



504 - eDiscovery MiniClinic: Your Roadmap to Success

Anne Milem
Associate General Counsel
Sallie Mae, Inc.

Eric Schwartz
Team Leader, Legal Technology Services
Ernst & Young Global Limited

Sonya Sigler
General Counsel
Cataphora Inc.

Evan Stolove
Associate General Counsel
Fannie Mae

Faculty Biographies

Anne Milem

Anne L. Milem is associate general counsel at Sallie Mae in Reston, Virginia. She manages Sallie Mae's litigation docket which includes class actions, individual consumer lawsuits and some commercial litigation. Ms. Milem also works on other litigation-related matters, including government inquiries, subpoenas, and e-discovery.

Before moving in-house, Ms. Milem was a commercial litigator with the law firm of Arent Fox in Washington, DC. She served in both first-chair and second-chair roles at jury trials. Ms. Milem's litigation experience was varied but included trademark disputes, government contract disputes, employment cases, and various pro bono cases. She was a member of Arent Fox's employment committee for several years. Ms. Milem clerked for Judge Stanley P. Klein at the Fairfax County Circuit Court for two years immediately following law school.

Since joining Sallie Mae, Ms. Milem has been very active in the ACC's Washington Metropolitan Area Chapter. She is a founder and co-chair of WMACCA's Litigation Forum. Ms. Milem served for ten years on the Board of the Northern Virginia Chapter of the Virginia Women Attorneys Association in different positions.

Ms. Milem graduated summa cum laude from Texas Tech University with a B.A. and is an alumni of Georgetown University Law Center. At Georgetown, she participated in the Domestic Violence Clinic and was competitively selected as a Law Fellow and later a Senior Writing Fellow in the law school's Legal Research & Writing Department.

Eric Schwartz

Team Leader, Legal Technology Services
Ernst & Young Global Limited

Sonya Sigler

Sonya Sigler is the vice president, business development & general counsel at Cataphora, Inc. in Redwood City, California. She actively works on intellectual property matters including negotiating and drafting various Internet, technology, software licensing and other agreements.

Ms. Sigler worked as an attorney at Sega supporting the business development, product development, marketing, and finance groups, as well as Sega Studios in Los Angeles. More recently, She worked at Intuit, where she evaluated and negotiated agreements and licenses for Intuit's industry-leading financial software products, Quicken, Quicken Mortgage, and QuickBooks. She also has extensive experience as a business development and legal consultant to start-ups such as Treasure Media, RealCommunities and IDO Systems, as well as established companies such as Sony and Intuit.

Ms. Sigler writes articles and speaks frequently on the subjects of electronic discovery, intellectual property, and other topics. She is a member of the ACC, the ABA, and The Sedona Conference Working Group 1 on Electronic Document Retention and Production. She is active in the community as a member of the Governance Council for the San Carlos Charter Learning Center and the chickens' ball steering committee. Past board work has included the Women in Interactive Entertainment Association, Women in Technology Advisory Board and the Nova Vista Symphony.

Ms. Sigler holds a J.D. from Santa Clara University and a B.A. from UC Berkeley and is a member of the

Evan Stolove

Evan S. Stolove is an associate general counsel in the litigation practice group at Fannie Mae in Washington, DC. He specializes in representing the Company in commercial disputes, shareholder litigation, regulatory investigations and other matters that involve potential headline risk. He is also the litigation practice group's preservation program and E-Discovery manager.

Prior to joining Fannie, Mr. Stolove was a partner with the law firm of Arent Fox, PLLC in its Washington, D.C. office. Mr. Stolove's practice involved commercial litigation with a specialization in intellectual property litigation, including trademark, copyright, cyberspace and trade secrets, as well as First Amendment litigation.

Upon graduation from law school, Mr. Stolove served as a judicial clerk to Judge John C. Eldridge of the Maryland Court of Appeals, and then to Judge André M. Davis of the United States District Court for the District of Maryland.

Mr. Stolove received his A.B. from the University of Michigan and his J.D., with honors, from the University of Maryland School of Law.

Informational Retrieval Glossary

Boolean

Boolean search is a common technique which allows terms to be logically connected using AND, OR and NOT to refine a query. For example: "(mercury OR saturn) AND NOT automobile" would return all documents containing either the word mercury or the word saturn and not containing the word automobile.

Bayesian Classification

In general, bayesian classification is based on the statistical probability of a class and the features associated with that class. This type of classification utilizes a training set composed of classes that have correctly assigned features. Once the probabilities of the training set features and classes have been stored, new data is compared against the training set. During this comparison of the "learned" classification of the training set with the new data, the new data's features are calculated and the new data are assigned classes whose probability of matching the training set's classes and features is highest.

Categorization

In general, categorization is the grouping of objects, people or ideas on the basis of some kind of "similarity". As applied to electronic discovery, it describes the grouping of documents according to some desired criteria. Categorization may be by topic or by legal criteria. Categorizing documents about specific products, or documents that relate to sales in a given country are examples of categorizing by topic. Legal criteria might include categorizing responsive, non-responsive, and privileged documents.

Clustering

Clustering is the grouping of information by some category or statistical similarity. This is done through various grammatical, semantic, and even punctuation algorithms designed by combination to detect topics rather than just individual keywords. Statistical clustering can be done by counting words and their frequency, then grouping those documents with similar statistics together in a cluster. When files are determined to be about the same or similar topics, they are clustered together, and usually displayed in some kind of graphical relationship that facilitates reviewing similar documents together.

Concept

Concept search attempts to find documents that address some concept that a user is interested in. To do so, it goes beyond *keyword* search for documents that contain a specified word or phrase, and tries to find other documents that address the underlying concept. For example, a concept search for "fiber" might return documents that refer to the concept of fiber using alternative terms such as cloth, material, cotton etc.

Keyword

Keyword search looks for documents that contain a specific word or phrase. Keyword searches may be further refined by using *Boolean* operators.

Latent Semantic Indexing

Latent Semantic Indexing has involves extracting multiple concepts from the data collections through a statistical semantic analysis of each file. The theory is that unstructured files comprise *latent* concepts that are not readily recognized and remain hidden until a more precise lexicon is developed out of the whole collection. These concepts then form a dictionary (lexicon) for the collection that can be weighted for both frequency of occurrence and relevance. At that point each file in the collection is compared to the concepts list, and it is assigned a *fingerprint* (or value) that uniquely defines the file according to those criteria. Searches can then be conducted by requesting files that are statistically similar, i.e. that have similar fingerprints, under the presumption they will be not just similar but conceptually related as well.

Informational Retrieval Glossary

Linguistic Techniques

Search or categorization techniques that are based on analysis of language features of documents, in contrast with *statistical* techniques. *Ontologies* are an example of a linguistic technique.

Neural Network

A neural network is a computer program whose operation is loosely inspired by the way a human or animal brain works (though the neural network is much, much simpler). A neural network can be "trained" by giving it sample inputs and the correct outputs associated with these. The network can analyze the difference between the answers it is generating and the "correct" answers. It can then automatically adjust its internal workings, until its answers on the training set adequately match the given outputs. The idea is that you can now feed it new inputs (the answers to which are unknown) and it should now be able to provide the correct outputs for these. For purposes of electronic discovery, the inputs might be information about documents and the outputs a *categorization* of those documents.

