

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

IN THE UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

KLEEN PRODUCTS, LLC, et al.,) No. 10 C 5711
)
Plaintiffs,)
)
v.)
)
PACKAGING CORPORATION OF AMERICA,) February 21, 2012
et al.,) Chicago, Illinois
) 9:00 a.m.
Defendants.) Evidentiary Hearing

TRANSCRIPT OF PROCEEDINGS
BEFORE THE HONORABLE MAGISTRATE JUDGE NAN R. NOLAN

APPEARANCES:

For the Plaintiffs: FREED KANNER LONDON & MILLEN LLC
2201 Waukegan Road
Suite 130
Bannockburn, Illinois 60015
BY: MR. MICHAEL J. FREED
MR. ROBERT J. WOZNIAK

THE MOGIN LAW FIRM, P.C.
707 Broadway
Suite 1000
San Diego, California 92101
BY: MR. DANIEL J. MOGIN

SCOTT & SCOTT LLP
707 Broadway
San Diego, California 92101
BY: MR. WALTER W. NOSS

TRACEY DANA McCULLOUGH, CSR, RPR
Official Court Reporter
219 South Dearborn Street
Room 1426
Chicago, Illinois 60604
(312) 922-3716

1 APPEARANCES CONTINUED:

2 For Defendant Packaging
3 Corporation of America:

KIRKLAND & ELLIS LLP
300 North LaSalle Street
Chicago, Illinois 60654
BY: MR. BARACK S. ECHOLS
MR. LEONID FELLER

5 For Defendant
6 International Paper:

FOLEY & LARDNER LLP
777 East Wisconsin Avenue
Milwaukee, Wisconsin 53202
BY: MR. JAMES T. McKEOWN

7

8

9

10

11 For Defendant
12 Temple-Inland:

FOLEY & LARDNER LLP
321 North Clark Street
Suite 2800
Chicago, Illinois 53202
BY: MS. JOANNE LEE

12

13

14 For Defendants
15 Cascades and Norampac:

MAYER BROWN LLP
71 South Wacker Drive
Chicago, Illinois 60606
BY: MR. ANDREW S. MAROVITZ
MS. BRITT M. MILLER

15

16 For Defendant
17 Georgia Pacific:

K & L GATES LLP
70 West Madison Street
Chicago, Illinois 60602
BY: MR. SCOTT M. MENDEL

QUINN EMANUEL URQUHART &
SULLIVAN LLP
51 Madison Avenue
22nd Floor
New York, New York 10010
BY: MR. STEPHEN R. NEUWIRTH

18

19

20

21

22 For Defendant
23 RockTenn CP, LLC:

FIGLIULO & SILVERMAN
Ten South LaSalle Street
Suite 3600
Chicago, Illinois 60603
BY: MR. JAMES R. FIGLIULO

24

25

WINSTON & STRAWN
227 West Monroe Street
Suite 4400
Chicago, Illinois 60606-5096
BY: MR. R. MARK McCAREINS
MR. JOSEPH L. SIDERS

1 APPEARANCES CONTINUED:

2 For Defendant
3 Weyerhaeuser Company:

McDERMOTT WILL & EMERY LLP
227 West Monroe Street
Suite 4400
Chicago, Illinois 60606-5096
BY: MS. RACHAEL V. LEWIS

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

1 THE COURT: Good morning, everyone.

2 MR. MAROVITZ: Good morning, Your Honor.

3 MR. MOGIN: Good morning, Your Honor.

4 THE COURT: Okay. So Lynette will call our case.

5 THE CLERK: 10 C 5711, Kleen Products versus
6 Packaging Corporation.

7 THE COURT: Okay. Good morning, ladies and
8 gentlemen. We're here for an evidentiary hearing, and this
9 case has been referred by Judge Shadur. So will the
10 plaintiffs' lawyers identify themselves, please.

11 MR. MOGIN: Good morning, Your Honor. Dan Mogin on
12 behalf of the plaintiffs.

13 THE COURT: Hi, Mr. Mogin. Welcome to Chicago.

14 MR. MOGIN: Thank you very much.

15 MR. FREED: Michael Freed, Your Honor. Interim
16 co-lead counsel on behalf of the direct purchaser plaintiffs.

17 THE COURT: Thanks, Mr. Freed.

18 MR. NOSS: Walter Noss, Your Honor, on behalf of
19 plaintiffs.

20 THE COURT: Hi, sir.

21 MR. WOZNIAK: Robert Wozniak on behalf of the
22 plaintiffs.

23 THE COURT: Good morning, Mr. Wozniak. And for
24 defendants will you identify -- we have seven defendants. So
25 will each of the defendants -- I'm going to see if I can keep

1 people straight here -- introduce yourself and what client you
2 represent, please.

3 MR. MAROVITZ: Good morning, Your Honor. Andy
4 Marovitz for Temple-Inland. I'm here with Britt Miller as
5 well.

6 THE COURT: Thank you.

7 MR. McKEOWN: Good morning, Your Honor. James
8 McKeown for International Paper, and my colleague Joanne Lee is
9 with me as well.

10 THE COURT: And how do you spell Miss Bee's last
11 name?

12 MR. McKEOWN: Lee, L-E-E.

13 THE COURT: Miss Lee. And you are for? I'm sorry,
14 Mr. McKeown.

15 MR. McKEOWN: International Paper.

16 THE COURT: International Paper. Okay. Our third
17 defendant. Good morning, sir.

18 MR. MENDEL: Scott Mendel for Cascades and Norampac.

19 THE COURT: Okay. Thank you.

20 MR. NEUWIRTH: Good morning, Your Honor. Stephen
21 Neuwirth for defendant Georgia Pacific. And I'm also joined
22 here today by Jim Figliulo.

23 THE COURT: Okay. Thank you. And next.

24 MR. McCAREINS: Good morning, Your Honor. Mark
25 McCareins on behalf of Rock Tenn. And I have to apologize. At

1 11:30 I have to go to another matter before Judge Cole, and
2 I'll be turning over the baton on behalf of Rock Tenn to Joe
3 Siders.

4 THE COURT: Okay. You're leaving your case in good
5 hands then.

6 MR. McCAREINS: Completely.

7 THE COURT: Okay. Mr. Siders. Will you return to
8 us, or are you going to -- just going for a status?

9 MR. McCAREINS: I hope to after probably you're done
10 with lunch.

11 THE COURT: Okay. Thank you. Good morning, sir.

12 MR. ECHOLS: Good morning, Your Honor. Barack Echols
13 on behalf of Packaging Corporation of America. And I have with
14 me my colleague Leonid Feller.

15 THE COURT: Okay. Mr. Feller, thank you.

16 MS. LEWIS: Good morning, Your Honor. Rachel Lewis
17 on behalf of Weyerhaeuser Company.

18 THE COURT: Okay. Do you need to be at the table?

19 MS. LEWIS: No.

20 THE COURT: Are you sure?

21 MS. LEWIS: Oh, do I? No, Your Honor.

22 THE COURT: Okay. Yes. Who's next? That's it.

23 Okay. Good. Okay. And you all met our court reporter Tracey
24 McCullough.

25 Are there any pretrial matters or any prehearing

1 matters anyone wishes to bring up?

2 MR. FREED: Your Honor, Michael Freed. I believe
3 this Friday we filed a motion for partial reconsideration of
4 Your Honor excluding the testimony of Miss Tenny. We're
5 perfectly prepared to wait for a ruling on that depending upon
6 what evidence defendants put into the record. But if Your
7 Honor is prepared to rule on that, that is before you.

8 THE COURT: Okay. So I think the record wise -- just
9 one moment.

10 (Brief pause.)

11 THE COURT: So since our last status at which we
12 decided or I ordered that today's hearing was going to proceed,
13 the defendants volunteered to go forward this morning. We
14 divided the time 9 to 12:30 for the defendants, an hour break
15 for lunch, 1:30 to 5 o'clock for the plaintiffs. We also, we
16 also -- and I guess I just want to stress again that I think we
17 have so much to do within the eight hours that all lawyer
18 arguments, I am going to be available for argument some other
19 day. Not today. And the scope of the referral from Judge
20 Shadur is for all discovery.

21 So we -- we're going to concentrate our efforts, I
22 think it is fair to say on the issue before the Court, is
23 search method. There are three other issues. And I suggested,
24 and it was really a strong suggestion that if your other issues
25 that were mentioned in the brief relate to the expert who's on

1 the stand, you may want to ask a couple questions. But nothing
2 is going to be precluded, including Miss Tenny is her name,
3 right, the linguistics lady?

4 MR. FREED: Yes.

5 THE COURT: Okay.

6 MR. FREED: I apologize for interrupting, but I
7 neglected to advise you she is here and available to testify
8 today should it be determined that she can.

9 THE COURT: Well, here's what my thinking was: At
10 the time we set today's schedule each of you had the five
11 people who are scheduled to testify today. You had their
12 resumes. You had -- I think you've even exchanged reports with
13 each other. And what happened with Miss Tenny, who is a
14 linguist, I'm not saying she may not -- we may need her
15 testimony. But I think in fairness there wasn't enough notice
16 for today. And I'm not precluding, I'm not precluding
17 anything. Okay. I just think we have our plate full on the
18 eight hours today. And my ruling was that she's not testifying
19 today. That's all. Okay. So if she does want to leave.

20 Now, here's my next suggestion. When I was on your
21 side of the podium and we had experts, the normal procedure is
22 that experts are not excluded from the courtroom. Because if
23 at the end of the day I have questions, you have questions, I
24 think they should hear the other people's testimony. Does
25 anybody have an objection to that?

1 MR. MOGIN: No objection by the plaintiffs, Your
2 Honor.

3 THE COURT: How about for the defendants?

4 MR. McKEOWN: No objection for the defendants, Your
5 Honor.

6 THE COURT: Okay. And are your experts intending to
7 stay all day? That was kind of a scheduling question I had.

8 MR. MAROVITZ: Ours are, Judge.

9 THE COURT: How about yours? Well, I guess they're
10 on this afternoon, so they better be, right.

11 MR. MOGIN: Yes.

12 THE COURT: Well, that's good. Okay. Because then
13 because the -- when we designate a time for the plaintiffs and
14 the defendants that's to do both direct and cross. So it would
15 be good if everybody were here if we have any cleanup at the
16 end. Okay.

17 So plaintiffs, you want to call your first witness,
18 please.

19 MR. MOGIN: The defendants, Your Honor.

20 THE COURT: I'm sorry. The defendants, right. Doing
21 things a little differently. Hi, sir.

22 MR. NEUWIRTH: Thank you, Your Honor. Again this is
23 Stephen Neuwirth.

24 THE COURT: Thank you for saying your name again too.

25 MR. NEUWIRTH: Thank you, Your Honor. And the first

1 witness called by the defendants is Mr. Kenneth Koch, K-O-C-H.

2 THE COURT: Okay.

3 MR. MOGIN: Your Honor, actually before Mr. Koch does
4 testify.

5 THE COURT: Yes.

6 MR. MOGIN: The Court may recall that last week,
7 perhaps Thursday, the plaintiffs filed a memorandum regarding
8 the testimony by experts invading the legal matters, testifying
9 about legal issues.

10 THE COURT: Right.

11 MR. MOGIN: And part of that was addressed to the
12 description, the very brief description that we received of
13 Mr. -- I think Mr. Koch.

14 MR. NEUWIRTH: Yes.

15 MR. MOGIN: Mr. Koch's testimony, where he
16 indicated -- where Mr. Neuwirth indicated that Mr. Koch would
17 be testifying about what we would consider to be legal matters.
18 I don't know if the Court has had an opportunity to take a look
19 at those pleadings.

20 MR. NEUWIRTH: Your Honor, I would just say that I
21 don't believe anything Mr. Koch is going to be testifying to
22 today constitutes a legal matter by any stretch. And certainly
23 if objections need to be made during the testimony, they can.
24 But I think in advance of the testimony it would be premature
25 for the plaintiffs to assert that Mr. Koch is going to be

1 testifying to legal matters.

2 THE COURT: Don't you think this whole hearing is
3 about the interplay between law and science? I mean in a way.

4 MR. NEUWIRTH: Well, certainly there is an issue of
5 interplay between law and science, and we are looking forward
6 to Your Honor's ruling on how the law applies in this context.
7 But I believe all of the witnesses for the defendants are going
8 to be testifying about factual issues that will bear on your
9 legal ruling.

10 THE COURT: So I think maybe what we should do, I
11 don't know how to do a pre -- this is kind of a -- this whole
12 area is a new area here in a way. So why don't you -- you have
13 a right to make an objection. So you've got your -- you've got
14 your record made. And I think we should take it all, let them
15 testify. And if we need to strike anything based upon that
16 type of objection, we can certainly do it afterwards. Okay.

17 MR. MOGIN: Very good, Your Honor.

18 THE COURT: Thank you.

19 MR. MOGIN: And in that regard since we're dealing
20 primarily with experts today --

21 THE COURT: Right.

22 MR. MOGIN: -- do we want to have a mild suspension
23 of the rule against leading questions so we can move this
24 along?

25 THE COURT: Yes, I wanted to -- in fact, I wanted to

1 do that, and you are all preserving Daubert challenges for
2 trial or for any other period here. Today is an evidentiary
3 hearing that is educational and investigatory and helpful to
4 ruling on -- I don't think I even have a motion in front of me,
5 though. Or do I have a motion? Is there actually a motion?
6 No.

7 MR. MOGIN: You do not, Your Honor.

8 THE COURT: So I am definitely in new territory here.
9 Okay.

10 MR. MOGIN: Very good.

11 THE COURT: You're not waiving anything. Okay.

12 MR. MOGIN: Thank you. And in light of that we won't
13 request to voir dire a witness. We'll just reserve that.

14 THE COURT: Right. Thank you. Okay. Okay, Mr.
15 Neuwirth.

16 MR. NEUWIRTH: Thank you, Your Honor. Does Your
17 Honor plan to have a procedure for swearing in the witness, or
18 can we just proceed?

19 THE COURT: Oh. No, we're not that casual.

20 (Laughter.)

21 THE COURT: Yes. The oath hasn't left us.

22 KENNETH KOCH, DEFENDANTS' WITNESS, DULY SWORN

23 DIRECT EXAMINATION

24 BY MR. NEUWIRTH:

25 Q Can you please state your full name for the record.

1 A Kenneth Charles Koch.

2 Q Mr. Koch, where are you currently employed?

3 A At KPMG.

4 Q And how long have you been at KPMG?

5 A I've been at KPMG for nine years.

6 Q And what is your current position there?

7 A I lead our forensics technology services practice in the
8 southeast U.S.

9 Q Now, you said you've been at KPMG for about nine years.
10 Can you generally describe your employment history at KPMG?

11 A Sure. We provide computer forensic and electronic
12 discovery services for our clients in matters such as
13 litigation, regulatory inquiries, and internal investigations.

14 Q And has that been the focus of your work over those nine
15 years?

16 A It has.

17 Q Tell me what you did before you joined KPMG nine years
18 ago.

19 A Prior to joining KPMG, I was in the United States Air
20 Force for 10 years.

21 Q And what was your last position in the Air Force prior to
22 working for KPMG?

23 A My last position was a special agent and computer crime
24 investigator with the Air Force Office of Special
25 Investigations.

1 Q And how long were you at the Office of Special
2 Investigations?

3 A For the last four years of my time in the Air Force.

4 Q And what was the nature of the work that you did in the
5 Office of Special Investigations during that period?

6 A I ran criminal and counterintelligence investigations.

7 Q And did that involve computers and electronically stored
8 information?

9 A It did. I specialized in computer crime investigations,
10 and we did primarily work in computer forensics area.

11 Q Okay. And briefly what is your educational background?

12 A I have an undergraduate degree in management information
13 systems from Colorado Christian University.

14 Q And in connection with the work you do on E discovery at
15 KPMG, do you take any steps to stay current on developments in
16 the area of ESI discovery?

17 A I do. I try to keep up on my reading, and I attend Legal
18 Tech. I attend the Georgetown Advanced E Discovery Institute
19 in November and various other conferences that may come up.

20 Q Do you have any connection to the Sedona Conference?

21 A I do. I'm a member of the Sedona Conference.

22 Q Now, at KPMG are there any steps that are typically taken
23 in matters involving discovery of ESI?

24 A Yes. Typically based on my experience we would go out and
25 identify where potentially relevant ESI might exist in the

1 client's enterprise. And it's usually based on a list of
2 custodians that's provided to us by counsel or the client. And
3 based on that list of custodians we might try to go out and
4 figure out, you know, where they store -- where they have the
5 capability to save documents and information. And then, and
6 then gather the information that has been identified as needing
7 to be gathered.

8 Once it's gathered we would put it through a process
9 to take, you know, the vast amounts of ESI and try to cull it
10 down to something a little bit more meaningful that you would
11 then prepare to set up in a repository where counsel could come
12 in and review the documents and ultimately make their
13 productions.

14 Q So is it correct that you just identified five basic
15 steps? Identifying where the ESI is located, potentially
16 relevant ESI, collecting it, having a process to narrow it down
17 from the large set to a smaller set which can then be reviewed
18 by attorneys and then ultimately produced?

19 A That's right.

20 Q Okay. And you mentioned at the start when you talked
21 about this first step of identification custodians. Can you
22 just flesh out what that meant.

23 A Sure. Usually it's -- when we're handed a custodian list,
24 it's usually, you know, key employees at the company that
25 counsel and/or the client think would be most likely to have

1 information that's responsive to whatever document request is
2 out there.

3 Q Okay. And the second step you mentioned was this
4 collection step. What happens in that step?

5 A Well, for all the data that's been identified, you would
6 go out and gather that data and make copies of it and collect
7 it and ultimately start to prepare it to put into a repository.
8 So you might apply some processes to it. You might apply some
9 filtering to it, including things like, you know, file type
10 filters to say if you want to see all the e-mail and office
11 type documents for this matter.

12 And if you have a certain date range that you can
13 work with, we might apply date filters to that information too.
14 And if you have certain keywords that you know that you want to
15 look for to try to cull down that set a little bit, we might
16 apply those keywords. And so we take a very large amount of
17 ESI and then cut it down to something a little bit more
18 meaningful prior to having folks start the review.

19 Q So it sounds like you talked about the second and third
20 step. The second step is the collection, is that correct?

21 THE COURT: We have an objection.

22 MR. MOGIN: I'm sorry, Your Honor. It's not an
23 objection. I'm just having a little bit of trouble hearing the
24 witness.

25 THE WITNESS: I'm sorry.

1 MR. MOGIN: If I could just ask that that last part
2 of that be repeated, the last part of his answer.

3 THE COURT: So, Miss McCullough, what's your
4 preference? Miss McCullough, can you read that back.

5 (Record read.)

6 BY MR. NEUWIRTH:

7 Q So it sounds like you were talking about there the second
8 and the third step. And this second step, the collection, is
9 it correct that you take the ESI from these custodians and
10 literally collect it and put it in a location where these
11 processes you described like filters and keywords can be
12 applied in the third step; is that correct?

13 A That's correct.

14 Q And when the ESI is collected, is it loaded into some sort
15 of platform for doing the processing you described?

16 A It is. That's generally how you would set it up for some
17 sort of content review. You'd have to upload it into some sort
18 of tool, some sort of platform.

19 Q And in this third processing stage you mentioned that you
20 might apply date filters. You mentioned other steps you might
21 take to take this broad set of ESI and narrow it down to
22 something more meaningful, I think was your term. And one of
23 the things you mentioned was keywords.

24 What were you referring to when you said keywords?

25 A Search terms that have been developed to identify things

1 that are likely to be responsive to the issues that are under
2 request.

3 Q Are you familiar with something called deNISTing?

4 A Yes.

5 Q And what is deNISTing?

6 A DeNISTing is basically applying a filter to the data to
7 remove any system type files that might exist there, things
8 that are generally not responsive.

9 Q And what is the reason for these various types of
10 processing steps you mentioned that you would do in this third
11 phase?

12 A Well, usually when you have a broad base collection, you
13 end up with quite a bit of ESI that you would then want to cull
14 down into something a little bit more meaningful before you
15 then put it into a repository and have folks start to go
16 through and review the information, because you don't want them
17 reviewing things that are garbage and spending a lot of money
18 there.

19 Q And when you say garbage, what are you referring to?

20 A Things that aren't responsive, just -- yes.

21 Q And you then mentioned a fourth stage, which is once
22 you've narrowed down the set of ESI, I think you said it's then
23 actually reviewed by attorneys; is that correct?

24 A That's correct.

25 Q And then the fifth stage you mentioned was from this

1 review there's a set of documents to be produced?

2 A Yes.

3 Q Okay. So let's turn now to this case. How did KPMG
4 become involved in this litigation?

5 A We were contacted in May of 2011 by Counsel On Call with
6 an opportunity to put in a bid at Georgia Pacific to help them
7 with their processing. They were looking for a processing and
8 hosting vendor and also someone to help them with the search
9 capabilities. So we were able to put together a bid, and
10 ultimately we were engaged in that same month.

11 Q Now, you mentioned that you were contacted by Counsel On
12 Call. What is Counsel On Call?

13 A Counsel On Call is a firm that provides attorneys, you
14 know, as needed to corporate clients, and they also have an E
15 discovery focus where they have a big E discovery review teams
16 and things like that.

17 Q Is Counsel On Call a company that KPMG has worked with
18 prior to this matter?

19 A Yes, we have worked with them for at least the last year,
20 and we've got several matters that we worked together with them
21 on across multiple clients.

22 Q Okay. Now, you mentioned that you put in a bid. When was
23 that bid submitted in this case?

24 A In May of 2011.

25 Q Okay. And just for clarification, I know you had

1 mentioned this briefly, but what was it that this bid that you
2 put in May 2011 covered?

3 A Georgia Pacific was looking for a vendor to help them with
4 processing and hosting and search, searching the data. So that
5 was the bid that we put in.

6 Q And is it fair to say that that was a bid that related to
7 the first three steps you mentioned, the identification,
8 collection, and processing of electronically stored
9 information?

10 A No. The data had already been collected. It was just to
11 process and search and host the repository for the information.

12 Q Okay. So this was a bid related to the process of taking
13 a broad amount of ESI and narrowing it down to a smaller set to
14 be reviewed?

15 A That's correct.

16 Q And you said KPMG was ultimately retained. And when was
17 that retention?

18 A In May of 2011.

19 Q And at the time that KPMG was retained in May of 2011 did
20 KPMG offer its clients any particular platforms? You had
21 mentioned earlier that platforms are used in this processing
22 phase to upload the data and work with it, the data and the ESI
23 generally.

24 Did KPMG in May 2011 offer its clients any particular
25 platforms for this purpose?

1 A We did. We offered Discovery Radar, which is also known
2 as DR, which is our own platform. And we also offered
3 Clearwell at that time, which is a commercial platform.

4 Q And why was KPMG offering Clearwell to its clients?

5 THE COURT: Wait. I have a question. Is a platform
6 a search method?

7 THE WITNESS: It's the tool, ma'am, that we would
8 load the data into. And it would be the tool to be able to
9 apply searches, but also ultimately to be able to review some
10 of the information as well.

11 THE COURT: All right. And the two that you were
12 using at that time was Clearwell and what else?

13 THE WITNESS: Discovery Radar.

14 BY MR. NEUWIRTH:

15 Q And I believe you said Discovery Radar was a KPMG
16 proprietary product?

17 A Yes, it's our proprietary platform. Also called DR.

18 Q And so what was the reason that in May 2011 in addition to
19 your own product KPMG was also offering Clearwell?

20 A Well, we definitely like our own product, of course. But
21 we also want to have something that's commercially available
22 that other folks like in the industry, and so we'll always have
23 a commercial product available as well. And we go out and do
24 the research to figure out what's sort of best in class and
25 then make our decisions as far as sort of what's best in class

1 and then what the majority of the industry likes.

2 And we decided that we would start hosting Clearwell
3 several years ago. I can't remember when we actually started,
4 but we've been hosting it ever since.

5 Q And as of May 2011 was it still KPMG's view that Clearwell
6 was best in class, as you have said?

7 A Yes, we have run several engagements through Clearwell and
8 had very good success with it.

9 Q Now, are you familiar with something called the Magic
10 Quadrant?

11 A I am.

12 Q And what is the Magic Quadrant?

13 A Well, from an E discovery perspective the Magic Quadrant
14 was Gartner's assessment this past year of E discovery vendors
15 that are out there.

16 MR. NEUWIRTH: Okay. Your Honor, if I could mark for
17 identification as Defendants' Exhibit 1, we have the Gartner
18 report, the Magic Quadrant report that the witness just
19 referred to. If I may hand a copy to the witness.

20 THE COURT: Surely.

21 MR. NEUWIRTH: And I also have copies for opposing
22 counsel and for Your Honor.

23 THE COURT: Thank you.

24 MR. MOGIN: May I just note for the record, Your
25 Honor, we have not seen this before.

1 MR. NEUWIRTH: May I approach, Your Honor.

2 THE COURT: Yes. Thank you.

3 MR. NEUWIRTH: I'm giving the marked copy to the
4 witness, and I also have copies for the court reporter and for
5 Your Honor.

6 BY MR. NEUWIRTH:

7 Q Now, if I could, can you just tell me what it is that I've
8 just handed to you.

9 A It's the Magic Quadrant for E discovery software.

10 Q And is there a date on this document?

11 A May 13th, 2011.

12 Q And is this the report by Gartner that you were referring
13 to in your prior answer?

14 A Yes.

15 MR. NEUWIRTH: If there's no objection, Your Honor,
16 we would move for the admission of this document as Defendants'
17 Exhibit 1.

18 MR. MOGIN: May we reserve objection, Your Honor.

19 THE COURT: Yes, until after cross.

20 MR. NEUWIRTH: That's fine, Your Honor.

21 BY MR. NEUWIRTH:

22 Q I'd like you, if you could, to turn to the second page of
23 this document. And do you see that there is a chart on the
24 upper right side of that page?

25 A Yes, I do.

1 Q Can you tell me what that chart is.

2 A Sure. That's the results of the assessment and where they
3 placed the different vendors.

4 Q And based on your review of this chart, where does
5 Clearwell fall in the assessment by Gartner of E discovery
6 software?

7 A Clearwell is in the leader area.

8 Q And how does the leader area compare to the other areas on
9 this chart?

10 A The leader area is definitely the ones that are on top,
11 the ones that understand the market and are very well respected
12 out there.

13 Q And is the Magic Quadrant report something that you in
14 your work at KPMG value?

15 A Yes, it is.

16 Q Now, are you familiar with the term analytics as it's used
17 with respect to E discovery platforms?

18 A Yes, I am.

19 Q And what is your understanding of the term analytics?

20 A There's a lot of different types of analytical tools that
21 can be used to help do things like group like documents
22 together. Some of the analytics that have been developed over
23 the past years have been able to group like documents together
24 based on their content. There's also analytics that allow you
25 to analyze e-mail and pull out certain web domains. There's

1 analytics that allow you to put e-mail threads together in
2 their context so you can get through, you know, the data a
3 little bit easier.

4 Q Now, did the version of Clearwell that KPMG made available
5 in May of 2011 include any features that had analytics?

6 A It did. It had several of those features that I just
7 mentioned. The topics feature is one that can group like
8 documents together based on their subject. It has a threading
9 feature. It has very robust domain analysis to pull out
10 different domains and things like that.

11 Q Now, you just said that this topics feature can group
12 documents together by their subject. And earlier you had
13 referred to analytics as grouping documents together based on
14 their contents.

15 Is the topics feature able to group documents based
16 on their content?

17 A Yes, it analyzes the content and groups them together.

18 Q And you had also mentioned e-mail threading. What is
19 that?

20 A It's where you can take a chain of e-mails with the
21 replies and the forwards and put the thread together and see it
22 sort of in its context.

23 Q And is that something that the Clearwell program that was
24 available in May 2011 was able to do?

25 A Yes.

1 Q Were there any other features of Clearwell that you would
2 put in the analytics category?

3 A It can analyze again, you know, domain names, e-mail
4 addresses. It's got several, several tools. I can't remember
5 them all off the top of my head, but they're right there on the
6 web site.

7 Q Now, I think you know that in this litigation the
8 plaintiffs are using the term content based advanced analytics
9 or CBAA. Prior to this litigation had you in your professional
10 experience or otherwise ever heard the term content based
11 advanced analytics or CBAA?

12 A I hadn't heard that specific term before this case.

13 Q Now, do you have an understanding of what content
14 analytics refers to?

15 A Well, I think generally it refers to what I described
16 before in being able to group like documents together, analyze
17 the content of the documents and then group like documents
18 together. And I saw some of the letters from the plaintiffs
19 when they were describing content based advanced analytics.
20 And it seemed to go between the technology that allows you to
21 group those like documents together, and then also later on in
22 the letter it seemed to start talking about some of the newer
23 technology that's out there that's commonly referred to as
24 predictive coding.

25 Q And what is predictive coding?

1 A Predictive coding is a fairly new technology that will
2 allow a more senior trained attorney with a good grasp of the
3 issues to sit down and start to train the software as to what's
4 responsive and what's not responsive. And the software watches
5 these actions, and after a while it gets to a point where it
6 understands and becomes stable so that you can apply those
7 across the entire population.

8 Q How exactly is it that an attorney trains the software?

9 A By literally sitting down and making judgment calls on a
10 sample set of documents to say this is responsive, this is not
11 responsive.

12 Q So the attorney takes documents, decides which ones are
13 responsive and which aren't, and shares those decisions with
14 the software?

15 A That's correct.

16 Q And the software learns from those decisions?

17 A Yes, that's right.

18 Q And is it correct that those decisions that the attorney
19 is making about which documents are or are not responsive are
20 subjective decisions by the attorney?

21 A Absolutely.

22 Q And so those subjective decisions are then shared with the
23 software, which learns from those subjective choices, is that
24 correct?

25 A Yes.

1 Q Now, does KPMG offer predictive coding at the present
2 time?

3 A We do. In our latest version of Discovery Radar we have
4 implemented Equivio Relevance into the back end of it, so we've
5 got the capability to provide some of the -- what's commonly
6 referred to as predictive coding. And we call it software
7 assisted first review.

8 Q Now, you said that KPMG offers this at the present time.
9 When did KPMG first make this type of predictive coding feature
10 available to clients?

11 A We first rolled it out in September of last year.

12 Q And that's September 2011?

13 A That's correct.

14 Q And so that's four months after you had been retained in
15 this case by Georgia Pacific in May of 2011, is that correct?

16 A Yes, that's right.

17 Q Now, in May of 2011, just to be clear, did the KPMG
18 software that you were offering include any of this, any of
19 these predictive coding features?

20 A No. We didn't have that until September.

21 Q And so at no time prior to September were they available?

22 A Correct.

23 Q Okay. And does the DR 4 predictive coding feature work
24 the way you described predictive coding to work with an
25 attorney making choices and sharing that information with the

1 software?

2 A It does.

3 Q And what is the benefit in your view based on your
4 experience of using predictive coding?

5 A Well, I personally don't have any matters that are using
6 the capability yet, but we do have some at KPMG, and we're --
7 it's still pretty new. But what we're hoping is that it's
8 going to save a lot of money on the review end. So instead of,
9 you know, first taking that broad set of ESI and once you kind
10 of cull it down to something meaningful, apply your date range,
11 your keywords, your file types, and you get that down to
12 something a little bit more manageable. Instead of then just
13 having a team of attorneys do this first pass review, we can
14 apply some of this technology and hopefully save some of the
15 costs of the review.

16 Q So just to be clear --

17 MR. MOGIN: Your Honor, objection. In light of the
18 witness' admission that he has no experience in using this
19 technology, I'd like to move to strike the prior testimony.

20 MR. NEUWIRTH: I think we can clarify this, Your
21 Honor. I don't think the witness said he has no experience
22 with this technology. It's something that's offered by KPMG.
23 And I think we can establish that he's familiar with it and
24 involved in offering it to clients. I think what he said was
25 that at the present time he wasn't working on any matters where

1 a client had chosen to use it.

2 MR. MOGIN: Well, I believe the record will reflect
3 that Mr. Koch testified that he has not worked on any
4 engagements that have used this technology.

5 THE COURT: All right. So in the spirit of what I
6 said before about we're going to let most things in and I'll
7 figure it out afterwards, why don't you ask these questions
8 again about his personal experience.

9 MR. NEUWIRTH: Sure.

10 THE COURT: Maybe his training, maybe his knowledge.
11 Maybe he read an article.

12 MR. NEUWIRTH: Sure.

13 THE COURT: I don't know what it's based on. Okay.

14 MR. NEUWIRTH: Okay.

15 THE COURT: I think it's called foundation.

16 MR. NEUWIRTH: Sure.

17 BY MR. NEUWIRTH:

18 Q So, Mr. Koch, do you feel familiar with the features, the
19 predictive coding features that KPMG now offers in its
20 proprietary platform?

21 A I do. I've been involved in discussing it with my clients
22 and provided demos to clients with this software. And we just
23 haven't had any takers yet.

24 Q And as this product came to market did you familiarize
25 yourself with it? Were you trained in how it operates, and do

1 you feel comfortable that you understand exactly how it works?

2 A Yes, I do. And my teams that run the day-to-day
3 operations have been through the training for it. Again, we
4 demonstrate this software and we're actively talking to our
5 clients about using it. It's just -- it's fairly new.

6 Q Right. And this new software, just to be clear although
7 you are familiar with it and you said you've demonstrated it to
8 clients, as of the present time have any clients of your
9 southeast region business chosen to use this new -- these new
10 predictive coding features?

11 A Not yet.

12 Q And are there any clients of KPMG throughout the United
13 States that have chosen to use it?

14 A Yes. There are a couple of matters that we have using the
15 predictive coding capabilities right now.

16 Q Now, when you say a couple of matters, is it fair to say
17 that that's probably five or fewer?

18 A It's fewer than five.

19 Q And how many matters -- how many E discovery matters does
20 KPMG typically handle throughout the United States at any point
21 in time?

22 A We have hundreds.

23 Q And so at the present time there may be a few, less than
24 five of all those hundreds of matters where clients have chosen
25 to use those predictive coding tools, is that correct?

1 A It is so far. Again, we just rolled it out in September,
2 and we're very hopeful that, you know, many more people are
3 going to start using it in the future.

4 Q Right. Now, given that you're familiar with this and
5 given your obvious familiarity based on what you testified
6 earlier about matters involving search terms, can you tell us
7 what you see -- just now that we have gone over your background
8 with this, can you just tell us what you see as the potential
9 benefits of using these predictive coding features.

10 A Yes. So the same as I said last time. We think that it's
11 going to be able to help clients save a lot of money on the
12 review end, to where instead of having a large team of sort of
13 first pass review folks going through the documents, that this
14 might be able to be applied to save some of that time and
15 money.

16 Q Now, just to be clear, you're talking now about review,
17 which was that fourth stage of the five you described. This is
18 where attorneys look at documents that have been identified as
19 potentially responsive and review them, is that correct?

20 A That's correct.

21 Q And what you're saying is that this predictive coding
22 software can be involved in doing that review rather than human
23 beings, is that correct?

24 A That's correct.

25 Q And that's why it would save money?

1 A Yes.

2 Q Now, in terms of the earlier stages where you take a broad
3 set of ESI and cull it down, is the KPMG predictive coding
4 software being used instead of search terms for the purpose of
5 culling down a broad set of ESI to something more narrow?

6 A I'm not aware of any matters now where they're using the
7 predictive coding capability instead of search terms. The
8 matters that I'm aware of they have culled down the larger set
9 of ESI with some terms prior to applying the predictive coding
10 to help on the review side.

11 Q Now, putting aside the benefit of costs, have you found
12 based on your work at KPMG or otherwise that predictive coding
13 would be better at locating potentially responsive documents
14 than search terms?

15 MR. MOGIN: Objection, Your Honor. At this point
16 we're far afield of the witness' --

17 THE COURT: Okay. I mean, not only has he not done
18 it, the company hasn't done it. Nobody -- I mean, I guess what
19 he said is -- I don't know what he's basing this on. I mean, I
20 mean, this is a real foundation problem here.

21 MR. NEUWIRTH: Yes, well, let me see --

22 THE COURT: Even for -- okay. I mean it's kind of
23 ironic that the plaintiffs are objecting when he's saying how
24 wonderful predictive coding is, but that's their -- you know,
25 that's kind of what happens at hearings. Okay.

1 MR. NEUWIRTH: Well, let me see if I can do this a
2 different way.

3 THE COURT: But on this specific question -- all
4 right. Try it again.

5 MR. NEUWIRTH: I can ask it a different way.

6 BY MR. NEUWIRTH:

7 Q You're involved in the marketing -- you've described that
8 you're involved in marketing predictive coding to clients,
9 correct?

10 A Yes.

11 Q In marketing predictive coding to clients have you told
12 clients that predictive coding is better than search terms for
13 taking a broad set of ESI and narrowing it to what you
14 described as a more meaningful set to be reviewed?

15 THE COURT: I'm going to object to that. I mean,
16 honestly, what he is saying in a marketing -- I mean, I --

17 MR. NEUWIRTH: Well, I think what we're trying to do
18 is determine whether the advantages -- what he considers to be
19 the reason to use predictive coding, which I think is different
20 from what the plaintiffs are saying are the reasons to use
21 predictive coding. And all we're trying to do is establish
22 what those are.

23 THE COURT: I think one question you could ask that I
24 have a question of, Mr. Mogin, is is this new software -- in
25 the review portion, in step one or step two, are they doing

1 word search; or is it all analytical search? And then is it
2 word search or department search? Those are the issues in this
3 case here. I mean, what he really is telling somebody in a
4 marketing setting is really different than here.

5 MR. NEUWIRTH: Well, that's fine --

6 THE COURT: Now, Mr. Mogin, is that question -- do
7 you have an objection to my question?

8 MR. MOGIN: I don't, Your Honor. But just to confuse
9 the matter as much as I possibly can.

10 THE COURT: Okay.

11 MR. MOGIN: We have no objection to Mr. Koch
12 discussing KPMG's marketing. My objection was the hearsay that
13 was inherent in what Mr. Neuwirth was trying to bring out, and
14 that Mr. Koch had gone far afield of his admitted lack of
15 expertise. But as to the marketing, if they want to discuss
16 that, we do not object, Your Honor.

17 MR. NEUWIRTH: I don't think there's been any
18 admitted lack of expertise. I think Mr. Koch described his
19 familiarity with the software. The only issue is how much it's
20 been used so far, and I think he's established that it's a new
21 software that's just come to market. And all I was trying to
22 do -- he explained that the main advantage I think was cost
23 savings at the review phase. So I think I can ask your
24 question and get directly to the point and move on.

25 THE COURT: Thank you.

1 BY MR. NEUWIRTH:

2 Q And the question is, is this software something that KPMG
3 is using to do these earlier phases of identification and
4 culling down of the broad set of ESI to something more narrow,
5 or is it being used for the review phase?

6 A It's being used for the review phase. Not necessarily for
7 search and retrieval.

8 Q And what is being used for search and retrieval?

9 A Well, in my experience --

10 MR. MOGIN: Objection, Your Honor. I believe the
11 witness testified that he was not involved in the search and
12 retrieval in this case. So again, we're well beyond his
13 personal knowledge.

14 MR. NEUWIRTH: No. No. He's testified -- we can go
15 over what KPMG did in a minute, but he's made clear what we're
16 talking about now is this phase of taking the broad set and
17 narrowing it down, which he described as the third phase of the
18 processing, which he said he was involved with.

19 THE WITNESS: That's correct.

20 MR. NEUWIRTH: This is just a question about what
21 KPMG is doing, which I think the witness has said he's familiar
22 with.

23 THE COURT: I want to hear the answer to this.

24 MR. NEUWIRTH: Thank you, Your Honor.

25 THE COURT: Thank you.

1 THE WITNESS: I'm sorry. Can you repeat it again.

2 BY MR. NEUWIRTH:

3 Q I think Judge Nolan had such a good question, which I
4 repeated. But let me do it one more time.

5 A All right.

6 Q Is KPMG -- I think you said that KPMG is using this for
7 the review phase. Is KPMG using this -- you said it's not
8 using this software for the process of narrowing the broad set
9 of ESI to a smaller one for review. And the question is what
10 is KPMG doing to narrow down the broad set to a more narrow one
11 in cases where this new software is being used for the review?

12 A In my experience the way to narrow down a large set of ESI
13 has been to apply certain filters. So date range, file type,
14 and keyword.

15 Q And keyword you mean search terms?

16 A Yes, search terms.

17 Q Okay. Now, what was the role that KPMG played in this
18 case in GP's, Georgia Pacific's process of taking a broad set
19 of ESI and narrowing it down to a more meaningful set?

20 A We processed data and hosted it in Clearwell, and helped
21 to -- helped in the search and testing of the search terms by
22 applying those searches that were provided to us in Clearwell
23 and then providing -- we would provide the results, the hit
24 count results. And then they would go back and make some
25 modifications and provide us with updated search terms as they

1 were sort of tuning them up. And then we would apply those and
2 sort of back and forth.

3 Q And what was your personal role in this process?

4 A I'm the engagement partner responsible for the engagement.

5 Q Now, when did this process begin?

6 A In May of last year.

7 Q Now, based on your personal experience, are there any
8 potential problems with using search terms to segregate out
9 potentially responsive ESI from a larger universe of ESI?

10 A Absolutely.

11 Q And what are those potential problems?

12 A Well, you could miss things by not having a very good
13 keyword list or you could have a keyword list that is going to
14 bring back a lot of false positives and cost a lot more money
15 to get through and have a lot of junk in the data set.

16 Q Okay. And were any steps taken here in the Georgia
17 Pacific process to address these potential problems?

18 A Yes, there were several iterations of key terms that we
19 applied and provided hit results back so that they could be
20 tweaked and tuned up to make sure that they were comfortable in
21 the end with the results that were coming out of the data from
22 the keyword -- from applying the keywords.

23 Q And was the topics function of Clearwell, this analytics
24 function you talked about, used in this process?

25 A It was. They used it in both the data that was returned

1 from keywords and the data that wasn't returned from keywords
2 to test and look through to determine if there should be any
3 other keywords that they might want to apply.

4 Q So you're saying that it was applied to the set of
5 documents that was hit by the key terms as well as the
6 documents that were not?

7 A That's correct.

8 Q But the topics function was used on both. Were there any
9 other steps that were taken to address these potential problems
10 that could occur with search terms?

