

## Metadata

(and other things that go bump in the night)



Catherine Sanders Reach, Director  
American Bar Association  
Legal Technology Resource Center  
July 27, 2006

# What is Metadata?

- “Data about Data”
- Useful for finding documents, reviewing, and sharing
- Invisible, but not too hard to find
- Reveals information about electronic documents beyond the printable text



# Metadata: Why is it such a big deal now?

- Widely known since 1998
- Mass movement to MS Word by attorneys
- Advent of efilng
- Advent of electronic discovery
- High profile examples

# Metadata: Establish Best Practices

- Get the facts
- Understand the implications
- Establish firm policies to protect yourself and your clients



# Metadata: Examples in the News

- October 2000 – WSJ reports critical “anonymous” email edited in MS Word to a candidate for the Senate reveals author – chief-of-staff of the opposing party
- February 2003 – 10 Downing Street dossier regarding Iraqi security found to be plagiarized from US researcher

# Metadata: More Examples in the News

- March 2004 – SCO Group, seller of Linux and Unix, files suit against DaimlerChrysler and Autozone. Metadata reveals the attorneys had originally identified Bank of America, but they were later removed from the pleading.
- December 2005 – Justice Department reveals social security numbers
- March 2006 – Google inadvertently reveals financial projections and info about projects in the works
- And more...

# Metadata: Ethics

- Unsolicited Receipt of Privileged or Confidential Materials
  - ABA Formal Opinion 92-368
  - Superseded by Rule 4.4 and Withdrawn by Formal Opinion 05-437
- Inadvertent Disclosure of Confidential Materials
  - ABA Formal Opinion 94-382
  - Withdrawn by Formal Opinion 06-440
- Use of Computer Software to Surreptitiously Examine and trace email and other electronic documents
  - NYSBA Opinion 749
  - NYSBA Opinion 782
- Florida weighs in

# MS Word – What Can Be Viewed

- Author name/ initials
- Company/organization name
- Subject, file type, location
- Date created/modified/last accessed
- Number of revisions/versions
- Previous document authors
- Total editing time



# MS Word – What Can Be Viewed (continued)

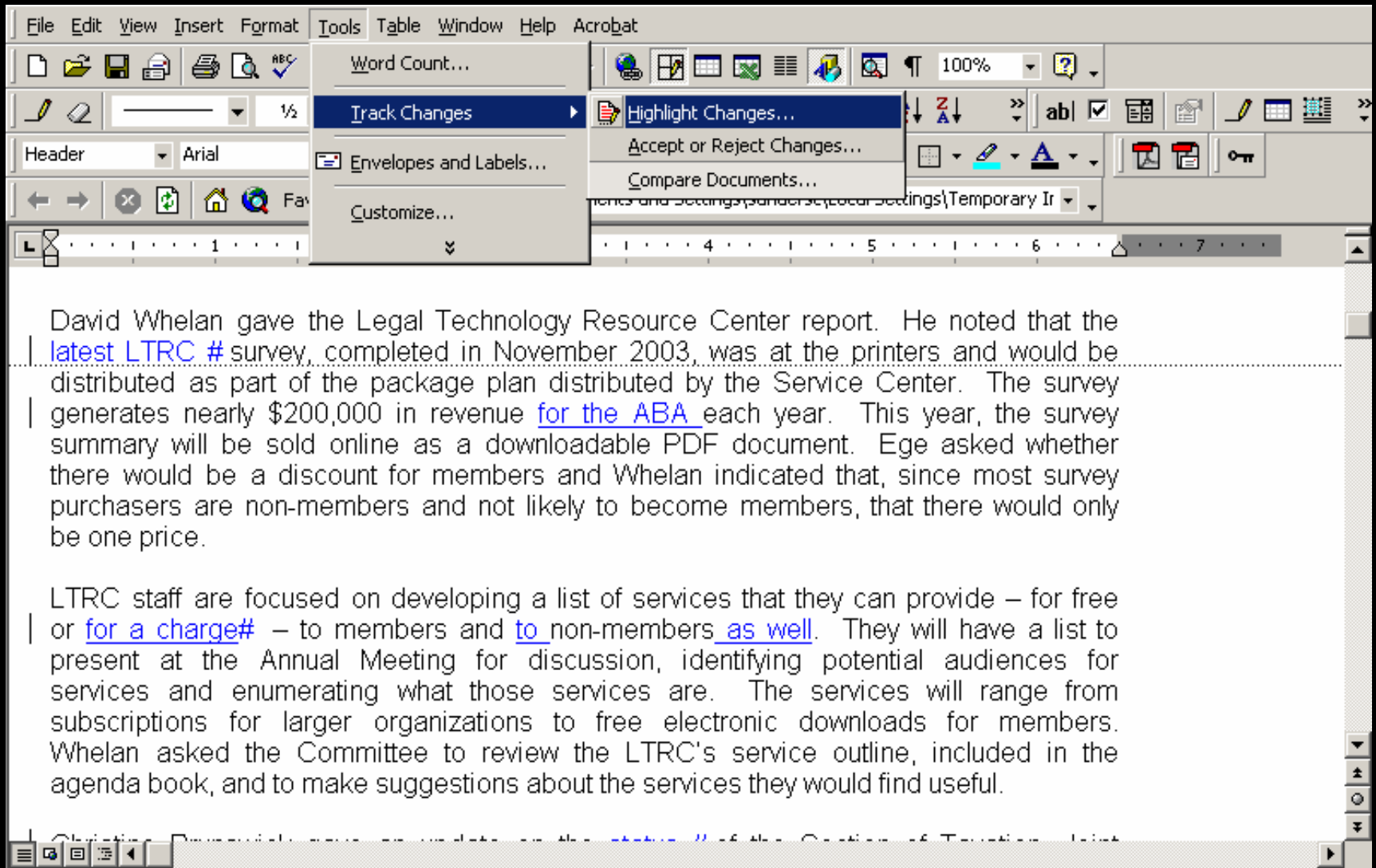
- Template information
- Hidden text (formatting)
- Comments
- OLE objects
- Macros
- Hyperlinks
- Redlining/changes

