



©iStockphoto.com/alexandercreative

# The Case for Native Production

Most digital evidence now exists as data within databases rather than the static images provided by traditional paper productions. Counsel must therefore stop thinking about discovery as the quest for “documents” and start focusing on how to obtain information in useable and complete forms.



## CRAIG D. BALL

PRESIDENT  
CRAIG D. BALL, P.C.

Craig is a trial attorney, certified computer forensic examiner and e-discovery consultant. He has served as a court-appointed ESI special master over 50 times. Craig authors *Ball in Your Court*, an influential blog on e-discovery. He is also an adjunct professor teaching an e-discovery and digital evidence course at The University of Texas at Austin School of Law.

Historically, the law seldom addressed forms of production because there were few alternatives to paper. When evidence became digital, however, flat forms of information acquired new dimension and depth, described and supplemented by metadata (data about data) that supports a user’s ability to find, use and trust digital information.

Today, the form in which information is supplied determines if it is intelligible, functional and complete. For example, digital photographs hold metadata revealing where they were taken and by what camera, while spreadsheets carry formulae supporting complex calculations. Microsoft Word documents store editorial histories and are laced with conversations between collaborators. Presentations feature animated text and rich media, including sound and video, and dynamic connections to other data. Databases do not “store” documents, but rather assemble information on demand. Even conversations, once just fleeting interactions, now linger as text messages on cellular and internet networks.

Given this digital evolution, counsel must reframe their views on discovery and focus on the best ways to obtain information in useable and complete forms. Critical content is often removed

when parties make traditional document productions using page images instead of producing records in their native forms. Although native productions offer both parties distinct advantages, many attorneys have resisted them because they are accustomed to working with paper or imaged productions, are unfamiliar with the proper processes for reviewing and producing native files or misunderstand what forms of production are defensible under the Federal Rules of Civil Procedure (FRCP).

Against this backdrop, this article explores:

- The rules addressing forms of production.
- The options available to parties when making productions in response to document requests or subpoenas.
- The arguments favoring native productions and rebuttals to common objections.
- Best practices for drafting requests for production, specifically those seeking native files.

## APPLICABLE RULES

The FRCP provides certain protocols for selecting forms of production. In particular, the FRCP requires counsel to:

- Before the first pretrial conference, address issues related to the forms in which electronically stored information (ESI) should be produced and, where possible, reach agreement on the forms of production (*FRCP 26(f)(3)(C)*).
- When making requests for production, specify the form of production for each type of ESI sought, whether in paper, imaged, native or near-native form, or a mix of these forms (*FRCP 34(b)(1)(C)*).
- When responding to requests for production, either:
  - supply the specified forms in the production, if not objecting to those forms;
  - object and designate the forms the producing party intends to use; or
  - state the forms used in the production, if the requesting party failed to specify the forms sought.

(*FRCP 34(b)(2)(D)*); see also *FRCP 34(b) advisory committee's note (2006)* ("A party that responds to a discovery request by simply producing [ESI] in a form of its choice, without identifying that form in advance of the production ... runs a risk that the requesting party can show that the produced form is not reasonably usable and that it is entitled to production of some or all of the information in an additional form.")

- Meet and confer to resolve a forms dispute when the requesting party receives a production that does not comply with the forms designated (*FRCP 37(a)(1)*).

The FRCP contemplates that counsel will understand the forms of ESI in their cases and resolve forms disputes before requests for production are served. However, because the FRCP prescribes no deadline for a producing party to object to the forms sought, a requesting party sometimes learns that the producing party refuses to produce certain forms only after the producing party has incurred the cost of processing an unacceptable production.

To prevent this, a requesting party should supply a clear and practical written specification of the forms sought before the initial Rule 26(f) conference so that the producing party can assess the feasibility, cost and burden of producing in the specified forms. Even a party who does not know the forms in which an opponent's data natively resides can anticipate the most common forms, for example, for e-mail, word processed documents, presentations and spreadsheets.

A requesting party should not wait until the response date to determine if an opponent will fail or refuse to furnish the forms sought. Counsel should press for a commitment from the other side and, if forms are in dispute, file a motion to compel asking the court to select the forms to be produced.



Search [Document Requests: Initial Considerations](#) for more on document requests under FRCP 34.

