

>> Jon Leibowitz: All right, I want to thank you all for coming this morning to day two of workshop session two, "How Will Journalism Survive the Internet Age?" Which some might actually call, "Will Journalism Survive the Internet Age?" I wonder about that myself. But in any event, we are -- of course, we had a long discussion about not saying that, and not making that the title, because we didn't want to be judgmental or conclusory. But I want to thank you all for being here this morning. At our workshop in December, panelists and presenters emphasized the importance of quality journalism to democracy. So for example, Paul Steiger, Propublica, highlighted the important duty of investigative journalists to expose corruption and give the public the power to affect change. Matthew Gentzkow explains his findings that fewer people vote when a newspaper closes. Which is not surprising, actually, because people who read the newspapers -- newspapers have a component about political discourse, and elections, and public policy. And people get an opportunity to read a newspaper probably become more interested in that. Rupert Murdoch and Arianna Huffington, despite very different perspectives on a number of things, I'm sure, both underscored a very similar need, which is for media organizations to continue to distribute the news that matters in a way that people can understand. But as news media organizations falter or continue to falter -- and we all know, this is of course going on -- consumers' needs are not always being met. In an effort to cut costs, layoffs continue. So just two weeks ago, ABC News announced that it would be reducing its staff, perhaps by as much as 25%. Coverage of state houses and local city halls is in a sad decline. And yesterday, I went and I visited something, a sort of start-up of refugees from the "Chicago Tribune" called the Chicago News Cooperative. It started by a interesting man Jim O'Shea, who had worked for "The Des Moines Register," "The Chicago Tribune." He had been the editor and chief of "The Los Angeles Times," and he -- their plan, and they have a little start-up in downtown Chicago. And their first contract was signed to give an insert, I think, once or twice a week to "The New York Times - Chicago Edition." But their plan is to put three journalists, have three reporters covering the state house, to have three reporters covering Cook County, very important in Chicago. Three reporters covering city hall in Chicago. And when they do that, and I'm confident that they will, they will have bigger bureaus in each of those three places than "The Chicago Tribune." I just want you to think about that. I mean, that tells you something about how a major American newspaper, "Chicago Tribune," you know, considered to be, maybe not in the highest echelon of newspapers, but certainly a very important and a very credible one is becoming basically a venue for advertising, more than anything else.

Plus, it's not really covering the news anymore. Advertising dollars for print newspapers are decreasing, as we all know -- as advertising and especially advertising continues to move online. And the move to online advertising really isn't surprising, because online, and mobile devices, are really where most people, and certainly more and more people are accessing the news. A recent study by the Pew Research Center found that 61% of the people surveyed accessed some kind of news online, compared to 50% who read the news in the local newspaper. Additionally, 33% of cell phone users access news via their cell phones. Consumers are also turning to their social circle for their news, with 73% of Internet users creating, commenting about or disseminating the news via social networking sites. And I would imagine -- I would imagine Rick Lane will be walking into this room soon. I imagine that those changes are even more dramatic when you look at different demographics, especially under 35 years old. In response, news organizations and journalists are searching for creative ways to either thrive or really remain alive and in business and stay relevant. Online experiments and competitive collaborations are beginning to surface. Nonprofit news organizations are sharing content with for-profit newspapers, many of which previously would never have considered using content from anyone outside their own staffs. Foundation-funded online news organizations are expanding their output in new directions. Existing and upstart news agencies are learning to use the internet in unprecedented ways, making access to information easier. Some news media organizations, including "The New York Times," are now taking tentative steps towards charging consumers for access to their news online. So I would say, you know, there's a lot of good experimentation going on that holds, I would say, some promise for the future of news, and certainly for the consumers who need to see it, and for democracy, which is so vitally attached to good substantive news and news organizations. It is too early to know for certain, though, whether these creative solutions can remedy the destruction of existing news organizations that's occurred thus far. What we do know for certain, though, is that there is no going back. Obviously, the internet is here to stay, and these changes, at least in part, will be here to stay as well. So today, we're concentrating on potential solutions in an effort to shed light on the problems, and where we go from here. And our question, through this workshop, is this -- are changed government policies needed to support the optimal, or I might even say, the minimally necessary amount of journalism that we need in our society? My own sense, by the way, is evolving -- and it's evolving, it's not there yet. And of course, we'll write a report and we'll be thinking about this with other commissioners, and with Susan DeSanti and the staff that's been

working on this so productively and so hard. But my own tentative conclusion is a qualified "yes." So long as, I would say, policies -- the policy is platform neutral and it doesn't unnecessarily entangle government and journalism. After all, we live in a Washington where, you know, the widget industry or any industry can lobby, sometimes very, very successfully for preferential tax treatment. It seems hard to me that we couldn't do the same. It seems hard for me to understand why we wouldn't do the same for, really, something that is so ultimately vital to the future of our democracy. And I know this puts me on the other side of probably both Rupert Murdoch and Arianna Huffington, and -- which I don't think is a bad place to be, necessarily. But you know, again, this is not the kind of creative destruction that we see when, you know, a travel agency is battered by -- by, you know, new internet direct ticket buying by consumers. That's a great thing for consumers and maybe a bad thing if you're a travel agent. But -- or, you know, and my staff heard me say is ad nauseam, Rob could come up and say it right now. But really, you know, if you own a hotel on Route 1, and they build Interstate 95, bad for you, great for society. I'm just not sure we want to take that risk, when it comes to the future of news. So yesterday and today, our workshop focused on policy solutions that have been proposed as ways to support journalism. On Tuesday, we discussed copyright issues, technology to lower the cost of journalism, and corporate and taxed approaches. And this morning, by the way, I was doing a back-and-forth. And Susan, I don't know if you saw this with the SBA, about whether SBA loans are available for start-up online news organizations. I think they ought to be if they aren't. And it's not entirely clear that they are. So hopefully we'll have an answer soon. And today, we're going to turn our focus to additional areas. First, we're going to hear from Jesse Shapiro. Is Jesse here? Thank you, Jesse -- who will discuss the evidence that shows there's a significant demand for news, in accordance with one's own political ideology. And I do -- maybe you can answer this question for me, because I do sometimes, and it's much more intuitive than empirical, think that that's sort of a problem, because, you know, too many people and too many lawmakers, you know, they watch MSNBC, or they watch Fox News. And you know, the twain rarely meet. And I think it really sort of does undermine -- and they talk to each other. But they don't talk to the -- the people who are watching MSNBC talk to the other people watching MSNBC. The people who are watching Fox News talk to the other folks who are watching Fox News. The twain rarely meet, and I think it really undermines bipartisanship, which I think is critically important to a functional democracy. We have a lot of bipartisanship at this age, and seeing it function very, very well. Joel Waldfogel --

Joel, are you here? Thank you. That's good, you're sitting next to Jesse -- will discuss media products, market failures, and media regulation. Our last discussion of the morning will focus on interactive data and using the Internet to lower the cost of journalism. We'll see a few examples of these ideas in action. And after lunch, we're going to review some journalism history. Chairman Ruth Goldway of the Postal Regulatory Commission, who I do not see in the audience, will discuss the past postal periodical subsidies, which were very important in the newspaper industry before and at the beginning of the rise of the penny press in the early 1800s. Geoffrey Cowan, from USC's Annenberg School, will discuss the history of government funding of the news. His discussion is based on a fascinating report released by the Annenberg School in January. And Robert McChesney is going to discuss additional proposals for public support of the news. And our last discussion today will focus on competitor collaborations that can help support news organizations. Our goal for this workshop is really to try to move this discussion in a positive direction. I think most of us now agree that there is a problem. Now we need to focus on what the potential solution should be. And let me just assure you, to everybody in the room, particularly everybody who is participating -- your participation is going to contribute greatly to our report. And I'm almost as confident -- actually, I'm equally confident that our report is going to contribute greatly to this very critical public policy debate that's happening here, really, around the country. And hopefully will happen soon, and I expect will happen soon, in Congress. So, again, I just want to thank you all for coming. This is really one of the most interesting issues we have tackled as an agency. We're looking forward to working with the Federal Communications Commission, which will be taking a slightly different approach within its own jurisdiction, but looking at some of these very same issues. And now I will turn it over to -- do you want me to turn it over to the first speaker, or I will turn it over to Susan DeSanti who has been doing a stellar job of running this -- running this series of workshops.

[ Applause ]

>> Susan DeSanti: Thank you very much, Chairman Leibowitz. I am required to give you all a security briefing first. And that may seem like overkill to you. But on 9/11, we were holding a conference in this room. So there's some precedence. So they're a little bit fussy about this. So here it goes. Anyone who leaves the building without an FCC badge will be required to go through

the magnetometer and x-ray machine prior to re-entry into the conference center. In the event of emergency, fire, or evacuation, the alarms will sound. At that time, please gather your personal belongings, if situation permits, and leave the building in an orderly fashion. Once outside of the building, orient yourself to the corner of 7th Street and Constitution Avenue Northwest, where you will enter the National Gallery of Art. That is our rallying point. Everyone will rally by floor, so please stay together. You will need to check in with the conference meeting coordinator. In the event that it is safer to remain inside, shelter in place, you will be advised where you should report to while inside the building. Information and updates will be distributed via the public address system. Please remain with conference meeting coordinator. If you spot suspicious activity, please alert the conference or meeting coordinator and/or security staff, and I would encourage you to notify the security staff. Our first presenter today, as the chairman has said, is Jesse Shapiro, who is Assistant Professor of Economics and Robert King Steel faculty fellow at the Booth School of Business at the University of Chicago. Dr. Shapiro's research focuses on the economics of communication and persuasion in the areas of industrial organization and political economy. Dr. Shapiro also is a faculty research fellow in labor studies at the National Bureau of Economic Research. And Jesse, I believe you did that research with Matthew Gentzkow that was referenced by the chairman, and maybe you can use a couple of additional sentences to give us a little more -- little better idea of that research. Jesse?

