yep, you can contract that away. It doesn't matter.

There are some cases that go the other way. I will just pick one. It was the case Vault v. Quaid[2] from 1988. The Fifth Circuit says the op-

1. Pro CD, Inc. v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996).
2. Vault v. Quaid, 847 F.2d 255 (5th Cir. 1988).

posite. They say, actually, if a state law action -- here it's under a statute; Louisiana had a statute, the Software License Enforcement Act. It's not common law contract, it's a statute. But they said, this statute is actually unenforceable. It's preempted by the Copyright Act. The Copyright Act meant to let people make backup copies of their software. Congress made that decision to write that exception in, and for a contract to interfere with that statute, that would be state law interfering with federal law.

So I won't get into the ins and outs of preemption law, but I just want to say that there is an academic debate, at least some difference in the case law, about whether a contract can really be so -- I don't know what word to use other than "grabby," that the contract can be so aggressive as to say, okay, users, you agree not to make fair use. You agree not to write a critical review. You agree not to make backup copies. You agree not to show this to your wife. You know, some courts are happy to enforce whatever the contract says. Some courts not so much.

I am going to hand it back over to Marcelo to tie this up in terms the practical advice.

MR. HALPERN: A couple things just to pick up on things that people were saying.

If you think about the idea of, oh, they copied the entire RedBook. That seems like such an obvious thing; no one would do such a thing. People are still doing that. There are cases out there now of screen scrapers, as the terminology is used, people who go out, and they'll run a program that goes out to someone's website and repeatedly queries things for prices of whatever they're selling or their inventory levels or whatever and grab that information. So they are effectively leaching information out of someone's database to do exactly that, to create their own products, to create their own services.

So it is still happening, sometimes selectively, sometimes they're trying to grab everything at all once. That's why you have those little CAPTCHA things, where you got to type in to prove that you're a human and not a computer, what have you.

The other thing is, on this last point of the extent of the contract law v. copyright law, that's also still an active area, and there's some relatively recent cases that have created a lot of controversy over -- more on the side of software and shrink wraps about how much you can restrict future sales that might otherwise have been permitted, creating a lot of controversy. I haven't seen applied data context yet, but you can see where the same principles are going to come up. So these are still very active areas.

Now, that said, we got to work with what we got. Right? So as a business lawyer, as someone who is advising clients, you work with the law that is front of you, and contract law is the way to go at this point.

So preliminary questions, you're trying to figure out what to do to protect your databases and the information that your clients are working so hard to create. The first thing is: what is it? Identify the data sets. What are you trying to protect? Where is it coming from? Is it coming from sources that you actually have control over? Do you, in fact, have rights in that data? There's lots of conversations in a lot of the agreements I work on, trying to identify whose data is this, where did it come from, what rights do we have with it, and are there strings attached to -- you can do certain things with the information, you can't do other things.

Then, of course, what do you want to do with it? The whole point of the business world is to do something that's going to be of value to your company, your shareholders, to make some money. Do you have to share it in order to accomplish the goals?

One of the underlying principles, when you're talking about data privacy issues, is never share more than you have to. All of the data privacy principles say only share what is necessary to do the transaction. At the same it is true in the non-data privacy world. If you're trying to protect information, don't give it to anybody. So only share the information that you have to.

Do you need to actually provide a copy of it or is some abstract or some inquiry access enough to do it? So think about how is the information moving from one point to another. Is there actually a separate copy of it residing on someone else's system where you have less control over what happens? Or is it on a query type of basis?

Of course, the most secure information is the data that no one knows you have. So people -- if you're collecting information, and I think someone's making disparaging comments about the information that Google is collecting that people don't realize is being collected, well, that's pretty secure information if no one actually knows that you got it.

So the next question is: so you've got these huge data sets, you compile it from a lot of different places, who owns it? This is an area that's still really up for debate, particularly when talking information about people. Does the data subject own their own information, or does the company that collected the information own it, or is there some joint ownership of it?

So does Amazon own the fact that I bought this book? Or do I own the fact that I bought this book? Or do we have some sort of joint ownership in that piece of information and the compilation that goes with it? It's Unclear. The law has not really weighed in on this in a lot of ways.

There's some -- I think in the EU there's a movement a little bit more toward the data subjects having at least control over their information, but no one has really gotten to that ownership question directly. It's still unresolved.

The other questions, if you're sharing the information, who owns it? Is it the collector of the information, or is it the recipient on the other end that owns that information?

When you're renting and buying data, similarly to the phone book example, you can buy lists, marketing lists. I need a list of all the lawyers in Chicago, because I want to market this conference to them. Nice list to have. You can buy that list from somebody. Usually, you can rent the list, because you are allowed to use it once for your purposes. They do the same thing that Feist did. They put fake names in there, so they actually get the invitation to this as well, and they know if that list is getting reused.