11 A Yes. There were tests run of the data that was not
12 returned by the keywords, and they called that the null set.
13 So they ran tests in -- by pulling random samples of that
14 information from the documents that weren't returned from the
15 keywords to test to see if they had any errors and see if they
16 needed to add any other keywords to their keywords.

17 THE COURT: I have a question. A null set, N-U-L-L,
18 Miss McCullough, what does that mean?

19 BY MR. NEUWIRTH:

20 Q Can you describe what that means?

21 A Sure. The null set -- when you have the list of keywords
22 and you apply the keywords to the data, it brings back certain
23 files that hit on the keywords. Everything else that's left
24 behind, that's what we have been calling the null set. So the
25 things that did not hit on the key words.

1 THE COURT: And then say again what -- so then when
2 you get a set of null set, then what did you do with the null
3 set?

4 THE WITNESS: They went in and pulled random samples
5 of documents from that null set to test to see if there were
6 any other words that they should be adding to the keyword list
7 to bring back any documents that may have been responsive but
8 not captured by the keyword list that they already had.

9 BY MR. NEUWIRTH:

10 Q And that was done including with the topics feature?

11 A That's correct.

12 Q That analytics tool in Clearwell?

13 A Yes.

14 Q And was there also testing done of the random sample to
15 see whether or not documents that were responsive were ending
16 up in the null set as a result of the search terms?

17 A Yes, that was the -- what I just went through a second
18 ago.

19 Q And --

20 THE COURT: I have another question. What's the
21 difference between a topics function and a keyword?

22 THE WITNESS: The topics function is an analytics
23 function that can analyze the documents and group like
24 documents together based on their content. And the keywords
25 is --

1 THE COURT: I know what keyword is. So topic is more
2 than one word?

3 THE WITNESS: Yes, it's more of an analytical tool.

4 THE COURT: Okay. But it includes more than a
5 keyword.

6 THE WITNESS: Yes, ma'am.

7 THE COURT: Okay.

8 BY MR. NEUWIRTH:

9 Q And, in fact, it would look at the entire subject matter
10 of the document and compare that to subject matter in other
11 documents, correct?

12 A It would.

13 Q And that was a function in Clearwell?

14 A Yes.

15 Q And do you understand if any other steps were taken in
16 this process to address the types of problems that can come up
17 with search terms?

18 A Well, I do know that they got input from other defendants
19 on the keyword list and also from the plaintiff as well.

20 Q And what was the nature of the feedback that was received
21 from the plaintiffs?

22 A They sent a letter. I remember seeing a letter where the
23 plaintiffs had a whole list of keywords that I think they
24 pointed out that GP had not considered. And so we took a look
25 at that letter and then added, I think added some keywords from

1 that list that they provided.

2 MR. NEUWIRTH: May I approach, Your Honor, with a --

3 THE COURT: Yes.

4 MR. NEUWIRTH: I am handing for identification
5 Defendants' Exhibit 2 to the witness. If I may approach.

6 BY MR. NEUWIRTH:

7 Q And what is it that I have given to you, Mr. Koch?

8 A The plaintiffs' preliminary analysis of defendants' first
9 sets of proposed search terms.

10 Q And is this what you were referring to when you said that
11 there had been feedback provided by the plaintiffs on the
12 search term list?

13 A Yes.

14 MR. NEUWIRTH: And, Your Honor, I understand you're
15 reserving on entry into evidence. We will move for entry into
16 evidence, but we will understand that Your Honor reserves.

17 BY MR. NEUWIRTH:

18 Q Now, what was your reaction to this document upon
19 receiving it?

20 A Well, we looked at the document and it seemed to take
21 issue with, with the keywords in general and provided a list of
22 things that weren't considered by GP.

23 Q And did you find that the suggestions that the plaintiffs
24 had provided were helpful?

25 A Well, it probably would have been a little bit more

1 helpful to say here's -- you know, here's definitively what we
2 think you should add from a keyword perspective. Some of the
3 things they list on here that weren't, you know, considered
4 include things that are pretty broad and you might find, you
5 know, in someone's signature file at the end of their e-mail,
6 which could potentially return like every e-mail. So for cell
7 and mobile and fax and e-mail and things like that, you know, I
8 have that at the bottom of every e-mail on my signature block.

9 Q But was any use made of these suggestions by the
10 plaintiffs?

11 A Yes, we did end up using a couple of the words out of the
12 list to add to the overall keyword list.

13 Q Okay. And ultimately what was the number of hours that
14 KPMG personnel devoted to this process at Georgia Pacific of
15 taking a broad set of ESI and narrowing it down to a more
16 meaningful set for review?

17 A We sent almost 500 hours.

18 Q Now, I take it you are aware that the plaintiffs at this
19 point have been suggesting that the search term process here
20 should be replaced with a content based analytics approach?

21 A I am.

22 Q Based on learning that, has KPMG recommended to Georgia
23 Pacific that it abandon this search term process that has been
24 implemented and instead use a content based analytics approach?

25 A No. I think the most important thing about using any kind

1 of search capability is how you test it and how you QC it in
2 the end to be comfortable with the results.

3 Q By QC what do you mean?

4 A Quality control testing. Testing the results.

5 Q And based on your involvement and KPMG's involvement, do
6 you feel that the process that -- well, what is your view of
7 the process that Georgia Pacific has implemented here through
8 its search term methodology?

9 A Sure. I thought it was very thorough, very robust. There
10 were several iterations of search term development, and they
11 tested and went back and tuned up the keywords several times.
12 And, you know, quite honestly that's more than I see in a lot
13 of the cases that I'm dealing with. A lot of times they'll
14 just -- both sides will agree to some key terms that really
15 haven't been tested a lot and bring back a lot of garbage. And
16 this was a fairly thorough process.

17 Q And so given --

18 THE COURT: Okay. We have --

19 MR. MOGIN: Well, I'm going to object, Your Honor --

20 THE COURT: Yes.

21 MR. MOGIN: -- in light of the fact that the witness
22 has already testified that he has no experience whatsoever with
23 the technology that he's now critiquing.

24 THE COURT: Wait. I thought he was critiquing what
25 he did.

1 MR. MOGIN: No, he just said --

2 THE COURT: I thought this was pretty self-serving
3 that he thought he did a pretty good job. But you go right
4 ahead. But that's what I thought he was critiquing right now.

5 MR. NEUWIRTH: Thank you, Your Honor.

6 THE COURT: Not analytics.

7 MR. NEUWIRTH: Thank you, Your Honor.

8 THE COURT: Wasn't that your question?

9 MR. NEUWIRTH: My question was what was his view of
10 the process --

11 THE COURT: Did he do a good job.

12 MR. NEUWIRTH: -- that was done here. Correct.

13 THE COURT: Yes.

14 BY MR. NEUWIRTH:

15 Q And so given that KPMG now has this new -- these new
16 features, these predictive coding features, why aren't you
17 recommending that Georgia Pacific use those instead of the
18 methodology that it's put into place?

19 A Well, I would love Georgia Pacific to use our new tool.
20 That would be fantastic. But the way we sort of go to market
21 with our tool is a cost savings measure around helping save
22 money in the review phase. And GP's already done the review of
23 the majority of the data, so there would be no cost savings.

24 Q Now, when you say a review of the majority of the data,
25 what are you referring to?

1 A All the documents that they have in their repository.

2 Q And so you talked earlier about the five steps. That the
3 first three involved taking the broad set of ESI and narrowing
4 it down. The fourth step that you mentioned is actually
5 reviewing that narrowed set of ESI that emerges from that
6 process.

7 Are you saying that those documents have already been
8 reviewed?

9 A Yes.

10 Q And they've been reviewed by human beings?

11 A Yes, that's correct.

12 Q And do you know roughly what percentage of those documents
13 have been reviewed to date by Georgia Pacific?

14 A I think it's just about all of them. 99 percent or more.

15 Q And so this review benefit you talked about from using the
16 predictive coding feature would not add any value here since
17 the documents have already been reviewed by human beings?

18 A That's correct.

19 MR. NEUWIRTH: We have no further questions, Your
20 Honor.

21 THE COURT: Would you wish to cross-examine?

22 MR. MOGIN: I do, Your Honor. But in light of the
23 technical subject matter and the fact that we don't have
24 reports or depositions, I wonder if we could take a few minutes
25 before we begin the cross-examination.

1 THE COURT: Sure. Sure. We can take our morning
2 break right now. So 10 minutes. Can you do that?

3 MR. MOGIN: Yes.

4 THE COURT: Okay. 10 minutes. Okay. Thank you.

5 (Short break taken.)

6 THE COURT: Okay. We're back on the hearing. Mr.
7 Mogin, are you ready?

8 MR. MOGIN: I am, Your Honor, subject to the caveat
9 that as I said before we've had no depositions, we don't have a
10 report from Mr. Koch. So this is a bit --

11 THE COURT: So you're doing it the old fashioned way.

12 MR. MOGIN: Yes, trial by ambush. Here we go.

13 THE COURT: That's true. Okay. Here you go.

14 CROSS-EXAMINATION

15 BY MR. MOGIN:

16 Q All right. Mr. Koch, you're with KPMG, correct?

17 A Yes.

18 Q KPMG is a worldwide enterprise, correct?

19 A Yes.

20 Q KPMG provides forensics services throughout much of the
21 English speaking world, correct?

22 A That's correct.

23 Q They provide them in England, correct?

24 A Yes.

25 Q They provide them in Canada, correct?

1 A Yes.

2 Q They provide them in the United States, correct?

3 A Yes.

4 Q Now, do you interface with any of the people from Canada?

5 A We do on occasion. They're a member firm, so we talk to
6 the folks in Canada.

7 Q And how about the UK?

8 A Yes, we do.

9 Q Now, is it your understanding that the idea of -- that
10 KPMG only recently got involved in predictive coding and
11 advanced analytics?

12 A In our U.S. firm, yes.

13 Q But that's not true for the rest of the firm, is it?

14 A I'm not aware of what every member firm does.

15 Q Isn't it true that in 2006 there was an acquisition of
16 Autonomy Software by KPMG?

17 A Not KPMG in the U.S., no.

18 Q But it was by KPMG the worldwide enterprise, correct?

19 A Well, KPMG is a global network of member firms. We're not
20 a -- we're not a global firm per se. But we're a global
21 network of member firms that are each sort of their own legal
22 entities, but we're a cooperative.

23 Q Do you share information?

24 A Certainly.

25 Q So are you aware of what's going on in the forensics

1 services in Europe?

2 A Yes.

3 Q In England?

4 A Yes.

5 Q In Canada?

6 A Yes.

7 Q So are you aware of the Autonomy acquisition?

8 A I had heard about it, yes.

9 Q When did you first hear about that?

10 A I don't recall. I mean, it's been a while.

11 Q A couple years?

12 A Sure. At least a year.

13 Q Maybe three years?

14 A I don't remember, but it's --

15 Q Maybe even 2006?

16 A Maybe. I don't remember.

17 Q Would it refresh your recollection if I were to show you a
18 KPMG press release announcing the acquisition?

19 A Sure.

20 MR. MOGIN: Your Honor, if I may. If I may approach,
21 Your Honor, and mark this just as an exhibit for demonstrative
22 purposes.

23 THE COURT: Thank you. To refresh recollection.

24 MR. MOGIN: Refresh. Thank you. Thank you for
25 refreshing my recollection.

1 THE COURT: Yes.

2 BY MR. MOGIN:

3 Q Okay. You'll see that this article is headlined "KPMG
4 Forensic Purchases of Leading Investigations Software
5 Solution," correct?

6 A Yes.

7 Q And the date line is June 2006, correct?

8 A That's correct.

9 Q All right. Now, does this refresh your recollection as to
10 the date of the Autonomy acquisition by KPMG?

11 A I hadn't seen this particular news release before, but now
12 seeing it here, yes, I can read it and I understand it.

13 Q Okay. Very good. Now, you're aware that that -- are you
14 not, that that Autonomy product that was purchased by KPMG has
15 many of the features that the plaintiffs have described as
16 CBAA, correct?

17 A I know it has analytic capabilities.

18 Q Okay. Are you aware of a technology called Latent
19 Semantic Indexing?

20 A I am.

21 Q And what do you understand that to be?

22 A Latent semantic indexing is the technology behind the
23 analytic capabilities to group like documents together based on
24 their content.

25 Q Very good. And how long has latent semantic indexing

1 technology been available?

2 A It's been available for years.

3 Q 1970s, is that correct?

4 A I don't know about that.

5 Q Would you say that predictive coding is a proven or
6 unproven methodology for use in eDiscovery?

7 A You know, I don't know that I -- in my experience I
8 haven't used the predictive coding capabilities yet of our
9 implementation to prove it or disprove it, so I wouldn't be
10 able to say.

11 Q Well, you've testified that you are involved in marketing,
12 correct?

13 A Yes.

14 Q And you're aware of KPMG's marketing in this area,
15 correct?

16 A Yes.

17 Q Now, do you know Mr. Chris Paskach, P-A-S-K-A-C-H?

18 A Yes.

19 Q And who is Mr. -- would you pronounce that for me?

20 A Paskach.

21 Q Who is Mr. Paskach?

22 A He is our national practice leader.

23 Q He's your boss?

24 A Yes.

25 Q And do you know Mr. Michael Carter?

1 A Yes.

2 Q And who's Mr. Carter?

3 A I believe Mr. Carter is a manager or director at our
4 Cyprus data center.

5 Q At your what?

6 A Cyprus, California data center.

7 Q Do you know Mr. Phil Strauss?

8 A Yes.

9 Q And who is Mr. Strauss?

10 A He's a director in our San Francisco office.

11 Q Now, all of these people are involved in KPMG's forensic
12 services just like you, correct?

13 A Yes.

14 Q Are they also project managers like you?

15 A Well, Chris is the leader. So he's not a project manager
16 per se.

17 Q He's the leader. Okay. So would you concede that a
18 marketing -- strike that.

19 Would you concede that a brochure that bears the
20 names of Mr. Paskach, Mr. Carter, and Mr. Strauss was an
21 authoritative statement by KPMG about its forensics services?

22 A Yes.

23 Q Let me, if I may, please first I would like to ask you to
24 look at -- well, we'll mark this as exhibit next.

25 MR. MOGIN: I guess this would be Plaintiffs' 2, Your

1 Honor.

2 THE COURT: Right.

3 MR. MOGIN: Plaintiffs' 2, please. Entitled, "The
4 Case for Statistical Sampling in eDiscovery."

5 THE COURT: Thank you.

6 BY MR. MOGIN:

7 Q Did I read that correct? This is called "The Case for
8 Statistical Sampling in eDiscovery?"

9 A Yes.

10 Q Have you seen this before?

11 A Yes.

12 Q This is an official KPMG brochure, correct?

13 A Yes.

14 Q It would be an authoritative statement by KPMG, correct?

15 A That's correct.

16 Q And if you look at the back page, it was prepared by the
17 people that we were talking about before; that is, Mr. Paskach,
18 Mr. Carter, and Mr. Strauss, is that correct?

19 A Yes.

20 Q Now, let me also show you --

21 MR. MOGIN: If we could have Plaintiffs' 3 marked
22 please, Your Honor.

23 BY MR. MOGIN:

24 Q Plaintiffs' Exhibit 3 is entitled "Making Document Review
25 Faster, Cheaper, and More Accurate," is that correct?

1 A Yes.

2 Q And the subtitle is "How Concept Searching can Change the
3 Way your Legal Teams Handle First Pass Review," correct?

4 A Yes.

5 Q And this was another official KPMG document, correct?

6 A This document looks like it, just from the cover of it
7 looks like it came from our firm in Canada.

8 Q But you've interfaced with the folks in Canada, right?

9 A Yes.

10 Q You've talked with them about forensic services, right?

11 A Yes.

12 Q But, by the way, Canada has a legal system that's very
13 similar to the United States, doesn't it?

14 A I suppose. Sure.

15 Q Do you know?

16 A I don't.

17 Q You don't know?

18 A No.

19 Q Well, is a tool dependent upon the legal system that it's
20 used in?

21 A I suppose it depends on how you're using the tool.

22 Q You're an expert in these eDiscovery tools.

23 A Sure.

24 Q Is it language dependent?

25 A Some are. Some can handle multiple languages.

1 Q Is Clearwell language dependent?

2 A It can handle multiple languages.

3 Q It can.

4 A Yes.

5 Q Are there other tools that you're aware of that can handle
6 multiple languages?

7 A Yes.

8 Q In fact, speaking of Clearwell, how many times have you
9 read that Gartner report that has the Magic Quadrant thing in
10 it?

11 A Once, and then I've skimmed it. I've skimmed it a couple
12 times.

13 Q Did you read it before you were engaged by KPMG -- I'm
14 sorry, by GP, Georgia Pacific?

15 A Yes.

16 Q When did you first see that?

17 A Well, right when it came out it was distributed to us, so
18 it was all on the same -- probably in the same month that we
19 were engaged by Georgia Pacific.

20 Q Isn't it true that Gartner has been sued several times for
21 not including respected eDiscovery vendors in that Magic
22 Quadrant?

23 A I don't know.

24 Q Are you aware of any criticisms of that report?

25 A Not off the top of my head, no.

1 Q Did you look at any other sources that were similar to
2 Gartner?

3 A Not that I can remember.

4 Q Are all of the eDiscovery tools listed in the Magic
5 Quadrant, do they all perform the same function?

6 A No. There's specific functions that some do and others
7 don't. But generally they perform a lot of the same functions,
8 sure.

9 Q You're certain of that?

10 A I'm -- well, I'm not certain about every tool in the Magic
11 Quadrant, no, not off the top of my head.

12 Q Do you have any experience with any of those tools?

13 A I have -- I don't recall what the -- you want me to flip
14 back to the Magic Quadrant and take a look?

15 Q Sure. Do you have any personal experience with the
16 Symantec tool?

17 A No.

18 Q Do you have any personal experience with the FTI
19 Technology tool?

20 A Well, which FTI Technology tool?

21 Q The one that's listed here in the Magic Quadrant.

22 A I'm not -- I don't know.

23 Q Okay. Do you have any experience with kCura?

24 A We're just starting to offer Relativity, so I have had
25 some recent experience with kCura.

1 Q Relativity is the name of the --

2 A The software --

3 Q -- particular software offered by kCura, correct?

4 A Yes.

5 Q And you understand that Relativity has a predictive coding
6 or supervised learning capability, correct?

7 A I thought I read that they're coming out with it and
8 implementing it, but that it's not quite there yet. I could be
9 wrong.

10 Q You could be wrong on that, right?

11 A I could be wrong.

12 Q How about Guidance? What is it that Guidance or Guidance
13 does?

14 A Guidance makes a variety of software tools to help with
15 eDiscovery and computer forensics.

16 Q But it's not having to do with the search technology, is
17 it?

18 A There's search capabilities in Guidance, sure.

19 Q Search as in searching for documents or searching as in
20 reviewing documents?

21 A Searching for documents.

22 Q Okay. So that's really the collection aspect, correct?

23 A No, it's the searching aspect.

24 Q All right. It's not a review based software, is it?

25 A You can review in Guidance.

1 Q Really?

2 A There's some review capabilities, sure.

3 Q Which part of the Guidance tool can you review?

4 A The piece that presents results on the screen that allows
5 you to review it.

6 Q Do you know the name of that piece?

7 A Not off the top of my head.

8 Q Now, is Autonomy in the Magic Quadrant?

9 A It is.

10 Q And what do you know about Iron Mountain's tool?

11 A I don't have any experience with it.

12 Q How about Kroll Ontrack, are you familiar with that?

13 A I am.

14 Q Have you worked with Kroll Ontrack or seen demonstrations
15 of it?

16 A I've seen it and I've logged into a repository and looked
17 at it briefly at one point.

18 Q Isn't it true that that's a well-respected tool in the
19 industry?

20 A It is. I mean, in my understanding and in my experience
21 it is a well-respected tooling company.

22 Q So maybe that should be in the upper level of the Magic
23 Quadrant?

24 A Well, I don't know. I'm not with Gartner, so ... They
25 did their own analysis.

1 Q Well, if you were doing an analysis based on what you
2 know, wouldn't it be up there?

3 A I don't have enough knowledge to rank them about their
4 particular services. I haven't done the research.

5 Q But you would concede that Kroll is well-respected?

6 A Yes, from my experience they're well-respected.

7 Q Now, let's go to Plaintiffs' Exhibit 2, please. This is
8 the KPMG statistical sampling document. I wonder if you could
9 go to page 1, please. Do you agree with the statement that
10 "Effective use of statistical sampling can help overcome the
11 concern about use of predictive coding technology by
12 quantifying the reliability of the system's technology assisted
13 predictive coding," as is stated there?

14 A Yes, I think that statistical sampling can be used to test
15 and measure the reliability of any sort of search method that
16 you might use.

17 Q And did you prepare the statistics that have been
18 presented in this case regarding the review of Georgia
19 Pacific's documents?

20 A No.

21 Q Do you know who did?

22 A Prepare the --

23 Q Who prepared the statistical report that has been
24 presented to plaintiffs and the Court regarding the review
25 process of the Georgia Pacific documents?

1 A I don't know.

2 MR. NEUWIRTH: Your Honor --

3 THE COURT: Yes. Hold on. We have --

4 MR. NEUWIRTH: When you refer to a report, can tell
5 us what you are referring to?

6 THE COURT: I had the same question.

7 MR. MOGIN: Your Honor, we can mark as Plaintiffs' 4
8 the November 22nd letter which I believe is Defendants' Exhibit
9 15.

10 THE COURT: So now wait. Why don't we identify --
11 okay. So it's the November 22nd letter. It is from who to
12 whom?

13 MR. MOGIN: This is a November 22nd letter addressed
14 to Mr. Freed, Mr. Kanner, and myself.

15 THE COURT: From the?

16 MR. MOGIN: From counsel for Georgia Pacific Mr.
17 Neuwirth.

18 THE COURT: Okay.

19 MR. MOGIN: And it was presented by the defendants as
20 Exhibit 15 in this matter in their opening brief.

21 THE COURT: Okay. Thank you. And we're now marking
22 this as Plaintiffs' 4, right?

23 MR. MOGIN: Yes.

24 THE COURT: Okay.

25 BY MR. MOGIN:

1 Q Have you seen this letter before?

2 A I do recall seeing this letter.

3 Q Okay. Now, if you'll look at page 5. Do you have page 5
4 before you?

5 A I do.

6 Q And you'll see there's a paragraph there that refers to a
7 validation process where Counsel on Call determined with
8 99 percent confidence that the final set of search terms had no
9 more than a 5 percent margin of error in identifying documents
10 as not responsive to plaintiffs' document requests.

11 Did I read that correctly?

12 A Yes, you did.

13 Q Have you seen this before?

14 A I do recall seeing this document.

15 Q Have you seen the underlying statistical report -- have
16 you seen any underlying data that relates to this declaration
17 of 99 percent confidence level?

18 A I haven't seen any under -- any other report. I've seen
19 this document.

20 Q Are you familiar with statistical reporting?

21 A Not extremely. I'm not a statistical person.

22 Q Do you understand what this statistic means?

23 A Only very generally.

24 Q A layman's understanding, correct?

25 A Sure.

1 Q Is that the type of statistical reporting that KPMG is
2 referring to in Plaintiffs' Exhibit 2? That is, the document
3 entitled "The Case for Statistical Sampling in eDiscovery."

4 A I have the document in front of me again. What was the
5 question?

6 Q Is the report that you just read from Plaintiffs' 5; that
7 is, the November 22nd letter, the type of statistical reporting
8 that KPMG is endorsing in this brochure "The Case for
9 Statistical Sampling in eDiscovery?"

10 A Well, we put out the document that talks about the case
11 for statistical sampling. So, yes, we are -- you know, as a
12 firm we are -- we put out a white paper about statistical
13 sampling and using it to test results of your searches.

14 Q Okay. Is this report compliant with what KPMG has --

15 MR. NEUWIRTH: Could you repeat the question, please.

16 BY MR. MOGIN:

17 Q Is the report in Plaintiffs' 5 compliant with the type of
18 statistical reporting urged by KPMG in this document?

19 A I don't understand your question. Tell me again which
20 document you're referring to. Is the report --

21 Q Is the statistical report that's made there in Plaintiffs'
22 5, the November 22nd letter, reporting on the results -- the
23 search results of the KPMG docu -- of the Georgia Pacific
24 documents in which KPMG participated, is that statistical
25 reporting compliant with the type of statistical reporting

1 urged by KPMG in the document discussing the use of
2 statistics -- a statistical sampling in eDiscovery?

3 A I don't know. I haven't analyzed and digested both
4 documents together to compare them. I couldn't tell you off
5 the top of my head.

6 Q But you've marketed statistical sampling to your clients,
7 haven't you?

8 A Yes.

9 Q So you have an understanding of statistical sampling,
10 don't you?

11 A Generally.

12 Q Okay. So can you read that report and make any sense of
13 it?

14 A I mean, I can read it just the same you did. I didn't
15 write it. So, you know, I can sit here and read it to you. I
16 guess what's the question? Does it make sense?

17 Q That's fair enough.

18 All right. Let's go to, if you would, page 5 of the
19 statistical sampling document. Now, you mentioned in your
20 direct testimony something about a random sample.

21 A Yes.

22 Q Were you using that term in a statistical sense, or were
23 you using that term in a more generic sense?

24 A In a generic sense based on how Clearwell implements that
25 technology and ability to pull a random sample.

1 Q But you understand that the term random sample is, in
2 fact, a term of art in statistics, correct?

3 A I do.

4 Q Now, and you will concede, will you not, that the random
5 sample that you discussed earlier was not a statistically
6 correct random sample?

7 A I couldn't say. I don't know. I'm not a statistics
8 person.

9 Q You did no statistical measurements to determine if the
10 statistical -- if the random sample complied with good
11 statistical practice?

12 A We weren't involved with the statistical piece of it. We
13 applied and pulled the random sample based on the information
14 we were given. So they would say, please create a random
15 sample with this many documents. And we would apply that in
16 Clearwell and then provide the results.

17 Q So you created the random sample?

18 A Technically we clicked the buttons in Clearwell to create
19 the random sample.

20 Q Do you know what you did besides clicking buttons?

21 A Yes. We received either a number of documents to use to
22 pull a certain number of documents randomly or a percentage.

23 Q In other words, a quantity of documents?

24 A That's correct.

25 Q Okay. And were they -- were the documents somehow

1 identified? Was there a numerical indicator on the documents?

2 How did you know which documents to pull?

3 A Oh. Based on whatever document set they asked us to go
4 to. So if it was the null set, they may say, you know, go pull
5 a certain number of documents from the null set. So that's
6 where we would go. That's the collection of documents we would
7 go to create the random sample.

8 Q A certain number of documents?

9 A Sometimes it was a certain number, but also sometimes it
10 was a percentage as well. There's two --

11 Q So how --

12 A Sorry. I wasn't finished.

13 Q Go ahead.

14 A There's two options in Clearwell that you can do, either a
15 number of documents or a percentage.

16 Q And beyond knowing that you followed the Clearwell
17 constructions, you're not aware of whether or not this was a
18 statistically valid random sample?

19 A No, we're not -- we weren't involved in that piece of it.

20 Q So if, in fact, that random sampling method was not
21 statistically valid, then your entire testimony about the
22 random sample would be invalid, wouldn't it?

23 A I don't know. I'm not, I'm not the statistics person. I
24 don't know what you mean.

25 Q So who gave you the instructions to pull a particular

1 number of documents?

2 A GP or Counsel on Call.

3 Q Do you recall which?

4 A It was -- in some cases it would have been Counsel on
5 Call. In some cases it would have been perhaps outside
6 counsel. I mean, everybody was involved, you know, during the
7 whole project as a team.

8 Q Do you recall who gave you the specific -- which person
9 gave you the specific instructions regarding the random sample?

10 A No, not off the top of my head. There were multiple
11 occasions where we were doing that.

12 Q How many random samples did you pull?

13 A I don't remember off the top of my head.

14 Q Well, do you have any idea?

15 A It was, it was several.

16 Q Two or three?

17 A I think it was more than that.

18 Q Is there any relationship in your understanding between
19 the number of random samples and the confidence -- strike that.
20 Confidence level in statistics?

21 A No. I don't, I don't know.

22 Q Let's go back, if we can, to statistical sampling. And to
23 save time I wondered if you would please look at page 5.

24 THE COURT: Of your -- of 2 or 3?

25 MR. MOGIN: This is --

1 THE COURT: Of Canada or the United States?

2 MR. MOGIN: This is the United States, Your Honor.

3 THE COURT: So that's -- so now we're on Plaintiffs'
4 2. What page?

5 MR. MOGIN: This is Plaintiffs' 2, page 5.

6 THE COURT: Page 5. Do you have it, sir?

7 THE WITNESS: I do, yes. I don't have stickers on
8 mine, so I had to pull the -- I've got it, though.

9 BY MR. MOGIN:

10 Q All right. Very good. "Measuring and assuring process
11 quality," do you see that?

12 A Yes.

13 Q And did you have some responsibility for quality control?

14 A Well, we were assisting in the search and the test and we
15 would apply, you know, the criteria that they gave us. So,
16 yes.

17 Q All right. And so did you apply statistical measures?

18 A We applied the numbers that they gave us to pull the
19 random samples. So if they told us to pull, you know, 500
20 documents, a random sample from this collection, we would make
21 sure that we plugged in the right number to pull.

22 Q Okay. So let's start at where it says "Measuring and
23 assuring process quality." KPMG in this document says, "The
24 ability to carefully select a sample and infer from it the
25 condition of a larger population with a high degree of

1 confidence in the reliability of the inference has tremendous
2 utility in electronic discovery," correct?

3 A Yes.

4 Q In KPMG's test projects the software consistently recalled
5 a greater number of relevant documents than the human reviews
6 did, is that correct?

7 A Yes.

8 Q And were you involved in any of those tests that are
9 referred to here?

10 A No.

11 Q Are you aware of the results of those tests?

12 A Generally, yes.

13 Q So KPMG has conducted tests on predictive coding software,
14 is that correct?

15 A We have.

16 Q And they did so for a number of years before they rolled
17 out their own proprietary product, correct?

18 A Well, sir, we don't have a, we don't have a proprietary
19 product in Discovery Radar. We implemented Equivio Relevance
20 into Discovery Radar. And so we -- sure, we did tests before
21 we put it in. You know, I don't think it was going several
22 years back, but definitely we sort of tested the system before
23 we implemented it into Discovery Radar, sure.

24 Q I'm sorry. I misunderstood. I had thought that Radar was
25 a proprietary product of KPMG's.

1 A Discovery Radar is a proprietary product of KPMG's, and we
2 licensed Equivio Relevance to put into Discovery Radar. So
3 that piece of it is Equivio Relevance, but it's implemented
4 into -- sorry if I wasn't clear before. It was implemented
5 into Discovery Radar.

6 Q Similar to what they do in the UK with the Autonomy
7 product?

8 A I suppose. I'm not familiar with how they use Autonomy
9 over there.

10 Q All right. Now, let's go to the next part of this, which
11 says, "Demonstrating and assuring process capabilities to
12 defend a technology assisted review is a matter of, 1, sound
13 design; 2, transparency; and 3, quantifiable results." Do you
14 agree with that?

15 A Sure.

16 Q Let's talk here about sound -- well, let's first talk
17 about transparency. What do you understand transparency to
18 mean in this context?

19 A Something that is, that is clear for those to see and
20 understand the process that's being applied.

21 Q And does transparency have some relationship to
22 replicability?

23 A To --

24 THE COURT: To what?

25 BY MR. MOGIN:

1 Q Replicability. The ability of another person to replicate
2 the results.

3 A Sure, I think it does.

4 Q All right. And can you tell me what has been done in this
5 particular case to assure transparency.

6 A Well, I know that the keywords were shared with others to
7 take input, and so that was a transparent process. And there
8 was input sought on how to build, you know, an appropriate
9 keyword list.

10 Q This is referring to transparency in the statistical
11 results, isn't it?

12 A Our document is, yes.

13 Q Yes. Okay. So let me then rephrase the question.

14 What transparency of the statistical results to your
15 knowledge has occurred in this case?

16 A Well, I mean, I guess this letter right here that you were
17 pointing to earlier. I don't have a sticker on mine, so I
18 don't know what exhibit it is. But I suppose that that lays
19 out, you know, what they did, which I would consider
20 transparent.

21 Q Are you aware of any other transparency in statistical
22 reporting that's occurred in this case?

23 A I'm not aware of all the communications back and forth
24 between GP and plaintiffs.

25 Q But the question was transparency in statistical

1 reporting.

2 A Well, I would assume that that's where it would come with
3 the communications back and forth.

4 Q What statistical reports have you seen?

5 A I've seen this document that you referred to earlier as a
6 statistical report, which is a letter. And I haven't seen any
7 other underlying statistical reports that have been generated.

8 Q Are you making some distinction between statistical
9 reports and some other document that has statistics in it?

10 A No. No.

11 Q Okay. So the November 22nd letter is the only statistical
12 report that you're aware of in this case?

13 A Yes.

14 Q From GP?

15 A Yes.

16 Q Now, you oversaw the project, correct?

17 A I oversaw KPMG's engagement to GP, which helped with the
18 processing and hosting of documents and applying search terms
19 and pulling random samples and things like that, sure.

20 Q Well, now, in fact, KPMG wasn't retained until May 4th, is
21 that right?

22 A We were retained in May.

23 Q You were retained in May. In fact, May is when you put in
24 a bid, correct?

25 A Yes.

1 Q Can you tell me the parameters of that bid?

2 A We put in a bid to provide processing and hosting and help
3 with the search, and I think that was, I think that was
4 essentially it.

5 Q It was a financial bid. You stated a price for these
6 services?

7 A Yes.

8 Q And was it your understanding that the lowest bidder was
9 going to win?

10 A No. In fact, I know we weren't the lowest bidder. They
11 let us know that right away.

12 Q Were there higher bidders?

13 A Higher than us?

14 Q Yes.

15 A I'm not sure.

16 Q Do you know who the other bidders were?

17 A I don't.

18 Q Do you know how many other bids there were?

19 A I certainly don't, no.

20 Q Were there more than two?

21 A I just said I don't know.

22 Q Well, when was this lawsuit filed?

23 A I don't remember off the top of my head.

24 Q Do you have any understanding?

25 A I don't remember when it was filed.

1 Q Was it before May?

2 A I'm sure it was if they were engaging us in May and had
3 already, you know, had data for us to process. Then I would
4 assume that it would be, yes.

5 Q In fact, you know it was in 2010, don't you?

6 A I don't know that for a fact.

7 Q Have you reviewed the complaint in this case?

8 A I'm sure I read it at some point, but I just don't recall
9 off the top of my head.

10 Q Would the case number of the case inform you of the year
11 that the case was filed?

12 THE COURT: Okay. I think they'll stipulate to that.

13 MR. NEUWIRTH: We'll stipulate.

14 THE COURT: Okay. All right.

15 BY MR. MOGIN:

16 Q All right. Having now heard that the case was filed in
17 2010, in fact, in September of 2010, in your professional
18 experience is that consistent with good preservation or
19 identification practice; that is, to wait approximately nine
20 months before engaging vendors?

21 MR. NEUWIRTH: Objection.

22 THE COURT: Okay.

23 MR. NEUWIRTH: There is -- now, there's a complete
24 lack of foundation. There hasn't been any testimony on the
25 steps that were taken to preserve documents or when.

1 THE COURT: All right.

2 MR. NEUWIRTH: And he testified he was not involved
3 in that step.

4 THE COURT: All right. A couple things. You can
5 answer that specific thing, but it is now a quarter to 11. I'm
6 a little concerned on our other two witnesses who are supposed
7 to be on this morning. I'm just reminding you. It's your
8 hearing, not mine. But I am reminding you of the other two
9 people. And can you rephrase. Take the keyword out. I mean,
10 I don't think -- I think of everything that's on the issue here
11 today I don't think we're talking about preservation. At least
12 nobody has said preservation until now. If you're interjecting
13 a new issue here.

14 BY MR. MOGIN:

15 Q Well, you testified, Mr. Koch, that it was important to
16 have proper --

17 THE COURT: They can't hear you. So if --

18 MR. MOGIN: I'm sorry.

19 THE COURT: The lawyers can't hear you. Okay.

20 BY MR. MOGIN:

21 Q Didn't you testify on direct about what was essentially
22 the five or six steps that were necessary in order to go
23 through a proper practice, a proper review that incorporated
24 best practices?

25 A Yes.

1 Q Okay. And you understood and didn't you testify about the
2 importance of proper identification of the documents as one of
3 the steps that was to be taken?

4 A Yes, of identification of ESI.

5 Q Of ESI.

6 A Yes.

7 Q Okay. And you talked in terms of custodians, correct?

8 A I did.

9 Q But you're aware that there can be other sources of ESI
10 within an enterprise, correct?

11 A Yes.

12 Q In fact, you in your practice I venture to say have found
13 relevant documents, responsive documents in places other than
14 custodial locations, isn't that true?

15 A I'd say responsive information, sure.

16 Q And what do you understand to be the nature of this
17 lawsuit?

18 A I understand at a high level that it's about price fixing.

19 Q Have you worked on any price fixing engagements before?

20 A I'm certain I have, but I can't recall any off the top of
21 my head. We work on just a ton of different matters, and a lot
22 of times we're not really deep into the substance of the
23 matters.

24 Q Okay. If I misspeak, Her Honor will correct me, but you
25 understand that the crux of this lawsuit has to do with

1 conspiracy; correct?

2 A Yes.

3 Q And does that have any implications with respect to the
4 identification of documents in your practice?

5 A Whether or not conspiracy has any implication?

6 Q That's right.

7 A Sure.

8 Q How so?

9 A Well, I suppose if you were doing an investigation and you
10 may, you know -- and you had a conspiracy, you would want to
11 make sure that you identify, you know, the appropriate
12 documents to gather.

13 Q All right. Very good.

14 MR. MOGIN: Your Honor, I will try to move as quickly
15 as I can in light of your admonition regarding the time. May I
16 suggest the following: In light of the time implications,
17 since we have Plaintiffs' 2 and 3 in, the two KPMG documents,
18 if I can move their admission in and if we can -- well, if
19 those are in evidence, then I can dispense of questioning the
20 witness with respect to these documents.

21 THE COURT: Do you have any objection, Mr. Neuwirth?

22 MR. NEUWIRTH: Well, I only have an objection to the
23 idea that the documents we introduced which are clearly ready
24 to be entered into evidence would be adjourned based on Mr.
25 Mogil's comments but his would come in. We have no objection

1 to everything coming in.

2 THE COURT: Right. Do you have any objection to
3 theirs? This is a little horse trading we have going on here.

4 (Laughter.)

5 MR. MOGIN: Good enough, Your Honor.

6 THE COURT: Good enough. Good enough. In Chicago
7 this is how we do it.

8 (Laughter.)

9 THE COURT: We let it all in. We'll worry about it
10 later. Okay.

11 (Whereupon, Plaintiffs' Exhibits 2 and 3 and Defendants'
12 Exhibits 1 and 2 were received in evidence.)

13 MR. MOGIN: May I have just a moment please, Your
14 Honor.

15 THE COURT: Sure. No, take your time. Take your
16 time.

17 (Brief pause.)

18 BY MR. MOGIN:

19 Q So you had nothing to do with -- or KPMG had nothing to do
20 with the collection and preservation steps, is that correct?

21 A I'm sorry?

22 Q Is it correct that KPMG was not involved in the collection
23 and preservation steps in this case?

24 A That's correct.

25 Q So you don't know whether the documents that were received

1 were from the proper custodians, correct?

2 A They had -- they provided us with a hard drive of data
3 that had already been collected. That's where we, that's where
4 we got involved.

5 Q And you don't know the locations from where that was
6 collected, correct?

7 A Only based on what we saw from the hard drive with the
8 data on it that had already been collected. I mean, there's
9 obviously mail files and things on there and things like that,
10 so we assumed that they got them from their mailbox.

11 Q But you don't know if backup drives were searched?

12 A We didn't go do the collection.

13 Q You don't know if the share spaces were searched?

14 A We didn't get involved in doing the collection.

15 Q And you're not aware of the time period for which the
16 search was conducted, are you?

17 A I'm sure I've seen it at some point, but I can't recall
18 what it is off the top of my head.

19 Q KPMG didn't have any input into the planning of the search
20 and collection process, correct?

21 A That's correct.

22 MR. MOGIN: All right. Your Honor --

23 THE COURT: Hold on. Hold on.

24 (Brief pause.)

25 THE COURT: Okay. Thank you. I don't have any other

1 questions.

2 MR. MOGIN: Thank you.

3 THE COURT: Okay. Mr. Neuwirth, do you have any
4 more?

5 MR. NEUWIRTH: We have no questions, Your Honor.

6 THE COURT: Okay. Thanks, Mr. Koch, for coming and
7 we hope you can stay. Okay.

8 THE WITNESS: Thank you.

9 (Witness excused.)

10 THE COURT: All right. And you can call your next
11 witness, please.

12 MR. MAROVITZ: Your Honor, Andy Marovitz for
13 Temple-Inland. We'd like to call Dan Regard, please.

14 THE COURT: Okay. Oh, you're not calling Counsel on
15 Call next?

16 MR. MAROVITZ: We were going to call Counsel on Call
17 after Mr. Regard.

18 THE COURT: Okay. Then we'll hear from Mr. Regard.
19 Come on up, Mr. Regard.

20 DAN REGARD, DEFENDANTS' WITNESS, DULY SWORN

21 DIRECT EXAMINATION

22 BY MR. MAROVITZ:

23 Q Mr. Regard, good morning.

24 A Good morning, sir.

25 Q Introduce yourself please to the Court.

1 A My name is Dan Regard.

2 Q Mr. Regard, where are you currently employed?

3 A I work at Intelligent Discovery Solutions in Washington,
4 D.C.

5 Q Is that also known as IDS?

6 A It is.

7 Q What's the business of IDS?

8 A We provide consulting services and technology services in
9 a litigation context to companies and to parties involved in
10 litigation.

11 Q What's your current position at IDS?

12 A I'm currently the CEO and managing director. I'm also one
13 of the two cofounders.

14 Q Have you been retained to reach expert opinions in this
15 case?

16 A I have been, yes.

17 Q Have you, in fact, reached such opinions?

18 A Yes, sir, I have.

19 Q What are those opinions?

20 A I've reached four opinions in this case. No. 1, I've
21 reached the opinion that the search methodologies used by the
22 defendants in this case were consistent with my experience and
23 my understanding of best practices. My second opinion is that
24 the custodial centric approach that the defendants used to
25 identify key players and documents within their organizations

1 is also consistent with my prior experience and with best
2 practices.