>> Jesse M. Shapiro: Thanks. Well thanks very much, Susan, and thanks to Chairman Leibowitz. This is joint work, as you mentioned, with my colleague at Chicago Booth, Matthew Gentzkow, who was here for the fall workshop. And at that time, he talked about historical research we've done studying the effect of exits and entries of newspapers on voting behavior, and in particular as Chairman Leibowitz mentioned, on voter turnout. Today I'm going to show you a different slice of our research. I'll talk about a paper called "What Drives Media Slant: Evidence from U.S. Daily Newspapers." This work, as I said, was joint with Matthew and is published. The published version is available publically on my website, if anybody is interested in reading more details. Today, I will give you kind of an overview of what we've done. So, as people in this room know much better than I do, two central principles that have been important for a long time in guiding U.S. policy towards the news media are, one, that a diversity of voices is good, and here's the Supreme Court in a very famous decision saying that one of the most vital of all general interests is

the dissemination of news from as many sources, and with it, as many different facets and colors as possible. So we want rich diversity in the marketplace of ideas. And two -- that left to its own devices, an unfettered media market may produce too little diversity. And one particular threat that's been identified over and over again to the diversity of opinion in the news media is the consolidation of ownership. The FCC, in this quote, has traditionally assumed that there is a positive correlation between viewpoints expressed and ownership of an outlet. And the commission has sought, therefore, to diffuse ownership of media outlets among multiple firms in order to diversify the viewpoints available to the public. This is a question that has been the subject of much research. But, despite a long history of effort to try to resolve this issue, I think the jury is still out on how important, in fact, ownership consolidation is in determining the degree of ideological diversity in the media space. And I think one of the primary reasons for that is fundamentally technological, which is that it's very difficult to quantify the content of a news outlet. News media content is obviously richly multidimensional. If it were possible to summarize it in a single number, the newspaper would be much shorter. It isn't. And as a result, most of the existing research uses some form of human hand coding to try to identify the partisan slant of news content. And that's been the limitation of the size of the sample people have been able to study. So a well-known study by Pritchard consists of only ten newspapers. A recent study in economics by Tim Groseclose and Jeff Milyo, that's been much cited. It talks about only six newspapers. So despite being of very high quality, these studies are limited in some sense by the number of media outlets they're able to consider. And that makes it difficult to answer some of the questions that are central to the policy debate. And as a result, the FCC has admitted that this remains an open empirical issue. They have said a larger number of independent owners will tend to generate a wider array of viewpoints in the media than would a comparatively smaller number of owners. We believe this proposition, even without the benefit of conclusive empirical evidence. The emphasis is ours. So this is where we try to come in and offer some contributions towards a possible resolution of this empirical question, whether in fact independent ownership is central to ideological diversity. And we're going to try to make several contributions in this paper. First -- and I'll say more about the details about this in a minute, we're going to try to make a methodological contribution and offer a content-based measure of media slant, and I'll say more about how we measure slant in a moment. Then I'm going to try to give you a little bit of a feel, you know, thinking of this like an economist, trying to look at the demand for slant and the supply

of slant. So we're going to look at whether, in fact, consumers demand ideological slant, and to what extent they do. And we'll use evidence on circulation data. And what I'm going to argue is that the data show that there's evidence that consumers value an ideological match between their ideology and the ideology of the news outlet that they consume. And then we'll turn and look at whether that demand is met with corresponding supply by news outlets. We'll show first that news outlets do respond to consumer demand -- that is that the kind of news demanded by the customer seems to play an important role in driving the heterogeneity across news outlets in the left-right slant of the news. And then I'll actually argue that in our data, there's little evidence that the owner plays an important role. That is, two newspapers with the same owner look no more similar in terms of their ideological position than two newspapers with different owners. Okay. So, first -- so first thing I want to tell you about is how we go about measuring slant. So one of the things that we face, there's a challenge in this study, giving our goal of trying to extend the sample size, is to try to find an automated, portable way to quantify the orientation of news outlets on a left-right political spectrum. And the way we decided to approach that is to study the strategic use of language. So let me give you an example that illustrates our approach, and then I'll try to overview in more detail exactly what we did. So in the Social Security debate in 2005, there was discussion about the strategic use of language. And one of the ways you can learn about that is from a widely circulated leaked memo by Frank Luntz, a famous political consultant, which was circulated to 2006 Republican congressional hopefuls. And in the context of the Social Security debate, it gave them the following advice: "Never say 'privatization' or 'private accounts.' Instead, say 'personalization' or 'personal accounts.' Two-thirds of America wants to personalize Social Security while only one-third would privatize it. Why? Personalizing Social Security suggests ownership and control over your retirement savings, while privatizing it suggests a profit motive and winners and losers." Okay, so Luntz is offering strategic advice. If you want people to support Social Security reform and you're a Republican, what you should do is avoid phrases like "privatization" and "private accounts" and focus on phrases like "personalization" or "personal accounts."

Well we went to the 2005 congressional record to see whether, in fact, republicans were heeding that advice and whether democrats were heeding the converse advice to emphasize privatization and private accounts. And in the 2005 congressional record, democratic members of Congress use the phrase, "personal account," 48 times, republicans use it 184 times. Democratic members of Congress use the phrase "private account" 542 times, and republicans used that phrase five times.

So I take two lessons from this example that we have tried to generalize from. First, as an economist, I'm not surprised to see that people who have a political agenda are trying to find a language that's gonna tend to get a listener to support that agenda. So language is used strategically, like every other tool. And it's being used here to get the listener to be more sympathetic to a policy position. Second, from a statistical standpoint, what I've learned from this is if you tell me nothing about a Congress person in 2005, except how often they said personal account and how often they said private account, provided they said one or the other, I can do a pretty good job of guessing which political party they belong to even if you tell me nothing else about them. So what we've learned is that the frequency of these phrases is a good predictor of a politician's ideology. And so what we're going to build off of is the idea that we can go into the congressional records and identify partisan phrases. And then what we're going to do, just to give you some kind of road map, is we're gonna go look for those phrases in the news media. So here's another set of examples. On June 23rd, 2005, "The Washington Post" talked about the house GOP offering a plan for Social Security, Bush's private accounts will be scaled back. On that same date, "The Washington Times," a somewhat more right-wing publication said that the GOP backs use of Social Security surplus funds finds funding for personal accounts. So here's that same contrast in use of language now lining up with what we would expect are differences in the political orientations of these two newspapers. So what we're gonna do is go look for a bunch of phrases like personal accounts and private accounts in the congressional record. And then once we have those phrases, we're gonna go find them in newspapers and use that to try to index newspapers according to whether they talk more like a republican or more like a democrat. So the way we get our set of phrases is not from the Luntz memo, but rather from a computer analysis of the 2005 congressional record. What we do is we download the congressional record using an automated script. We identify with another script who is speaking at each point in time, and we match that to data about that speaker, including their political party. For every two and three-word phrase spoken in 2005, we count how often it was said by democrats and how often it was said by republicans, and then we select the 502 word phrases and the 500 three-word phrases for a total of 1,000 phrases that are most diagnostic of the party of the speaker. That is these are the phrases that for which knowing these phrases is most useful about identifying the party of the person using them. And -- so, one thing that I want to stress is, when you do something like this, there are pros and cons. So this is a very scalable method, and it's possible for me to state an algorithm that

produces the list of phrases but because we didn't have a human do it, there's going to be some messy examples and instances of things that don't quite make sense because we're trying to teach a computer how to read, and that's a messy business. So here are -- we have all of the phrases, posted online, but here are the top ones that come up, and just to highlight a few examples, here's our Social Security debate, so republicans talked about personal retirement accounts, democrats talked about private accounts, private accounts turns out to be the single most democratic two-word phrase spoken in 2005 in the sense of how diagnostic it is about the party of the speaker. Democrats also spoke frequently about the Social Security trust fund. In the context of foreign policy, the republicans were talking about the global war on terror in 2005, the democrats talking about the war in Iraq. If someone said war in Iraq in 2005, that somebody was very likely to have been a democrat. In the context of fiscal policy, more generally, of course the republicans use the phrase debt tax more often than the democrats did. Republicans use the phrase tax relief where the democrats use phrases like tax break and tax cuts for the wealthy. So, once we have our set of 1,000 highly partisan phrases, what we do is we match how often each speaker in congress says each phrase to an index of their ideology and a few of these, it doesn't matter very much which one we use, we use both the ideology of their constituency, that is how did their constituents vote in the 2004 presidential election. We also used some standard indices based on their roll call voting records, and those two things so are highly correlated with one another that which one we use is not very important. And then we use a statistical model to basically allow us to predict, based only on speech, how right wing or left wing is this Congress person. So we have a statistical mapping that says, based on the number of times this speaker said each of these 1,000 phrases, this is what -- how partisan or how right wing or left wing this speaker is likely to be. So we basically are able to identify a predicted ideology based only on speech and confirm that among Congress people, this measured as a reasonably good job of picking up variation in political partisanship just based on speech alone. Then we need to apply this to newspapers. So another computer script goes online and searches the full text record of 433 U.S. daily newspapers accounting for more than 70% of the circulation of daily newspapers at the time in the U.S. and counts the uses of each of these 1,000 phrases in news text wherever possible. And then we apply the same statistical model to the newspapers that we've applied in Congress and asked the question of the model, if this newspaper were a member of Congress, how republican a member of Congress would it be? Would this newspaper be a very right-wing member of Congress or a very left-wing

member of Congress? We get a number for every newspaper. Now, again, there are things we like about this method and things we have to live with because of the things that we like. This method has the virtue, again, of being scalable, reproducible in the sense that it's performed by computer algorithm and not by subjective judgment. It's portable across context, so any environment in which there parliamentary text, that you can identify with the partisan orientation of the speaker you can use to calibrate a method like this. And it doesn't require the employment of armies of undergraduate and graduate research assistants, which is maybe bad from a jobs perspective, but very good from the point of -- view of scalability, so it allows us to compute a number for over 400 newspapers which is very nice. The downside of that is that because this method introduces a lot of noise, and because the computer isn't taking account of context and meaning, for any one newspaper our method is going to be more noisy than a method that really involves a closed reading of the newspaper's text. So, for the aggregate of newspaper this method, I think is very good and very defensible. For getting a specific ideology measure for a single newspaper, I think this isn't the ideal method and you want to use something that involves some more significant human involvement. Nevertheless, this method does a reasonably good job of correlating with people's subjective opinions. There's a website called Mondo Times, where users are able to submit ratings of the ideology of newspapers, and Mondo Times was good enough to share their number with us and indeed, the news sites that are rated as more conservative on Mondo Times also show up as more right wing, according to our slant index. So, that is an overview of how we measure the slants of 400 and some newspapers in the U.S., then we want to use that to test some economic hypotheses. The first economic hypothesis is that people are drawn to newspapers that are an ideological match for them, which is something that I think we expect to be true. We wanted to see if the data, in fact, support that, and whether that's an economically significant force in this industry. To test that, we obtain data from the audit bureau of circulation in 2005, on circulation by zip code for a large fraction of U.S. daily newspapers. The audit data didn't include all newspapers, and then we proxify for a zip code's ideology by the share of donations, in the federal election commissions public register that go to republicans, as a fraction of all donations to republicans and democrats. And then we want to test again the key hypothesis that consumer demand is going to be greater for newspapers with a slant similar to consumer ideology. And I won't go into the details of how we implement this test, but we try to look within a newspaper's market, so say within a metro area are the more right-wing newspapers doing relatively better in zip codes that are relatively right

wing within that market and are there more left wing newspapers doing relatively better in more left-wing zip codes within their market? And this builds on methods that Joel has developed in the media context to study racial patterns of news consumption. We're here using, instead of for race, to look at political meaning. So in particular, what we're test is whether more republican newspapers circulate more in more republican zip codes and we find strong support for that hypothesis in the data. So again, this is something I'll skip the details of, but I think the evidence is very clear. At least in my reading of it that this is an important force and I think that's consistent with our prior expectations about the way people make consumption decisions in this space. So that raises the question, do newspapers respond in a manner consistent with profit maximization to this incentive. That is, do they tailor the news to match the predispositions -- ideological predispositions of their customer base? The first way we're going to test that, and the main thing that we do in the paper to test that is to study whether newspapers that circulate in more conservative areas, like more conservative metropolitan areas, measured by the voting patterns of their readers, have more conservative language, that is have a slant that's to the right of newspapers circulated in less conservative areas. And we find strong support for that hypothesis variation in the fraction voting republican in the 2004 election, explains about 20% or more percent of the variation in our slant measure which is actually fairly substantial fraction given the fraction of the variation that's likely just computer noise. So there's strong evidence that newspapers are tailoring the news to match the ideological predispositions of their customers and in fact, in our data, using some modeling, we're able to actually calculate for each newspaper what choice of slant would maximize the newspaper's profits and the data are consistent with the hypothesis that newspapers are acting to maximize their profits or circulation given the ideological predispositions of those in their market. The next substantive hypothesis of importance is the one that I started with, that two newspapers with the same owner have a more similar slant than two newspapers with different owners. Once we control for our proxies for the ideology of newspapers' customers we find no support for this hypothesis. So on the X-axis here is, again, after controlling for the ideology of the consumer, this is the ideological leaning of the newspaper, and on the Y-axis is the average ideological leaning of the other newspapers that have the same owner. And we find no evidence the two are related at all. That is two newspapers with the same owner do not appear to be more similar, in our data, at least statistically speaking, to two newspapers with a different owner. A good example might be "The New York Times," in 2005 owned the Spartanburg Herald-Journal, in