Ontology

An ontology is an arrangement of words, phrases and search terms under a *concept*. Here is a simple example:

AIRCRAFT CONCEPT

- Boeing
- 747
- Cessna
- Glider

By reading documents and establishing whether they contain any of the four terms listed under the AIRCRAFT CONCEPT, we can determine whether any of the documents discuss the concept of aircraft. This process can be automated, so that a computer does the work. If the computer finds a document that contains one or more of the four terms, it concludes that the document is (at least partially) about aircraft. The document might also discuss other concepts, but a reference to the concept of aircraft is clearly present in the document.

Relevant

"Having some reasonable connection with, and in regard to evidence in trial, having some value or tendency to prove a matter of fact significant to the case." Finding relevant documents more effectively than our competitors is a key value that we provide.

Related

Documents that may be related to each other in a number of ways, such as addressing the same issue, or being created by a significant *actor* at a crucial time. Most electronic discovery software can find only documents which are similar to each other, or which belong to the same *email thread*. Cataphora *Discussions* are the first technology to pull together related documents, not just similar ones.

Statistical

Categorization may be based on statistical analysis of the similarity of documents. For this, a document is mathematically represented by a set of features such as the occurrence of words, or their proximity to other words in the documents. Different weights (levels of importance) may be assigned to the various features. Documents are then deemed to be similar (and therefore belong to the same category), based on the degree to which their features resemble each other.

Taxonomy

Taxonomy is the practice and science of classification. Taxonomies, which are composed of *taxonomic units* known as *taxa* (singular *taxon*), are frequently hierarchical in structure, commonly displaying parent-child relationships.

Vector Space Modeling

Vector Space Modeling (VSM) is a concept that first came into favor in the early 1970s and it has provided some additional guidance in automated document review even to this day. It is based on building vectors that describe the relationships between each search query and each file in the collection. Each vector, by its magnitude and direction then maps to other files that are closest to it in relation to the same *feature* as emphasized by the search query. Each file thus becomes a compilation of *features* that place it in a multi-dimensional construct. That construct can be realized in a graphical display depicting all the relationships as vector lines between and among separate files.



Document Discovery Costs

**What about
volume?**



Costs Hypothetical – Hardcopy Process

- Standard process is to collect all relevant hard copy materials relevant to the matter from identified individuals
 - Work areas
 - OP and related staff areas
- Discipline libraries
- File rooms
- Records Center
- (Also collect removable electronic storage and portable computing devices at the same time)



Costs Hypothetical – Hardcopy Costs

- 500,000 pages collected (125,000 documents)
- Collection costs
 - Out of pocket
 - Two lawyers: 172 hrs @ \$250/hr= \$43,000
 - Paralegal: 86 hours @ \$125/hr= \$12,900
 - Travel, lodging & meals for 4: \$18,275
 - Internal
 - Paralegal: Soft cost 86 hours
 - Custodians: Soft cost 150 hours
 - Associated lost opportunity
 - Shipping
 - Boxes: 250 boxes @ \$35/box = \$8,750
- Subtotal **\$82,925**



Costs Hypothetical – Hardcopy Costs

- 500,000 pages collected (125,000 documents)
- Processing costs
 - Prep/image/OCR/endorse (@ .36/pg)
 - 500,000 (@ \$0.36) \$180,000
 - Document coding (@ \$1.15/document)
 - 125,000 (@ \$1.15) \$143,750
 - Oversight (@ \$75/hr) \$3,000
 - Other \$1,000
 - Return 250 boxes @ \$35/box = \$8,750
- Subtotal **\$336,500**



Costs Hypothetical – Hardcopy Costs

- 500,000 pages collected (125,000 documents)
- Review costs
 - 500,000 pages @ 100 pages/hour = 5,000 hours
 - 5,000 hrs @ \$250/hr = \$1,250,000
- Subtotal \$1,250,000
- Produce relevant documents (assume 20%)
 - Electronic (TIFF / PDF) @ \$0.04/image
 - 100,000 relevant images = \$4,000
 - Paper @ \$0.11/page
 - 100,000 relevant pages = \$11,000
 - Shipping 50 boxes @\$35/box \$1,750



Costs Hypothetical – Hardcopy Costs

- 500,000 pages collected (125,000 docs)
- Collection \$82,925
- Processing \$336,500
- Review \$1,250,000
- Produce \$16,750

- TOTAL \$1,686,175



Costs Hypothetical – Electronic Process

- Standard process is to collect for each individual
 - Email exchange
 - DWS Backup
 - S drive
 - Typical custodian has ~1.5 GB of data
- Standard process is to collect file shares and other sources with relevant data as well
 - File share size varies



Costs Hypothetical – Electronic Process

- Facts relevant to electronic collection
 - 50 document custodians
 - 1.5 GB per custodian = 75 GB
 - 1 GB = 75,000 pages
 - Page count = 5,625,000
 - 7 file shares
 - Data stored = 50 GB
 - 1 GB = 75,000 pages
 - Page count = 3,750,000
- Total page count 9,375,000
- Total GB count 125



Costs Hypothetical – Electronic Context

- If the collection were processed using *basic* data extraction techniques and then reviewed *online* the costs would be:
 - Data extraction @ \$3,000/GB
 - For 125 GB \$375,000
 - Review 9,375,000 pages
 - Review rate @ 200 pages/hr = 46,875 hrs
 - 46,875 hrs @ \$250/hr = \$11,718,750
 - (20 staff for 59 weeks)
 - Produce relevant content (assume 20%)
 - Electronic (TIFF / PDF) @ \$0.04/image
 - 1,875,000 images = \$75,000
- Total \$12,168,750



Costs Hypothetical – Electronic Context

- If the collection were processed using *basic* data extraction techniques, deduplication and then reviewed *online* the costs would be:
 - Data extraction @ \$3,000/GB
 - For 125 GB \$375,000
 - Assume 20% deduplication of 9,375,000 pages
 - 7,500,000 pages @ 200 pages/hr = 37,500 hrs
 - 37,500 hrs @ \$250/hr = \$9,375,000
 - Produce relevant content (assume 20%)
 - Electronic (TIFF / PDF) @ \$0.04/image
 - 1,500,000 images = \$60,000
- Total \$9,810,000



Costs Case Study – Electronic Costs

- If the collection were processed using *advanced* electronic search techniques and *online* review the costs would be:
 - Intake processing of 125GB @ \$4,440/GB
 - Assume 30% volume reduction
 - 87.5 GB at \$4,440/GB = \$388,500
 - Search algorithm development \$112,750
 - (175 hrs linguist, 200 hrs engineer)
 - Search review and refinement \$140,250
 - (200 hrs linguist, 270 hrs engineer)
 - Identify review set (20%)- 17.5 GB
 - [17.5 @ 75,000 = 1,312,500]



Costs Hypothetical – Electronic Context

- Lawyer search development time
 - 100 hrs @ \$350/hr = \$35,000
 - Lawyer review time
 - 1,312,500 @ 1,000 pages/hr = 1312 hrs
 - 1312 hrs @ \$250 = \$328,125
 - Total to establish response set \$1,004,625
 - 1,312,500
 - Produce response set
 - Electronic (TIFF / PDF) @ \$0.04/image
 - 1,312,500 images @ \$0.04 = \$52,500
-
- Total cost \$1,057,125



Costs Hypothetical

	Hardcopy Conventional	Electronic Basic	Electronic Basic with Deduping	Electronic Advanced
Collect	\$82,925			
Process	\$336,500	\$375,000	\$375,000	\$388,500
Filter Development				\$288,000
Review	\$1,250,000	\$11,718,750	\$9,375,000	\$328,125
Produce	\$16,750	\$75,000	\$60,000	\$52,500
Totals	\$1,686,175	\$12,168,750	\$9,810,000	\$1,057,125

Sample Resources

ACC Annual Meeting 2007
Chicago, IL

Session 504 - eDiscovery MiniClinic: Your Roadmap to Success

Tuesday, October 30, 2007 11:00 AM - 12:30 PM

1. Books

The Electronic Discovery Handbook: Forms, Checklists, and Guidelines by Sharon D. Nelson, Bruce A. Olson and John W. Simek (Published by ABA Law Practice Management Section): This book is more practical and has helpful checklists.

