

# Resist Native Production!

The Safer Way to Conduct eDiscovery

Presented on May 29, 2008 at the  
Strafford Publications Teleconference on

## Metadata: Avoiding Ethical & Malpractice Exposure

by James Berriman, Esq.  
CEO, Evidox Corporation  
[www.evidox.com](http://www.evidox.com)

© 2006-2008 James Berriman

**Evidox**<sup>TM</sup>  
Electronic Discovery & Trials

# Resist Native Production

## eDiscovery: A different set of metadata risks

- In the context of eDiscovery, you are concerned with the documents of the parties rather than the documents you generate as counsel.
- The documents of the parties constitute substantive evidence on which your client's claims and defenses will be based. This evidence can include metadata.
- You have ethical obligations to conduct discovery properly.
- For evidentiary documents, the scrubbing approach often applied to documents you generate as counsel may not be appropriate because it alters evidence.
- On the other hand, producing evidentiary documents that contain unreviewed or irrelevant metadata may not be appropriate either.

# Resist Native Production

## eDiscovery: A different set of metadata risks

- Thus, during eDiscovery, you have a different set of metadata risks to address:
  - Complying with the duty to preserve evidence.
  - Avoiding inadvertent spoliation or alteration of evidence.
  - Implementing appropriate document control measures:
    - Document numbering, confidentiality endorsements.
  - Reviewing the evidence for relevance and privilege before production.
  - Limiting the scope of production to that which is properly discoverable.
  - Resisting unduly burdensome requests.

# Resist Native Production

The threshold question: Is the metadata discoverable?

- The new Federal Rules of Civil Procedure regarding electronically stored information did not actually enlarge the substantive scope of civil discovery (although they added some procedural requirements and protections).
- This is because electronically stored information was already fully discoverable under the earlier rules (if it was within the proper scope of discovery).
- Rule 26(b)(1) still governs the "Scope and Limits" of discovery and still imposes the requirement of "good cause" to obtain a discovery order.
- Thus, the threshold questions of scope and discoverability should still be assessed before producing metadata. As always, these are judgment calls.
- Fortunately, your existing expertise in making objections to overly-broad discovery requests is fully transferable to the realm of electronic documents and metadata.

# Resist Native Production

## When might metadata be properly discoverable?

- As a practical matter, when might discovery of metadata be appropriate?
  - When a file's authenticity is legitimately disputed (not just its substantive content)?
  - When a file's history or provenance is independently relevant?
  - When alteration or spoliation are suspected?
  - When crucial information is not available through other means?
  - When the burden of obtaining, reviewing, and producing the metadata is not excessive?
- In short: when there is good reason to look behind the face of a document.

# Resist Native Production

## When might metadata not be properly discoverable?

- As a practical matter, when might discovery of metadata be objectionable because it is unjustified or abusive?
  - When the dispute is only about the substantive content of the documents, not their electronic history or provenance?
  - When the authenticity of the documents is not legitimately at issue?
  - When alteration or spoliation are not reasonably suspected?
  - When the marginal utility of obtaining metadata is outweighed by the burden?
  - When the information can be more easily obtained through other means?
- In short: when there is no good reason to look behind the face of a document.

# Resist Native Production

## Protect yourself pending resolution of discovery disputes

- Even if you believe that production of certain metadata is objectionable, you must still protect yourself and your client from the risk of spoliation pending resolution of your objections.
- Make sure you have issued litigation hold letters to the client and its relevant employees to preserve the relevant electronic document in their existing states.
- Make sure you have preserved copies of the original native files that form the basis of the dispute. Preserving is much less expensive than processing them for eDiscovery review and production.

# Resist Native Production

## What are "native" files?

- We use the term "native" to describe electronic files in their original formats as found on a company's mail servers, file servers, local computers, and other digital storage devices.
- There are many native file formats. In most cases, the formats that matter are those associated with user-created data (as opposed to system files, program files, help files, etc.).
- The user-created file formats most often encountered in the context of eDiscovery include the following:
  - Microsoft Office files, including Word, Excel, PowerPoint, Access, Visio
  - Mailbox files, including Outlook PST and Lotus Notes NSF files
  - Adobe PDF files
- There are many other formats, but those listed above constitute the great majority of relevant native files typically found during civil eDiscovery in a business case.

