eDiscovery – Search and Information Retrieval Strategies

Broadcast – November 14, 2007

© by the respective presenters and their firms

Jim Daley

Tom Stevens

Rob Brownstone

Agenda

overview

- I. Setting the Stage
- II. Search and Retrieval Plans
- III. Implementation Best Practices

- Q and A

© 2007

I. Setting the Stage

- A. Information-Technologists are from Mars and Attorneys are from Venus*
  - A Discovery Revolution
  - Old Methods No Longer Work
  - The Myth of Human Review
  - The Limits of "Key Words"
- Sedona Conference Commentary on Search & Information Retrieval Methods (2007) – <www.thesedonaconference.org>

* DOAR, *Top Ten E-Discovery Challenges & Solutions* (8/6/03) <www.doar.com/apps/uploads/literature11_The_Top_10_Challenges_and_Solution.pdf>

© 2007

I(A). Sedona Conference

- Nonprofit 501(c)(3) research and educational institute dedicated to the advanced study of law and policy in the areas of antitrust, intellectual property rights and complex litigation
- "Transcending" traditional continuing legal education courses to combine experience with dialogue, outside of the adversarial system, to contribute to the reasoned and just development of law and policy
- Supporting think tanks (Working Groups) operating under the umbrella of The Sedona Conference, to develop principles, guidelines and best practices in targeted areas "ripe" for development

© 2007

I(A). Search / Retrieval Commentary

2007 THE SEDONA CONFERENCE JOURNAL 189

THE SEDONA CONFERENCE® BEST PRACTICES COMMENTARY ON THE USE OF SEARCH AND INFORMATION RETRIEVAL METHODS IN E-DISCOVERY

*A Project of The Sedona Conference® Working Group on Best Practices for Document Retention and Production (WG1), Search & Retrieval Sciences Special Project Team**

August 2007 Public Comment Version

Editor-in-Chief:
Jason R. Baron

Executive Editors:
Richard G. Braman
Kenneth J. Withers

Senior Editors:
Thomas Y. Allman
M. James Daley
George L. Paul

© 2007

Sedona Practice Point 1

- **In many settings involving electronically stored information, reliance solely on a manual search process for the purpose of finding responsive documents may be infeasible or unwarranted. In such cases, the use of automated search methods should be viewed as reasonable, valuable, and even necessary.**

© 2007

Sedona Practice Point 2

- **Success in using any automated search method or technology will be enhanced by a well thought out process with substantial human input on the front end.**

© 2007

Sedona Practice Point 3

- **The choice of a specific search and retrieval method will be highly dependent on the specific legal context in which it is to be employed.**

© 2007

Sedona Practice Point 4

- **Parties should perform due diligence in choosing a particular information retrieval product or service from a vendor.**

© 2007

Sedona Practice Point 5

- **The use of search and information retrieval tools does not guarantee that all responsive documents will be identified in large data collections, due to characteristics of human language. Moreover, differing search methods may produce differing results, subject to a measure of statistical variation inherent in the science of information retrieval.**

© 2007

Sedona Practice Point 6

- **Parties should make a good faith attempt to collaborate on the use of particular search and information retrieval methods, tools and protocols (including as to keywords, concepts, and other types of search parameters).**

© 2007

Sedona Practice Point 7

- **Parties should expect that their choice of search methodology will need to be explained, either formally or informally, in subsequent legal contexts (including in depositions, evidentiary proceedings, and trials).**

© 2007

Sedona Practice Point 8

- **Parties and the courts should be alert to new and evolving search and information retrieval methods.**

© 2007

I. B. Obstacles to Information Retrieval

- Ambiguity and "Elasticity" of Language
- Need for a Suitable Search Methodology
  - One Size does not Fit All
  - *Caveat Emptor* – Let the Buyer Beware
- Current Limitations of Search Technologies
  - No "ready for prime time" Artificial Intelligence
  - Need for Human Process

© 2007

I. C. Legal Requirements

- Federal Statutes; FRCP E-Discovery Amendments; and Common Law
- 1. Duty to Preserve vs. Duty to Produce
- 2. Good Faith Discovery Efforts
  - Reasonable – Perfection as the Enemy of Good
  - Consistently Applied
  - Rapid Development of Common Law (Case Law)

© 2007

I(C). 3. Searching Cooperation?