# Metadata: How Do I See It?



- Enable “Track Changes”
- View Comments
- File – Properties and look behind the tabs
- Text editors
- Save as HTML and view source
- Third party software

# Metadata: Track Changes



# Metadata: View Comments

The screenshot shows the Microsoft Word 2003 interface. The 'View' menu is open, with 'Comments' selected. The document content is as follows:

## Law Office 2002<sup>[chs7]</sup>: Technology and Demographics

The American Bar Foundation performed research in 1995 on the demographics of lawyers, by firm size. That research found that nearly half of all lawyers were sole practitioners, with another third practicing in law firms with 2 to 40 lawyers.<sup>12</sup> Of the # estimated 1 million active attorneys in the United States only approximately 108,000 work in law firms with 155 lawyers or more.<sup>13</sup>

Comments From: SandersC

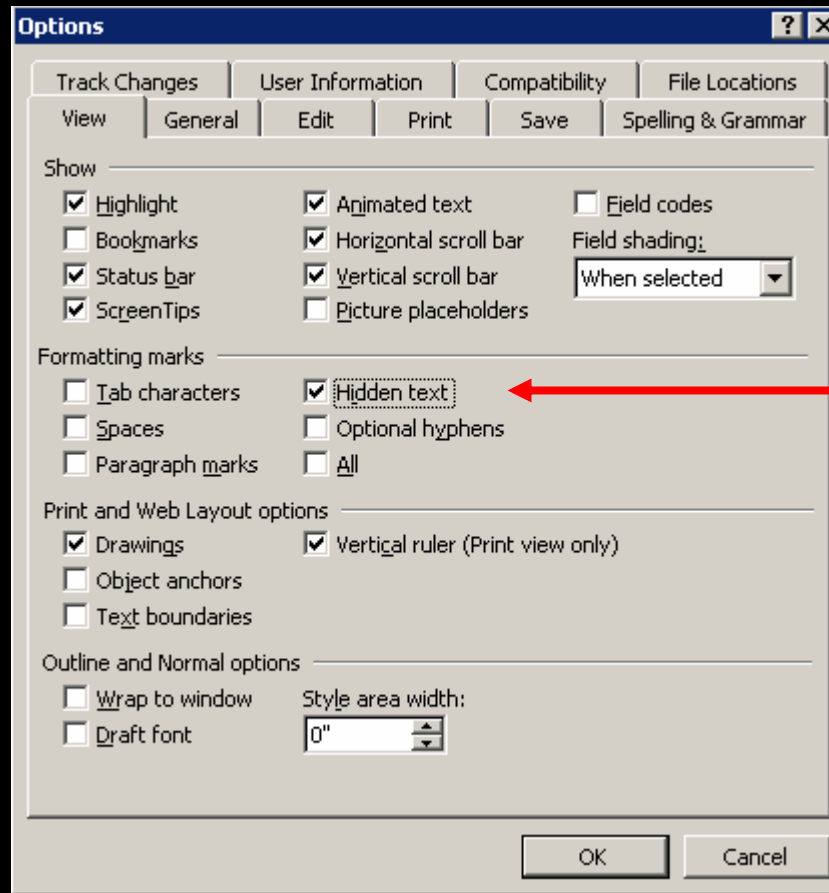
[chs7] FH questions the validity of a lot of the statements in this section. Again, might as

# Metadata: Hidden Text

## Defining the Digital Library

There is no universally accepted definition of what a digital library is or what it contains. Clifford Lynch refers to it as "shorthand for digital collections in libraries."<sup>8</sup> He suggests that it is an "exten[sion] of the collection to encompass substantial amounts of digital materials, making use of information systems to provide access to them, and providing a coherence of access between their digital materials and their printed materials."<sup>9</sup> # | For the purposes of the research, we have defined a digital library as a library whose collection is wholly electronic – Internet-based, CD and DVD, and software-accessible (e.g., brief banks) – and yet

# Metadata: Hidden Text



# Metadata: Hidden Text

## Defining the Digital Library

There is no universally accepted definition of what a digital library is or what it contains. Clifford Lynch refers to it as "shorthand for digital collections in libraries."<sup>6</sup> He suggests that it is an "exten[sion]" of the collection to encompass substantial amounts of digital materials, making use of information systems to provide access to them, and providing a coherence of access between their digital materials and their printed materials.<sup>7</sup>

This broad definition incorporates both the collection and the services.<sup>8</sup> For the purposes of the research, we have defined a digital library as a library whose collection is wholly electronic – Internet-based,

