In the United States District Court
for the District of Maryland

PRINCIPLES FOR THE DISCOVERY OF
ELECTRONICALLY STORED INFORMATION IN CIVIL CASES

GENERAL PRINCIPLES

Principle 1.01 (Purpose)

Electronic discovery is now routinely encountered in civil litigation. At the same time, the Court is aware that the discovery of ESI is a potential source of cost, burden, and delay. The purpose of these ESI Principles is to encourage reasonable electronic discovery, in cases where it is appropriate to conduct such discovery, with the goal of reducing cost, burden, and delay and to "secure the just, speedy, and inexpensive determination of every action and proceeding" pursuant to Fed. R. Civ. P. 1. These ESI Principles also promote the avoidance or early resolution of disputes regarding the discovery of ESI without Court intervention. While parties are encouraged to discuss these ESI Principles in individual cases, compliance with them is voluntary and not required by the Court.

Principle 1.02 (Cooperation and Exchange of Information)

The Court recognizes the principles of The Sedona Conference® Cooperation Proclamation[1] and expects cooperation on issues relating to the preservation, collection, search, review, production, integrity, and authentication of ESI. The Court particularly emphasizes the importance, of cooperative exchanges of information about ESI at the earliest stages of litigation. An early exchange about ESI that will be relevant to the case may help ensure that conferences between the parties, as well as agreements between the parties, are meaningful.

[1] https://thesedonaconference.org/cooperation-proclamation

Each case is different, and the type of information exchanged should be tailored to best meet the needs of the case. Depending on the case, the parties may consider exchanging a data map (either in list form or visual) and information about the following types of technologies, systems, tools, or protocols as used by the parties: software applications or platforms, including databases; document management, mail, and messaging systems; types of computing devices (including portable computing and storage devices); use of home computers or personally-owned devices; the identity and rights of individuals to access the systems and specific files, services, and applications; network and database design and structure; use of cloud, off-site, or other third-party services, including social media and personal email; and backup and recovery routines, including backup media rotation practices. The parties may also consider exchanging organizational charts for key custodians of ESI and relevant policies, including those relating to computer usage, document management, ESI, or document retention or destruction.

Principle 1.03 (Proportionality)

The parties should apply the proportionality standard set forth in Fed. R. Civ. P. 26(b) to all phases of the discovery of ESI, including the identification, preservation, collection, search, review, and production of ESI while maintaining the integrity of the ESI. To assure reasonableness and proportionality in electronic discovery, parties should consider the factors described in Fed. R. Civ. P. 26(b). To facilitate adherence to the proportionality standard, requests for production of ESI and related responses should be prepared in consultation with custodians, IT custodians, and/or IT administrators so the resulting discovery is reasonably targeted, clear, complete, accurate, and as particularized as practicable.

ESI CASE MANAGEMENT PRINCIPLES

Principle 2.01 (Preservation of ESI)

a) Parties should take measures to preserve ESI as required by law. Parties should discuss preservation of ESI as early in the litigation as feasible. Such discussions should continue to occur periodically as the case and issues evolve.

b) In determining what ESI to preserve, parties should apply the proportionality standard referenced in Principle 1.03.

c) Parties are not required to use preservation notices to notify an opposing party of a preservation obligation, but if a party does so, the notice should apply the proportionality standard referenced in Principle 1.03 and be reasonably targeted, clear, complete, accurate, and as specific as practicable.

d) If there is a dispute concerning the scope of a party's preservation efforts, the parties should comply with the process outlined in Local Rule 104.7 and fully discuss the reasonableness and proportionality of the preservation. If the parties are unable to resolve a preservation issue, then the issue should be promptly raised with the Court.

e) Consistent with Proportionality Principle 1.03, the parties should discuss limiting the preservation, search, review, and production requirements imposed on each party by determining what ESI sources can be excluded from preservation and production because they are marginally relevant or not reasonably accessible.

Principle 2.02 (Conference of the Parties)

a) In cases involving ESI, a conference of the parties is helpful. Before such a conference, counsel should discuss who will participate with their clients and each other to ensure the participation of one or more persons for each party who are well-informed concerning the potentially relevant systems and data.

b) Topics the parties should be prepared to discuss include:

1) The sources, scope, and type of ESI that has been and will be preserved, including: date ranges; identity and number of potential custodians or sources; preservation and production by third parties in possession of relevant ESI, and their costs, capabilities, and policies; and other details that help clarify the scope of preservation;

2) The appropriate form and forms of production;

3) Any difficulties or exceptional costs related to preservation;

4) Search and culling methodologies (including keywords or technology assisted review, as appropriate) and suitable methods to query and produce responsive ESI;

5) The phasing of discovery, where appropriate, to prioritize discovery from custodians or sources most likely to contain discoverable information, including ESI, and those accessible at the lowest cost; and, as warranted, to defer or avoid discovery from sources unlikely to contain discoverable information or that are costliest to access;

6) The potential need for a protective order (see, e.g., Local Rule 104.13 and Appendix D), "clawback" agreement, and any procedure pursuant to Fed. R. Evid. 502(d) or (e), including a Rule 502(d) order; and

7) Opportunities to reduce costs and increase the efficiency and speed of the discovery process.