- Case law regarding lack of cooperation between litigants
- HAVE to Winnow/Cull in some way . . .
- Strategy – zealous representation vs. obstruction
- Credibility with Court
- Efficiency
- *See generally* Robert D. Brownstone, *Collaborative Navigation of the Stormy e-Discovery Seas*, 10 Rich. J.L. & Tech. 53 (2004) <http://law.richmond.edu/jolt/v10i5/article53.pdf> . . . cited at .pdf p. 14 of above-cited *Sedona Best Practices Commentary*

© 2007

I(C)(3). Searching Cooperation – Cautionary Decisions

- *Zurich Am. Ins. v. Ace Am. Reins.*, 2006 WL 3771090 (S.D.N.Y. Dec. 22, 2006) (prior handling of similar claims – discoverable but arguably undue burden):
  - "sophisticated reinsurer . . . operat[ing] multimillion dollar business . . . entitled to little sympathy for utilizing . . . opaque data storage system, particularly when, by the nature of its business, it can reasonably anticipate frequent litigation.
  - At the same time, the volume of data accumulated . . . makes a search of its entire database infeasible.
  - The parties shall therefore propose a protocol for sampling."

© 2007

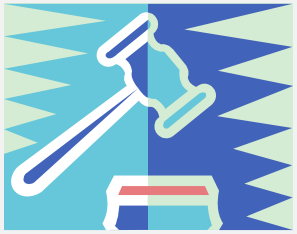
I(C)(3). Searching Cooperation?

- Impasse can lead to time/resources revisiting same scope-of-discovery issues . . . and to adverse order:
  - more searches not limited by relevancy objections
    - *Amsted ERISA Lit.*, 2002 WL 31844956 (N.D. Ill. 12/18/02)

      <http://Amsted-ND-Ill-12-18-02.notlong.com> (WESTLAW password needed)
  - production of Δ's senior executives' e-mails to π's consultant for searching based on agreed-upon terms
    - *Tulip Computers v. Dell*, 2002 U.S. Dist. LEXIS 7792 (D. Del. Apr. 30, 2002)
      <www.ded.uscourts.gov/RRM/Opinions/Apr2002/00-981.pdf>

© 2007

I(C)(3). Searching Cooperation?



- **Adverse Orders (*c't'd*):**
  - **responder to search additional custodian's e-mails**
    - ***McPeek v. Ashcroft* (D.D.C.) (LEXIS ID needed)**

      202 F.R.D. 31 (D.D.C. 2001) ("*McPeek I*") <http://McPeek-DDC-2001.notlong.com>

      212 F.R.D. 33 (D.D.C. 2003) ("*McPeek II*") <http://McPeek-DDC-2003.notlong.com>
  - **responder to use requestor's Boolean search terms; other search terms addable by mutual agreement only**
    - ***Medtronic Sofamor Danek v. Michelson*, 2003 WL 21468573 (W.D. Tenn. 5/13/03) <http://Medtronic-WDTenn-5-13-03.notlong.com>**
- ***See generally* <http://law.richmond.edu/jolt/v10i5/article53.pdf#page=18>**

© 2007

I(C)(3). Search Cooperation *(c't'd)* – Cautionary Decisions

- **Judges begin to factor in attorneys' prior representations that no additional responsive e-mails/files could even exist.**
  - ***See, e.g., Tulip*** **slip op. at 11 <http://Tulip-at-11.notlong.com>**
- **Recent decision: judge chastised both sides for inability to agree re: search parameters re: D's ESI. Court came up with its own protocol and imposed it on litigants**
  - ***Williams v. Taser Int'l*****, 2007 U.S. Dist. LEXIS 40280 (N.D. Ga. 6/4/07) <http://Williams-Taser-NDGa-6-4-07.notlong.com>**
- **Implicit finding of violations of ethics rules**

© 2007

II. Search/Retrieval Plan – A. Factors to Consider

- Do the issues in the case lend themselves to searching?
  - How complex are the issues?
  - How many relevant documents are there?
    - Use statistical sampling to determine "richness"
      - √ Must be random
      - √ Mathematical statistics is a very mature field
      - √ Provides targets for search results
      - √ Can assess appropriateness of culling/winnowing

© 2007

II(A). Search/Retrieval Plan – Factors to Consider

- **Heterogeneity of the universe of documents**
  - **Can all ESI be "normalized" and searched simultaneously?**
  - **Can or should the population be subsetted?**
    - **Generally speaking, homogeneity improves searching**
      - √ **At what point do you reach diminishing returns?**
  - **Text documents v. Non-text or No-text documents**

