Best Practices for Native File Review and Production

By Jim Wagner, Co-founder and CEO, DiscoverReady LLC
I. Context and Current Environment

The world of e-discovery is quickly going native. While paper or TIFF were the readily accepted standards for producing electronic documents for many years, several factors are now driving the industry to deliver more native documents instead. Offering increased efficiencies and a more thorough understanding of a matter, this trend towards producing documents in their native format was greatly encouraged by new federal e-discovery rules that took effect in December 2006.

Production and review of native documents present both unique opportunities and challenges for lawyers and clients. Lawyers, in particular, need a complete understanding of production standards and strategies in both litigation and regulatory settings.

This white paper reviews the history and context of native file review and production, and then offers our recommendations and tips for the best practices to employ during production and review of native documents for regulatory requests, investigations and litigation.

Historical Background

The rapid migration from the use of paper to electronic documents in legal matters has created problems as it has solved others. Our early paper beginnings seemed chaotic enough with copious amounts of physical copies of discovery documents distributed to all involved with a case. The 1990’s found paper being converted to scanned images, which created even more paper as those images were transformed back into hard copies (blowbacks) for productions sets and to use at depositions and trial. And when the importance of e-mail was eventually acknowledged, these electronic files were “blown back” and produced an even more staggering amount of paper, which was converted back to scanned images, which then was often printed out again and … Well, you get the picture.


Why were we so slow to adjust? For one, we had little incentive to change. With few decisions involving poor e-discovery practices in the headlines (unlike today) there was little awareness of organizations being punished for poor production practices. There was a degree of ignorance or lack of focus by requesting parties. And there were plenty of lawyers who looked at the complication and expense of e-discovery and opted instead for the ostrich approach: “We won’t ask if you don’t.”

Also, e-discovery is technical. It poses challenges that not enough people – whether lawyers, clients or even vendors – fully understand. And, of course, the legal community is not known for its progressive use of technology. Lawyers like paper. Our focus is not necessarily on achieving efficiency or cutting costs.

But, perhaps, the biggest reason for the slow adoption of these practices was the lack of understanding and awareness by in-house counsel. Not only were they slow to catch on to the cost ramifications of sticking with outmoded processes for document production and review, but they were also very reluctant to turn over control of their electronic data files to vendors and other third parties.

The Drive to Native

In retrospect, it seems obvious that paper-based e-discovery made little sense. It was inefficient, ineffective and time consuming. Eventually, lawyers and clients alike declared: “Enough is enough!”
As often happens when change is slow in coming, it arrived with a bang. In response to the overly conservative e-discovery practices, now lawyers took a much more aggressive approach. That meant that, where paper and TIFF once sufficed, in many cases, now only native files would do.

This drive to native received an official seal of approval with the December 2006 amendments to the FRCP, which require parties to produce electronic files in the form in which they are ordinarily maintained – translation: “native” – or in a form that is reasonably usable.


In addition to the FRCP and the courts, government regulators and investigators are demanding full access to files in their native formats to use in their own applications.

What This Means to Responsible In-House Counsel

Today, native is where the evidence is. That is now the norm and it is not going to change in the foreseeable future. More to the point, courts and requesting parties have caught on. They react viscerally to non-compliance and to mountains of barely usable paper – think piles of printed spreadsheets.

Added to that, overwhelming volumes of data equals overwhelming cost, and that equal unhappy in-house counsel. Poor planning and processes equal poor (sometimes devastating) results. That makes the role of in-house counsel more important to the discovery process than ever before. The general counsel is now a key part of the litigation team, with greater say in how outside counsel handles the process.

What This Means to Responsible Outside Counsel

For those lawyers still hiding behind reams of paper, this new environment will require them to make quantum leaps in their practices. But for others, who have kept more current with technological advances in the way information is created and stored – and with the regulations and law that have developed to treat those advances – this drive to native presents opportunities to distinguish themselves and better serve their clients.

For outside counsel, the new landscape provides the opportunity to take the lead on behalf of clients and say, “This is how it should be done.” Lawyers and their clients must be prepared to define discovery processes at the outset of a matter. To do this, they must understand and be prepared to defend their clients’ collection, review and production processes. To the extent that all or some portion of the production will be in native format, they also must understand and select applications capable of executing native file reviews and productions in a defensible manner. At the same time, they should be prepared to advise clients on the risks associated with these processes.