A more detailed checklist of information that may be helpful in guiding such discussions is included as Appendix 1: Suggested Topics for ESI Discussion. The Court encourages the parties to address any agreements or disagreements related to the above matters in the status report required by the scheduling order.

Principle 2.03 (E-Discovery Liaison)

In many cases, and where consistent with the proportionality factors in Rule 26(b), the discovery of ESI will be aided by the participation of electronic discovery liaisons. In addition, if a dispute arises that involves technical aspects of electronic discovery, as part of its obligations under Local Rule 104 concerning discovery disputes, each party should consider appointing an ESI liaison who will be well-informed concerning the relevant systems and information. An ESI liaison should be knowledgeable about the location, nature, accessibility, format, collection, searching, authenticity, integrity, and production of ESI in the matter. The ESI liaison should, at a minimum:

a) Be prepared to participate in the resolution of any discovery disputes relating to ESI so as to limit the need for Court intervention;

b) Be knowledgeable about the party's ESI discovery efforts;

c) Be familiar with, or gain knowledge about, the party's electronic systems and capabilities in order to explain those systems and answer related questions; and

d) Be familiar with, or gain knowledge about, the technical aspects of electronic discovery in the matter, including electronic document storage and organization, form/format issues, accessibility, and relevant information retrieval technology (including search methodology).

e) The failure to appoint an ESI liaison in a case where one is appropriate is one factor the Court may consider in granting relief in any discovery dispute or request for sanctions.

Principle 2.04 (Production of ESI)

a) Production Format: Production will be (1) in any form or forms agreed to by the parties, or (2) if no agreement is reached, in any reasonable form or forms specified by the requesting party if such format is consistent with Proportionality Principle 1.03, including native production. However, no party shall be compelled, except by Court order, to accept production in a form that substantially degrades or jeopardizes the utility, integrity, and/or authenticity of ESI. The parties may wish to discuss the use of a mutually accessible third-party service for the storage and sharing of discovery documents to minimize potential costs. Sample production protocols are attached as Appendix 2.

b) Privilege Logs: The parties should confer about the nature and scope of privilege logs for the case, including whether categories of information may be excluded from any logging requirements and whether an alternative to a document-by-document log will suffice.

c) The Discovery of Search Methodologies and Litigation Hold Material: Depending on the circumstances of a particular case, communications implementing or otherwise facilitating efforts to comply with the duty to preserve information, review for privileged information, or cull for responsive documents may or may not be protected from disclosure and discovery under Fed. R. Civ. P. 26. Unless the parties reach an agreement as to the production of this material, questions of discovery of this material are a matter of substantive law that will be decided on a case-by-case basis. Parties discussing these issues may wish to consider the use of a Fed. R. Evid. 502(d) order.

d) Metadata: Metadata is an important part of ESI and should be considered for production in every case. The production of metadata should be consistent with the proportionality principles of Fed. R. Civ. P. 26 and Principle 1.03. A detailed discussion of metadata can be found in Appendix 3: Metadata Reference Guide.

e) Cost-Shifting: Parties are generally responsible for their own costs of production of ESI. However, electronic discovery costs may be shifted in accordance with the applicable provisions of Fed. R. Civ. P. 26. Likewise, a party's nonresponsive or dilatory discovery tactics may prompt cost-shifting considerations. Cost-shifting can be negotiated by agreement of the parties or requested by appropriate motion to the Court.

f) Integrity of ESI: Parties should discuss how to produce the metadata and/or native files so that ESI maintains its integrity from when it is collected until when it is used in proceedings so that the parties have a method to confirm the integrity of the ESI throughout the litigation.

Principle 2.05 (Disputes Regarding ESI)

Disputes regarding ESI that the parties are unable to resolve shall be presented to the Court at the earliest reasonable opportunity. If the Court determines that any party or counsel has failed to cooperate and participate in good faith in electronic discovery or the Local Rule 104 process (including by the failure to appoint an ESI liaison under Principle 2.03, where appropriate), the Court may require additional discussions between the parties, order the appointment of an ESI liaison, and, if warranted, may consider discovery sanctions, including costs to the aggrieved party.

EXPECTATIONS OF COUNSEL

Principle 3.01 (Preparedness of Counsel)

It is expected that counsel for the parties, including all counsel who have appeared, as well as all others responsible for making representations to the Court or opposing counsel (whether or not they make an appearance), will be familiar with the following:

a. The electronic discovery provisions of the Federal Rules of Civil Procedure, including Rules 26, 33, 34, 37, and 45, and Federal Rule of Evidence 502;

b. The applicable rules of professional responsibility and other duties of counsel that are relevant to electronic discovery; and

c. The Local Rules and Discovery Guidelines (Appendix A) of this Court.