© 2007

II(A). Search/Retrieval Plan – Factors to Consider

- **Is there enough time to perform a particular search and retrieval method?**
  - **Advanced search methods typically require upfront "training" and iterative feedback for tuning**
- **Status of ESI in the matter**
  - **Is all ESI on hand or will the collections or productions be rolling over time?**
  - **How does this impact statistical sampling and validation?**

© 2007

II. B. Search/Retrieval Plan – Search and Retrieval Methods

- **High reliability culling:**
  - De-duplication: eliminate duplicate documents
    - Within source
    - Across sources
  - Date filters
  - System file filters
  - Other file type filters
  - Other processes that defensibly reduce document volume upstream of text culling

© 2007

II(B). Search/Retrieval Plan – Search and Retrieval Methods

- More aggressive culling:
  - How many documents are left? Do we really need to cull further?
  - Keyword and Boolean
    - Words and phrases in conjunction with “and,” “or” and “not” logic
  - Statistical sampling for validation
    - Recall
    - Precision

© 2007

II(B). Search/Retrieval Plan – Search and Retrieval Methods

- "Bleeding Edge" Alternate culling methods:
  - Bayesian Classifiers
    - Mathematical formulas that place values on words, their interrelationships, proximity and frequency
  - Fuzzy Search Models
  - Statistical Clustering
  - Semantic Representation
  - Thesauri, Taxonomies and Linguistic Ontologies
  - Analytics – Social Networking and Other Patterns

© 2007

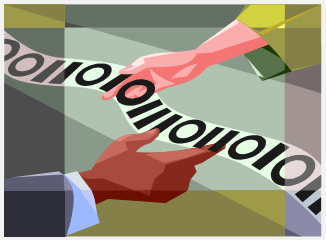
II. C. Search/Retrieval Plans – Court Treatment

- **De-Duplication, then Searches**
  - ***See In re CV Therapeutics Secs. Litig.*, 2006 U.S. Dist. LEXIS 63155 (N.D. Cal. 8/22/06) (*"CV II"*)**
    - **rejecting π's objections to use of de-duplication AND search terms**
    - **finding that use of these processes struck reasonable balance between needs of, and burdens on, parties**

<https://ecf.cand.uscourts.gov/cgi-bin/show_case_doc?391,14318,,,,1026,1>

© 2007

II(C). Search/Retrieval Plans – Court Treatment



- ***Disability Rights Council (DRC) v. WMATA*, 242 F.R.D. 139 (D.D.C. 6/1/07)** <https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2004cv0498-161>
  - **FACTS:**
    - **WMATA failed to stop e-mail system from deleting (automatically purging) all e-mails > 60 days old – 2-3 years after suit filed**
    - **π's requested restoration and search of back-up tapes**
    - **WMATA raised:**
      - √ **FRCP 37(f) safe harbor; and**
      - √ **FRCP 26(b)(2)(B) undue burden/cost**

© 2007

II(C). Search/Retrieval Plans – Court Treatment

California Lawyer 10/07

- **HELD:**
  - "Indefensible" lack of good faith under 37(f)
  - Overwhelming "good cause" under 26(b)(2)(B)
  - "I bring to the parties' attention recent scholarship that argues that concept searching, as opposed to keyword searching, is more efficient and more likely to produce the most comprehensive results."
- **ANY REASONING?**
  - Pays only lip service to GroupWise e-mail being odd type
  - Cites law review article but punts on search details once pertinent e-mails restored from back-up tapes

© 2007

II(C). Search/Retrieval Plans – Court Treatment

- *Treppel v. Biovail Corp.*, 233 F.R.D. 363 (S.D.N.Y. 2/6/06) <http://www.nysd.uscourts.gov/courtweb/pdf/D02NYSC/06-00983.PDF>
  - FACTS:
    - Δ refused to produce any ESI because π would not get involved in search-terms selection or provide guidance as to which of Δ's employees' files should be searched
  - HELD:
    - when faced with π's "recalcitrance" and absence of an agreement, Δ should have proceeded unilaterally, using own list (per Sedona Principle 11, now at <http://www.thesedonaconference.org/content/miscFiles/7_05TSP.pdf>)
    - Δ to: promptly conduct "interim step" of a search; produce responsive results in π's requested format (native); and explain the search protocol it used