II. Executing Native Productions

Under amended FRCP 34(b)(ii), barring specific agreement, parties must produce electronic information either in the form in which it is “ordinarily maintained” or in a form that is “reasonably usable.”
The rule does not give a party the right to require that documents be produced in multiple formats. However, a party does have the right to request different files in different formats.

The notes to FRCP 34(b) are particularly important. Along with the text of the rule, they make clear:

- The requesting party may ask for different forms of production for different types of electronically stored information.
- The responding party must state the form it intends to use for producing electronically stored information if the requesting party does not specify a form or if the responding party objects to a form that the requesting party specifies.
- The option to produce in a reasonably usable form does not mean that a responding party is free to convert electronically stored information from the form in which it is ordinarily maintained to a different form that makes it more difficult or burdensome for the requesting party to use.

With these changes in the federal rules, the meet-and-confer conference takes on new importance. How you produce will significantly alter how you manage and review. Before you go into your conference, you need to have considered a discovery strategy that will allow you to produce documents in reasonably usable formats while also protecting against disclosure of privileged information and trade secrets. You need to have thought about what you want to achieve and what you and your client are prepared to accept.

Best Practices for Production Standards

The jury remains out on precisely what the rules mean by “reasonably usable.” It is fair to say, however, that you should not expect or accept less than native or “native-like.” In this context, a reasonably usable format (other than native) should include full metadata for all relevant fields and full text, along with an image showing all hidden information, such as comments and tracked changes.

You do not need native fields for everything. In many cases, TIFF files will be fully adequate and perhaps more practical. One reason for this is that the litigation support industry is very effective at handling TIFFs and many in-house litigation support tools are geared for TIFFs. Another is that TIFF files are easier to number and work with at depositions and at trial.

But if you are going to deliver files in TIFF format, you have important decisions to make beforehand about how you will handle various fields. Take an example as common as an e-mail address. It may display in an e-mail as only a name, as only an address, or as a name followed by an address. Do you want to include the full name and address? This is an important practice point, because anything less than a full name and e-mail address may leave you guessing as to identity and may be more challenging to search.

Prior to making any production request, we recommend that you develop and use a form similar to that shown here. (Figure 1.) Of course, you would customize the form and its fields to suit your own purposes. But a form such as this helps ensure that you think about various types of documents and their information fields before you go into the meet-and-confer.
Still, for certain types of file formats, TIFFs would seldom be considered “reasonably usable.” Examples are spreadsheets created in Microsoft Excel and presentations created in Microsoft PowerPoint. Also, TIFFs often present challenges in showing the parent/child relationship between documents.

Given this, we often recommend using a hybrid approach, delivering TIFF files when sufficient and delivering native files where necessary. Although litigants are not entitled to receive information in multiple formats, in certain instances they may even wish to stipulate to the use of both native files and TIFF files with metadata. That way, the TIFF files will enable the parties to work from same documents and Bates numbers for trial and deposition preparation and these files will “play well” with most litigation support tools. Meanwhile, the native files will address the unique attributes of certain file formats such as Excel; preserve color, action and context in PowerPoint files; and ensure the functionality of files from programs such as Microsoft Project and Microsoft Access.

But remember the dog that caught the bumper. Do not ask for native files if you do not have the tools and processes to deal with native files. And be aware that it may cost more to reprocess native documents.

Unique Files/Unique Scenarios

Keep in mind that unique file types call for unique solutions. At the outset of the discovery process, reserve the right to revisit any file types that you later find to be difficult to properly display or convert and that require other augmentation.

Databases present unique problems. They are virtually impossible to review in their native state through litigation support applications. The common practice is to run reports or convert the data into a usable format. Other options are to use neutral parties to query and report on the data or to recreate the database environment. Typically, we suggest reporting or accessing the data in a controlled environment.

Other areas in which you are likely to encounter unique files are engineering, IP development, and practice-management and collaboration tools. In any of these circumstances, the key is to preserve your options if the original production method proves inadequate.

Producing in the Regulatory Environment

Production in the regulatory environment is becoming increasingly difficult, as different agencies, and sometimes individuals within the same agencies, are using different tools to review data and setting specific, but varying, preferences for production format.
While a surprisingly high number of regulators still request or accept TIFF productions (together with text and metadata), the overwhelming trend in this area is for regulators to request native format. And, although expectations regarding production format vary, beware that regulators have developed tremendous impatience when presented with electronic files blown back to paper as they have become increasingly sophisticated and have an increasing desire to “see all” in the context of their investigations.