But once you've taken that list and you created a derivative set from it -- I filtered it down; I applied it to something else; I have cross-referenced that list against a different list to refine it further -- who owns the resulting copy of that? Is that mine? Is it the original owner's? What do you do with it? How do you determine what that derivative data set accomplishes? It goes, again, to some of the issues Peter raised in terms of the application of the algorithms that come up with this number. That's a number that was derived from a data set. Of course, we did it to a bunch of data sets, so we've got lots of numbers that were derived from those algorithms.

So that mining, the network effect, that combined data set, the analysis of it, who owns that? Is it the owner of the course data or is it the person who did the actual analysis that came up with that number, that came up with the coin price or that came up with the average?

So what do you do? Contract. So Pro CD taught us that contracts work, at least for now until someone tells us they don't, and trying to set up the contracts in a way that makes sense. So think again about who owns the data that is being shared, the, identify it in the contract.

I am sharing information with you as between the two of us. I own it. I may have gotten it from someone else, but that doesn't matter, as long as I've got the rights to give it to you. Of course, you want the reps and warranties and so forth that I, in fact, have the right to give it to you. Then identify who owns the derivatives of that data.

So set it out in the contract. You're going to use this information to do X, Y, and Z, and you own that derivative data or you have some limited rights in that, or I have limited rights and the algorithms that go with it. A lot of the business that gets done is, I have a proprietary algorithm, but my algorithm isn't any good unless I've got the data to apply it against, so I am going to try and acquire that data from someone else that's collecting it so that I can produce my, for example, coin prices.

I don't have to – I'm not the one who bought and sold the coins that created those prices. I need the collection of that sales data in order to create my expected price using my proprietary algorithm.

Look at all its limitations of use, this is not -- these are not trivial

things to identify in a contract, because you're drawing some very, very fine lines sometimes as to what you can do and what you can't. Of course, the Internet is full of business models that haven't been developed yet.

I've got an idea. I think I want to do this. But, look, I did this. That would be really cool to do, too. Let me do that with it. Can I do that with the data? Identify where those limits are, and that includes the right to combine the information, the analyzing, the mining of the data, and the ongoing distribution of it.

If I have given you a copy of my data set as part of our contractual relationship, do you now have a right to give a copy of or a subset of that data to somebody else? If you think back to that Ghostery chart of how many lines of information are going out, remember a lot of those people are actually collecting the same information. So there may be three different companies that have collected the fact that you clicked on that ad. Who has the rights to that data, and who gets to control that information.

It's not easy, and you got to think through all of those positions, then look at how the restrictions apply if the elements overlap with data obtained from another source. So a classic situation is, we are sharing information about mutual customers. We both have the customer's name. It's my customer, and it's also your customer. Now we're sharing information that each of us has acquired in the course of our own independent relationship with customer. Well, what about those overlapping elements? If you told me something I already know; can you lay claim to that or ownership to that in some way?

Then consider this; do you really care if someone else has the data? There's a lot of discussion that goes on around theoretical use of information. Sometimes you look at it and, if you step back, you say, is this information really that confidential? Is it dated? Is it something that's going to become irrelevant, --If I give to you and in a month this information is going to be useless because it's really specific to current events? Or is the real issue the confidentiality of the analysis of that data?

Where is the value? Think about what is driving the value of the business transaction and set up those restrictions accordingly.

Algorithm software might be protectable under a copyright patent or trade secret, pick your poison, and that's part of the debate, too, in these algorithms and the idea of the one number or not, is, should that be more accurately described as a patentable idea as you have developed this algorithm, which would not be subject to copyright, or is the number that's at the end of that algorithm a copyrightable piece of information?

If you don't tell people the algorithm, they can't develop that number. They may come close. They may try and reverse engineer it. But if

you keep it as a trade secret and only divulge the end result, that's a way of maintaining control over the end data. So I will give you the results rather than I will you give the data, is another way to go about it.

Remember, when business people look at all of this, the data at rest has no intrinsic value. You know, if you have a phone book sitting on your shelf, it's not valuable unless you need to look up a phone number, and then it's potential value that we're looking to derive in the contractual relationships.

And I think we have a few minutes for questions.

FROM THE FLOOR: There are a lot of cases, working their way up the courts now as to not just who owns what, but what's the sale, what's a license, what's involved. I'm just wondering, one of the first cases like you're talking about from what we call the present activist, the Supreme Court, got rid of about 97 years of prior cases saying that vertical price restraints are not a, per se, antitrust violation.

Is that case going to help or give people a glimpse of the decision of the Supreme Court on the kind of cases that you're talking about?