Electronic Discovery and Evidence by Mike Arkfeld (self published at www.arkfeld.com)

The Discovery Revolution by George Paul, (Published by ABA Press): this book focuses on FRCP rule changes but the first chapter a detailed historical perspective)

Practical Guide to E Discovery by Mary Mack (Published by Fios) This book has an excellent overview of electronic discovery processes.

2. Associations and Groups

Association of Corporate Counsel ("ACC") Virtual Library Free to Members, and includes: 1) "Sample Electronic Discovery Interrogatories and Requests for Production of Documents;" 2) InfoPak on "Records Retention;" and 3) "Ten Tips for Electronic Discovery."

The Sedona Conference Website
Free, Research and educational institute dedicated to the advanced study of law and policy in the areas of antitrust law, complex litigation, and intellectual property rights. Working Group 1 writes on topics such as search and retrieval technologies as well as the Sedona Principles. www.thesedonaconference.org

General Counsel Roundtable
For Members: Creating a Data Map
<https://www.gcr.executiveboard.com/Members/DecisionSupportCenters.aspx/63163918>

3. Whitepapers and Articles

Discovery of Digital Information by Ronald J. Hedges (ret. NJ)
<http://www.kenwithers.com/articles/hedges092704.pdf>

Jenner & Block Meet & Confer whitepaper -
http://www.jenner.com/files/tbl_s20Publications%5CRelatedDocumentsPDFs1252%5C637%5CElectronicDividends.pdf

Leveraging Content Analytics to Reduce E-Discovery Risks and Costs (Free Whitepaper available through www.KahnConsultingInc.com)

Managing Preservation Obligations After the 2006 Federal E-Discovery Amendments, by Thomas Y. Allman (Published in 13 Rich. J.L. & Tech. 9 (2007)) <http://law.richmond.edu/jolt/v13i3/article9.pdf>

4. Websites and Vendor Newsletters

Cataphora's "Discussions" - <http://cataphora.com/information/newsletters.php>

Discovery Resources

- Free, electronic discovery resources <http://www.discoveryresources.org/>
- Blog listings - http://www.discoveryresources.org/02_law_blogs.html

FIOS Newsletters - <http://www.fiosinc.com/counselor/>

FileNet's "Compliance Roadmap" - www.filenet.com

Includes: 1) ROI Calculator; 2) Whitepapers & Podcasts; and 3) Records Management Guide.

K&L Gates Electronic Discovery Blog - www.ediscoverylaw.com

Kroll Ontrak newsletters - www.krollontrack.com/newsletters/

5. Sample Documents

American Bar Association - <http://www.abanet.org/litigation/committees/pretrial/forms.html>

Aptara - http://www.aptaracorp.com/legal_forms.htm



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

Data Collection, File Types and Costs Checklists	
1. Electronic Data Collection Checklist	
Types of Data Stores	Places to Search for Relevant Electronic Data
Shared Servers/ or Network	All network nodes
Computers	Desktop(s) Laptop(s)/ Notebooks At work and at home
Handheld devices and PDA (Personal Digital Assistant)	TREO Blackberries Mobile Phones iPods or MP3 players
Memory Devices	Thumb Drives Memory Stick Memory cards (from digital cameras, phones, other devices)
Removable Media	DVDs CDs Floppy drives Tapes Other removable media cards Detached hard drives Zip drives Security Camera tapes
Archive or Backup media	Backup Tapes Storage Locations: - ~ File shares - ~email devices - ~archival tapes - ~hosted emails, - ~ Attachments Archives Document Managements systems
Databases	Credit card/debit card Databases VOIP/Phone logs IM databases



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	Business application databases (Quicken, Calendar, Address Book, etc.) Other Data Entry systems
Online	Audit Logs Access Logs Web Pages Blogs Printer memory caches Copier/Scanner memory caches
Deleted/Unused space	Slack Space Deleted/Recovered Files
Other	Desk Drawers File Cabinets Closets Warehouses Obsolete Computer Equipment 3 rd Parties Partners/Licensors Escrow placeholders Be creative in your thinking here



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

2. File Type Checklist		
Overview of File Type	Probable Extensions	
Documents	.doc .pdf .pps .ppt	.rtf .txt .wpd .xls
Images	.gif .jpeg .jpg .png	
Web Pages	.htm .html .mht .mhtml .shtml	
Media	.adf .avi .mov .mp3 .mpeg .mpg .ra	.ram .rm .swf .wav .wma .wmv
Databases	.mdb	
Applications	.qdf	
Others	www.filext.com	



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

Specific File Type	Summary	Typical File Extensions
ARCHIVE CONTAINER FILES		
ARCHIVE	Unknown compressed archive. This file type can contain one file or many files in multiple directories.	varies
ARCHIVE.ACE	A compressed archive. This file type can contain one file or many files in multiple directories. ACE is a proprietary data compression archive file format with superior compression compared to the ZIP file format , but at the cost of a lower compression speed.	.ace
ARCHIVE.ARJ	A compressed archive. This file type can contain one file or many files in multiple directories. ARJ is an archiving program created by Robert Jung for IBM-compatible computers. The letters stand for "Archive Robert Jung."	.arj
ARCHIVE.BZ2	A compressed archive. This file type can contain one file or many files in multiple directories. Bzip 2 UNIX Compressed File	.bz2
ARCHIVE.CAB	A compressed archive. This file type can contain only one file or many files in multiple directories.	.cab



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

ARCHIVE.EXE.ZIP	A self extractable zip archive or any executable that may contain zip archive.	.exe
ARCHIVE.GZ	A compressed archive which is generated using the GNU zip program. This program is used to compress the size of a single file. If more than one file has to be compressed, they must be tarred together first.	.gz
ARCHIVE.ISH	A compressed archive. This is not very commonly used.	.ish
ARCHIVE.ISO	An optical disk image or a Universal Disk Format (UDF) image. It contains all the data files and the filesystem metadata in a single file.	.iso
ARCHIVE.JAR	A compressed archive. JAR (for "Java ARchive") is a file format that bundles multiple files into a single archive file. Typically a JAR file contains the class files and auxiliary resources associated with applets and applications.	.jar
ARCHIVE.LZH	A compressed archive which uses the Lempel-Ziv and Haruyasu compression algorithm. It is used mostly in Japan.	.lzh
ARCHIVE.RAR	A compressed archive. RAR is a compressed file like ZIP, but uses a different encoding format	.rar
ARCHIVE.TAR	An archive. This file type can contain only one file or many files in multiple directories. Tar (for "Tape ARchive") is a Unix shell command that creates a single file called an " archive " from a number of specified files or extracts the files from such an archive. A tar archive has the file suffix ".tar". The files in a tar archive are not compressed,	.tar .tar.gz .tar.z .z