South Carolina. That newspaper was more similar to other newspapers in South Carolina than to "The New York Times," itself. Another hypothesis of importance that I haven't mentioned before but that is often talked about is that incumbent politicians pressure newspapers, either through access, or other means to slant their news towards their ideological preferences. And so, we've tested in the data whether having a republican governor or a republican representative, locally seems to be correlated with the newspaper slant, controlling again for consumer ideology and once again, we find no support for that hypothesis in the data. So, what I hope this paper contributes, most of all, is the methodology for measuring slant, that's portable to other context and widely applicable and allows us to look at data on a larger scale than was possible with human coding. I think we find evidence in the data that there's significant consumer demand for slant. That is, consumers have a preference for, or exhibit demand for consuming ideologically similar news. Firms, in turn, respond to consumer preferences by tailoring the news to match the predispositions of their customer base. Once we've accounted for consumer preferences, there's little evidence that the identity of the newspaper's owner is correlated with its content and little evidence that the party affiliation of incumbent politicians is correlated with the newspaper's content. Thank you.

[ Applause ]

>> Susan DeSanti: I'm interested in how you controlled for an alternative hypothesis, which is that voters are influenced by the way the newspapers present the news and that's what accounts for -- so, in other words, it's the chicken and the egg, so obviously, you've accounted for this and how did you do that?

>> Jesse M. Shapiro: Do you want me to use this microphone here?

>> Susan DeSanti: Go right ahead.

>> Jesse M. Shapiro: So, the way we try to account for that issue is to exploit variation across markets and ideology that's related to demographic factors and are unlikely to be directly affected by the newspaper. A good example would be church attendance. So, the markets where people attend church more regularly also tend to vote more republican and those markets have more right-

leaning newspapers. It's conceivable that the reason people are attending church more often is because of the politics of the newspaper. But if you're prepared to rule out that hypothesis, then this method kind of resolves that issue in favor of the hypothesis that the consumers are the chicken, I guess. Okay. Thanks very much.

>> Susan DeSanti: Thank you. Our second presenter today is Joel Waldfogel, who is the Joel S. Ehrenkranz Family Professor and Professor of Business and Public Policy at the Wharton School at the University of Pennsylvania. Within industrial and economics, Dr. Waldfogel has conducted empirical studies of price advertising, media markets and minorities, the operation of differentiated product markets and issues related to digital products, including piracy, pricing and revenue sharing. And he's going to discuss challenges for media markets and possible regulation. Joel?

>> Joel Waldfogel: Thank you very much. And thank you for involving me in this conversation. Some of what I'll say, since I didn't know what others would say, will just be the same depressing stuff on what's happening, and some of it maybe will be some other things. All right, why don't we just go. I thought of this last year, I guess its 2010 now, I should note that I'm aware of that. Okay, so I want to talk a little about a few things. The nature of media products -- so this will be in some sense the economic history, as well as economics going forward. Then I'll talk about recent technological challenges that I think are bringing us all here. And then, finally, I'll talk about I guess what I think of as implications for the well being of the industry and consumers, using the FTC words diversity, competition and localism, but they're really just words about the well being of various participants, including the consumers of news. I probably won't talk about things to monitor. All right, so a little about the nature of media products. Media products are special in the sense that they inform us, sometimes they entertain us. They're special, they're unlike say, usual widgets, because usually, we just care about the buyer and seller of widgets. Of course, in media product, most of which are advertised or supported, the buyers and sellers are the buyers of advertising and the displayers of advertising. Here, we have this very important third party, the consumers of information. Now, they are in some sense buyers for some of these products, they pay for newspapers. But really, they're not the major source of revenue. So this is a special product, because there's this third party who's not directly involved in the important economic transactions. You know, 80% of newspaper revenue, we heard yesterday from Hal, is advertiser

revenue. Now, as products, even apart from that special feature, news product; as products, they're unusual and, I think, interesting in two other respects. Media products have really high fixed costs. That is costs independent of how many copies, or if it's a physical copy, or how many people consume the product. You have high fixed costs. And that means, you know, if you think about it, fixed costs have to be covered with revenue per -- sort of per user. That revenue per user times the number of users has to be big enough to cover that big fixed costs. So what's available depends essentially on the value that advertisers attach to users, not the value that users attach to content, and you need a lot of users. You have these two kind of perverse aspects. Oddly, whether we get informed, therefore, depends not on our appetite for information by and large, but on whether we're in a large segment that's sought after by advertisers. Now of course there are exceptions to this. You know there are magazines that are largely user-supported. And there are kind of a spectrum of kinds of products. But for newspapers in particular, by and large we're talking about advertiser support. That's in the traditional mode. So whether we get a newspaper depends on having a lot of folks like us who also want such a newspaper. And so as a consequence of the way these markets work, large groups tend to get more products. So you know, I've studied, for example, the provision of newspapers targeting minorities versus non-minorities, and it's clearly the case that people in small groups who have different preferences either get no product or small product, and so forth. So being in a big group favored by advertisers is an advantage for getting this byproduct called information. All right, so what are the ensuing market failures? And again, I'm just talking about history here. You know, we were thinking, I think yesterday, and the way we talk about this industry, we think that everything was wonderful before, but actually, even before, there were some things that made this a complicated environment. So let's talk about that. The market failures we could have expected already, we could sometimes have expected inefficient under provision. Now why is that? Well, whether something had ought to get provided is determined by whether the area under the demand curve, or the value that people would attach to this product, exceeds the cost of providing it. Whether it did get provided depended on whether the revenue exceeded the cost. Now of course, it's possible if not all of that willingness to pay can be captured as revenue for things that ought to get provided to not be provided. All right? Now usually, economists say, "Well, if price discrimination's possible, that's not much of a problem." Of course, turns out price discrimination, while it's possible, still isn't going to get all the area under the demand curve as revenue. We don't know much about that, although I've been doing some work on it. So in any

event, there was a possibility that small groups with different preferences wouldn't get stuff or would inefficiently not get stuff. Now, one context that you can think about -- and actually, I was reminded of it in yesterday's discussion of possibly granting tax advantage to newspapers, of course, we have public broadcasting, right? And one way to think about public broadcasting, which isn't really public, but that's a different story, is simply a way to correct market value. In the old days, we also wondered whether it was simply cannibalizing commercial activity. But of course, if commercial activity goes away, you gotta have something. So anyways, at the same time that we have inefficient under provision as a possibility, we also have the ironic possibility of excess entry. So the way to think about this is that when a private, say a second entrant enters, the benefit he experiences is partly new demand, but partly just diverted demand from existing firms. The social benefit really is reflected by the new demand, whereas the other part is a loss to one party, a gain to another. The private benefit tends to exceed the social benefit so it's entirely possible to have too much entry in markets like media markets. And probably not so much in newspapers today, but certainly in radio markets, there are situations like a Dallas that has six country stations. Sorry if I offend country fans. They're not really literally the same, but they're rather similar. All right, so what do we know about -- now, it shouldn't so much be what do we know about regulation, but stuff that is affected or might be affected by regulation? We know that ownership matters for content. Now, notwithstanding the paper we just heard, which I actually agree with and believe quite firmly, but let me talk about the distinctions between this and that. First of all, we know that ownership concentration can matter for product positioning in multi-product environments. And my evidence for this first is the radio consolidations that occurred pursuant to the Telecommunications Act of '96. We had huge increases in local ownership concentration. And what we got as a result of that was a kind of a spring out of products in the product space. Monopoly tended to reduce duplication of formats and enhance variety or at least per station available. So we know ownership concentration matters in these multi-product environments. Monopoly environments like the ones that Matt and Jesse have been talking about are a little bit different. We also know that in some contexts, the owner demographics actually matter for targeting. So my evidence of this is again from radio markets where black-targeted and Spanish/Hispanic-targeted stations, Spanish language stations, the number of black owners in a market has an affect on the number of black-targeted stations over and above the man characteristics. So there's some reason to -- and that's a little bit ironic, because on the one hand,

virtually every black-owned station was black-targeted. But on the other hand, most black-targeted stations weren't black-owned, right. So you didn't need black ownership to have black targeting. Nevertheless, the evidence seemed to show that ownership changes that changed the number or reduced the number of black owners over and above everything else that was happening seemed to reduce the total amount of black-targeted stuff. So again, just stepping back, ownerships can matter for the targeting of these products. Secondly, maybe more important, content matters for behavior. And so this is true in Matt and Jesse's paper on closing newspapers. I've done a bunch of studies on related phenomena. And let me just say, the presence of group-targeted content affects voter turnout. The reason I focus on group-targeted stuff is that, you know, newspapers are generally ubiquitously available. It's very hard to study things that are ubiquitously available. If they're always there, there's nothing to look at, you know. But things like Spanish language local television news is not ubiquitously available. It's available in about 25 U.S. markets, up from about 10 15 years ago. And so in that context, you can see what happens to Hispanic voter turnout in places that get a Spanish language local newspaper, and sure enough, it seems to go up. So that may be no surprise to normal people, but economists need evidence for thing that seem obvious to everyone, otherwise we just don't feel good. So this conversation matters, I think. The provision of information affects stuff that we very well might care about, as if you didn't know that. All right, technological change. The internet -- and some of this we heard yesterday -- the internet brings a number of innovations all at once. Most of which, frankly, are just plain good. Right? So this is, you know, we're complaining about something that's mostly good. One innovation is market enlargement. So now consumers anywhere can get access to products anywhere, here or elsewhere, which means the geographic size of the market that I can serve isn't just my local physical market, but it can be wherever. Internet brings reduced distribution costs. This has got to be good news for both consumers and firms. It also brings a greater ease of switching. Instead of having to subscribe to a different daily paper and wait for it to arrive on my doorstep, you know, I can just put in a different URL and look at a different one. Surely this is good for consumers. This is kind of the "more competitive aspect" of it. But, and this is, I think, the big "but", is the availability of these easy access to substitutes makes it much harder for sellers to charge for stuff, especially stuff that's not particularly distinctive. There's a paradox here, I think. Digital distribution makes information. And by information, I mean all the zeros and ones, any form of information you like, ubiquitously available. I mean, music, video, audio, it's all there on my telephone. And you walk around