In short, you should expect only limited opportunities to negotiate production format in the regulatory context. At the same time, you may have more leeway to negotiate over the scope of the production related to the pertinent time span, number of custodians, etc. If you understand the context and the timeline, you may be able to go back to the requesting regulator with calculated estimates of the amount of information to be produced and money to be spent if the request were followed as presented and use those estimates to negotiate a more manageable scope and, occasionally depending on cost and timing issues, format.

To make matters more complicated (and costly), we have seen a recent trend among regulators to request to see “all data” from custodians for significant time periods. Without the use of keyword or date range filters, these requests can involve hundreds of thousands of files and perhaps as much or more in dollars to cover the cost of review and production. These “produce all” requests demonstrate the increasing need to have a cost-effective method in place for screening electronic documents prior to a large-scale production.

An important practice point related to “produce all” reviews—employees often reveal much about their personal lives in the e-mails they send at work. Some of that information can be highly embarrassing. One need only look at the Enron e-mails for ample evidence of this. Thus, in the produce all context, screening should weed out privileged communications, of course, but organizations should also consider screening for highly personal information. So think privilege, but also think personal privacy.

Best Practices for Collecting Native Documents

Historically, companies have often used home-grown techniques for collecting documents— for example, simply dragging files to a common production folder; however, this approach will change the path structure in which a file resides and change its “last modified” date and other key information.

For a paper blowback production or a production of TIFFs without metadata, the modification or elimination of this type of metadata is insignificant, because this information is not reflected in the image or piece of paper produced to the requesting party. But in native-file production, poor document collection practices come front and center. When you deliver native files, a key component is the metadata, one component of which is the file folder where the document resides.

Consider the example in Figure 2. Two folder windows each show the same document, “StandardPricingModel.doc.” In the first, the file path shows that the document resides in a folder titled “Form Documents.” But in the second, the file path shows it to be in a folder named “Form Documents from Prior Employer.” The difference between the two folders’ names could mean all the difference in a litigation scenario.

This is not to suggest that every document collection should become an exercise in forensics. Rather, the point is that you should carefully consider and monitor the implications of the collection practices you employ and that native file productions tend to reveal more about the underlying methods by which electronic files are stored, copied for review and produced.
Best Practices for Numbering Native Documents

While Bates numbering is easy with TIFF documents, it is anything but with native documents. So how should you stamp a Bates number on a native file?

There is no ideal method. The most common practice is to insert a unique identifier in the file path in some way. The downsides to this method are that the identifier is more difficult to see than a traditional Bates stamp and having it in the file path obstructs downstream sorting and review.

Another approach is to provide a separate XML or CSV file showing a unique document ID. This would show the ID as a unique field in a litigation database, but the significant downside here is that the ID is not contained or viewable within the actual file.

A third approach is to provide an ID in a unique field, such as document properties or something similar. Downsides of this approach are that the IDs would not be easy to view and the process of inserting them could overwrite other data or change the document’s last-modified date.

Whichever approach is followed, the numbering system must be established in advance. If you are producing documents for the SEC, be sure to stipulate how you will ID files so you know you are on the same page. Once you have established a method, deliver data samples to the requesting party to confirm that the method will be acceptable. Any documents that will be used at trial or in depositions will most often be converted to TIFF format, endorsed with Bates numbers, and printed to paper.

Best Practices for Redacting Native Documents

When redaction is required, it is because documents contain sensitive data. This makes “getting redaction right” the highest-risk element of any production. But for native documents, redaction is not easy – in fact it can be quite complicated.
Native file redaction requires substantial planning and training. Reviewers involved in the native file redaction process must have a sophisticated grasp of the application in which the redaction will take place.

They must be meticulous both in their review and elimination of potentially sensitive information, and they must also be extremely diligent in avoiding unintentionally deleting or modifying other information that is relevant and not subject to redaction. Further, the technical team and reviewers in a native file redaction process must follow detailed steps (e.g., creation of all-new files so as to avoid modification of the original, using some method of track changes to monitor those portions that have been redacted, devising and executing queries to execute with the QC process, etc.) in order to defend the production process and to reduce the risk of inadvertent disclosure.

Given the cost, complexity and risk involved, the best practice with regard to native file redaction may be to fight your hardest to avoid it. Redaction (rather than full withholding) should occur only if the requesting party can demonstrate a legitimate need that cannot otherwise be met and if the parties can identify a reasonable and certain process for accomplishing it.