# Metadata: Document Properties

KJE20808 MEM.DOC Properties

General Summary Statistics Contents Custom

Title: DRAFT

Subject: ABA SCOTIS Minutes

Author: David P. Whelan

Manager:

Company: ABA

Category:

Keywords:

Comments:

Hyperlink base:

Template: Normal.dot

Save preview picture

OK Cancel



# Metadata: Document Properties

The screenshot shows a software window with the following metadata:

- Created:** Monday, April 14, 2003 2:35:00 PM
- Modified:** Wednesday, September 22, 2004 3:22:32 PM
- Accessed:** Wednesday, September 22, 2004 3:22:32 PM
- Printed:** Monday, April 14, 2003 8:47:00 AM

Additional information:

- Last saved by: SandersC
- Revision number: 4
- Total editing time: 3 Minutes

**Statistics:**

Statistic name	Value
Pages:	26
Paragraphs:	55
Lines:	363
Words:	3262
Characters:	17431
Characters (with spaces):	20831

# Metadata: Document Properties

metadataLTRC.ppt Properties

General Summary Statistics Contents Custom

Title: No Slide Title

Subject:

Author: Kathryn Thompson

Manager:

Company: Personal

Category:

Keywords:

Comments:

Hyperlink base:

Template: C:\Program Files\Microsoft Office\Templates\LTRCAnii

Save preview picture

OK Cancel

# Metadata: Document Properties

**KJE20808 MEM (2).rtf Properties**

General | Summary | **Statistics** | Contents | Custom

Created: Wednesday, September 22, 2004 1:30:00 PM  
Modified: Wednesday, September 22, 2004 1:30:41 PM  
Accessed: Wednesday, September 22, 2004 4:24:58 PM  
Printed: Thursday, May 20, 2004 11:28:00 AM

---

Last saved by: SandersC  
Revision number: 2  
Total editing time: 2 Minutes

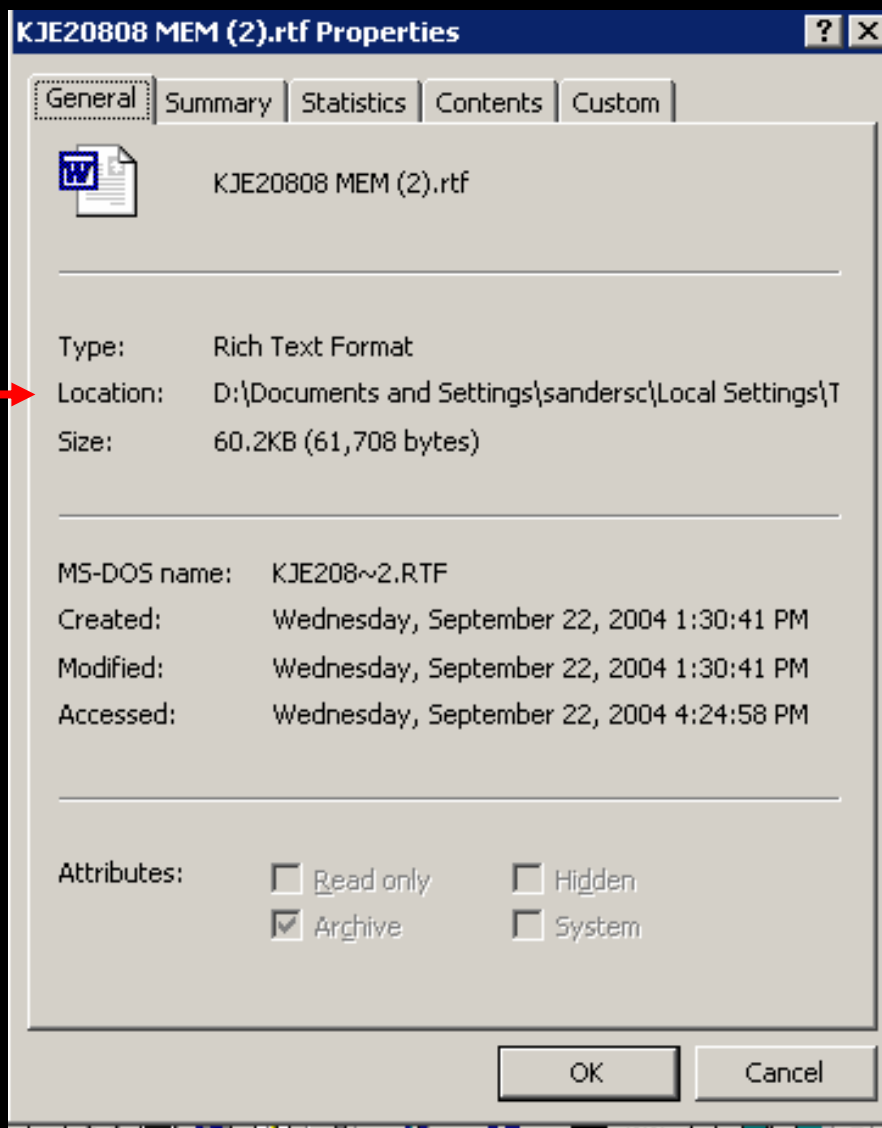
---

Statistics:

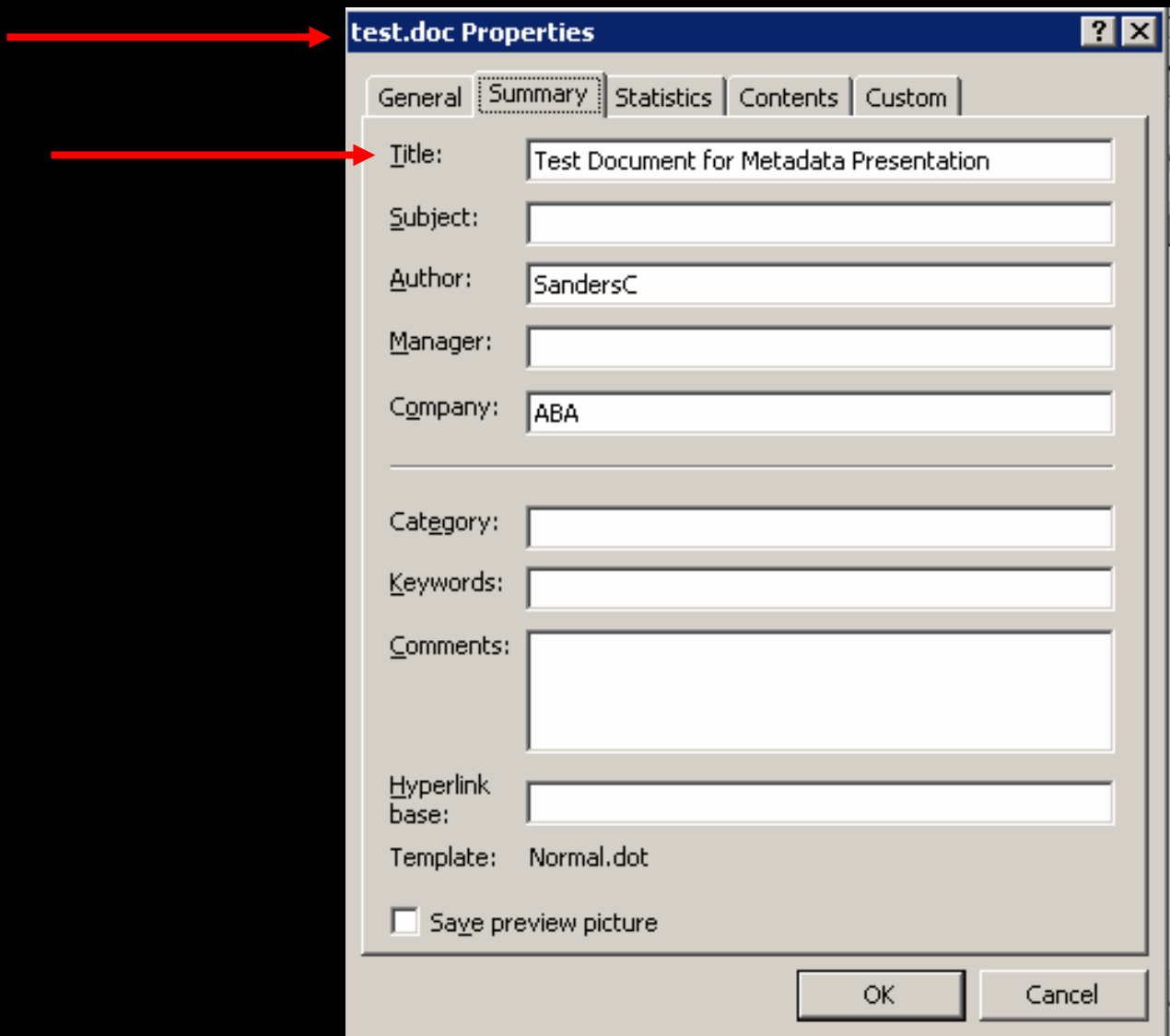
Statistic name	Value
Pages:	7
Paragraphs:	42
Lines:	301
Words:	3259
Characters:	16773
Characters (with spaces):	20120