Washington, or the Penn campus, and you see everybody listening to music for hours and hours a day and staring at their devices as they walk. The service flow coming off of these digitizable products strikes me as being at an all-time high, but because of the challenges to appropriability, the sellers of these things are really hurting. And the consumer perspective is "Why should I pay for things I can get elsewhere for free?" The seller perspective, as Hal pointed out with the Bertrand Paradox, is "How can I charge for things that other guys are giving away for free?" You know, in music, it's literally piracy that is the problem. In the case of newspapers, I think there may be an analogy to piracy, but it's just not quite piracy. It's just that much of what I can get in the thing I used to have to pay for is available for free and better elsewhere. But the paradox, though, is that the value of this product category is at an all-time high, and yet, the appropriability seems to be at currently an all-time low. All right. So what are the implications? Given the nature of media product, what are the implications of this changed environment -- let's just say for the well being indexed by these words that the FCC tends to use. Diversity, first of all, this is a word that has multiple meanings, or maybe no particular meaning, but one version of diversity is variety of content appealing to different sorts of people. Certainly, market enlargement is great for this, right? If all of a sudden any product can see a national or international market, there's room for a great variety of products serving many different niches of consumers. So market enlargement would seem to promote enormous diversity, especially in nonlocal product markets. Reduced distribution costs as well promote the proliferation of products, which again should at least favor variety, if not some deeper notion of diversity that I'm not quite in control of. Competition is more complicated, I think. There's a stylized idea of competition based on sort of a text book widget market. And the internet does bring good news for textbook markets in the sense that they bring lower costs and more elastic demand, which we would think should drive prices toward costs in textbook markets. And particularly, if we were thinking about markets with u-shaped cost curves and increasing marginal costs, this seems just like great good news. But I think the nuance in media markets is that not only are fixed costs high, they're indigenously high -- I'll explain that in a minute. And so I think that the textbook notions of competition that produce benefits for consumers don't so obviously produce good news in these kinds of markets. Because the question is, in these contexts can firms still cover their costs and produce quality products. So, let's talk a little bit about quality competition. You know, how is it that newspapers compete? And I would argue that it's not mainly by prices -- it's not mainly by the prices they charge to consumers, but rather by the content

that they prepare for consumers in the hopes of attracting readers and then in addition, advertisers. When quality is produced with fixed costs, then -- you know, what's interesting is that a better product doesn't need to charge more than a lower quality product because it hasn't increased its marginal costs. I mean, of course, it has to cover its costs, so what I said isn't quite true. But its marginal costs aren't higher, and if customers -- or consumers agree on what constitutes quality and what's better -- for example, if they agree that more -- you know, more pages, more content, you know, is better, then a better product can attract a large market share even if the market grows large. So, instead of -- now, the textbook idea, there's some fixed cost of operating. The market gets big, we can just accommodate more competitors. But if that's not necessarily true, as the market gets big, I can invest more in quality and undercut the prices of my competitors, I can stay big. So, these are arguments that are not mine, but John Sutton's from his famous book, "Sunk Costs and Market Structure." Well, I think this characterization fits some media products really well. The quality is produced with the fixed costs, not the marginal costs. Probably more so newspapers than radio. You know, in radio we have these different formats and bigger markets have a proliferation of formats. But think about newspapers and this, again, is becoming economic history. But across, for example, US cities -- bigger cities don't have very many more papers. A little bit more -- I mean, New York arguably has three dailies that are targeting the whole metro area, depends how you count. But Fergus Falls, Minnesota, population 12,000 has one. New York's about a thousand times bigger and it doesn't have a thousand. It has about three, or maybe ten, depending how you count. So what really is happening across these markets is a different investment and quality. They have lots more reporters and editors, and therefore bigger endogenous -- they choose to have higher fixed costs. They have more Pulitzer Prizes for reporter and reader and so forth. Now, the thing here that's maybe a little new is the notion that it's not really obvious that having more products is better for consumers in a situation like this. It might be better to have one high quality product than a bunch of low quality products. And, you know, I think about this when people talk about, you know, the amazing fragmentation that's going to happen with the internet. I'm not so sure that's right. I'm not sure we're going to have a thousand people hanging out a shingle on the internet producing, kind of, products that can attract anything. Maybe you have 500 people who get together or maybe you have one guy who hires a bunch of reporters and editors and can actually make a much better local product than the singletons who hang out a shingle. It's just -- it's not quite clear that we're going to have the super fragmentation.

All right, so let me talk a little bit about, specifically about localism and newspapers, and some of this I think we've heard before this depressing news. I mean, we used to have a local product with lots of monopoly power, and no real easy substitutes, or at least, I mean, that's a too strong a statement 'cause of course there's always radio and there's always television. But this product, this local newspaper -- bundled both local news, national news, local sports, nonlocal other kinds of information, as well as classified ads, and there was this cross-subsidization that was going on. I mean it's entirely possible what we lived through up till 10 years ago, was a situation in which owners were able to indulge their preference for producing lots of news, and sort of a golden age maybe producing too much from a strictly economic perspective, maybe too much information. Although, it may well be that when you take account of the spillover benefits, for example, the fact that it gets people up to vote maybe it wasn't too much. But again from the standpoint of market participants, maybe it was too much. Now, the challenges that have arrived with us are the threats from the classified markets from Craig's List. The National online products that -- I mean, you know what's interesting, again, I think we subsidize information provision with advertising. But now there are these new ways to advertise where the advertisers don't necessarily feel they have to produce information, along with their -- along with the advertising they have to produce. They're looking at how. But Craig's List and so forth. And so it's really the weird way we happen to finance news that is sort of the problem. And so the force on bundling from technology that allows folks to produce advertising without producing news is what causes this whole thing to unravel. All right. It's a little about competition and localism, and traditionally, when people think about regulation, one of the big forms of regulation is merger regulation, and is traditionally been a concern about prices paid by advertisers, not usually as much worry as there ought to have been about the information consumers who after all aren't the big economic participants in the transaction. And there had traditionally been concerns about, you know, mergers, from  $n$  in the market down to  $n$  minus 1, and threats to voices and prices, but I think right now the concern is an  $n$  of zero, as opposed to a sort of 3-2, that kind of thing, and so the question is, do we have a source of local information. And this is where we get back to Matt and Jesse. Is one even bad in the U.S. context? It seems as though one isn't bad from the standpoint of producing bias. I don't want to suggest that we should let all of the barrel and sconies off the hook with this bit of evidence, but still, some kind words for Monopoly here, if you're worried about these mergers, could even be the cross-ownership rules are no longer are well rationalized, or prohibitions on cross ownership. All

right, so kind of finally my business school professor hat -- you know, this is a lot of experimentation going on that firms are being exhorted to do. If you think historically this is really what the economic historians and technology historians call a era of ferment. You know, think back to 100 years ago, the automobile. There were like literally hundreds of firms making cars. It wasn't clear whether cars would be steam powered, electric powered, gas powered. Wasn't clear whether bicycle manufacturers would be the winners, whether whoever would be the winners. Now, I don't -- that's not that comforting, but the point is this kind of uncertainty is not that unusual when you have this sort of revolutionary technological change. So it is true right now, dailies are shrinking, and that seems bad. I mean, I would think it's bad. It's bad for the outcomes that I claim to have evidence that it affects, and that we all care about. But at the same time, local weeklies seem to be growing, or at least shrinking a lot less. Old media journalists -- we hear a lot of stories about this -- are launching hyper local products. It's not so clear yet whether this is going to work, but there's, I think, some reason for optimism if you think about it. You know, if you need to be -- if you want to be successful, you need to have a product with few substitutes. And so, hyper local maybe has fewer substitutes. It could be that Twitter from the football game could kill you, but I don't know, maybe, maybe not. Plus, you know, hyper local product has a natural targeting of ads to local firms, you know, local advertisers. So there's some reason for optimism. But, you know, it's sort of too soon to tell what's going to happen. So I guess the point here is that it's not clear whether -- how much government involvement we're going to need, you know, if we look back to other episodes of this big kind of technological change. It's very hard to predict what's going to happen. Okay. That's it, for me.

[ Applause ]

>> Susan DeSanti: Thank you very much, Joel. Now we are going to move on to a presentation on our next topic, which is "Interactive Data and the Semantic Web," which I'm going to let our speakers explain to you, because they can do that much better than I ever could. We've heard a lot about how the Internet has challenged many traditional media models. But it is also enabled new ways of collecting, aggregating and analyzing information. And as Sarah Cohen explained to us yesterday, the hope is that new ways of aggregating data and making it available to the public will enable journalists to discover stories more easily, and therefore, reduce the costs of journalism.

The next three presentations, and the discussion that follows, will explore how government might use interactive data technologies, in conjunction with the Internet to enable professional journalists and the interested public to reduce those costs of doing investigative or accountability reporting. And our first presenter is David Blazkowsky, who is Director of the Office of Interactive Disclosure at the Securities and Exchange Commission. And he's going to tell us about the SEC's experience, in implementing interactive data for the reporting of public company financial data. Interactive financial data provides the public, such as investors, quicker access to the information they want in a form that's easily used, and it also can help companies prepare the information more quickly, and more accurately, as well as workers at the agency to analyze what's going on. David.