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

	just gathered together in one file.	
ARCHIVE.UUE	A compressed archive. This file type can contain only one file or many files in multiple directories. Compressed ASCII file archive created by UUENCODE (uuexe515.exe)	.uue
ARCHIVE.Z	A compressed archive. This file type can contain only one file or many files in multiple directories. UNIX Compressed Archive File	.Z
ARCHIVE.ZIP	A compressed archive. This file type can contain only one file or many files in multiple directories. Zip files compress data and therefore save time and space and make downloading software and transferring e-mail attachments faster	.zip
EMAIL CONTAINER FILES		
EUDORA	An email archive. May contain one or more email messages, calendar items, etc.	.mbx
LOTUSNOTES	A Lotus Notes database file. This type contains email folders, calendar items, tasks, custom forms, contact information, or other files that are stored in the database. Lotus Notes and Domino are a fundamental component of IBM Workplace.	.nsf
MBOX	An email archive. RFC-822 compliant format that may contain one or more email messages, calendar items, etc. This file type is a standardized	.mbx .eml



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

	email format.	Nescape Mozilla
OE	An Outlook Express email archive. This type includes all email folders, calendar items, tasks or contact information that Outlook Express collects.	
OE.DBX	An Outlook Express email archive. This type includes all email folders, calendar items, tasks or contact information that Outlook Express collects.	.dbx
OE.IDX	An Outlook Express email archive. This type includes all email folders, calendar items, tasks or contact information that Outlook Express collects.	.idx
TEXT EXTRACTABLE FILES		
ASCII	A text file with pure ASCII text and no formatting.	.txt, .asc, .prn
CODE.C	C/C++ Program File	.cpp, .c, .h
CODE.JAVA	Java Source Code	.java
CODE.JS	JavaScript Source Code, This extension can be captured by a number of applications including: Windows Script Host, Dreamweaver MX, Notepad, Netscape Navigator, PavScrip, UltraEdit. The files are generally text files. Note: This file type can become infected and should be carefully scanned if someone sends you a file with this extension.	.js
CODE.SQL	Structured Query Language Data	.sql
CONCORDANCE	Single page tiff formats bounded by summation meta file documents.	.dat/.dii
EMAIL	Individual email messages or components (calendar items, tasks, etc.).	



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

	Culled from Exchange or Notes.	
HTML	A Hypertext Markup Language file. Any text data can be encoded in HTML.	.htm .html
MSOFFICE	Desktop application files that are part of Microsoft Office other than Excel, Word or PowerPoint**	unknown
MSOFFICE.DOC	Individual desktop application files. Files that are identified as part of Microsoft Word family.	.doc
MSOFFICE.MPP	Microsoft Project	.mpp
MSOFFICE.PPT	Individual desktop application files. Files that are identified as part of Microsoft PowerPoint family.	.ppt
MSOFFICE.XLS	Individual desktop application files. Files that are identified as part of Microsoft Excel family.	.xls
PDF	An Adobe <u>Acrobat portable document format</u> . The PDF format has become a standard for document transfer between computer system architectures.	.pdf
RTF	A Rich Text Format File	.rtf
VSD	A Microsoft Visio file. A visual/graphics program that helps in creating business and technical diagrams.	.vsd
WPD	A <u>WordPerfect document</u> . WordPerfect was once a substantial word processing application.	.wpd
XML	An Extensible Markup Language File. Any text data can be encoded in XML.	.xml



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

IMAGE FILES		
IMAGE.BMP	A Windows OS/2 bitmap graphic. Common image file type.	.bmp
IMAGE.GIF	A Graphic Interchange Format. Common image file type.	.gif
IMAGE.JPG	JPEG/JIFF Image. Common image file type.	.jpg .jpeg .jif
IMAGE.PCX	A PC Paintbrush bitmap graphic. Common image file type.	.pcx
IMAGE.PNG	PNG (for Portable Network Graphics) is a file format for image compression	.png
IMAGE.TIF	A Tagged Image Format File (tiff). TIFF images are a common format when converting electronic/paper materials into a standardized storage or review mechanism.	.tif
AUDIO FILES		
AUDIO	An audio file format is a container format for storing audio data on a computer system. The general approach towards storing digital audio formats is to sample the audio voltage in regular. Therefore sample rate , resolution and number of channels are key parameters in audio file formats.	.669,.6cm,.8cm,.8med,.8svx,.ahx, .aifc,.aiff,.al,.alaw,.amr,.ape,.au,.avr, .cda,.cel,.cmf,.cwp,.d00,.d01,.dss,.far, .flac,.iff,.it,.kar,.kin,.ktp,.m3u,.m4a,.med, .mmf,.mod,.mp2,.mpc,.mpu,.mp_., .mtm,.nap,.ogg,.pca,.pcm,.pvc,.rmf, .rmi,.rmx,.s3m,.saf,.sd2,.sdx,.sf, .sfl,.sib,.smf,.snd,.sng,.svx, .sw,.swa,.tfmx,.thx,.tsp,.uax,.ub,.vmf,.voc, .voi,.vox,.wave,.wax,



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

		.wrk,.xm,.xmi,.xmk,.wrk,.xm,.xmi
AUDIO.AAC	Highly compressed audio file, similar to MP3, that maintains quality nearly indistinguishable from the original audio source	.aac
AUDIO.AIF	CD-quality audio file, similar to a .wav file	.aif
AUDIO.MID	Musical Instrument Digital Interface standard file; contains music data such as what notes are played, when they are played, how hard the key is pressed, how long it is held, etc; does not contain actual audio	.mid
AUDIO.MIDI	Musical Instrument Digital Interface file; contains music data such as what notes are played, when they are played, how hard the key is pressed, how long it is held, etc; does not contain actual audio	.midi
AUDIO.MP3	Compressed audio format developed by the Moving Picture Experts Group; near-CD quality sound (stereo, 16-bit)	.mp3
AUDIO.MPA	Compressed audio file using MPEG Layer I, II or III compression	.mpa
AUDIO.RA	Audio file created by RealPlayer	.ra
AUDIO.RAM	Audio file created by RealPlayer	.ram
AUDIO.WAV	CD-quality audio file, similar to AIFF	.wav
AUDIO.WMA	Audio file compressed with Windows Media compression	.wma
VIDEO FILES		
VIDEO	An video file format is a container format for storing video data on a computer system.	.3g2,.3gp,.3gp2,.3gpp,.3mm,.avs,.by,dir,.dv,.dvrms,.dvr,.fla,.flc,.fli,.flv,.flx,.gl,.grasp,.ifo,.ivf,.ivs,.isx,.m1v,.m4e,.m4u,.mkv,.moov,.movie,.mpe,.mpeg,.mpv2,.mvc,.omf,.prx,.qtch,.rmvb,