APPENDICES

Appendix 1: Suggested Topics for ESI Discussions

Early discussions are often helpful in cases involving ESI. Potential topics for the parties to discuss may, in the appropriate case, include the following, subject to the proportionality analysis contained in Rule 26 of the Federal Rules of Civil Procedure and Proportionality Principle 1.03:

Preservation

1. What are the key factual issues of the case?
2. What are the sources of potentially responsive ESI? Who are the custodians?
3. Can the custodians/sources be prioritized?
4. What are the date ranges for which data should be preserved?
5. Is an organizational chart encompassing the potentially responsive custodians available?
6. Is a data map encompassing the potentially responsive custodians available? What ESI sources exist from which data should be preserved? This could include, but not be limited to, data that is on premise, off-site and in the cloud; structured and unstructured data; network and standalone equipment; applications; removable storage; phones, tablets, mobile devices; social media; voice messaging; and instant messaging systems.
7. What repositories may contain relevant data, but are not reasonably accessible because of undue burden or cost? Will such repositories be preserved?
8. What repositories may contain relevant data, but will not be preserved?
9. What are each party's pertinent information management policies, computer usage policies, retention and destruction policies, "Bring Your Own Device" (BYOD) policies, and any other policies related to information management or governance?
10. Which non-custodial repositories should be preserved? Examples include department share drive, ShareFile locations, etc.
11. Has automatic deletion and purging of potentially responsive ESI been suspended?
12. What methodologies will be used to preserve and collect ESI? Will they account for chain of custody, integrity of ESI, and pertinent metadata and audit trail information?
13. Are there third parties who may possess potentially responsive ESI? If such third parties exist, how will that data be preserved?
14. Are there any disputes related to preservation that need to be presented to the Court for resolution?

Liaison

1. The parties should discuss whether each side will designate an ESI liaison for the duration of the litigation; and
2. If so, how they will be utilized.

Collection

1. What has been preserved; what will be collected?
2. How will it be collected?
3. How will it be processed?

4. Will phased collection and processing be efficient for the case?
5. Is there an agreement on a method for dealing with collection exceptions for which remediation is impossible or too costly?

Search

1. What methods of searching the data will be used to identify responsive ESI and filter out ESI that is not responsive?
2. Parties may discuss, if and as applicable, search and review methodologies and technologies.
3. Parties may discuss whether or not a search protocol should be presented to the Court for prior approval.

Production

1. In what forms and formats will ESI be produced, including decisions concerning:

    a. Which metadata fields, if any, will be provided;
    b. Whether OCR should be produced for non-text searchable files;
    c. The form and format of load files, if any, accompanying the production of documents;
    d. The naming conventions and Bates numbering of produced documents, including native files, full-text documents, OCRed documents and images;
    e. What, if any, files should be produced in native format;
    f. The image format, if any, to be produced;
    g. Whether the parties shall produce ESI in phases; and
    h. The media upon which the ESI productions will be delivered.

2. Are there any security or privacy issues applicable to any produced ESI?

Privilege

1. The parties should discuss a plan for dealing with privileged information, including obtaining an order from the Court pursuant to Fed. R. Evid. 502, if necessary.
2. The parties should discuss, if necessary, the production, exchange, and format of privilege logs.

Appendix 2: Sample Production Protocols

One of the easiest ways to minimize waste and unnecessary dispute is for parties to reach early agreement on the form or forms of production. Where the parties have not already agreed upon a production protocol, these sample production protocols are offered as a starting point for negotiation of the form or forms in which electronically stored information ("ESI") is exchanged. Any production protocol should be tailored to the needs of the parties and to the types of systems and data subject to discovery. If appropriate, the parties may discuss the procedure for maintaining the integrity of produced ESI throughout the litigation.

These sample protocols attempt to suggest best practices as of the writing of this appendix. As the types of ESI and the tools used to support electronic discovery evolve over time, so too must the manner in which ESI is produced. An overview of each sample is included below.

Appendix 2.1: Hybrid Production Protocol – This protocol permits the conversion of ESI to static image format. By creating a static image of each page, the parties are able to cite to a normalized representation of each page, aiding in creating a clearer record. Though searchability and application metadata is stripped away by image conversion, it is largely restored by the production of attendant extracted or OCR text and metadata in ancillary "load files." Imaged production protocols necessitate upfront expenditure to convert records, much of which may never be used in proceedings. Furthermore, the conversion of all produced ESI to image increases the size of the files ultimately exchanged, which has the potential to increase downstream processing and storage costs. To ameliorate some of these shortcomings, this hybrid production protocol provides for production of certain ESI in native formats, cross-referenced to Bates numbered image placeholders. This protocol assumes the parties have access to the resources and litigation support software required to generate and work with images and load files.

Appendix 2.2: Native Production Protocol – This protocol recognizes that conversion of ESI from its native format may impose an undue burden on the parties and may render the production less complete and usable. A native production permits technically-proficient parties to make more efficient use of the production and enables parties with limited resources to utilize low-cost and commonly-available tools to conduct search and review, eliminating the need to procure additional software required to pair images with text and metadata. Moreover, native productions offer greater flexibility, and because of their smaller size, native formats can reduce the cost to process and store data on a per-gigabyte basis. For use in proceedings, parties may wish to convert selected native documents to static images or present the information digitally. In the case of the former, the parties may consider reaching agreement on the procedure for stipulation to the image format.

Appendix 2.1
Sample HYBRID PRODUCTION PROTOCOL

1. "Information items" as used here encompasses individual documents and records (including associated metadata), whether on paper, 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, text messages, word processed documents, digital presentations, social media posts, webpages, and spreadsheets.

2. Responsive electronically stored information ("ESI") (except for spreadsheets, presentation files, or other information items containing speaker notes, animated text, embedded comments, or tracked changes) should be converted to image, Bates numbered, and produced with fully searchable text. A single-page TIFF placeholder bearing the Bates number for each record not converted to image shall also be produced. This Protocol describes the specifications for producing hybrid productions and attendant load files.