# Resist Native Production

## Why are native documents a problem?

- Many native files are chock-full of metadata, including metadata that you can't see.
- When conducting eDiscovery, you must of course review your client's documents for responsiveness and privilege, and then produce the appropriate subset of those documents to the opposing party.
- How should you do this?
  - Should you open Word documents in Word?
  - Should you open PowerPoint files in PowerPoint?
  - Should you open mailbox files in Outlook? Etc.?
- And should you produce these files to the opposing party in native format?
  - No. Not unless there is a good reason or you are compelled to do so.
- There are many problems and risks with reviewing and producing electronic files in their native formats.
- You should resist native production for the following reasons.

# Resist Native Production

## Risk No. 1: Inadvertent alteration of evidence

- When you review native files in their native applications and produce them natively to the opposing party, you may have serious evidentiary problems. This is because the content or appearance of native files will sometimes change when opened or printed by you or the opposing party:
  - Application metadata (e.g., "Document Properties" in MS Office files) can be altered as a result of opening or printing a file. This can include the "last printed" date, the "last accessed" date, and the "last edited by" field.
  - Autocoded fields, such as the automatic date feature available in Microsoft Word, will automatically change when the document is opened and printed.
  - Some documents are configured to import data automatically from other files or external sources, or to recalculate figures. If this occurs during review or after production, the document will no longer contain the same data and content as it did when last used by your client.

(continued)

# Resist Native Production

## Risk No. 1: Inadvertent alteration of evidence (continued)

- The displayed username in Outlook will print the name of the currently logged-in user rather than the original user. This puts the name of the reviewer on every printed email rather than the name of the original custodian.
- Changes in formatting (such as the colors in a PowerPoint chart) can arise from local settings that differ from the original user's settings.
- Changes in pagination can arise from the selection of a different printer driver than the one used by the original creator.
- Changes can arise from opening the file in a different version of the application than the one in which it was created. For example, user comments in a PowerPoint 2003 file are not visible if the file is reviewed in PowerPoint 2000.
- The main concern is not that you or the receiving party will deliberately alter evidence, but that the evidence will be inadvertently altered and the parties will not realize that the evidence is no longer authentic or reliable.

# Resist Native Production

## Risk No. 2: Loss of document control

- When you produce native files, you may have serious document control problems:
  - Native electronic files cannot be directly endorsed with control numbers in an acceptable way.
  - This would involve substantively editing the files and changing their metadata.
- Thus, when a copy of the document is later used as evidence at trial or to support a motion, it may not be possible to determine with confidence where the document came from, especially because there are often so many variations of a particular electronic document in a collection.
- The document control issue compounds the problem with inadvertent alteration of native files. If a document's content and appearance have been inadvertently altered, and the document has not been endorsed with a control number, it can become especially difficult to determine the source and authenticity of a native file.

# Resist Native Production

## Risk No. 3: Inability to endorse as confidential

- When you produce native files, you have very limited ability to provide appropriate protection of confidential documents:
  - Native electronic files cannot be directly endorsed with confidentiality legends in an acceptable way.
  - This would again involve substantively editing the files and changing their metadata.
- Thus, restrictive legends such as "Subject to Protective Order" or "Attorney's Eyes Only" cannot be placed on the face of the documents.
- At best, only the media on which the native files are produced can be marked.
- If such a document were later printed for use as an exhibit at a deposition or in motion papers, a critical protection would be missing.

# Resist Native Production

## Risk No. 4: Inability to redact appropriately

- When you produce native files, you cannot redact them in an acceptable way:
  - Text can only be removed from native files by substantively editing the content.
  - This would again involve altering the evidence and changing their metadata.

# Resist Native Production

## Risk No. 5: Inability to review metadata adequately

- As a general rule, you do not want to produce anything that you and your review team have not reviewed for relevance and privilege.
- But if you produce native files, you are also producing the application metadata contained in those files.
- Application metadata is not always easy to review. This is especially true for "hidden" metadata that is not accessible via the application interface.
- This means that the review team cannot easily conduct an effective review prior to production.