## FORMS OF PRODUCTION

It is rarely necessary or advisable to employ a single form of production for all ESI produced in discovery. Counsel should tailor the forms sought to the data requested and may ask for a hybrid production combining different forms. This may include:

- Paper.
- Page images.
- Native files.
- Near-native files.

## PAPER PRODUCTION

Printing searchable ESI onto paper is usually an unnecessary and costly step backward. Still, paper productions are a reasonable choice and most useful when:

- The data to be produced is maintained in paper format.
- The volume of data is small.
- The ability to search electronically is not required.

## IMAGED PRODUCTION

In imaged productions, parties produce digital pictures of documents, e-mails and other electronic records, typically using Portable Document Format (PDF) or Tagged Image File Format (TIFF) images. Because converting an electronic record to a series of TIFF images strips the record of its electronic searchability and metadata, imaged productions are accompanied by load files holding searchable text and selected metadata (see *Box, Load Files*). Searchable text is obtained by extracting text from an electronic record or a scanned paper document using optical character recognition (OCR).

Imaged productions are often best suited to materials requiring redaction or those that include scanned paper records, microfilm or microfiche. However, these productions limit the discovery of certain information, including:

- ESI holding embedded information, such as collaborative content or formulae in spreadsheets.

## Load Files

Load files are composed of delimited text, which follows a predetermined sequence and is separated by characters like commas, tabs or quotation marks. Load files first appeared in discovery in the 1980s to add electronic searchability to scanned paper documents. Different review platforms use varying load file formats to order and separate information according to guidelines known as load file specifications.

When making or receiving an imaged production using TIFF images, the parties must negotiate the organization and substantive content of the load files to populate document review platforms. This includes whether they will be

pegged to review software like CT Summation, LexisNexis Concordance and Opticon, or Westlaw Case Logistix, and what information about the TIFF images will be provided, such as custodial and date information.

To understand the structure of a load file, consider a table created to keep track of documents. The table may use the first two columns to number the first and last page of each document. Subsequent columns may include additional information about the document, such as its name, modified date, author and file type.

For example:

BEGDOC	ENDDOC	FILENAME	MODDATE	AUTHOR	DOCTYPE
0000001	0000004	Contract	01/12/2013	J. Smith	docx
0000005	0000005	Memo	02/03/2013	R. Jones	docx
0000006	0000073	Taxes_2013	04/14/2013	H. Block	xlsx
0000074	0000089	Policy	05/25/2013	A. Dobey	pdf

The horizontal and vertical lines act as delimiters, delineating one field of data from the next. In place of lines, a computer interpreting a load file uses characters like commas or tabs. The same data as a load file may be delimited as follows:

```
BEGDOC,ENDDOC,FILENAME,MODDATE,AUTHOR,DOCTYPE
0000001,0000004,Contract,01/12/2013,J. Smith,docx
0000005,0000005,Memo,02/03/2013,R. Jones,docx
0000006,0000073,Taxes_2013,04/14/2013,H. Block,xlsx
0000074,0000089,Policy,05/25/2013,A. Dobey,pdf
```

Load files that use commas to separate values are called comma separated value or CSV files. More commonly, load

files adhere to formats compatible with the Concordance, Opticon or Summation review tools.

Overlay load files are used to update or correct existing database content by replacing data in fields in the order in which the records occur. It is crucial that the order of data within the overlay file match the order of data replaced. Data must be sorted in the same way, and the overlay must not add or omit fields.



Search [The Case for Native Production](#) for more on load files.

- Non-printable information, such as voicemails, videos, animation files or structured data.

### NATIVE PRODUCTION

Native productions are most beneficial when producing word processed documents, spreadsheets or electronic presentations. These productions require parties to produce the actual data files containing responsive information, such as producing Word documents in their native .DOC or .DOCX formats, Microsoft Excel spreadsheets as .XLS and .XLSX files and Microsoft PowerPoint presentations in native .PPT and .PPTX formats.

When employed by competent counsel using proper tools, native productions are almost always cheaper and faster than imaged productions because parties save on image conversion,

text extraction and OCR (see below *Advocating for Native Production*). Native productions also provide more information than other types of production. Using native review tools or applications similar to those used to create the data, requesting parties can see the evidence as it appeared to the producing party, with all the original embedded commentary and metadata. Additional benefits include:

- Easier deduplication.
- The ability to thread e-mail messages into conversations.
- Normalization of time zone irregularities.