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

		.rp,.rts,.scm,.smil,.smv,.spl,.svi,.vfw,.vid,.viewlet,.viv,.vivo,.vob,.vro,.wm,.wmd,.wmx,.vwx
VIDEO.ASF	Media file developed by Microsoft; can be compressed using a variety of codecs	.asf
VIDEO.ASX	Shortcut to an audio/video file (typically in .ASF format); references a URL for the location of the actual media	.asx
VIDEO.AVI	High-quality video format; uses less compression than similar formats such as .mpeg and .mov	.avi
VIDEO.MOV	Common multimedia format (mainly used for video); cross-platform, but developed by Apple	.mov
VIDEO.MP4	Video clip or movie that uses MPEG-4 compression	.mp4
OTHERS		
CLASS	Java Bytecode	.class
DLL	A dynamic link library (DLL) is a collection of small programs, any of which can be called when needed by a larger program that is running in the computer. The small program that lets the larger program communicate with a specific device such as a printer or scanner is often packaged as a DLL program (usually referred to as a DLL file).	.dll
EXE	A program file. Virtually all programs that run under Windows or DOS are in the .EXE format.	.exe
KNOWN_OS	File belongs to a collection of digital signatures of known, traceable software applications. For reference see http://www.nsr1.nist.gov	Various



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

UNKNOWN	Unable to identify type of the file.	Various
XSL	XSL (Extensible Stylesheet Language), formerly called Extensible Style Language, is a language for creating a style sheet that describes how data sent over the Web using the Extensible Markup Language (XML) is to be presented to the user.	.xsl
MSACCESS	Microsoft Access Database	.mdb
	Microsoft Active Server Page	.asp
	Macintosh Binary File	.bin
	Cascading style sheet for web pages	.css
	Comma Separated Values; sometimes used as a spreadsheet format or for import/export of data	.csv
	dBASE	.dbf
	MS Word Template format	.dot
	Autocad Drawing	.dwg, .dxf
	Macintosh BinHex archive	.hqx
	A compressed file archive	.lha, .lhz
	Lotus WordPro	.lwp
	Microsoft Access queries and reports	.maq, .mar
	MacWrite	.mw
	Open Office Spreadsheet	.ods, .sxc
	Open Office Documents	.odt, .ott, .sxw



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

	Aportis (Palm) document	.pdb
	Postscript files	.ps, .eps
	Microsoft Publisher	.pub
	Pocket Excel	.pxl
	StarWriter	.sdw
	Standard Generalized Markup Language	.sgml
	Corel Presentations format	.shw
	Macintosh Stuffit archive	.sit, .sea
	Netscape Mail	.snm
	Shockwave Flash	.swf
	Corel Quattro Pro spreadsheet format	.wb3
	Lotus 1-2-3	.wk1 or any#, .wks, .wku
	MS Works spreadsheet	.wks
	MS Works document	.wps
	MS Write	.wri
	MS Excel spreadsheet template	.xlt
	Yahoo Messenger	.ymg



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

3. Electronic Data Cost Consideration Checklist

Single vs. Multiple Vendors	<ul style="list-style-type: none"> - Using different vendors for various aspects of the projects might decrease the cost. - More efficient to use vendors that can integrate with in-house applications to reduce loading and hosting costs.
General Fees	<ul style="list-style-type: none"> - Set up Costs <ul style="list-style-type: none"> o Fixed fee o Hourly - Consider a cap on fees - User license fees <ul style="list-style-type: none"> o (month vs. months) - License subscription fees
Training Costs	<ul style="list-style-type: none"> Live Training <ul style="list-style-type: none"> • Day rates vs. hourly rates • Travel costs • Vendor's waiver for some of training sessions Web based training <ul style="list-style-type: none"> • Session fee vs. hourly fee Free web-based training by vendor
Bundling	<ul style="list-style-type: none"> Identify what terminology vendor is using and what activities are included in price. <ul style="list-style-type: none"> • Are services bundled together for one fee, or assessed separately? • What is included? Be Specific
Data Loading and Processing Costs	<ul style="list-style-type: none"> Loading Costs <ul style="list-style-type: none"> • Native • TIFF/PDF • Other file types- Costs Assessed Other processing/loading issues Common for volume discounts. <ul style="list-style-type: none"> • Negotiate reduced rates as volume increases
Hosting Costs	<ul style="list-style-type: none"> Data hosting charged per volume vs. monthly basis. Negotiate
Best Practice to Reduce Hosting Cost	<ul style="list-style-type: none"> Reduce volume of data as much as possible <ul style="list-style-type: none"> • Custodians of Interest • Date Range • Deduplication methods <ul style="list-style-type: none"> o Horizontal



ACC Annual Meeting
October 29-31, 2007 Chicago, IL
Data Collection, File Types, and Costs Checklists

	<ul style="list-style-type: none"> o Vertical o Across data set
Hosting Shared Databases	Review sources of data prior to loading to exclude duplicate data
Costs of Archiving Data	Use vendor to hold unique copy of document
Costs of Production	Ability of vendor to archive data for a dormant case for a reduced hosting fee Cost for re-activating database
Professional Services Fees	Dependent on format Reduced rate to reproduce earlier production Additional Costs Paper production Costs (per page)
Cost Savings in Professional Service Fees	Charged per hour. <ul style="list-style-type: none"> • Unpredictable costs • Dedicated Resources • Pre-pay for a bulk set of hours (bulk purchase for lower rate)
Termination Costs	Other persons (in-house counsel, outside counsel, litigation support personnel or consultant) can perform certain services. Sliding fee/level of experience Blocks of hours for reduced fee Build in level of support, or hours of support, into loading charges.
Service Level Agreements	Costs for closing down database or transferring the data and work product to another vendor
Other Cost Considerations	Service Level Agreements provide for certain level of service guaranteed by financial penalties. See Sedona RFP document http://www.thesedonaconference.org/content/miscFiles/RFP_plus_july05ver.pdf

FRCP Changes Summary

Effective Date – December 1, 2006

<http://www.uscourts.gov/rules/newrules6.html#cv0805>
http://www.uscourts.gov/rules/EDiscovery_w_Notes.pdf

[Civil Rule 5.1](#) (Constitutional Challenge to a Statute - Notices, Certification, and Intervention) and [24](#) (Intervention)(sets forth the process when a party draws into question the constitutionality of a federal or state statute)* (PDF)

[Civil Rule 9](#) (Pleading Special Matters) (conforming amendment pertaining to Supplemental Rule G) (PDF)

[Civil Rule 14](#) (Third-Party Practice) (conforming amendment pertaining to Supplemental Rule G) (PDF)

[Civil Rule 16](#) (Pretrial Conferences; Scheduling; Management) (establishes process for the parties and court to address early issues pertaining to the disclosure and discovery of electronic information) (PDF)

[Civil Rule 26](#) (General Provisions Governing Discovery; Duty of Disclosure) (requires parties to discuss during the discovery-planning conference issues relating to the disclosure and discovery of electronically stored information) (PDF)

[Civil Rule 33](#) (Interrogatories to Parties) (expressly provides that an answer to an interrogatory involving review of business records should involve a search of electronically stored information) (PDF)

[Civil Rule 34](#) (Production of Documents and Things and Entry Upon Land for Inspection and Other Purposes) (distinguishes between electronically stored information and "documents") (PDF)

[Civil Rule 37](#) (Failure to Make Disclosure or Cooperate in Discovery; Sanctions) (creates a "safe harbor" that protects a



ACC Annual Meeting
 October 29-31, 2007 Chicago, IL
 Data Collection, File Types, and Costs Checklists