3. Images

    a. Images should be single-page, Group IV TIFF files, scanned at 300 dpi.
    b. File names cannot contain embedded spaces.
    c. The number of TIFF files per folder should not exceed 2,000.
    d. If an information item contains color, it shall be produced in color, unless the color is merely decorative (*e.g.*, company logo or signature block).

4. Image Cross-Reference File

    A comma-delimited image cross-reference file (*e.g.*, .OPT or .LFP) to link the images to the metadata and text should be supplied. Such a cross-reference file typically consists of nine fields per line, with a line for every file in the database.

    For example, the .OPT format is as follows:

```
ABC00000001,VOL0001,\IMAGES\0001\ABC00000001.TIF,Y,,,4
ABC00000002,VOL0001,\IMAGES\0001\ABC00000002.TIF,,,,
ABC00000003,VOL0001,\IMAGES\0001\ABC00000003.TIF,,,,
ABC00000004,VOL0001,\IMAGES\0001\ABC00000004.TIF,,,,
```

5. Text

    Searchable text of the entire document must be provided for every record, at the document level.

    a. Searchable text must be provided for all documents that originated in electronic format but are not produced in their native forms. Text files should include page breaks that correspond to the pagination of the image files. Any document in which text cannot be extracted must be processed using optical character recognition (OCR), including PDFs without embedded text.

b. OCR text must be provided for all documents that originated in hard copy format. A page marker should be placed at the beginning, or end, of each page of text, *e.g.*, *** IMG0000001 *** whenever possible. The data surrounded by asterisks is the ImageID.

c. For redacted documents, provide the full text for the redacted version.

d. Text should be delivered as multi-page ASCII text files with the files named to conform to the ImageID field. Text files should be placed in separate subfolders with each subfolder limited to 500 files.

6. Data File

The data file (*e.g.*, .DAT or .CSV) is another delimited file containing all of the fielded information and associated metadata for each information item produced.

a. The first line of the data file must be a header row identifying the field names.

b. Date fields should be provided in the format: MM/DD/YYYY.

c. All family relationships should be preserved, and all attachments should sequentially follow the parent document/email.

d. All metadata associated with email, audio, and native electronic document collections must be produced per the table below.

e. In some cases, it may be appropriate to specify the data file delimiters for certain litigation support systems. For example, default .DAT file delimiters for Concordance are:

Comma	,	ASCII character (020)
Quote	þ	ASCII character (254)
Newline	®	ASCII character (174)

The text and metadata of email and attachments, and all other native file document collections, should be extracted and provided in a data file using the field definition and formatting described below:

Field Position	Field Name	Type	Description/Metadata
1.	**BEGDOC**	Paragraph	Beginning bates number
2.	**ENDDOC**	Paragraph	Ending bates number
3.	**BEGATTACH**	Paragraph	Beginning bates number of family
4.	**ENDATTACH**	Paragraph	Ending bates number of family
5.	**ATTCOUNT**	Paragraph	Attachment count

Field Position	Field Name	Type	Description/Metadata
6.	**PARENTID**	Paragraph	Bates number of family parent
7.	**DOCDATE**	Date	Date of document or creation date (MM/DD/YYYY)
8.	**DATESENT**	Date	Date Email Sent (MM/DD/YYYY)
9.	**TIMESENT**	Time	Time Email Sent (HH:MM:SS AM/PM)
10.	**DATERECEIVED**	Date	Date Email Received (MM/DD/YYYY)
11.	**TIMERECEIVED**	Time	Time Email Received (HH:MM:SS AM/PM)
12.	**TIMEZONE**	Paragraph	Time zone used to process custodian data
13.	**AUTHOR**	Paragraph	Who created document (LASTNAME, FIRST)
14.	**FROM**	Paragraph	Who is document sent from (LASTNAME, FIRST)
15.	**TO**	Paragraph	Who is document sent to (LASTNAME, FIRST)
16.	**CC**	Paragraph	Who is copied on document (LASTNAME, FIRST)
17.	**BCC**	Paragraph	Who is blind copied on document (LASTNAME, FIRST)
18.	**DOCTYPE**	Paragraph	What type of document this is (*e.g.*, Message or attachment)
19.	**FILEEXT**	Paragraph	File Extension (e.g., .msg or .doc)
20.	**EMAILSUBJECT**	Paragraph	Email subject line
21.	**EMAIL MESSAGE ID**	Paragraph	Message ID for email
22.	**FILENAME**	Paragraph	Original file name
23.	**LASTMOD**	Date	Date last modified (MM/DD/YYYY)
24.	**CUSTODIAN**	Paragraph	Custodian (LASTNAME, FIRST)
25.	**SOURCE**	Paragraph	Where did document come from?
26.	**ORIGFOLDER**	Paragraph	Original file folder (*e.g.*, Personal Folders\Deleted Items\)
27.	**PAGES**	Number	Number of pages in document
28.	**DOCLINK**	Paragraph	This will be used if there is a native, path to folder where data LINK record is located
29.	**HASH**	Paragraph	MD5 or SHA Hash Value (unique file signature)
30.	**HASH DE-DUPLICATE**	Paragraph	Instances of hash de-duplication (by full path)

Field Position	Field Name	Type	Description/Metadata
	INSTANCES		
31.	**CONVERSATION INDEX ID**	Paragraph	Microsoft Conversation index number generated by Microsoft Outlook to identify email conversations.

