

ASSOCIATION OF CORPORATE COUNSEL

Title: *Electronic Data Discovery Technology & Terminology – A Primer for In-House Counsel*

Date: July 26, 2007

Presented by ACC's Litigation, sponsored by Steptoe & Johnson

Faculty: **Mike Bergeron**, Of Counsel, Steptoe & Johnson, LLP; **Jose Ramon Gonzalez-Magaz**, Partner, Steptoe & Johnson LLP; **Stephanie Mendelsohn**, Director of Corporate Records and Electronic Discovery, Genentech, Inc.; **Bill Mooz**, Vice President, General Counsel, Catalyst Repository Systems, Inc.; **Sonya Sigler**, General Counsel, Cataphora, Inc.

(Bill Mooz): Hello everyone. My name is (Bill Mooz). I'm Vice President and General Counsel of (Catalyst Repository Systems), Inc. On behalf of the ACC litigation committee and today's panel, I'd like to welcome all of you to today's webcast on electronic data discovery, technology and terminology.

Before I introduce the program, I'd like to go through a couple of housekeeping matters. First, during the course of the program, you can ask us questions by simply typing in your question in the box in the lower left hand corner of the screen and clicking on send. We'll do our best to get all questions answered during the presentation but if we run out of time to reach all of them, we will answer them in writing and the answers will be posted on the ACC web site.

This web cast is being recorded and also will be made available on the ACC web site.

Finally, I'd like to ask everyone in advance to complete the evaluation form which is the first attachment in the box in the middle of the left hand side of the screen at the end of the presentation. This is very important to the ACC and to the speakers to ensure that we continue to deliver the highest quality presentations for your consumption.

With that, I'll introduce today's panel. Today's program is the latest in a series of programs being run by the litigation section of the ACC on what inside counsel needs to know in order to handle the requirements created by the new Federal rules of civil procedure governing the discovery of electronic evidence and the recent presentation that we had regarding the meet and confer conference is where the requirements in terms of what has to be produced and in what format get established. What we're going to talk about today is how you go about complying with those requirements in an efficient format and we are going to focus first on the processes that you need to run an electronic discovery project efficiently.

The first speaker will be (Stephanie Mendelson) who heads up corporate records and e-discovery for the legal department at Genentech, which is a leading bio technical company in the San Francisco Bay area. (Stephanie) is responsible for developing and implementing Genentech's strategies around the management of corporate records and

she also heads up the e-discovery activities within Genentech. Prior to joining Genentech, she was a litigation partner at a global law firm and has been a frequent speaker and author on topics related to records management and electronic discovery.

Then we're going to turn it over to (Jose Gonzalez Magas) and (Mike Ferguson) of the law firm of (Steptoe and Johnson) to talk about the processes that you can use to effectively review e-discovery data. Jose is a partner with the firm. His practice covers mass (fort) litigation and international arbitration. He's been involved in a number of large-scale e-discovery projects and is one of the attorneys who supervise the operations of (Steptoe's) complex litigation and discovery center. (Mike Ferguson) is of counsel with the firm and he runs the complex litigation and discovery center.

We'll then go to the second portion of the program and turn to the technology and the associated myriad of buzzwords that go with that that you will need to go and automate the way that you go through an electronic discovery process.

And then (Sonia Sigler) will lead off talking about the technologies you use for collecting, calling and analyzing data. (Sonia) is the Vice President and General Counsel at (Catafora). She spent the last five years deeply involved in electronic discovery matters as well as intellectual property and licensing issues. She's a frequent speaker on these topics and you may have heard her at last year's annual meeting of the ACC.

Then finally I will finish out talking about the technologies that you can use in the review and production phase of an electronic discovery project and as I mentioned earlier, I am Vice President and General Counsel of (Catalyst Repository Systems).

We will then close out with a session on questions and answers and as I mentioned earlier, if we are unable to get to your question during this session, we will answer it in writing and post it on the web site.

With that, I will turn it over to (Stephanie).

(Stephanie Mendelson): Thank you very much.

I'm going to start with a poll and you'll see the poll at the top of this slide. You just answer yes or no. The question is how many of you have started a data catalog where you describe your computer system and identify sources of electronic data that might be relevant to future litigation. So if you started a data catalog put yes. If not, no and then we'll get back to those results.

My message today as I talk about identification, preservation and collection processes is really to think about your pre-litigation effort. The pre-litigation data identification will help you comply with later e-discovery obligations and we'll talk about how.

Specifically we're going to talk about three things. First, what you do before litigation to identify your data sources and that's people who might have potentially relevant

electronically stored information, ESI, and then the locations for that ESI. Two, the process for preserving ESI once your preservation obligation is triggered and then three, the process for collecting ESI for review and possible production.