However, the applications needed to view obscure data formats, such as specialized engineering applications or enterprise database software, may in some cases be prohibitively expensive. Further, counsel reviewing native productions must

take special care to avoid altering evidence. This often can be done by making copies of the data files and reviewing the copies, rather than the files themselves.

### NEAR-NATIVE PRODUCTION

Some ESI cannot feasibly or prudently be tendered in true native forms. Near-native files can preserve the essential utility, content and searchability of native files even though they are not, strictly speaking, native files. Some examples of ESI that make good candidates for near-native productions include:

- **Enterprise e-mail.** When messages are exported from a corporate Microsoft Exchange mail database to a container (which holds collections of e-mail as well as content supporting other Outlook features), the container is not native to the mail server. However, the container replicates the pertinent content and essential functionality of the mail server.
- **Database content.** Exports from databases are often produced in delimited formats like CSV files (where data values are separated by commas) that are not native to the database, but support the ability to interpret the data in ways faithful to the source.
- **Social networking content.** It is not possible to precisely duplicate the content from social networking sites like Facebook, where the content is stored in the cloud, so near-native forms seek to capture that data's essential utility, completeness and searchability by, for example, using HTML or PDF.

### ADVOCATING FOR NATIVE PRODUCTION

As discussed above, native files carry more information than their imaged counterparts and are inherently functional, searchable and complete. Moreover, native files are described by more and different metadata, which is invaluable in identifying, sorting and authenticating evidence. Yet many producing parties are reluctant to produce native or near-native files due to misconceptions about what the FRCP requires or because of how they have performed productions in the past.

### DISADVANTAGES OF IMAGED PRODUCTIONS

Converting a client's ESI from its native state as kept in its ordinary course of business to TIFF images adds needless expense in several ways. For example, it requires the producing party to:

- Pay for:
  - converting native forms to TIFF images;
  - stamping Bates numbers; and
  - generating load files.
- Produce multiple copies of documents like spreadsheets that are virtually incapable of being produced as images.
- Incur increased vendor costs to process and host TIFF images and load files because their file sizes are much larger (usually five to 40 times larger) than their native counterparts.
- Forgo full deduplication due to the difficulty and unreliability attendant to deduping page images.
- Reproduce some information, for example, if the receiving party requests embedded metadata that has been stripped from the native files.

### REBUTTING ARGUMENTS AGAINST NATIVE PRODUCTION

Producing parties usually assert certain justifications for refusing to produce ESI in native and near-native forms, specifically that these productions:

- Are not required by FRCP 34.
- Cannot be Bates labeled.
- Raise the risk that evidence will be altered.
- Require a broader review, for example, because native files include additional user-generated content.
- Limit the producing party's ability to redact privileged or otherwise protected information because the act of redacting alters the record.

None of these justifications withstands scrutiny.

### Interpreting FRCP 34

A producing party intent on an imaged production, despite a timely request for a native production, often justifies that choice by referencing its obligation to produce ESI in "the form or forms in which it is ordinarily maintained or in a reasonably usable form or forms" (FRCP 34(b)(2)(E)(iii)).

At least one court, however, has suggested that this reliance is misplaced, noting that a producing party may elect an imaged production "only if the requesting party declines to specify a form." Otherwise, files are "ordinarily maintained" in native format. (See *Anderson Living Trust v. WPX Energy Prod., LLC*, 298 F.R.D. 514, 525-26 (D.N.M. 2014) (noting that if the requesting party fails to designate a production format, the producing party has a choice under FRCP 34(b)(2)(E)(i) and (ii) to produce in native format, the form in which it is ordinarily maintained, or in a reasonably useful form or forms).)

### Bates Labeling Native Files

Making the transition to the more modern, native forms of production requires counsel to accept that printouts and images of ESI are not the same as ESI. Though a party cannot use Bates-style identifiers on discrete pages of a native file until it is printed or imaged, many native forms (for example, spreadsheets, social networking content, video and sound files) do not lend themselves to imaged formats and would not be Bates labeled in any event.