>> David Blazkowski: Thank you very much, Susan, for that kind introduction. And thank you, again to you, and to Chris, and of course to Chairman Leibowitz for the opportunity to be here today. I mean -- I click, I'm there? Wonderful. Talking about technology, it's nice to know that technology actually works for this kind of thing. As I said, it is a privilege to be here as the member of the staff of the SEC, which was, itself, birthed by the FTC 76 years ago. Just wanted to go on its own, as we are raising, really, our game from mere disclosure -- which is what we were created for -- to something to true transparency for information. Now, my last journalism experience was a long time ago, as a college freshman. But I do get -- I do remember that as a journalist, what matters are facts. And where facts are more plentiful, accessible and robust in meaning, analysts of all kind, including journalists, get new opportunities to change the game, to redefine the game, and interactive data, which is liberating facts broadly, financial information broadly, right now and other kinds in the future -- broadly and usefully promises that kind of impact, at least in business journalism, and I would submit probably beyond. Now here, I'll discuss a bit about what interactive data is for those who are not familiar with the term, how it works and how it might apply to journalism. But first as an SEC employee -- you've heard many of these. You don't see anything at the FTC, I am sure. But as a -- as a matter of policy, the SEC disclaims responsibility for the private statements of its employees. Therefore, the views that I'm expressing today are solely my own, and do not reflect the view of the Commission, the commissioners or of any employees other than myself. So back to the program -- let's take another look at the title, "Better Data, Better Reporting." I'm not mentioning cost anywhere in this, and I'm going to avoid that issue -- please behave -- because I think interactive data, because I believe that the issues of

cost -- because I believe that interactive data such as SBRL, which I'll explain in a moment, and technologies like it might cut costs profoundly. I think they will. But more importantly, they can bring the kinds of dynamic, new and compelling insights with the added advantage of low costs that make news, news. And you know, as the chairman pointed out in his trivia example, innovation is as important. It's value-add, in business jargon. You know, predictions of demise or limits are dangerous -- that's for anything, media and non-media. Let's not forget that in the '50s, the chairman of IBM was predicting that the market for computers was going to be exactly 30. 25 years earlier, the head of one of the largest mid-sized -- one of the largest computer manufacturers, DEC, at the time, thought that the personal computer was a silly idea. And of course, there is no DEC today, but there are, well, lots of PCs all around the room, aren't there? Financial reporting really does go back 4,000 years to clay tablets and cuneiform, which you might agree with me, is the hardest of hard copy. Frankly, that media paradigm took us up through-- hard copy took us up through parchment and paper, and into the late '80s, microfiche and other forms of films for recording and managing and distributing content. This paradigm really is all about documents. Documents are important, not just to lawyers, but to all of us -- documents that are filled with facts, but documents nonetheless, where you do have to read everything, pretty much linearly, to find what it is that you need. From the '60s into the '90s and still today, electronic formats like ASCII, through even the more well-known to most of us HTML and PDFs, they are still documents, numbers, terms -- concepts are all just strings of character devoid of semantic meaning. They are electronic, therefore more transportable, more convenient. But they're still pictures engraved in virtual clay rather than the physical stuff. The test -- you can't convert a table of data from a PDF into a spreadsheet any more easily than you can stuff a clay tablet into a hard drive -- certainly not without a whole lot of smart stuff going on behind the scenes. Whether you're a hedge fund analyst or a journalist, you can buy it, you can transcribe it, but either way you still have to convert it manually at some point. Even spreadsheets and databases -- easier to download and use, but hard to repurpose for the way you might want to use them, given the naming conventions, lack of standards and proprietary software and protocols that are involved with both -- with database standards and spread sheets. The world is better with them, but still not good enough. The same folks who brought you the web finally wrote a language that did for data what HTML did for images and pictures. XML, Extensible Mark-Up Language, is generic a standard made to transport data across the Internet. And it's been adapted to XBRL, designed expressly for business and

financial reporting and securities. Together, they are what I will define in their media presence, in the XML family, are what I'll be referring to as "interactive data." So what is XBRL, or Extensible Business Reporting Language? It's computer language, nothing more, really. But it facilitates the use of financial information by people, rather than merely shepherding it through the innards of a computer. It's a concept for really liberating financial concepts from the documents with which they're associated, to be found, to be taken, and to be used by those of us who need the information. Very important here is that it's a standard -- it really is a standard. It's a standard that can operate -- or interoperate, even better -- across platforms, applications, countries, and national borders. It really doesn't care what language it is. It's supervised. There are regulators, such as ourselves, standards setters, who are involved in making sure that this standard really is standard across the world. Something that's flexible, or extensible -- thus, the X in XBRL -- that's appropriate for things such as U.S. reporting, financial reporting, which is subject to many judgments by those who are preparing financial statements. Not everything fits into a box that can be identified. And finally, the transcends financials -- and I think this is where it begins to interact even more with journalism. Even beyond financial-type items, such that text, such that narratives, nonfinancial numbers, as well as things with a dollar sign or a Euro sign in front, can be represented. And of course, it's something -- this is a standard, it's not new to the world. It's something that's already used widely by financial regulators here in the States, such as us, and the FDIC, and by international regulators. So what XBRL does, and why it's useful to those of us who want to use financial information, business information in our work, whether we're reporters, or government folks, or investors, is that it adds structure and meaning to what otherwise is a flat and dead document. To represent the financial statements, such as you see on the right, a little excerpt - I know it's not readable. It's a financial statement -- an income statement, to be precise. So XBRL is really a language. It's a, for these purposes, a syntax by which to identify real world items, such as net sales, with vocabulary, or tags that come from a standardized list called a "taxonomy," see on the left. And the taxonomy here describes U.S. generally accepted accounting principles, better known as G.A.P., which is almost as complicated as the taxonomy of life, where you've probably taxonomy last in ninth grade biology. Because accounting is hard -- XBRL accounting is hard and complicated, but it has structure, which supports meaning in the branches of the income statement, balance sheet, equity state and footnotes, just as being in the mammal branch of the animal kingdom, means you are going to have four limbs and hair or fur. But the important

things that are implied by that, silly example, except that where you are really determines what you are about. Therefore, there's structure, there's syntax and there's vocabulary that can be used to describe where you are and how you relate to everything else in that taxonomy. And together, they enable the semantic value of context, rules and content. Interactive data -- and here, specifically XBRL, means that producers of info such as companies can tag their financial reports using this standard. That's basically the proposition. Each tag comes with a name like a bar code -- it's unique -- and metadata to describe its precise coordinate in the taxonomy and in the company's information. Here, net sales in the report correspond only to a gap definition in tact of net sales in the -- in the red oval, with a unique computer name in the red box in the top, and with metadata to describe its particulars, so that it can be found and used with precision. For example, currently, year, amount -- is it audited? Those are all examples of metadata. And they are all searchable, too -- as much as the actual numbers that are being used. What if the company has unusual or unique characteristics? You can create extensions, which are a complication to comparability, but are themselves of a measurable value all the same. Of course, each extension can be a red flag by itself. I'm begging for investigation, say, by a journalist, or a business school professor, or grad student. Comparability is important, but so are the things not necessarily comparable and stick out, or anomalous. And of course, this tagging applies to all of the good stuff that's in the footnotes, as well. And if I may foreshadow a bit, it's this very structuring built in at the front end, into XBRL, and then applied to each fact, that makes the data in XBRL so easy to transform via software into insights, into analysis, into reportage. Today, as we've said, all a company can do until now is submit their financials as a document, and leave the poor users on the other end of our financial disclosure system, known as EDGAR -- gem of disclosure that it is, to either print them out, transcribe them or pay for someone else to do the transcription as part of the service. It kills trees, it kills time, it kills dollars, it kills productivity. And if you submit the same thing to a machine readable XBRL, it's just gobbledygook in the technical sense of the word -- don't worry about not being able to read that, it's just code -- into the same EDGAR. And -- those would be snowflakes floating around, of individual facts. All of those individual facts, and you have multiple exciting means of access through viewers, download spreadsheets, RSS feeds, specialty applications -- right into the analysis that you have to do, for even the most granular or arcane fact, for whatever you're trying to do, whatever you're trying to use to find information to turn into analysis. That's what is different about XBRL -- it's intrinsically usable content, ready to consume by machines and by

humans both. So let's hit on some benefits for users. For companies -- starting there just for a moment, companies can distribute information better. That's the interactive part of interactive data. It's not just one way, it goes both. All companies with no time lag -- and let me emphasize, all companies, not just largest, with no time lag, all of the information that they present, whether they're large or small. But for the user -- let's start with faster, cheaper, better. The whole story -- the whole story, as granular as it comes with no filtering or softening by intermediaries, right, or, at least as right as the filer itself made it. But even more important is the functionality. And to you, the functionality, the applications, the feeds, the alerts, the ability to innately transform that data into other more usable forms, and looking at the productivity bullet, that you can do more with less. This is my one touch on the cost line -- both more and less. Productivity, but by increasing the numerator, news quality and quantity, even as the denominator cost can be reduced. But together, the two sides contribute to making the market efficient, which is good mechanics, as well as transparent, which makes for a good basis for understanding for the journalist and for the investor alike. Just a very quick look at our most well-known rule for interactive data -- for all publicly listed companies to provide their financial information and XBRL, their registrations and their periodic statements, especially the most well known, the 10Ks and the 10Qs, and their foreign equivalents. We're rolling out this rule in phases, starting with the largest companies. What we're requiring from all U.S.-listed companies is complete tagging of their financials, using the U.S.-gap taxonomy of all the primary financials, and in the first year, simple block taggings -- one tag for the whole block -- of their footnotes and schedules. Of course, and you know as well as I do, the good bits in the footnotes. They always are. And every company will have to use in their second year of tagging under the rule, the standard tags to tag all of the amounts, all of the facts of their footnotes and associated schedules. Think of that. Everything in the -- in that -- in the footnotes, everything in the financials, accessible, from litigation details, to derivatives, to lines of business, to accounting policies -- all available at the click of a button. The largest 500 have already started. You can come to SEC.gov, to our EDGAR site, and come on board. And you will find, already, three periods, including the annuals already filed for these largest 500 or so companies. Another 1200 companies are starting to file this coming June -- the midcaps order, called large accelerated filers. And then, the last 10,000 will start filing next year, June 2011, including even the smallest public nanocap, and all companies reporting to international financial reporting standards. So that's quite a rapid pace. Now, that will all be an immense, unique data set for analysis, investigation,

and comparisons, whether by journalists, or by academics, or, of course, by the professional and retail investor communities. But this is the only information on our public rule -- public company's rule. Or mutual funds will start reporting their risk and return, and performance information in 2011, and ratings companies will have to report all of their ratings very shortly. But in the end, it's what you can do with it. Excuse me. We, at the SEC, have seen cool, off-the-shelf tools -- tools that can and should be on the market for you soon. Better, we're starting to see free tools, widgets and more, emerge. Better -- emerge, and we even suggest come and visit our free viewer, which is on the SEC's website. Or looking at some of the points here, I mention "faster, better, and cheaper" -- again, referring not just to the cost element, but to the better element, the value add element. These are all opportunities to put every user of information, whether it's the largest hedge fund or the blogging journalist sitting at home. To be able to be, on, really the same basis for the first time, the information that's coming out of this for companies is actually better than even the largest hedge fund might be able to create today when you take into account that there's still a lag to do transcriptions, there's still a lag to normalize and standardize information. Metadata -- metadata, allowing you to look for unusual tags as a lead, looking for a precise item, looking for the kinds of things that might allow you to get the story -- the numbers behind the numbers, the story behind the numbers. For precision and providence, as well as machine -- let's call it "crunchability" and applications. Let me give you an example. Just on our own, at the SEC, we built, in the course of a morning, or one of my folks built in the course of a morning, a simple widget to be able take all of the footnotes, the common footnotes for all of the competitors in a particular sector, and just lay them out in one spreadsheet in real time, and add more as they continue to be filed -- an act which, while not profound in its nature, it's not necessarily hard, but would require today scrolling through to find that particular footnote and each one individually cutting and pasting and lifting it in, or writing very complicated scripts in order to make that happen, probably with a significant error rate. Now, when we start requiring the detail tagging of the footnotes, starting this summer for those largest companies, not only will you get the footnote with its full text, but you get all of the details that are within it, able to be laid up next to each other. I would argue that going up to faster, better and cheaper, those are opportunities to be able to get the story found, to get the story verified and get the story out, with all of those three attributes in place. Joining and mashing up with any structured data set, XBRL and XML, that's really a whole family of kinds of structured data. And the big story is there is structured data. The structures can be compared, can be aligned and using