So thanks for responding to that poll and we'll go back and look at the results in a bit.

Turning now to our focus on how you identify data sources. Again, this pre-litigation and by pre-litigation I mean you know also before audits, before investigation. The pre-litigation data identification will help you comply with your later e-discovery obligations. Under the amended federal rules of civil procedure 26F, prior to a scheduling conference that's held under 16B, parties must discuss issues related to the disclosure of ESI including the form or forms of production and the committee noted that it may be important for counsel to become familiar with information systems before the conference. The committee further noted that in appropriate cases, individuals with special knowledge of the computer systems may be helpful. So before the roles we may have started to prepare for discovery by identifying you know information about our computer systems. Now that the rules have been amended, we really must prepare ahead of time. In this – in this circumstance though I think we really have to remember that perfect should not be the enemy of good. Once you start talking to your IT folks and telling them that you need to you know catalog your infrastructure and identify your data sources and maybe map potential custodians to those data sources or servers, they are going to laugh at you. The task is totally daunting and practically impossible if you're, you know, a very large corporation, Fortune 100, Fortune 500 certainly.

So really the approach to take is a practical one. Some practical advice is you know start big at the enterprise level. This is something that IT should be able to help you with. Describe the enterprise servers, you know how many you have, where are they located and what type they are, identify enterprise databases in categories such as accounting and finance, HR, sales, production and operations, identify corporate applications for email and calendar, identify back up procedures and media – back up media rotation cycles and then identify the most knowledgeable people about your infrastructure, about your email, applications and processes and about your back up processes, the areas that are most likely to come up in initial discovery questions. Then after you've started at this very high level with a very kind of practical and doable approach, then you can identify key business units that maybe previously or frequently are involved in litigation or investigations and for those you can start drilling down more to identify key applications and databases.

So let's go back for a second and look at the poll results and you know we have some mixed results. Not everybody responding yet but we do have some folks that are starting already to do the data catalog. All right.

So after you have done kind of your initial work, you can before any legal hold also document your preservation and collection process. Documenting your – these processes will help you improve your compliance and your ability to quickly initiate a legal hold and it will also help you demonstrate your good faith should any ESI be inadvertently

lost. Their documentation might include you know of course triggers for when a hold is initiated, who issues the hold, a template for the hold language, how the hold will be distributed and how it will be tracked. The hold documentation can also include specific contacts for suspending any recycling or automatic deletion and processes for how specific types of data like email will be or even voice mail will be preserved.

After a legal hold obligation is triggered and my slide says after a legal hold is initiated. Really we're talking about even before the initiation of hold, right at the point that your obligation is triggered, that's the point that you start identifying your custodians who may have potentially relevant documents and electronically stored information.

And then the next step is to identify their data sources. Based on your pre-litigation work, you can start by – you hopefully have information already about any enterprise applications and databases that are involved and then as you drill down through custodian interviews, you learn more about specific applications and databases they use and where they store their email and potentially relevant files.

Move on now to the actual preservation process. There are really four steps. First two, initiate the hold; two, to follow up with the custodian interviews; three, to help facilitate the hold; and then four, to document. So starting with step one, initiate the hold; commonly it's done through email now and that may or may not include confirmation of receipt. The hold of course should sufficiently describe issues involved so that employees can determine what information is relevant and the hold should also indicate that relevant information may exist in many forms including paper, email, database, et cetera, and can be found in many different locations. Your next step following up with the custodian interviews; in order to really ensure that you're preserving as you are required to and that you have information about where you're potentially relevant data exists, this step is key. Typical questions related to the preservation of ESI include what databases do you use, what applications do you use, where do you store relevant files, do you use shared directories, do you store relevant material on your desktop, do you use collaboration tools like e-room, are you frequently contributing to a blog perhaps.

The next step is that to help facilitate how this preservation is going to happen. Nothing in the case law indicates now that it's not generally sufficient to rely on employee's discretion to identify relevant material and not delete it. You're initiating the hold. You're notifying them perhaps by email of the hold and then you're following up with custodian interviews to make sure they understand it and so you know that may be sufficient good faith. However, sometimes you may feel it's necessary or maybe just less burdensome to employees to provide a repository where relevant material can be stored. For instance, IT can set up a mailbox related to the litigation for each employee and all relevant email can be moved to that mailbox for preservation and then later collection, if needed. Our technologists will caution later in this presentation that you have to beware when moving or copying data that some of the hidden data that we all now know as metadata may be impacted so you have to investigate that issue. And then step four, document, document, document. Again to demonstrate your good faith preservation

efforts, you need to be in this area very precise about the names and dates and activities related to your preservation efforts.