7. Linked Native Files

Spreadsheets must be produced in their native electronic formats. Also, Microsoft Office files, or other information items containing speaker notes, animated text, embedded comments, or tracked changes must be produced in their native electronic formats.

a. Native file documents must be named per the BEGDOC (beginning bates number).
b. The full path of the native file must be provided in the .data file for the DOCLINK field.
c. The number of native files per folder should not exceed 2,000 files.

8. Image Handling

For any records converted to image, the following settings should be applied at conversion.

Microsoft Word		
Option	**Setting**	**Description**
Show Track Changes	Yes/No	If yes, 'Final Showing Markup' will be used. If not, 'Final' view will be used.
Show Hidden Text	Yes/No	If yes, text marked as hidden will be printed.
Show Comments	Yes/No	If yes, comments will be printed.
Print Headers	Yes/No	If yes, headers will be printed.
Print Footers	Yes/No	If yes, footers will be printed.
Print Field Codes	Yes/No	If not yes, fields containing PRINT code are cleared to prevent output TIFF corruption.
Use SavedDate Instead of CurrentDate	Yes/No	Any auto date/time fields will be replaced with Saved Date/Time instead of current date.
Use Filename Only for Auto Filename Fields	Yes/No	If yes, any auto filename fields will be printed with just the filename, not the path.
Disable Auto Hyphenation	Yes/No	If yes, auto hyphenation will not be used for foreign language docs.
Microsoft Excel		
Option	**Setting**	**Description**
Unhide Columns	Yes/No	If yes, all hidden columns will be printed.
Unhide Rows	Yes/No	If yes, all hidden rows will be printed.
Unhide Worksheets	Yes/No	If yes, all hidden worksheets will be printed.

Unhide Charts	Yes/No	If yes, all hidden charts will be printed.
Print Order	Over Then Down	This is the order that excel pages are printed.
Print Orientation	Portrait/Landscape	This will enforce the print orientation to portrait or landscape.
Paper Size	Letter/Legal	This will force the paper size to letter or legal.
Print Comments	None	Choose where to print comments on the converted image.
Unhide Formulas	Hidden/Visible	If set to Hidden, the cell values will be displayed. If set to Visible, formulas will be displayed.
Set Scaling to Fit	Yes/No	If yes, the width of the Excel file will be squeezed to fit on one page.
Autofit Column and Row Sizes	Yes/No	If yes, height and width is increased to fit contents.
Disable Custom Filters	Yes/No	If yes, custom filters are disabled.
Black Font	Yes/No	If yes, font color of all cells is set to black so that content is displayed.
Reset Print Area	Yes/No	If yes, the print area is reset.
Set Header Margin	0.5	Top margin is checked and adjusted to prevent truncation.
Margin Handling Header	Keep Offset	Define how the margin of the header is calculated.
Set Footer Margin	0.5	Bottom margin is checked and adjusted to prevent truncation.
Margin Handling Footer	Keep Offset	Define how the margin of the footer is calculated.
Use Filename Only For Auto Filename Fields	Yes/No	If yes, auto filename fields will be printed with just the filename, not the path.
Show Auto File Name	Yes/No	If yes, the English code will be shown, not the value.
Show Auto Date	Yes/No	If yes, the English code will be shown, not the value.
Show Auto Time	Yes/No	If yes, the English code will be shown, not the value.
Limit Output to ### Pages	250	The output for each file will be limited to the given number of pages (0 means no limitation)
Microsoft PowerPoint		
Option	**Setting**	**Description**
Print Hidden Slides	Yes/No	If yes, all hidden slides will be printed.
Scale to Fit the Paper	Yes/No	If yes, the converted slide will be scaled to fit the page.
Print Comments	Yes/No	If yes, comments will be printed.
Print Type	Unchanged	Number of slides per page. Notes page will print both the slide and the notes on the same page.
Print Notes at End	Yes/No	If yes, all notes will be displayed at the end of the document.
Use Default Theme	Yes/No	Default theme can be used to display text that will not print because it blends within the image.

Appendix 2.2
Sample NATIVE FORMAT PRODUCTION PROTOCOL

1. "Information items" as used here encompasses individual documents and records (including associated metadata), whether on paper, as discrete "files" stored electronically, optically or magnetically, or as a database, archive, or container file. The term should be read broadly to include all forms of electronically stored information (ESI), including but not limited to e-mail, messaging, word processed documents, digital presentations, social media posts, webpages, and spreadsheets.

2. Responsive ESI shall be produced in its native form; that is, in the form in which the information was created, used, and stored by the native application employed by the producing party in the ordinary course of business.

3. If it is infeasible or unduly burdensome to produce an item of responsive ESI in its native form, it may be produced in an agreed upon near-native form; that is, in a form in which the item can be imported into an application without a material loss of content, structure, or functionality as compared to the native form. Static image production formats serve as near-native alternatives only for information items that are natively static images (*i.e.*, faxes and scans).