MR. HALPERN: I think so.

MR. DiCOLA: I think you're talking about an antitrust case. These cases don't typically involve antitrust. The way antitrust intersects with copyright is through the misuse defense, which is really underdeveloped and pretty sketchy at this point. So the connection, even though it's certainly true that IP intersects with antitrust in different ways, I don't think the jurisprudence in the antitrust has affected it.

FROM THE FLOOR: But that case involved vertical price restraints. So what you're talking about is if someone owns something or if they bought it rather than licensed it, can the person who originally made it -- like the Chicago Bears, can they stop somebody from reselling a ticket two or three times?

MR. DiCOLA: Yeah, I'm not disputing that they're conceptually connected. I'm just saying, you asked a question about a prediction about what the Supreme Court will do and if it gives us any indication, and I'm telling you that I have never seen a case that connects those two bodies of law, yet.

So I agree with you that lots of courts in IP cases are making decisions that affect the structure of competition and affect those things. But the developments you are talking about are likely to come from antitrust and not from copyright.

MR. HALPERN: Yeah, I think that's right. The issue that people look at in copyright and in other owned property restrictions, that they give, to some extent, a de facto monopoly on -- patents are always talking about, it's a licensed monopoly on that idea. Copyright could, to some extent, be considered the same thing. I got a monopoly on my book. I'm the only one that has the rights to it. I can divvy up the rights.

When you start putting them into the context of antitrust, it becomes a really different analysis. I don't think the courts -- there are a lot of overlapping issues in antitrust and IP, but I don't think this really comes into play all that much.

MR. DiCOLA: Yeah, I think that's why they failed, like Burner versus Autodesk, which says -- that the First-Sale Doctrine can be trumped by contract. The courts in copyright cases just don't. They maybe want the two, but they just don't talk about antitrust law or they don't talk about competition economically in any kind of structured way.

MR. HALPERN: Other questions?

MR. DiCOLA: The Burner case is an interesting one if you're not familiar with it.

MR. HALPERN: Yeah, it is worth clicking on.

FROM THE FLOOR: I think this is on point, but how would you advise a client that hosts a website or that has a website on whether or not they should be protecting themselves from advertising beacons and stuff that is planted on their website?

MR. HALPERN: Well, generally if it's your website and you control the website, then you are usually able to control what beacons go on, because those are contractual relationships with the advertisers that are doing it.

That said, there are certainly persistent systems. In fact, it was some -- was it Verizon that had the sort of permanent cookie on the phones or something like that recently that they finally said they are going to take it off.

Again, the way that you can protect it, to the extent that you can, first you need to try and detect it and stop it if it happens. It's a whole technological question, how do you do that?

But in the website in the terms and conditions, you can certainly put language in there that says you can't do this, that anybody that is doing that, and that raises up a breach of contract claim.

That is also how the screen scraping issues are being handled, you

put it into the terms and conditions, you can only use this information for your personal use, whatever the subject matter is, and you forbid the wholesale collection of information the way that a screen scraper would.

Now, that doesn't stop them from doing it. What that does is gives you a contractual cause of action against them if you catch them doing it. So there's always the technological side of can you stop them from -- prevent them from doing it technologically and then the legal side of what claims do you have against them if they do it.

It gets you out of what a lot of people have tried to use before is the trespass, the chattels and things like that that have had, shall we say, mixed results. It creates a much stronger position to go into court and say, you have signed a contract when you came to this site and you violated that contract, and we've got remedies for that.

FROM THE FLOOR: What about the other side of that coin? So do you need to put something in your license agreement to protect you from, for instance, the users or consumers that are coming to your site that end up getting -- their information gets captured by the beacons or for past time.

MR. HALPERN: I'm not sure I really understood the question.

FROM THE FLOOR: Again, I am the owner of the website. Should I have something in my end-user license agreement that protects me from users who might sue me because their information was collected by Facebook, for instance?

MR. HALPERN: If a user came to your site by virtue of the web beacons and advertising on your site, their information is being reused in other ways. That tends to come out -- the short answer is yes, you should address it in your terms-of-use policy. It tends to show up from the privacy policy where you're talking about what you will and won't do on your site.

But that's a disclosure issue. Again, when the user comes, they're told, this is what's going to happen with your information. They have the ability to use the site, not to use your site, or to use somebody else if they want. It's like part of that 70-page document that people click through and don't read.

PROFESSOR SORKIN: I am going to have to, once again, cut it off at the point where it's getting even more and more interesting, but I'd like to thank our panelists.

We don't have a break scheduled now. So we are going to get set up as quickly as possible for the next panel.

While we do that, I do need to mention regretfully because the Bears are mentioned in response, that any transmission or account of this last panel requires express written permission of the National Football League.