3 My third opinion is that the manner in which ESI has
4 been produced by the defendants in this case is sufficiently
5 robust to provide the receiving parties significantly
6 equivalent access to those documents as the defendants had in
7 the usual course of business. And my fourth opinion is the
8 manner in which the defendants have handled the process for
9 considering offline media is consistent with best practices.

10 Q We'll talk about the specifics of those opinions in a
11 moment, but first let's find out about your experience and
12 expertise and your qualifications in order to offer those.

13 Tell us something about your college education.

14 A I have an undergraduate degree in computer science, with a
15 math minor from the University of Southwestern Louisiana. If
16 you look that up, the university has changed its name a few
17 times. It's now the University of Louisiana in Lafayette.
18 After college I spent a few years consulting on computer
19 science and litigation issues. And then I pursued and earned a
20 masters of business from Tulane University, as well as a law
21 degree from Tulane. I have a third -- it's not a degree. It's
22 a certificate of specialty in European legal practice.

23 Q Other than your education, what work, if any, has helped
24 you gain the necessary experience and expertise to lead IDS?

25 A Well, from a technology and litigation perspective, I've

1 been working with technology and computers in a litigation
2 context since the late 1980s, developing financial forecasting
3 and courtroom exhibits. I've had a number of my own companies,
4 including during and after law school for scanning and coding
5 and preparation of document repositories. In terms of
6 leadership, I've been responsible at Deloitte & Touche, FTI
7 Consulting, and LECG, three companies specializing in
8 consulting and expert services, developing and leading
9 regional, national, and international teams for electronic
10 discovery services.

11 Q Mr. Regard, have you ever lectured to lawyers or to law
12 students on eDiscovery issues?

13 A Yes, sir, I have.

14 Q On what occasions?

15 A I'm called upon on a fairly regular basis to lecture at
16 various law schools up and down the eastern seaboard. Most of
17 the time that's George Washington, Georgetown, George Mason,
18 American, Temple, Penn Law. I make presentations for CLE
19 credit or for -- just for presentation purposes to a variety of
20 law firms. And I've been a present lecturer at the Georgetown
21 Advanced Institute for eDiscovery, at the Masters Conference,
22 and at special events put on by Sedona.

23 Q Do you belong to any professional associations?

24 A I do. I belong to the Louisiana Bar Association, although
25 I'm not a practicing attorney. I belong to the American Bar

1 Association and the International Bar Association. I belong to
2 the Sedona Working Group 1 on domestic issues for electronic
3 discovery. I belong to Sedona Working Group 6 on international
4 issues for disclosure, discovery, and privacy. I belong to the
5 High Tech Computer Crimes Investigation Organization. I belong
6 to the International Computer Forensics Professionals
7 Organization.

8 I'm a director of the Georgetown Advanced Institute
9 for eDiscovery. I am on the cabinet of the -- the educational
10 cabinet for the Masters Conference for Legal Professionals.
11 I'm a member of the American College of eNeutrals, a director
12 as well. There may be a few others on my resume.

13 Q Have you ever offered testimony to the Rules Committee?

14 A I have. I had an opportunity to speak to the Rule
15 Committee in Dallas in preparation for the amendments
16 ultimately that made it into the 2006 changes.

17 Q Have you written any scholarly articles on eDiscovery
18 topics?

19 A I've written a number of articles that have appeared in
20 trade magazines. But most recently I was fortunate enough to
21 co-author a chapter on eDiscovery of databases in an eDiscovery
22 desk manual published under the ABA by Judge Grimm and Michael
23 Berman and Courtney Barton.

24 I've participated in Sedona since its inception,
25 Working Group 1. So the original Sedona principles, I

1 participated in the drafting of those. Most recently I've been
2 on the smaller drafting steering committee for the
3 international principles through Working Group 6. It's a lot
4 of stuff. And we also published at the end of last year the
5 database principles for database discovery.

6 MR. MAROVITZ: Your Honor, may I approach.

7 THE COURT: Yes, of course.

8 BY MR. MAROVITZ:

9 Q Mr. Regard, I'm handing you what's been marked as
10 Defendant's Exhibit 3 for identification. Identify for us, if
11 you will, Defendants' Exhibit 3.

12 A This is a copy of my resume.

13 Q Who prepared it?

14 A I did.

15 Q Is it current or very recent in time?

16 A It's very recent. For example, I noticed that some of the
17 publications listed as pending have now been published.

18 Q To the best of your knowledge does it accurately summarize
19 your educational and professional eDiscovery experience through
20 the date on which it was drafted?

21 A Yes, sir, it does.

22 MR. MAROVITZ: Your Honor, we'd like to move to have
23 admitted Defense Exhibit 3 into evidence.

24 THE COURT: Any objection?

25 MR. MOGIN: No objection.

1 (Whereupon, Defendants' Exhibit 3 was received in
2 evidence.)

3 BY MR. MAROVITZ:

4 Q Mr. Regard, what sorts of real world eDiscovery services
5 does IDS provide to companies seeking eDiscovery advice?

6 A Well, let me first clarify we work with individual
7 litigants, not just companies. We provide advice in the early
8 onset of litigation in helping companies identify potential
9 locations for electronically stored information. We help them
10 understand from a computer perspective their architecture. We
11 help them collect information from a variety of sources,
12 whether it's individual computers, laptops, servers. We work
13 with IT departments to export data off of larger systems. We
14 work to define and extract data off of enterprise applications
15 or databases. We also take possession of that data very often,
16 and we will transform it or process it.

17 We provide also hosting services either directly
18 through tools that we license or through partnership with other
19 companies that do the hosting for us or for our clients. We
20 help clients segregate documents. We help them execute
21 searches across large collections of documents. We help them
22 prepare the documents for review. And then we're often called
23 upon to prepare the documents for production, to Bate stamp
24 them, to package them up, to export them.

25 Q Have you ever been qualified as an expert in court in

1 eDiscovery matters?

2 A Yes, sir.

3 Q And has any court ever rejected your qualifications as an
4 expert?

5 A No, sir.

6 Q Have you ever served as a court appointed neutral expert
7 in eDiscovery?

8 A I have.

9 Q On what occasions?

10 A I had an occasion quite a few years ago to be a court
11 appointed forensics neutral in a dispute between two parties to
12 actually go in and execute forensics activities as a neutral.
13 But more recently I was appointed in Washington D.C. as a court
14 appointed special master on an ESI dispute involving an FTC
15 investigation.

16 Q And who was the judge?

17 A Judge John Fasciola.

18 MR. MAROVITZ: Your Honor, may it please the Court.
19 We tender Mr. Dan Regard as an expert witness in the field of
20 ESI.

21 THE COURT: Mr. Mogin, can you do any questions about
22 his qualifications on your cross?

23 MR. MOGIN: Your Honor, I haven't been provided
24 sufficient information to do so, but I will attempt to do so
25 and I will not at this point stipulate.

1 THE COURT: Good. Thank you.

2 MR. MOGIN: And I would repeat the arguments that we
3 made earlier in our filing regarding legal opinions dressed up
4 as ESI opinions.

5 THE COURT: Right. And I took it very seriously.
6 And certainly in a decision will take that into consideration.
7 For today -- we just have so much on our plate here today. I'm
8 not going to try to distinguish one from the other. Okay. All
9 right. Do you agree with that? Whether you agree with it or
10 not, that's the ruling.

11 (Laughter.)

12 MR. MOGIN: I understand, Your Honor, but that
13 basically requires me then to say I will just have a running
14 objection and I won't be jumping up.

15 THE COURT: I was just going to say you can certainly
16 make objections if, particularly if something's -- your record
17 is preserved that you -- in general it's preserved, okay.

18 MR. MOGIN: So shall I jump up and make objections?

19 THE COURT: Yes, start off jumping up. Why not.
20 Okay. Try not to have legal conclusions, though, because that
21 is frankly my job.

22 MR. MAROVITZ: We agree, Your Honor.

23 THE COURT: Okay. Thank you.

24 BY MR. MAROVITZ:

25 Q Mr. Regard, what role have you had in defendants'

1 discovery protocol here?

2 A Well, I was first hired by the defendants sometime in the
3 spring or summer of last year.

4 THE COURT: The defendants or a defendant?

5 THE WITNESS: A single defendant, Your Honor.

6 THE COURT: And who was that?

7 THE WITNESS: That would have been Temple-Inland.

8 THE COURT: Oh, Temple. Okay. Thank you.

9 THE WITNESS: Temple hired me to advise them on some
10 of their self-collection techniques and the tools that they
11 were using, which I did. And then more recently I was brought
12 back into this case in order to provide the opinions for
13 today's hearing.

14 BY MR. MAROVITZ:

15 Q What information did you receive with respect to the ESI
16 process?

17 A With respect to preparation for this hearing?

18 Q And to render your opinions in this case.

19 A I've been granted copies and access to correspondence, to
20 pleadings, to document requests. I've been granted copies of
21 30 (b) 6 disclosures made by the defendants, and I've had an
22 opportunity to talk to the defendants' -- counsel for the
23 defendants and consultants for the defendants.

24 Q And how did you go about obtaining that information that
25 you've just described?

1 A Well, the documents were provided to me by counsel for
2 Temple-Inland. The conversations, I was permitted access to
3 scheduled conference calls and speak to them directly.

4 Q Did you receive sufficient information first with respect
5 to Temple-Inland and its consultants to allow you to evaluate
6 its ESI process?

7 A Yes, sir.

8 Q Would you have simply assumed as part of your work that
9 all the other defendants' ESI processes were the same as
10 Temple-Inland's?

11 A Well, there's a certain amount of cooperation is my
12 understanding that the defendants engaged in to develop their
13 initial set of search terms, so there was some similarity on
14 the work product. Downstream from that I found differences,
15 and I did not make the assumption. I think the differences are
16 spelled out in the 30 (b) 6 disclosures. And it's also what I
17 confirmed or learned in the course of my conversations with the
18 individual defendants.

19 Q And by the 30 (b) 6 disclosures, do you mean the letters
20 that counsel sent to the plaintiffs to describe processes in
21 response to a 30 (b) 6 deposition notice received by the
22 defendants?

23 A That's my understanding. Both the letters and I believe
24 in the case of some of the defendants, at least one, subsequent
25 addendums or additional information sent.

1 Q With respect to all of those other defendants, did you
2 receive sufficient information to allow you to evaluate their
3 ESI processes?

4 A Yes, sir.

5 Q Mr. Regard, are you being paid for your testimony today?

6 A I am.

7 Q At what rate?

8 A 525 per hour.

9 Q Does that rate or the amount that you're paid depend in
10 any way on the outcome of this hearing or the outcome of this
11 case?

12 A It does not.

13 Q I want to talk about ESI best practices now with you. Mr.
14 Regard, what are the industry best practices by which ESI
15 gathering and production are measured?

16 A Well, I think it's the principles that we measure them by.
17 As to best practices, we execute to achieve those measurement
18 principles. The principles, and there are many, but the ones I
19 find are relevant here are the first principle is that we
20 expect that parties will do a diligent job to find electronic
21 discovery, but there's not an expectation they will find every
22 single document.

23 MR. MOGIN: Objection, Your Honor. I think when
24 we're talking about -- he's divided this into principles and
25 best practices. I think that principles are derived from law,

1 and that's a legal matter. Now, the best practices that are
2 used to implement the law or to comply with the law is a
3 different issue. But I don't think that this witness can
4 testify as to the principles.

5 MR. MAROVITZ: I disagree, Your Honor. And why don't
6 you give us -- Your Honor with your indulgence, we could ask
7 where those principles are derived.

8 THE COURT: Well, I wanted to know where they came
9 from.

10 MR. MAROVITZ: Sure. So Mr. --

11 THE COURT: So let's figure out where they came from
12 first.

13 THE WITNESS: Well, when I talk about principles, I
14 rely heavily on the work that I've participated in at the
15 Sedona Conference. But not everything that we've done at the
16 Sedona Conference has been written up as a -- under the title
17 principle. We have a number of papers that we've published
18 that have recommendations, best practices, or scholarly work on
19 the area of electronic discovery. So that when I say a
20 principle, I'm talking generically what I derived from the work
21 that we'd done in the Sedona Conference and in other
22 conferences and think tanks I participate in.

23 The Sedona Conference is --

24 THE COURT: Well, Mr. Mogin -- let me ask Mr. Mogin.
25 Well, when a person is both a consultant expert in an area, and

1 that seems like Mr. Regard's primary work is in eDiscovery or E
2 consulting. When he's using the word principle, I mean what --
3 tell me what your objection is here so just the record is
4 clear. Tell me what your objection is.

5 MR. MOGIN: I think that principles, Your Honor --
6 there's a distinction between principles and best practices.
7 And that when you unpeel the onion, that principles really are
8 matters of law. That is what the Courts and Congress have said
9 are the requirements of law that apply to discovery whether
10 it's eDiscovery or other discovery. Now, how that gets
11 implemented would be best practices. So if I could use an
12 example.

13 Miranda in a criminal context would be a principle.
14 The giving of a Miranda warning as proper police procedure
15 would be a best practice.

16 MR. MAROVITZ: Your Honor, if I may.

17 THE COURT: Yes.

18 MR. MAROVITZ: I think Mr. Mogin's testifying. The
19 fact is that Mr. Regard is prepared today to explain where the
20 material that he replies upon comes from, the life's work he's
21 done in ESI, and the fact that he is not trying to take your
22 job in interpreting what the law is. The Court is well able to
23 interpret his opinions today and apply it to the law.

24 THE COURT: I think that it was his choice of words
25 that started this debate. I mean, I think if he is -- I think

1 what Mr. Mogin said is if he is talking about best practices,
2 okay, he can talk about best practices. Then on
3 cross-examination we're going to find out if -- you know, what
4 does he base his best practices on, I think. Is that correct?
5 Is that what you were saying?

6 MR. MOGIN: Depending, yes, Your Honor.

7 THE COURT: And his experience. I mean, I think
8 that's it. I think it was the principle word that got us
9 tripped up. Okay.

10 MR. MAROVITZ: Let me reload.

11 THE COURT: Yes.

12 BY MR. MAROVITZ:

13 Q We were talking a minute ago about the industry best
14 practices by which ESI gathering and production are measured.
15 And to provide context, I think your answer in general was that
16 there were several. One of which was that you can't expect to
17 get every single document wherever it's located in the company.
18 So was that right, and what were the other ones that you were
19 testifying to?

20 A So that is correct. As a best practice litigants and
21 producing parties should strive to produce as much electronic
22 discovery as possible, but must realize they cannot produce a
23 hundred percent of every document because of the complexity and
24 diversity of our systems. That No. 2, as a best practice
25 litigants should strive, or parties should strive, not

1 necessarily litigants -- it could be third parties -- to use
2 technology to cope with the volumes of electronic discovery and
3 electronic documents that we have today. And that also as a
4 best practice when that technology is used, there should be
5 some type of quality assurance testing to verify or to satisfy
6 concerns that the technology is applied correctly.

7 Q And does that last step that you spoke of, does that
8 include in some cases testing the remaining corpus or what's
9 been called today the null set?

10 A That has more recently emerged as a best practice.
11 Absolutely.

12 Q Now, with those best practices in mind, let's discuss what
13 you did in this case to determine whether or not they were met.
14 Please provide the Court with a general overview of what you
15 learned with respect to defendants' processes here.

16 A Well, a lot of what I'm going to talk about has already
17 been described in the disclosures and the correspondence, but
18 I'm happy to repeat it. My understanding is that during the
19 course of the litigation the filing of the complaint, the
20 correspondence, the pleadings, and the requests for production,
21 that the defendants have assembled together a list of terms
22 that have been made known to them through the allegations and
23 requests. And that I call this in my own vernacular the seed
24 set of search terms and queries have been put together through
25 a collaboration with the various defendants. And actually the

1 task was taken by Georgia Pacific to test those.

2 And that testing, as earlier testimony indicated, was
3 an iterative process where search terms were applied against a
4 corpus of documents. The results were examined. The null set
5 was examined. The search terms were revised and tested again.
6 At some point in that process the search terms, it's my
7 understanding, were shared with the plaintiffs. Feedback was
8 received on the quality or perceived lack of quality of those
9 search terms, and they were modified as a result. And then at
10 some point subsequent the search terms, again the seed set now
11 revised and modified, was distributed to different defendants.
12 And each of the defendant groups took those search terms,
13 adapted them to their individual organizations, and ran them
14 against information they had collected as potentially
15 responsive to this litigation.

16 And that when they finished applying these search
17 terms to that collection that they had, that each of the
18 defendants in their own way tested the residual documents,
19 again their individual null sets, if you will, as a verifica-
20 tion process to collaborate the success of the search terms.

21 Q Mr. Regard, I'm handing you what's been marked as
22 Defendants' Exhibit 4 for identification.

23 A And I just said collaborate. I should have said
24 corroborate.

25 MR. MOGIN: Again, Your Honor, for the record this is

1 not a document we have seen before.

2 BY MR. MAROVITZ:

3 Q Mr. Regard, do you recognize Exhibit 4?

4 A Not in this size. Yes, sir, I do recognize it.

5 Q My eyesight is getting worse, and so we've got to make
6 things wider.

7 A It is unfortunately.

8 Q What do you recognize Defense Exhibit 4 to be?

9 A This is an exhibit that I had pulled together. I actually
10 set it up, and my team and I helped pull this document
11 together. This reflects some of the information I gathered
12 from the pleadings, the disclosures, and my conversations with
13 individual defendants.

14 Q Did you oversee the document's creation?

15 A I did.

16 Q Is Defense Exhibit 4 a fair and accurate summary of the
17 facts elicited during your interviews and that you culled from
18 the documents that you reviewed that relate to the issues that
19 are contained on Exhibit 4?

20 A It is.

21 Q And to the best of your knowledge are the facts and -- are
22 the facts that are contained on Exhibit 4 accurate?

23 A To the best of my knowledge, yes.

24 MR. MAROVITZ: Your Honor, we move for admission of
25 Defense Exhibit 4 into evidence.

1 MR. MOGIN: Your Honor, we object. This is a bit of
2 an ambush to be frank about it.

3 THE COURT: It's what?

4 MR. MOGIN: An ambush. We were supposed to have
5 received the demonstratives on last Thursday. There was no
6 indication that we would receive a document anything like this.
7 We have no idea when this document was compiled, and we can't
8 at this point in the proceedings even check the accuracy of it.
9 I note that there's some statements about disclosures, about
10 the search terms by the defendants to the plaintiffs. And in
11 my view it's grossly inaccurate. And if we had the time and
12 the ability, we could go through and we can prove it. But
13 having been ambushed, having just received this document, we
14 don't have any ability to do that. We'd object to the
15 admission of this document.

16 MR. MAROVITZ: Your Honor, if I may.

17 THE COURT: Yes.

18 MR. MAROVITZ: A few things. First we mentioned --
19 when Your Honor set up this hearing, Your Honor mentioned that
20 exhibits could if they had already been prepared be turned over
21 in advance, but that there may be some that weren't fully
22 prepared. This wasn't fully prepared at the time of Thursday.

23 Second, the information that's contained on the
24 exhibit for the most part has already been shared with the
25 plaintiffs in the 30 (b) 6 and other letters. This is a

1 summary document, Your Honor. And it's fine with us frankly if
2 plaintiffs wish to reserve on this. I just want to make sure
3 that I make the motion to get it admitted into evidence. We
4 don't have a jury here, Judge. It's simply a matter of using
5 it with this witness to allow the witness to identify and
6 describe what he did.

7 MR. MOGIN: Your Honor, there's an awful lot of
8 detailed information in this that the plaintiffs would simply
9 need to look at to verify before.

10 THE COURT: Well, I assume this is like an outline of
11 what Mr. Regard's testimony is going to be about what he did.
12 This is kind of his help to -- this is like his exhibit book.
13 So the first is I think he's able to say on the stand what he
14 did. Okay. I do think he can say what he did. So I'm going
15 to reserve ruling on the admissibility of this.

16 Now, I don't -- I wouldn't use the word ambush. But
17 just as I said clearly on Friday that -- to the plaintiffs that
18 I thought they didn't have enough notice on Miss Tenny. Okay.
19 I do think we could have at least been told that he is going to
20 bring a spread sheet that involves six defendants. And I have
21 a question. Why isn't GP on this? I mean, GP is not on this.
22 So it's not in -- did he only review the six people who are the
23 six companies that are on here?

24 MR. MAROVITZ: Right. His -- the answer to that
25 question, Your Honor, his testimony primarily will be for the

1 other six because the first witness --

2 THE COURT: GP had their own.

3 MR. MAROVITZ: Right. Has already testified
4 generally as to the GP process. Secondly, I just -- I bristle
5 a little bit at the charge that this is an ambush. We wrote in
6 our letter --

7 THE COURT: I didn't, I didn't use --

8 MR. MAROVITZ: No. No. I'm not --

9 THE COURT: -- Mr. Mogin's word.

10 MR. MAROVITZ: I'm well aware of that.

11 THE COURT: Okay. But I am saying that, I am saying
12 that they at the last minute wanted to add something for today,
13 and I didn't allow them to do it. So I think his point is well
14 taken. But we'll just --

15 MR. MOGIN: Your Honor, I don't care to wrangle about
16 this on the record in light of the time limitations that we
17 have.

18 THE COURT: Right.

19 MR. MOGIN: But you have reserved ruling.

20 THE COURT: I have reserved ruling.

21 MR. MOGIN: And I would urge you before ruling to go
22 back and please take a look at the letter that Mr. Marovitz
23 submitted where he describes the testimony that Mr. Regard is
24 going to give.

25 MR. MAROVITZ: That's fine.

1 THE COURT: I think the content of what he's going to
2 say, I mean, I'm going to let -- there is a difference between
3 an exhibit and the content. I don't think the content is a
4 surprise. I think the exhibit and pulling it all together is
5 quote, unquote were. But let's start with the content because
6 it's now 11:20.

7 MR. MAROVITZ: Very good, Your Honor.

8 BY MR. MAROVITZ:

9 Q Mr. Regard, let's jump directly to the content.

10 A Yes, sir.

11 Q Tell us exactly what you did in terms of evaluating the
12 defendants' ESI protocol.

13 A Well, No. 1 I wanted to understand the process that I
14 recently described prior to this dialogue. Basically the
15 process they went through in originating search terms that were
16 ultimately used, the process used to modify those, to
17 reconsider them, to test them at the first level. And then as
18 they were distributed to individual defendants to understand
19 how the defendants tested those against the null set. And
20 really that was the most important thing to me.

21 Because as I learned in this process and that I have
22 seen in other cases, individual parties can follow a variety of
23 paths from the beginning of the collection of ESI to the
24 ultimate production. But it's the testing of the application
25 of technology that's of particular importance today. And so it

1 was the testing of the null set that I thought was the most
2 important part of corroborating, corroborating the ability of
3 defendants to rely upon their processes. And so this chart
4 reflects my focus on gathering from the defendants the specific
5 information as to the null set testing that they conducted.

6 Q Mr. Regard, plaintiffs claim that defendants' reliance on
7 keywords at the outset of this process compromised the results.
8 Do you agree with that?

9 A I do not.

10 Q Why not?

11 A Two reasons. At least two initially. No. 1, in my
12 experience it's been very common, in fact, in every case that I
13 reflect upon keywords have been used in one fashion or another.
14 And I find that keywords are very common in dealing with the
15 volumes of ESI that we deal with today. There was some earlier
16 testimony about filtering of documents. The filtering of
17 documents by date ranges is a type of a keyword. The selection
18 of custodians is a type of a keyword or key players.

19 Again, the use of keywords to help separate documents
20 from that which is a viable corpus of documents to that which
21 is obviously or sometimes not so obviously not relevant is a
22 very useful technique. We use keywords both to include
23 documents as well as to exclude documents in shaping up the
24 document corpus.

25 The second reason is because the keywords the way

1 they have been used in this particular case I find have
2 followed best practices. Best practices in that there was
3 significant human input at the beginning. Counsel
4 collaborated. There was significant review of documents from
5 both keyword hits as well as nonhits. And again, I'm repeating
6 some of the testimony from this morning from Georgia Pacific
7 which drove through KPMG and Counsel on Call the testing of the
8 initial set. That iterative processing, again is a best
9 practice to try the keywords, to look at the results, to use
10 the results to modify the keywords and try them again and
11 again.

12 And then finally when the keywords were distributed
13 to defendants, they were modified again and tested. So I find
14 the application of keywords both appropriate in my experience
15 and the way that they were applied consistent with best
16 practices.

17 Q Have you undertaken any independent studies of -- well, of
18 anything that relates to keywords in your area of expertise?

19 A Well, I mentioned earlier that my background is in
20 computer science, my undergraduate degree. One of the things
21 that we pride ourselves on at IDS is that we have a computer
22 science focus. In fact, sometimes we're called upon to testify
23 on issues of not ESI but computer science and how computers
24 work at both the forensics level, various enterprise
25 applications, et cetera.

1 As a computer science student, I've gone and studied
2 academic papers on various forms of machine learning and
3 information retrieval, the TREC papers. But more importantly
4 papers going back to space vector machines back in the 1960s
5 through the '70s, the '80s. Latent semantic indexing,
6 probabilistic latent semantic indexing, latent Dirichlet
7 allocations, and more recently work by Google and Yahoo on
8 search engine optimization.

9 Q So that goes to the first point about keywords, your
10 background and experience in using them. The second point that
11 you made was that essentially they comported with best
12 practices here. I want to ask you a question about that.

13 How is it in this case that you understand the
14 defendants used keywords?

15 A In my understanding in this case the keywords have been
16 used to create a, I'll call it a perimeter of documents from
17 which to review for production of ultimately documents
18 responsive to the document requests. And I use the word
19 perimeter because that's really what it has been. The
20 documents that were collected -- and that's just an artifact of
21 the collection process. Because of the way we collect
22 documents, tend to overcollect documents always. When you
23 collect an entire PST or an entire my docs folder, you're
24 overcollecting.

25 And so we use filtering processes of it was mentioned

1 earlier file extensions, date ranges, and custodians. It's
2 also very useful, and again, in my experience very common for
3 parties to create a perimeter of documents by using keywords
4 and then taking that corpus and reducing it further through
5 document review.

6 Q Were the -- oh, I'm sorry. Mr. Regard, go ahead. I
7 didn't mean to cut you off.

8 A Well, I was going to add to that, one of the indicators of
9 the perimeter is responsiveness or the precision of documents
10 that are inside the perimeter. I know through my talks, my
11 discussions with the various defendants, that once this
12 perimeter is drawn, the tighter you draw the perimeter the more
13 that you can exclude noise documents but also the greater the
14 likelihood is you might exclude relevant documents.

15 And so when I look at things like precision, I look
16 at how broad is the perimeter. And in this case from my
17 discussion with the defendants, the perimeter has been fairly
18 broad which has resulted in what we call the precision being
19 relatively low. And so again that confirms to me this was a
20 conservative perimeter that was drawn to identify documents for
21 subsequent document review.

22 Q Based upon all of this, Mr. Regard, do you have an opinion
23 on whether defendants' search protocol, which included but was
24 not limited to keywords, met or exceeded best practices?

25 THE COURT: Now I'm in a dilemma. So far Mr. Regard

1 has been talking about in general word search. You get down to
2 the specifics here, and you keep saying the defendants. I mean
3 we did not know, and I did read every piece of paper in this, I
4 did not know how many custodians each -- I didn't know any of
5 this factual information.

6 MR. MAROVITZ: That's a fair point, Your Honor.

7 THE COURT: I mean, I really don't know whether --
8 you know, what Mr. Regard has been saying so far has been
9 pretty general. I don't know whether we shouldn't bring Mr.
10 Regard back or do Mr. Regard a separate time so that the
11 plaintiffs can have a chance to absorb some of this
12 individual -- this is he's now testifying about six separate
13 systems here. This doesn't seem fair when they got this chart.

14 MR. MAROVITZ: If I may, Your Honor, a couple things.
15 First I'd be happy to walk through with Mr. Regard the
16 information so that it's clear on the record.

17 THE COURT: Well, did you know the underlying -- let
18 me say maybe it's just me. And I mean be straight. Did you
19 know these specifics?

20 MR. MOGIN: No, Your Honor.

21 MR. MAROVITZ: Your Honor, we've identified our
22 custodians months and months and months ago.

23 THE COURT: Well, we know. But if -- I mean, what
24 the last witness said is there has to be some kind of
25 statistical, if you will, if we're going to find out if this

1 method is valid, accurate as much as human beings can make it.
2 I mean I can't even read this and listen to Mr. Regard at the
3 same time.

4 MR. MAROVITZ: Well, I'd be happy, Judge, if you
5 will, to have him walk through what these things mean so that
6 it's clear on the record for the Court. There really -- we
7 tried to make this a summary. It may be the case that there
8 are stray things here and there that were not provided
9 previously, but this is a summary of the give and take that the
10 plaintiffs and defendants have had over a series of months.
11 And we prepared it really to orient, as Your Honor pointed out,
12 Mr. Regard's testimony here. So I think it's a great
13 observation that in order to enlighten this, to allow him to go
14 through and identify what each of these things means. We'd be
15 happy to do that. We want the record to be clear about this.

16 MR. MOGIN: Your Honor, if I might. Both in the
17 descriptive letter from Mr. Marovitz of February 16th, as well
18 as in their defendants' motions they tell us that Mr. Regard is
19 going to testify concerning the testing and validation of
20 search terms. We don't have that testing. We don't have that
21 validation. And the first time that we've seen the statistics
22 is in this document that counsel is trying to introduce. We're
23 simply not prepared to deal with this witness without that sort
24 of underlying information.

25 MR. MAROVITZ: Judge, just to be clear, Miss Miller

1 just handed me, tab 2 is a good example.

2 THE COURT: Of what? Tab 2 of what?

3 MR. MAROVITZ: Tab 2 of our initial -- exactly. The
4 brief contains 44 tabs essentially of all of the background of
5 the back and forth between the parties. There are letters from
6 all defendants, from plaintiffs. Tab 2 is an example of some
7 of that back and forth, including the custodians.

8 THE COURT: The August 11th letter, is that what --

9 MR. MAROVITZ: Pardon?

10 THE COURT: You're talking about an August 11th
11 letter?

12 MR. MAROVITZ: That's exactly right.

13 THE COURT: Okay.

14 MR. MAROVITZ: So if, Your Honor, if you -- for
15 instance, if you go towards the back, you could see all of the
16 different custodians who were there first from Georgia Pacific
17 and from Norampac. Then from PCA. Then from Rock Tenn. And
18 then from our client Temple-Inland. These were provided to the
19 plaintiffs back in August of 2011. There's a series of
20 additional letters that were provided back and forth between
21 the parties. As I say, it provided them a lot of this
22 information.

23 And I really, I guess I would ask the Court simply
24 to -- since Mr. Regard is here, to hear him out. Plaintiffs
25 can cross him. If plaintiffs can make a showing later that

1 there's some real material information here that they haven't
2 had access to or couldn't have had access to, then the Court
3 certainly will act on that. But --

4 MR. MOGIN: Your Honor, the entire statistical
5 presentation at the bottom of this chart is brand new
6 information.

7 THE COURT: All right. So here's what I think we
8 should do. They've got another hour how they want to use their
9 time. If you want, you could do more of the direct of
10 Mr. Regard, but definitely Mr. Regard must have other cases in
11 Chicago. We'll do the cross another time when they have some
12 time to hear it, or you can put on CAC and we'll do, we'll do
13 the whole thing another day. I mean, I don't know what else to
14 do because I'm surprised, and I don't have to get up and do the
15 cross.

16 MR. MAROVITZ: Right. Well, Your Honor, apologies to
17 both you and to plaintiffs' counsel for the surprise. We had
18 written in our letter just to be clear that any additional
19 exhibits which had not yet been prepared will be provided
20 Tuesday morning.

21 THE COURT: It's not a technical -- it's not
22 technical.

23 MR. MAROVITZ: Okay.

24 THE COURT: I mean, some exhibits are nothing. Okay.
25 So then that's kind of what -- I mean, this is -- this could be

1 a whole day testimony here.

2 MR. MAROVITZ: Well, if you'll permit me --

3 THE COURT: You want to talk to about the process --

4 MR. MAROVITZ: If you'll permit me --

5 THE COURT: You want to talk about Mr. Regard's
6 process a little bit, kind of give a little bit overview of the
7 process or you want to call CAC? Whatever you want to do.

8 MR. MAROVITZ: If you'll permit me 30 seconds.

9 THE COURT: Yes. Sure.

10 (Brief pause.)

11 THE COURT: Yes, sir.

12 MR. MAROVITZ: Judge, thank you for your indulgence.
13 I think with your permission what we'd like to do, we think it
14 would be important to get Mr. Regard's testimony in at one
15 time. We're a little concerned about breaking it up. And
16 frankly the way we have this set is we were going to have Mr.
17 Regard go and then have Sandy go after that. So it probably
18 makes sense if Mr. Regard has to come back anyway for us to
19 essentially stop the defendants' presentation here, allow the
20 plaintiffs to put their experts on, and then set another date
21 for our two witnesses to be completed.

22 THE COURT: Well, I think -- here's the other thing
23 is I know the least about these other six parties. Because,
24 you know, there was a choice made before I got into the case
25 that GP was kind of taking the lead or something. So I

1 factually am interested to know kind of what went on with the
2 six. I'm sorry, Mr. Regard.

3 THE WITNESS: Not at all, Your Honor.

4 THE COURT: But you get another great trip to
5 Chicago, okay.

6 THE WITNESS: Yes, ma'am.

7 THE COURT: So anyway -- can you stay, though? Are
8 you staying today --

9 THE WITNESS: Yes, ma'am.

10 THE COURT: -- at least part of the day anyway. I
11 think you ought to call CAC, okay. With the understanding that
12 we're going to recall Mr. -- we're not going to take back
13 everything he said here today.

14 MR. MOGIN: Of course.

15 THE COURT: But we'll give you a do over.

16 MR. MAROVITZ: We appreciate it, Your Honor.

17 THE COURT: Okay.

18 MR. MAROVITZ: And I don't know if Your Honor -- our
19 preference would be to call Counsel on Call after Mr. Regard.
20 So we wonder whether it would make more sense simply to move to
21 the plaintiffs' presentation now. Currently our Counsel on
22 Call examination is relatively brief. So the ordering that we
23 have would be Mr. Regard and then Counsel on Call, and we'd be
24 happy to turn the floor over to the plaintiffs now for their
25 experts so we can do both of them on another day afterwards.

1 And Counsel on Call could be here on another day.

2 THE COURT: No. No. Why aren't we going to do
3 Counsel on Call right now?

4 MR. MAROVITZ: It's just -- it's the way that we had
5 it set up, but we'll -- if Your Honor would prefer it that way.

6 THE COURT: You were going to do Counsel on Call
7 after their presentation?

8 MR. NEUWIRTH: No.

9 MR. MAROVITZ: No, ma'am. What we had intended to do
10 was to have Mr. Regard go first.

11 THE COURT: Right.

12 MR. MAROVITZ: And then have Counsel on Call go.

13 THE COURT: Well, that's what I'm saying. So why
14 don't do -- just do -- we've got 45 minutes on Counsel on Call.

15 MR. NEUWIRTH: Your Honor, we're certainly happy to
16 proceed in whatever way Your Honor determines makes sense. I
17 think that what Mr. Marovitz was trying to suggest to the
18 Court, but we'll proceed as you say, is that we think that the
19 length of Mr. Brown's testimony from Counsel on Call could be
20 reduced, and it would make sense logically to have Mr. Regard's
21 testimony prior to Mr. Brown's testimony. And so I think what
22 the suggestion was that if Mr. Regard is going to come back on
23 another day, to have Mr. Brown come back on --

24 THE COURT: Didn't Counsel on Call just do GP?

25 MR. NEUWIRTH: Yes.

1 THE COURT: Did they do everybody?

2 MR. NEUWIRTH: They did GP.

3 THE COURT: Well, then --

4 MR. MOGIN: Well, Your Honor, if I can refer you to
5 the chart --

6 MR. NEUWIRTH: We can proceed.

7 MR. MOGIN: -- we've been discussing --

8 MR. NEUWIRTH: We can proceed.

9 MR. MOGIN: -- it appears that Counsel on Call was
10 also involved with another one of the defendants.

11 MR. NEUWIRTH: But not Mr. Brown.

12 THE COURT: Not Mr. Brown.

13 MR. NEUWIRTH: So we'll proceed however Your Honor
14 wants. And if you would like to hear Mr. Brown now, we can.

15 THE COURT: Let me -- you know, I do what I do all
16 the time incorrectly. Mr. Mogin, is this the way you want to
17 proceed? This is actually your hearing. Do you want Mr.
18 Regard just to continue on today, do your cross, and we'll be
19 finished? Is that what you want?

20 MR. MOGIN: Quite frankly, Your Honor, I'd like to
21 hear Mr. Brown's testimony. We don't have -- we have very
22 little information from Mr. Brown, but we're ready to proceed.

23 THE COURT: Well, I see no reason not to do Mr. -- I
24 mean Mr. Brown seems like a follow-up from the first witness.

25 MR. NEUWIRTH: We're happy to go, Your Honor.

1 THE COURT: So I think that the problem is all the
2 information that came with Mr. Regard just was a little fount
3 of information here that --

4 MR. NEUWIRTH: We're ready to go, Your Honor.

5 THE COURT: Okay. So I say, Mr. Regard, you step
6 down and let's call Mr. Brown.

7 MR. FREED: Your Honor, if I may. Michael Freed.

8 THE COURT: Yes.

9 MR. FREED: It may be implicit in what you have said,
10 but before Mr. Regard gets his next opportunity to visit
11 Chicago this winter, we would like to get the kind of
12 information that we will need in order to cross-examine him at
13 that time. So we're going to want to have an opportunity to
14 learn more about what he's done. Now, that -- and that's the
15 dilemma. He doesn't have to do it now if they can give us the
16 information before he returns. So if they wanted to go through
17 what he has done up until now, because otherwise we're not
18 going to be able to deal with what's on this chart anyway.

19 So I'm saying maybe we -- and I apologize to Mr.
20 Marovitz. It's his witness. But maybe we should learn what he
21 did and then recall him after that so at least we'll be working
22 off some information.

23 THE COURT: You know, one of the reasons expert
24 discovery is different than other discovery, and there
25 really -- and we were -- we're definitely in a hybrid situation

1 here, because this was not following all of the rules, but it's
2 exactly the reason that experts unlike many laypeople are going
3 to be testifying to more complicated subject matter. So if you
4 wanted to ask some preliminary questions of Mr. Regard about
5 his process, that would help us for the next time. Because I
6 certainly don't want this to happen the next time.

7 I think what he said so far is he got hired. He
8 worked with -- he worked with the defendants early in the case
9 and then he came back in order to opine on their search
10 methodology. Right.

11 MR. MAROVITZ: I think that's right. Judge, I just
12 want to -- you're exactly right. This is an unusual situation.
13 We didn't have --

14 THE COURT: But I'm not even being critical. What
15 I'm saying is let's use a couple minutes here since Mr. Regard
16 is here. He can tell us kind of what he did. No cross of it.
17 And then when we have him back again at least they'll know what
18 the process was. Thank you.

19 MR. MAROVITZ: That's fine. Thank you, Your Honor.
20 I appreciate it. We're all finding our way in this new regime.

21 THE COURT: Right.

22 BY MR. MAROVITZ:

23 Q Mr. Regard, you get all that?

24 A Luckily I have a trailblazer in the form of a competent
25 counsel, so I'll follow the questions and we'll do the best we

1 can, Your Honor.

2 Q Very good. Let's -- just to give an idea of the way this
3 worked, and I know counsel and the Court will guide me if I go
4 astray. Let's take a look at the Temple-Inland line --

5 A Yes, sir.

6 Q -- on the chart. Maybe you can walk through the
7 Temple-Inland line there and give the Court an idea of what you
8 did and how you did it.

9 A And maybe I can clarify. This information as to the
10 defendant Temple-Inland, counsel, the consultants that worked
11 on the Temple-Inland matter in terms of processing and
12 participating in the application of search terms, the
13 technology that Temple-Inland used, the number of custodians
14 that Temple-Inland has collectively collected, and then the
15 smaller number -- a set of custodians that I would call the
16 first level key players that were used for the testing of
17 search terms.

18 A footnote about the totality of the custodians, that
19 it includes assistants. In this case my understanding is
20 assistants who worked as assistants to various officers or
21 people with responsibilities within the organization. Whether
22 or not search terms were used in the collection of documents
23 from the native systems. When modifications or search terms
24 were disclosed to the opposing party, the requesting party.
25 Various footnotes of information that I learned from the

1 pleadings, the correspondence or through discussions. And then
2 a summary at the bottom.

3 Again, like each of the defendants I spoke to,
4 Temple-Inland received the search terms from the collective
5 effort of the defendants driven by Georgia Pacific to test and
6 then modified those for Temple-Inland. Temple then used
7 information that had been collected from custodians but culled
8 down through de-duplication, through deNISTing, which was
9 described with the earlier testimony, to a corpus of documents
10 that they then applied the search terms against and that
11 created a null set. So they took a corpus of documents. They
12 applied the search terms, they created a hit set, and the
13 documents that were not hits we called the null set. And then
14 from the null set they applied a random selection of documents.
15 They reviewed those documents. So that would be the 500, Your
16 Honor, if we're on the same place in that first column.

17 And then they reviewed those 500 randomly chosen
18 documents, of which they found 7 of them in their opinion to be
19 responsive to their document requests. And then I calculated a
20 percentage, which I also believe has been calculated and
21 produced in various correspondence by Temple-Inland. I didn't
22 create the result. I merely reperformed the mathematical
23 calculation of the percentage of potentially relevant documents
24 they found in the random sample chosen from the null set.

25 My goal in creating this column of information was

1 more to make sure that I had at my reference the information
2 that I found in these pleadings and discussions than to create
3 a novel calculation. My other goal in this was to document to
4 some small degree for myself again as a reminder of the process
5 they went through, the iterative process both in stage 1 and in
6 stage 2 of developing these search terms and then applying them
7 to a corpus of documents that ultimately was used for document
8 review.

9 Q And maybe, Mr. Regard, we'll await another day on the
10 specifics of all the other columns that are here. But
11 generally as you've just put it, is that the general process
12 that you used with respect to the defendants in this case,
13 obviously other than Georgia Pacific?