4. Examples of agreed-upon native or near-native forms in which specific types of ESI should be produced are:

Source ESI	Native or Near-Native Form or Forms Sought
Microsoft Word documents	.DOC, .DOCX
Microsoft Excel spreadsheets	.XLS, .XLSX
Microsoft PowerPoint presentations	.PPT, .PPTX
Microsoft Access Databases	.MDB, .ACCDB
WordPerfect documents	.WPD
Adobe Acrobat documents	.PDF
Photographs	.JPG, .PDF
E-mail	.PST, .MSG, .EML [1]
Webpages	.HTML

[1] Messages should be produced in a form or forms that readily support import into standard e-mail client programs; that is, the form of production should adhere to the conventions set out in RFC 5322 (the Internet e-mail standard). For Microsoft Exchange or Outlook messaging, .PST format will suffice. Single message production formats like .MSG or .EML may be furnished if source foldering metadata is preserved and produced (*see* paragraph 13). For Lotus Notes mail, furnish .NSF files or convert messages to .PST. If your workflow requires that attachments be extracted and produced separately from transmitting messages, attachments should be produced in their native forms with parent/child relationships to the message and container(s) preserved and produced in a delimited text file.

5. Where feasible, when a party produces reports from databases that can be generated in the ordinary course of business (*i.e.*, without specialized programming skills), these shall be produced in a delimited electronic format preserving field and record structures and names. The parties will meet and confer regarding programmatic database productions, as necessary.

6. Information items that are paper documents or that require redaction shall be produced in static image formats, *e.g.*, single-page .TIF or multipage .PDF images. If an information item contains color, it shall be produced in color unless the color is merely decorative (*e.g.*, company logo or signature block).

7. Individual information items requiring redaction shall (as feasible) be redacted natively or produced in .PDF or .TIF format and redacted in a manner that does not downgrade the ability to electronically search the unredacted portions of the item. The unredacted content of each redacted document should be extracted by optical character recognition (OCR) or other suitable method to a searchable text file produced with the corresponding page image(s) or embedded within the image file. Parties shall take reasonable steps to ensure that text extraction methods produce usable, accurate and complete searchable text.

8. Except as set out in this Protocol, a party need not produce identical information items in more than one form and may globally deduplicate identical items across custodians using each document's unique MD5 or other mutually agreeable hash value. The content, metadata, and utility of an information item shall all be considered in determining whether information items are identical, and items reflecting different information shall not be deemed identical. Parties may need to negotiate alternate hashing protocols for items (like e-mail) that do not lend themselves to simple hash deduplication.

9. Production should be made using commercially reasonable electronic media of the producing party's choosing, provided that the production media chosen not impose an undue burden or expense upon a recipient.

10. Each information item produced shall be identified by naming the item to correspond to a Bates identifier according to the following protocol:

    a. The first four (4) or more characters of the filename will reflect a unique alphanumeric designation identifying the party making production.
    b. The next nine (9) characters will be a unique, consecutive numeric value assigned to the item by the producing party. This value shall be padded with leading zeroes as needed to preserve its length.
    c. The final six (6) characters are reserved to a sequence beginning with a dash (-) followed by a four (4) or five (5) digit number reflecting pagination of the item when printed to paper or converted to an image format for use in proceedings or when attached as exhibits to pleadings.
    d. By way of example, a Microsoft Word document produced by ABC Corporation in its native format might be named: ABCC000000123.docx. Were the document printed out

for use in deposition, page six of the printed item must be embossed with the unique identifier ABCC000000123-00006.

11. Information items designated "Confidential" may, at the Producing Party’s option:

a. Be separately produced on electronic production media or in a folder prominently labeled to comply with the requirements of paragraph __ of the Protective Order entered in this matter; or, alternatively,
b. Each such designated information item shall have appended to the file’s name (immediately following its Bates identifier) the following protective legend: ~CONFIDENTIAL-SUBJ TO PROTECTIVE ORDER IN CAUSE MDL-13-0123.

When any “Confidential” item is converted to a printed or imaged format for use in any submission or proceeding, the printout or page image shall bear the protective legend on each page in a clear and conspicuous manner, but not so as to obscure content.

12. The producing party shall furnish a delimited load file supplying the metadata field values listed below for each information item produced (to the extent the values exist and as applicable):

Field BeginBates
EndBates
BeginAttach
EndAttach
Custodian/Source
Source File Name
Source File Path
From/Author
To
CC
BCC
Date Sent
Time Sent
Subject/Title
Last Modified Date
Last Modified Time
Document Type
Redacted Flag (yes/no)
Hidden Content/Embedded Objects Flag (yes/no)
Confidential flag (yes/no)
E-mail Message ID
E-mail Conversation Index
Parent ID
MD5 or other mutually agreeable hash value
Hash De-Duplicated Instances (by full path)

13. Each production should include a cross-reference load file that correlates the various files, images, metadata field values and searchable text produced.

Questions and Answers about the Native Production Protocol

Q. If our company used a PDF or TIFF file in the ordinary course of business, do we have to convert that to some "native" form?

A. No, if the information item originated natively in the usual course of business (such as by scanning a paper document to PDF or a receiving a fax as a TIFF image), those forms are the native forms and should not be converted to another form.

Q. If we have a printout of a document and an electronic version that we think is the file used to create the printout, do we have to deduplicate them? Which do we produce?