### III. Executing Native Reviews

**Managing Review of Microsoft Word**

**Figure 3**

![Image](image1)

**Figure 4**

![Image](image2)

In Microsoft Word, the view makes a huge difference. Figure 3 shows a document with “track changes” disabled. Figure 4 shows the same document with “track changes” enabled. What the naked eye may not see in the document may be there nevertheless and, as in this case, that “hidden” text may be material.

The lesson in this is that, when reviewing native documents, make sure you are looking at all the text the document contains. The safest way to see all the text is to view the document in its native application and in track changes mode. Avoid using “near native” viewers, such as Quick View Plus, that cannot show documents in “Final Showing Markup.”
Equally important is that you train your reviewers on the critical need to review as a default in a “full text” format, such as Attenex Patterns’ “text-and-hits” mode, which generally shows all of the text of a document, including “hidden” information.

Managing Review of Embedded Objects

With many Microsoft applications, what you see may not be all you are getting, thanks to their ability to allow embedding of files. A pie chart contained within a PowerPoint presentation (Figure 5) may actually have an Excel spreadsheet embedded within it (Figure 6). In your review process, you should be prepared for the potential of embedded objects.

If your review application does not separately identify embedded objects, then you need to confirm that embedded file contents will show in a full text mode and encourage users to stay in this mode. This tends to pick up embedded information. At the same time, teach reviewers to be on the lookout for embedded files and, if using a true native file viewer, to “right click” and open the underlying source files. Once again, it is best to avoid using near-native viewers that cannot open the embedded source files.

Figure 5

![Figure 5](image1)

Figure 6

![Figure 6](image2)
Figure 5 is a slide from a PowerPoint Presentation. Figure 6 is the embedded source file for the chart in Figure 5. Note from Figure 6 that the chart displayed in Figure 5 is but a small part of a larger, multi-tab XLS spreadsheet. To access the full XLS the review application must automatically separate embedded files or must allow the reviewer to “right click” on the graph and open the source file.

Managing Review of Excel

Excel documents also contain notes and embedded data. These components, such as cells, sheets and even entire workbooks can be “hidden” from view. Here again, it is best to avoid the use of near-native viewers that cannot open hidden columns and rows. (See Figure 7 and Figure 8).

Figure 7

![Figure 7 Image]

Figure 8

![Figure 8 Image]
Figure 7 is an XLS as opened by the reviewer. Figure 8 shows what the diligent reviewer finds when “unhiding” Row 18.

It is also important that you aggressively train your reviewers on the need to review in a full text mode as their default. Generally, such a mode will show “hidden” information. If reviewing in native-file mode within Excel, reviewers need to select the option to “unhide” all hidden rows and columns.

In addition to dealing with hidden information, be mindful that reviewers sometimes make the mistake of looking at only the top page of a spreadsheet but not the subsequent tabs. Make sure your reviewers are trained to look at all sheets within a workbook. Additionally, careful keyword screening can be used to ferret out important and potentially privileged files.

Be Wary of Thread Suppression

Series of e-mail messages often form threaded conversations. I write you, you write back, I respond. Many applications combine related e-mails into threads. By creating a single chain of messages, they save reviewers a lot of time by significantly reducing the volume of e-mail they must review. (Figure 9)

But there is a risk here, particularly in the context of privilege reviews. Thread suppression only presents the last e-mail in a chain. But all who participated as senders and recipients through the entire thread may not be included in that last e-mail.

When screening for potentially privileged information, it is common to search for documents to or from a particular lawyer. If that lawyer participated somewhere along the thread, but not at the end, your search of authors and recipients may not identify that lawyer’s involvement.

The key step in avoiding missing a potentially privileged communication in such a scenario is simply understanding that, depending on your review application, thread suppression may allow names to be dropped off an e-mail chain. Look for applications that effectively search all “to” and “from” fields in e-mails and that consistently track discussion participants. To the extent that you do not use such a tool, develop careful procedures to account for that shortcoming.

Figure 9
Using Keyword Screens Prior to Production

As a final protective measure prior to production, consider using keyword screens against your proposed production set to ensure that you have caught everything of interest from the standpoints of privilege and relevance.

Do this by identifying critical terms for highest level of relevance and comprehensive terms for privilege. Using these terms, execute searches against the planned production set. Then review the resulting set in a "show hits" mode to identify those critical terms within the set.