OK Cancel

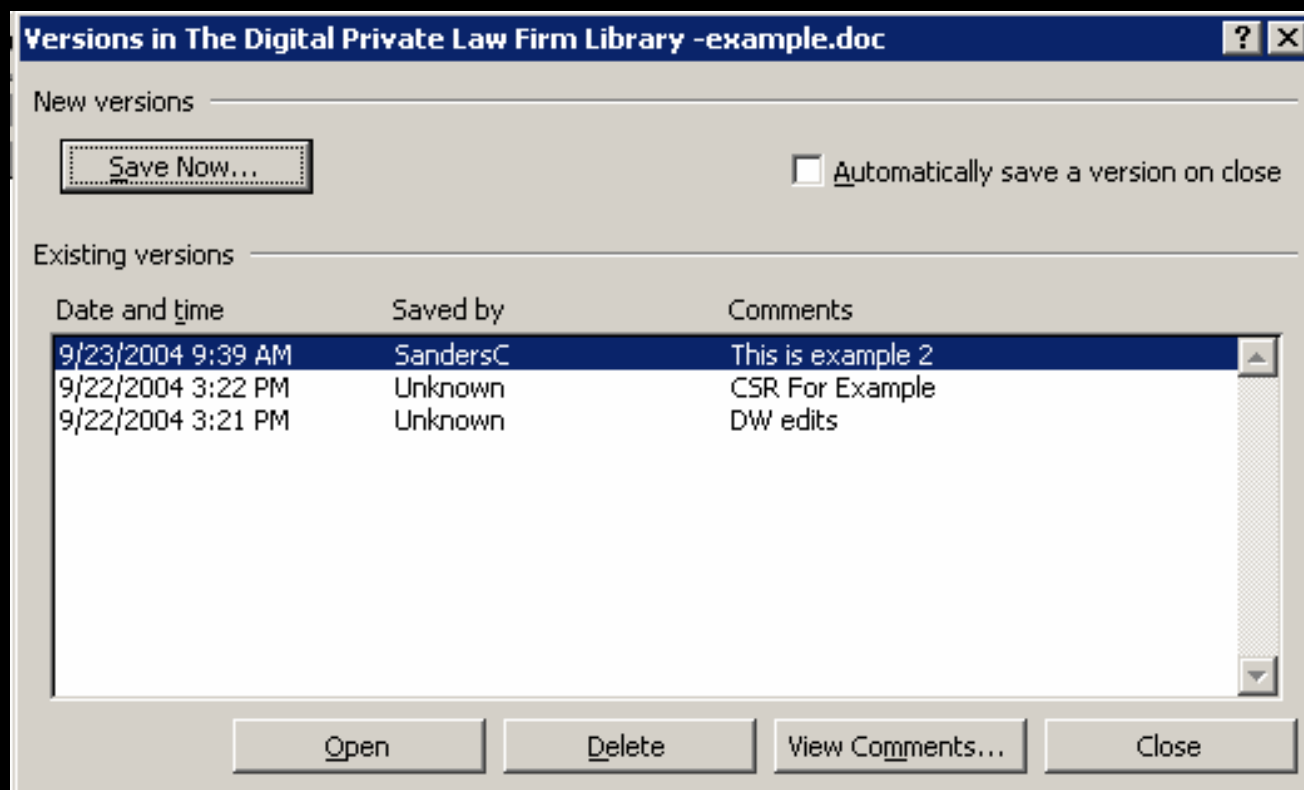
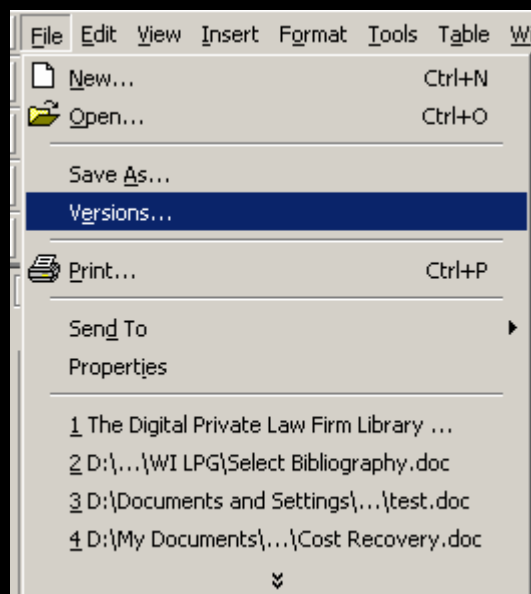
# Metadata: Document Properties



# Metadata: Document Properties



# Metadata: Versions



# Metadata: Fast Save

Options

Track Changes | User Information | Compatibility | File Locations

View | General | Edit | Print | Save | Spelling & Grammar

Save options

- Always create backup copy
- Allow fast saves
- Prompt for document properties
- Prompt to save Normal template
- Embed TrueType fonts
  - Embed characters in use only
- Save data only for forms
- Allow background saves
- Save AutoRecover info every: 1 minutes

Save Word files as: Word Document (\*.doc)

Disable features not supported by Word 97

File sharing options for "Document1"

Password to open:

Password to modify:

Read-only recommended

OK Cancel

# More Ways to View Metadata

- Save document as html. Click view/source.
- Recover text from any file. Choose file/menu/open/select “recover text from any file” from “from files of type”
- Open document in Notepad or other text editor