So I touched briefly on the kinds of things that must be preserved and what you need to notify your employees about. It's important to recognize that what gets preserved is very broad and as you work through the e-discovery process from preservation through collection and review, production and then of course ultimate presentation to the court, the universe gets increasingly smaller but this initial step is a very broad – has a very broad scope.

Beyond the possible data sources that we saw in the previous slide, you know we're seeing new data sources all the time. This slide reflects the OIP and now companies are considering adding unified messaging capability where your voice mail is converted to a digital .wav file. Unified messaging may be a particular concern for companies because the voice mail ends up in multiple locations. Since you can't control the distribution of messages like email, it's hard to track and preserve and then digitized voice mail becomes another source of network volume. The bottom line is whether this functionality is important to your business and if so, how to manage the risk. Thinking about whether you need disaster recovery backups for unified messaging, maybe you don't; about to minimize retention periods before system deletion, will help reduce the volume that's available when a legal hold obligation is triggered. And then of course once a hold is initiated, whatever exists must be preserved.

Finally, the collection process; things to think about are initially who performs the collections. Do you have internal resources for the collections or do you bring in an e-discovery provider? What's the priority of the collection? What's the scope of the collection? There's a balance to be struck between not getting more than you need to process but getting as much as you need to avoid a future and perhaps burdensome collection. Ways to limit the collection include identifying file extensions that you don't want to collect such as image files, music downloads, et cetera, deciding on date restrictions and collecting for specific custodians rather than for example collecting an entire mailbox. So again, all of the information that you gather pre litigation will help you be more specific and more compliant if you have e-discovery obligations related to preservation and collection.

With that, I will now turn the slides over to Jose.

(Jose Gonzalez Magas): Hello. Good afternoon.

We have already covered and discussed the first stage of the EDD process, of course what we need to do to be able to collect the data and we'll switch in a few minutes to how to deal with the management of those files and the electronic processing of those files but I'd like to now bring us to the latter part and last part of the process which is the review of the data. At some point in the process, it will be necessary for an attorney to take a look at files and electronic documents and make legal determinations as to responsiveness, privilege, coding for purposes of the substance of the case. Now a few

years ago at (Steptoe) we analyzed this (intuition) regarding e-discovery and realized that the new rules and the trends would actually yield such (aluminum) volume – large volume of electronic files, even in small cases, that we needed to come up with a way to actually process the review in a manner that would be both accurate and reliable but also cost efficient. So we established actually a facility dedicated only to that purpose and that allowed us then to come up with certain considerations as to what would be the cost factors that effect a document review or the review of electronic files and what would be the quality control measures that we can implement to actually allow for a more reliable review and I'd like to go through some of those major cost factors.

First, as you can see on the slide, the first one is the reviewing platform to be used. It is important that the platform be previously tested and be properly de-bugged. In our experience, this is a major cost consideration is the review or start having problems with the platform, it will effect the schedule. It will affect the budget particularly since the providers of the legal services I believe tend to be, as we are, flexible in terms of the platform to be used. Those platforms selected usually by clients must be reliable. Also we have noticed that this makes a huge difference, the native file review goes a lot faster and is a lot less costly than the image space review and this is so significant that we have the (tech that are ratio) sometimes of three to one even four to one. The complexity of the coding form makes also a substantial difference because it's directly (determinacy) upon the time the reviewer will spend on each particular file. I recommend doing objective coding electronically so that the reviewers actually already have a lot of the coding done and can concentrate on the legal assessments and the subject of more substance of coding of the documents. I would also mention that as a litigator, I tend to at the end of the case look back at what we did and determine whether the coding we performed was in fact what we needed or did we do more coding that was necessary because I realized it can be if the volume is large enough a very expensive proposition. So I would recommend that as you engage with your attorneys in this assessment you ask the question, do we really need all that coding and how can we actually narrow it perhaps. The degree of experience and specialization of the reviewers, obviously you don't want the more senior attorneys in the case reviewing the simpler emails. You want perhaps junior associates reviewing the simple materials even perhaps contractors. Temporary attorneys I know are sometimes used for that purpose and furthermore the specialization consideration is of great importance because we've noticed that a lot of the documents that we review are actually in foreign languages and we need reviewers to be able to tackle those appropriately. The training of the reviewers is also a major consideration. If you train them properly in the substance and the techniques to conduct a review, things will go a lot more smoothly and also you will be able to gain certain efficiencies in the project and that includes the preparation of a project manual. I listen to clients comment that originally they thought the training would not be as necessary but then they spend the day with the reviewers analyzing what's actually provided in terms of training and afterwards clients uniformly agree that this an appropriate thing to do under the circumstances that gains benefits.