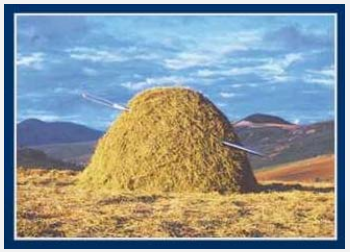
© 2007

II(C). Search/Retrieval Plans – Court Treatment

- *In re Seroquel Prods. Liab. Litig.*, 2007 WL 219989 (M.D. Fla. 1/26/07) ("purposefully sluggish" Δ searched "in secret;" "key word searching . . . must be a cooperative and informed process; "[c]ommon sense dictates . . . sampling and other [QA] techniques" required "to meet requirements of completeness")
- *Flexsys Americas, LP v. Kuhmo Tire U.S.A., Inc.*, 2006 WL 3526794 (N.D. Ohio Dec. 6, 2006) <http://Flexsys-Kuhmo-NDO-12-6-06.notlong.com> (upon finding parties had not come to agreement regarding scope of eDiscovery according to new FRCP guidelines, resolving discovery dispute by ordering π to produce all electronic documents from ten individuals to be chosen by Δ)

© 2007

III. Implementation Best Practices



- **A. Relevance needs analysis (defining target universe)**
  - **If concept-searching worked as well in practice as in theory, if lawyers understood it and if there were enough time to do it right . . . .**
  - **Huge potential value at beginning of case in finding real key descriptive words and names**
  - **That way technology would help define word lists rather than lawyers insisting on certain terms**

© 2007

III(A). Implementation Tips – Relevance – Smart Culling

Most Innovative Use of Technology by a Law Firm
LTN Law Technology News

- "File Identification Narrowed by Definition"
- > 95% of broad collection may be non-responsive
- If so, only < 5% responsive or privileged
- So, Reduce set before attorney review
  - ↓ number of custodians
  - limit time frame
  - convert e-mail to de-duplicate-able format
  - aggressively de-duplicate and filter

© 2007

III(A). Analyzing Relevance – Key Points/Takeaways

- Involve tech team with legal/lit team from Day 1
- Connect tech team with client's IT ASAP
- Plan for enough processing – and indexing – time for first data set to be ready for attorney review
- Plan for enough days after review is complete before production is ready to go to other side (dependent on terms of original stip/agreement)
- Reviewer training important
- Legal team responsible for QC of productions

© 2007

III. Implementation Tips
B. Legal/IT Iterative Process

- **Collaboration/dialogue as to key decisions, *e.g.*,:**
  - **1. Culling ESI by Date Range**
  - **2. De-Duplication**
    - **By custodian?**
    - **Across collection?**
  - **3. Picking Search Terms**
    - **picking appropriate search terms; and**
    - **testing if they actually return expected results**
    - **tech team can offer help in telling you what *not* to search for but rarely has insight into what you should search for**
  - **4. Integrity of parent/child relationships**

© 2007

III(B). Implementation Tips – Dialogue re: Choices *(c't'd)*

The	my	are
d *	when	so
of	there	*
his	is	it

- **Be careful with "noise words"**
- **In addition to the ones in above screenshot, some "why bother?" search terms include:**
  - **IBM, Dell, etc.**

    Craig Ball, *Unlocking Keywords*, LTN (1/1/07)
    <http://www.lawtechnews-digital.com/lawtechnews/sample/?pg=56>
  - **AI (or any other two-letter term)**
  - **Microsoft, Sun, etc.**
- **Several simple searches often better than one complex Boolean search**

© 2007

III. C. Just one method may not be sufficient

- **Concept-searching**
  - Great in theory; not necessarily workable
  - Most purported concept-search tools do not work once hit certain volume of data
  - Only options may be very costly
- **Word-searching**
  - For better and worse, is state of art
  - Short list gets most of what you need. Long lists – especially with Boolean operators – are often counterproductive
  - BE VERY CAREFUL WITH PROPER NAMES
  - Boolean operators vary tool to tool . . . * . . . ~ . . . ? . . .
  - QC the list itself!

© 2007

III. D. Educating the judge, who may not really "get it"

- As noted above re: *DRC*, judges arguably starting to endorse "concept searching" or AI without fully understanding the issues
- Sampling the way it will likely go . . . . though will be disputes about it
- Unless someone not only "gets" sampling techniques but also deeply understands the nature of the documents, difficult to convince judge to conclude "best" way to search
- *"I know not with what weapons World War III will be fought, but World War IV will be fought with sticks and stones."* – Albert Einstein

© 2007

Conclusion/ Questions

- Q and A

- M. James Daley

Redgrave, Daley et al.

<jdaley@rdrw.com>

- Tom Stevens

ACT Litigation Services

<tomstevens@actlit.com>

- Robert D. Brownstone

Fenwick & West LLP

<rbrownstone@fenwick.com>

© 2007