software brought together and made to work. Remember, before I said the structuring up front is very important. It might even seem excessive to some. But the structuring up front makes it much easier to be able to pull together the information at the end. So it makes building software easier, using software easier. And of course, for the retail, the guy at the end of the day who might not be a computer whiz, to be able to just get to the analysis, just find the message. Also thoughts and alerts are easy to create. Social media friendly and interactive tools. Let me just give you an example there, there are all kinds of free widgets popping out on to the web. There's, for example, in the ratings space, there's already a site you can go to and you'll find ratings models that are out there that are being pumped with XBRL data from our financial site, into ratings formulas, that can, in true social media sense, can be modified and amended and commented on and compared by any participant, of course, you can download what you like. Don't download what you don't like. But it's an opportunity to bring real information into doing real analysis. It might supplement issues that exist in the current market in this case in the ratings market. And the night is still young, as is XBRL, XBRL is easy to use, and like with the web free ware will emerge but most important is what curious-driven folks such as reporters and teams will do. Like the chemicals commercial. XBRL won't alone do the job but in the hands of a talented journalist with tech savvy will make them immeasurably better. And probably a lot cheaper through saved time and travel, etcetera. But the real story will be the better story and are we talking just about financial data? No, we're not. We're talking about applications beyond as well, but I'm limiting my comments to the financial stuff. Point is, moving to the world of easily transformable data, we are moving from mere disclosure of documents into their transparency. Just as a parting page, a last page. Lets look forward to where this could go, even without SEC pushing it. SEC has currently five series of interactive data. Company stuff I mentioned before. Mutual funds and ratings, also in XML, ownership data. But other kinds of financial reporting information are able to be considered, whether by SEC action or by other means, more voluntary means. For example, in the case of proxy's, which have been recommended by our investor advisory committee, proxy voting to have that information clearly available, but for investor decision making and for analysis. These have been advanced by commentators in the press, SEC committee, standards, organizations and of course by journalists and other public interest entities. Not about the SEC only though our experiment is quite robust. Other agencies, the FDIC here in the U.S., private entities and other governments producing or planning such sets. Using the same standard, of course, you can mash

and interact these across the data sets, again because of structuring up front, the analytical side is much easier. And there's a network affect here as well, in terms of kinds of analyses, complexity and tools. The bottom line, through interactive data, journalists can have at their fingertips, in as good a form as anyone can better than any today, huge amounts of vital information. And they'll have the means to extract meaning and interesting analyses. In investments these would be called buys or sell divisions. To the staff or commission those are called leads. To the journalist I guess they are called leads too but I think they're talking about the stories that they would drive to you. And to their readers news they can use. To me, as a non journalist, talking about journalism does seem to be a very exciting story. Thank you very much. [ Applause ]

>> Susan DeSanti: Thank you very much, David. We agree it's a very exciting story. Our next speaker is Noel Hidalgo, who is director of technology innovation at the New York senate office of the chief information officer. And he'll discuss the chief information officer's experience in making the public website more useful to the public, including through the use of interactive data. Noel.

>> Noel Hidalgo: Well now that a third of the room has left, I feel less stressed. Thanks for sticking around and looking forward to hearing the conversation about how open media is used in the state legislature. What I will first preface the situation is what, the first series of slides I'll show you is what we came with, then the last series of slides is where we're heading to. Just about every single state legislature across the United Nations, which there are 110 different governmental bodies that represent states, territories or districts, this is the problem that they are all facing. And you can go through, state by state, legislature by legislature, city council by city council, and you find we all face the same set of tools and same number of problems. This is 100 years ago in Albany. Sadly to say not much has changed. The buildings have been destroyed, but the state house pretty much looks as epic as it ever has been. And it continues to be a -- the center of New York state's government, even though the city will continue to try to be an upstart, we have constant issues up in Albany that really produce the very interesting balance within the state. Particularly here is the inside of the New York state Senate chamber. Sadly, not much has changed since this photo has also been taken. And you can still only cram about 100 people in the galleries, which are on the second floor. We have 62 Senator, which represent roughly around 300,000 people, within New York. Most legislatures when you come to them these days have a lotus notes

web-based system. I don't know how many are from state legislature yourselves or deal with this type of information disbursement, but it's raw. It's not easy to get to. You then go to their legislative bill look up service and it's also raw, it's not easy to get to. If they're providing information, it's normally behind pay wall that is expensive. For example, the organization that runs our joint commission, the LDBC, 1/12 of their budget is generated by income which is going out to lobbyists and newspaper organizations and charging them money for information which should be within the public domain. Within the state Senate, we then, for many years, for the 40 years that the majority was under republican control, they had a news clipping system which was all about paper. It cost \$1.5 million a year to run, and people would show up at 4:00 in the morning. Stacked with newspapers from around the state. And they would exacto-knife them to bits and photocopy them and redistribute them around the state, tagging each issue in the upper right-hand corner with some type of tag. Our e-mail system was very much the same. It was lotus notes. It was -- really hadn't changed since the -- IBM had introduced the lotus notes system. Our internal intranet was also similar based on lotus notes, provided very little flexibility for us to communicate dynamically within the legislature. And kind of the saddest thing, but kind of the most telling of all of this, is that our constituent management system, tracking all of our issues, are the constituent's issues to the state Senator, is held on a green screen terminal like this. It wasn't until just three or four years ago that the democrats had access to this particular system, so there was a huge disparity in resources that were being given to state legislatures. So in 2008, the state house switches parties, from democrat to republican, and the question was how do we reform this? We had been consistently seen as the worst state legislature in all 50 states. So how do we bring this about? Well, to the majority, it was obvious. You have to hire a rocket scientist. And Andrew Hoppen was hired as the first CIO of the New York state Senate to think about using technology tools, to bring about greater access information to not just the citizens but to everybody. So, when we start this whole process, we started under what is open government. Because the New York state legislature had been considered an un-open government. Well, thank goodness, we didn't have to do too much home work, because Barack Obama, one of his very first initiatives, was to focus on the open government initiative. And my -- the test in the lower right hand -- left-hand corner is unreadable. But there are three general principles about them. It's about collaboration. Well, the first one is about collaboration. Just like you see here in the public street, where it is accessible to pedestrians, car, commerce, to cyclists, to individuals, to storefront. We need to think about data

as a collaborative experience, that it is part of the enterprising streetscape. Anybody should be able to get to these pieces of information and use them. It also brings about transparency, it promotes accountability, it brings and provides information for citizens, what the government is actually doing, and finally, it's participatory in nature, fundamentally participatory in nature, government is about participation, regardless of whatever some of these elected officials think. And so, how do those three things -- how does it really impact the state like New York, which is, you know, I think it's the 16th or 17th largest state? Third largest in population, Albany and New York City are separated by three hours, via train. It's a two-day walk. A day and a half long carriage ride. Just from around the state alone, Albany is pretty much centrally located. Whether you're going out from Buffalo on the eastern part of the state, whether you're coming up from New York, or coming from the north, up in Plattsburgh, so the digital technology eradicates that distance. Especially with the understanding that in the north country, you have -- you tend to have cell phone reception, you don't have to have -- tend not to have high-speed internet kind of activity. In the west, you have pockets of higher speed internet connectivity, and other methods of interactivity. What we've considered it, we've created a flat playing field that enables citizens of New York from across the state regardless of whatever technology tool they're using to access the information of the state Senate. Fundamentally, it's based upon open content. We take a look at web content at a very serious level. We focus on web, mobile and SMS you can pull out your phone right now, type in [NYSenate.gov](http://NYSenate.gov). You get a mobile version of our website. Fundamentally, the Senate is required by our radical rule, our rules were re-radicalized. Once there was a coup last year, where the republicans and some democrats, switched parties, and they created a whole separate Senate and became a month-long drama that still plays out to this day. But one of the cool things in this coup is the rules changed drastically. We're forced to change legislation put all of the content of our Senators committee meetings and public hearings, all of that has to be streamed live. We also took pro active steps that we're displaying our web traffic, so you can see how Senators' offices are actually using this type of information. We're the first state house to adopt a create a common license. We have feeds for just about everything and developing an API which, for each Senator, which, on demand, you'll be able to call up their contact information, most recent contact -- contact information, content that they're putting up there, and more or less you have this information, at the tips of your fingertips. What has it brought about is open questions. By providing an interactive website, the Senators have started posing the questions, more and more questions to their

constituents. On the Staten Island ferry, Senator Savino surveyed her constituents asked what do you want to change within your transit. And they said we need quiet zones. Senator Voelker and a majority of minority Senators got together and they all posed the same question which is, do we really want to have \$25 mandatory registration fees and they were able to get the governor to reverse that in the budget. We've produced -- because the Senate website is the culmination of 62 different Senator sites and 32 different committee, schedules now publicly available. And we're forced to do Senate session committee hearings and public hearings. Yeah, got them. Senators have now seen it's great to have everything else up there. And now that they've realized that their calendars are being out there and that stuff is being out there on the internet, they start to change their performance in regards to video. Our number one streamed event was the debate for marriage equality. When we started the debate, it was one of the few bills, and I'll repeat this again. It's one of the few bills that has ever been introduced to the floor of the state Senate that there wasn't a pre determined outcome. Pause for that. So we recognize the unique opportunity here to stream this debate that was going on in Albany to the general masses, across the U.S. and across the world on such a heated issue. We had 16,000 concurrent streams. It ended up that the number one viewed video from all of that is on the verge of half a million views on Youtube right now. That's something that would have never been possible if you had asked individuals to come to Albany to participate in all of this. All of our videos are online for you to take, to retake, and to sort of mash up into whatever system you want. We've discovered that now, by just providing these videos online, blogs and newspapers, particularly the "Daily News", "New York Post" and "New York Times" are really sucking in this content and using it on a day-to-day basis. Routinely, the blogs that have the political blogs of the "Daily News" and the "Post" are linking to live stream videos of different committee meetings over and over and over, as well as to our new legislative bill system which I'll show you in a minute. So once the Senators have realized that their schedules are out there and content is out there, they have really bought into this whole notion of social networking. To the point that there have been a few videos kind of critiquing different Senators and their how do I say this? More or less their online attitude of what they're doing online some Senators were discovered playing mafia wars on Facebook. Some Senators, such as Senator Marziaz uses this as a vocal point to combat what the liberal governor is speaking about. These Senators have engaged in online conversation. Senator Sitcowski has an ongoing conversation with Robert Harding on a frequent basis. There's this intense value giving these local elected officials a platform to continue

this conversation. Once again all of these feeds that you're getting from Facebook and Youtube and Twitter, as well as all aggregated within the grand API, so that way you can suck it into the database and monitor all this content. We continue to do further outreach and participation through online town halls and conferences. On conferences are these groovy idea where you pretty much propose a question and people come from whomever. Really you can open it up to whomever, come in and hang out, talk about the issues that are at hand. Here within D.C. there's quite a number of them that focus on the government 2.0 and open government. If you are interested in that, I can tell you more about that later. But we're breaking ground hosting conferences and sight of the state house. We've been really pioneering how we can do online town halls. Senators are required to spend half of the year in Albany. So how do we engage their constituents while they are up in Albany. We are going to have three online town halls on a Sunday come at the end of the month, which is going to be pretty interesting. The last thing that we're working on is idea generators and this is where it enables direct constituent feedback. Enabling individuals to, once again, you propose the question and let the constituent answer it. This is about crowd sourcing, but this is about finding out a way to have a -- not a moderated conversation but at least an educated conversation about what are the issues that are pressing and all of these tools that we've written are based on open-source tools. So Drupal is our main management system. We use the hub, which is an online social repository so that way you can download the code as we've created it and use it. We're pretty big in understanding just like as I showed you, the streetscape, is that code itself is part of that grandiose streetscape. It enables for commerce to happen like the laws and rules that govern the existence of our physical space. Like Lawrence Essex says 'law is code'. Code governs online space. We've adopted a very unique license where we dual license all our software code. One is initially put into the public domain under a BSD license which essentially enables you to run free and then to do whatever you want with it. If you want to continue to enhance the product in a collaborative, communicative manner, we license all of our code under a GPL 3 so it ensures there's no privatization, while allowing individuals to capitalize on tax dollars investment. Where does this all lead? It leads to fundamental re-understanding how you get through legislation. We've really been looking at a lot of systems that do this type of code. David talked about the security and the financial framework of all of this. But what is the legislative framework. If you go to open.NY Senate.gov, which is our legislative look-up search, we take a lot of cues from Google. It's designed specifically for allowing the end user to get easy access. Because bills are readable in