A. No, this protocol recognizes that they are not the same. The electronic file holds more information than the printed page (*e.g.,* comments and application metadata) and the printout may reflect different information (*e.g.*, signatures, highlighting, and margin notes). Furthermore, the electronic version is inherently searchable and sortable by metadata, where the paper document is not. If responsive, you produce both, as they are not identical under the protocol.

Q. So, what items are identical and must be deduplicated?

A. Only items with matching hash values are deemed sufficiently identical that just one instance need be produced. If you have been deduplicating in other matters or producing as TIFF images and load files, computing and matching hash values is something you already do. If not, it's a very low-cost undertaking that saves a lot of wasted effort and money.

Q. Won't it cost more to produce in native and near-native forms?

A. No. The forms of production in this protocol require considerably fewer steps because there is no need to convert the items from the forms in which the parties use and store them in the ordinary course of business to other, less utile and complete forms. Further, producing in native and near-native forms minimizes the expensive and error-prone processes of extracting searchable text and converting it to images. Especially with Microsoft Office productivity formats (Excel, Word, and PowerPoint documents), conversion to image formats significantly downgrades utility and completeness of the evidence.

Q. But won't we lose the ability to Bates number production? I want my Bates numbers!

A. Not at all. Electronic productions are "Bates numbered" consecutively, and when items are printed out or imaged for use in proceedings or as exhibits, they will bear embossed Bates numbers, page numbers, and protective legends, just as they always have. What changes is that you don't have to emboss all that on each page until you actually need that information in a paginated format. Still, the electronic forms always carry a Bates number (in their file name) and even a protective legend for items designated "confidential." It's a little different than paper, but then, ESI is a lot different than paper. This protocol saves a great deal of money without adding complexity, so the difference is a change for the better.

Q. Footnote 1 states: "[T]he form of production [for e-mail] should adhere to the conventions set out in RFC 5322." What does that mean?

A. It's just a shorthand way to tell your technical people they shouldn't downgrade the e-mail for production. RFC 5322 is the current international Internet standard that sets out what needs to be present in an e-mail for it to be complete and functional. By using any of the everyday forms of e-mail that are RFC 5322-compliant (*e.g.*, PST, MSG, EML, EMLX, MBOX, etc.), you will be preserving the content and structure of the e-mail that allows it to be reviewed in any of the tools that support e-mail, including all major e-discovery platforms. These forms afford the parties maximum flexibility at lowest cost. Plus, they are less costly because they come straight out of the mail servers and archives in RFC 5322-compliant formats. Conversion to TIFF and load files requires costly parsing and processing of e-mail contents with the result that, *e.g.*, message header values needed for threading conversations and message IDs helpful to deduplication are lost or corrupted. Moreover, family relationships between messages and attachments that support efficient review are often lost or misplaced. Trying to dissect and rebuild e-mail messages as TIFF images and load file data often leads to contentious motions, expensive experts, and sanctions, all of which could have been avoided by sticking to the forms e-mails are intended to take.

Q. Why do we have to extract searchable text and embedded metadata values from native and near-native files?

A. You don't. Unlike TIFF images, native and near-native forms are inherently electronically searchable and carry application metadata within the files. So there's no need to extract text for search as it's already in the file produced. The metadata production requirement speaks to production of fields "as applicable." If the metadata is in the file produced, extracting the same data to a load file is redundant and, accordingly, not "applicable."

Q. Our lawyers don't have the tools to review native forms. Their review tools are pretty old and only support review of TIFF images. What do they do?

A. They can keep on using their tools. Native and near-native forms are easily downgraded to forms that lawyers with older tools can manage. That's what they've been doing and one reason why e-discovery has been so costly. Any party who needs downgraded forms of production can go on paying to convert the data for their use. This protocol serves to eliminate that cost and hardship to those capable of dealing with the evidence in the same forms in which the witnesses and parties do. If you don't mind the higher cost, use any old tool you want to review; just *produce* in native and near-native forms.

Q. We want to produce on CDs. Is that an "appropriate" medium of production?

A. That depends upon the volume of data you're producing. If your production can fit on 2-3 CDs, it's appropriate. If your production will span 20 CDs, it's a waste of everyone's time and money to spend hours extracting from 20 CDs what would have taken minutes to pull from a ten buck thumb drive.

Q. We prefer to produce as TIFF images because then no one can see the hidden metadata—like collaborative comments, speaker notes, formulas, tracked changes, and such. Isn't that just metadata?

A. The information listed is user-generated content, and dismissing it as "just metadata" doesn't justify its eradication. It is evidence, like margin notes on paper documents and comments written on Post-Its. If you've been ignoring it without consequence, consider yourself lucky. This protocol treats it as part and parcel of the ESI to be produced.

Q. If we don't convert everything to TIFF or PDF, what will prevent you from changing the evidence? Aren't TIFF and PDF images harder to alter than native forms?

A. Nothing prevents a dishonest litigant from seeking to change the evidence, save the certainty that any change important enough to impact the outcome of a case will be checked against the source and exposed. Because of the ability to digitally fingerprint or "hash" native and near-native productions, it's far easier to quickly and reliably detect alterations. Contrary to popular misconceptions, it's simple to alter TIFF and PDF files in ways that are difficult for a reader to detect. Adobe Acrobat has supported extensive editing of PDF files for years. TIFF images are just pictures, so can be modified using the same off-the-shelf tools used to enhance snapshots. It's an urban myth that producing TIFFs and PDFs is more secure.