4. Other Checklists	
Search Types	Basic Search Types <ul style="list-style-type: none"> • Phrase • Boolean • Proximity • Directed Proximity • Phonic • Stemming • Numeric • Macro • Wildcard • Fuzzy • Concept/Synonym/Thesaurus • Combining Search Types Advanced Search/Information Retrieval Methods <ul style="list-style-type: none"> • Ontologies • Clustering (Concept Searching) • Combination of these methodologies (Context searching)
Report Types	<ul style="list-style-type: none"> • Exception Reporting: Report what files and locations could not be indexed • Network Report: Report what locations are found, what are currently attached, what devices and locations are no longer attached • Search Report: Report the results of a query including metadata for found documents and the location of "hits" within the document. Reports can be generated in the background. • Location of Interest: Show all locations and files that have been selected as having documents of interest. Show the approximate size and contents of each location • Audit Report: List of all queries that were run and who ran them • Privilege Log – all Priv markings, made by whom, when, and why • Redaction Log – all redactions, made by whom, when, and why • Foreign Language Report – a list of all foreign languages found in the data set • Analytics Reports – various reports about the data and what is found in it from an analytical standpoint

ACC Annual Meeting 2007

Session 504

party from sanctions for failing to provide electronically stored information lost because of the routine operation of the party's computer system) (PDF)

[Civil Rule 45](#) (Subpoena) (technical amendments that conform to other proposed amendments regarding discovery of electronically stored information) (PDF)

[Civil Rule 50](#) (Judgment as a Matter of Law in Jury Trials; Alternative Motion for New Trial; Conditional Rulings) (permits renewal after trial of any Rule 50(a) motion, deleting the requirement that a motion made before the close of all the evidence be renewed at the close of all the evidence) (PDF)

[Civil Rule 65.1](#) (Security: Proceeding Against Sureties) (conforming amendment pertaining to Supplemental Rule G) (PDF)

[Form 35](#) (Report of Parties' Planning Meeting) (technical revision reflecting the proposed amendment to Civil Rule 26) (PDF)

[Supplemental Rule G](#) (Forfeiture Actions in Rem) (establishes comprehensive procedures governing in rem forfeiture actions) (PDF)

[Supplemental Rule A](#) (Scope of Rules), [Supplemental Rule C](#) (In Rem Actions; Special Provisions), [Supplemental Rule E](#) (Actions in rem and Quasi in Rem: General Provisions), and [Rule 26](#) (General Provisions Governing Discovery; Duty of Disclosure) (conforming amendments pertaining to proposed Supplemental Rule G) (PDF)

ACC Annual Meeting 2007

Session 504

Case Law Updates

Cenveo Corp. v. Slater, 2007 WL 442387 (E.D. Pa. Jan. 31, 2007)
Court Sets Out Imaging, Recovery and Disclosure Protocols for Imaging of Defendants' Computer Equipment by Plaintiff's Computer Forensics Expert (**Rule 26 Meet & Confer, Privilege**)

Doe v. Norwalk Community College, 2007 WL 2066497 (D. Conn. July 16, 2007)
Rule 37(f) Safe Harbor Provision Requires a Routine System in Place and Some Affirmative Action by Party to Prevent System from Destroying or Altering Information

Disability Rights Council of Greater Wash. v. Wash. Metro. Area Transit Auth., 2007 WL 1585452 (D.D.C. June 1, 2007)
Magistrate Judge Facciola Orders Production of Email from Backup Tapes in Light of Party's Failure to Suspend Email System's Automatic Deletion Feature (Rule 37)

Guy Chem. Co. v. Romaco AG, 2007 WL 1521468 (N.D. Ind. May 22, 2007)
Search and Production Costs of \$7,200 Render Non-Party's Ordinary ESI "Not Reasonably Accessible Due to Undue Burden" (**Rule 45** Not Reasonably Accessible, Cost Shifting)

Kingsway Fin. Servs., Inc. v. Pricewaterhouse-Coopers LLP, 2007 WL 1837133 (S.D.N.Y. June 27, 2007)
Attachment of Protected Email to Service Copies of Motion Requesting its Return Constitutes Deliberate Disclosure to Adversaries, Waiving Any **Privilege**

O'Bar v. Lowe's Home Centers, Inc., 2007 WL 1299180 (W.D.N.C. May 2, 2007)
Court Sets Out Detailed Guidelines for Discovery of ESI, Adapting "Suggested Protocol" of the District of Maryland (**Rule 26 Meet & Confer**)

Scotts Co. LLC v. Liberty Mut. Ins. Co., 2007 WL 1723509 (S.D. Ohio June 12, 2007)
2006 E-Discovery Amendments Do Not Require Forensic Computer Search as a Matter of Course; Court Orders Parties to Meet and Confer on Certain Issues (**Rule 34, 26**)

Williams v. Taser Int'l, Inc., 2007 WL 1630875 (N.D. Ga. June 4, 2007)
Court Selects Search Terms and Sets Out Detailed Electronic Discovery Protocol in Light of Parties' **Inability to Collaborate (Rule 26 Meet & Confer)**

Select cases cited from www.ediscoverylaw.com



Agenda

- FRCP Changes Overview/Update
- Preservation of Data
- E-Discovery Procedures
 - Assembling a Team
 - Defining a Repeatable Process
- Preparing for a Meet & Confer
- Review and Production of the Data
- Q&A

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



FRCP 1

● Rule 1. Scope and Purpose of Rules

These rules... shall be construed and administered to secure the just, speedy, and inexpensive determination of every action.

- ... two out of three ain't bad

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



FRCP Changes

- Rule 16 - Scheduling
- Rule 26 - Meet & Confer
- Rule 33 - Access to ESI
- Rule 34 - Sampling
- Rule 37 - Safe harbor?
- Rule 45 - Ordinarily Maintained, Not Reasonably Accessible

ACC's 2007 Annual Meeting:

Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Follows Stages of Civil Litigation

- Commencement (FRCP 3 – 6)
- Pleadings (FRCP 7-15)
- **Pre-Trial / Scheduling Conference (FRCP 16)**
- **Discovery (FRCP 26-37)**
 - **Of documents, including electronic data**
 - **Depositions**
- Trial (FRCP 38-53)
- Post-trial (FRCP 54 – end)

ACC's 2007 Annual Meeting:

Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Case Law Update for FRCP

- Rule 37(f) Safe Harbor
Doe v. Norwalk Community College, 2007 WL 2066497 (D. Conn. July 16, 2007)
- Rule 26 Meet & Confer
Williams v. Taser Int'l, Inc., 2007 WL 1630875 (N.D. Ga. June 4, 2007)
- See Handout for More Cases

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Part I - Preservation

Evan Stolove
Associate General Counsel
Fannie Mae

ACC's 2007 Annual Meeting: Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Duty to Preserve

- Sources of duty
 - Statutory
 - Common Law

- Amendments to FRCP largely quiet as to timing and scope of duty

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



When the Duty to Preserve Arises

- Judgment Call
 - Establish Procedure
 - Investigate Facts and Circumstances
 - Consider Prior Experiences
 - Experienced Decision Maker

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Guidance on Issuing Hold Notices

- Issuing a Proper “Litigation Hold Notice” is paramount
- Arguably Best Guidance is Still the *Zubulake* line of cases: *Zubulake IV*, 220 F.R.D. 212, and *Zubulake V*, 220 F.R.D. 422

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Before Issuing The Hold Notice

- Review Facts/Pleadings
- Identify Key Players/Business Areas
- Speak with Key Players

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Before Issuing Hold Notice (cont.)

- Determine How to Handle:
 - Preserve in Place or Collect?
 - Existing Disaster Recovery Backup Tapes Scheduled to Be Recycled?
 - Dynamic and Disappearing Data?