And regarding the training let me just go through the sort of training that the reviewers need to be able to conduct the project properly. There has to be training as to the

techniques for analyzing documents, files, and data in an efficient and reliable manner. Of course, analyzing the information provided by email takes a certain methodology that would be different from the technique used for memos, for letters, for data charts and the like.

We have analyzed how this has an impact and it can be substantial. It's also important for the reviewers to understand the particular legal precepts as to the (defection) of privileged materials and what sort of privilege log will result. Often just getting the reviewers trained in that substance will accelerate their assessments considerably and will actually shorten the time needed later to deal with the privilege log which can be, as you know, hundreds if not thousands of pages long.

The coding of the data review, as we discussed, can make a substantial difference and for purposes of the (indemnification) and handling of close call data, what we do is we have a (Steptoe) associate and I think this is where law firms are going to go in the future if they're not there already to actually supervise as project managers and provide over the shoulder supervision for the reviewers so that if the reviewer has a question, if the reviewer is confused, that can be identified, can be dealt with, can actually be used as a training tool for the other reviewers and thus expedite the process of the review but it often happens that we have to actually contact you as in house counsel because you are the folks with the best position to understand your company's documents and with that guidance, we can accelerate and expedite the process even more.

The substance parameters of the review of course the reviewers must be trained in the substance of the legal matter or litigation but what I'm actually referring to here is can we narrow the parameters in such a way so that complying with the legal responsibilities the reviewers can actually factor out certain considerations of the document and thus perform the review more quickly and of course the reviewers must be trained in the use of the technology and the equipment particularly of the platform to be used.

Now this is something that I think is important to in house counsel that is to receive progress reports on a regular basis even on a daily basis. We think it's an important tool on both sides of the fence to work together to get this done quickly and appropriately and that report can provide for instance a review rate of substance. That is, are the reviewers' rates where it should be? Are they going too fast and thus we may be sacrificing reliability and accuracy or are they going too slowly and thus it may be that we're going to have a problem meeting the budget and schedule provided and speaking of budget, that's the second item? I know it's of primary importance. That progress report also provides some budget tracking stating at this point of the process, here's where we are in terms of the budget. Do we expect at this point to meet it correctly? You can also have your attorneys provide substantive reports. A couple of years ago I had a client heavily involved in the litigation and she wanted to know what were the ten top documents that were detected by the reviewers and that way she gained a sense of what is it that we were looking at, what was the importance of the documents we were looking at and provide some sort of daily analysis even in shorthand as to where the litigation would go on the basis on the documents found at that point and this progress reports also function to as

(Stephanie) mentioned document, document, document being (mineralized) the stages and developments in the project especially for those situations in which it may be necessary to show up in court and explain the methodology used for analysis during the review and how it was performed. I think this chart is useful in terms of displaying how the review rate can actually be assessed. This is based on some source information and also our personal experience at the center. As you see if you take a look at the column for the average, the one in the middle, you'll see that electronic files we have determined that generally at 600 file review rate per reviewer for an eight hour day is appropriate but that rate goes down substantially to 200 if the review involves documents that were for instance hard copies and then scanned and required the reviewers to actually input objective coding. This is of course different from the review rate for electronic which is a native format review. The number can actually go up a little and this is actually something that I recommend to clients that we actually do both. Have the electronic discovery service provider perform electronically the creation of the objective coding and then our reviewers can actually concentrate on the assessment of the documents even if they are hard copy scanned documents and then for them to input only the subjective coding and thus concentrate on the substance of the documents. I think this sort of analysis can be very helpful in your assessments as to budget considerations and schedule considerations and I would recommend that before you begin a large document review project with your attorneys you actually discuss this sort of assessment with them.

Now in terms of quality control of data review, the big fear of course in a document review is that during the electronic processing phase of the project which we're going to be discussing in a couple of minutes and during the document review, there's always a risk of that stray document not having been detected somehow and in order to avoid that risk we have developed and I think this is going to become the standard, some quality control measures to actually confirm the reliability of the review and that includes the supervision during the review, that includes an assurance that all the data was reviewed. This is particularly important when you have a large number of custodians, a large number of sources for the data to be reviewed. There has to be communication with the electronic discovery management service providers, the attorneys and in house counsel as to making sure that all responsibilities have been complied to in this way. I also recommend for the project manager to perform a random check of the coding performed every day performed by each reviewer so that if there's a problem with the coding being performed, we can address it quickly and solve it before it grows. If there's a need for (cycle) level review, that is we've reviewed the file once but for some reason something is not right about the coding and we need to review it again, we need to make sure that that was done and that was done hopefully as quickly as possible upon detection and this last item I would recommend to you, there should be a confirmation with the client that the parameters were properly applied because those parameters may change during the process of the review and again, you as in house counsel are in a particularly good position to assess how these parameters should work and how they should be approached.