14 A Well, again there are nuance differences as to how many
15 documents they started with and the way that they modified the
16 search terms for their individual organization or the software
17 that they were using, but that is what I have learned is
18 generally the way that each of the defendants handled their
19 approach to this stage of document review and production.

20 MR. MAROVITZ: Your Honor, I have -- maybe I can
21 propose that we proceed in this fashion. I have a number of
22 questions that I would ask to Mr. Regard that are not really
23 related to this issue. And if you'd like me to reserve those,
24 I can. Or if you'd like me to fire away, I can do that.

25 THE COURT: Why don't you start off. I mean, they

1 don't have to do with this issue, but --

2 MR. MAROVITZ: It really doesn't.

3 THE COURT: Okay.

4 MR. MAROVITZ: If we can reserve the rest of that
5 issue for another time.

6 THE COURT: Right. Okay.

7 BY MR. MAROVITZ:

8 Q Mr. Regard, without getting into the specifics of what you
9 did that we'll talk about on another day, how does defendants'
10 search methodology compare to the content based advanced
11 analytics that the plaintiffs have offered in this case? Let
12 me ask you a --

13 A There's a lot in that question.

14 Q Yes. Let me ask you a more detailed question. Are you
15 familiar with the phrase concept based searching or other
16 methods of computer based advanced analytics?

17 A I am familiar with the idea of content based analysis and
18 analytics. The exact phrase content based advanced analytics
19 is new to me in this case. It doesn't surprise me. I've seen
20 a lot in the marketplace of various companies and software
21 providers promoting content based analytics. It's a common
22 term in marketing. It has a variety of different definitions.

23 But if one takes the testimony this morning to say
24 it's based on somehow analyzing the words in a document and
25 using that to make decisions about that document or to

1 understand the decisions on that document combined with the
2 analysis to make a projection of what another document might
3 be, then yes, I'm very familiar with that.

4 Q Plaintiffs say in their reply brief that when compared
5 head to head plaintiffs' content based search methodology is
6 far superior in this case to defendants' Boolean search
7 methodology. Do you remember seeing that?

8 A I recall seeing that, yes.

9 Q Do you have an opinion on whether plaintiffs' content
10 based search methodology is far superior in this case to
11 defendants' Boolean search methodology?

12 MR. MOGIN: Objection, Your Honor. I think this is a
13 question for you ultimately, but also I don't believe that
14 there's any foundation based upon the prior answer to the
15 question that Mr. Regard has the expertise to compare the two
16 systems in the manner that the question seeks.

17 THE COURT: Well, hold on. Hold on. I mean, the
18 reason I'm allowing these, folks, is to help us because I'm not
19 an expert, okay, in the two systems. Okay. But now I'm
20 really -- I'm so sorry I'm interrupting you. But, you know,
21 you asked him if he's aware of it. Do you think that's --

22 MR. MAROVITZ: Maybe I can help, Your Honor.

23 THE COURT: Yes.

24 MR. MAROVITZ: The plaintiffs have made submissions
25 in this case.

1 THE COURT: Right.

2 MR. MAROVITZ: And Mr. Regard will testify that he's
3 had a chance to review those submissions, and that's really
4 what I'm asking him about. I'm certainly not asking him
5 whether -- I'm not asking him more than that.

6 THE COURT: And what's your objection again?

7 MR. MOGIN: Well, I think the question has just been
8 changed. What the question was he was asked to compare CBAA as
9 the plaintiffs have used it with Boolean search as the
10 defendants have proposed it.

11 MR. MAROVITZ: Oh, no, that's not my question. My
12 question is as plaintiffs have proposed it. I'm not aware that
13 the plaintiffs have actually used it. So my question needs to
14 be clear. And if my misspoke before, I apologize. But my
15 question is as proposed, is it the case that when compared head
16 to head plaintiffs' content based search methodology is
17 superior to defendants' Boolean search methodology.

18 MR. MOGIN: And my point, and putting aside the legal
19 issue, my point is is that based upon the witness' prior answer
20 there's no foundation for him to comment on plaintiffs'
21 proposed methodology because he hasn't testified -- there's not
22 a proper foundation at this point as to any knowledge or
23 expertise that this witness has with the advanced analytics.
24 He said he's aware of it. As I heard that, it sounded like
25 he's done some reading in the area. Well, I've done some

1 reading too, and nobody wants to hear my expertise.

2 MR. MAROVITZ: The problem, Judge, the problem,
3 Judge, is that plaintiffs have presented advanced analytics as
4 though it is a discrete methodology that everybody knows. I
5 think Mr. Regard can debunk that, and that's why I'm asking him
6 the question the way I've asked it.

7 THE COURT: All right. I'm going to -- I think that
8 the whole first 15 minutes of Mr. Regard's resume was to
9 qualify him as an expert, if you will, in the field of -- in
10 this field. In this field. And I think based upon that, he is
11 going to give an opinion here -- you know, I so much look at
12 this as a real work in progress. Okay. This is not a Daubert
13 hearing. This is not -- we have got a very practical issue
14 here we've got to address. So I really wanted to keep this as
15 specific as possible. Okay. Not a, as I said before, a
16 treatise on what might be applicable in other cases. I'm just
17 talking about right here. So if you want to ask him -- I'm
18 going to overrule your objection is what I'm doing, and you may
19 ask the question. Okay.

20 BY MR. MAROVITZ:

21 Q Do you have an opinion, Mr. Regard, on whether as applied
22 here in this case for this Court, whether the plaintiffs'
23 content based search methodology that's been offered that
24 you've read about is superior in this case to the defendants'
25 search methodology that the defendants used in this case?

1 A I do.

2 Q What's that opinion?

3 A My opinion is that it is not demonstratively superior.
4 That, in fact, the process used by the defendants uses many, if
5 not all of the aspects the plaintiffs have even asked for. And
6 that when one thinks of content based analytics, that the
7 application of keywords is content based analytics. The use of
8 the topics functionality on Clearwell is content based
9 analytics. It's the process we need to look at, not the
10 individual technologies.

11 Q And what do you mean by that? Why are keywords content
12 based analytics?

13 A A lot of the pleadings and the discussion as I look at it
14 revolves around the issue of predictive coding. Predictive
15 coding is the ability to look at a few documents to make
16 decisions, and to come up with a methodology, a technology that
17 will take those decisions and apply them to a different body of
18 documents. And that's what keywords do. When you look at
19 documents and you say these keywords are choosing documents
20 that I believe help me segregate my documents and find ones
21 that are likely to be relevant and I apply those keywords which
22 look at the full text of the documents I'm applying them
23 against and segregate those documents, it's making a decision.
24 That is a form of predictive coding.

25 And then we go and we confirm the way that technology

1 is working by looking at the null set. I would also suggest,
2 not suggest. I would also acknowledge, observe that the
3 defendants used other types of content analytics in developing
4 and testing their search terms. They used the topic feature in
5 Clearwell. This is a feature that uses software technology to
6 analyze the words in individual documents and create
7 relationships between those words and the way those words
8 appear on a frequency basis between multiple documents and in
9 proximity to each other -- not proximity, but in co-appearance,
10 if you will, in one document versus another. And again uses
11 that to group documents, define similar documents, and make
12 better decisions about how to segregate one set of documents
13 from another.

14 So from that aspect what the defendants have used is
15 content based analytics, and they've used it in a predictive
16 coding manner. So I don't find the methodology suggested by
17 the plaintiffs to be superior. I find in many ways it
18 overlaps. I find that they didn't recommend any particular
19 tool or technology. And that the tools and technologies the
20 defendants have used have already incorporated those concepts.

21 Q So in many respects is the dichotomy on the one hand
22 between Boolean searching and on the other -- well, why don't
23 you comment on the way this has been set up on the dichotomy on
24 the one hand between Boolean searching and the other on
25 predictive coding as it applies to this case.

1 A Again, I have worked with clients on cases to use Boolean
2 tech -- Boolean search strings in a predictive coding manner,
3 where we have used iterative searching and evaluational
4 searches to create a set of search terms that we feel will
5 successfully and reasonably segregate a corpus of documents
6 into those that we're going to review and those that we're not
7 going to review. That is a prediction on that corpus based on
8 these search terms.

9 And I also want to say that, you know, we talk about
10 search terms as if we're talking about one or two searches in
11 Google. That's not what we're doing here. We're talking about
12 very complex search terms that the defendants put together
13 here. I mean, I've looked at some of these. These have dozens
14 of terms. And when you look at the Boolean structure of those,
15 those translate into many more searches. It's not a case of
16 running five or six searches and saying that was it and,
17 therefore, it's so simple it's not reliable. It was actually
18 extraordinarily complex and varied.

19 MR. MAROVITZ: Your Honor, if I may. I want to
20 reserve time for Counsel on Call because I know Mr. Regard is
21 coming back anyway at some point. So might I suggest that we
22 essentially allow Mr. Regard to come back on another day. I
23 would finish off his direct examination with the remaining
24 points that we just -- we haven't gotten into because we're all
25 going through this new area of procedure and law, and we allow

1 Counsel on Call to give its testimony at this point.

2 THE COURT: All right. Let's do Counsel on Call, and
3 I'll think about what we're doing -- I'm going to think about
4 how we're handling the rest of this at lunch. Okay.

5 MR. MAROVITZ: That's, that's --

6 THE COURT: But let's call Counsel on Call right now.
7 Okay.

8 MR. MAROVITZ: Right. Thank you, Your Honor.

9 THE COURT: Thanks, Mr. Regard. Don't leave please.

10 THE WITNESS: Thank you, Your Honor.

11 (Witness excused.)

12 MR. NEUWIRTH: If it pleases the Court, Your Honor,
13 we would call Samuel Brown from Counsel on Call.

14 SAMUEL W. BROWN, DEFENDANTS' WITNESS, DULY SWORN

15 THE COURT: Have a seat and please state your name
16 for the record. Okay.

17 THE WITNESS: My name is Samuel William Brown.

18 MR. NEUWIRTH: Thank you, Your Honor.

19 DIRECT EXAMINATION

20 BY MR. NEUWIRTH:

21 Q Mr. Brown, for whom do you work?

22 A Counsel on Call.

23 Q And briefly what does Counsel on Call do?

24 A Counsel on Call is a company that provides attorneys on a
25 contract basis to its clients, which are typically large law

1 firms and corporations. The company has a division in which it
2 provides attorneys that work on substantive matters to
3 corporate clients, and then they have an eDiscovery division.

4 Q And how long have you been at Counsel on Call?

5 A Since April of 2006.

6 Q And what's your current position there?

7 A Senior attorney and project manager.

8 Q And can you tell me the types of clients that you've
9 worked for at Counsel on Call.

10 A Well, currently obviously Georgia Pacific, currently
11 AT & T, AT & T Mobility, Cox Communications, Wal-Mart, and in
12 the past Coca-Cola.

13 Q And these are all clients that you've worked on with
14 eDiscovery projects related --

15 A That's correct.

16 Q -- related to ESI? And very briefly what is your
17 educational background?

18 A I have a bachelors degree and a JD from the University of
19 Tennessee.

20 Q And you say you have a JD. Have you practiced law?

21 A Yes.

22 Q And are you admitted in any states?

23 A Licensed to practice in Tennessee and Georgia.

24 Q And have you ever practiced at a law firm?

25 A Yes.

1 Q Which firm?

2 A My own firm. First of all, I was an associate and then a
3 partner in a firm in Knoxville, Tennessee for I think 11 or 12
4 years.

5 Q Now, what was the timing of Counsel on Call's retention in
6 this case?

7 A I believe we were retained in May of 2011.

8 Q Okay. Now, in the interest of time I'm just going to move
9 very quickly. Can you tell me were you personally involved in
10 the work that Counsel on Call did here related to the
11 development and testing of search terms?

12 A Yes.

13 Q And is that a process that began in May 2011?

14 A Yes.

15 Q Now, when this process began are you familiar with the
16 concept of a sample set for developing and testing search
17 terms?

18 A Yes.

19 Q And was a sample set created here?

20 A It was.

21 Q And what was in that sample set?

22 A Initially there were four high, what were considered to be
23 high priority custodians.

24 Q And who were those four custodians?

25 A Christian Fisher.

1 Q And who is Mr. Fisher?

2 A Mr. Fisher is an executive vice president and in charge of
3 the Container Board Division.

4 Q Who was the second one?

5 A Molly Hilliard.

6 Q And who is that?

7 A She is involved in the trades area.

8 Q And the third?

9 A Scott Denton. He is a pricing analyst.

10 Q And the fourth?

11 A Travis Ballard. And he is involved in inventory and
12 modeling down time.

13 Q And is there anyone else whose ESI was included in the
14 sample set that was used for development and testing of search
15 terms?

16 A We also included a fifth individual, Robert Bellinger.

17 Q And why was this fifth individual included?

18 A Mr. Bellinger was or is a plant manager. It was felt that
19 his ESI should be less -- well, it was felt that with respect
20 to the first four custodians they would be a very target rich
21 environment.

22 Q And, in fact, those four custodians were from among the
23 custodians that Georgia Pacific had identified to plaintiffs as
24 custodians whose ESI would be searched, correct?

25 A That's correct.

1 Q Mr. Bellinger was not on that list, correct?

2 A That's correct.

3 Q So you had this sample set of the five custodians. And
4 can you describe for the Court how the process of developing
5 the certain terms began.

6 A Well, the first thing we did was we received a set of
7 search terms and strings from I believe it was International
8 Paper's counsel.

9 Q And what did you do with that list?

10 A Well, we had our test corpus, which at that time consisted
11 of about 94,200 documents or so. We then --

12 Q And those were the documents that came from the files of
13 these five custodians, the ESI of these five custodians. And
14 is it correct that those have already been through some of the
15 steps talked about today, like de-duplication and deNISTing?

16 A Yes. Those were provided --

17 Q And 94,000 was what was left over, correct?

18 A Those were provided to KPMG. They were deNISTed,
19 de-duped, and made available to us.

20 Q And you ran the, you ran those --

21 A We ran the set of -- the initial set of search terms and
22 strings and looked at the hit counts and then began a linear
23 review of the result set from those searches.

24 Q And what then did you do? Just generally describe the
25 process.

1 A As a result of looking at those, we discovered that they
2 were, in fact, highly responsive. They pulled back responsive
3 documents, but they did not pull back a very large number of
4 responsive documents. So we then began a process of
5 identifying additional terms that we could then add to those
6 terms and strings.

7 Q And did you do this just once or did you do this multiple
8 times?

9 A We did it repeatedly. It was over a two-week period in
10 which we worked with internal counsel at Georgia Pacific. We
11 worked with Quinn Emanuel. We worked with KPMG. And it was
12 through a series of iterative steps that we eventually arrived
13 at a very elaborate, very robust sort of strings.

14 Q Now, in this initial process what was the time period that
15 this took place? You said it started in May and ran till
16 roughly when?

17 A I believe that we got the initial set of terms from IP in
18 about the second or -- toward the third week of May. And it
19 would have been -- we would have been finished with that
20 process around the 16th till say the 20th of June.

21 Q And in this initial process was the topics function from
22 Clearwell used as part of the work you did to develop the
23 search terms?

24 A Yes.

25 Q And was it applied to the null set?

1 A It was applied -- at that time it was applied both to look
2 at the documents that were positive hits, and then it was
3 applied generally to the reviewed corpus. We didn't have a
4 null set of that topic.

5 Q So it was applied to both the hit set and the corpus as a
6 whole?

7 A That's correct.

8 Q Now, you said that the results by mid-June were robust.
9 Did you do any testing to determine that?

10 A Okay. At that point we felt that the results were robust,
11 and so we began the validation phrase in which we created what
12 I referred to as the combined composite set and then the null
13 set. The combined composite set was the set of search strings
14 that would hopefully pull back, you know, positive hits from
15 the test corpus. The null set was defined as the test corpus
16 less the combined composite set.

17 Q And did you create a sample to test?

18 A At that point we created sample sets.

19 Q And how big was that sample set?

20 A For the null set it was 660 documents.

21 Q And you ran -- and what testing did you do with that set
22 of -- that sample set?

23 A We reviewed them linearly.

24 Q Meaning one by one?

25 A One by one.

1 Q And did you -- and was the purpose of that -- what was the
2 purpose of that review? What were you trying to determine?

3 A The purpose of that was to look for false negatives.
4 Basically --

5 Q You were trying to see if any documents that had been in
6 the null set were --

7 A That were responsive had slipped through and into the null
8 set.

9 Q And based on that testing, what was the percentage of
10 documents in the null set that you determined could have been
11 considered as documents that should have been hit?

12 A It was 4.1 percent.

13 Q So out of all the documents just 4.1 percent --

14 A Yes.

15 Q -- were ones that you would consider that could have been
16 treated as hits?

17 A Might have been treated as marginally responsive.

18 Q Okay. Now, after this testing, what was done next?

19 A Well, then we also did the same, essentially the same
20 exercise on the combined composite set. There we looked at 400
21 documents, and we determined that out of the 400, 218 of them
22 were responsive. And so we had a 58 percent responsiveness
23 rate.

24 Q All right. So that's in the set that was hit by the
25 search terms, 58 percent of the documents were actually

1 responsive as opposed to just potentially responsive?

2 A That's correct.

3 Q Okay. Now, after this round of testing in mid-June, what
4 was the next part of the process?

5 A It was the -- the results of the search strings themselves
6 were given to the defendants.

7 Q And did the defendants give you feedback?

8 A They did.

9 Q And did you use that feedback to further develop the
10 search terms?

11 A Yes, we did.

12 Q And did you come up with revised sets of search terms
13 based on that?

14 A We did.

15 Q And in this process did you again use the topics function?

16 A Always.

17 Q Did you use it both on the hit set and the null set?

18 A We used it on the hit set, and we used it especially on
19 the null set.

20 Q And then did that lead to a further revised set of search
21 terms?

22 A It did.

23 Q Did you test them again and validate them?

24 A We did.

25 Q And this time in the validation what was the percentage

1 result you got?

2 A At that time we actually ran the validation twice. We had
3 4.7 percent on one set and 4.1 on the other.

4 Q And that was testing of the null set?

5 A That was testing of the null set.

6 Q And then what was the next step in the process after this
7 testing that was done?

8 A Well, and again we did the combined composite set.

9 Q And what was next?

10 A That would have been around I think the 5th of August.
11 And those search strings were turned over to the plaintiffs.

12 Q And then did you receive feedback back from the
13 plaintiffs?

14 A On about the -- I believe it was about the 15th of
15 September we received a letter.

16 Q Could you hold up -- you should have there in front of you
17 Defendants' Exhibit 2, which is the plaintiffs' preliminary
18 analysis of defendants' first set of proposed search terms.

19 A Yes.

20 Q And is this what you received back? Is this what you were
21 referring to?

22 A It is.

23 Q And did you use this in any way to try to add anything to
24 the search terms?

25 A Yes.

1 Q And what did you do?

2 A Well, we reviewed the letter, and we determined -- there
3 were several examples where, in fact, for example, we missed
4 the name of a trade association. So we added that to one of
5 the search strings. We then began a process, an iterative
6 process discussing with the other defendants how we could
7 accommodate some of the, the concerns that the plaintiffs
8 raised in the letter principally by modifying some of the
9 strings and by modifying the proximity of connectors so that we
10 would capture a larger set of documents.

11 Q As a general matter did you find this plaintiffs' input
12 helpful to the process?

13 A As a general matter I did not.

14 Q But you did try to extract this information?

15 A Absolutely.

16 Q And did this lead again to go another development of a
17 further revised set of search terms?

18 A It did.

19 Q And did you validate those search terms again?

20 A Yes.

21 Q What were the results of that validation process?

22 A At the end of the vali -- at the end of that validation
23 process the null set validated to 4.2 percent.

24 Q So you now have had four tests all in the range of
25 somewhere between 4 and 5 percent when you tested the null set?

1 A That's correct.

2 Q And did that lead to a final set of search terms?

3 A Yes.

4 Q Okay. Now, was that final set of search terms then
5 applied against GP's full review corpus?

6 A It was.

7 Q And is it correct that that corpus included ESI from 17
8 custodians?

9 A I believe that's correct.

10 Q And did it also include ESI from other sources that were
11 not specific custodians?

12 A There were specific -- I think there were reports that
13 were pulled from other custodians --

14 Q Well, was it from other --

15 A Other sources.

16 Q From other sources, right. So there were both custodian
17 ESI --

18 A And source ESI.

19 Q -- and ESI from other sources within GP, correct?

20 A That's correct.

21 Q And so you now -- the search terms were applied to that
22 and you got a set of documents to review, is that correct?

23 A That's correct.

24 Q And do you recall how large that set of documents was,
25 that ESI set that was going to be now reviewed by attorneys?

1 A The total set including everything that you mentioned, all
2 17 custodians plus the additional ESI sources was about 140,000
3 records.

4 Q And --

5 MR. MOGIN: Your Honor.

6 THE COURT: Yes, sir.

7 MR. MOGIN: I don't want to interrupt the proceedings
8 in light of the time, but this also regarding the application
9 of the 17 custodians is new information.

10 MR. NEUWIRTH: I just have one more question. The
11 17 -- those custodians have all been disclosed. We're not
12 trying to make any big point. I just have one follow-up
13 question on this.

14 THE COURT: Well, you could --

15 MR. MOGIN: The 17 -- first off, the 16 custodians
16 have been disclosed to us. However, the fact that the
17 custodian has been disclosed is not an issue. The fact that
18 this test was run with respect to those custodians is the new
19 information.

20 MR. NEUWIRTH: He didn't say the test was run. He
21 said the search terms were run against the custodians to get a
22 body of documents to be reviewed. And all I want to do is have
23 Mr. Brown tell the Court where GP is in the review process.
24 That's it.

25 THE COURT: Okay. To me what you're saying is also

1 subject to cross-examination. This is one person talking about
2 something he did himself. Okay. So there's a little bit
3 easier to control the cross on that. And you may certainly
4 follow up whatever you want to follow up.

5 MR. NEUWIRTH: I just would have one follow-up
6 question.

7 BY MR. NEUWIRTH:

8 Q Of those approximately 140,000 documents of ESI that you
9 mentioned that were the result of applying the search terms to
10 the full body, those were reviewed by Counsel on Call as a
11 first level of review, correct?

12 A That's correct.

13 Q And what percentage of those documents have now been
14 reviewed as of today?

15 A Over 99 percent.

16 Q Okay. How many hours did Counsel on Call spend on the
17 process of developing and testing and validating the search
18 terms you referred to earlier?

19 A I believe in excess of 900 hours.

20 Q 900?

21 A 900.

22 Q Thank you. Just one final question. You mentioned
23 earlier that when the set, the sample set was set up, it had
24 these five custodians, four from the list of custodians that GP
25 had identified as being likely to have responsive ESI that was

1 shared with the plaintiffs and one who was not on that list?

2 A That's correct.

3 Q And you mentioned today that you did four rounds, at least
4 four rounds of this validation process?

5 A Right.

6 Q Were all of those rounds -- did all of those rounds of
7 testing involve all five of those custodians in the sample set?

8 A The final validation phase after we received the
9 plaintiffs' input did not include Mr. Bellinger.

10 Q So the final round of testing was just on the four
11 custodians that you identified earlier, the four senior
12 executives that GP, Georgia Pacific had identified as being --

13 A That's correct.

14 Q -- likely to have responsive documents and this so-called
15 controlled custodian Mr. Bellinger's files were not included in
16 that validation test?

17 A They were completely removed and suppressed. We did
18 not --

19 Q And that was the test that had the result of 4.2 percent?

20 A Yes.

21 MR. NEUWIRTH: We have no further questions, Your
22 Honor.

23 THE COURT: You can cross.

24 MR. MOGIN: Can we possibly delay the cross, the
25 beginning of cross till after the lunch hour in light of the

1 fact that we're bringing Mr. Regard back?

2 THE COURT: Cross when -- you're saying after lunch?

3 MR. MOGIN: Yes.

4 THE COURT: Oh. And then that's going to cut into
5 your time?

6 MR. MOGIN: Well, I understand that, but Mr. Regard's
7 going to come back so there will be some second sessions.

8 THE COURT: No, I -- I actually, I'm going to talk to
9 Chris about what we're doing. I'm also -- I'm just so confused
10 on this. Do you want to take a break for lunch now, is that
11 what you're saying? Because if you want to take a break for
12 lunch, that's fine with me. We were going to go -- and then
13 we'll -- you've got the two witnesses this afternoon.

14 MR. MOGIN: Yes, Your Honor.

15 THE COURT: Okay. So let's take a break. We're
16 taking one hour, so we're back here at 1:20. Okay.

17 MR. MOGIN: Thank you, Your Honor.

18 THE COURT: And we will do cross-examination then.
19 Okay. Thank you.

20 (Whereupon, said trial was recessed at 12:20 p.m., until
21 1:20 p.m.)

22

23

24

25

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

IN THE UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

KLEEN PRODUCTS, LLC, et al.,) No. 10 C 5711
)
Plaintiffs,)
)
v.)
)
PACKAGING CORPORATION OF AMERICA,) February 21, 2012
et al.,) Chicago, Illinois
) 1:20 p.m.
Defendants.) Evidentiary Hearing

TRANSCRIPT OF PROCEEDINGS
BEFORE THE HONORABLE MAGISTRATE JUDGE NAN R. NOLAN

APPEARANCES:

For the Plaintiffs: FREED KANNER LONDON & MILLEN LLC
2201 Waukegan Road
Suite 130
Bannockburn, Illinois 60015
BY: MR. MICHAEL J. FREED
MR. ROBERT J. WOZNIAK

THE MOGIN LAW FIRM, P.C.
707 Broadway
Suite 1000
San Diego, California 92101
BY: MR. DANIEL J. MOGIN

SCOTT & SCOTT LLP
707 Broadway
San Diego, California 92101
BY: MR. WALTER W. NOSS

TRACEY DANA McCULLOUGH, CSR, RPR
Official Court Reporter
219 South Dearborn Street
Room 1426
Chicago, Illinois 60604
(312) 922-3716

1 APPEARANCES CONTINUED:

2 For Defendant Packaging KIRKLAND & ELLIS LLP
3 Corporation of America: 300 North LaSalle Street
4 Chicago, Illinois 60654
BY: MR. BARACK S. ECHOLS
MR. LEONID FELLER

5 For Defendant FOLEY & LARDNER LLP
6 International Paper: 777 East Wisconsin Avenue
7 Milwaukee, Wisconsin 53202
8 BY: MR. JAMES T. McKEOWN

9 FOLEY & LARDNER LLP
10 321 North Clark Street
11 Suite 2800
12 Chicago, Illinois 53202
13 BY: MS. JOANNE LEE

14 For Defendant MAYER BROWN LLP
15 Temple-Inland: 71 South Wacker Drive
16 Chicago, Illinois 60606
17 BY: MR. ANDREW S. MAROVITZ
18 MS. BRITT M. MILLER

19 For Defendants K & L GATES LLP
20 Cascades and Norampac: 70 West Madison Street
21 Chicago, Illinois 60602
22 BY: MR. SCOTT M. MENDEL

23 For Defendant QUINN EMANUEL URQUHART &
24 Georgia Pacific: SULLIVAN LLP
25 51 Madison Avenue
22nd Floor
New York, New York 10010
BY: MR. STEPHEN R. NEUWIRTH

FIGLIULO & SILVERMAN
Ten South LaSalle Street
Suite 3600
Chicago, Illinois 60603
BY: MR. JAMES R. FIGLIULO

WINSTON & STRAWN
227 West Monroe Street
Suite 4400
Chicago, Illinois 60606-5096
BY: MR. R. MARK McCAREINS
MR. JOSEPH L. SIDERS

1 APPEARANCES CONTINUED:

2 For Defendant
3 Weyerhaeuser Company:

McDERMOTT WILL & EMERY LLP
227 West Monroe Street
Suite 4400
Chicago, Illinois 60606-5096
BY: MS. RACHAEL V. LEWIS

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

1 SAMUEL W. BROWN, DEFENDANTS' WITNESS, PREVIOUSLY SWORN

2 THE COURT: Okay. Mr. Mogin.

3 MR. MOGIN: Thank you, Your Honor.

4 CROSS-EXAMINATION

5 BY MR. MOGIN:

6 Q Mr. Brown, let's be clear, you are a lawyer representing
7 or assisting in the representation of a defendant in this case,
8 is that true?

9 A I'm an attorney licensed in Tennessee and Georgia, and I
10 am a contract attorney that works as an independent contractor
11 through Counsel On Call that is engaged by Georgia Pacific. I
12 am not an attorney of record in this matter.

13 Q Do you have a duty of loyalty to Georgia Pacific?

14 A I would think so, yes.

15 Q Did you act on that duty in connection with this
16 engagement?

17 A Yes.

18 Q So I want to make sure that I understand the chronology
19 that you've outlined. Someone presented you in May of 2011
20 with a set of search terms, is that correct?

21 A That's correct.

22 Q And who presented those to you?

23 A It came from Quinn Emanuel.

24 Q And was it developed by Quinn Emanuel?

25 A No. My understanding was that it was developed by counsel

1 for International Paper.

2 Q That would be the -- at that time the Gibson Dunn firm?

3 A I'm sorry?

4 Q At that time the Gibson Dunn firm?

5 A All I know is that it came from International Paper.

6 Q Well, did you talk to anybody who had created the search
7 string?

8 A No.

9 Q How many terms were on that search string?

10 A It was a set of around I want to say between 12 and 16
11 strings that had, you know, a number of terms, 3 or 4 terms for
12 each string.

13 Q Did you retain a copy of that in your file?

14 A Yes.

15 Q Do you have it with you today?

16 A No.

17 Q Were you instructed not to bring it with you today?

18 A No.

19 Q And then the next thing that happened was you reviewed
20 those search terms, correct?

21 A Correct.

22 Q And you made some suggested changes?

23 A The first thing that we did was we tested those terms
24 against our test corpus. We ran -- we set the terms up in the
25 syntax appropriate for Clearwell, and then we ran those terms

1 against the 94,000 document test corpus that we had set up in
2 Clearwell. We then reviewed --

3 Q No, sir. My question was -- had to do with the search
4 string.

5 A Yes.

6 Q Who made the next set of modifications to the search
7 string?

8 A We all did.

9 Q Who's we all?

10 A We were working collaboratively with Quinn Emanuel, with
11 KPMG, and with Georgia Pacific.

12 Q Anyone else?

13 A Only the people on my team and Counsel On Call.

14 Q And how many people were on your team at Counsel On Call?

15 A At times another person or two people. Myself and two
16 others.

17 Q Was that true of when you engaged in the review process as
18 well?

19 A No, there -- you mean the document review?

20 Q Yes.

21 A No. There are I think 14 attorneys on that project.

22 Q 14. Okay.

23 A Yes.

24 Q Now, going back to this chronology, if you will. So when
25 is the second iteration filed?

1 A Okay. After we -- the next step in the process --

2 Q If you could just tell me when the next iteration of
3 search strings came out, please.

4 A Well, I'm not sure I understand your question.

5 Q Okay. First you got the set from IP, and then you did
6 something. And then you came up with a new set of search
7 strings or an amendment to the first set, correct?

8 A Correct.

9 Q Okay. And when was that?

10 A That would have been about a week after we received the
11 first set of strings.

12 Q Okay. And so the collaborators in that process were
13 Counsel On Call --

14 A Georgia Pacific --

15 Q -- Quinn Emanuel, and KPMG?

16 A And internal counsel at Georgia Pacific.

17 Q I'm sorry?

18 A And internal counsel at Georgia Pacific.

19 Q Internal counsel from Georgia Pacific. Was anybody from
20 the IT department involved?

21 A Not in the development of the search terms, no.

22 Q And then that's when you ran your first tests, correct?

23 A After we developed what we consider to be a robust set of
24 search terms, we then ran our first validation exercise, and
25 that would have been around the latter -- the middle part of

1 June I believe.

2 Q Okay. And do you have the results of those tests? You
3 maintain those in your files?

4 A Yes.

5 Q Do you have them with you today?

6 A I do not.

7 Q Okay. And then what happens next in the process of
8 developing the search strings? You're up to June now.

9 A At that point we handed the set of search terms and
10 strings off to the defendants.

11 Q To which defendants?

12 A The defendants would have been -- well, I don't know
13 exactly which defendants. I know it would have been at least
14 International Paper. That was through Quinn Emanuel.

15 Q So you gave them back to Quinn Emanuel?

16 A Correct.

17 Q And Quinn Emanuel made some circulation, the details of
18 which you're not aware of?

19 A That's correct.

20 Q All right. And then what happened?

21 A Then we received feedback through Quinn Emanuel from the
22 other defendants.

23 Q How do you know it was from the other defendants?

24 A That's what I was told.

25 Q Was there any indication that, for example, Rock Tenn

1 suggests the following terms or PCA suggests the following
2 terms, anything like that?

3 A I do recall one instance when International Paper
4 suggested that we add the term, I think it was rationalize.
5 But generally speaking, the conversations were cordial and were
6 not as specific as to this defendant says we should do X, that
7 defendant says we should do Y, the next defendant says we
8 should do Z. Rather, my understanding was that there was a
9 discussion that was then presented to us.

10 Q All right. We're up to June now. So then what happened?

11 A Okay. Then at that point we revised the terms based on
12 the input from defendants. We ran a rationalization process of
13 our own in which we sought to condense the search terms. At
14 this point -- prior to that time there were -- it was a certain
15 level of redundancy because each term was developed either for
16 a specific concept or a specific RFP. But we decided to reduce
17 the duplication. And then we ran another -- after the
18 completion of the rationalization process we ran another
19 validation exercise.

20 Q Do you mean to suggest that sometime after the
21 June 15th-ish event that there was some sort of a list that
22 existed that linked specific requests for production of
23 documents to specific search strings?

24 A We documented all of our work.

25 Q Do you have any of your documentation with you today?

1 A No.

2 Q Did you ever create any document that linked specific RFPs
3 to specific requests for production of documents to specific
4 keyword strings?

5 A Yes.

6 Q Do you still have that document?

7 A I do.

8 Q Do you have that document with you?

9 A No.

10 Q Does counsel have that document?

11 A I would -- I honestly don't know if they have them today
12 or not.

13 Q How many requests for production of documents do the
14 search strings attempt to capture?

15 A I believe it was around 15 or 18.

16 Q 15 or 18 requests for production of documents?

17 A That's correct.

18 Q Now, you're aware, are you not, that at some point the
19 defendants requested -- strike that.

20 Were you aware that in January of 2011 at the
21 defendants' request the plaintiffs provided a list of
22 categories?

23 A I don't believe I was aware of that.

24 Q Were you aware that plaintiffs provided a revised list of
25 categories in March of 2011?

1 A I do not believe I was aware of that.

2 Q So you never saw either of those category lists?

3 A If I did, I have no recollection of them.

4 Q Well, okay. Were you aware that at the defendants'
5 request the plaintiffs provided a list of document requests
6 that they labeled as the conduct requests as opposed to
7 transactional requests?

8 A Not specifically, no. I'm aware that, I'm aware that
9 there were document production requests. I believe that there
10 were around 90 or 95, 96 requests that were -- that we did
11 receive a copy of those.

12 Q Well --

13 A But in terms of how those were subsequently discussed
14 between the parties and divided, I have no knowledge of that.

15 Q Okay. So I believe that -- well, I guess plaintiffs'
16 requests for production of documents have not been placed in
17 evidence yet. Let me see if I can get a copy of those and put
18 them before you, please.

19 A I'm sorry, but if you could speak up. I'm only getting
20 about half of what you're asking me.

21 Q I'll speak up.

22 A Thank you.

23 Q But I'm not yelling at you.

24 A Very good.

25 MR. MOGIN: We would ask, Your Honor, that this be

1 marked. This is plaintiffs' first request for production of
2 documents directed to all defendants. And this I believe would
3 be Plaintiffs' 5. And for clarity of the record, Your Honor,
4 during the break I realized that I had been referring to
5 Plaintiffs' 4 previously as Plaintiffs' 5. And that was the
6 letter of November --

7 THE COURT: 23rd.

8 MR. MOGIN: Yes.

9 THE COURT: That's what I have as 4.

10 MR. MOGIN: Thank you. So this will be No. 5, and
11 this is plaintiffs' first request for production of documents.

12 BY MR. MOGIN:

13 Q So, Mr. Brown, can you tell me which of plaintiffs'
14 document requests the 15 search strings that you've testified
15 about relate to.

16 MR. NEUWIRTH: Can I ask for clarification. Are you
17 talking about the 15 that came from International Paper or
18 something else?

19 MR. MOGIN: Well, we're up to June, correct?

20 THE WITNESS: Correct.

21 MR. NEUWIRTH: The testimony -- the only reason I'm
22 asking this question for the interest of everybody getting this
23 right is that I believe the testimony about 15 search strings
24 related to what was received from International Paper at the
25 very beginning of the process.

1 MR. MOGIN: All right. Let me clarify that.

2 THE COURT: Thank you.

3 BY MR. MOGIN:

4 Q So by the end of June do you recall how many search
5 strings there were?

6 A Somewhere between 15 and 25.

7 Q Was it -- let me represent to you, if I may, that the
8 final search string that we received from Georgia Pacific,
9 which is in evidence as Exhibit 5, has 21 strings. It's
10 actually 17 strings, 4 of them have subparts. Does that ring a
11 bell?

12 A Yes.

13 Q So it's 21 slash 17, correct?

14 A Yes.

15 Q All right. Now, please refer, if you would, to Exhibit 5
16 and tell me which requests for production of documents the
17 search strings, those search strings relate to. And if you
18 like, Mr. Brown, on the very last page there should be a list
19 of categories with subheadings, and that might assist you.

20 MR. NEUWIRTH: Your Honor, the witness has been
21 requested to answer a question about what I believe is now the
22 final set of search terms and which requests they relate to.
23 As Mr. Mogin notes, the final set of search terms is something
24 that was an exhibit to one of our filings. And so I don't --
25 it may be helpful to the witness to have that with him when he

1 answers this question. But I defer to everyone. I just wanted
2 to make that option available.

3 THE COURT: Do you know where that document is?

4 MR. NEUWIRTH: Yes. It is, it's Exhibit 5 to our --
5 to docket No. 288.

6 THE COURT: You want to show that to -- Mr. Mogin,
7 you want to show that to Mr. Brown?

8 MR. MOGIN: Well, I want to be clear about something
9 first, please, Your Honor. I believe Exhibit 5 is the October
10 list. I think the June list that the witness is referring to
11 is Exhibit 4.

12 MR. NEUWIRTH: Well, you -- Mr. Mogin, in fairness
13 when the witness answered this question, you then referred to
14 the final list. When the witness said he recalled somewhere
15 between 18 and 25 in June, you then said that the final list
16 had X number. And so it wasn't clear. It seemed to me you are
17 now asking about the final list.

18 MR. MOGIN: Well, let's then -- I will be happy to
19 clarify it. If I can get --

20 MR. NEUWIRTH: We have -- there are two lists.
21 Exhibit 3 to our filing docket No. 288 is the list from
22 August 5th, 2011, which I believe there was testimony was sent
23 to the plaintiffs. Exhibit 5 is the final list, which I don't
24 have -- it doesn't have an exact date on it, but it is the
25 final list. So both of those are available.

1 BY MR. MOGIN:

2 Q All right. We'll skip ahead. We'll move forward from
3 June. We'll move forward to sometime in August.

4 A And let me just add additionally that when we initially
5 started the search term process, perhaps this will clear up
6 some of the discussion, we for organizational purposes and
7 because the search strings that we were provided were organized
8 by RFP, that is how we started the process. But eventually we
9 felt that it would be a better approach to organize the search
10 strings around broad concepts that encompassed the RFPs but
11 went beyond -- that were more of a concept approach, especially
12 after the rationalization occurred.

13 So by the end of the process, by the final set that I
14 believe you were given, it would not be correct to say that
15 they were an RFP specific set of search strings. But rather
16 they were responsive to concepts that were embodied in the
17 RFPs.

18 MR. MOGIN: I'm going to move to strike that answer
19 as being nonresponsive, if I may, Your Honor.

20 MR. NEUWIRTH: Your Honor, that's --

21 THE COURT: All right. I'm going to leave -- okay.
22 We're going to leave it in. I'm overruling the objection, and
23 we're leaving it in. Can we go back to -- can we get a little
24 context here. Okay. Because I think -- I mean, I'm confused
25 so the witness may also be confused here. So which one of the

1 lists of search terms do you want him to refer to? And let's
2 give him a copy if we have that. If it is June, August -- so
3 far I have June, August, or October. And then that I guess is
4 tied to an RFP is what your question is.

5 MR. MOGIN: What I had hoped to do, Your Honor, was
6 to use the August 5th list of search terms, which I think
7 relates to his June process that he's talked about. I'm trying
8 to establish the chronology myself.

9 THE COURT: Okay.

10 MR. MOGIN: And then -- but we don't apparently have
11 that with us here in the courtroom. So in order to move the
12 proceedings I'll modify my questioning, and I'm going to skip
13 ahead a little bit, but I might have to come back in order to
14 get the chronology correct.

15 THE COURT: Okay.

16 MR. MOGIN: So I will now place before the witness
17 what I'll ask to be marked as Plaintiffs' 6 which -- and even
18 though, Your Honor, there is a fly -- there's a fly sheet here,
19 Your Honor, that says Exhibit 5, but that relates to the
20 Exhibit 5 to the motion that was submitted to you.

21 THE COURT: So if you'll give it to Miss McCullough
22 and give one to the witness. Thank you.

23 (Document tendered.)

24 MR. NEUWIRTH: I'm sorry. Exhibit 5 is the final
25 list, not the August list. Exhibit 3 is the August list.

1 MR. MOGIN: I think I tried to specify that, Your
2 Honor. We don't have a copy of the August list here in the
3 courtroom or sufficient copies, so we're skipping ahead.

4 THE COURT: Okay. So this is the -- just tell me
5 what month this is.

6 MR. MOGIN: The plaintiffs received this in October.

7 THE COURT: Okay. Got it.

8 BY MR. MOGIN:

9 Q So I'll tell you what, before I start asking questions
10 about these two documents, let's do a little back and fill on
11 the chronology, shall we.

12 A Sure.

13 Q So in June you completed an iteration and then did some
14 testing. And then do you understand that that document was
15 provided to the plaintiffs in early August, August the 5th?

16 A I believe around the 5th of August, yes.

17 Q Were any changes made between the testing and validation
18 process that you described taking place in June and August the
19 5th?

20 A No.

21 Q Okay. Did anything happen with respect to development of
22 the search strings during the month of August?

23 A No.

24 Q So the next event in this chronology would be the receipt
25 of plaintiffs' critique of the first set of the search terms,

1 is that correct?