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Hold Notice Recipients

- End-users
- IT Staff
- Records Managers
- Outside Contractors/Vendors

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Contents of Notice

- Matter
- Obligation to Preserve
- What to Preserve
- Suspend Deletion Practices
- Time Period
- Contact

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Contents of Notice for IT Staff

- Advisory Committee Note to FRCP 34(a)
“The wide variety of computer systems currently in use, and the rapidity of technological change, counsel against a limited or precise definition of electronically stored information. Rule 34(a)(1) is expansive and includes **any type of data that is stored electronically.**”

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Contents of Notice for IT Staff

- Potentially Overlooked Sources:
 - Legacy and Retired Media
 - Audio Recordings, Including Voicemail
 - Pagers, Cell Phones and PDAs
 - Instant Messaging
 - Removable Media
 - Facsimile and Copy Machines

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Contents of Notice for IT Staff

- “When a party is under a duty to preserve information because of pending or reasonably anticipated litigation, **intervention in the routine operation of an information system is one aspect of what is often called a ‘litigation hold.’**”
-- Advisory Committee Note to FRCP 37

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Contents of Notice for IT Staff

- Interruption of Routine Processes:
 - Recycling of Backup Tapes
 - Email
 - Software and Hardware Upgrades
 - Virus Scanning and Defragmenting
 - Wiping Media Upon Termination

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Monitoring Compliance with Hold

- Documentation
- Chain of Custody Logs
- Issue Reminders and Supplements
- Alert New Employees

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Mitigation of Risk Before - Prepare

- Map Data and Take Out the Trash
- Create eDiscovery Response Team
- Develop Relationship with IT Personnel
- Policies & Procedures – Holds and Records Management
- Move Mobile Data Onto Servers
- Education and Training

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Mitigation of Risk – Once Obligation is Triggered

- Negotiate with Opposing Counsel (FRCP 26(f))
- Seek Protective Order (FRCP 26(c))
- Preserve Documents in Native File Format (FRCP 34)
- Move Quickly and Document Your Efforts

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Motivating the Business to Care

- Can You Say, “*Morgan Stanley?*”
- Adverse Publicity
- Potential Criminal Sanctions
- Regulator Scrutiny
- FRCP 37 Sanctions

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Good Faith, Reasonable Efforts

“Responding to e-discovery is like trying to catch a glass of water with your bare hands. You’re not going to get it all, and what you get will be a mess.” – Craig Ball

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Part II - E-Discovery Collection Procedures

David Wetmore
Risk Advisory Services
Ernst & Young

ACC's 2007 Annual Meeting: Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Assembling an E-Discovery Team

- Include Team Members with Cross-Functional Skills in the Process, Bringing Specific Priorities and Responsibilities to their involvement with identifiable roles
- Identify effective, flexible, experienced candidates
- Include inside and outside expertise on legal and technical issues
- Ensure right mix of expertise for each project, each phase and avoid weak links
- Understand that the team mix and/or roles may vary from project to project

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



The Team

- Project Team v. Legal Team
- Essential Components
 - Inside Counsel
 - Lead Attorney - outside counsel
 - Corporate Stakeholder
 - IT Representative
 - Project Leader

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Other Players

- Others - Corporate
 - Records Management
 - Compliance, Security, Risk Management
 - HR
 - Finance
 - Accounting
 - Operations
- Others - Beyond Corporate
 - Legal team - core group of associates, paralegals, in-house expertise
 - Litigation support expertise - vendors, in-house, consultants, outside counsel
 - Forensic/ data collection expertise
 - Experts - testifying to process

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Defining Roles

- In-house Counsel
 - Time and Focus
 - Custodian for the Processes
 - Assures that Legal Judgment Directs the Project

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Defining Roles

- Lead Attorney - Outside Counsel
 - Must Know at Least a Little About Every Step of the Process
 - Understand Overall Goals
 - With Other Corporate Stakeholders
 - Analyze Spending & ROI
 - Identify Gaps and/or Weak Links
 - Evaluate Process
 - Suggest Process Improvements

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Defining Roles

- Corporate Stakeholder
 - Risk Management - Representing Fiscal Interests
 - Analyze the Spending and ROI
 - Identify Gaps and/or Weak Links
 - Evaluate Progress
 - Suggest Process Improvements

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Defining Roles

- IT Representative(s)
 - Thorough Understanding of the Corporate IT Infrastructure
 - Ensure Familiarity with Actual Practices, not just Process and Procedures
 - Manage Internal IT Resources Essential to the Process
 - Leverage to Locate and Identify Data, but Utilize Those More Qualified and Experienced to Gather, Process, and Review Data

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Defining Roles

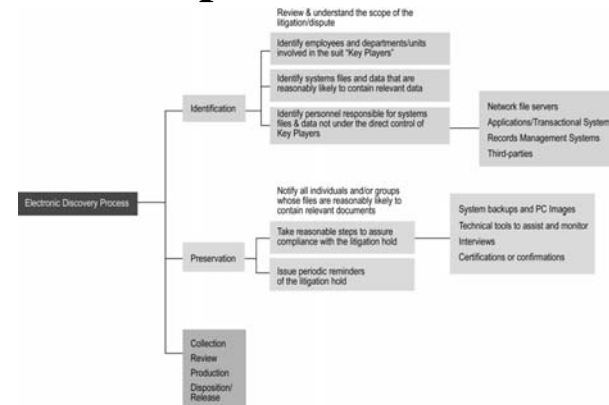
- **Project Leader**
 - Involved in Every Step of the Process
 - Communications Between Legal Team and Project Team go Through Project Manager Exclusively
 - Serves as a Funnel for ALL Decision Making, Process Changes, etc.
 - Manages the Project Team, NOT the Legal Team

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Define a Repeatable Process



ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Communication and Coordination

- Have a Plan (and a Plan B) and Communicate it to ALL Involved
 - Discovery Plan
 - Project Plan
- Have Kick-off Meeting, Face to Face, if Possible - Team Members See Broader Context and Where They Fit In
- Designate Those with Authority to Make Decisions - Both *Process* and *Budget* Related
- Utilize Project Leader as Funnel

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Document Every Step

- Scope and Criteria - Communicate Modifications as Project Progresses
- Create an Audit Trail for Changes
- Utilize Forms for Tracking (Chain of Custody, Status, Disposition, etc.); Make Sure ALL Team Members Have *Access and Use It*
- Designate Shared Repository - Extranet, Public Folder, DMS, etc. - Include ALL Documentation

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Part III - Preparing for a Meet & Confer

Anne Milem
Associate General Counsel
Sallie Mae



Preparing for a Meet & Confer

- Rule 26(f)
- Locating Data
 - Data Map
 - Is Data Reasonably Accessible?
- 30(b)(6) Witness
- Review & Production Scope, Timeline
- Production Format



Highlights of Rule 26(f)

- Discuss “any issues relating to **preserving discoverable information**, and . . . develop a proposed discovery plan”
- Reach Agreement as to Hold Obligations During the Litigation

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Confer About the Following:

- The Subjects on which Discovery May Be Needed
- When Discovery Should be Completed
- Conduct Discovery in Phases?
- Any Issues Related to Disclosure or Discovery of Electronically Stored Information
- Claims of Privilege After Production

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Locating Data

- Assembling a Data Map
- Identify and Understand Your Company's Information and Computer Systems
- Handout/Checklist

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Is Data Reasonably Accessible?