And with that, I turn it over to (Sonia). Thank you so much.

(Sonia Sigler): Thank you Jose. I'm going to talk about the technology and terminology in three areas, collection, calling and analysis and talk about the technology in particular and then also any particular terminology in each of those areas that you need to be aware of.

In the first area on the actual collection of the data, (Stephanie) talked about the process in particular and when you actually go to collect the data most people do mirror images of the hard drives or servers. Some follow a self selection method where they allow people to pull documents and things off of their computers and put them in a folder on a server and then you can use a combination of the methods.

On the forensic side, they do a bit by bit copy and that's usually referred to as the mirror image. I'm seeing new data mapping tools pop up in the industry in particular two ((inaudible)) legal tech in January and in June were (Cazion) and (Deep Dive). Those are appliances that actually sit on the server and can map what they find on the server and provide an index that's searchable so that when you do go to collect the data you can actually be very specific on what you want to collect and from whom.

It looks like about a third of you from (Stephanie's) survey have started this process of mapping your data and are probably already investigating tools in this area. So those two to keep in mind as well.

On the forensic analysis, sometimes you need to figure out if data's been deleted and if it's important in your matter then you'll be talking to forensics companies to do that type of collection and you would use those also in matters where you don't find data. For example, if you're collecting data from the CEO and you're not finding things that you want to, you're going to need to do forensic analysis on that type of situation.

Going to collection philosophy and (Stephanie) touched on this a little bit in her presentation, if you want to do a narrow based collection, then you can definitely pull the data by custodian or by date range or by specific keywords if you know at the time what your matter is about or (Stephanie) mentioned the file types definitely eliminate system files, application files and not use those in your – you know not have them muddled in your collection. The other philosophy is to do a broad based collection, just collection it all, do a mirror image of the laptop or desktop or servers and then call it and I'll talk about those methods in a minute.

We're finding that corporations tend to do that. Law firms don't usually recommend it because they don't want to review all that data. There are many, many vendors out there that will call the data after you get it or do it on the front end.

When you call data that you've collected, you need to look at the goals that you have. With the change in the federal rules, you now need to list your readily accessible data as (Stephanie) mentioned in terms of mapping it and you also need to be able to list the categories of data and the locations and then finally you need to also figure out what's not reasonably accessible and list that as well and cost is a big consideration in that area. The point of using calling tools is to actually reduce the data set so when you get to review it

as Jose mentioned it you're not reviewing everything in the kitchen sink. You want to make it a manageable process. No one wants to pay to review 10 million documents so you use calling techniques to make it reasonable and then the amount of data in corporations is only growing so these calling methods actually help you get to a management set of data.

The first from a technology terminology standpoint is de-duplication methods. People mean different things when they de-dup the data. They refer to different methodologies. They use different algorithms. Vendors use different algorithms and the most common that I see used in the industry is MD5 patch values and that's just an algorithm and that's a name for the algorithm. There's also (Shaw 1) and do you really need to know what it is? Not from a technical standpoint but you do need to be aware of the methodologies and then I would definitely ask vendors if you don't understand what they're using to explain it clearly.

In the de-duplication of data sets, I've seen lots of different things happen. You can de-duplicate just among custodian in the own set of data collected for that person or across custodians or across the entire data set. There's also new technology referred to as mirror duplicates. This will refer to drafts of documents and similar drafts and the changes made within them. It will also refer to standard agreements that you just changed the customer name and something else in it so mirror duplicate technology can find all of those types of documents and separate them out so that you can review them together.

I'm just going to briefly talk about this graphic in terms of duplication rates. In a broad based collection, you may find 25 to 40 percent duplicates. If you're restoring backup tapes you'll find probably about 90 percent duplicate especially if it's weekly or daily backups. If it's daily, it's probably even a higher rate.

Let's go on to calling methodologies. There are two main methodologies here. One is linguistic methods and these all are word based and the word has to be there to find it and those are referred to as keyword search or (ontologies) and they're statistical based methods and these are all number based. So it's counting the number of times a word appears in a document or it's counting the number of times the words appear together with other words and those types are called topical clustering and you look at statistically similar documents and then counting the words and the prints together leads to latent (semantic) indexing and in the links on the left side, number 10 is a glossary. It's an information material glossary and each of these terms is defined in more detail there so that's a very helpful document to look at later.