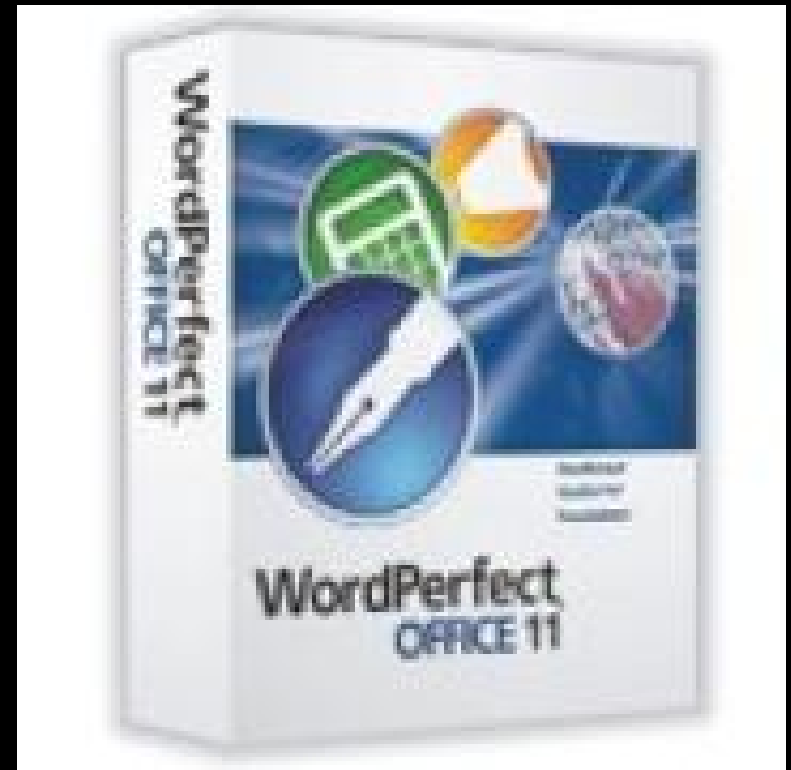


# More Ways to View Metadata

- Computer Forensics experts
- Binary files
- Meta Data Reviewer
  - <http://www.princetonsoftwarecompany.com>
- Unix/Linux users can use Antiword or Catdoc to view text file

# What About Word Perfect?

- Undo command reveals changes
- Comments
- Document properties similar to MS Word
- Versions



# Metadata: How Do You Get Rid of It?

- MS offers tools and tricks for removal
- Save as PDF
- Third party software tools



## Microsoft Removal Tools

- Knowledgebase articles for Word, Excel, PowerPoint from Office 97 forward
- Option in MS Word 2002
- Add on for MS Office XP/2003

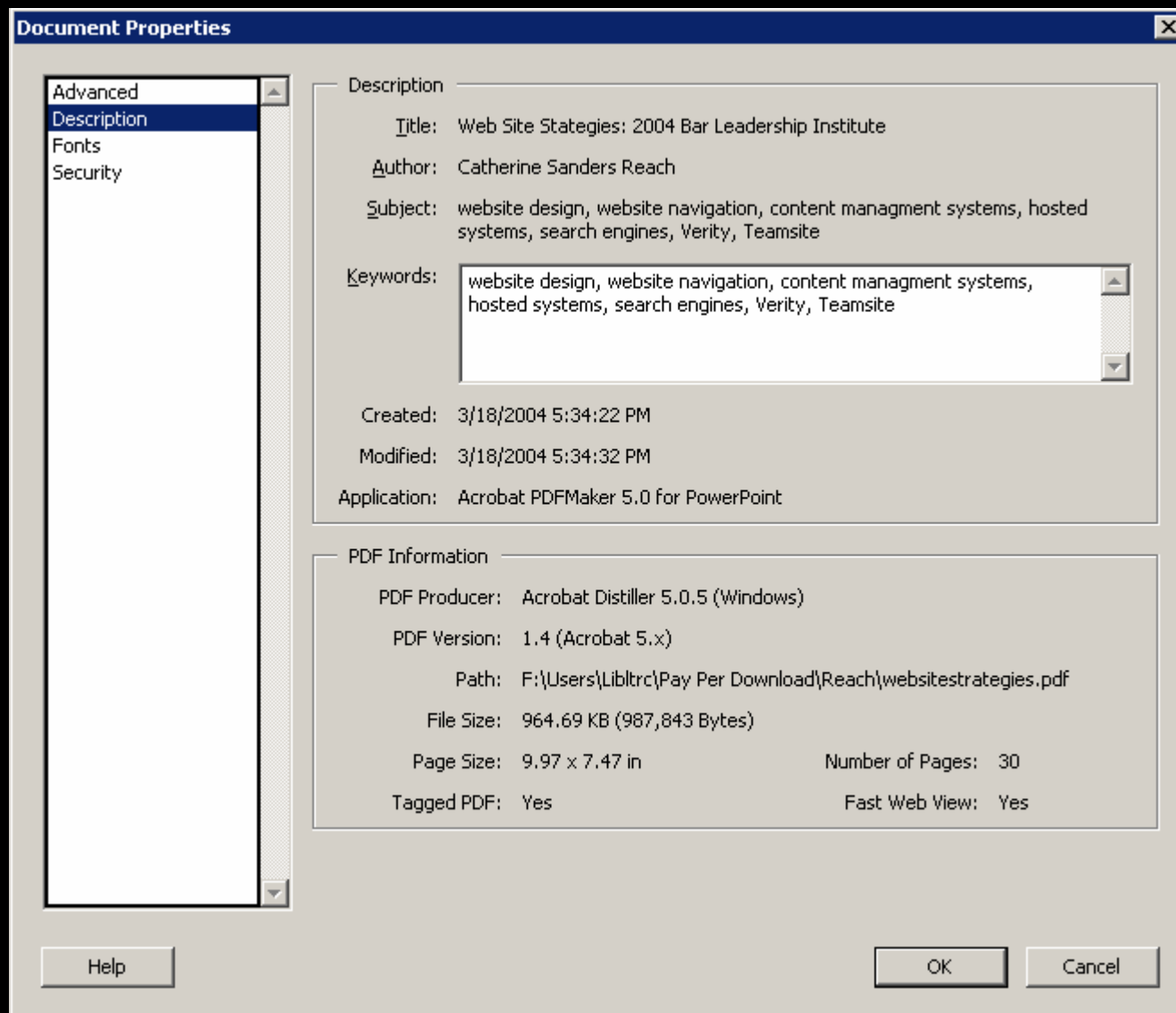
## WordPerfect Removal Tools

- Macro from MentorDocs
- WordPerfect X3 Option
- Instructions for removal in WP 12