RSS, SML, CSG, JASON you can put it into whatever system you want, we saw "The New York Times" represent, and we said, well, you should no longer -- you should stop scraping the government data, we should be giving it to you, so you can go ahead and add enhancements to the constituents. We also allow people to vote, leave their comments and you can also see the actions of the Senate where these bills go, where they are the same as, in the assembly and most importantly, you can see the actual vote on these bills which, before hand you had to file a Foil request for. One of the last things that we've been working on is just like the news clip system that we used to pay \$1.5 million a year for, is figuring out how do we do a news, like a Google news system for the state Senate. We have, as I showed you, a very large state. We have multiple repositories of news that go from traditional news publications all the way down to small blogs. How do we incorporate those things into this news aggravation system? We've been slowly building this application, based off of day life, that incorporates blogs as well as major-named newspapers to incorporate both sides of views. And one thing that is interesting is that when a Senator has their name mentioned in any one of those different news feeds, they proactively reach out to those news organizations, even if it's a smaller blog, they'll have their communications person reach out to them and say, "Hey, thanks for writing that really great article." Or, "You were misinformed in your scathing criticism of me." So where does this come from or where does this go to in regards to state government as a platform? Well, as I've said, it's about multidisciplinary or multi interactive ways of communicating with the constituent. From voice calling, traditional constituent services, you pick up the phone, you get someone on the line, they answer your question; to SMSing and to the iPhone and ANDROID and smart phone apps. Instant messaging clients, where you have the ability that you can IM bill information. So you can sit there and AOL chat and just keep on pinging the system over and over and over. We have Twitter notification and Twitter replies. So we proactively send out information on Twitter and then we have a reactive system that also comes in that does queries. Everything below -- well, actually everything on the page except for some of the stuff that we're doing within the iPhone and ANDROID app was developed outside of the state Senate. It was developed by an individual in Virginia, marqueeed on a weekend, because he thought it was really, really cool, what we were doing. And we continued to leverage his investment and time and support his work, because he saw our work was valuable to plug into. So our one primary developer who's working on this legislative system, if you are in a state or location that is looking for a hand up, to make this a really better system, you should

contact Nathan Freitas. He's on Twitter as nathanfreitas. And he's really the mastermind behind understanding what is this legislative bill look-up service. And thinking about how the API can access in four different restful manners and how we can build this truly multi dynamic system that enables great interaction. And, like I said, everything is creative commons where we can. So thank you for your time. Thanks for this great opportunity.

[ Applause ]

>> Susan DeSanti: I want to thank you and David. It's really exciting to see what can be done in government. Even when it's having a few problems like the one in New York. Our final speaker is Jim Snider, who is president of isolon.org. He will present ideas on how interactive might improve journalists ability to report on public officials conflict of interest. He's going to discuss how reporting for public officials can be required in a way that's analogous to how the SEC is enabling better and less expensive financial reporting. Jim.

>> J.H. Snider: I have an accompanying working paper, close to -- I know I'm the last speaker, I'm going to try to raise the energy level. I'm quite enthusiastic about what I'm presenting so I think that's going to be easy to do. So, the outline of the presentation, I'm going to talk a little bit about what this excitement is about this briefly, the technology. We'll look briefly at three case studies where it's being used, including overlaps in the last two that we just heard but with a different twist, and then I'm going to present what I call the bias ontology, which is a way to automate conflict of interests reporting in a really powerful way. Hopefully it will take your breath away when you're done. So, the basic story line is simple of the semantic web. Great gains and efficiency for publishing government data and making it more accessible. The difficult part is the implementing it. But I'm going to argue, the payoff is worth that pain. So, here is a slide that I think really nicely summarizes, the semantic web. Now, when you look at this, it's a little bit jarring, because, by the time you're generally three or four years old, you know that a house is a house. And shirt and pants, you know what all those things are. Pants, you know what all those things are. And here they're labeled. But a machine, looks at that picture, and doesn't know that a house is a house and a shirt is a shirt. So this illustrates the core ideas of the semantic web from my perspective. The labels on these objects are metadata. We were introduced in the metadata in the XBRL description.

XBRL is a type of metadata. So these different items are tagged, that's metadata. To make it more machine reading able so a machine who looks at this picture can under it. And ontology, our logical linkages between the metadata. So an ontology would be the door is a part of a house. That would be a house ontology. Very simple ideas. Let's apply it to the SEC's page, where it has an address. So here I have highlighted some structure data on the page. And you could then add metadata to that. The title "Federal Trade Commission" is the organization. Washington, D.C. could be tagged as a state. The zip code is tagged. This is all done in the source code, behind the scenes. It doesn't affect what you're looking at. Then the whole set of metadata together forms an address ontology. So this is the transition that we're beginning to see between a human readable document centered web, and this machine readable web, which is what all of the XBRL and whatnot is about, allowing you to do all sorts of things that weren't previously possible. So now, so we have this three levels, we've got the data, that's the web we're familiar with right now, it's what we see. We add on top this metadata. And then the metadata are linked logically on ontologies. And then ontologies also can be linked together in sophisticated ways. Ontology is the glue that allows you to take all of this data that's scattered across the web, making it for sort of one virtual relational database. That's why the semantic web is called the web in the sky. You're taking all of this data and the experience that you might have had with an individual relational database, let's say, and you can apply it across the entire web. So, Here it is, a critical piece, it's not just well-structured data with all of these tags but it's posted to the web. And so again, the web, because a web of semantic data, data that's enhanced with meaning and that you can search across the entire web. So we've heard a lot about new technologies lowering the cost of journalism. The story I'm more interested in is vast improved journalism, as a result of this technology. Let's say we went back to the great days of high revenue journalism and we multiplied it by ten. Hell, let's make the whole economy devoted to journalism. I would still argue that with semantic web technologies, we can dramatically improve the quality of --, because we're talking about, in many cases, improving by a factor of a million or more, things that are just prohibitively expensive to do today, suddenly become possible. So we're talking about a qualitative leap in the quality of investigative research. Also, it's of really important problems that no journalists, even the wealthiest organization, even at the national level would do, because it's just too expensive, too much of a pain to do, suddenly becomes possible to do at a click of a button. So I'd say, "Let's forget about-- the cost side is important but it's a giant leap in quality which I would focus on as a result of this." And this goes

well beyond just political conflict of interest reporting. It also covers, all sorts of business applications and consumer protection issues as well. We'll see. This is a powerful language but it's simple. Okay. So, the technology is not the critical problems. The economics, significant market failure in the development of these type of ontologies, and real political obstacles; because people don't like to be made accountable. They are not going to push this, so you gotta have an external part of you doing it. So there's real economic and political market failure problems and political failures for making this come about. So, now we're gonna look quickly at three ontologies that are sort of shaking up the world. One is Google's snippet which was introduced last May of 2009. There are several components to it. Simple ontologies, they have a product line, a name on it, organization one. This is product reviews. Very simple. So in the old days, when you are doing a search, this is Capital Grill, it's about a mile or half a mile here from the FTC, nice restaurant, you get the old type of snippet is this type of snippet here. Here's the Capital Grill, and here's a little description of it. Here's the new type of snippet you get with product reviews. They rate them by stars. This is a summary of all of the ratings for that restaurant here. Then you can click on it and you can see all of the reviews from all of the different entities that they've aggregated there. It's a much more useful set of information. Now you can take -- once you have the structured data, you can do things with it, like mix it with a map. So, here's the FTC down here, and these are some restaurants, three stars and above that are in easy walking distance, then you can drill down. And you can do all sorts of wonderful things when you have structured data to mash up together in an automatic way. So, that whole thing is based on just basically these six fields here. The item reviews, six tags, the name of the items, so that would be the Capital Grill, a description, the rating, how many stars an individual reviewer gave? The name of the reviewer, the date reviewed, and then the description of the review. And from this basic data, you can do all of these wonderful things. And that is pretty simple. This is what is behind the hood, underneath the hood, in the source code. Google has a number of different ways they basically allow vendors to put it -- mark up your data in micro formats or RDFa, but it's actually very simple. These are the tags added to the data that you see. So now, the second one we've seen, is XBRL, I can't tell you how ecstatic I was when saw XBRL. My first job coming out of college, I work for the professor of entrepreneurship at the Harvard Business School, writing case studies. And when I wanted SEC data, I had to go to the library stacks, write to the SEC and get them. And then, later in the '80s, when they came out and made it available on the web, in document format. Now at PDS, "Oh,

that's the greatest thing in the world! I'll still have to cut and paste and put it together to make sense of it." But now, it's like all of this data is one giant database that you can search through, because it's tagged. It's unbelievable what that means. The one point I want to highlight on XBRL is how difficult it was, how time consuming it was to make it happen. Charlie Hoffman came up with the idea of XBRL in 1998. And it's 2009 and SEC has just started rolling out on a four-year rollout. This is a big deal. They had to form a standards body and get everybody on board. And it's not just SEC, basically, every advanced country in the world is rolling out XBRL, China, Netherlands, Israel, Singapore, Chile, you name the country, they are either already adopted in the last year or two, or planning on adopting it in the near future. This is just a crazy thing, if you told me a few years ago that something as incredibly complicated as XBRL was just in the United States, more than 14,000 tags could -- this type of thing could happen, I would say, "You're crazy." Now, this legislation ontology, we got a little hint of that in the presentation here. This is also very exciting. A few years ago, I was at the Harvard Kennedy School, the Shorenstein Center and I did a 126 legislators. How they made their information available? And I can tell you, the New York Senate was the absolute worst. My research assistant spent weeks going back and forth, just to do a basic roll call. They had a box for it but it didn't work. And it's amazing what they are proposing to do but basically what they're doing is taking a mess of data, these unstructured type of document forms, and tagging it so it can be searchable. But, the approach that is being taken in the United States is different than what's being taken by others legislators. In the United States, each legislature is doing this on their own. There's no common ontology. It's great what they're doing. It's just a huge improvement. But in Europe, in the antiparliamentary union, which is more than 150 legislators, has come up with a common standard. I can't speak about its adoption. And they have, you know, a legislative XML University there where, you know, I.T. folks can go all over and learn this standard. And that is a much more powerful way to approach this. But you know, that's a little bit too far for the Americans, for whatever reason it is. So it's a very different approach than XBRL. So now, just to summarize some of the differences with the product review ontology -- very simple, private industry, implemented on its own. It worked great. It's not very -- I don't believe it's an extensible type of approach. The financial ontology -- very complex. We have private -- you know, accountants, all of the financial players got together with the government agencies, the FDIC, the SEC, the governments of other countries to design this standard. The government incentive was very high to do this, and the private players' incentives were very strong