When working with native documents where so much information can be hidden or difficult to see, this is a critically important step. It reduces the risk of having sensitive information fall through the cracks.

Summary: Best Practices for Native Review

- Train reviewers on use of a “full text” mode. Make sure they understand the technology and the limitations of the technology. When reviewing, use the text mode first to be able to review “hidden” data.
- Train reviewers to identify and open “hidden” information in Microsoft applications. Make sure they are trained in the way to review in native applications.
- Where possible, avoid use of near-native tools, such as Quick View Plus, that do not allow reviewers to access “hidden” information.
- Monitor and ensure quality control of reviewers’ compliance with instructions.
- Use keyword searches on documents to be produced to screen for privilege and other issues.
- Strike a careful balance between paralyzing/non-viable analysis (e.g., looking at every potential field of metadata in an Office application) and appropriate diligence (e.g., reviewing in full text mode, using the native application to show track changes and hidden information).
- Allow in-house counsel to participate in defining the review and production process — there are real risk and cost implications involved and ultimately many of these are business as much as legal decisions.
- Follow consistent processes that you can teach your colleagues, explain to your clients and defend in court. For in-house lawyers, make sure your outside counsel understand them and that you apply them consistently throughout your company.

IV. Other Key Practice Points

Selection of Technology

E-discovery technologies vary in significant details. Make sure the tool you select has the ability to:

- Distinguish between near-native and actual native. For security and other purposes, your preference should be to avoid having to download native documents to the desktop.
- Identify embedded files and present them for review as part of compound family.
- For Bates numbering purposes, insert file ID numbers into native applications or create a separate load file reflecting these IDs.
• Review compound e-mails (parent and attachment) in context for relevance and to maintain their relationship for production purposes.

• Use trumping rules for compound e-mail families (i.e., if a parent is privileged and a child is responsive, the privileged parent trumps) to avoid inadvertent production of privileged files within a document family.

• Conduct keyword queries to ferret out potentially privileged and important information, showing the resulting hits in a “highlighted” mode.

Keys to Planning and Training

• Know your capabilities both to review and to produce before you go into any pretrial conference.

• Establish standard formats in which you request data and in which you offer to deliver data.

• Develop repeatable processes so you can advise clients of the likely results (time, expense, etc.) of your efforts.

• Expressly negotiate clawbacks for inadvertently produced privileged or personal data.

• Where possible, negotiate for the opportunity to supplement your request for additional data fields. (Review opposing parties’ data early to avoid being precluded from requesting additional production.)

• Train your team – have a process they can understand and successfully execute – and monitor their efforts.

• For outside counsel, advise clients and let them make business decisions in proportion to their risks.

• Attack the issue and convert it to an asset rather than a liability or an unknown.

V. Conclusion

Production and review of electronic documents in their native formats present both opportunities and dangers. In this new environment, lawyers should thoroughly understand production standards and strategies as they relate to litigation and regulatory settings. No set of best practices can substitute for judgment and experience. This white paper should serve as a guide, but it is up to you to carry appropriate methodologies and processes into practice.
About the Author

Jim Wagner is a co-founder and CEO of DiscoverReady LLC (www.discoverready.com), a national provider of discovery management services. He is a frequent public speaker on best practices for effectively gathering and reviewing electronic discovery.

Prior to forming DiscoverReady, Jim served as the Chief Operating Officer and a member of the Board of Directors of Document Technologies, Inc. (www.dtiglobal.com), a 1,000-employee integrated document solutions provider focused on the legal vertical. In his role as COO he was responsible for the day-to-day oversight of Document Technologies, Inc.’s national operations and the development and implementation of strategic initiatives. Significant accomplishments during Jim’s tenure with Document Technologies, Inc. include the development and oversight of LitWorks® (www.litworks.net)—the industry’s first and most comprehensive certified litigation support training program.

Jim began his legal career as an attorney in the corporate finance group of Powell Goldstein LLP. His general corporate practice focused on mergers and acquisitions, equity and debt financings, and the representation of private equity funds and high growth business ventures. From 1998 to 2000, his practice focused principally on the negotiation of international outsourcing transactions with the world’s largest telecommunications and electronics companies.

Before joining Powell Goldstein LLP, Jim worked for Glaxo Inc. in Research Triangle Park, North Carolina, as a tax associate and legal intern.

Jim was graduated with Honors from the Duke University School of Law and is a Phi Beta Kappa graduate of the University of Tulsa. Jim can be reached at jim.wagner@discoverready.com.