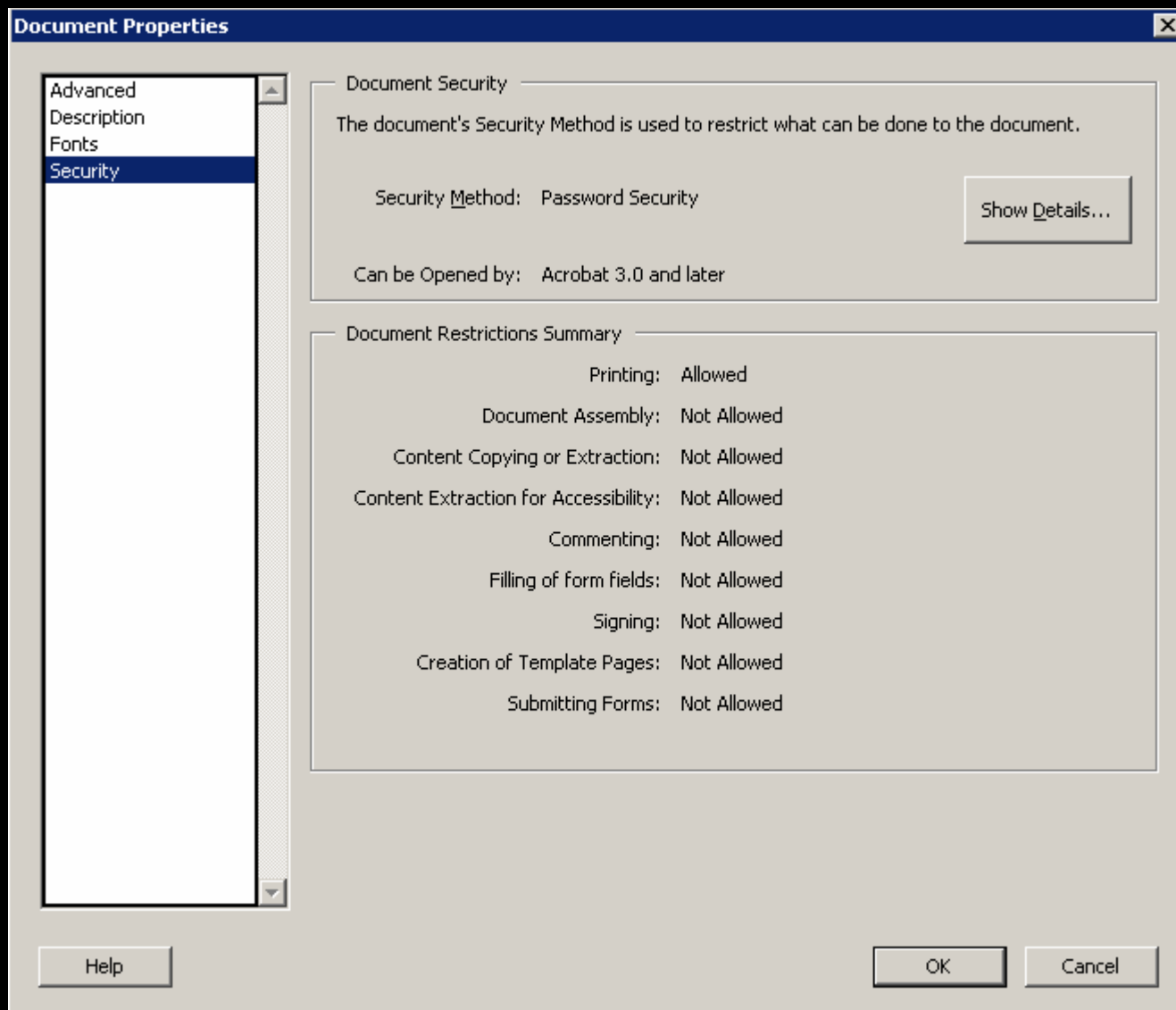
# The “Save as PDF” Solution

- PDF = Portable Document Format
- Can't edit
- Still saves many document properties
- Use Adobe Acrobat = \$\$
- Use GoBCL.com or OpenOffice = free
  - But no bells and whistles

# The "Save as PDF" Solution



# The “Save as PDF” Solution



# Metadata: Removal Software

- Payne's Metadata Assistant
- Esquire Innovations: iScrub
- Kraft Kennedy & Lesser: ezClean
- Workshare: Workshare Protect
- SoftWise: Out-of-Sight
- Appligent: GetInfo (also Redaction) for PDF