- Rule 26(b)(2)(B)
- Need NOT Produce ESI from Sources that You Identify as “not reasonably accessible because of undue burden or cost”
- You Must Identify Them and Continue to Preserve Them

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Is Data Reasonably Accessible?

What is “reasonably accessible” will change over time as technology continues to change.

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Is Data Reasonably Accessible?

- The Advisory Committee did not define the technological features that may affect the burdens and costs of accessing ESI
- Identified 3 types of information systems:
 - Designed to provide “ready access” to ESI used in regularly ongoing activities
 - Designed to provide “ready access” to ESI that is NOT regularly used
 - Designed to retain information that is accessible only by incurring substantial burdens or costs

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Inaccessible Data

- Party Withholding ESI Bears Burden to Show It's Not Reasonably Accessible
- ESI that is Not Reasonably Accessible
 - Still be Subject to Discovery
 - Upon Showing of "good cause"

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



30(b)(6) Witness

- Identify, Involve and Prep Rule 30(b)(6) Witness
- Advisory Committee Notes Suggest the "identification of, and early discovery from, individuals with special knowledge of a party's computer systems may be helpful"

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Review & Production Scope

- Time Period of Data to Review
- Custodian List
- Topic List (Not Limited to Keywords)

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Review & Production Timeline

- Rolling Production
- By Custodian or Topic?

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Production Format

- Discuss Upfront
 - Native
 - Converted
- Hard to Access Items
 - Databases (Take a Peek)
 - Proprietary Info (Escorted Review, Clean Room)

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Part IV - Review & Production

Sonya Sigler
General Counsel
Cataphora

ACC's 2007 Annual Meeting: Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Purpose of review

- Relevant Documents Need to be Selected
- Not ALL Documents are Responsive and/or Relevant – estimated 85% are NOT
- Identify Privileged and Confidential Documents
- Understand the Issues of the Case
- Locate Smoking Guns, if any

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



“E” in Electronic-Review

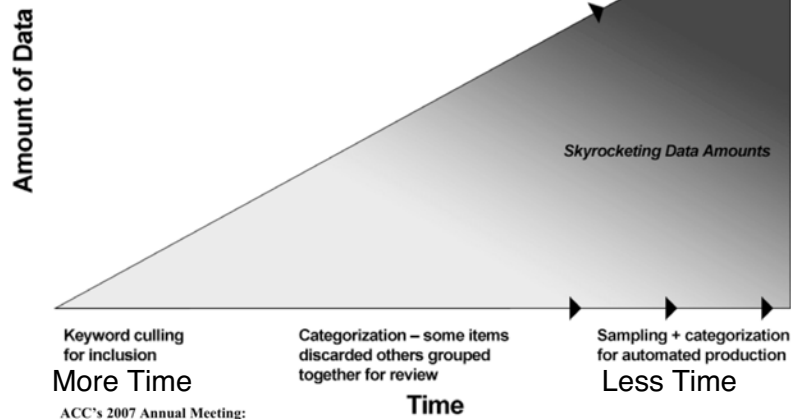
- Electronic Documents are Treated Differently from Paper
- Many Capabilities for Electronic Documents
 - Searching
 - Sorting
 - De-duplication
 - Near Duplicate Detection
- Take Advantage of Tools Available
- Review in Native Format vs. Imaged Conversion (pdf, tiff, html, etc.)

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Inevitable Progression



ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Review Methodology - Upfront

- Identify Issues in the Case
 - Formulate Queries and Approaches for Finding Responsive Documents
 - Formulate Relevancy and Responsiveness Guidelines
- Identify Primary Participants
- Select or Triage Documents for Review

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Review Tools for Relevancy Assessment

- Keyword Searches, Culling
 - Slices of Data are Reviewed
- Categorization of Data
 - Entire Dataset is Categorized
 - Review Targeted Data
- Automated Review
 - Categorization of Dataset
 - Random Sampling (Statistically Significant)

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Keyword Culling

Pro

- Word Stemming
 - Hous* - house, housemate, household
- Easy to use/explain/agree
- Familiar
- Fast results



Con

- Over-inclusive
- Disambiguate
- Under-inclusive
- Word must be present
- Hard to craft
- Ineffective with short messages, IMs

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Categorization of Data for Review

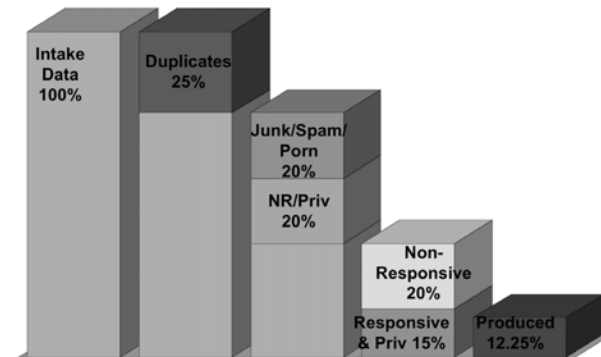
- Categorize Entire Data Set
 - Spam/Porn/System Files
 - Personal/Private Data
 - Non-relevant Business Data
- Business Data
 - Relevancy Assessment by Topic
 - Privilege Review
- Keyword, Topic Analysis - Overlap, Holes

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Getting to a Manageable Review Set



These figures vary based upon the data set received

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Categorization Methods

- Linguistic Methods(Word Based)
 - Keyword (Culling Method)
 - Ontologies
- Statistical Methods (#s based)
 - Topic Clustering
 - Statistical Similarity
 - Counting #s of words, appearance together
 - Latent Semantic Indexing

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Automated Review



ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Review risks

- Failure to find responsive documents
- Failure to recognize responsive documents
- Failure to recognize privileged documents
- Inconsistent treatment of documents (e.g., duplicates)
- Failure to complete project in a timely manner

- Sophisticated Tools
 - Understand What They Do and Don't Do Well
 - Inform Yourself, Speak to References, Consultants

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Evaluating the Review Process

- Accuracy of Review Depends on Many Factors
 - Human Factors
 - Quality of Tools
 - Query Quality
 - Processing accuracy
- Post Mortem - Must Evaluate ALL Parts of the Process
 - Sampling Techniques are Available to Assess Quality of Project

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



Possible Production Formats

- Paper (if the other side asks for it this way, be happy to oblige. It is the least useful format in which to receive a production.)
- Paper-like (TIFF or PDF images)
 - TIFF images without any searchable data at all are increasingly unacceptable.
- Native Files
- Hosted “production” areas of the producing party’s web repository.

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



FRCP 34 [amended, continued]

- **Rule 34. Production of Documents, Electronically Stored Information.**
 - (b) PROCEDURE. ... the request may specify the form in which the electronically stored information is to be produced....
- ...[the responding party may object] to the requested form, stating the reasons, and the form it intends to use [instead].
 - (ii) if a request does not specify, the responding party must produce in the form it is *ordinarily maintained* or in an *electronically searchable form*.

ACC's 2007 Annual Meeting:
Enjoying the Ride on the Track to Success

October 29-31, Hyatt Regency Chicago



FRCP Changes for Production

- FRCP Default - Rule 34
 - Requesting Party Specifies or Parties Agree
 - If Nothing Specified or Agreed to, then
 - Ordinarily Maintained
 - Reasonably Usable
 - Not More than One Format
- Native
 - Meta-data and embedded data
 - Unique ID
- Shared Database