Let's just take keyword calling because that's what most people do and most people are familiar with. There's all kinds of smaller techniques that you can use like the stemming where you use part of a word and a star to keep going and that will return other words. That's an easy use of that and people are familiar with it so they're comfortable with it. The con on keyword searching is that it can be over inclusive leading you to have to (disambiguate) words and that means if you're referring to George Bush and to (disambiguate) that you would figure out if it's George Bush Senior or George Bush

Junior the second and if it's the president or someone else. So that's called (disambiguating) when you figure out what it's referring to even though it's using the same word.

Keywords are under inclusive in the sense that the word actually has to be there to find it and it cannot retrieve documents that use a different word to refer to the same concept and keyword can be ineffective when used with short messages or IMs that refer to things as it or with other pronouns. So people are starting to use other methodologies to call the data and concentrate on reviewing a smaller subset of data.

So linguistic methods, (ontologies) are groups of related things and it's all based on words so you basically have a thesaurus that you build and if you look in the information retrieval glossary it gives an example of aircraft and if you find the word Boeing or 747 or (Sethna) it will all retrieve those documents under that (ontology) and those (ontologies) can be used to automate the review and Jose referred to this briefly in terms of being able to pre-code everything and in automated review, all of your documents come back with a categorization as responsive or non responsive based on this technology use. It's all referred to as technology assisted review.

On the statistical methods, there's topical clustering and this would find and the example I used there with Royalty, and Disney and (Hi), the hosting in the same document would come up as statistically similar. Latent (semantic) indexing, it would be the technology that's finding more like this document as using latent (semantic) indexing where you use words next to each other or in the same document and if they occur enough times then it's likely to be the same concept theme spoken about.

So last thing I'm going to talk about is the technology and terminology related to the analysis of the data and this can be done up front if you're trying to figure out which custodians to pull data from or what file types there are or the volume of data that's been collected and it could also be used on the back end or during review when you're trying to figure out the substance of the data and locate specific evidence for exhibits and the slides that I'm going to show here are just examples. There are hundreds of analysis methods. The first one is the closest analysis and that is this particular graph is showing the top five closest actors to the particular custodians and it's filing all of the other identities for those particular custodians. This particular analysis is showing who is communicating with who in this particular email set and it's mapped by color for the group that it's talking about so if you're looking at what the legal group is talking about, you can focus in on that with more specificity.

The mind share analysis is looking at the topics that people are talking about in email or on their hard drives and it maps them by topics so you can see that help and save the environment and internal audit work, the purple topics that were of concern were being talked about a lot.

If you have a question you can post it in the left corner in the bottom and just type it in there and hit send and we can address that as we go along.

The next one and this is just an example of tone detection using the words and the emotional tone of people and capitalization in the emails to figure out if the tone is neutral or angry. So these are just examples of the analysis that can be done in the beginning or in the middle or end of the review or the collection phase to figure out what you need – what is there in the data and finally I just want to say don't be afraid to ask vendors about their technology and what exactly they use and then you can use various resources to educate yourself and one of the other links is to a resource list as well.

I'm going to turn it over to Bill.

(Bill Mooz): Thank you (Sonia). So one of the things that I think everybody on this call is probably painfully aware of is that the cost of reviewing documents in discovery is rapidly becoming the number one cost item on most law departments or many law departments budgets and a net result of that is we as in house counsel are facing two pressures. There's the obvious pressure to make sure that you comply with the rules that you do an effective review. Don't let privileged documents slip through the net but there's the added pressure of making sure that you control costs as much as possible and the key to controlling the cost in the review phase as has been alluded to by a number of the other presenters is to have a good plan and you really need to spend some time in getting your plan laid out, putting in place the processes for the review like Jose talked about and making sure that you have the technology that's going to go and enable you to execute on that plan very effectively and the plan for review begins with many of the things that (Sonia) covered here in terms of collecting and calling. You want to collect a reasonable amount of documents so you're not starting out behind the eight ball. You want to call the documents in an intelligent fashion and in fact, the calling phase and the review phase are increasingly getting merged. We had a matter recently where we were working with the clients in response to a second request from the FTC. They started out with over a (terabyte) of data and through some sophisticated key terms searching and (ontologies), we were able to produce about 64 percent of the documents that ultimately ended up being produced without even having further linear review, having human review of the documents. So you can cut your costs significantly if even before you do the review you pay attention to the earlier phases that we've talked about earlier.