# What to Look For in Metadata Removal Software

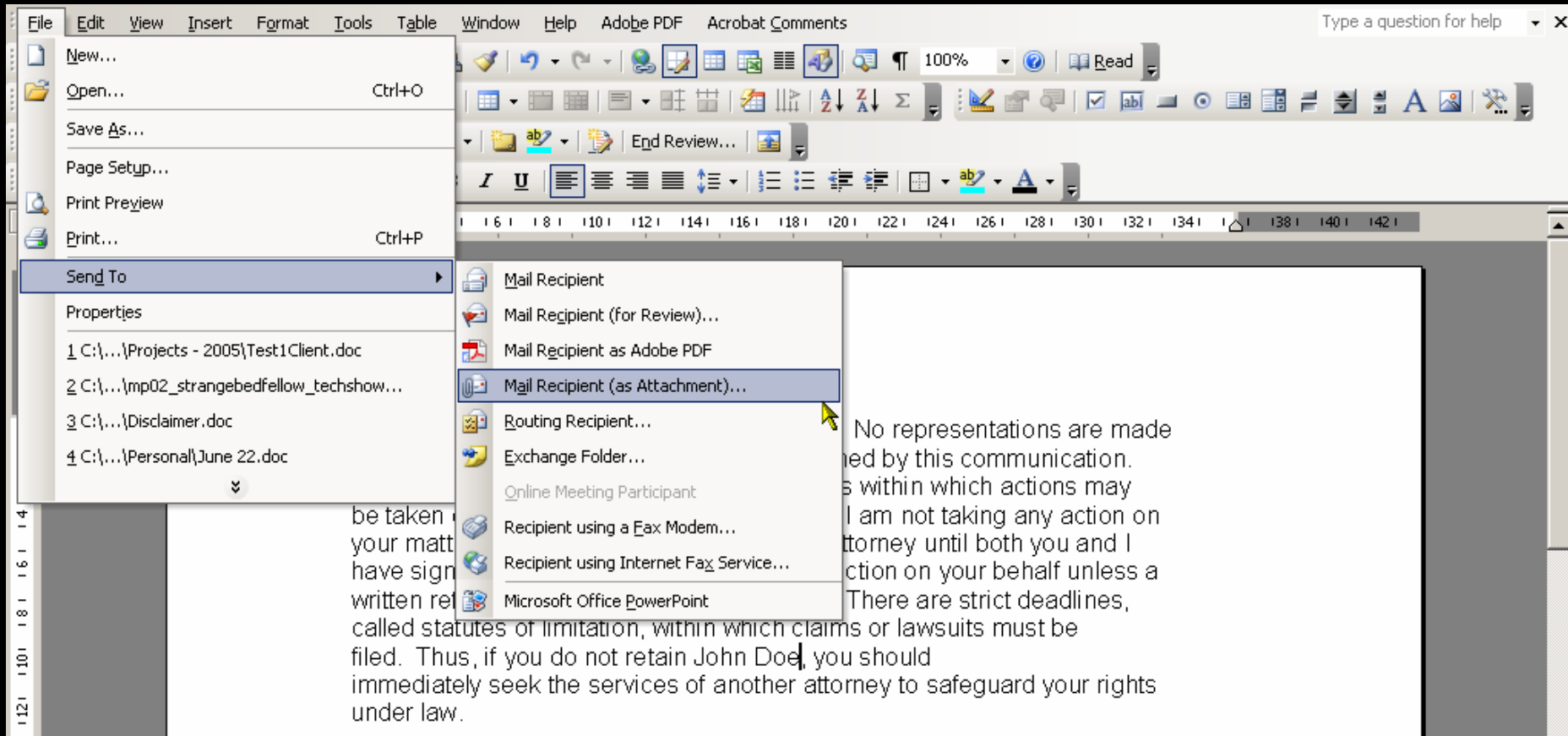
- Analysis of all products (MS Office Outlook, Word, PowerPoint, Excel)
- Individual or batch
- Automatic or user intervention
- Integration with document management software
- Cost/licensing

# Metadata: Don't Forget Email



- MS Outlook using MS Word as editor
- Header information reveals author, route, IP address
- Attachments

# Metadata: Don't Forget Email



# Electronic Discovery and Metadata

- Metadata is a critical component of document
- Metadata as a source of evidence
- Court described printed email as “dismembered”
- Discovery documents? Scrubbing = shredding
- Many grey areas

# Email Security

- Encryption: ABA Formal Opinion 99-413
  - Disclaimers
- Chat rooms/e-mail discussion lists/web forums
  - Recorded, searchable conversations
    - With other lawyers
      - Be aware that even closed lists are not completely private
    - With people seeking legal help
      - Be careful, follow state ethics when responding to online message boards to the general public
- Misaddressed e-mails



# Internet Security

- Your website
  - Consider click through assent to disclaimer statement to avoid conflicts or unintentional perception of formation of an attorney-client relationship
  - Disclaimer to protect articles and presentations being construed as legal advice
  - Ethics and advertising rules
    - KY blogger runs afoul of advertising rules
- Internet Use Policies
  - Staff's use of the internet can jeopardize the firm. Create and mandate internet use policy



# Understand Electronic Discovery

- Amendments to the Federal Civil Rules of Civil Procedure become effective December 1, 2006
- Understand e-discovery or else
  - UBS Warburg, Zubulake, Metropolitan Opera
  - Native files vs. print
- Retain experts
- Create an e-discovery strategy
- Test on smaller cases

# Legal Research

- Fee vs. Free
  - Make sure to check for authority (who is writing this stuff?) and currency (last updated when?)
- Print vs. Digital
  - Some digital materials are more up-to-date and easy to use
    - Eg. Shepards online or Shepards in print
  - Learn when to use print, when to go online
  - JUST ASK® a law librarian





Thanks!

Catherine Sanders Reach  
[Sandersc@staff.abanet.org](mailto:Sandersc@staff.abanet.org)



American Bar Association  
Legal Technology Resource Center  
<http://www.lawtechnology.org>