2 A That would be correct.

3 Q And as I understood your testimony, you looked over that
4 critique, correct?

5 A That is correct.

6 Q And, in fact, that critique is before you as Plaintiffs'
7 Ex -- I'm sorry, as defendants' search term -- it's Defendants'
8 Exhibit 2.

9 A Correct.

10 Q So you looked that over and you adopted -- can you tell us
11 precisely which of the suggestions of plaintiffs were adopted
12 from the next iteration?

13 A Well, I think we -- the first thing that we did was
14 conceptually try to understand what the -- you know, the core
15 objection was to what the plaintiffs were saying with respect
16 to our search terms. So as I read the first page of the first
17 paragraph, it's reading from the document: "In addition to the
18 above nonexhaustive list of core terms, the following items and
19 derivatives that appear in the RFP definitions do not appear in
20 the defendants' list." And there is a list of terms, which I
21 took to be search terms. They appeared to have been taken from
22 the plaintiffs' RFPs. So the first thing that I did was to
23 actually build a search string that encompassed this entire
24 list. And --

25 Q One search string?

1 A One search string. Because it is -- this appears to be a
2 search string, is the way it is structured. And I ran that
3 against our test corpus.

4 Q May I ask what led you to believe that that was intended
5 to be one search string? Do you see any logical connectors
6 there?

7 A Well --

8 Q Do you see any parentheses there?

9 A No, but it is -- well, I see a number of parentheses -- I
10 see a number of -- not parentheses, no.

11 Q Do you see any *ands* in that string?

12 A It was a set of terms that we were missing.

13 Q Do you see any *ands* in that string?

14 A No.

15 Q Do you see any *ors* in that string?

16 A No.

17 Q Do you have any proximity indicators within that string?

18 A Nope.

19 Q Do you see any groupings of documents within that string?

20 A Actually there is one *and*.

21 Q There's one *and*, and where is that *and*?

22 A At the very end.

23 Q At the very end. So that indicates to you, does it not,
24 that this was not indicated to be a single search string?

25 A No, it was indicated to be a single search string. But it

1 was indicated that our search strings were deficient by lacking
2 all of these terms.

3 Q But you then created one search string and ran that, is
4 that right?

5 A Yes.

6 Q Using all of these terms?

7 A That's correct.

8 Q And what did you conclude on that basis?

9 A Well, it pulled back approximately 67,700 hits from our
10 94,000 document test corpus.

11 Q The one search string did?

12 A With all these terms.

13 Q Got two-thirds of your documents?

14 A Yes.

15 Q And so did that make it a valid search string in your
16 opinion?

17 A It did not.

18 Q Now, let's talk about your opinion of search strings. Do
19 you have any training in linguistics?

20 A No.

21 Q Do you have any training in statistics?

22 A Not for many years.

23 Q Do you have any training in the science of information
24 retrieval?

25 A Not beyond -- as in terms of like a degree or something,

1 no, I do not.

2 Q How about national language processing?

3 A I know what the term means, but I -- no.

4 Q So you applied your lawyerly training and expertise to
5 development of the search strings, is that right?

6 A Correct.

7 Q So are you an expert in search strings, is that your
8 contention?

9 A I am -- I have created and ran hundreds, if not thousands
10 of searches on a variety of different subjects in a variety of
11 different litigations.

12 Q Are you an expert on the creation of Boolean search
13 strings?

14 A I consider myself to be knowledgeable and experienced. I
15 do not consider myself to be an expert.

16 Q Do you know what the term Boolean search string refers to?

17 A Yes.

18 Q What does it refer to?

19 A It's from Boolean algebra in which keyword terms are
20 connected to one another by various connectors that either
21 cause the computer that's running the search string to pull in
22 combinations of documents based on the instructions provided by
23 the connectors.

24 Q Are you familiar with the term polysemy, P-O-L-Y-S-E-M-Y?

25 A No.

1 Q Pardon me?

2 A No.

3 Q Are you familiar with the term synonymy, S-Y-N-O-N-Y-M-Y?

4 A Well, synonym in the sense of like a word that has the
5 same meaning as another word.

6 Q Do you know?

7 A If that's the word that you're using, that's how I would
8 define it, yes.

9 Q Have you heard the word before?

10 A Yes.

11 Q Have you engaged in exercises in connection with
12 developing Boolean keyword strings that involve polysemy or
13 synonymy?

14 A I may very well have, but not using that nomenclature.

15 Q What are the effects in developing Boolean keyword strings
16 of either of those phenomena?

17 A I'm sorry?

18 Q What are the effects when you develop a Boolean query of
19 either of those phenomena?

20 A Well, like I said, I don't know what the terms mean.

21 Q All right. Well, if you can refer please to Plaintiffs'
22 Exhibit 2, page -- I believe it's page 4. You'll see on the
23 left-hand side there's a box.

24 A I'm sorry. I don't know which one Plaintiffs' Exhibit 2
25 is.

1 Q It's entitled "Making Document Review Faster Cheaper and
2 More Accurate."

3 THE COURT: What page did you say? What page did you
4 say, Mr. Mogin?

5 MR. MOGIN: Page 4, please, Your Honor.

6 THE WITNESS: I'm sorry. Which page?

7 BY MR. MOGIN:

8 Q Page 4. In the middle of the page you'll see that there's
9 a subheading, "So What Exactly is Concept Searching?"

10 A Yes.

11 Q Do you see the box on the left-hand side?

12 A I do.

13 Q All right. And do you see the definition of polysemy that
14 appears there?

15 A Let me read the entire paragraph. All right.

16 Q All right. Now, having read that paragraph, do you have
17 an understanding of the impacts of either polysemy or synonymy
18 in connection with the development of Boolean search queries?

19 A They can introduce false positives and false negatives
20 into the --

21 Q Did you know that before you read this document?

22 A I knew that the concept that the English language isn't --
23 well, I know that the English language isn't perfect. And I
24 know that care must be taken in the construction of Boolean
25 strings to -- and testing of the results of running the strings

1 needs to be taken in order to ensure the results that you wish.
2 I have not specifically read KPMG's promotional literature, and
3 I have not heard those two terms.

4 Q All right. Now, when you constructed what eventually
5 became Plaintiffs' Exhibit No. 6; that is, the search strings,
6 did you review Plaintiffs' Exhibit No. 5, the RFPs?

7 A I reviewed the RFPs, yes.

8 Q Did you look in the defined terms?

9 A Yes.

10 Q And did you attempt to incorporate into your Boolean
11 strings each of the defined terms?

12 A No.

13 Q Now, let's go back then to exhibit -- to the final search
14 string list Exhibit 6. And take a look, if you would, and
15 compare that to Plaintiffs' 5, the RFPs. And please tell us
16 which of the 15 or so RFPs that you have identified are
17 reflected in the search strings -- in each of the search
18 strings in Exhibit 5.

19 A Well, by this time --

20 MR. NEUWIRTH: Your Honor, I have to object to this
21 question. This question is mixing up things that were said
22 earlier. There's been no testimony that this related to just
23 15 of the RFPs. That was a question that was asked about a
24 different point in the process.

25 THE WITNESS: Right. By this time by the end of --

1 THE COURT: Mr. Brown, just one minute, because I
2 want the record to be clear here. So let's -- why don't you --
3 I think now we're to the right month and to the right document.
4 So why don't you ask your question. I have Plaintiffs' Exhibit
5 5 is the RFP that was directed to all defendants. I don't have
6 a -- let me see. So this was May 3rd you sent it out.

7 MR. MOGIN: Correct.

8 THE COURT: And I think Plaintiffs' Exhibit 6 is a
9 document that, I don't know whether the record is clear, but I
10 think this is a record of search strings that you sent to the
11 defendants.

12 MR. MOGIN: No, Your Honor.

13 THE COURT: Oh.

14 MR. MOGIN: This is the end products of defendants'
15 search strings. This is what was sent to the plaintiffs by
16 Georgia Pacific in October of 2011 after we went through the
17 iterations that were described in the chronology of the
18 witness' testimony.

19 THE COURT: So now you are at -- so now ask your
20 question. Now that I'm straight on that, why don't ask your
21 question about the two documents.

22 MR. MOGIN: All right.

23 THE COURT: Okay. So Mr. Brown has a chance now to
24 see what goes with what. Okay.

25 BY MR. MOGIN:

1 Q So, Mr. Brown, how many of the plaintiffs' RFPs from
2 Exhibit 5 are represented in your opinion in search strings in
3 Plaintiffs' Exhibit 6?

4 A I don't think I can answer that question because the
5 method that we used by the time this final set of search
6 strings was prepared had become significantly divorced from the
7 RFPs themselves. Basically, by way of background defense
8 counsel made -- responded to these RFPs as I understand it.
9 They objected to certain RFPs as being objectionable for some
10 reason. Other RFPs were identified that would be answered by a
11 sufficient to show methodology or where a specific custodian
12 could be identified or specific source of data at Georgia
13 Pacific could be identified that would satisfy the RFP.

14 Some RFPs were duplicative or it was unclear. And so
15 the end result of the process, though, was that there was a
16 collection of RFPs that would be used that we would have to
17 develop search terms for to search the ESI. And that process
18 resulted in this set of search strings.

19 Q Is the bottom line of that testimony that you're unable to
20 link specific search strings in Plaintiffs' 6 with specific
21 document requests in Plaintiffs' 5?

22 A And again, I'm sorry. I did not hear what you said.

23 Q I said is the result of the process that you've just
24 described that you are unable to tell us which of the search
25 strings in Plaintiffs' Exhibit 6 correspond to the RFPs or to

1 particular RFPs in Plaintiffs' Exhibit 5?

2 A No, I could go through the RFPs one by one and potentially
3 locate search strings that may have had a bearing on that, on
4 the specific RFP.

5 Q But in the process of creating the RFPs, in other words,
6 you would have to do that as of now; is that correct?

7 A That's correct.

8 Q There's not a document that you could refer back to that
9 would link these final RFPs; that is, Exhibit 6 to specific
10 document requests in Exhibit 5?

11 A Not at the end of the iterative process, no.

12 Q Now, did you do anything to assure yourself that all of
13 the defined terms in Exhibit 5 were incorporated into Exhibit
14 6?

15 A Did I do anything to determine that all of the terms in
16 the RFPs were included in the search term string?

17 Q I just asked about the defined terms. If you'll refer to
18 Exhibit -- well, do you understand the concept of defined
19 terms --

20 A I do.

21 Q -- as it relates to RFPs?

22 A Right.

23 Q Okay.

24 A You have defined the meaning as specified by the
25 paragraph.

1 Q So you see the section that says definitions beginning on
2 page 1?

3 A Yes.

4 Q And it continues, does it not, on to -- through page 10?

5 A Correct.

6 Q And did you review those definitions?

7 A I did.

8 Q And did you do anything to assure that those defined terms
9 were incorporated within the final search string that is
10 Exhibit 6?

11 A We read at the beginning of the process the entire set of
12 RFPs. We read the definitions of the definitional section.
13 And all of that information was included in our deliberations
14 throughout the entire process. I did not go back at the end of
15 the process and check off the boxes, if that's what you're
16 asking.

17 Q So the answer is that you don't know whether all of the
18 defined terms are reflected in the search strings?

19 A No. My answer is that I read the RFPs cover to cover. I
20 read the defined terms, and I used that information to build
21 the search strings.

22 Q Are all --

23 A And, in fact, to then, you know, test the search strings.

24 Q Are each of the defined terms included in the search
25 strings?

1 A I'm sure they are not.

2 Q Do you have any documentation that would show us which
3 ones were included and which were eliminated?

4 A On the defined terms, no.

5 Q What did you do in terms of search terms when plaintiffs
6 asked for all documents relating to a particular concept? How
7 did you pick up using your Boolean logic the term relating?

8 A What we did was we started with a process. We had a
9 process. We started at the beginning of that process with the
10 International Paper search terms. We then went through a
11 number of rounds of iterations, of discussions, of revisions,
12 of testing of the individual terms. The upshot of which was at
13 various stages in the process we would have a collection of
14 search strings with terms that we felt covered the ESI that we
15 were charged with locating. At that time we then tested the
16 result of that.

17 MR. MOGIN: Your Honor, may I have an answer to my
18 question, please.

19 THE WITNESS: I'm answering your question.

20 THE COURT: Hold on. I think Mr. Brown is answering
21 your question. Okay. You can continue, Mr. Brown.

22 THE WITNESS: Thank you. At each stage of the
23 process after the cumulative discussions were completed we
24 engaged in a set of testing of validation with our test corpus
25 that I had already described. We had what we considered --

1 what we called the combined composite set and the null set.
2 The combined composite set being the collection of hits that
3 resulted from all of the search strings. The null set being
4 the quantity of the -- the ECA corpus that were left over after
5 those hits were removed. We then picked statistically
6 significant samples of each one of those bodies and did a
7 linear review of them based on our experience as review
8 attorneys. Again having read the RFPs, having read the
9 definitional section, having read the complaint.

10 And we went through and we actually coded those
11 documents in Clearwell so that, you know, on the -- for
12 example, on the -- in the first round of iterations we -- or in
13 first round of development we looked at 400 documents from the
14 combined composite set and determined that 218 of those had
15 been pulled in by the search terms and were responsive. We
16 then looked at the null set and we looked through 660 of those,
17 and from that initial 660 set, if memory serves, there were two
18 documents that were technical problem documents that couldn't
19 be identified. There were 27 documents that were marginally
20 responsive, and the remaining set of documents were
21 nonresponsive.

22 So to answer your -- so to bring all this back
23 together and to specifically answer your question, our goal in
24 creating a set of search terms was in effect to create
25 something like a net or a mesh that we could run through the

1 ECA corpus and pull out every document that was responsive to
2 those hits, and then most importantly go back and validate the
3 process by testing the null set.

4 BY MR. MOGIN:

5 Q Okay. You had a process. Who designed the process?

6 A The process was developed by me and was developed in
7 coordination with my superiors at Counsel On Call.

8 Q What input did the plaintiffs have in the process?

9 A In the process that we used we solicited -- the input that
10 we received, for example, in the September 15th correspondence
11 we utilized that in defining our terms. But in terms of the
12 process itself, it was a process that the defense group
13 developed.

14 Q Okay. And did you ever become aware that plaintiffs sent
15 a second letter in November regarding what's been marked as
16 Exhibit 6?

17 A Would that be this document?

18 Q November 10th, 2011.

19 A 22nd.

20 Q Let's mark it. Perhaps it hasn't been marked yet. This
21 will be Plaintiffs' 7. Let me direct your attention to the
22 bullet points beginning on page 3.

23 A Yes.

24 THE COURT: So this is Exhibit 7?

25 MR. MOGIN: This is Plaintiffs' 7, Your Honor.

1 THE COURT: Plaintiffs' 7.

2 THE WITNESS: Yes.

3 BY MR. MOGIN:

4 Q Yes, you have seen this letter before?

5 A I believe I have, yes.

6 Q Okay. And after receipt of this letter were the search
7 terms subsequently modified in any way that reflects the
8 concerns expressed in the letter?

9 A No. I don't believe so because this letter was sent on
10 the 10th of November, and I do not believe that we changed any
11 of the search terms subsequent to that time.

12 Q Now, you said that the search strings in part reflected
13 defendants' objections to plaintiffs' requests for production
14 of documents, is that right?

15 A Again, I do apologize. I'm sort of hard of hearing in one
16 ear.

17 Q I said is it correct that defendants' objections to
18 plaintiffs' requests for production of documents are reflected
19 in your search term list?

20 A In the sense that once an objection was made to the
21 request, that request was not considered in the construction of
22 the strings. That doesn't mean to say, however, that these
23 search strings would not necessarily pull up documents that
24 might be responsive to one of those.

25 Q So who identified the ESI corpus that you were given?

1 A I'm sorry. You're going back to the beginning of the
2 process?

3 Q At the beginning of the process who identified the corpus?

4 A The four custodians were identified by Georgia Pacific.

5 Q Did you have anything to do with the collection of
6 documents from Georgia Pacific?

7 A No.

8 Q Did Counsel On Call participate in that process?

9 A We do not do that.

10 Q So if a mistake was made in the collection of the process,
11 in your opinion would that have impacted the search results?

12 A It would depend on the nature of the mistake.

13 Q Documents that should have been included were eliminated.

14 A If documents that were -- that should have been included
15 were eliminated, then those would not have been in the test
16 corpus. And it is possible that a search term that would have
17 been developed otherwise might not have been developed.

18 Q And a responsive document or a document that might
19 otherwise have been responsive couldn't have been found no
20 matter how good your Boolean strings were, correct?

21 A Not necessarily, no. The -- your argument is premised --
22 the question as I understand it is premised on the idea that a
23 unique search term would be required for a unique document.
24 And it's very possible that that document that was not
25 collected -- I mean, if it's not in the corpus, it's not going

1 to be produced.

2 Q Exactly. Thank you. Now --

3 A If that's what you're saying me.

4 Q That's precisely what I was asking you. If it's not in
5 the corpus --

6 A Okay. I'm sorry. I wasn't trying to make it more
7 difficult. I just didn't understand what --

8 Q If it's not in the corpus --

9 A If it's not in the corpus, it's not going to be produced.

10 Q That's right. No matter how good your search strings
11 are --

12 A Right.

13 Q -- correct? Okay. Now, in your experience in managing
14 all of these document reviews, have you ever found documents
15 from sources other than custodians that were responsive?

16 A Absolutely.

17 Q Thank you. So if you failed to search in those areas, it
18 wouldn't be reflected in the corpus, would it?

19 A It wouldn't be reflected -- in the test corpus those
20 documents would not be present.

21 Q Thank you. Now, did you use the same sets of reviewers
22 for all of the tasks that you've described?

23 A Are you speaking about the document review or about the
24 ESCA portion?

25 Q In the review.

1 A I'm sorry?

2 Q In the review. That is, when you looked at the four
3 custodians, did one group look at the null set and a different
4 group look at what you have called the combined set?

5 A Okay. I refer to that as the ECA portion. Yes, I and two
6 other attorneys were responsible for that portion of the
7 exercise.

8 Q For both the combined set and the null sets?

9 A That's correct. The way we did that typically was we were
10 very concerned about the issue of bias. So null sets, for
11 example, were divided into -- whatever number that we needed to
12 test were divided into assignments, and those assignments were
13 split between the attorneys.

14 Q And so were the attorneys that -- were the attorneys aware
15 of which set contained -- which set was the null set and which
16 set was the combined set?

17 A Yes.

18 Q Were you involved in the determination of which filters to
19 apply?

20 A By filters do you mean like the date range?

21 Q Date range, domain names, anything like that.

22 A Not really, no.

23 Q Do you know who was?

24 A I would assume Georgia Pacific.

25 Q Are you familiar with Expedia.com?

1 A Yes. It's a website that deals with travel matters,
2 travel reservations, travel packages, things like that.

3 Q Okay. And do you know whether Expedia.com was a domain --
4 strike that.

5 Do you know whether ESI that related to the domain
6 Expedia.com was excluded from the review set that you were
7 given?

8 A It was not.

9 Q It was not excluded?

10 A It was not excluded.

11 Q So you found documents from Expedia.com?

12 A I know that their domain name was not excluded. I don't
13 know if there were documents that had Expedia in them.

14 Q Would you agree with me that if there were RFPs that were
15 seeking information about the travel arrangements or actual
16 travel engaged in by certain executives, that Expedia.com
17 should be included within the corpus?

18 A I honestly don't know how Georgia Pacific organizes their
19 travel information, so I don't know if they use Expedia or not.
20 The way the Expedia came up was when the documents were first
21 moved into Clearwell, there were certain processes that were
22 performed. And standard in every eDiscovery matter that I've
23 ever worked in, specifically de-duplication and deNISTing,
24 which I believe those terms were discussed this morning, it is
25 not uncommon to apply a domain -- a set of domain name filters

1 that are historically regarded in the industry as being, for
2 lack of a better term, junk e-mails or junk domain names. Such
3 as, you know, ESPN.com, WSJ, The Wall Street Journal.com, maybe
4 Expedia, Orbitz, that kind of thing.

5 But in this particular case that was not done. We
6 looked at that possibility, and we just didn't find that many,
7 many domain names of that nature that it warranted the exercise
8 of suppressing them.

9 Q So did you look through the corpus to determine the number
10 of domain names referring to travel sites were there before you
11 determined whether or not to exclude that?

12 A We have a list of domain names.

13 Q Okay. Are you the person who is responsible for doing the
14 statistical testing and reporting?

15 A Yes.

16 Q So if you would refer please to Plaintiffs' 5 -- I'm
17 sorry. The November 22nd letter.

18 A Thank you. I would appreciate it if you'd refer to the
19 letters by their dates. Yes.

20 Q On page 5 there I believe you'll see a reference to a
21 confidence level?

22 A Yes.

23 Q And did you help author that sentence?

24 A I did not author this letter, no.

25 Q Just the sentence.

1 A I provided information that was undoubtedly used in the
2 construction of the sentence, yes.

3 Q Does the letter accurately represent the information that
4 you provided?

5 A No.

6 Q Do you have the results of the statistical testing with
7 you here today?

8 A I can tell you what it was.

9 Q Do you have the results here with you today?

10 A I do not.

11 Q Do they remain in your files?

12 A Yes.

13 Q Did Georgia Pacific's counsel have access to them before
14 November 22nd, 2011?

15 A I don't know. Georgia Pacific's internal counsel would
16 have. I honestly don't know.

17 Q You don't know whether Quinn Emanuel did or not?

18 A I would have -- I don't know the answer to that question.
19 What this sentence -- there's another -- well, let me go back
20 by way of explaining this sentence --

21 MR. MOGIN: Your Honor --

22 THE WITNESS: -- if I may. The work that we did was
23 to -- in most of our projects we strive for an error rate that
24 is perhaps 10 percent. Meaning that in the null set we don't
25 wish to have more than 10 percent false negatives.

1 BY MR. MOGIN:

2 Q If I understand what you've testified here today, then the
3 statistics that are generated based upon review of a null set
4 are incredibly important to your process, is that true?

5 A That is absolutely correct.

6 Q Now, you testified that there were, what, three reviewers,
7 yourself and two other reviewers that were involved in the
8 subsequent review?

9 A Yes.

10 Q And you were involved in the search string development as
11 well. Were the others?

12 A Yes.

13 Q So all three of you helped develop the search strings and
14 then you reviewed both the null set and the combined set?

15 A On the last validation exercise I personally did not do
16 either in fact. The other two attorneys did. In the other
17 validations I would have done at least one of them, one of the
18 samplings.

19 MR. MOGIN: If I could have just a moment, Your
20 Honor. I think I'm --

21 THE COURT: Sure.

22 (Brief pause.)

23 BY MR. MOGIN:

24 Q Yes, one -- a couple of things. You did mention that you
25 conducted a random sample, is that right?

1 A Yes.

2 Q And was that a statistically valid random sample?

3 A I asked KPMG to prepare a statistically valid random
4 sample using -- first using a software tool called Raosoft,
5 which would estimate the sample size that I would need for a
6 certain confidence interval and a certain margin of error.
7 After I -- and that program gives me the sample size. I then
8 gave that information to KPMG and asked them using whatever
9 tool Clearwell possesses or whatever process Clearwell
10 possesses to generate a true random sample of the size that I
11 requested.

12 Q Do you know how it was done?

13 A Do I know how Clearwell randomizes documents?

14 Q Do you know how --

15 A I have no idea. I have no idea how Clearwell randomizes
16 documents.

17 Q Do you understand that the term random sample is a term of
18 art in statistics?

19 A Yes.

20 Q And do you believe that your process complied?

21 A I do.

22 Q Now, when you ran the null set, you found, what is it, 15
23 documents that were marginally responsive?

24 A Which null set?

25 Q Well, if you look at the letter of November 22nd.

- 1 A I'm sorry. Which letter?
- 2 Q The letter of November 22nd.
- 3 A Okay.
- 4 Q You see on page 3?
- 5 A Okay.
- 6 Q The last bullet point. "Georgia Pacific in consultation
7 with COC --
- 8 A Yes.
- 9 Q -- sought to validate the null sets."
- 10 A Yes. There's 27 marginally responsive documents.
- 11 Q 27 marginally responsive documents in a null set of 660
12 documents?
- 13 A That's correct.
- 14 Q So what percentage is that?
- 15 A That would be about 4.1 percent.
- 16 Q So the confidence interval would be what?
- 17 A It would be -- well, it's 99 percent level of confidence.
18 The margin of error is 5 percent. So it would be 4.1 percent
19 plus or minus 5 percent.
- 20 Q Okay. Now --
- 21 A So between 91 and a hundred or 9 -- yes.
- 22 Q So 9 percent of the documents that are, as you've termed
23 them, marginally responsive --
- 24 A The --
- 25 Q -- within the null set, correct?

1 A 4.2 percent of the null set as we tested it were
2 marginally responsive, the 4.1 percent.

3 Q And you used the term marginally responsive.

4 A Yes.

5 Q Okay. Now, responsiveness if you're going to use the
6 terminology of information retrieval is a binary decision, is
7 it not? It's either responsive or it isn't. It's either a
8 true positive or a false positive, correct?

9 A In information science I think that would be correct.

10 Q So marginally responsive is your lawyer characteriza-
11 tion --

12 A Absolutely.

13 Q -- of the relative responsiveness of a document?

14 A That is correct.

15 MR. MOGIN: Thank you. I'm done, Your Honor.

16 MR. NEUWIRTH: We have no questions, Your Honor.

17 THE COURT: Okay. Well, thanks, Mr. Brown.

18 THE WITNESS: Thank you.

19 THE COURT: Thanks for coming. Can you wait a little
20 longer?

21 THE WITNESS: Oh, yes.

22 THE COURT: All right. Good. Have a seat.

23 (Witness excused.)

24 THE COURT: Okay. You want to call your first
25 witness.

1 MR. WOZNIAK: We'll call Tim Hanners.

2 TIMOTHY HANNERS, PLAINTIFFS' WITNESS, DULY SWORN

3 THE COURT: Yes, sir.

4 MR. ECHOLS: I'm sorry, Judge. Just one point of
5 order.

6 THE COURT: Sure.

7 MR. ECHOLS: This morning you made clear to us that
8 with the experts here that any of the four topics were good to
9 be allowed to have testimony received on them, but you also
10 made clear that there was one topic on preservation which was
11 not at issue, not anticipated for this now. When we filed our
12 motion in limine earlier, we didn't have Mr. Hanners' report,
13 which we got Thursday night. And he spends a bit of time on
14 the topic of preservation. It's not at issue for this
15 proceeding. It's not before the Court. It raises individual
16 defendant specific issues that would require each of us to
17 cross-examine him.

18 And I just wanted to be clear that that's a line that
19 we believe isn't one of those areas that ought to be allowed
20 for these purposes.

21 THE COURT: Were you intending to go into
22 preservation?

23 MR. WOZNIAK: Absolutely. The scope of the search
24 and the efforts of defendants to identify and preserve
25 documents have, in fact, been at issue.

1 THE COURT: Can you do your -- can you without -- I'm
2 sort of stepping on everybody's toes here. Can you do your
3 examination of Mr. Hanners and leave that towards the end. I
4 need to review -- I have to review what you're talking about.
5 I have Mr. Hanners here. I'll look at it now.

6 MR. WOZNIAK: I can say that with some level of
7 confidence that I don't believe we are going to get into any
8 defendant specific issues that we have identified at least by
9 name. We will be talking about some opinions that Mr. Hanners
10 has reached based on written responses that defendants have
11 provided to us.

12 THE COURT: Okay. I'm just asking you in the next 10
13 minutes can you talk about something else?

14 MR. WOZNIAK: Sure.

15 THE COURT: So I get --

16 MR. WOZNIAK: Other than preservation. Certainly we
17 can --

18 THE COURT: Right. Right.

19 MR. WOZNIAK: Sure.

20 THE COURT: I just need to go back -- I just must
21 have missed that. Okay.

22 MR. WOZNIAK: Okay. I'll do my best, Your Honor.

23 THE COURT: Right.

24 MR. WOZNIAK: I think this may be a little tricky,
25 but I think we can certainly go through Mr. Hanners'

1 qualifications and other matters.

2 THE WITNESS: Is that going to preclude the other
3 topic of discussion as well in my report?

4 MR. WOZNIAK: Well, let's first --

5 THE COURT: Well, you'll jump up. You'll jump up if
6 we're getting there.

7 THE WITNESS: Okay.

8 MR. ECHOLS: Okay.

9 THE COURT: And I made a broader statement. It's not
10 if it's one question as part.

11 MR. ECHOLS: Absolutely, Judge. I understand that.

12 THE COURT: If it's not question. If it's one
13 question, it's not worth it. We ought to move on, okay.

14 MR. WOZNIAK: Your Honor, for the record Robert
15 Wozniak for the plaintiff.

16 THE COURT: Thanks, Mr. Wozniak.

17 DIRECT EXAMINATION

18 BY MR. WOZNIAK:

19 Q Mr. Hanners, can you please state your full name for the
20 record.

21 A Timothy Drew Hanners.

22 Q And, Mr. Hanners, what --

23 MR. NEUWIRTH: Your Honor, could we ask for the
24 witness to be sworn in.

25 MR. WOZNIAK: I think he --

1 THE COURT: I did.

2 THE WITNESS: I already did.

3 THE COURT: I did. Didn't I swear you in?

4 THE WITNESS: Yes, ma'am, you did.

5 MR. NEUWIRTH: I didn't hear it.

6 THE COURT: No, actually the court reporter just
7 asked me too. Either I have such a calm little voice you
8 didn't hear you.

9 MR. NEUWIRTH: Thank you.

10 BY MR. WOZNIAK:

11 Q All right. So you've stated your name. Mr. Hanners, can
12 you please state your current occupation.

13 A I am the owner of a computer forensics investigations
14 company called 1st Forensic Consulting located in Lucas, Texas.

15 Q What do you mean by computer forensics?

16 A I provide computer forensic services. And the term
17 computer forensics is generally associated with the
18 preservation -- identification, preservation of computer based
19 evidence for a matter before the court.

20 Q And what was the last part?

21 A In a matter that's coming before the court, legal matter
22 or hearing.

23 Q And so do you have any specialized training in the area of
24 computer forensics?

25 A Well, my experience in computers started back in 1982

1 where I was working with the Stealth fighter program. I worked
2 on aircraft related computer systems. And then in 1989 I
3 joined the Air Force Office of Special Investigations, United
4 States Air Force Office of Special Investigations, where I
5 became a general investigator with them. I later specialized
6 in doing computer crime investigations, commonly termed as CCI
7 as a special agent with them.

8 I conducted a large number of investigations
9 involving federal, state criminal matters, foreign intelligence
10 services, any number of crimes that involved a computer at the
11 time. Since that time -- while with them I attended
12 specialized training in the field of computer forensics. I
13 also attended the Federal Law Enforcement Training Center and
14 other specialized computer forensics training courses.

15 Upon my retirement from the United States Air Force I
16 worked with Bank of America, where I attended additional
17 computer forensics courses. I worked as a special -- or vice
18 president of information security located in Dallas, Texas.
19 Part of my duties there involved the collection and
20 preservation of information within the bank. After that period
21 of time I worked as a contractor for Computer Sciences
22 Corporation out of D.C., working with the United States Postal
23 Inspection Service lab in Pittsburgh running their cases. They
24 basically went in and cleared up a three-year backlog.

25 As part of that, we received additional training

1 there. I have attended other courses and classes. After
2 working with the Postal Inspection Service, I worked for a
3 company called Vogon International, who taught computer
4 forensics. I went through their computer forensics training,
5 as well as instructed in their computer forensics training, to
6 include U.S. Secret Service, U.S. Treasury Service, federal and
7 state law enforcement agencies both in the States and overseas.
8 After they were bought out by Kroll Ontrack, I worked with
9 another company called Xact Data Discovery, which is the EDD
10 Company based out of the Dallas area for the last four and a
11 half years till last August, when I started up my own company.

12 Q I'm going to hand the court reporter what I believe will
13 be marked as Plaintiffs' Exhibit 8.

14 Mr. Hanners, I've just handed to you your resume,
15 which was previously provided to the Court as an exhibit to
16 plaintiffs' opening brief in this matter. We've already
17 covered quite a bit of your experience and your background.
18 I'd like to turn your attention to the third page. You list
19 some court experience there, and it indicates that you have
20 testified as an expert in several cases?

21 A Yes, it does.

22 Q Without going into the detail of each case, can you just
23 briefly tell me what those cases involved and what the nature
24 of your testimony was.

25 A The nature of my testimony was similar in all. It all

1 revolved around computer based evidence and my review of that
2 evidence in all but one, which was a capital murder case where
3 I testified both on the computer evidence as well as
4 evidentiary searches for physical evidence.

5 Q Any of these cases involve large scale ESI collections?

6 A The last one there, John Doe versus Peters School District
7 involved over 30,000 users. Their information was reviewed in
8 that matter. In addition to the ones that are documented here,
9 a large number of my cases were settled prior to going to full
10 trial. And those cases involved extremely large sets of data
11 and include 420 terabytes of data in past cases with
12 information coming off tape in those cases to up and to more
13 recent last year 24 terabytes of data at alive on servers that
14 was being reviewed for ESI.

15 Q Have your credentials as a testifying expert in the field
16 of computer forensics ever been challenged?

17 A No.

18 Q And you've never been excluded or prevented from
19 testifying as an expert?

20 A No, I have not.

21 MR. WOZNIAK: Your Honor, I'd like to move that
22 Plaintiffs' 8 be offered into evidence, and we would also
23 tender Mr. Hanners as an expert in the field of computer
24 forensics.

25 THE COURT: Okay. What's your position?

1 MR. MENDEL: Your Honor, we'd like to have you
2 reserve ruling until after cross-examination.

3 THE COURT: Sure. That's fine. Thank you.

4 BY MR. WOZNIAK:

5 Q Mr. Hanners, why were you retained by plaintiffs in this
6 matter?

7 A I was retained to review the defendants' replies as they
8 relate to doing a targeted collection of custodians for ESI
9 versus that of a different technology that might be used or a
10 bigger scope or volume of information.

11 Q Did you review any materials in connection with the work
12 that you have done in this matter?

13 A Yes, I did.

14 Q Did you review the complaint?

15 A Yes, I did.

16 Q Did you review plaintiffs' document requests that were
17 served on defendants?

18 A Yes, I did.

19 Q Did you review the ESI or production format stipulation?

20 A Yes, I did.

21 Q Did you review the 30 (b) 6 deposition notice that was
22 served on defendants?

23 A Yes, I did.

24 Q And did you review written responses that were provided by
25 defendants to plaintiffs in connection with that 30 (b) 6

1 notice?

2 A Yes, I did.

3 Q And you reached some conclusions and observations in
4 connection with your review of all of those materials?

5 A Yes, I was able to.

6 Q And those conclusions and observations are written in your
7 preliminary report of findings that was submitted to the Court
8 last week, isn't that right?

9 A Yes, they are.

10 Q Is there anything that you've reviewed -- well -- yes, let
11 me ask you. Is there anything that you have reviewed
12 subsequent to submission of your report in connection with the
13 30 (b) 6 deposition process?

14 A There was a deposition of one of International Paper's
15 employees that I reviewed. I believe it was Mary -- I'm trying
16 to recall her last name. It started with S. I'm sorry. I
17 don't recall her last name.

18 Q Do you know whether that was a rough or final transcript
19 of that deposition?

20 A It was a rough transcript of the deposition. It was made
21 available to me within the last couple of days.

22 Q And did your review of that transcript change any of your
23 conclusions or findings that are included in your report?

24 A No, it did not.

25 MR. WOZNIAK: Your Honor, we would move to include --

1 well, actually let me -- I'm sorry. I'm getting ahead of
2 myself. Let me hand up and we'll have marked as Plaintiffs'
3 Exhibit 9 I believe we are on, Mr. Hanners' preliminary report
4 of findings. And that was submitted to the Court last
5 Thursday.

6 MR. MENDEL: Your Honor --

7 MR. WOZNIAK: I'm not moving it into evidence quite
8 yet, but --

9 THE COURT: Did you want to say something?

10 MR. MENDEL: Yes, we do object to any use of this
11 document at this stage of the proceedings. It is all about
12 preservation. That's all this document addresses.

13 MR. WOZNIAK: Well, I would say it's about
14 identification and preservation. And it also goes directly to
15 the heart of a key dispute up to this point, which is whether a
16 custodian based approach to identifying and collecting
17 documents is sufficient as compared to a subject matter
18 approach, which is one plaintiffs had been urging from the
19 outset.

20 THE COURT: And how does that -- we're going to have
21 to take it one step at a time on preservation. Here's what
22 this gentleman says: "None of the defendants provided
23 sufficient detail about their identification and preservation
24 process to permit any reasonable assessment of their processes
25 as adequate." Okay. We've heard six hours of identification.

1 Okay. I think anybody sitting here could tell you something
2 about what you folks did on identification. There hasn't been
3 anything about preservation, which is the reason I said what I
4 did. He says he didn't have enough to say anything, so I don't
5 know what he's going to talk about.

6 MR. MENDEL: That's exactly our point, Your Honor.
7 Then he goes on and writes a whole report about something he
8 says he doesn't have any basis for.

9 THE COURT: Well, that goes to something else, okay.
10 I mean, I -- I guess we have to take it on a
11 question-by-question basis. Okay.

12 MR. MENDEL: May I have a standing objection, Your
13 Honor, to this report.

14 THE COURT: And you can keep objecting when we get to
15 preservation. On identification I think we've had information
16 on your identification process. Okay.

17 MR. MENDEL: I think the identification process that
18 was described was how we take a corpus of documents and
19 identify the relevant documents. What I believe Mr. Hanners is
20 going to be talking about is identifying that corpus initially
21 and the preservation of those documents. That is a different
22 issue. And it's much earlier in the process, and it is not
23 what this testimony today has been about.

24 THE COURT: All right.

25 MR. MENDEL: It's not about search terms. It's not

1 about predictive coding.

2 THE COURT: Can we have him ask the first question so
3 then I can figure out where we are. Okay. So we're not
4 moving -- the report is certainly not -- you can talk to him
5 about the report.

6 MR. WOZNIAK: Before I even attempt to move it in as
7 evidence I was going to lay a foundation and ask him some
8 questions about the report and other work that he's done to
9 this point.

10 THE COURT: Right.

11 BY MR. WOZNIAK:

12 Q So, Mr. Hanners, if I could turn your attention to the
13 last, very end of your report. There's an attachment there, a
14 chart, a diagram, if you will.

15 A I'm there.

16 Q Okay. And you recognize this diagram?

17 A Yes, it's an electronic discovery reference model.

18 Q And what is that?

19 A It's a concept that was put forth by the Evergreen Group.
20 It's called the EDRM model. It was a group that was
21 self-established back in the early 2002 time frame I believe to
22 bring forth the best practices for doing ESI collections
23 together and formulate plans and -- or godets for doing general
24 type of ESI collections.

25 And you'll notice in the first steps of it, the first

1 step actually applies to a company as it would maintain its
2 information on a regular course of business to ensure it was
3 properly handling information. The second step labeled
4 identification is the very first step in the ESI process
5 according to the EDRM model.

6 Q And this is an area -- in your experience in computer
7 forensics your work is primarily focused on which of these
8 boxes?

9 A Well, generally speaking, for large eDiscovery cases such
10 as this my chief focus is in the identification and preserva-
11 tion steps, the steps that are taken prior to any work being
12 done with the information to subdivide it. It's basically an
13 evidentiary identification and preservation process to identify
14 evidence to make sure it doesn't change.

15 Q And why is that important?

16 A Well, with the volatile nature of computers, for instance,
17 say, a desktop as an example, information can change as a user
18 changes -- opens Word documents, saves Word documents. He
19 could accidentally delete a document not out of intent, but out
20 of mistake. E-mail could be changed. E-mail can be saved on
21 the local system by the user. What we're basically trying to
22 do is go in there at the very beginning of the ESI process and
23 grab a snapshot of that information so it can be set aside and
24 preserved for additional work.

25 MR. MENDEL: Your Honor, if I may. Again, this is

1 all about preservation. It's not about any of the issues we've
2 talked about this morning.

3 MR. WOZNIAK: If I may, Your Honor. Using this
4 diagram as an example, I mean this is designed to show a
5 process by which an ESI collection comes about. And without
6 identifying and preserving data, one can't go on to the next
7 step of collecting and processing that data. So these are
8 inextricably tied together. We've made clear throughout in
9 both our position statement submitted to Judge Shadur going
10 back to mid-December and continuing forward to the present,
11 we've been very consistent that the scope of the search in this
12 case is directly at issue.

13 And defendants have consistently taken the position,
14 for instance, that they are only obligated to search, for
15 example, live active servers in terms of searching for --
16 identifying as a first step potentially responsive ESI. So
17 that's one example of how I think Mr. Hanners' testimony is
18 directly on point here.

19 MR. MENDEL: Your Honor, the scope of the search and
20 whether that search is to involve search terms or predictive
21 coding is an issue before Your Honor. Document preservation,
22 which is what Mr. Hanners has been talking about, what his
23 report is all about, is not in issue here. There's no motion
24 before this Court about whether or not the defendants have
25 properly preserved their documents.

1 THE COURT: Well, that's what I'm starting to -- you
2 have to tell me what you mean. I mean -- what I mean by
3 preservation is document preservation. Okay. Are you folks
4 using preservation in a different way?

5 MR. WOZNIAK: We are using preservation as it's
6 connected to identification. I mean, you can't preserve what
7 you haven't identified, and you can't search what you haven't
8 preserved. So it's all tied together. And as Mr. --

9 THE COURT: I mean, but you've just now interjected
10 backup tapes. I mean, I just -- before I go down this road and
11 get in -- I thought you meant document preservation in the
12 sense that there was an accusation that documents weren't
13 preserved here. And I don't think any of the motions I have
14 have to do with whether or not documents have been preserved.

15 MR. WOZNIAK: We -- if you're talking about a
16 spoliation motion or something of that nature, no, we don't
17 have any motion on the table.

18 THE COURT: Right.

19 MR. WOZNIAK: What we do have is our very clear
20 position that was set out in the December 15th position paper
21 that was submitted to Judge Shadur where we talk about the
22 scope of the search and the nature of the search that should be
23 conducted if one is going to adequately search for potentially
24 responsive ESI. And we have -- for example, Mr. Brown just
25 testified that if something is not in the corpus, it can't be

1 found. So I think that just serves to reinforce our view that
2 these issues are all tied together.

