N. Scott Sacks, Attorney (D.C. Bar No. 913087)
Jessica N. Butler-Arkow, Attorney (D.C. Bar No. 430022)
Anna T. Pletcher, Attorney (California Bar No. 239730)
Adam Severt, Attorney (Member, Maryland Bar, numbers not assigned)
Ryan Struve, Attorney (D.C. Bar No. 495406)
Shane Wagman, Attorney (California Bar No. 283503)
United States Department of Justice
Antitrust Division
450 Fifth Street NW, Suite 7100
Washington, DC 20530
Telephone: 202-307-6200
Facsimile: 202-616-8544
Email: scott.sacks@usdoj.gov

Attorneys for Plaintiff United States of America

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

UNITED STATES OF AMERICA,
 	
                         Plaintiff,
 	
v.
 	
EBAY INC.
 	
                         Defendant.
	Case No. 12-CV-05869-EJD
	ATTACHMENT B TO JOINT CASE MANAGEMENT STATEMENT AND [PROPOSED] ORDER: DOJ STANDARD SPECIFICATIONS FOR PRODUCTION OF ESI

U.S. Department of Justice Antitrust Division
Standard Specifications for
Production of ESI and/or Hard Copy as Images and Text

Introduction:

This document describes the standard specifications and procedures for making an image-based production to the Antitrust Division in the form of load files**.**

- To ensure the efficient processing and review of any electronic production, Division legal, economic, and technical staff need to resolve the details prior to production, and preferably before you or your vendor begin to gather and process responsive documents.
- Care should be taken to ensure that all responsive data and metadata are preserved in the collection process.
- These are not Unicode compliant specifications and do not cover production of translations. Please contact Division staff if either of these is anticipated.

A. Categories of Documents

Discussion regarding the details of an electronic production should focus on seven categories of documents: (1) email and other electronic messages (*e.g.,* instant messaging), (2) other electronic documents, (3) hard copy documents, (4) shared resources, (5) databases, (6) audio and video data, and (7) foreign-language materials. General requirements for each category of document are outlined below. For information regarding document-specific metadata and bibliographic information (identifying information), please refer to the enclosed Metadata Table of Requested Fields.

1. Email, Attachments, and Other Electronic Messages

Email and other electronic messages (*e.g.*, instant messages (IMs)) should be produced as image files with related searchable text, metadata and bibliographic information. Depending on how the company's systems represent names in email messages or IMs, we may require a table of names or contact lists from custodians.

Email repositories, also known as email databases (*e.g.*, Outlook .PST, Lotus .NSF), can contain a variety of items, including messages, calendars, contacts, tasks, etc. For purposes of production, responsive items should include the "Email" metadata/database fields outlined in the Metadata Table, including but not limited to all parent items (mail, calendar, contacts, tasks, notes, etc.) and child files (attachments of files to email or other items), with the parent/child relationship preserved. Similar items found and collected outside an email repository (*e.g.*, .MSG, .EML, .HTM, .MHT) should be produced in the same manner.

Each IM conversation should be produced as one document.

   A. Attachments. Pay special attention to the PARENTID and ATTACHMENTIDS fields, which are used to track email families. While this example relates to email families, all attachment relationships for all responsive documents are to be produced in this format.

EXAMPLE: Consider ABC-JD-00000001 a 10-page parent email, with records ABC-JD-00000011 to ABC-JD-00000015, ABC-JD-00000016 to ABC-JD-00000020, and ABC-JD-00000021 to ABC-JD-00000025 as its attachments. Fields should be populated exactly as follows using the semicolon as the multi-entry delimiter for ATTACHMENTIDS:

BEGDOC#	ENDDOC#	PARENTID	ATTACHMENTIDS	FAMILYRANGE
ABC-JD-00000001	ABC-JD-00000010		ABC-JD-00000011; ABC-JD-00000016; ABC-JD-00000021	ABC-JD-00000001 – ABC-JD-00000025
ABC-JD-00000011	ABC-JD-00000015	ABC-JD-00000001		ABC-JD-00000001 – ABC-JD-00000025
ABC-JD-00000016	ABC-JD-00000020	ABC-JD-00000001		ABC-JD-00000001 – ABC-JD-00000025
ABC-JD-00000021	ABC-JD-00000025	ABC-JD-00000001		ABC-JD-00000001 – ABC-JD-00000025

2. Electronic Documents

Electronic documents include word-processing documents, spreadsheets, presentations, and all other electronic documents not specifically discussed elsewhere. Production of these items should include image files with related searchable text, metadata, and bibliographic information.   All passwords and encryption must be removed from electronic documents prior to production. Please note the following:

A. Spreadsheets
Spreadsheets should be produced in native format (*e.g.*, as .XLSX files), with searchable text for the entire document, metadata, and bibliographic information. **Provide only a single image of the first page of the spreadsheet or provide a single placeholder image. The placeholder image must contain at a minimum the BEGDOC#, FILENAME, and FILEPATH.** The Bates range for a spreadsheet should be a single number (*e.g.*, ABC-JD-00000001 – ABC-JD-00000001). The linked native file name should match the BEGDOC#/DOCID with the appropriate file extension.