Q. Why must MD5 hashes of each production item be furnished?

A. Though parties are free to negotiate an agreement to produce alternate metadata, parties are cautioned to always calculate, supply, and preserve the hash value of each electronic information item produced as a simple and reliable method by which to ascertain if an item has been inadvertently or deliberately altered following production.

Appendix 3: Metadata Reference Guide

Metadata is information that helps us use and make sense of other information. More particularly, metadata is information, typically stored electronically, that describes the characteristics, origins, usage, structure, alteration, and validity of other electronically stored information ("ESI"). Metadata occurs in many forms within and without digital files. Some is supplied by the user, but most metadata is generated by systems and software.

Some define metadata simply as "data about data," where others characterize metadata as data that is not user-generated but is created by a computer system or application to keep track of a file's attributes. However, even user-generated data may qualify as metadata. For example, a Bates number is metadata, although assigned by counsel.

Because metadata is defined so broadly, a blanket request for the production of metadata may be unhelpful. The metadata values associated with a particular file or information item vary according to the nature of the item and its use. For example, the relevant metadata from a word processed document differs from e-mail metadata and from metadata pertinent to a database.

Metadata is unlike almost any other discoverable information because its import may flow from its probative value as relevant evidence, its utility in functionally abetting the searching, sorting, and interpretation of ESI, or both. If the origin, use, distribution, destruction, or integrity of electronic evidence is at issue, the relevant "digital DNA" of metadata is probative evidence that should be preserved and produced. Likewise, if the metadata materially facilitates the searching, sorting, and management of ESI, it should be preserved and produced for its utility.

Absent a specific agreement between parties or instruction from the Court as to the form or forms of production, parties typically produce information in the form or forms the information is ordinarily maintained or in some other reasonably usable form. In determining what form or forms to produce data, a producing party should take into account the need to make metadata as accessible both to display and to search, for the receiving party as it is to the producing party, where appropriate and necessary, after consideration of proportionality factors outlined in Principle 1.03.

Metadata can be generally categorized as System Metadata or Application Metadata.

System Metadata reflects context, being information about a file that is not embedded within the file it describes, but is stored externally by the computer's file management system, which uses system metadata to track file locations and store demographics about each file, *e.g.*, file name, size, creation, modification, and usage. System metadata may be crucial to electronic discovery because so much of our ability to identify, find, sort, and cull information depends on its system metadata values. For example, system metadata helps identify the custodians of files, when files were created or altered, and the folders in which they were stored.

Other metadata, called Application Metadata, reflects content. It is information that the software application creates and stores within the file. As an example, Microsoft Word stores the date when a document was last printed and the time expended editing the document.

The following are suggestions for producing different types of metadata.

1. Application metadata is, by definition, embedded within native files; so native production of ESI obviates the need to selectively preserve or produce application metadata. When ESI is converted to other forms for production, the producing party should assess what metadata will be lost or corrupted by conversion and identify, preserve, and extract relevant or useful application metadata fields for production. The extracted metadata is produced in ancillary production formats called "load files," designed to be ingested by tools used to review electronic documents. Not all metadata lends itself to production in load files because some metadata (like tracked changes in a Word document) must be seen in context within the native application or an e-discovery review platform.

2. For e-mail messages, this is a fairly straightforward process, notwithstanding the dozens of metadata values that may be introduced by e-mail client and server applications. The metadata essentials for e-mail messages are typically:

   - Custodian – Owner of the mail container file or account collected;
   - To – Addressee(s) of the message;
   - From – The e-mail address of the person sending the message;
   - CC – Person(s) copied on the message;
   - BCC – Person(s) blind copied on the message;
   - Date Sent – Date the message was sent;
   - Time Sent – Time the message was sent with UTC/UMG offset;
   - Subject – Subject line of the message;
   - Date Received – Date the message was received;
   - Time Received – Time the message was received;
   - Attachments – Name(s) or other unique identifier(s) of attachments;
   - Mail Folder Path – Path of the message from the root folder to the mail folder (to permit the threading of messages as a "conversation");
   - Message ID – Microsoft Outlook or similar unique message identifier; and
   - In-Reply-To – Microsoft Outlook or similar unique message identifier.

3. Other Mail Metadata: E-mail messages that traverse the Internet contain so-called "header data" detailing the routing and other information about message transit and delivery. Header data may be useful to address questions concerning authenticity, receipt, or timing of messages. Certain header values are essential to support the ability to thread messages into intelligible conversations. Metadata essentials may also include metadata values generated by the discovery and production process itself,

such as Bates numbers and ranges, hash values, production paths, extracted or OCR text, family designations, and time zone offset values.

4. The system metadata values that should typically be considered for preservation and production include:

   - File name;
   - File size;
   - File path;
   - Last modified date and time; and
   - Source or custodian.

5. Parties should discuss the production of metadata at an early practicable stage in the litigation and use proportionality principles in determining the scope of such production. The fields of metadata to be produced, if any, and the form(s) of production should be addressed by the parties and memorialized in a written agreement.