Now when you come into the review phase itself, as Jose alluded to, one of the key questions is what are you going to use in the way of a review platform and a review platform is simply a software system that's going to store the data that's being reviewed and enable to the reviewers to get through it as quickly as possible. There's a lot of tools out on the market right now. they differ dramatically in terms of age and currency into what they have but with respect to any of them before you get to features, you're going to need to make a core decision of whether you want to acquire enterprise software which is software that you run in house on your own hardware with your own IT department or whether you want to use a hosted solution that's run either within your law firm or within a service provider. You may not have a choice in this matter if you're dealing with tight deadlines and it's not really an option to go and get IT involved and get a set of software installed, acquired, up and running, get people trained and such, you may have no choice

but to go to the hosted software. When you go to hosted software, what you have to look at next is how are you going to access the system. The simplest way which is something that's offered by most of the new systems is web based access. All you need a browser. You type in the URL. You go to the log in page. You have your credentials. You log on and you have complete access to the application as though it's running on your desktop. Some of the older systems use what are called terminal services like (Citrix) where they set up a software lair that in essence enables you to access remotely into the server. Those typically require that you load certain software on each user's computer and it will take a little bit more involvement from IT to pull that off.

So once you figure out what tool you're going to use, you need to consider what your plan is for the data involved and as Jose mentioned earlier, the current standard is to review the data in a native format and what that does is it enables you to avoid the requirement that prevailed for a number of years with some of the older technology that you convert the documents from say Word format into an image format like TIF or PDF. That's both expensive and it takes quite a bit of time. So you want to make sure that your technology gives you the ability to do a native review.

The second thing that you need to think about is what is the metadata that you need to extract from the document. As Jose indicated, if you can go and using technology have the significant metadata field extracted from the document for you, that will speed up the review considerably so it's things like what date was the document created, who was the author, who was the recipient. You want to be selective in how you do that. We recently saw a situation where somebody was proposing to pull out more than 200 metadata fields many of which were of fairly limited relevance and that was going to cause the cost and the time associated with the review to go up significantly. Then you will look at what does it take to process the data and whether you're using a native review or a review in something like TIF, you will have to process the data to get it into a format that it can go into the review machine and give you all of the extracted data that you want to work with as part of the coding process.

Then the final thing you have to think about in terms of data is just how are you going to get it into the system and you can transfer stuff over the wires using file transfer protocols if you have small volumes of data, say two gigabytes in size or less but if you have a lot of data, you're going to be mailing either CDs, DVDs or hard drives and you need to keep that time requirement in mind as you go and you set what your process is and what your expectations are for turnaround.

The next thing that you need to look at in putting your plan together are work flows and there's really two things that you're looking at here. The first is who's going to do what work and the second is what are you going to give them in the way of data and tools to do it and this is probably the area where in the review process you have the greatest opportunities to save money. The first thing you need to look at which is something that Jose alluded to earlier is who are you going to use to do the review and it's increasingly common that you use some form of contract attorney. They may be either on shore or off shore to do the first pass review in large document cases and then you have the higher

priced trial attorneys making the more critical determinations such as final privilege reviews, obviously what exhibits are going to be introduced at trial and things of that sort. In order to do this, you need to make sure that your technology has the features that will enable you to break up the tasks into these different phases and assign them effectively and I'll talk about those in just a minute here.

The second thing you need to look at is what I call batching which is how are you going to organize your documents for assignment out to review teams so those review teams can be as effective as possible and there are a number of studies out there looking at this. The most aggressive ones that I've seen indicate that you can cut your costs on the review phase by as high as 10 times if you go and you organize your documents in logical batches around common themes. Now the theme could be a custodian. It could be a subject matter. It could be a date time, anything like that but the theory is that if you group the documents in this fashion, you are more likely to be able to make determinations in bulk that apply to an entire file of documents for example looking at a file of invoices and saying these are clearly irrelevant. I'm going to code them all as irrelevant at once without looking at each individual document or when you have to look through each individual document, that you can get through it more quickly because you're familiar with the subject matter that's there. There are a couple of different ways that you can organize and (Sonia) talked about these. You can use equations within the system to do it for you which is known as concept clustering or you can go and put together searches called key term searches or fielded searches to go and organize them yourselves.

The next key piece of technology that's required to go and organize your review are folders and forms and what this is really about is where you're going to put things and what you let people do. Folders, it's a pretty simple concept just like dealing with paper. You create a folder of documents that you give people access to. That's the documents that they can see. You want to make sure that security rights and access rights are administered at that level so people see what they're supposed to see and don't see what they're not supposed to see.

The second thing that's key here are review forms and this is especially important if you're using different attorneys to do the first pass, the second pass and the third pass and what you want to do is give people the opportunity to go and enter information in a fixed predictable fashion through things like check boxes that are very efficient and you don't want to clutter them up with things that they don't need to see but you want to give them the things that they do need to see. A second thing that's important here are the fields. If you are going to be sharing a repository among a joint defense group or something like that, you want to make sure that you can have private fields where you can put your comments in, in a fashion that nobody else can see them.