3 MR. MENDEL: That has nothing to do with
4 preservation.

5 THE COURT: All right. All right. That's why I
6 would say that's inadequacy of the search method myself as
7 opposed to the term preservation, but -- and I think that the
8 term preservation sets off all kinds of bells on people, legal
9 bells and whistles that -- I thought you were talking about --
10 I thought this fellow had reviewed what each of them had done,
11 and he was going to opine that he didn't think they did such a
12 good job, and one of the reasons they didn't do such a good job
13 are they didn't include everything that maybe could have been
14 included.

15 Now, you're calling that preservation. I'm not
16 calling that preservation. I'm talking about the accuracy of
17 the search that was done.

18 MR. WOZNIAK: Fair enough, Your Honor. And I
19 would --

20 THE COURT: And I think it's the preservation -- I
21 mean, you were expecting him to say you didn't do such a good
22 job.

23 MR. MENDEL: Your Honor, we had read his expert's
24 report.

25 THE COURT: Right.

1 MR. MENDEL: Which you have before you, Your Honor.

2 THE COURT: Right.

3 MR. MENDEL: It has nothing to do with the adequacy
4 of the search. It is all about preservation.

5 THE COURT: Well --

6 MR. ECHOLS: And, Judge, this is -- part of the
7 reason this affects all of us and all of us differently is that
8 we understood when Mr. Hanners was identified that there was a
9 dispute about whether backup tapes, for instance, could or
10 should be searched or not. Now, that was one of those topics
11 for a later day. We understood as Your Honor said in ruling on
12 our motion in limine that there is an issue -- there's a
13 potential issue concerning if the identification of those
14 custodians used for developing a search methodology were
15 sufficient.

16 That's totally separate from whether any of us
17 defendants used appropriate preservation methods in identifying
18 the sources, locations of ESI to be preserved. You're
19 absolutely right, that's never been raised either before Judge
20 Shadur or before Your Honor. The source of collection of
21 documents, yes, that may be disputed. But the preservation is
22 an individual issue. We've had requests from plaintiffs for
23 information from us. We've responded to those requests. It's
24 ongoing. Some depositions have taken place. Letters are being
25 written, but it's entirely individualized, and nothing is ripe

1 and certainly nothing before this Court on the preservation
2 issue the way you define preservation, Judge.

3 MR. FREED: Your Honor, may I be heard on this.

4 THE COURT: Yes. Maybe you can help us out here.

5 MR. FREED: I hope so, Your Honor. Because I think
6 what's happening here is the ultimate question is whether the
7 defendants will be producing the documents which are responsive
8 to the document requests. And that starts the five step
9 process, which their own expert started talking about, and we
10 can't jump up and say that wasn't appropriate for discussion.
11 If they have the best search methodology in the world but
12 they've excluded 95 percent of the relevant documents, there's
13 a deficiency in their production.

14 And what this witness is going to go to is the whole
15 collection effort, the location effort, the identification
16 effort. The discussion of ESI is only a portion of the big
17 picture, which is are they giving us everything that they
18 should be giving us? Have they employed all of the techniques
19 that they should have employed? Did they do the right thing
20 when they started the process? Did they do the right thing
21 when they searched the documents? He's giving a background on
22 that. And he -- and that is something which is relevant to the
23 ultimate determination which I believe Your Honor is going to
24 have to make, is have they responded correctly and properly to
25 the requests for production?

1 He's prepared to explain why in very material
2 respects he has seen evidence which indicates they haven't.
3 Now, you can't just compartmentalize and say, no, I just want
4 to talk about search. That's all I want to talk about, because
5 we have maintained from the beginning that the whole body, the
6 corpus -- they keep talking about this corpus -- is this all
7 encompassing? The corpus is not all encompassing. The corpus
8 is a fraction perhaps of what they should have been doing. And
9 you can't understand the application of the search terms and
10 how they could apply to the rest of the corpus, which should be
11 analyzed as well unless you know what the whole corpus ought to
12 be.

13 So I don't think you can just make that artificial --
14 you know, it's like saying -- well, I'll leave it at that.

15 MR. MOGIN: Your Honor, I would add very briefly that
16 if you go back and you look at the plaintiffs' statement of
17 position as well as the joint statement of position and the
18 defendants' statement of position, all of which were submitted
19 to Judge Shadur in connection with the December 15th hearing,
20 you will find that the first issue is defined in two parts.
21 Part one, to summarize, is Boolean versus content based, and
22 part two is custodians versus subject matter.

23 THE COURT: Right.

24 MR. MOGIN: And, in fact, subject -- and Mr. Hanners
25 is here to talk about custodians versus subject matter.

1 MR. ECHOLS: And, Judge, we have no objection --

2 THE COURT: And there's no problem.

3 MR. ECHOLS: We have no objection to that.

4 THE COURT: No problem.

5 MR. ECHOLS: Custodians and subject matter.

6 THE COURT: It's no problem with that at all. Okay.

7 I mean, I want him -- I want to hear, I want to hear on that
8 topic, okay. I think it's when --

9 MR. MENDEL: And, Your Honor, his entire -- none of
10 his report addresses --

11 THE COURT: Well, let's not get there yet. Now, come
12 on. We don't even know what he's going to say here today. I
13 mean, I -- I mean, I want to hear what he's got to say. I
14 was -- my back went up on the problem is they didn't preserve,
15 okay. And you don't know if they -- because I don't -- I think
16 what this gentleman says in his report is he doesn't know what
17 they preserved because it's not there. Okay. No, he doesn't
18 know.

19 And the backup tapes or archive or not reasonably
20 accessible. The reason I was cutting you off is we have enough
21 to do today on what has been done already and you guys are
22 jumping to something we haven't done. That was the reason I
23 was cutting you off on that. I'm not saying on another day.
24 We obviously on a long term discovery have to talk about that.
25 We're not ready for it right now. I've been interrupting

1 lawyers all day, and I hated it when judges did it to me.
2 Okay. I don't want to be micromanaging this gentleman's direct
3 examination. I know what you mean by preservation. I wish you
4 could find -- where's that linguist?

5 (Laughter.)

6 THE COURT: Let's get her in here, okay, to give me
7 another term other than the P word, and you can talk about it.

8 MR. MOGIN: Does that mean you're granting our motion
9 for partial reconsideration?

10 THE COURT: Sure. At 12:00 o'clock tonight.

11 (Laughter.)

12 THE COURT: Now, let's just see what you can do --
13 let's start to hear what he did, what his study heard. I'm
14 hearing you on this. If nothing else let's find out what it is
15 this gentleman's going to say. Okay. Sit down, and you're
16 more than welcome if you want to make it for the record jump
17 up. And I'm sorry to interrupt you. Okay.

18 MR. WOZNIAK: That's understood, Your Honor. No
19 problem. I will do my best to proceed --

20 THE COURT: Right.

21 MR. WOZNIAK: -- and elicit testimony --

22 THE COURT: Right.

23 MR. WOZNIAK: -- that will not cause defendants to
24 keep jumping on their feet.

25 THE COURT: Right.

1 MR. WOZNIAK: We'll see how that goes.

2 BY MR. WOZNIAK:

3 Q Mr. Hanners, I'm going to refer back to your report only
4 for the limited purpose of talking about the first section
5 where you identify what you call general principles for the
6 reliable identification of preservation processes. And the
7 only question that I have is whether those general principles
8 as you list them are based on anything more than your own
9 opinion?

10 A Yes, they are. The EDRM Group was a collaboration of --
11 it basically read like a who's who of who's on eDiscovery. If
12 you looked down, you saw all the significant players, the
13 KPMGs, all the other major significant vendors, as well as
14 attorneys and practitioners of computer forensics were
15 attending and making inputs. And they basically came up with
16 some standards and some guides. One being an identification
17 guide to help you through the ESI discovery process.

18 Q Let's talk about identification. And specifically you
19 understand that the defendants in this case, and we've heard
20 testimony to this effect already today, the defendants have
21 taken a custodian based approach to identifying potentially
22 responsive ESI?

23 A Yes.

24 Q Do you agree with a custodian -- that a custodian focused
25 approach is the best way to locate all potentially responsive

1 ESI?

2 MR. MENDEL: Objection, lack of foundation, Your
3 Honor.

4 BY MR. WOZNIAK:

5 Q Let me ask you this way: Are you familiar with an
6 approach other than a custodian based approach that one might
7 take in attempting to identify all potentially relevant ESI in
8 a given lawsuit?

9 A Well, if I can equate this to just as a general
10 investigator. If I look at a matter that's before me, whether
11 that be electronic evidence or any other form of evidence, if I
12 don't go in and consider all the available evidence that's
13 presented to me, then I'm liable to rule out information before
14 I've ever considered it.

15 A case may be that where I go into a home to search
16 for computer information. If I never go into a room, I don't
17 know if there's computer media in there. I don't know if
18 there's a floppy disk sitting in there that I haven't searched.
19 So if I take a targeted search and equate say the living room
20 to being a custodian and I only search there, I'm not searching
21 for anything else. If I do a more topical or subject matter
22 search or consider potential evidence being in the whole house
23 or the whole business, then I go in with that in mind. My
24 focus is wide open, my eyes are wide open. I'm looking. I'm
25 asking the relevant questions.

1 And my part in the initial identification process is
2 to help ensure that the right questions are asked. Working
3 with inside counsel, working with outside counsel, working with
4 maybe HR, IT, and others within the business that are key
5 players that have a corporate knowledge that I don't possess.
6 My goal is to go in and ensure that they're addressing all
7 their available media areas, to include how the business works,
8 how do they communicate, how do they store information? And
9 then from that we would assign tasks for individuals to go out
10 and do.

11 Part of that might be a mapping to go out and collect
12 where is this information sent. A lot of businesses have
13 diagrams of computers, but they may be months or years old. So
14 we ask them to update that to bring back to the table so an
15 informed decision can be made of all the possible storage
16 locations. An example would be during an identification phase
17 I would remind the legal counsel that when they're talking to
18 people, you need to ask them do they store information on USB
19 devices, on external hard drives, on CDs, do they work from
20 home? Is that information being preserved?

21 And we generally would say there's some volatile
22 information out there. PCs are volatile. Stuff that's in
23 rotation is volatile. Archives that can be overwritten would
24 be considered a volatile area that you need to look at
25 preserving that information. Take the necessary steps to stop

1 activity from happening that's going to change information.

2 Q Okay. You've said a lot there, and let me --

3 A I'm sorry.

4 Q -- try to bring this back to the defendants taking a
5 custodian based approach in this case. And what I've heard you
6 say is that you would recommend taking a broader approach?

7 THE COURT: Okay. You have a gentleman standing up.
8 What would you like to say, sir?

9 MR. MENDEL: Your Honor, I object to the question
10 being unclear as whether now he's asking about a custodian
11 based approach to preservation or a --

12 MR. WOZNIAK: Purely the best approach --

13 THE COURT: No, he said --

14 MR. WOZNIAK: Purely for purposes of identification.

15 THE COURT: For identification. I think that's what
16 he set up. There's a custodian based approach to
17 identification and there's a subject matter approach. And I
18 think what this gentleman was referring to was subject matter
19 approach or mapping he called it.

20 BY MR. WOZNIAK:

21 Q I'll rephrase. What I think I heard you say is that a
22 subject matter approach would result in a greater volume of
23 potentially relevant ESI being collected or identified. Let me
24 use identified.

25 A During the identification phase you want to consider all

1 available sources of information based on activity and based on
2 the matter, not based upon preselecting a target group of
3 people to go out and look for because you've identified a much
4 smaller group of -- than the possibly available evidentiary
5 material.

6 Q If you were to start from scratch, I'm just going to ask
7 you to run me through the steps of how you would sort of soup
8 to nuts go about identifying -- and you've done some of this
9 already. But if I -- say you're brought in as a computer
10 forensics expert and it's a large scale ESI project. You are
11 asked to identify -- and we'll put it in the context of this
12 case. It's a price fixing conspiracy case. You're asked to go
13 about identifying all potentially responsive ESI. What's the
14 first step you take?

15 A The first step I would take would be as previously
16 mentioned, getting together a team within the company or
17 working with a team in the company. That being I'm generally
18 called in by counsel or management at some point and introduced
19 to the rest of the team. I would ensure that that team
20 includes a good corporate knowledge from IT. They know where
21 most information resides.

22 I would make sure that someone from HR was inclusive.
23 They know the current people working at the company and the
24 ones that are former employees. You can identify that based on
25 relevant time period. I would ensure that the -- any number of

1 people from management that might have knowledge of the matter
2 and that have the sufficient clout to ensure that the processes
3 that we identify can be followed and that there's not going to
4 be any interruptions in those processes. That a senior manager
5 can say that they're going to do it and it's going to do it
6 appropriately.

7 I would try to identify the immediate -- my immediate
8 concern would be let's identify the type of media that's out
9 there that can be changed while we're sitting here talking and
10 let's take the immediate steps to stop that from changing. I
11 would then -- for things that would be missing from that
12 group's information base such as the active known computers
13 within the company, I would send them out to gain that
14 additional process through doing an active network mapping to
15 identify systems that are connected and turned on to their
16 system, to actively seeking out the sources of available ESI.

17 Q Let me stop you there. You've talked -- if I could
18 rephrase a part of what you just said. You said -- I think it
19 sounds like you said you would assemble basically a cross
20 functional team as a first step.

21 A That's correct.

22 Q And what if management of the company in question was
23 potentially involved or was alleged to have been involved in
24 some wrongdoing, would that impact in any way the scope of your
25 identification effort?

1 A Well, generally speaking, from an investigative
2 background, my background, I would go in and tell -- talk with
3 someone that's not involved from legal. Whether that's inside
4 counsel or outside counsel and advise them that anyone
5 connected to the matter should not be a decision point during
6 the initial identification part of the process. That someone
7 else should be appointed for that period of time to be able to
8 make those decisions that -- if it's the CEO, somebody lesser
9 in the company should make the decision. And if it's somebody
10 else in the company as far as IT is involved, they should be
11 locked out of the process.

12 Q And if I might follow up, why? Why would you want someone
13 other than those that were potentially involved to be a part of
14 the identification process?

15 A Well, potentially you're asking for a person that may be
16 named later as part of the allegation to identify documents and
17 collect documents that may potentially prove they did
18 something. It's not generally -- the investigator tells me
19 that's not the smartest thing in the world to do. I would
20 equate that to having a physical crime scene, somebody --
21 somewhere where somebody may have gotten stabbed and asking
22 the person that did the stabbing to go in and collect the knife
23 for me.

24 MR. WOZNIAK: Your Honor, I might have -- I wanted to
25 ask a question about a portion of Mr. Hanners' brief where he

1 says -- and I don't intend to get into this in great detail,
2 but he has a section in here that talks about preservation
3 being only as good as the identification project which preceded
4 it. And I wanted to simply ask him what he means by that.

5 THE COURT: Don't you want to know what he means by
6 that?

7 MR. WOZNIAK: Well, I'm just -- I guess I'm --

8 THE COURT: No, the gentleman behind you who's
9 standing up.

10 (Laughter.)

11 MR. MENDEL: Your Honor, I think it's pretty clear
12 he's going to be talking about preservation.

13 THE COURT: He doesn't know what you did. How can he
14 say anything? I mean, duh. I mean, by what we know about
15 preservation or what --

16 MR. WOZNIAK: I simply want to ask him that question
17 and --

18 THE COURT: Does he know what each of the eight
19 defendants -- has he ever seen a litigation hold? Has he ever
20 done anything about the preservation that actually happened in
21 this case?

22 MR. WOZNIAK: Well, he -- I mean, he has reviewed the
23 written responses that the defendants provided to us in
24 response to our 30 (b) 6 notice, and those written responses
25 were supposed to provide information about their identification

1 and preservation efforts in this case, including the
2 information relating to litigation holds.

3 THE COURT: I really want to hear what he says from a
4 forensic standpoint, okay, about the identification and the
5 search that was done. I think that's why we dragged him here.
6 Not whether or not they had -- but it's your case. It's your
7 case. I just got your last point about self-selecting, about
8 potential people doing self-selecting of word search. Okay. I
9 did get that. Okay. I did get that. That was in your brief.
10 Now, I don't know -- I still don't understand where the
11 preservation is coming from, which is why this gentleman keeps
12 jumping up.

13 MR. WOZNIAK: Well, I think in a general sense if I
14 can just say -- make the point that if something -- you've
15 identified something as potentially responsive.

16 THE COURT: Right.

17 MR. WOZNIAK: Okay. If you don't preserve that or
18 you choose not to preserve that because you -- well, let's say
19 you haven't identified it. Okay. You've taken a custodian
20 based approach and so you've ignored certain chunks of
21 potentially responsive ESI because you've made the
22 determination that only certain custodians are going to have
23 responsive documents in this case.

24 THE COURT: Right.

25 MR. WOZNIAK: Okay. So there's this --

1 THE COURT: Void.

2 MR. WOZNIAK: -- group of documents that are not part
3 of the corpus.

4 THE COURT: Right.

5 MR. WOZNIAK: And I simply want to ask Mr. Hanners
6 some questions that go to exactly what he's written in his
7 report, that preservation can only be as good as the
8 identification project that precedes it. And I think that's a
9 general question that's fair game, and it doesn't get to --

10 THE COURT: Go ahead and ask it. Just go ahead and
11 ask it because we have to move on to the next topic.

12 MR. WOZNIAK: Fair enough.

13 THE COURT: Okay.

14 BY MR. WOZNIAK:

15 Q What do you mean, Mr. Hanners, when you say preservation
16 is only as good as the identification project which preceded
17 it?

18 A Well, if your identification fails to document all
19 available sources of information, sources like you could be
20 talking about backup tapes, you could be talking USBs, or how
21 the company works, how the individuals work, if you fail to
22 address that and bring that back to the table and identify it
23 within your working documents that you're going to carry
24 forward to the preservation process, then you're going to be
25 missing information. You're not going to collect on that. And

1 as a result, relevant ESI could be left behind at the onset
2 before any searches or other processes are done with the
3 information.

4 And part of that you basically take the list that you
5 generated in the I.D. section once you define what is going to
6 be preserved to the preservation process. And generically --
7 and once I start preserving information I'm going to document
8 all that information. Everything is preserved, who did it,
9 when, how. So it's building my chain of custody up. This is
10 how I identified it. This is what I did to keep it from
11 changing condition, called a preservation step or whatever
12 terms you choose to apply. And that information typically sits
13 aside and is retained so it's never changed or modified,
14 following the best evidence rules.

15 You go seek the best evidence, you gather it, you
16 store it in a method that it can't be changed or modified. And
17 typically with electronic media we're going to work off a copy
18 of that information so you're never going to touch that
19 original evidence. It's done for the simple point of fact that
20 we want to be able to go back to the beginning. We have the
21 documentation that says how we identified, what we identified,
22 who we identified, or the process we used. And it's defensible
23 from that viewpoint. Excuse me if I ran on, Your Honor.

24 Q A quick follow-up. You said about potentially altering
25 documents I believe you said or images.

1 A Right. Well --

2 Q And I actually have a follow-up question on that.

3 A Okay.

4 Q I want to know if what you meant by that has to do with
5 metadata that's associated with a given electronic file, for
6 instance?

7 A Well, metadata is information about documents, about other
8 electronic information, and it's contained within the scope of
9 ESI. It contains information like last saved by, other
10 information. If you're not properly identifying it and not
11 properly, quote, the term preserve generically, I'd consider it
12 collecting it and setting it aside, if you're asking someone to
13 look and see if that document's relevant, just the mere fact of
14 opening it could change the last access date.

15 If I moved it to another computer for whatever reason
16 as part of the ESI process, then I've changed the date created
17 for that file. So that information is lost to the reviewer at
18 some future date. They would assume that the date created is
19 true and accurate or they would have to question and backtrack
20 and see if the information is still there. Certain types of
21 ESI are volatile, and that comes out in the identification
22 process up front. We know that PST files are volatile that sit
23 on people's systems. We know that other information are
24 volatile based on just how a computer works. We know that we
25 can't really as a whole trust that information is going to be

1 there tomorrow.

2 What happens if the individual's hard drive crashes?
3 What happens if the computer is taken away by somebody and is
4 no longer available? A laptop that disappeared, stolen. So we
5 want to go ahead and preserve that information -- collect that
6 information at the initial or identify it for collections.

7 Q We'll call it collect and set aside.

8 A Collect and set aside. I'm sorry.

9 Q Let me ask you specifically, with respect to metadata if a
10 document is, for instance, accessed in the process of trying to
11 determine whether it's responsive and the date last accessed is
12 changed, the date last modified just as an example. That
13 document is then produced to a party in litigation. Does the
14 party receiving that document have any way of knowing -- I mean
15 can you tell then whether that document was produced as it was
16 kept in the ordinary course of business by the producing party?

17 A Well, if the path statement is changed as part of your
18 identified process that you're going to use later, then it's no
19 longer available to the end user. At some point it's pulled
20 into these ESI collection tools, and it's based upon what's
21 represented to the ESI tool as to where it originated at.

22 Say I moved a file from my laptop to a server and I
23 didn't include the full pathing information, the full date
24 range and all the stuff we do with computer forensic imaging
25 process as we collect, then that information's changed and

1 modified. If it's taken from that location and later
2 represented in an ESI tool, it's going to show that as the
3 location where it came from with the dates associated and any
4 other changes that occurred.

5 Q You mentioned earlier including archived media sources, I
6 believe you called them as part of the identification process,
7 is that correct?

8 A That's correct.

9 Q What are some examples -- well, let me just cut to the
10 chase. Are backup tapes an example of an archived source of
11 ESI?

12 A Yes, it is.

13 Q And what is typically contained on a backup tape? Let's
14 say you have a backup tape of an exchange server. What would
15 be contained on that tape?

16 A Well, a backup is a snapshot in time of information,
17 whatever was targeted. Whether that be -- say, an example
18 would be exchange server. If I targeted the EDB file to back
19 it up so I have everybody's e-mail that's contained within that
20 EDB file, that's a snapshot from whenever it's made. So if the
21 tape was made yesterday, it's reflective of information that's
22 fairly current. If it's a three-year old tape of e-mail, then
23 it probably has no comparison or very low comparison to what's
24 actually in the person's current e-mail. The older you go, the
25 further -- more differences you would expect to see.

1 Q So it fair to say that if one was seeking to go back eight
2 or nine years, one would likely find nonduplicative e-mail on a
3 backup tape as opposed to what's contained on an active server?

4 MR. MENDEL: Objection, Your Honor. Lack of
5 foundation.

6 THE COURT: What's the objection?

7 MR. MENDEL: Lack of foundation, Your Honor. And
8 also relevance to this matter.

9 THE COURT: I will overrule it. He can answer it. I
10 don't know that this has to do with choosing between word
11 search and a concept analysis, but --

12 MR. WOZNIAK: Well, it goes more to the scope of the
13 search, which is again something that we've placed at issue I
14 believe pretty consistently. But I only have a few more
15 follow-up questions.

16 THE COURT: Go ahead.

17 THE WITNESS: I'm sorry. Can you repeat the
18 question?

19 BY MR. WOZNIAK:

20 Q The question was, the likelihood that an eight-year-old
21 tape, let's say, would contain duplicative or nonduplicative
22 information as opposed to what's contained on a live active
23 server.

24 A Well, if you look back eight years, you're going to see
25 many generations of changes that take place in an eight-year

1 time frame. The whole e-mail system could have swapped out
2 multiple times if you're referring to e-mail. People that come
3 and go from a company may no longer be there, so their
4 information on an eight-old-year tape would be certainly
5 nonduplicative of information present say on a file server or
6 found within the e-mail server.

7 Q Based on your review of the written responses, the 30 (b)
8 6 letters from defendants, do you know whether defendants have
9 revealed the existence of backup tapes in their possession?

10 A Yes, they have.

11 Q And did they provide information about the content of
12 those tapes?

13 A Some did.

14 Q Okay. For those that did not provide any information, is
15 there a way of going about determining the content?

16 A Yes. Generically it's called cataloging of tapes. There
17 are -- and whether they have the equipment or not, there's
18 companies that produce catalogs of tapes that are out there.
19 And a catalog is nothing more than what you would typically see
20 in Windows as a directory structure. It would show you the
21 full path down to file name but do so in a text basis. It
22 would show you the file name, the size of the file, and maybe a
23 MD5 hash of the file. And that's doable for the entire
24 contents of the tape.

25 Q And would that then allow you to see whether the

1 likelihood that there were -- there was duplicative or
2 nonduplicative ESI on that tape?

3 A With the use of an MD5 hash, yes, you could de-dupe the
4 files based upon that.

5 Q And do you have personal experience cataloging backup
6 tapes?

7 A Currently I use a company called RenewData out of Austin,
8 Texas. They do the cataloging work for me. And past companies
9 I've worked with we did it in-house.

10 Q What are the typical costs in your experience associated
11 with cataloging a backup tape?

12 A Well, for a generic tape I've received quotes this year of
13 approximately \$50 per tape. Typically the company I work with
14 if you submit a hundred tapes or more, that cost comes down.

15 Q How common is it in your experience to access backup tapes
16 in the context of litigation?

17 A Well, when I was working with a larger company, we did it
18 on a fairly routine basis.

19 Q Which means? Can you elaborate?

20 A Well, companies that are involved in litigation -- I've
21 had banks come to me with a couple of thousand tapes. I've had
22 a company here in Chicago I did work for in the past that
23 bought me 3700 tapes to have the information as part of an SEC
24 finding saying their information wasn't available. They needed
25 to de-dupe, actually de-dupe and retain one copy, but they

1 wanted to retain a historical record of where all -- that copy
2 may have existed among all the possible custodians. That
3 process is certainly out there and available.

4 Q In your experience is a broad approach to identifying and
5 collecting potentially relevant ESI significantly more
6 expensive than taking a more narrow approach, for example,
7 looking only at certain selected document custodians as opposed
8 to basing your identification on subject matter or an entire
9 department?

10 A Well, lacking -- or just stating a fact, last year in St.
11 Louis I went in to collect say, I think the figure was 50
12 custodians. While I was on-site -- and I had projected three
13 days to do so. While I was on-site during that collection
14 during the three days they added, I think the total came out to
15 be 107 custodians. The cost to the client was nil because I
16 was able to work it in in the same time frame that the normal
17 custodians I was there to collect. Their machines were
18 available at the same time. So there was no appreciable cost
19 for that.

20 Q And would it make a difference if you take a -- let's say
21 you take a subject matter approach to searching or for
22 identifying potentially relevant ESI. How does that compare to
23 the costs of -- again if you had to sort of estimate, how much
24 more costly, if at all, it would be to take a more narrow
25 custodian based approach?

1 MR. MENDEL: Objection, Your Honor. Lack of
2 foundation.

3 THE COURT: All right. Lack of foundation as to the
4 type of subject matter. I mean --

5 MR. WOZNIAK: I'm willing to scratch the -- or pull
6 the question.

7 THE COURT: Okay.

8 MR. WOZNIAK: I don't think it's important.

9 THE COURT: Thank you.

10 BY MR. WOZNIAK:

11 Q Strike it. And really what I want to get at is just to
12 sum up, based on your experience you typically take a subject
13 matter or a broad approach to identifying potentially
14 responsive ESI?

15 A I always take a much broader approach than a targeted
16 approach. I would go with a subject matter approach or global
17 approach to collecting ESI within a division, group, or
18 department.

19 Q And again, if you could just quickly tell me, summarize
20 why that is. Why you believe it's important to do so.

21 A Well, early on and in a litigation matter you may know or
22 may not know the full scope of the information that may be
23 requested. So we try to collect as much information as we can,
24 and we don't know at a later on date that additional
25 individuals could be named. And if you don't collect it up

1 front, you don't have an opportunity to build a defensible
2 process saying we preserved everything that was available to
3 us.

4 If you have to go back at later date it's more
5 costly. I generally point that out. And my recommendation is
6 always collect broadly. And you're not processing it. You're
7 setting it aside until a determination can be made.

8 Q And you said more costly. If, for instance, you wanted to
9 add a particular custodian, if one had taken a custodian
10 focused approach and one later learned that there was another
11 person that should have been included but you didn't learn this
12 until a year later, you might not even be able to do that
13 realistically, isn't that right?

14 A Well, if they've upgraded the person's computer since
15 then, the prior information could have been lost. Certainly if
16 you were looking for latent data, that would be gone with the
17 previous computer. E-mail systems. A person could have left,
18 a person could have died. Any number of factors could come
19 into play there that really make it no longer a viable
20 solution.

21 MR. WOZNIAK: I don't think I have any further
22 questions, Your Honor.

23 THE COURT: Okay. Do you have any cross?

24 MR. MENDEL: Yes, Your Honor. Can we take a break,
25 Your Honor.

1 THE COURT: Yes. Why don't we take 10 minutes.
2 Okay.

3 (Short break taken.)

4 MR. MENDEL: In light of the fact, Your Honor, that
5 there are ongoing 30 (b) 6 discussions between the parties that
6 are about a lot of what Mr. Hanners just testified about --

7 THE COURT: You have no questions?

8 MR. MENDEL: No questions.

9 THE COURT: Okay. Thanks, Mr. Brown. You're
10 excused.

11 THE WITNESS: I guess I'm done.

12 THE COURT: You're excused.

13 THE WITNESS: Pleasure meeting you, Your Honor.

14 THE COURT: Thank you.

15 MR. MOGIN: This is Mr. Hanners.

16 THE COURT: Mr. Hanners. I'm sorry.

17 THE WITNESS: That's okay. I'm from a large family.

18 (Witness excused.)

19 MR. MOGIN: Your Honor, we'd like to call Dr. David
20 Lewis, please.

21 DAVID LEWIS, PLAINTIFFS' WITNESS, DULY SWORN

22 MR. MOGIN: Your Honor, with respect to the time, if
23 I could have a little latitude with respect to the form of the
24 questions, I think we can get this information in before our
25 deadline.

1 THE COURT: Sure. Absolutely.

2 DIRECT EXAMINATION

3 BY MR. MOGIN:

4 Q Dr. Lewis, would you please tell us your educational
5 background.

6 A I have a bachelors degree in mathematics and a bachelors
7 degree in computer science from Michigan State and a masters
8 degree and Ph.D in computer science from the University of
9 Massachusetts at Amherst.

10 Q Are you also a fellow of the American Association for the
11 Advancement of Science?

12 A Yes, I am.

13 Q Could you tell us a little bit about that.

14 A I was elected in 2006 for contributions to the algorithms
15 and evaluation of the use of supervised learning and
16 information retrieval.

17 Q And just a one sentence definition of supervised learning
18 and information retrieval, please.

19 A Supervised learning is learning from examples. People say
20 this document, say, belongs to one category, this document
21 belongs to another. And computer algorithms learn to tell the
22 difference between the two.

23 Q Okay. And how are you currently employed?

24 A I'm an independent consultant.

25 Q And can you tell me the nature of your consulting

1 practice, please.

2 A I consult on information retrieval, machine learning,
3 natural language processing, and the statistical evaluation of
4 systems in these areas.

5 Q Okay. And natural language processing, would you give us
6 the one sentence definition of that, please.

7 A Computer analysis of language.

8 Q Do you currently work with any eDiscovery vendors?

9 A Yes, I do.

10 Q And what type of eDiscovery vendors do you work with?

11 A I work with a company called Kroll Ontrack. I'm a
12 consultant for them, and they provide eDiscovery services. I
13 recently started consulting for a small Chicago company called
14 Nextpoint, which provides Cloud based eDiscovery services.

15 Q What do you do, or what did you do for Kroll?

16 A I have designed algorithms for supervised learning and
17 statistical ranked retrieval, and I have also designed
18 algorithms for statistical evaluation of those technologies.

19 Q And you mentioned a small company that you're working with
20 now. What have you done for them?

21 A I've done some preliminary work for them on how they might
22 use supervised learning, but that has not been implemented yet.

23 Q Have you done any teaching in the field of information
24 retrieval?

25 A I've taught a number of tutorials at conferences in

1 information retrieval, computational linguistics, and
2 statistics.

3 Q Have you published any papers that have been peer reviewed
4 in scientific journals?

5 A Yes, I've published a number of peer reviewed journal
6 articles.

7 Q And are the articles that you have so published attached
8 to your resume that was submitted to the Court?

9 A Yes, they are.

10 Q And is that resume true and correct and accurate?

11 A Yes. I haven't updated it for a few months. I noticed
12 that there's an article in the *Journal of Artificial*
13 *Intelligence and Law* which is listed to -- as to appear, but
14 that actually has appeared both in print and online now.

15 Q And what was that article?

16 A That's an article on the evaluation of information
17 retrieval systems in electronic discovery.

18 Q Do you have any patents in the field?

19 A Yes, I believe eight of them have issued so far.

20 Q And can you tell us in general terms what those patents
21 are for?

22 A They're patents on information retrieval and machine
23 learning.

24 Q Very good. Now, very briefly, Dr. Lewis, could you give
25 us the definition of the science of information retrieval.

1 A Information retrieval is the science of developing methods
2 for better access to data such as textual data, where the
3 meaning of the data is somewhat subjective and the information
4 needs are complex subjective information needs.

5 Q Are there other disciplines that are involved in
6 information retrieval besides computer science?

7 A Yes. Issues of computational linguistics come to bear,
8 issues of statistics come to bear. Database technologies
9 sometimes come to bear.

10 Q And could you give us a one sentence definition please of
11 computational linguistics.

12 A Well, I used the term natural language processing earlier.
13 The two are essentially the same. They're the study of how to
14 develop computer systems that perform meaningful tasks on
15 linguistic data.

16 Q And what is the relationship of the field of statistics to
17 the science of information retrieval?

18 A So statistics is used in two ways in information retrieval
19 systems. First it's used for -- in various techniques for
20 improving the performance of the systems. And second, it's
21 used in various ways to evaluate the performance of the system.
22 And those are two different bodies of statistics.

23 Q Please tell us how it's used to evaluate the performance
24 of a system.

25 A Well, you would typically draw a random sample from some

1 universe of documents that one needs information access to.
2 Review elements of the -- review the elements in the random
3 sample to determine which categories each element should belong
4 to. You then need to run the information system on the
5 universe of documents and see what the outputs of the system
6 are on that universe of documents. And then finally, you need
7 to compute some statistical estimate of the effectiveness of
8 the system by comparing the system's decisions to the manual
9 judgments.

10 Q And how would you express that last statistic?

11 A Excuse me. I didn't quite hear the question.

12 Q How would you express that last statistic that you
13 described?

14 A Oh, well, it's common to use a confidence interval for
15 expressing these forms of estimates.

16 Q In layman's terms what's a confidence interval?

17 A So a confidence interval consists of three quantities. So
18 first it's worth mentioning a confidence interval is an
19 estimate of a particular statistic; that is, a particular value
20 on the universe of documents. The confidence interval has
21 three parts. It has a central value, the sort of expected --
22 the point estimate expected value. It has a margin of error,
23 which is sort of a degree of uncertainty on that estimate. And
24 then it has a confidence level, which is an expression of how
25 confident you are that the size of the sample drawn was

1 representative of the universe of documents.

2 Q So confidence level refers to the random sampling process,
3 correct?

4 A Yes.

5 Q Now, could you please tell us what is from a statistical
6 perspective as it's used in information retrieval, what is a
7 random sample?

8 A A random sample of a given size or a simple random sample
9 of a given size, because there's several random sampling
10 techniques. So a simple random sample is a sample that's drawn
11 in a fashion that every set of that size from the universe has
12 an equal probability of ending up being the sample.

13 Q All right. And what is the margin of error in layman's
14 terms, please?

15 A The margin of error is simply how certain you are about
16 the estimate.

17 Q And what is -- again, what is the estimate as you have
18 just used that term?

19 A So the estimate would be typically a confidence interval
20 is expressed as a particular central value plus or minus a
21 margin of error, so that that point estimate is that central
22 value. It's -- typically it's the value that would be most
23 probable to be the actual value. But that somewhat depends on
24 the details of the sampling.

25 Q And what is the central value that's important to

1 information retrieval in the context of this case?

2 A The most important statistic here is recall. My
3 understanding of the discovery process is that it's very
4 important to find most of the responsive documents. Recall
5 is -- recall is the proportion of all the responsive material
6 in the universe to be searched which the system has managed to
7 find.

8 Q And you've heard other witnesses today use the term
9 precision, correct?

10 A Yes, I have.

11 Q And precision is a term of art in information retrieval,
12 correct?

13 A Yes, it is.

14 Q It has a different meaning than it does in everyday
15 conversation, correct?

16 A That's correct.

17 Q And what is -- precisely what does precision mean?

18 A Precision is of all the material that an information
19 retrieval system found and identified as being interesting,
20 what proportion of it was actually responsive. So it's a
21 measure of how much junk was in the stuff that the system
22 found.

23 Q So why is recall, understanding recall more important than
24 understanding precision in this context?

25 A Well, recall goes to the extent to which the need to find

1 responsive documents has been satisfied. Precision is largely
2 an issue of what it's going to cost to go through the stuff
3 that the system found. So you can always compensate for
4 failures in precision by doing more manual review, running
5 through the material with other tools, but there's no way to
6 compensate for a failure in recall. If you're never given the
7 material, there's nothing you can do about it.

8 Q Now, in this particular case have you had the opportunity
9 to review the defendants', first Georgia Pacific's proposed
10 search methodology?

11 A Yes, I have.

12 Q And can you tell us in general terms what you have
13 reviewed.

14 A I reviewed the November 22nd letter and there were
15 subsequent briefs and documents which provided a couple other
16 versions of the same Georgia Pacific process that was described
17 there. I've also looked at the complaint in the case and the
18 requests for documents. I've looked for the -- looked at the
19 Boolean search strings from the various defendants. I've
20 looked at Dr. Tenny's and Mr. Hanners' reports. And I listened
21 to the testimony here in court today.

22 Q All right. Now, with respect to defendants' search
23 strings, do you understand those to be Boolean queries?

24 A Yes, they appear to be Boolean queries.

25 Q And can you give us a little bit of your background from

1 information retrieval about what Boolean queries are, what they
2 do, and what are their limitations.

3 A So a Boolean query is a logical expression on search
4 terms. It combines search terms which are intended to be
5 exactly matched against the corpus. And a search term might
6 have things like a wild card, so that would be an exact match
7 against words that end in several ways. Combined with logical
8 operators *and* and *or* and *not* often there are proximity
9 operators which allow expressing that certain words are near
10 each other. So that's what a Boolean query is. It's a
11 technology that's been used in information retrieval for a very
12 long time.

13 Some of the limitations of it are it requires an
14 exact match against the document so that the user using a
15 Boolean query has to specify exact matching conditions.

16 Q What are the ramifications in a document review situation
17 of the requirement of an exact match?

18 A Well, it would depend on the request for proposals -- I'm
19 sorry, request for documents.

20 Q Could you elaborate, please.

21 A Sure. So, for instance, some of the requests for
22 documents in this case refer to ideas that can be expressed in
23 a very wide range of fashions linguistically. Notions like
24 raising prices, changing capacities of factories. There's a
25 very -- these are very broad topics, a very wide range to talk

1 about them. So it is difficult to express these sorts of
2 notions in an exact match framework.

3 There's also requests for documents related to
4 alleged illegal activities. One can expect that given that my
5 understanding is that some of this -- in this industry there
6 has been previous litigation, that people would be cautious, if
7 they were undertaking illegal activities, they would be
8 cautious in the way that they referred to them. They would be
9 unlikely to use easily anticipatable terminology when
10 discussing such activities.

11 Q Are you saying then that one of the limitations of Boolean
12 queries are that it requires the user to anticipate in advance
13 of reviewing the corpus the exact words that might be found in
14 documents?

15 A That's correct.

16 Q Now, Dr. Lewis, are you involved in something called TREC?

17 A Yes, I am.

18 Q And TREC is a project sponsored by the National Institute
19 of Standards, is that right?

20 A That's correct.

21 Q And can you tell us about your involvement in TREC.

22 A Yes. I was one of the founding members of the TREC
23 program committee in 1992, and I've served on the TREC program
24 committee most, but not all of the years since then. I was
25 also one of the cofounders of the TREC Legal Track. TREC Legal

1 Track was a subtask within TREC that was focused on studying
2 the behavior of information retrieval systems on simulated
3 electronic discovery tasks.

4 Q And TREC stands for Text REtrieval ...

5 A Conference.

6 Q Conference. All right. So in the TREC studies have there
7 been any findings regarding -- strike that.

8 TREC has conducted certain studies where they've
9 compared Boolean queries, the effectiveness of Boolean queries
10 to content based queries, is that correct?

11 A Yes.

12 Q And what have been the results of those studies?

13 A Well, there's two series of studies that are relevant
14 there. There was one series of studies that had to do with
15 comparing Boolean querying with statistical ranked retrieval.
16 Statistical ranked retrieval was found to be more effective in
17 the last of those studies where the methodology had been
18 refined.

19 Q Let me stop you right there. Give us the one sentence
20 definition of statistical ranked retrieval, and then we'll go
21 back to what TREC did.

22 A Okay. Statistical ranked retrieval refers to technologies
23 that accept a query, usually simply a list of words, and
24 produce a ranking of documents taking into account the
25 statistical properties of words and phrases in the entire

1 collection of documents.

2 Q All right. Now, you were telling us about TREC and its
3 comparison of Boolean to other methodologies.

4 A Right. So there was a -- there was comparison of
5 statistical ranked retrieval with Boolean queries. There was
6 also a series of three studies that compared supervised machine
7 learning methods with Boolean queries. And again in the last
8 of those three when the -- and which was the one that had the
9 most refined methodology, the best supervised learning system
10 had an effectiveness level of 250 percent of the Boolean system
11 that was compared.

12 Q So now you've referred so far in your testimony to three
13 different search systems, is that correct?

14 A That's correct.

15 Q Boolean, statistically ranked retrieval, and what you've
16 called supervised learning; correct?

17 A That's correct.

18 Q And what's the relationship from your perspective of
19 supervised learning to what other people have described as
20 predictive coding?

21 A Predictive coding is a term that is used in the eDiscovery
22 industry. My understanding is it refers to the use of
23 supervised learning in eDiscovery. There is -- I should
24 mention I prefer not to use the term predictive coding because
25 there's some legal controversy. A company called Reconnind has

1 claimed certain trademark rights to that phrase, so I prefer to
2 use the term supervised learned.

3 Q Very good. Now, are you familiar with or have you become
4 familiar with a term called content based advanced analytics?