to do it. Legislative ontology, quite different -- sort of medium complexity, I would say purely government, and very low incentive to do this. So that's -- that's the comparison. Now, this is the core of what I wanted to get across to you, is this bios ontology, which is a way to describe conflicts of interest in a well-structured way, to revolutionize all sorts of accountability type of information. So just the basic story is -- and again, a little bit of terminology. You know, the division of labor is a source of wealth in our society. The division of labor requires that you have principles that delegate cast agents -- this is terminology used in economics, and management, political science. And when you delegate a task to an agent, you want to ensure that the agent does not have conflicts of interest, to the extent you can, or you want them to disclose it so you can take that into consideration. This is a pervasive type of relationship we have in modern society. It's becoming increasingly important as the marketplace gets more complex, or you're delegating things to elected legislation. You don't really know what they're doing, the outcomes, and you don't know where the outcomes are. The conflicts of interest become absolutely essential to the accountability system. This is why our fiduciary laws and whatnot have exploded in the last generation. So the key elements of bios ontology is first, you have to designate who has put together the agency claim. The agency claim -- agent always claims to act on the behalf of the principle, that's the agency claim. So the government can put together the agency claim. The SEC has also its agency requirements. The FTC -- basically, almost, there are literally thousands of agencies at the local, state and national level that have these agency claims. And also you can have third parties doing agency claims. In the case of the media, the journalist epic disclosures that would be an agency claim put out by the agent. So the part that I'm going to focus on here, the five major components of agent claim, you have to specify the principle, the agent, the agent's covered interest -- that's the agent's income and assets, and we'll see how this works in a second, so it's very vivid -- and the covered actions. The covered actions are what the agents do on behalf of the principle. Then these, the false settings allow you to really speed up the way it looks in practice. So now, here's an example, it's a very simple example, powerful earmarks. So here we have the agent, is Senator Shelby, who is apparently the king of earmarks in the Senate. He's the agent, the covered action is he's given money to the University of Alabama for an earmark. The covered interests are the independent contributions he's received, the PAC contributions and the lobbying there. And this language allows you to integrate them in a well-structured, very efficient way. Now, this is provided when my board members in the executive director for the Center for Responsive Politics,

you'll see their stuff all the time in "The Washington Post," "The Wall Street Journal," "Politico," you name it -- all the time when reporters want information, they tend to go at the national level, for the Center of Responsive Politics to get this type of linkage and information. So there are great advantages of this type of languages, economies of scale and application markets. We're talking about our -- today these applications are done, sort of database by database. We're talking a factor of tens of thousands, applications that will work broadly to revolutionize the economics of providing these type of applications. As we saw at XBRL, this type of ontology allows for more efficient data entry and integration. What I want to focus on is more efficient semantic search. You can do now in one query what usually would have taken tens of thousands of queries to do, even assuming if the data was available on the web -- which for many of the databases, we're not even to Web 1.0. It's not. Assuming it's there, with one simple well-structured search, you can do amazing, well-structured searches. And even though we're focusing on journalism here, which is the impact for information intermediaries, we'll see that for direct search -- for example, the FTC's blogger rules, that can have a huge impact on how those things are done. So now, the critique of the current mechanisms, like the earmark one that look nice, but actually underneath the hood, it was incredibly labor intensive -- the way that was actually done, is each member of the house posts on their home page their earmarks in a PDF-style document, something you got to go through to parse that information to allow you to structure these links, and to make those links between those databases and earmarks is a nontrivial task, it's quite complicated. So I'm talking about an increase of efficiency, of maybe a million fold, for certain types of applications, and I'm going to show you that type of application now. The earmark is not one of those, it's a relatively simple application. So here, we're going to look at the state of Arkansas, its budget, okay? And we want to do one of these agency queries. This is the type of -- to find conflicts of interest linking items in the budget to covered interest. You can specific this in a very simple way. Now, if you do this today, it would require for a medium-size city or small state -- literally millions of queries. Now it can be done instantaneously. So here's a simple -- this is what we're all familiar with, the budget, you know. You've got the line items, you've got the costs, whatever it is, this is what we have today -- no bios ontology. Suddenly, you can link all of the budget items to covered interest, just like we saw on the earmarks. But budgets tend to be a little bit more complicated. And then we can sort of do all sorts of budget manipulations. We can drill down and see the individual line items. You know, budgets can be viewed in many different ways, by program -- that was a programmatic view. By object --

you know, salary, benefits, capital expenditures and whatnot. By location -- you can see integration with geographic information systems, how all these items do it, by political district, whatever it is. You can do all of this. And once you got the data in there, in a well-structured way, you can manipulate it in incredible ways that nobody does today, because it's just -- it would be insanely difficult to do. So what I focused on here, in the description, is elected officials and voters. You know, I could have gotten one of a billion cartoons on this subject. You know, this is a big thing. But there are many, many other applications of this type of ontology here -- doctors and patients. You know, today's doctors are rife with conflicts of interest, in prescriptions and whatnot. This type of technology could make a huge difference in really user-friendly ways that you can barely imagine today. You know, journals -- all sorts of problems with writers and their undisclosed conflicts of interest. I work in a think tank world, you know, rife with hidden conflicts of interest. Real estate agents -- we've got all these highfalutin laws about conflicts of interest that everybody just ignores. You know, this would be a way to really rethink how real estate ethics works. Now here, we've got some people at the FCC that are doing a look at the public interest obligations of broadcasters. They've got a whole bunch of conflicts of interest rules. They're all implemented with Dark Ages technology. And you know, this is -- when they look about, not just how programming data can be changed, but the metadata associated with the programming data, with all the digital streams. And you know, this can be a revolution. You know, they're supposed to disclose product placements. But you know, they have barbaric type of disclosure approaches that they take to doing this. You know, so we got all these product placement types of thing, sometimes where there's laws, sometimes where there aren't. This is the one that I think is one of the most exciting areas, the FTC introduced blogger rules, effective last December, where bloggers who do reviews of products have to disclose conflicts of interest in the blogs. This got a lot of attention. I'd say it was fun for the cartoonist. I could have picked, also, many other cartoons on this one here, because it was viewed as so -- sort of controversial. So here's a very different type of display than the one I've shown you so far. So we have a document -- this is the actually the way I envision most of these would work. So here's a review of a baby stroller. The conflict of interest item, the review is linked there. Which is -- you click on it, and then you can see the conflicts of interest that are related. Now, critical thing here, is this gets the government out of the graphic design business deciding, you know, how conflicts of interest should be displayed, because the reader can choose how to view the conflict of interest. The font, the box, the highlight, whether it's

in the text, whether it's summarized in separate boxes -- all under the control of the reader. The reader does it once, for any type of log, sets the parameters, and then the display is automatic. So here, for the proposed bios ontology, compared to the other ontologies we've seen -- relatively complex, more complex than the product review, but I think quite a bit less complex than the XBRL ontology. For some of these applications, this huge -- you know, elected officials aren't going to be at all excited about it. But for certain other ones, when it comes to occupational licensing, as a new type of vehicle to manage conflicts of interest. Maybe for the FTC's blogger rules and whatnot, there I see less of a conflict of interest. And here, in some cases, you know, private entities are like government officials -- nobody wants to be made accountable. They are not that excited about a technology that is going to make them disclose all sorts of things that will reduce their market power. But it varies by the type of application. So there's some problems with relying on the market to come up with these ontologies, it varies by a product. Standards development is expensive. There are significant free-riding problems. And when it comes to ontologies, there are very significant network effects, and positive externalities associated with doing these things. Now we -- Google, you know, it has been a great solution moving this ahead in a lot of areas. But you can't rely on Google for everything. And they largely solved a lot of these coordination problems, because they can -- more than 70% of the market. When Google says, "this is the way what we do product reviews," all of the, you know, trip advisers, everybody that gets in shape and online, the posts are dated that way, they solve the coordination problem. But we don't, you know, always have a monopolist that can do that type of thing. And then a big issue, why I think there has to be a government involvement, is this real significant conflict of interest problem in implementing these. Market players don't mind vague epic statements and what not as long as they aren't enforced and that they're feel-good. But when you talking about something that makes it simple to enforce and track what's going on, that's a completely different story. So that's a reason for government involvement. And now, there's different flavors of government involvement. One type of flavor is, like we've done with the legislation, agency by agency comes up with their standard -- I don't like that approach, as I think you probably have here. Another is the government-wide, and the Department of Homeland Security and our whole criminal justice system has worked really hard to come up with a number of ontologies that work across the government. That's great. That's better than doing it agency by agency. But the gold standard for me, for this type of ontology, I think you need to have a government private partnership, because its scope is so

great, and XBRL is the case study of how you can do that. So with that, I want to just make a few caveats, conflicts of interest are subtle. You know, this is a powerful technology. But at the end of the day, you know, you can only prove correlation. You can't prove causality. All sorts of things -- there are limits to the technology. We're talking about a significant, long-term type of framework to make this happen, like XBRL. This is not the type of thing that can happen today, that we can do some really neat experiments in the short term to get a sense as to what this will do. But this is a, you know, like a chicken and the egg -- the more people that use it, the more powerful it becomes. And I think this has great implications for open government and media reform. This is not the way the open government and media reform community -- I mean, I've been to countless events like this. This is not how they frame issues. And one reason they don't frame issues like this, from my perspective, is the folks are looking for quick hits, quick results, the foundation -- anything that smacks of requiring work that's a little high tech and what not, they're scared of, the foundations don't like it. But, you know, when you want to do great things, you know, developing skills, unit capital, whatever it is, it takes a long term commitment. And this is it. And I would encourage these communities to start re-conceptualizing how they frame a lot of these issues, with standards making essential to what is done. So I'm trying to put together a standards group, we'll get the stakeholders together to do this. If you've got any thoughts, you know, I would welcome. And I have also this working paper that I'm releasing today, which largely covers what's here. But it focuses much more on the FTC's blogger conflict of interest rules, as in this presentation. It's just a vivid illustration, largely, of how this allowed us to rethink how we've conceptualized these types of public policy. So that's it. And thank you for --

[ Applause ]