Further, most items produced in discovery are not used in proceedings. If and when Bates-style identifiers are needed on pages for later use at trial or in motion papers, a requesting party can simply require that file identifiers and page numbers be stamped on printouts or images. In practice, that impacts only a small subset of a production.

Finally, the name of an electronic file can be changed without altering its content, hash value (a unique signature or digital fingerprint on each electronic record) or last modified date. It is both simple and inexpensive to replace or append an incrementing Bates-style identifier to a filename. A producing party can even include a legend stating that the materials are subject to a protective order.

### Protecting Against Evidence Manipulation

Evidence tampering is not a new hazard in litigation or one unique to e-discovery. Though any form of production, hypothetically, is subject to manipulation by unscrupulous opponents (including imaged productions), native productions support quick, reliable ways to prevent and detect alteration. For example, producing native files on read-only media like CDs or DVDs guards against inadvertent changes. Alterations are easily detected by comparing hash values of suspect files to the files produced.

Counsel savvy enough to seek a native production should be savvy enough to both detect evidence manipulation and refrain from handling native files in a way that may alter that evidence.

### Reviewing All User-Generated Content

Native forms hold user-generated content, including collaborative comments and tracked changes in Word documents, animated text in presentations and formulae in spreadsheets. This content is not seen in an imaged production because the process of converting the file to TIFF format typically strips it of this information. Though rarely reviewed, this content is similar to Post-it notes and e-mails and may be equally (or more) responsive to requests as the rest of the file.

### Removing Privileged or Protected Content

Redaction necessarily involves a change to a file, namely to remove privileged or protected information while disclosing responsive information. The forms of production for items requiring redaction should be the forms best suited to efficient removal of privileged or protected content without rendering the remaining content wholly unusable.

Some native file formats support redaction well while others do not. Ultimately, the volume of redacted items in a production tends to be insignificant. Accordingly, the form selected for redaction should not dictate the broader forms of production when native forms have distinct advantages for the entirety of the production. Instead, a producing party may redact in an imaged format and then restore searchability after redaction, using native or near-native forms for the remainder of the production.

## DRAFTING REQUESTS FOR PRODUCTION

Although the rules permit requesting parties to specify the forms in which ESI should be produced (*FRCP 34(b)(1)(C)*), it is common for requests for production to be wholly silent on the forms of production, despite pages of detailed definitions and instructions.

To ensure that the requesting party receives the information sought, in the forms it desires, counsel should create clear, detailed requests by:

- Removing dated boilerplate definitions and requests.
- Specifying the precise forms of production sought.
- Negotiating the content of accompanying load files.
- Providing guidance on any deduplication or redaction procedures.
- Identifying the preferred production media.

### REMOVE UNNECESSARY LANGUAGE

The term “document” must be revised to an alternate term like “information items.” Instead of the usual thesaurus-like litany of types of information, counsel should consider using language such as the following:

“‘Information items’ as used here encompass individual documents and records (including associated metadata) whether on paper or film, as discrete files stored electronically, optically or magnetically, or as a record within a database, archive or container file. The term should be read broadly to include e-mail, messaging, word processed documents, digital presentations, spreadsheets and database content.”

Counsel should also remove redundant prose like “including, but not limited to” and “any and all.” These terms do not add clarity. Instead, when incorporating examples of responsive items in a request, counsel should add an instruction that:

“Examples of responsive items set out in any request should not be construed to limit the scope of the request.”

Similarly, rather than including “any and all” in a definition, counsel can add an instruction that:

“Requests for production should be read so as to encompass any and all items responsive to the request.”

Finally, before serving a request, counsel should check all included definitions to ensure that:

- Only those terms that are actually used are included.
- Terms in the request are used only in ways that are consistent with the definitions.

### SPECIFY THE FORMS SOUGHT

The most common error in requests for production is a party’s failure to specify the forms sought for ESI. A greater error is to employ legacy boilerplate specifying forms the requesting party does not want.

To guard against this, a request for production should specify the forms of production sensibly and precisely. Counsel should not assume that a reference to “native format” is clear or sufficient and should instead identify the exact formats sought for common file types. Further, requests or definitions can state that information that exists in electronic form should be produced in native or near-native formats and should not be converted to imaged formats.