5 A Yes, I have.

6 Q And have you developed an understanding of that term?

7 A Yes. My understanding of content based advanced analytics
8 is that it's one of several terms that's used in information
9 technology to refer to technologies that include statistical
10 ranked retrieval, supervised machine learning, and a variety of
11 unsupervised learning methods such as clustering and latent
12 semantic analysis.

13 Q Now, very good. Now, based -- is there anything else that
14 you can inform the Court with respect to the relative
15 capabilities of these three methods before we begin to discuss
16 Georgia Pacific's specific methodology; that is, Boolean versus
17 statistically ranked retrieval versus supervised learning?

18 A The major thing that I think is important to understand
19 about the three technologies is the relative burden that they
20 put on the need of a user to anticipate what the language is in
21 the universe of documents to be searched. A Boolean query
22 system requires specifying an exact match on search terms
23 use -- expressing that in logical operators.

24 A statistical ranked retrieval system relaxes the
25 demands on the user by allowing them to simply provide a list

1 of words and phrases. And a supervised learning system relaxes
2 that demand even more by allowing the user to express their
3 information need in the form of examples. This is responsive,
4 this is not responsive. With the computer algorithm doing the
5 work of figuring out which words distinguish responsive from
6 nonresponsive documents and how much weight to pay attention to
7 those words.

8 So if we compare supervised learning back to Boolean
9 querying, the identification of which words are important can
10 be done by the algorithm. The identification of how much
11 attention to pay to those words is done by the algorithm. And,
12 in fact, in a Boolean framework you can't even express this
13 sort of relative importance of words.

14 A final thing I would mention is that a supervised
15 learning algorithm is essentially unbounded in the amount of
16 information that it can make use of. You can provide it more
17 and more and more labeled examples, and it will get more and
18 more and more value out of them.

19 Q Now, in your --

20 THE COURT: I have a question, though.

21 MR. MOGIN: Certainly.

22 THE COURT: Is the SRR word based or concept based,
23 though?

24 THE WITNESS: Statistical ranked retrieval -- so I do
25 not like to use the word concept based when talking about

1 information retrieval systems because it's so ambiguous. It
2 has been used to refer to a wide range of technologies. A
3 statistical ranked retrieval system can take advantage of
4 words. It can take advantage of phrases. It can take
5 advantage of the metadata if the system is configured in that
6 fashion. But I would prefer to avoid using the word concepts,
7 Your Honor.

8 THE COURT: But it does involve words?

9 THE WITNESS: It does -- yes, it can use words. It
10 can use phrases. It can use metadata.

11 THE COURT: Okay. All right.

12 BY MR. MOGIN:

13 Q When you say it can use words, do you mean Boolean type
14 words or other configurations of words?

15 A Well, I mean words -- the words are words. The question
16 is what the information retrieval system does with them. A
17 statistical ranked retrieval system uses the statistical
18 properties of language and can pay differing attention to
19 different words. So again, the statistical ranked retrieval
20 system is different from Boolean in that sense. Boolean,
21 either the word is used or not in certain logical combinations.
22 A statistical ranked retrieval system computes a numeric weight
23 for each word and determines how much attention to pay to the
24 word.

25 Q Can you think of an example when we compare Boolean to

1 statistical ranked retrieval that would apply to this case
2 perhaps based upon the RFPs that are in issue?

3 A Well, sure. So if you were -- excuse me. If you were
4 looking for one of these concepts that's difficult to express
5 like a price increase, there's many many different ways to talk
6 about price increases. You could -- so you would either have
7 to anticipate that in a Boolean query, or you could use, you
8 could use a query to a statistical ranked retrieval system and
9 be able to match documents at least based on all of the words
10 you listed, which could be a big long list of words. You don't
11 have to combine them with logical operators. And there would
12 be differential weighting of those.

13 And then if in addition there's a supervised learning
14 capability, or for that matter if there was certain forms of
15 unsupervised learning, such as latent semantic indexing, you
16 also would be able to match on words that were not anticipated
17 in the original query.

18 Q What do you mean match on?

19 A Match on.

20 Q Match on.

21 A Match on. Sorry. Yes.

22 Q Okay. Now, you used the term anticipate. What does
23 anticipation involve in this context?

24 A Well, the point being that if you're using an exact match
25 system, you have to anticipate exactly the combinations of

1 words that will appear in the documents that you're looking
2 for. In a statistical ranked retrieval system you still have
3 to anticipate something there. You have to anticipate at least
4 some of the words that are going to occur in the documents
5 unless you're applying -- again, there's some of these other
6 technologies like latent semantic indexing where you could get
7 a match even with no identical matches on the query.

8 And then finally, in supervising learning there's no
9 anticipation at all. You simply look at examples and you say
10 this is responsive, this is not responsive, and the system does
11 the rest.

12 Q Now, you're aware that in this case one of the examples
13 that the plaintiffs have put forward is the phrase *they are*
14 *with us*, correct?

15 A I am. I understand that.

16 Q Could you please tell us how each of these systems would
17 handle -- could you please tell us the likelihood of each of
18 those three systems being able to find such a document
19 containing that phrase *they are with us*.

20 A Well, I would say that it's important to recognize that
21 the individual words *they*, *are*, *with*, and *us* are very all high
22 frequency terms. And that's going to pose a difficulty for any
23 system that's using only the words in the query. So it's going
24 to be extremely difficult to handle in a Boolean query system.
25 You know, if the Boolean query had access to metadata, you

1 know, maybe you could, you know, gin up something with the
2 metadata to try to get at those relevant documents.

3 The statistical ranked retrieval system, you know,
4 you would have a somewhat better chance if you are using the
5 metadata in a statistical fashion. But, you know, frankly
6 getting something like that is -- you're only even going to
7 have much of a decent shot with a supervised learning system,
8 and for that matter with a supervised learning system that has
9 access to the metadata so that it could latch onto things
10 besides the words. You know, time of day, custodians, file
11 path names, you know, was something stored in an unusual place,
12 things like that.

13 And these are the kinds of systems -- these are the
14 kinds of things that are very difficult for a person to
15 anticipate, but where a supervised learning system has some
16 hope of finding patterns in the data and finding them.

17 Q Now, you heard Mr. Hanners' testimony about the
18 possibility of metadata being altered or not properly
19 maintained, correct?

20 A Yes, I have.

21 Q And could you tell us if that situation were to occur,
22 what would happen with the use of these systems of metadata as
23 you've described it.

24 A Well, you know, any time the metadata is distorted, the
25 system is losing evidence that it could otherwise use to find

1 responsive documents.

2 Q Now, would supervised learned be able to find a document,
3 label it as responsive bearing the phrase *they are with us*?

4 A Yes, if it has access to the metadata. I think it's
5 unlikely it would find it if it had only the text to work with.
6 You know, maybe if there was routine use of these -- of that
7 particular phrase in the context of responsive documents, it
8 might get it. But it's going to be pretty hard. Those are
9 high frequency words.

10 Q The same with *they're with us*?

11 A Yes, I mean, it's the same problem. The words have very
12 high frequency and they occur in a lot of different documents.

13 Q *They're okay with that*?

14 A It's the same issues.

15 Q Okay. Now, have you reached a conclusion -- some
16 conclusions about Georgia Pacific's proposed search
17 methodology?

18 A Yes, I have.

19 Q Can you tell us what your conclusions are.

20 A My conclusions are first that it cannot be relied upon to
21 find a substantial proportion of responsive documents. And
22 second, that it cannot be relied upon to produce a
23 statistically valid estimate of its own effectiveness.

24 Q Let's start with the first. It cannot be relied upon to
25 produce -- I'm sorry. What did you say?

1 A To find a substantial proportion of the responsive
2 documents.

3 Q It cannot be relied upon to find a substantial portion of
4 the documents. Please explain your reasons.

5 A Well, there's -- there's several reasons. The first is
6 that the process of developing their queries was done on --
7 without having collected all of the sources of responsive
8 material.

9 Q Can you explain that further.

10 A Yes. My understanding from the report and the testimony
11 of Mr. Hanners is that there are substantial sources of
12 potentially responsive material that are not associated with
13 particular custodians. And thus were not material that was
14 present at the time Georgia Pacific did their -- applied their
15 protocol.

16 Q Okay. And you said you had a number of reasons. What's
17 your next reason?

18 A Yes. The second reason is that Georgia Pacific did the
19 development of their queries on a set of five arbitrarily
20 selected custodians.

21 Q And can you explain what's wrong with that?

22 A Well, yes. The purpose of developing a search query is to
23 distinguish responsive documents from nonresponsive documents.
24 The ability of the query to do that depends on the distribution
25 of words in the responsive and nonresponsive documents. By

1 using an arbitrary subset of the five custodians they've
2 created an artificial distribution of words which is not
3 representative of the entire body of material to be searched.

4 Q So it wasn't diverse enough? Is that what you're saying?

5 A Yes, it's not diverse but it's also not representative.

6 Q Why isn't it representative?

7 A Well, because it's each person's vocabulary is somewhat
8 different. They use language differently based on their life
9 experiences, their job functions, and so on. So if you choose
10 any fixed set of the custodians as a -- it's not going to be
11 representative of the universe of documents that you need to
12 search.

13 Q Okay. And the impact of that would be?

14 A The impact is that again it would degrade the
15 effectiveness of the Boolean queries that they developed by
16 their iterative procedure.

17 Q Okay. Is there any other reason that you believe that the
18 methodology cannot be relied upon?

19 A The third was the question about the spam filtering, the
20 junk filtering that was brought up earlier. That filtering on
21 the string Expedia.com may have removed responsive documents.
22 Now, I was a bit unclear from Mr. Brown's testimony at what
23 point that was no longer being done. In the November 22
24 document it was still described as part of the process.

25 But to the extent to which those responsive documents

1 were removed during the iterative process of query development,
2 that would again degrade the effectiveness of the queries that
3 were produced.

4 Q So what kind of spam or junk are you referring to? Is
5 there any spam or junk that it would be legitimate to remove?

6 A Well, one can think of some obvious examples we all get in
7 our e-mail boxes. You know, if you had a very, very highly
8 accurate junk mail filter and it was applied in a fashion where
9 there was some validation of its effectiveness, you know, I
10 could imagine a responsible use of something like that.

11 Q All right. And are there any other reasons?

12 A Okay. Well, the fourth reason was, is -- okay. Scratch
13 that one because that's an evaluation one. The next reason is
14 the use of the -- there was a process that they undertook
15 iteratively to develop the search queries. And it involved
16 taking samples and reviewing those samples for whether
17 documents were responsive or not. That process as was
18 described by Mr. Brown was done with the knowledge by the
19 people doing the reviewing of whether the system had retrieved
20 the document or not.

21 That is, was the document in what they called the
22 combined composite set or was it in the null net? So the
23 reviewers potentially are biased by the fact that they know
24 what the right answer they're supposed to find is.

25 Q And what are the impacts of that?

1 A Well, review for responsiveness is a very complicated
2 subjective decision. And it's easily affected by contextual
3 factors. And obviously an important contextual factor is
4 knowing what answer is going to make the system look good.

5 Q Now, with respect to the five custodian process that
6 you've heard described and then later iterations of that, is
7 that a random sample as you understand it?

8 A No.

9 Q Could you explain why.

10 A There was no statement that the custodians were chosen
11 randomly or even chosen in any fashion that their documents
12 were meant to be representative.

13 Q So from an information retrieval statistical perspective,
14 what's a random sample? How should it have been done in this
15 case?

16 A Well, the first thing would be to identify the entire
17 universe of documents to which the queries are going to be
18 applied. And then you would draw, for instance, a simple
19 random sample from that universe of documents. And as I
20 mentioned earlier, a simple random sample is a sample such that
21 all samples of that size have an equal probability of ending up
22 being the sample.

23 Q Now, were there any other issues with respect to the
24 first, that is, the methodology and its ability to find
25 responsive documents? Was there any issue of overfitting?

1 A Well, that's really one of --

2 THE COURT: Of what? I'm sorry. Of what?

3 MR. NEUWIRTH: Can you repeat that question.

4 BY MR. MOGIN:

5 Q Was there any issue of overfitting?

6 A That's an evaluation issue.

7 Q All right.

8 A That's not a, that's not an effective -- that's not an
9 issue with the effectiveness of the queries. That's an issue
10 with the evaluation.

11 Q We'll come back to that then. Did they make effective use
12 of the analytical tools that were available to them?

13 A No.

14 Q With re -- go ahead and explain that.

15 A Well, so there was a description in the process and also
16 we heard testimony that Clearwell's topics page tool was used
17 in the process. I examined the Clearwell manual, and the
18 topics page tool is a tool that does document clustering and
19 then also attempts to sort of pull out which words are
20 representative of each cluster.

21 Q And go ahead and explain as much detail as you need
22 document clustering and how it compares to the other
23 technologies you've described.

24 A Well, document clustering falls in what is scientifically
25 referred to as an unsupervised learning process. That is, it

1 is a machine learning technique which finds relationships among
2 documents without any human guidance. So it attempts to group
3 together documents that have similar topic, but without sort of
4 any human guidance as to what those topics are.

5 Q And what's the impact of the lack of human guidance?

6 A Well, the lack of the human guidance is that document
7 clustering methods will find some sort of relationships between
8 documents, but those relationships don't necessarily have
9 anything to do with the particular information need of
10 interest. I've done experiments where I've run, oh, at least
11 10 different document clustering algorithms on the same corpus,
12 and they produced 10 different clusterings of the documents.
13 Each one has its own statistical bias as to what an interesting
14 pattern is, but these don't necessarily line up with the
15 distinctions that people want to make among the documents.

16 Q And so because of those limitations on the concept
17 clustering, are you saying that Georgia Pacific used the tool
18 improperly?

19 A That's limitations on document clustering. No, actually
20 there were two other reasons I felt they used it
21 inappropriately or ineffectively would be a better way to put
22 it. First, and again this was something that I again have some
23 unclarity after Mr. Brown's testimony. In the November 26th
24 letter it was said that the topics page tool was applied
25 separately to the combined composite set and to the null set.

1 The difficulty with that is that it means that documents that
2 were missed by the query cannot be clustered together with
3 documents that were hit by the query. And, of course, the
4 documents hit by the query are the ones that are, you know,
5 presumed to be the most, the most rich in responsive documents.

6 Q You know, let's get to the whole idea of the null set. As
7 you heard Mr. Brown testify, their whole thing from their
8 perspective, their process hinges on the null set. Do you have
9 any opinion with regard to that?

10 A My understanding of Mr. Brown's testimony and of the
11 description in the November 22nd letter to the extent that I
12 could make out what they were saying in that letter, is that
13 their validation process is focused on determining how many
14 responsive documents are in the null set. And that can be a
15 reasonable way to evaluate an information retrieval system.
16 There's some caveats that are important there, however.

17 Q What are the caveats?

18 A Well, the caveats are that it's extremely important if you
19 do you that, that the review of the documents in the null set
20 be consistent with the review of documents that's done for
21 production.

22 Q And did Georgia Pacific follow that process?

23 A Well, I'm not aware of who did the review for
24 production -- well, no, there's a whole -- that's right,
25 because Mr. Brown testified about that. Mr. Brown and two of

1 his colleagues reviewed the null sets, or the samples, excuse
2 me, samples from the null seats. And there was a team of I
3 think 15 lawyers or something that were doing the review for
4 production. So what would be critical is to, you know, have
5 some statistical guarantee that the two sets of reviewers are
6 actually making comparable decisions.

7 The serious problem that could arise is that if you
8 use a stringent criterion for evaluating responsiveness in the
9 null set and then a more liberal criterion during review, if
10 you then actually compared the quantities, you could think,
11 well, you found some, you know, incredibly large proportion of
12 the documents that were out there, but that ratio would not be
13 correct because they were not being reviewed comparably.

14 Q Now, you've heard Mr. Brown's testimony with respect to
15 the number of, as he put it, marginally responsive documents
16 within the null set?

17 A Yes.

18 Q And how does that fit into the paradigm that you just
19 expressed as between stringent and liberal?

20 A Well, that would be -- it's sort of saying that certain
21 documents -- it's really sort of saying certain documents fall
22 into a different class of responsiveness. It would seem to be
23 trying to define certain responsive documents as, well, not
24 really being responsive. And I guess I don't know if
25 marginally responsive has got a legal meaning or something.

1 But I think the thing that most struck me about that was that
2 there was no notion of marginally responsive when they were
3 reviewing the documents the system found. There was only a
4 notion of marginally responsive when they were reviewing the
5 documents that the system had missed.

6 Q Very good. Now, let's go to your other area of criticism,
7 if you will, which was that the methodology cannot be relied
8 upon to produce a statistically valid estimate of
9 effectiveness.

10 A That's correct.

11 Q Can we begin with the reporting of the statistical events.

12 A Okay. So could I see the November 22nd letter?

13 Q It should be there on the witness stand.

14 A I have only stuff from Mr. Hanners here. Oh, is it up
15 here? Yes, I have it.

16 THE COURT: So just for the record you're talking
17 about Plaintiffs' 4, right?

18 MR. MOGIN: Plaintiffs' 4, correct, Your Honor.

19 THE WITNESS: Okay. I have the letter.

20 BY MR. MOGIN:

21 Q You've reviewed this letter previously?

22 A Yes, I have reviewed it several times.

23 Q Okay. Now, I believe that the statistical reporting is on
24 page 5, is that correct?

25 A Yes, it is.

1 Q Now, go ahead and read in for the record what you
2 understand to be the statistical reporting in that letter?

3 A The report takes this form. It is the sentence which
4 says, "Based on this validation process, COC determined with
5 99 percent confidence that the final set of search terms had no
6 more than a 5 percent margin of error in identifying documents
7 as not responsive to plaintiffs' document requests."

8 Q What is your response upon reading that?

9 A That's not a statistical statement.

10 Q Well, what kind of statement is it?

11 A It's a statement which uses statistical terminology, but
12 does not actually express a statistical result.

13 Q Why not?

14 A Because a confidence interval, as I mentioned earlier,
15 requires three things. It requires a confidence level. It
16 requires a margin of error. But most critically, it requires
17 the actual value at the center of the confidence interval.
18 When I took at this statement -- well, it's like this: Suppose
19 you wanted to decide what proportion of the vehicles on the
20 road were pickup trucks, and you did a random sample, and you
21 went and you told somebody, well, there are plus or minus
22 5 percent pickup trucks on the road. And that's analogous to
23 what this sentence says.

24 We don't know if there's 95 percent plus or minus
25 5 percent pickup trucks or 5 percent plus or minus 5 percent

1 pickup trucks or 13 percent. This statement to the extent it
2 communicates anything is completely consistent with every
3 document in the null set being responsive.

4 Q Say that again. I'm sorry.

5 A This statement is consistent with every document in the
6 null set being responsive.

7 Q Is there any -- if you look at the November 22nd letter,
8 you'll see that there is a reference to 27 marginally
9 responsive documents within the null set.

10 A Yes, I have that on page 3.

11 Q And how did you -- what's your reaction to that report?

12 A Well, as I mentioned earlier, the notion of marginally
13 responsive seems to have appeared only in the null set. You
14 know, other than that, it's a -- if we took away the word
15 marginally, then it's a measure of the number of responsive
16 documents they found in a sample from the null set.

17 Q All right. Now, putting aside the reporting issues --
18 well, are there other reporting issues that you have?

19 A Well, they omitted the technique that they used to compute
20 the confidence interval. They also omitted the raw data that
21 the confidence interval was computed from, so there's no way to
22 check their calculations.

23 Q So that implies a lack of transparency, is that correct?

24 A That would be a fair description.

25 Q And in your understanding of information retrieval as it's

1 applied in eDiscovery, there is some element of transparency
2 that is required to comply with best practice, is that correct?

3 A That would be my understanding, yes.

4 Q Are there other issues with respect to the fact that it
5 cannot produce a statistically valid sample of effectiveness?

6 A Well, yes. Yes. I mean, all of the upstream problems
7 that I mentioned earlier, the fact that the collection had not
8 been identified, the fact that there was an arbitrary selection
9 of five custodians or maybe four at other points, the fact that
10 the spam filtering may have removed responsive documents, the
11 fact that the review of the documents was potentially biased by
12 the fact that reviewers knew what the right answer was. Then
13 there's an additional factor which is particular to the
14 statistical validity of the evaluation.

15 Q And that is?

16 A That's the factor of overfitting. And overfitting in the
17 statistical sense means fitting a statistical model to the same
18 data that you're going to evaluate it on.

19 Q Could you explain that further, please.

20 A Yes. So in this case the Georgia Pacific process involves
21 tuning a search query. And Mr. Brown described in some detail,
22 and it's also described here, the fact that a number of
23 iterations were done to improve the performance of the search
24 query on this set of five custodians. The same -- a sample
25 from the same set of five custodians was then used to evaluate

1 that query. So the queries were customized to the data that
2 was going to be used to evaluate them.

3 It's sort of like if you had a class where the
4 students studied the final exam for the entire semester, and
5 then they were tested on the final exam, and then you felt that
6 if they did well that certified that they understood the field
7 and would be able to understand questions about that in the
8 future. The implication is that even if everything else had
9 been done correctly in this process, you would not be able to
10 extrapolate the effectiveness level produced by this procedure
11 to any other data.

12 Q So with respect to a project of the nature described by
13 Mr. Brown and Mr. Koch earlier, what would you have expected
14 the statistical reporting to look like?

15 A Simply the reporting part I would have expected to -- I
16 would have expected to see a description of how the random
17 sample was drawn and presumably a description that described an
18 appropriate process. I would have seen an explanation of how
19 the random sample was reviewed, and I would have expected to
20 see an unbiased review having been done.

21 And then I would have expected to see explicit counts
22 of the four possibilities, the number of true positives, false
23 positives, false negatives and true negatives, because one
24 could then verify any effectiveness measures that were
25 reported. And then finally the reporting of the estimates of

1 effectiveness should have taken the proper form. The procedure
2 described by which a confidence interval would be produced and
3 the central value, the margin of error, and the confidence
4 level.

5 I would have also expected to see frankly in a
6 computing setting an exact count of the size of the population.
7 There's several expressions here like something was more than a
8 certain amount or less than a certain amount. And if it's all
9 in a computer, you can count it exactly.

10 Q So if the other defendants were to report, as they did in
11 their opening brief, which I think has been marked, that the
12 search term effectiveness confirmed by this testing which
13 yielded margin of errors in the range of 1.4 to 3.9 percent is
14 comparable to the compelling results of the GP testing, your
15 reaction would be?

16 A If they had used the same process?

17 Q Yes, sir.

18 A I would not find those results to convey any useful
19 information about their information retrieval system.

20 Q Now, is there anything else that you can say regarding the
21 defects in the proposal -- the defendants' proposed methodology
22 and statistical reporting? Have we covered the main things?

23 A I think we've covered everything.

24 Q All right. Now, have you had an opportunity to examine
25 the plaintiffs' proposal?

1 A Yes, I have.

2 Q And have you reached any conclusions with respect to the
3 plaintiffs' proposal?

4 A Yes, I have.

5 Q And what are those conclusions, please?

6 A I conclude first that plaintiffs' proposal is likely to
7 find a greater proportion of responsive documents with less
8 human effort. And second, I conclude that plaintiffs' proposal
9 will provide a statistically valid estimate of the recall of
10 the information retrieval process.

11 Q And what's the basis for those conclusions, please?

12 A Well, I have examined the proposal and have examined it in
13 the context of my knowledge of the literature on research and
14 development in information retrieval and machine learning. I
15 have consulted textbook references on statistical sampling to
16 verify the techniques that are used for computing estimates.

17 MR. MOGIN: I'd like to mark please, Your Honor, as
18 Plaintiffs' 10 what has previously been tendered in connection
19 with Dr. Lewis' report, which is the proposed CBAA search
20 process for each defendant.

21 BY MR. MOGIN:

22 Q Dr. Lewis, is this the proposed process that you have put
23 together?

24 A Yes, I aided the plaintiffs in the development of this,
25 and in particular I'm responsible for the technical aspects of

1 this proposal.

2 Q So after reviewing plaintiffs' proposal you made certain
3 modifications in the explanations, is that correct?

4 A Yes, I did.

5 Q And the result is Plaintiffs' Exhibit 10, correct?

6 A This document, yes.

7 Q All right. Could you explain your proposal, please. Step
8 1, collection.

9 A Really all that's assumed here is that all the reasonable
10 sources of responsive documents are collected and made
11 available to the system.

12 Q And I guess we'll defer for another day what those
13 reasonable sources may be. Step 2, test set creation, could
14 you explain that, please.

15 A Yes. There are two proposals here for how to create the
16 test set; that is, the set of data that will be used to
17 evaluate the system. And those proposals are called the
18 indirect method and the direct method.

19 Q Could you first explain what the test set is and how it
20 operates.

21 A Yes. The test set is a simple random sample of the data,
22 and it's used to produce statistically valid estimates of
23 recall of the information retrieval system.

24 Q So that would be a random -- simple random sample drawn
25 according to recognized statistical technique?

1 A That's correct.

2 Q Okay. And now that we have established what the test set
3 is, how would we go about creating it?

4 A Well, there's two methods proposed here. The first is the
5 indirect method, and the indirect method is somewhat similar to
6 the ideas -- basically it's based on the idea of doing a sample
7 to find out what the number of responsive documents -- well,
8 the proportion and thus the number of responsive documents is
9 in the entire universe.

10 You then can compare the number, the estimated number
11 of responsive documents in the universe with the number of
12 responsive documents that have been found in production. Now,
13 as I mentioned earlier in commenting on Mr. Robbins' testimony,
14 it's very important in that case that the review of the test
15 set be done in a fashion that's consistent with the review for
16 production.

17 Q Now, having heard the testimony of both Mr. Koch and Mr.
18 Brown, does the Georgia Pacific methodology do that? Do you
19 recall it was Mr. Brown and two people who did the --

20 A Yes. Well, I mean it assumed -- the Georgia Pacific
21 methodology is implicitly assuming that there's consistent
22 review of their samples and review during production.
23 Obviously there's many other characteristics that are different
24 in the Georgia Pacific process than in this process.

25 Q Okay. Now, is consistency in any way a function of the

1 number of reviewers?

2 A That's a difficult question. It's obviously -- it is more
3 difficult to achieve consistency the greater number of
4 reviewers, but there's also issues of the training and skills
5 of the people. So it's hard to make a definitive statement
6 about that.

7 Q But you can say that the greater the number of reviewers,
8 the greater the likelihood of variability?

9 A That's fair to say.

10 Q So, for example, if there were three reviewers
11 consistently used well trained as compared to 13 contract
12 attorneys, we'd be more likely to get better results using the
13 direct method, is that right?

14 A I wouldn't want to make any statement about the skills of
15 contract attorneys. But just the larger number of people makes
16 it more difficult.

17 Q Fair enough. Okay. Is there anything else that you need
18 to tell us about the indirect or the countdown method?

19 A Only, only to reiterate the danger that's involved if the
20 reviews are not consistent. Because what can happen in that
21 setting is that you could think that you had much more higher
22 recall than you actually did.

23 Q And in the TREC studies what have you found about
24 people's -- attorneys' intuition about their own effectiveness?

25 A Well, that hasn't really been looked at in the TREC

1 studies, but there's a classic paper that actually was
2 mentioned earlier today by Blair and Moran, which was looking
3 at the searches that were done in the context of a discovery
4 setting. And the result was that --

5 MR. McKEOWN: Your Honor, I'm going to object. I
6 think this goes beyond his personal knowledge if he's going to
7 testify to this report, as to the results of this particular
8 study with respect to comparison.

9 THE COURT: Well, do you adopt -- I mean, let's find
10 out if that's his opinion or if he agrees with this study, and
11 then you can question him about it.

12 BY MR. MOGIN:

13 Q Do you agree with the Blair and Moran study?

14 A I will simply report what the Blair and Moran study found.
15 It's one of the most famous studies in information retrieval.
16 It found that the attorneys in the case believed that they had
17 found on the order of 75 percent of responsive documents. And
18 a proper statistical evaluation found that they had found less
19 than 25 percent.

20 Q Let's move on then to the direct method.

21 A Okay.

22 Q Would you describe that, please.

23 A Sure. The direct method is involved -- basically works by
24 taking random samples from the -- again, from the entire
25 universe of documents and reviewing them until a particular

1 number of responsive documents has been found. Based on the
2 desired confidence level and margin of error in this proposal
3 that number of responsive documents is 385.

4 So those samples as they're found are reviewed, and
5 the intention being that they be reviewed in a careful fashion
6 such that there's good agreement between the assessors. And
7 then that set of 385 responsive documents is what's used to
8 evaluate the recall of the information retrieval process.

9 Q And how did you calculate that 385 documents would be
10 necessary?

11 A So this was based on using a margin of error of 5 percent
12 on recall and a confidence level of 95 percent. And then
13 choosing a sample size, given that we don't actually know -- we
14 don't know how many responsive documents there are in the
15 collection, I used a conservative calculation based on a
16 binomial proportion.

17 Q Can you explain that a little further in layman's terms.

18 A Sure. So the question is the size of the sample you need
19 is based on sort of how wide you want this confidence interval
20 to be and how confident you want to be that your sample is
21 representative. The exact sample size is based -- in theory it
22 would depend on -- let's see. In a funny sense if you knew
23 what the proportion of responsive documents was and you knew
24 that it was fairly extreme, you might be able to draw a smaller
25 sample.

1 Since we don't know what the proportion of responsive
2 documents is, we make the conservative assumption that half of
3 the documents are responsive. That gives you then a sample
4 size which will give this margin of error and confidence level
5 regardless of the actual proportion of responsive documents.

6 Q Now, this is according to standard statistical technique
7 as applied in information retrieval, correct?

8 A Oh, yes. And it's applied in many other fields of
9 science, engineering, finance, and other areas.

10 Q All right. So now once we have this total of 385, what
11 happens next?

12 A Okay. So, so again the -- and the proposal is a little
13 bit unclear here. If the indirect method is used, we have step
14 2-A and then step 3. In the direct method -- so basically you
15 draw a sample once and you review it. In the direct you're
16 drawing several samples, reviewing each of them. So step 3
17 actually happens several times.

18 The important thing to mention about the review is
19 that the review needs to be unbiased. That is the personnel
20 who are reviewing the test set should not know what the
21 information retrieval process has been finding. And indeed
22 these three processes, the review of the test set and the
23 beginning of actually looking for responsive documents could be
24 done in parallel.

25 Q Did you hear Mr. Brown's testimony in that regard?

1 A Yes, I did.

2 Q And what was your conclusion?

3 A That the reviews that were done by Georgia Pacific were
4 potentially biased because the reviewers knew what the right
5 answer was supposed to be.

6 Q Another example of teaching to the test?

7 A Well, this is -- it's a little different. It's sort of
8 more, you know, just human review is affected by many factors,
9 including, you know, potentially self-interest.

10 Q All right. And then what happens in your protocol here?

11 A Okay. So in step 4 then we're now on to actually
12 searching for responsive documents. And the protocol is based
13 centrally around the use of supervised learning. To begin to
14 use the supervised learning set one wants an initial training
15 set. What's sometimes referred to as a seed set, though I
16 heard that term used in a different fashion earlier.

17 The important thing about the initial training set --
18 and I should acknowledge that many different vendors provide
19 supervised learning. Each of them will have their own best
20 practices for how to initialize machine learning. This is a
21 description of a procedure that, you know, is fairly generally
22 applicable, should produce good results, but might be modified
23 according to vendors' understanding of their particular
24 systems.

25 This particular procedure draws on four sources of

1 documents to produce the initial training set. The first is a
2 sample of responsive documents already found by the defendants.
3 The second is a sample of responsive documents available to
4 plaintiffs either that they've somehow obtained or that they
5 have created as simulated responsive documents to -- as a
6 representative of documents that might potentially exist. The
7 third is --

8 Q Just let me stop you there. So, for example, the
9 plaintiffs could create a document that said they are with us?

10 A Indeed. And they could also create, you know, based on
11 their best understanding what metadata might look like if such
12 a document was responsive. The third source is leveraging the
13 work that the plaintiffs -- excuse me, the defendants have done
14 in producing their Boolean queries. While we don't know how
15 effective those Boolean queries are, they were produced with
16 some intent to try to find some responsive documents, and so
17 they can be used to retrieve sets of documents from the entire
18 universe of documents, and some sample of those could be
19 reviewed and included in the training set as another source.

20 Q So if this were applied then, Georgia Pacific or any other
21 defendant would not have to go back to square one and begin
22 writing on a blank slate, is that correct?

23 A Oh, no. They would have gotten some considerable value
24 out of the Boolean queries that they have. And then finally
25 the fourth source is to simply take the words from defendants'

1 Boolean exercise and use them as queries to a statistical
2 ranked retrieval system and review a sample of the top ranked
3 documents. And this would take advantage of the statistical
4 ranked retrieval system's ability to not require exact match.
5 Again, one would take advantage of the work that's been done in
6 developing the search terms, but use them in the more powerful
7 context of a statistical ranked retrieval system.

8 Q All right. That Chicago accent got me again.

9 A Oh, I'm sorry.

10 Q You said match?

11 A So you would not need to do an exact -- the statistical
12 ranked retrieval system does not require an exact match. It
13 can use the words that were in the Boolean query but rank the
14 entire document collection by sort of degree of strength with
15 respect to those words.

16 Q All right. Is there anything else that you would do in
17 order to utilize statistical ranked retrieval in this protocol?

18 A Well, I think that would be the main way to use it. But,
19 of course, I should be clear that supervised learning systems
20 themselves produce a ranking of the collection. They -- almost
21 all supervised learning systems, certainly the ones I'm aware
22 of in eDiscovery, produce models. They learn terms and learn
23 term weights, and then can use those term weights to rank
24 documents just like you would rank from a natural language
25 query in a statistical ranked retrieval.

1 Q So in supervised learning you get a little bit of both?

2 A You do.

3 Q All right. Now, step -- we'll do the next step.

4 A Okay. So the next step is actually training the system,
5 and this simply means executing the supervised learning
6 capability of the software. The point to be stressed here is
7 that the maximal amount of information should be available to
8 the supervised learning. In particular the supervised learning
9 should have access to both the content and the metadata of the
10 documents.

11 Q And that would be the original metadata of the documents?

12 A Well, whatever the most informative metadata is.

13 Q Okay. And then what happens after we've trained the
14 system?

15 A So then after you train the system, it can then be used to
16 find additional documents. And those documents can be
17 reviewed. Responsive documents, you know, become part of the
18 production -- obviously responsive and nonprivileged documents
19 become part of the production set. Some of the new responsive
20 and nonresponsive documents that are found can also be added to
21 the training set.

22 Q All right. And then once you've done that what happens?

23 A Well, so now we've got some documents in the production
24 set, and we can estimate the recall. We can compute a 95
25 percent confidence interval on the recall of the production

1 set.

2 Q And this is estimation in a statistical sense, correct?

3 A Yes. And again, one thing that's important to stress is
4 just as the people reviewing the sample should not know what
5 the decisions of the system are on those documents, it's
6 important that the people running the system not know what the
7 behavior of the system on the sample is. And indeed it's
8 preferable if they're not even aware of the current estimate of
9 recall on the test set.

10 They may well want to have other samples of data that
11 they use to tune their system. But if they tune their system
12 to the test data, then they'll invalidate the statistical
13 validity of the results, and this is the overfitting process I
14 mentioned.

15 Q Now, with respect to the 95 percent confidence interval,
16 if the system is able to achieve a 95 percent confidence
17 interval within the specified margin of error, does that mean
18 that the system has discovered 95 percent of all responsive
19 documents within the corpus within that margin of error?

20 A No. No. The 95 percent confidence is a measure of the
21 representativeness of the sample that's drawn. The process
22 produces a confidence interval on recall. So, for instance,
23 the result of step 7 might be there's a 95 percent confidence
24 that the recall of the system is 13 percent plus or minus
25 5 percent.

1 Or you might have a 95 percent confidence that the
2 recall of the system is 78 percent plus or minus 5 percent. So
3 it's the estimate of recall which is what's telling you how
4 much of the responsive documents you've found.

5 Q So can you put that as close to in lay terms or give us a
6 very simple example of how a confidence interval is properly
7 used in this context.

8 A Well, you would -- so I mean we could go on to step 8, and
9 what would happen in step 8 is you would, you would look at
10 this confidence interval. The confidence interval would say
11 something like, you know, the system has found -- we think
12 the -- you know, the estimate is the system has found, or the
13 whole production process has found 13 percent of the responsive
14 documents, and we have some uncertainty. It's plus or minus 5
15 percent. So we're between 8 percent and 18 percent of the
16 responsive documents at this point.

17 And then there would be a cost benefit analysis. Is
18 13 percent enough? How much it would cost to find some more.
19 And, you know, this can be done many times in an iterative
20 process. The process as described here says that if review is
21 not terminated, you return to step 6. Actually it would be, it
22 would be more clear to say you could go back to any of steps 4,
23 5, or 6. If things were going really badly, it might be the
24 case that the whole initialization of the seed set would need
25 to be done again. Otherwise you might go back to step 5 maybe

1 to retrain the system some more, or maybe you just go back to
2 step 6. You think the system is already working pretty well
3 and you just sort of use it in its current state to find more
4 of the responsive documents.

5 Q Now, Dr. Lewis, the defendants have contended that this
6 process that the plaintiffs have proposed is unprecedented.
7 It's new, it's untested. What's your reaction to that?

8 A Well, people have been using supervised learning in
9 information retrieval since the early 1960s, so it's not
10 unprecedented in the field of information retrieval. We have
11 been studying supervised learning on eDiscovery, simulated
12 eDiscovery tasks in TREC since 2007 I think. There are a
13 number of vendors who provide supervised learning capabilities
14 in their software or review services. I don't -- maybe there's
15 a legal definition of unprecedented I'm not familiar with.

16 Q Well, do you know how long these current vendors have been
17 offering supervised learning as an eDiscovery tool?

18 A I haven't kept close track of that. I know that Orcatech
19 has been offering it since 2010. I know that the Kroll system,
20 which I helped design some of the algorithms for, has been
21 offered since 2010. I haven't really kept track of the others
22 in the industry.

23 Q Can supervised -- strike that.

24 Can supervised learning be used on different review
25 platforms?

1 A Yes, there's several different review platforms that use
2 it.

3 Q Is supervised learning used outside of the eDiscovery
4 area?

5 A Oh, yes. I mean, it's becoming ubiquitous in almost any
6 application that involves text data. To give an example, you
7 know, when you see advertisements on web pages, many of those
8 advertisements are placed by supervised learning systems. And
9 the reason is that benefits in accuracy of predicting whether
10 you'll click on an advertisement lead to millions of dollars in
11 improvements in revenue for large companies like Google and
12 Yahoo and whatnot, and they inevitably use supervised learning
13 systems. Some of them on a quite immense scale. Some of these
14 systems are trained literally on billions of training examples.

15 Q Are you familiar with Amazon.com?

16 A Yes. So Amazon would be another one. The recommenda-
17 tions, you know, when you go there and it recommends you might
18 be interested in buying these books, it has used supervised
19 learning to learn from large numbers of purchase decisions what
20 you might be apt to buy.

21 Q And do you believe that your proposal or the plaintiffs'
22 proposal is superior to defendants'?

23 A Yes, it is superior in that first it is likely to produce
24 a higher level of responsive documents with less manual review.
25 And second, it provides a statistically valid estimate of its

1 recall, where defendants' process does not provide a
2 statistically valid estimate.

3 MR. MOGIN: One moment please, Your Honor.

4 THE COURT: Sure.

5 (Brief pause.)

6 MR. MOGIN: Your Honor, there is only the issue of
7 Dr. Lewis' report. For the life of me I don't know what we're
8 doing with reports. I would move also to qualify Dr. Lewis as
9 an expert in the field of information retrieval as it's applied
10 in eDiscovery.

11 THE COURT: Okay. And, Mr. McKeown.

12 MR. McKEOWN: Your Honor, with respect to the area of
13 information retrieval, I think that's one area. I think the
14 application of information retrieval in eDiscovery, we would
15 ask you to reserve ruling on because this is a fairly novel
16 area. I'm not sure there is an expert in that area.

17 THE COURT: How about his resume coming in?

18 MR. MOGIN: Very good. And his report, Your Honor?

19 THE COURT: What about his report? Do you have a --

20 MR. McKEOWN: Reports typically don't come in.

21 That's usually the testimony of the experts.

22 THE COURT: Right.

23 MR. McKEOWN: And his exhibits.

24 THE COURT: Well, his resume's coming in. You're not
25 objecting to his qualification on search retrieval, correct?

1 MR. McKEOWN: On information retrieval, Your Honor.

2 THE COURT: Information retrieval.

3 MR. McKEOWN: Information retrieval.

4 THE COURT: Okay. Is that good enough for you?

5 MR. MOGIN: Yes, it is. Thank you, Your Honor.

6 THE COURT: Thank you. Are you going to cross, Mr.
7 McKeown?

8 MR. McKEOWN: Yes, Your Honor, I am. In addition,
9 with respect to the portions that are specific to Georgia
10 Pacific, Mr. Neuwirth will handle that portion of the
11 cross-examination.

12 THE COURT: Okay. That's fine.

13 CROSS-EXAMINATION

14 BY MR. McKEOWN:

15 Q Good afternoon, Dr. Lewis.

16 A Good afternoon.

17 Q My name is Jim McKeown. I represent one of the
18 defendants, and I have some questions for you about your
19 testimony.

20 Do you have in front of you Exhibit 10, the protocol
21 you were just discussing?

22 A Yes, I do.

23 Q Did I understand your testimony to be you modified part of
24 this, but you did not write it originally, is that correct?

25 A I did all of the technical design of this protocol based

1 on the descriptions -- my conversations with plaintiffs'
2 attorneys on their intended use of CBAA technology.

3 Q And when were you first retained in this matter?

4 A I believe it was January 18th.

5 Q January 18th of 2012?

6 A That's correct.

7 Q And have you spoken to any of the plaintiffs as opposed to
8 the plaintiffs' lawyers?

9 A No, I have not.

10 Q And when did you prepare the protocol we see marked as
11 Plaintiffs' Exhibit 10, or your edits to it?

12 A The, I believe the final version of this went in last
13 Thursday.

14 Q And when did you first see it?

15 A Well, I first started working on it -- well, it was, oh,
16 within a few days of having been retained I started working on
17 it. You know, basically I looked at the descriptions of what
18 the plaintiffs have been asking for. Discussed with them what
19 they meant by CBAA and, you know, started work on ideas for
20 this.