# Resist Native Production

## Other risks and problems with native production

- **Viruses.** Native files can contain viruses. This means that both your review set and your production set may be infected and may cause further infection during review or after production to the opposing party.
- **Review environment.** A collection of native files in mixed formats cannot be easily reviewed in a single integrated environment and may require the use of different applications to be reviewed.
- **Software licensing.** The party who reviews and the party who receives native files may not own the applications necessary to open the files. This is especially true for specialized filetypes (like CAD files) that require very expensive licenses.
- **Inefficiency.** If a collection of files is reviewed and produced in native format without eDiscovery processing, many of the benefits of eDiscovery will not be realized. These benefits include deduplication, keyword filtering, automatic extraction of header data, coding, conversion to a uniform review format, and compatibility with litigation review tools.

# Resist Native Production

What is the accepted alternative  
to native production?

# Resist Native Production

## Rule 34 has two alternatives for production

- Rule 34 of the Federal Rules of Civil Procedure now addresses the forms in which electronically stored information may be produced.
- Unless the producing party is required to produce the information in a particular form, the producing party has two options:
  - To produce the information in the form in which it is kept in the usual course of business;

OR

- To produce the information in another reasonably usable form.
- See F.R.Civ.P. 34 (b)(2)(E)
- In the current state of litigation practice, printing the files to paper is generally not considered to be "another reasonably usable form" (nor should it be).

# Resist Native Production

## What do these alternatives really mean?

- As a practical matter, Rule 34 means that the producing party may produce their electronic documents in one of two forms:
  - The files can be produced in their original native formats (as Word documents, Excel spreadsheets, PowerPoint presentations, mailbox files, etc.) because this is the form in which they are kept in the usual course of business.

OR

- The files can be produced in the industry-standard litigation-support database format that preserves the relevant original content and that preserves the appearance and searchability of the original native files.
- This industry-standard format is called "TIFF/Text" or "TIFF/Text/Meta."

# Resist Native Production

## What is TIFF/Text/Meta format?

- "TIFF/Text/Meta" (or "TIFF/Text") is the most widely used format for electronic production in civil litigation.
- It is long-established and non-controversial.
- It is routinely accepted by the SEC, the DOJ, and every major law firm that I am aware of.
- It includes the following elements:
  - TIFF images created from the native files.
  - The searchable full text extracted from the native files.
  - Optionally, it can include fielded data and metadata from the native files to the extent chosen, agreed upon, or ordered.
  - A "load file," which is an index file that links these various elements together.

# Resist Native Production

## What is the "TIFF" of TIFF/Text/Meta?

- In this context, the "TIFF" is a digital image that is the equivalent of a printed page.
- It has the same appearance that a page of the native file would have had if printed on a black and white printer. In effect, the native file is "printed" to an image rather than to paper and results in the same pattern of fine-grained dots.
- These are the advantages:
  - It provides an authentic and accurate image of the face of the native file.
  - The image can be endorsed with control numbers and confidentiality legends.
  - It can be configured with the same options as a printed page (e.g., with headers and footers displayed; with comments displayed or not, etc.).
  - The process does not alter the metadata embedded in the native file.
  - The image can be redacted. Further, after the redactions are done, they are "burned in" with no hidden layers so the redactions cannot be "peeled off."

# Resist Native Production

## What is the "Text" of TIFF/Text/Meta?

- The "text" is the full searchable text contained in the body of the original native file.
- It is extracted by a process that does not alter the dynamic content of the file.
- It is extracted to allow the content of the file to remain fully searchable when it is reviewed and produced (since the graphical TIFF image alone is not searchable).
- Unlike OCR text obtained by processing scanned TIFF images, extracted text is perfect since it is take directly from the original file.
- CAVEAT: If you redact the TIFF image for a document, you must remove the corresponding text for that document. This is typically done in two ways:
  - By not producing text for redacted documents (globally); or
  - By performing OCR of the redacted TIFF image and using the OCR text in place of the original extracted text.