Requesting parties can also specify that documents that do not exist in native electronic formats or require redaction should be produced in searchable PDF formats or as single page TIFF images, with unredacted OCR text furnished, and logical unitization and family relationships preserved.

Native format requires production in the same format in which the information was customarily created, used and stored in the ordinary course. Some examples of the native or near-native forms for specific types of ESI include:

- For Word documents, .DOC and .DOCX.
- For Excel Spreadsheets, .XLS and .XLSX.
- For PowerPoint Presentations, .PPT and .PPTX.
- For Microsoft Access Databases, .MDB and .ACCDB.
- For WordPerfect documents, .WPD.
- For Adobe Acrobat Documents, .PDF.
- For images, .JPG, .JPEG and .PNG.

Unless the entire content of a database is responsive, counsel should request that the producing party extract responsive content to a fielded and electronically searchable format that preserves metadata values, keys and field relationships. If doing so is infeasible, the producing party should identify the database and supply information concerning the schema and query language of the database along with a detailed description of its export capabilities to facilitate crafting a query to extract and export responsive data. For example, counsel may use the following language:

“E-mail messages should be produced in a form or forms that readily support import into standard e-mail client programs, or the form of production should adhere to the conventions set out in the internet e-mail standard. These may include, for example:

- For Outlook messages, .PST format. Single message production formats like .MSG or .EML may be furnished with folder data.
- For IBM Notes e-mail (formerly Lotus Notes), .NSF format or converted .PST format.

If it is necessary to extract attachments and produce them separately from their transmitting messages, they should be produced in their native forms with parent/child relationships to the message and containers preserved and produced in a delimited text file.”

When deciding on a form of production for e-mail, counsel may ask if the form produced can be imported into common e-mail client or server applications as a litmus test to distinguish native forms from less functional forms. If the form of the e-mail is so degraded that e-mail programs cannot recognize it as e-mail, there is a strong indication that the form of production has strayed too far from functional.

#### INDICATE THE LOAD FILE FORMATS

Every electronic file has system metadata that resides in the file table of the system or device storing the file. Different file types have different metadata. For example, every e-mail message has fields of information in the message header that support better searching, sorting and organization of messages. This data may be either responsive in its own right or simply advantageous to managing and authenticating electronic evidence. Either way, a requesting party should specify the metadata to be supplied in load files.

The parties should develop a comprehensive production protocol tailored to the case and serve it with any discovery requests. Simply demanding “the metadata” fails to identify exactly what counsel is seeking. Counsel should always specify,

in the protocol or in the requests themselves, the metadata and header fields sought.



Search [The Case for Native Production](#) for information on the metadata and header fields counsel should consider requesting.

#### CLARIFY DEDUPLICATION AND REDACTION PROCEDURES

The requesting party may wish to specify whether the production should be deduplicated.

Because redaction tends to impact only a small part of most productions, it is important that it does not override the requested form for the remainder of the production. Counsel should consider using the following instruction to make this clear:

“Information items that require redaction shall be produced in static image formats, e.g., single page TIFF or multipage PDF images with logical unitization preserved. The unredacted content of each document should be extracted by optical character recognition (OCR) or another suitable method to a searchable text file produced with the corresponding page image(s) or embedded within the image file. Redactions should not be accomplished in a manner that serves to downgrade the ability to electronically search the unredacted portions of the item.”



Search [The Case for Native Production](#) for more on TIFF redactions.

#### IDENTIFY THE MEDIUM FOR PRODUCTION

A well-crafted request should address the medium of ESI production or the means used to convey the electronic production to the requesting party. For example, if the production involves 100 gigabytes of data, the requesting party likely will not want it tendered on 143 CDs.

Counsel should also consider negotiating a hosted production. A hosted production permits the producing party to effect production using a secure website instead of via physical media like a DVD or hard drive. After the producing party has uploaded its responsive ESI to the website, the requesting party can access that data using a web browser and may search, view and annotate the data on the website, or download it.



Search [Document Requests: Drafting and Serving the Request](#) and [Document Requests: Drafting an RFP Checklist](#) for more on drafting and serving effective document requests.

Search [Document Requests: What to Expect in Response to an RFP](#) and [Document Requests: Common Problems with an RFP Response](#) for information on how a requesting party can determine whether all relevant information was produced by the producing party.