21 Q And if we go back to February 6th, you may recall that was
22 the date that the parties made their submissions to court on
23 their respective positions. Prior to February 6th, how much
24 work, how many hours had you spent on this matter?

25 A Prior to February 6th. I would have to go back and look

1 on my time records, but, you know, it would have been -- you
2 know, it would have been something -- it would have been at
3 least 10 hours. I'd really have to go back and look at my
4 records to know, you know, what happened before or after
5 February 6th.

6 Q Prior to January 18th when you were retained in this
7 matter, had you ever heard the term content based advanced
8 analytics?

9 A I'm not sure. I've heard many terms like content based
10 analytics. I'm not sure if I've heard it with the advanced in
11 it. You sometimes hear text analytics. You sometimes hear
12 content analytics. So there's a number of terms like this in
13 the industry. I'm not sure if I've heard that exact term
14 before.

15 Q And have you done anything to assist the plaintiffs in
16 their response to document requests with respect to the
17 protocol they're going to use?

18 A Could you repeat the question.

19 Q Sure. You have here Exhibit 10, which is plaintiffs'
20 proposal for search process for each defendant. You also
21 understand, don't you, that the defendants served document
22 requests on the plaintiffs?

23 A Oh, yes. I see what you mean.

24 Q Did you prepare a protocol for the plaintiffs to comply
25 with their document responses?

1 A No, I have not.

2 Q Have you done anything to assist the plaintiffs with their
3 document responses?

4 A No, I have not.

5 Q And you said that the Kroll system, that you helped write
6 one of the algorithms for, a couple of the algorithms for was
7 launched in 2010?

8 A That's correct.

9 Q Are you aware of that system having ever been approved by
10 a court for use in terms of finding responsive documents?

11 MR. MOGIN: Objection. I'm not sure the courts
12 approve this.

13 THE COURT: Well, if he knows -- I mean, if he knows.
14 I mean, if he knows. I don't know. I mean, I don't know
15 whether you know. Do you know, Mr. McKeown? Do you know the
16 answer to that question?

17 MR. McKEOWN: I do not believe it has ever been
18 approved, Your Honor.

19 THE COURT: Okay. Well, ask him if he knows. Okay.

20 MR. MOGIN: Shouldn't counsel have a good faith basis
21 for the question?

22 THE COURT: Sure.

23 MR. MOGIN: Has a court ever approved any particular
24 platform?

25 MR. McKEOWN: I'm talking about the --

1 THE COURT: The Kroll system --

2 MR. McKEOWN: The new system that has his algorithms.

3 THE COURT: -- that actually Dr. Lewis has worked on,
4 yes.

5 THE WITNESS: I've been informed by Kroll that the
6 system has been used in a number of eDiscovery matters. I'm
7 not aware of any of the, the legal issues around that, if there
8 was a court approval. I'm simply not aware of those issues.

9 BY MR. McKEOWN:

10 Q This concept of -- and I know you don't want to use the
11 word predictive coding because of the trademark. And you
12 prefer learning, right?

13 A Supervised learning.

14 Q Supervised learning. This concept of supervised learning
15 is a fairly new development, correct?

16 A No.

17 Q In the legal context, would you agree with that?

18 A If by new you mean within the past 10 years, sure.

19 Q Do you think the use of supervised learning as a means of
20 identifying documents for purposes of producing them in
21 response to document requests is a fairly new development?

22 A It is a relatively new development in the industry to my
23 understanding.

24 Q And you talked about TREC, correct?

25 A Yes, I did.

1 Q And that was Text REtrieval Conference, T-R-E-C, is that
2 right?

3 A That's correct.

4 Q And I think you said you were a founding member?

5 A I was one of the initial PC members and I was one of the
6 cofounders of the TREC Legal Track.

7 Q And the -- I believe you said that the TREC Legal Track
8 was started in 2007, is that right?

9 A I believe we had the initial discussions in 2005, and I
10 believe the first year that there were results for was 2006.

11 Q And one of the things that you do at TREC is you have this
12 interactive task, is that correct?

13 A Yes, there is an interactive task at TREC.

14 Q And as part of this interactive task, various teams look
15 at this collection of documents you have for purposes of
16 testing various approaches to finding documents, correct?

17 A Yes.

18 Q And 2008 was the first year that there was this
19 interactive task exercise, correct?

20 A I'd have to go back and look to check that. That sounds
21 reasonable.

22 Q Well, do you recall that there were only four teams that
23 first year?

24 A Again, I would have to go back and check the document. If
25 you have the document, I'd be happy to look at it.

1 Q I do. Let me just show you what we're going to mark for
2 identification purposes as Exhibit No. -- Defendants' Exhibit
3 No. 5. And I just would like to ask you to turn to page 26 for
4 purposes of refreshing your recollection.

5 A Okay. So this document is a *Law Journal* article by Maura
6 Grossman and Gordon Cormack called "Technology Assisted Review
7 in eDiscovery Can Be More Effective and More Efficient Than
8 Exhaustive Manual Review."

9 Q Right. And my question was if you could look to page 26
10 to see if that refreshes your recollection that 2008 was the
11 first year of the interactive task.

12 A Yes, it is.

13 Q And, in fact, in that first year of the four groups that
14 went, only one had a recall over 20 percent, is that correct?

15 A Are you referring to a particular mention of that in this
16 document?

17 Q That's actually what's going to be marked as Defendants'
18 Exhibit 6. Defendants' Exhibit 6 is an article that you helped
19 write, correct?

20 A That's correct.

21 Q And if we turn to page 24.

22 A Okay. I'm on page 24.

23 Q You report the results of the 2008 study, correct?

24 A Uh-huh. Yes, this is the reports of the 2008 interactive
25 task.

1 Q And if you look at the very top of the page on 24, the
2 last sentence there says, "One team, notably the one that made
3 the most use of TA time, obtained a relatively high recall,
4 .62. While the other three all making significantly less use
5 of TA time obtained recall values below 0.20," is that correct?

6 A I see that, yes.

7 Q And when we talk about recall, you said that recall had to
8 do with respect to how many of the total population of
9 responsive documents were found through this system, correct?

10 A That's correct.

11 Q So that when you ran the interactive task in 2008 and four
12 teams ran it, the best team did over 60 and the other three did
13 less than 20 percent of recall, is that correct?

14 A Well, you said when you ran. I was --

15 Q I'm sorry.

16 A I was not involved in running the Legal Track that year.

17 Q My apologies. When TREC had its interactive task for 2008
18 and four teams entered, three of the four terms had recall of
19 less than 20 percent, is that correct?

20 A That's correct.

21 Q Now, in 2009 were you involved in that interactive task?

22 A No, I was not.

23 Q Do you recall that in 2009 that the database that was used
24 was a collection of Enron e-mails?

25 A Yes.

1 Q And presumably this Enron e-mail production was an e-mail
2 production to FER (phonetic), correct?

3 A That's my understanding.

4 Q And presumably that had been collected somehow. Do you
5 know if search terms were used to collect that grouping of
6 e-mails that were used for the Enron -- or excuse me, the TREC
7 interactive task in 2009?

8 A I'm not familiar with the collection processes for the
9 Enron data.

10 Q And again, there were a number of teams, and they would
11 look at what recall they could achieve with respect to pulling
12 documents for particular requests that were crafted for
13 purposes of that exercise, correct?

14 A My understanding is that the teams were attempting to
15 optimize I believe it was the F measure. TREC, in no case are
16 the teams focusing solely on recall.

17 Q And if you look again at what we have marked as Exhibit 6,
18 your article.

19 A Yes.

20 Q And you turn to page 4.

21 THE COURT: Which article?

22 MR. McKEOWN: I'm sorry, Your Honor. I need to
23 give --

24 THE COURT: Defendants' 6?

25 MR. McKEOWN: Yes.

1 THE COURT: Okay.

2 MR. McKEOWN: I need to give him a different exhibit,
3 though.

4 THE COURT: Oh, okay.

5 BY MR. McKEOWN:

6 Q You have Exhibit 7 in front of you?

7 A Yes. It's a document entitled "Overview of the TREC 2009
8 Legal Track."

9 Q And is this a document you have seen before?

10 A Yes, I have seen this before.

11 Q By the way, I meant to ask you before, is H5 supervised
12 learning or predictive coding?

13 A It's actually unclear what technologies H5 uses
14 internally.

15 Q If you look at Exhibit 7.

16 A That's the TREC 2009 Legal Track.

17 Q The TREC 2009 Legal Track.

18 A Yes.

19 Q And we turn to page 4. Although I see they don't have
20 page numbers on this version.

21 A Yes, this does not seem to have page numbers.

22 Q If you look at Section 2.21.

23 A Okay.

24 Q Which is on the fourth page. And you look down at the
25 second to last paragraph. It says, "The steps we took to

1 process the collection are as follows." Do you see that?

2 A I do see that.

3 Q It appears that Clearwell was the process that was used,
4 is that correct, for the TREC?

5 A That's what's stated there, yes.

6 Q And so the interactive task group used Clearwell for
7 purposes of their exercise in 2009, correct?

8 A That does appear to be correct.

9 Q And the images that were used were TIFF images, is that
10 correct?

11 A I'm not seeing that here. Could you show me where you're
12 referring to.

13 Q That was just a question. It's not on the page. Are you
14 familiar with the fact that the images were TIFF images?

15 A I'm not aware whether TIFF images were used during 2009 or
16 not.

17 Q Do you recall that there were 24 runs undertaken in 2009?

18 A In which task are you specifying that?

19 Q If we could turn to page 20 -- page -- you don't recall
20 which tasks were undertaken or how many in 2009?

21 A There were several tasks undertaken. I don't remember the
22 exact number of participants. I'd be happy to look at the
23 section that you're referring to.

24 Q Do you recall that one of the tasks was related to fantasy
25 football?

1 A I believe there was a -- one of the topics used in one of
2 the tasks. I don't actually remember which of the tasks that
3 topic was used in.

4 Q And sitting here today you don't know that there were 24
5 tasks run at that time?

6 A I'm certain there were not 24 tasks run. The tasks are
7 the high level groupings, such as interactive, batch, or
8 supervised learning, and whatnot. There would never be more
9 than three of those in a year. I'm not sure what you're
10 referring to then.

11 Q Perhaps I'm using the wrong term. Does the term run mean
12 something to you?

13 A It's certainly possible there were 24 runs. If you could
14 point me to the particular task that you want to ask about, I'd
15 be happy to look at that.

16 Q If you turn to -- and again, unfortunately there are no
17 page numbers in this study. But if you turn in Exhibit 7 to
18 section --

19 MR. McKEOWN: I apologize, Your Honor. Could I just
20 have a moment.

21 THE COURT: Sure. Sure.

22 (Brief pause.)

23 BY MR. McKEOWN:

24 Q If you could look right above Section 2.3.4 -- excuse me.
25 Right before Section 2.3.5.

1 A Okay. I see Section 2.3.5.

2 Q Okay. And if you look at the paragraph above that.

3 A Where it says, "We return to these data below?"

4 Q I'm sorry. The first full paragraph there. That there
5 were 24 runs.

6 A Okay. The first full paragraph. Yes, there were -- it's
7 discussing that there were 24 runs. I have to look back and
8 see which task this is for.

9 Q And you could find the runs assigned by topics in Section
10 2.2.3.

11 A Okay. So this is under the interactive tasks. And sorry,
12 what was the section you just mentioned?

13 Q 2.2.3.

14 A All right. All right. So 2.2.3 is a list of the research
15 groups that took part.

16 Q And there's a list of the runs as well, correct?

17 A Oh, yes. There's the -- the table on the next page then
18 shows which topics each team submitted results for.

19 Q And you see there were topics 2001 through -- or 201
20 through 2000 -- or 207, correct?

21 A Yes.

22 Q And if we turn back to Section 2.2.2, it gives the topics
23 and topic authorities, is that correct?

24 MR. MOGIN: I'm sorry. Could I have that question
25 repeated, please.

1 MR. McKEOWN: Sure.

2 BY MR. McKEOWN:

3 Q If we turn back to Section 2.2.2. I apologize. There are
4 no page numbers on this document.

5 A Yes, I'm at Section 2.2.2.

6 Q There is the list of topics, correct?

7 A Yes.

8 Q And if you look at the topics, these are topics that were
9 created by the interactive task team as potential document
10 requests for purposes of this study, correct?

11 A Yes, that's my understanding.

12 Q And there are just seven topics in this exercise, correct?

13 A That's correct.

14 Q And they're applied to this Enron group of e-mails that
15 were collected for this task, is that correct?

16 A Yes.

17 Q And isn't it true that out of the 24 runs that were done
18 only 5 of the runs had a recall over 70 percent?

19 A Are you referring to a particular table or a description
20 here?

21 Q I was referring to your article, which I think is Exhibit
22 6.

23 A At which page?

24 Q At page 24.

25 A All right. And where on this page are you referring?

1 MR. McKEOWN: I apologize, Your Honor. Let me move
2 on. I'll come back to that.

3 THE COURT: Okay.

4 BY MR. McKEOWN:

5 Q Isn't it true that the total number of documents amassed
6 in all four years of the TREC legal study are less than the
7 number routinely assessed in even a real world eDiscovery
8 problem?

9 A Well, my understanding is that real world eDiscovery
10 problems vary to a considerable degree. However, large scale
11 problems routinely involve review of very large numbers of
12 documents. And I do believe that there are large scale
13 eDiscovery problems that would have reviewed more documents
14 than were reviewed in the TREC evaluations.

15 Q If you look at page -- again, on Exhibit 6, page 28. And
16 Exhibit 6 is your article, correct?

17 A Absolutely.

18 Q And I direct your attention to Section 5.5 on page 28.
19 The second paragraph. And again we're talking here about the
20 2009 project, correct?

21 A No. This is talking about the history of the TREC Legal
22 Track from the beginning.

23 Q Okay.

24 A Up to, I believe up to 2009.

25 Q And you wrote, "The Legal Track was humbling also for the

1 insight it has provided the IR research operational eDiscovery
2 settings. While the scope of the TREC relevance assessment
3 process was large by TREC standards in geography, technology,
4 personnel, and data set preparation, the total number of
5 documents assessed across all four years of Legal Track has
6 been far smaller than the number routinely assessed in even a
7 single real world discovery project."

8 Isn't that what you wrote?

9 A Yes, it is.

10 Q And if you'd turn back to page 24 at the very bottom of
11 the page. It says, "The post-adjudication results for the 2009
12 topics showed some encouraging signs. Of the 24 submitted runs
13 aggregating across all 7 topics, 6 obtained an F1 score of 0.7
14 or greater," is that correct?

15 A That's correct.

16 Q So that only 6 of the 24 were 0.7 or greater, correct?

17 A That's on the F score, yes.

18 Q And is there a recall score there as well?

19 A The next sentence says, "In terms of recall of the 24
20 submitted runs, 5 distributed across 4 topics attained a recall
21 score of 0.7 or greater." And then, "Of these five runs, four
22 distributed across three topics simultaneously obtained a
23 precision score of 0.7 or greater."

24 So for the systems that were attempting to do a
25 single Boolean classification optimizing the F1 score, that's

1 the results that were achieved.

2 Q And for recall there are only five above .7 or greater,
3 correct?

4 A That's correct. Of course, the systems were not trying to
5 optimize recall.

6 Q Okay. And of the five, two of the five were topic 207, is
7 that correct?

8 A I would have to go back and look. Do you have the
9 point --

10 Q Well, let me point you first to what 207 is, which I think
11 you'll find on page 24.

12 A Yes, I'm already open to that page.

13 Q And topic 207 called for all documents or communications
14 that describe, discuss, refer to, report on, or relate to
15 fantasy football, gambling on football, and related activities,
16 including but not limited to football teams, football players,
17 football games, football statistics, and football performance;
18 is that correct?

19 A That's correct.

20 Q And that was the document request that was run against the
21 Enron e-mails. And two of the five that hit over 70 percent
22 recall were the fantasy football as opposed to the other
23 requests, correct?

24 A Are you citing a particular one of the documents for that
25 statement?

1 Q We'll pull that up out of there later. Let me move along.

2 A Well, I will look forward to being pointed to that when
3 you find it.

4 Q Are you familiar with the concept of stemming?

5 A Yes, I am.

6 Q What is stemming?

7 A Stemming refers to a variety of techniques for allowing
8 matches between words that ignore to some extent the endings of
9 the words.

10 Q And you talked earlier about your concern with Boolean
11 searches was in part that you had to have precise word matches,
12 is that correct?

13 A Yes. And as I believe I mentioned in my testimony,
14 Boolean queries often allow a wild card operator that allows
15 matching on words to some extent ignoring their word endings.

16 Q And they allow for more than just word endings, don't
17 they?

18 A Stemming typically is focused on reduction to a stem form
19 that removes what would be called suffixes. So it's
20 essentially stemming is focused on word endings.

21 Q Have you seen the stemming reports in this case?

22 A The stemming reports. I've seen the Boolean query
23 descriptions. I don't recall if I've read the stemming
24 reports.

25 Q Are you familiar with the fact that each of the defendants

1 in addition to having these search terms had a whole number of
2 stems that were also searched as part of the Boolean search?

3 A Well, yes. The Boolean queries include wild card
4 operators. Often indicated by a star, but sometimes by other
5 endings, which allow ignoring word endings during the matching.

6 MR. McKEOWN: Your Honor, my colleagues have asked me
7 to inquire about the Court's plan for this evening. We didn't
8 have any questions that we were asking from 1:20 until probably
9 about 20 minutes ago.

10 THE COURT: I'm not going anywhere.

11 MR. McKEOWN: Okay.

12 THE COURT: Let we find out my court reporter,
13 though.

14 (Off the record discussion.)

15 THE COURT: I think our court reporter could stay
16 till 6. If we had to go later than 6, we could go down to my
17 courtroom and turn the tape on.

18 MR. McKEOWN: Okay.

19 THE COURT: Mr. Neuwirth.

20 MR. NEUWIRTH: Yes, I think there is a possibility
21 that we might need to go past 6. As you heard, there were --
22 in addition to the general issues, there were a number of
23 issues about Georgia Pacific's --

24 THE COURT: Well, Mr. Lewis fortunately is in
25 Chicago, so -- and we're going to have to come back for

1 Mr. Regard.

2 MR. McKEOWN: I'm happy to go for a while longer as
3 well, Your Honor.

4 MR. NEUWIRTH: We're happy to --

5 THE COURT: I'm not going anywhere.

6 MR. NEUWIRTH: We're happy to stay, but we just would
7 want to make sure that this is a process that in addition to
8 Mr. McKeown's cross-examination --

9 THE COURT: That you have enough opportunity.

10 MR. NEUWIRTH: -- that there's an opportunity to go
11 over Georgia Pacific's specific cross-examination.

12 THE COURT: No, right. Well, should we have Mr.
13 McKeown finish up and then we'll do Georgia Pacific the next
14 time? Would that make sense?

15 MR. NEUWIRTH: That's fine with us if it would please
16 the Court.

17 THE COURT: Okay. Why don't you continue on, Mr.
18 McKeown. Because you're -- I mean, you're on a roll right now,
19 and I'm remembering things, so ...

20 BY MR. McKEOWN:

21 Q Dr. Lewis, you have been handed what's been marked Exhibit
22 8. I'd like to direct your attention to the attachment to the
23 cover letter.

24 A Okay.

25 Q And if you look at the attachment to the cover letter, do

1 you see there are three pages that have the International Paper
2 search terms attached?

3 A That's these first three with the string Nos. 01 to 21?

4 THE COURT: Mr. McKeown, just one moment. You know,
5 this is kind of -- I asked the other experts to stay till the
6 end of the hearing. And obviously we're not going to be
7 recalling or asking questions. Mr. Regard, we need to know on
8 date. Does anybody else need any of the experts for any other
9 reason?

10 MR. McKEOWN: Present now, Your Honor?

11 THE COURT: Pardon me?

12 MR. McKEOWN: You mean to keep them present now?

13 THE COURT: Right, because otherwise if they're
14 catching planes and stuff other than Mr. Regard who we've got
15 to figure out date wise, I don't know if anybody else is here.
16 We had our linguistics lady, Mr. Brown, and Mr. Hanners. So
17 can we relieve them? Going, going --

18 MR. NEUWIRTH: Yes.

19 MR. FREED: Your Honor, also I apologize, but I have
20 a plane I have to catch, so I may have to leave before they
21 complete the examination. But I have plenty of adequate
22 counsel here.

23 THE COURT: Well, maybe it will take us a half hour
24 to figure out a date. I mean, maybe that's what we really need
25 the half hour for. What time's your plane, Mr. Freed?

1 MR. FREED: 7:10. So it's getting a little ...

2 THE COURT: Yes. Oh, darn it. How much more do you
3 have would you say?

4 MR. McKEOWN: I would probably go till at least a
5 quarter to 6.

6 THE COURT: Okay.

7 MR. FREED: I'm not suggesting that he shouldn't
8 proceed. I'm just apologizing for having to leave early.

9 THE COURT: Well, I think the whole question is,
10 though, on picking this next date. It was hard enough to get
11 us all together at one time, and I want you to be here, so ...
12 All right. 10 minutes more and then we're stopping. Okay.
13 Wherever you are in 10 minutes.

14 MR. McKEOWN: Understood, Your Honor.

15 THE COURT: Thank you.

16 MR. MOGIN: Your Honor, if I may interrupt for just a
17 moment.

18 THE COURT: Sure.

19 MR. MOGIN: I am concerned about this exhibit that's
20 been marked as Defendant 8. There seems to be lengthy
21 attachments to it. There's no indication that I've been able
22 to see in the few minutes that I've been able to examine the
23 letter. The attachments were in the original. It doesn't
24 indicate any such thing after the signature line, and I can't
25 see anything in the content of the letter that indicates so.

1 Perhaps I'm missing something. But if the attachment wasn't
2 part of the original and we don't have any indication of
3 that --

4 MR. McKEOWN: Well, why don't we just take --

5 MR. MOGIN: -- I don't think this is a proper
6 exhibit.

7 MR. McKEOWN: We could take the cover letter off. My
8 questions are all about the exhibit in any event.

9 MR. MOGIN: Well, as I said, I don't know that we've
10 seen this exhibit, Your Honor. Certainly not in its totality.

11 MR. McKEOWN: Well, if you look at the second page of
12 the letter, the second paragraph from the bottom, the middle of
13 the paragraph, it says the search terms, search terms and
14 associated stemming report are attached hereto as Exhibit 1.

15 MR. MOGIN: May I inquire of counsel if these are the
16 same search terms that were attached to the defendants' opening
17 brief.

18 MR. McKEOWN: The search terms, yes.

19 MR. MOGIN: But not the stemming report?

20 MR. McKEOWN: The stemming was not in there.

21 MR. MOGIN: Very good.

22 THE COURT: Do you still have an objection now that
23 you know where they're from?

24 MR. MOGIN: No, I don't, Your Honor.

25 THE COURT: Okay. Fine. Then continue on, Mr.

1 McKeown.

2 BY MR. McKEOWN:

3 Q Dr. Lewis, looking at the page that goes landscape mode
4 with the search terms starting with 01, 01.1, et cetera.

5 A Yes.

6 Q Have you seen this list of search terms before?

7 A Well, I don't remember if I've seen exactly this list.
8 These are very complicated Boolean queries. I'd have to go
9 back and compare them to the documents I've looked at before.

10 Q And if you turn past those first three pages to the --
11 what is behind it. Do you recognize that as a form of a
12 stemming report?

13 A I have not seen this particular output format before. It
14 looks like something that could be a stemming report.

15 Q So that if we look, for example, just under search 01.1
16 where we have Weyerhaeuser stemmed as the first line, do you
17 see that there?

18 A Yes, I do.

19 Q And again I'm on the stemming report.

20 A Yes.

21 Q It lists numerous spellings of Weyerhaeuser, including
22 Weyerhaeuserization. And what's your understanding of what
23 would happen in a Boolean search if some other form of spelling
24 were included in a stemming report?

25 A I really don't know how to interpret this report. This is

1 the -- yeah, you know, I would have to know what the software
2 that produced this report was configured to do.

3 Q Okay. Well, let me ask you a different question. Your
4 understanding when you talked about the precision that was
5 needed for purposes of searching with Boolean search terms,
6 it's not the case that if the word is -- someone wants to know
7 about the price -- if prices were raised, right, that if the
8 term raise, R-A-I-S-E, is included and there's a stemming that
9 has a whole variety of different variations of the word raise,
10 those variations would all be picked up by the Boolean search;
11 isn't that correct?

12 A It would depend on the exact details of the stemming
13 algorithm. But certainly the intent of stemming algorithms
14 that are included in Boolean search systems is to allow some
15 degree of matching on morphological variations of a word. And
16 I'm sorry. I shouldn't use that word. But just different,
17 different words that are derived from the same root.

18 Q And so when you were suggesting before in your direct that
19 Boolean searches you have to have a precise match, it's not
20 just whatever the one word is that's in that search string,
21 it's whatever other stems may also exist with respect to the
22 words in that search string; isn't that correct?

23 A It would depend on how the stemming operations of the
24 Boolean search are implemented. In some cases a Boolean search
25 system will require an explicit operator at this end of the

1 stem, which then specifies the set of extensions of the stem
2 that can be matched. In other Boolean search systems there may
3 be a configuration of the software that would turn on stemming
4 for all of the words in the Boolean query.

5 Q But you haven't seen the stemming reports for any of the
6 defendants, is that correct?

7 A Not to my recollection.

8 Q Okay. So you don't know what words were included or not
9 included as variations of the words in the search term string,
10 correct?

11 A Under the assumption that the Boolean search algorithms
12 used stemming, I don't know what variations of the stems or the
13 words in the Boolean queries were being matched on. It would
14 depend on the configuration of the stemming operations of the
15 system.

16 When I discuss Boolean query systems as being exact
17 match systems, I am referring to the fact that the -- each
18 word -- the combinations of words that are expressed in the
19 Boolean query need to be matched exactly. There often in
20 information retrieval systems is some degree of stemming that
21 allows a small amount of matching on other words with the same
22 root or with the same beginning depending on how
23 morphologically accurate the stemming is.

24 Q Let's talk a little bit about supervised learning that you
25 are proposing. And your hope in the broader scheme beyond this

1 case would be to reduce the costs of discovery, is that
2 correct?

3 A My hope is to see statistical technology information
4 retrieval used to solve people's needs for information. Many
5 different factors go into people's needs for information.
6 Completeness of the information found, costs, amount of human
7 effort, degree of desirability of types of human effort.
8 There's a huge variety of factors that one optimizes in
9 information retrieval systems. Obviously costs is one of the
10 factors that's of considerable interest.

11 Q And one of the things that you'd like to do is reduce the
12 costs of eDiscovery through supervised learning?

13 A I believe that supervised learning has considerable
14 potential for reducing costs of eDiscovery through the
15 possibility of leveraging the manual effort in review to
16 produce -- you know, in order to produce strong predictive
17 models that bring more responsive documents to the fore,
18 enabling the responsive documents to be reviewed with less
19 manual effort. So it's certainly one of the desiderata.

20 Q You've heard of a linear review, is that correct?

21 A Yes, I have.

22 Q And a linear review is when someone looks at each document
23 one after the other, correct?

24 A Right. My understanding of linear review and I -- my
25 understanding is also that there's some variation in what

1 people mean by linear review. But it is usual -- my
2 understanding is this usually referred to a process where some
3 set of documents is identified and each of those documents is
4 looked at by some person.

5 Q And if the defendants had some groups of documents that
6 were reviewed by linear review, you're not objecting to that,
7 are you?

8 A Well, I'm not here to object to anything. I'm here to
9 provide my analysis of defendants' process and my analysis of
10 plaintiffs' proposed process.

11 Q And you're not criticizing a linear review that may have
12 been done with respect to segments of documents, isn't that
13 correct?

14 A I'm not here to criticize any linear review. I have not
15 been asked to analyze such a thing.

16 Q So that if there were some group of documents that were
17 collected and put before the contract attorney to have every
18 document reviewed individually, your testimony about supervised
19 learning would be not directed at that, is that correct?

20 A I'm not sure that I understand your question. Could you
21 repeat it.

22 Q Sure. Let's assume that there is a server, and the server
23 has some folders that are dedicated to a topic that through the
24 process counsel has determined may have potentially relevant
25 documents. You're not suggesting that instead of saying that a

1 contract attorney ought to review every document in that
2 particular folder, that they can't do that, that instead they
3 ought to use supervised learning for those documents?

4 A Well, again, I'm not saying that anybody can or can't do
5 anything.

6 Q When you talk about recall and precision, I just want to
7 make sure I have the terms correct. When you talk about
8 recall, you are assuming that if there is some universe of
9 documents that are responsive, recall defines what percentage
10 of those are captured by the search process that was
11 undertaken, is that correct?

12 A Yes. Recall describes the extent to which a system has
13 found all of the responsive documents that are available in
14 whatever the appropriate universe to be considered and the
15 particular problem is.

16 Q So that, for example, if there were out in a collection of
17 a million documents 100,000 documents that were actually
18 responsive and your system collected 90,000 of that hundred
19 thousand responsive documents, you would say the recall is
20 90 percent, correct?

21 A Yes, that's correct.

22 Q And I think you testified earlier this afternoon that
23 recall is the most important factor, correct?

24 A My understanding in the discovery context is that there is
25 a great premium on the finding of responsive documents. And so

1 my impression would be then that recall is the most important
2 effectiveness measure in a discovery setting. Obviously
3 precision affects what the costs of subsequent review are going
4 to be. But in some sense recall seems fundamental given the
5 need to find the majority of responsive documents.

6 Q And in the TREC studies, the interactive tasks, again, of
7 the 24 runs in 2009 completed just a couple years ago, only 5
8 of the 24 got over 70 percent recall, is that correct?

9 A I believe that's the case. Of course, the systems were
10 not being told to optimize recall.

11 Q And in the field of information technology there is no
12 defined minimum level of recall, isn't that correct?

13 A That's correct. That was something -- that would be
14 something that's a task specific. What would be a sufficient
15 level of recall or any other effectiveness measure depends on
16 the details of the task.

17 Q Now, let's go back to my example before where we had the
18 million documents and we had found through our search
19 methodology the 80,000 of the 100,000 that are responsive. But
20 our search methodology has actually pulled in 200,000 documents
21 so that the search methodology has 200,000 documents, of which
22 80,000 are responsive. Are you with me so far?

23 A I believe you said 90,000 were found before. Are you now
24 saying 80,000?

25 Q Let's use 90,000. I was trying to make the math easier

1 for myself, but that's okay. We can go with 90. Let's assume
2 that there are 90,000 responsive out of the 200,000 that are
3 pulled in by your search methodology.

4 A Okay.

5 Q Your precision measure in that case would be 45 percent,
6 is that correct?

7 A Oh, well, I'd have to sit down and do the math. But let's
8 see. You're saying that there's -- well, let's go back to your
9 80,000 then if we're going to do that computation.

10 (Laughter.)

11 THE WITNESS: So you would be saying that of the
12 200,000 the system found 80,000 were responsive, 120,000 were
13 nonresponsive. So we'd have 12 over 20, which would be -- it
14 would be 6 over 10. So that would be a 60 percent precision.

15 BY MR. McKEOWN:

16 Q 60 -- oh, so the precision measures the number that are
17 nonresponsive or the number that are responsive?

18 A Oh, no, I'm sorry. I got it backwards. It's the -- yes,
19 80 -- I'm sorry. 80,000 over the 280,000. So it -- yes, it's
20 40. It's 40 percent.

21 Q And if I took those 200,000 and then gave them to contract
22 attorneys and had them review them for responsiveness so that
23 ultimately 120,000 documents were produced -- well, if I had
24 them --

25 A You mean the other way around, right?

1 Q I'm sorry. I'm going to change the numbers to make it
2 easier.

3 A Okay.

4 Q All right. My search methodology has collected 200,000.

5 A Okay.

6 Q Out there in the mystical world where we know exactly what
7 is responsive, there are 80,000 responsive documents. We're
8 changing the math to go to 80, right?

9 A Okay. So we have a new example here. Could you give me
10 all the parameters of the new example.

11 Q Certainly. My search methodology has drawn in 200,000
12 documents, correct?

13 A So could you first tell me what the size of the collection
14 is.

15 Q 1 million.

16 A Okay.

17 MR. McKEOWN: It's not going to be that hard of a
18 question at the end, and I know I'm very close to my 10-minute
19 limit, Your Honor. And I will wrap this up quickly.

20 THE COURT: I know you will.

21 BY MR. McKEOWN:

22 Q So I have a universe collected of a million documents. My
23 search methodology has brought back 200,000 documents.

24 A Okay. 200,000 hits.

25 Q 200,000 hits. There are only 80,000 that are truly

1 responsive.

2 A Out in the whole 1 million.

3 Q Out of the whole -- well, out of the 200,000.

4 A There's 80,000 in the 200,000 hits.

5 Q That are responsive?

6 A Okay.

7 Q All right. If I use contract lawyers to review those
8 documents and mark them responsive and nonresponsive, and the
9 contract attorneys mark the 80,000 documents that are
10 responsive as responsive and the other 120,000 as
11 nonresponsive, the set that I have produced to the other side
12 has a 100 percent precision rate, correct?

13 A If they do that with complete accuracy, that would be the
14 case.

15 MR. McKEOWN: Your Honor, this would be a good time
16 to break.

17 THE COURT: Okay. Thank you. So you may step down,
18 Dr. Lewis, but don't go too far because you're involved in the
19 next hearing too. Okay. Now, if we solve this, we could
20 settle the case. Monday, March 5th, next Monday, I'm assuming
21 that's probably not a go. But a week from Friday, March 9th,
22 could people switch things around? Mr. Regard, can't do it.
23 Okay.

24 MR. MAROVITZ: Your Honor, it's Andy Marovitz for
25 Temple-Inland. I have Mr. Regards' availability. During the

1 break I had a chance to grab it. We'll work around yours
2 obviously to the extent we can. Mr. Regard has some child
3 caring issues that he's trying to work around.

4 THE COURT: How is Wednesday, March 28th?

5 MR. MAROVITZ: That's fine, Your Honor.

6 THE COURT: Can you guys do March 28th?

7 MR. MAROVITZ: It's fine with Mr. Regard.

8 THE COURT: Right.

9 MR. MAROVITZ: I shouldn't speak for the other
10 defendants.

11 THE COURT: Dr. Lewis, it's a Wednesday March 28th.

12 DR. LEWIS: I believe that will be fine, Your Honor.

13 THE COURT: Okay. And we just have to find out from
14 our plaintiffs.

15 MR. MOGIN: Your Honor, I hate to be a fly in the
16 ointment. I have a -- would we be concluding on the 28th?

17 THE COURT: Yes, I can also do the 29th if the 29th
18 is better.

19 MR. MOGIN: Well, what I'm trying to schedule around,
20 Your Honor, is that the Southern District of California's
21 District Conference is the 29th.

22 THE COURT: Oh, so you want to get back. Well, I'd
23 like to start very early in the morning actually on the 28th.
24 And I think we -- I would like to suggest three things for the
25 agenda. We have to finish -- we're going to do Mr. Regard's

1 cross-examination. We are going to do Mr. -- well, actually
2 direct. I mean direct and whatever. And then we're going to
3 finish up Dr. Lewis', and I would like to have a meeting. I'd
4 like to have a Rule 16 conference with the lawyers. If we're
5 all going to be together, it can be 15 minutes or it can be
6 longer, but I would like to sit down with you. And I didn't
7 get an opportunity at the beginning of the case.

8 MR. NEUWIRTH: Just one point of inquiry, Your honor.
9 I don't know if that agenda was meant to be the order in which
10 we would do things. But since we are in the middle of the
11 cross of Dr. Lewis, it would seem to make sense to start with
12 finishing that.

13 THE COURT: Dr. Lewis, can you get down here at
14 8:00 o'clock in the morning?

15 DR. LEWIS: Sure.

16 THE COURT: Okay. He's close.

17 MR. NEUWIRTH: Thank you.

18 THE COURT: He's my neighbor. I'll give him a ride.

19 MR. FREED: Can I add Dr. Tenny to the mix? Because
20 at this point her report has been in the defendants' hands for
21 quite a long time. And we did have a motion for
22 reconsideration. They knew everything that they needed to know
23 on the 16th. If there was some surprise relative to this
24 hearing today, that's one thing. But to a hearing that is out
25 another month, Your Honor --

1 THE COURT: What new is she going to add to what Dr.
2 Lewis said?

3 MR. FREED: Well, she has certain issues which have
4 been raised with respect to search strings, which is precisely
5 an area of expertise which she has analyzed. And I would at
6 least like to keep the option open. And believe me if we feel
7 it's redundant, we're not going to impose on Your Honor or the
8 defendants. But --

9 MR. ECHOLS: The only question, Judge, Barack Echols
10 on behalf of Packaging Corporation of America, would be with
11 respect to any Daubert issues. At least based on what we saw
12 when we got the report on Thursday it's not clear that there's
13 any appropriate qualification or fit for this witness'
14 testimony in this particular circumstance. I don't know how
15 you would want those to be raised and addressed.

16 MR. FREED: We'd be very, you know, happy to do that
17 in advance so that there wouldn't be an issue about again
18 having her attend and then not be qualified.

19 MR. ECHOLS: Exactly. That was part of what you had
20 said the last time, Judge, not to waste anybody's time.

21 THE COURT: Why don't you talk to each other and see.
22 I don't want to be preclu -- I'm trying not to preclude
23 anybody, but I don't think after nine hours today that we need
24 any redundancy here.

25 MR. FREED: We don't, Your Honor. And we will

1 seriously consider it. I just want to keep the option open.

2 THE COURT: This gentleman needs to get to
3 California, and I have a 4 o'clock with Judge Holderman that
4 day, so I need to get there. So that's why we're starting at 8
5 o'clock in the morning. And I do want to have enough time for
6 this Rule 16. Yes.

7 MR. MAROVITZ: Judge, may I raise just two issues
8 just of housekeeping.

9 THE COURT: Absolutely.

10 MR. MAROVITZ: First, 8 o'clock in morning is just
11 fine. We intend at least a week before March 28th to give the
12 plaintiffs, and we'd like to submit to Your Honor as well,
13 additional information about the chart that was tendered. We
14 don't want there to be any claim of surprise or lateness or
15 anything else. So of the same kind of information that was
16 submitted beforehand whether it's called a report or something
17 else, we just --

18 THE COURT: And that goes to Mr. Regard's testimony
19 anyway.

20 MR. MAROVITZ: Correct. Correct.

21 THE COURT: Okay. That's fine.

22 MR. MAROVITZ: Second, in that connection also I
23 don't know if it applies generally here, but we would like a
24 suspension of the witness rule. We do need to talk to Mr.
25 Regard about that. And technically we'd like to be able to do

1 that.

2 THE COURT: And that's fine. And if they need to
3 talk to Dr. Lewis, that's fine too. Is that okay with
4 everybody? Yes?

5 MR. MOGIN: Yes.

6 THE COURT: Okay. Yes.

7 MR. McCAREINS: Mark McCareins on behalf of Rock
8 Tenn. I just talked to counsel for plaintiffs. I think we've
9 agreed that he would kind of fish or cut bait within the next
10 couple weeks --

11 MR. FREED: Absolutely.

12 MR. McCAREINS: -- as to whether he would even need
13 that additional expert.

14 THE COURT: Just think about it. Okay.

15 MR. FREED: We are, Your Honor. And we don't want to
16 impose on the Court or other counsel. We will give it very
17 serious consideration.

18 MR. MAROVITZ: And, Your Honor, finally. On
19 Mr. Regard there is obviously some more direct to go, but not
20 very much. There is obviously the point of the chart that we
21 didn't have a chance to go through today. And then Mr. Regard
22 offered opinions at the very outset about a couple other issues
23 that we just didn't get to. It's a --

24 THE COURT: Well, that's not a problem. I mean, I
25 cut, I cut that off. Okay. And we didn't finish with Dr.

1 Lewis. So that makes a lot of sense.

2 MR. MAROVITZ: Thank you, Your Honor.

3 THE COURT: So we'll get all that done. Then we'll
4 be all finished. Do you go back to Judge Shadur also? I meant
5 to ask that. Do you have a date with Judge Shadur?

6 MR. NEUWIRTH: March the 15th, Your Honor.

7 MR. MAROVITZ: Yes, March 15th at 8:45.

8 THE COURT: Do you physically come on March 15th?

9 MR. NEUWIRTH: Yes, Your Honor.

10 MR. MAROVITZ: We do.

11 MR. McKEOWN: Well, I think the -- if there's this
12 convention that has occurred that if Judge Shadur is satisfied
13 with the status report, we don't need to appear on the 15th.
14 And perhaps if the Court could suggest to Judge Shadur that we
15 don't need to appear because we're coming here on the 28th.

16 THE COURT: That we have more to do as far as the
17 purpose of the referral.

18 MR. McKEOWN: Yes.

19 THE COURT: Okay. I'll certainly do that.

20 MR. McKEOWN: Thank you, Your Honor.

21 MR. MAROVITZ: Thank you, Your Honor.

22 MS. MILLER: Thank you, Your Honor.

23 THE COURT: Okay. Well, we'll see everybody then.

24 We're saying Wednesday, March 28th, right. Okay. 8 o'clock in
25 to morning.

1 MR. MAROVITZ: Thank you, Your Honor.

2 MR. MOGIN: Your Honor, I had completely forgotten
3 completely about the Judge Shadur matter. But perhaps if we're
4 going to be here the 15th, we're all scheduled up anyway. If
5 the Court's got the availability.

6 MR. MAROVITZ: Not the experts, Your Honor.

7 THE COURT: I'm on criminal duty that week, which
8 means I have to be available all day, or I'm going to put you
9 all in jail. Okay.

10 MS. MILLER: The 28th it is, Your Honor.

11 THE COURT: All right. I have to be available the
12 entire day. Okay. Everybody, thank you.

13 MR. MOGIN: Thank you, Your Honor.

14 MS. MILLER: Thank you, Your Honor.

15 MR. MAROVITZ: Thank you.

16 (Whereupon, said trial was recessed at 5:45 p.m., to
17 reconvene on 3/28/12, at 8:00 a.m.)

18 CERTIFICATE

19 I HEREBY CERTIFY that the foregoing is a true,
20 correct and complete transcript of the proceedings had at the
21 hearing of the aforementioned cause on the day and date hereof.

22

23 /s/TRACEY D. McCULLOUGH

February 27, 2012

24 Official Court Reporter
25 United States District Court
Northern District of Illinois
Eastern Division

Date