# Resist Native Production

## What is the "Meta" of TIFF/Text/Meta?

- As a preliminary matter, in the eDiscovery realm, the term "metadata" is used loosely to mean any of the following:
  - File system metadata, i.e., fielded data about the file that is not contained inside the file but is contained in the file system in which the file is stored.
  - Application metadata, i.e., fielded data contained within the file, not visible on its face but intended to be viewed in the application environment.
  - Hidden metadata, i.e., data embedded within the file but not intended to be viewed (e.g., internal caches, internal editing artifacts).
  - Visible email headers, i.e., fielded data that is visible when the email is viewed in the normal email environment (from, to, cc, bcc, subject, sent).
  - Email transmission headers, i.e., fielded data that contains transmission and protocol data, often not visible by default (it is arcane).
  - Batch coding. This is a type of metadata deliberately added during eDiscovery processing and is sometimes produced (e.g., custodian or source).

# Resist Native Production

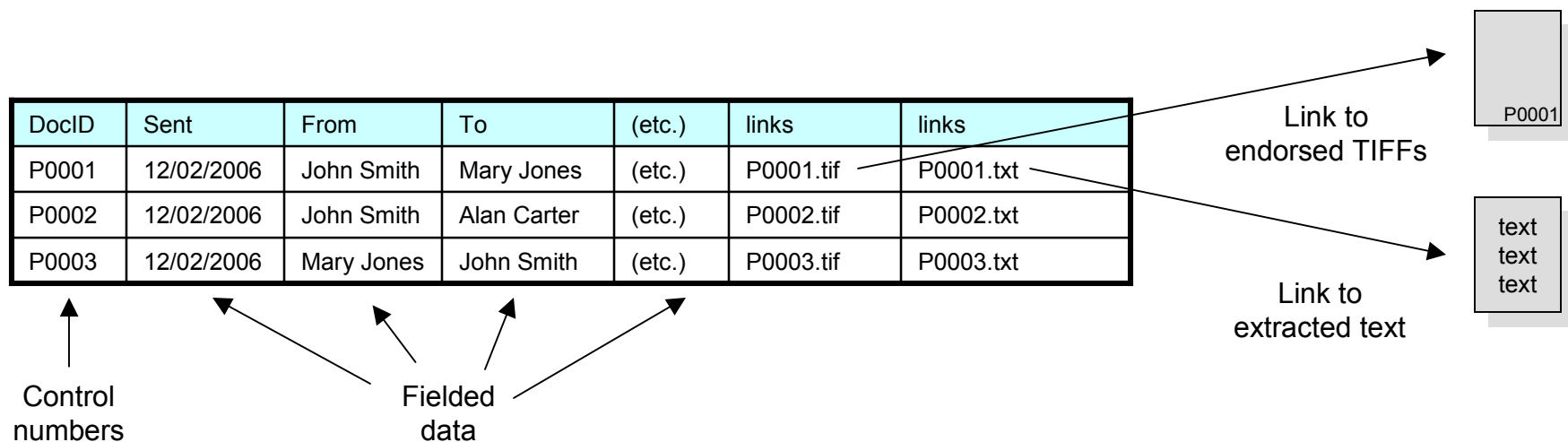
## What is the "Meta" of TIFF/Text/Meta?

- In the context of "TIFF/Text/Meta," the term "meta" is used to identify any of the available fielded data that the parties may choose (or agree, or be ordered) to produce.
- It does not include all possible types of metadata. In particular, it does not include the "forensic" categories (like hidden metadata).
- "Meta" may include the following:
  - Visible email headers
  - File system metadata
  - Application metadata to the extent available
  - Batch coding
- Or it might include none of these except the bare bones of the load file (e.g., DocID and other structural elements).

# Resist Native Production

## What is the "load file" provided with TIFF/Text/Meta?

- The "load file" is an index that links the other elements and allows them to be loaded into a review system like Summation, Concordance, CaseLogistix, Introspect, etc.
- Think of a load file as a table of data with row of data for each file.
- The data contained in each row will include whatever fields of metadata you have chosen (or agreed, or been ordered) to produce, along with the unique control number applicable to each file and links to the TIFF image and searchable text for each file.