When you finally get to the review itself, you're going to be looking for a couple key pieces of technology. One I mentioned earlier is the ability to bulk tag documents, in other words to code an entire folder of documents all at once if that's appropriate as opposed to having to go hand in hand or document by document and the second thing is

as Jose alluded to earlier, your cost is going to go up considerably if you have poor quality. You want to make sure that you have reviewer level audit trails so if you see something with an individual reviewer you can follow it through and correct it very quickly.

Just a couple seconds on multi language reviews, there are three big (gotchas) that occur when you have documents in languages other than English. The first is typically or historically documents were processed into an ask key format which was a computer format designed for English. It doesn't have enough code to handle languages with lots of characters like Kanji and if you don't process into UTF8 which is the new standard, you're going to get errors in your searches.

The second thing is you need to be able to tokenize languages like Chinese, Japanese and Korean where you don't have spaces between the individual characters. The program that you're using, the review tool, needs to be able to go and put those spaces in the appropriate places and then finally you need to make sure that your own computers or those of the reviewers are up to date to handle the current languages and you do that by installing multi language packs and I've given you the URL as to where you can download those.

With respect to productions, the historical way of doing it was either to produce in paper or to convert native files into PDF or TIF. Both of those methods are very expensive and are out of favor particularly under the new federal rules. The common way to produce these days is increasingly in a native document format or even to just set up a private sub collection or a private site on the same review tool that you're using and just to give your opponent access to that portion knowing that they're walled off from anything else.

So with that I will turn it over to (Sonia) for questions and answers.

(Sonia Sigler): OK Bill. The first question is going to go to you. Thoughts on archiving data.

(Bill Mooz): Well OK. The key question I think that you have to ask in archiving data is what is the likelihood that the data is going to be needed again and if you are in a situation where there's potentially an appeal being filed or it's part of serial litigation where the same issues may come up with the same documents the same custodians in another case, it probably makes economic sense to archive it. If there's little likelihood that it's ever going to be used again, you're probably better off just archiving it on disk or even deleting it.

(Sonia Sigler): OK. Thank you.

(Stephanie) I have a question for you. Is there a resource for identifying the necessary fields for a data catalog or the IT map?

(Stephanie Mendelson): In a word no and I'll host a written response but the case law has not really developed yet on that point and the committee notes related to the amended rules

of federal procedure discuss the need to gather system information but they are also not specific. You know the main point is that no two data maps will be the same so as I mentioned during my presentation you know what we need to take is a practical approach and you know you do the best you can and take an approach that makes sense in the context of your organization. The content of your data map will depend on your particular computer system and infrastructure so in a highly decentralized environment the very best that you can do may be to identify the type of high level information that I've reviewed in the presentation. If you have a more centralized system and a less sophisticated infrastructure, then you may well be able to be more specific in those circumstances. You may be able to drill down to identify potential custodians who will have relevant information in a particular litigation such as litigation related to a particular product and then you might be able to identify the applications they use and where data from those applications is stored on your servers but in most cases you'll need to start and possibly stay fairly high level and ultimately you – the approach is just to prioritize and peel the onion. The best – while there aren't any written kind of standards at this point, taking advantages of corporate counsel, benchmarking opportunities and resources like ACC is a really good use of those resources in connection with a topic like this.

(Sonia Sigler): Thank you. There's a question on here of what's meant by coding. There are two things meant by this. One is it's referring to the identification and documentation of either objective or subjective information about the document and this can include author, recipient, the to, the from, the BCC field, the subject, all of that type of information can be coded objectively and this is usually done when you have converted paper documents into TIF or PDF. When you use native data, that's already in there because it's most of the metadata fields that can be extracted when you index the data. The other thing it may refer to and it's also called categorization of the data is to pre categorize it all as in some fashion as responsive to a particular issue or non responsive or potentially privileged so that's a way to think about coding data to data.

There's one more question down here I want to answer. Somebody asked us doesn't native data risk impacting the metadata. The answer is no because you're working on a copy of the data and when you go to produce it you go back to the original to actually copy and produce that one and when you review the native data it's only in read only fashion so it does not get changed.

We are running out of time. I would like to thank the ACC and (Step toe and Johnson) for sponsoring this web cast and I'd like to thank each panelist, (Stephanie) from Genentech, (Jose Gonzalez Magaz and Mike Feron) from (Step toe and Johnson), and (Bill Mooz) from Catalyst. Thank you very much.

END