# Resist Native Production

## The advantages of TIFF/Text/Meta format

- The TIFF/Text/Meta format addresses almost every concern regarding native production:
  - The risk of inadvertent alteration of evidence is reduced or eliminated.
  - Document control is maintained and numbers can be applied to each page.
  - Confidentiality legends can be applied to each page.
  - Redactions can be made to the TIFF images.
  - No metadata is produced except that which can be reviewed by your team and that which you know you are producing.
  - Viruses are eliminated.
  - All files can be reviewed in the same integrated environment.
  - All files are in the same format, so no native file licenses are required.
  - All of the efficiencies of eDiscovery processing and review are preserved.

# Resist Native Production

## The "native review" approach

- Although you want to resist native production, the eDiscovery process allows a hybrid form of review called "native review."
- "Native review" involves locking down the native files into a "read only" format but also extracting the full text and meta fields, then loading everything into an integrated review environment (Summation, Concordance, CaseLogistix, etc.).
- The attorneys can then conduct their review on the native files with all the benefits of searching and sorting on the text and fielded data.
- When their review is complete, the TIFF conversion and endorsement is applied only to the subset of files selected as responsive.
- The opposing party then receives the TIFF/Text/Meta production set.
- This is the most cost-effective way to prepare for making a document production in TIFF/Text/Meta format. It avoids the cost of TIFF conversion and endorsement except for the subset of files deemed responsive.

# Resist Native Production

## When is native production justified?

- Nevertheless, there are some circumstances where native production is justified.
- These circumstances might include:
  - Where a forensic analysis must be performed on "hidden" metadata or other aspects of a native file that cannot be captured in TIFF/Text/Meta format.
  - When a particular file must be manipulated in its native format to be useful to the receiving party (e.g., a spreadsheet that contain hidden formulas not visible in the tabular data; an executable file that must be run).
- Such circumstances are very limited.
- Accordingly, the only documents that should be produced in native format are those that occupy these special categories.
- By default, everything else should be produced in TIFF/Text/Meta format.
- You may have to educate opposing counsel about this approach.

# Resist Native Production

## Conclusion

- Opposing counsel often does not even seek meta production.
- If so, then your default choice should be a simple TIFF/Text (no meta) production. As a practical matter, this generally satisfies the "reasonably usable" rule.
- You should limit the "meta" fields that you produce to those that you believe are appropriate. Unless there is a good reason otherwise, less (or none) is better.
- The scope of "meta" production can be negotiated with opposing counsel. If opposing counsel seeks too much, you can object.
- If opposing counsel seeks native production, you should object to the extent that there is no good justification for the request.
- If there is some justification for the request, you should seek to limit the native production only to that subset of the collection to which the justification applies.
- Sauce for the goose is sauce for the gander. Remember that you may want metadata from opposing party.

## Bio



**James Berriman, CEO**  
Evidox Corporation  
www.evidox.com  
207 South Street  
Boston MA 02111  
jberriman@evidox.com  
617-654-9061

Mr. Berriman is CEO and co-founder of Evidox Corporation. He was formerly Senior Counsel and Director of Litigation Technology at Goodwin Procter LLP, an AmLaw 100 firm of over 650 attorneys. He has been developing litigation technology systems since 1982. Before becoming an attorney, he developed a discovery database system on behalf of a Fortune 100 client involved in a nationwide series of lawsuits and developed a billing and timekeeping system for litigation attorneys.

Upon graduating from law school in 1990 and joining Goodwin Procter, Mr. Berriman began incorporating technology into his litigation practice. In 1999, the firm asked him to take responsibility for litigation technology firmwide. He founded the firm's Litigation Technology Group and grew it to a team of 15 specialists occupying a 4,000 square foot state-of-the-art facility. He coded and implemented the firm's litigation case management system and established the firm's scanning and digitizing labs, its e-discovery review facilities, and its trial technology systems.

Mr. Berriman co-founded Evidox Corporation in 2006 to provide e-discovery and litigation technology services.