B. Presentations
Presentations should be produced in full slide image format along with speaker notes (which should follow the full images of the slides) with related searchable text, metadata, and bibliographic information. Presentations should also be produced in native format (*e.g.*, as .PPT files). The linked native file name should match the BEGDOC#/DOCID with the appropriate file extension.

C. Hidden Text
All hidden text (*e.g.*, track changes, hidden columns, hidden slides, mark-ups, notes) shall be expanded and rendered in the extracted text file. For files that cannot be expanded linked native files shall be produced with the image files.

D. Embedded Files

All embedded objects (*e.g.,* graphical files, Word documents, Excel spreadsheets, .wav files) that are found within a file shall be produced so as to maintain the integrity of the source document as a single document. For purposes of production the embedded files shall remain embedded as part of the original source document. Hyperlinked files must be produced as separate, attached documents. Any objects that cannot be rendered to images and extracted text (*e.g.,* .wav, .avi files) must be produced as separate extracted files treated as attachments to the original file.

E. Image-Only Files
All image-only files (non-searchable .PDFs, multi-page TIFFs, Snipping Tool screenshots, etc., as well as all other images that contain text) shall be produced with associated OCR text, metadata, and bibliographic information.

F. Proprietary File Types and Non-PC or Non-Windows Based Systems
Proprietary file types, such as those generated by financial or graphic design software, should be discussed with the Division in advance of production to determine the optimal format of production.

File types from non-PC or non-Windows bases systems (*e.g.,* Apple, UNIX, LINUX systems), should be discussed with the Division in advance of production to determine the optimal format of production.

G. Archive File Types
Archive file types (*e.g.,* .zip, .rar) must be uncompressed for processing. Each file contained within an archive file should be produced as a child to the parent archive file. If the archive file is itself an attachment, that parent/child relationship must also be preserved.

3. Hard-Copy (or Paper) Documents

Hard-copy documents are to be produced as black-and-white image files, except where noted below, with related searchable OCR text and bibliographic information. Special attention should be paid to ensure that hard-copy documents are produced as they are kept, reflecting attachment relationships between documents and information about the file folders within which each document is found. In addition, multi-page documents must be produced as single documents (*i.e.*, properly unitized) and not as several single-page documents. Where color is required to interpret the document, such as hard copy photos, and certain charts, that image must be produced in color. These color images are to be produced as .jpg format. Hard-copy photographs should be produced as color .jpg, if originally in color, or grayscale .tif files if originally in black-and-white.

4. Shared Resources

Shared Resources should be produced as separate custodians if responsive custodians have access to them or if they contain responsive documents. The name of the group having access would be used as the custodian name, *i.e.* Marketing Execs or Accounting Dept. The company will separately provide a brief description of each shared resource that includes a list of the custodians who have access to that shared resource.

5. Database Productions

Production of enterprise databases are not addressed in these specifications and must be discussed with the appropriate government legal and technical staff to determine the optimal production format; these will usually fall outside the scope of an image-based production. Care must be taken to ensure that all responsive databases and their metadata are preserved.

6. Audio/Video Data

These specifications do not address the production of audio/video data. Care must be taken to ensure that all responsive audio/video data and their metadata are preserved. These data types may be stored in audio or video recordings, voicemail text messaging, and related/similar technologies. However, such data, logs, metadata, or other files related thereto, as well as other less common but similar data types, should only be produced after consultation with and written consent of the Division as to the format for the production of such data.

7. Foreign-Language Materials

Foreign language materials should be produced in accord with the specifications set forth in the Division's Guidelines on Production of Translations and, if appropriate, the Division's Specifications for Unicode Productions. Both are available upon request.

B. De-duplication

Before doing any de-duplication, provide the Division with a written description of the method used to de-duplicate (including which elements are compared and what hash codes are used), and what is considered a duplicate. Then confirm that your approach is acceptable to the Division. The Division does not allow de-duplication of hard-copy documents, or that of "loose" electronic documents (*e.g.,* presentation slides located on the custodian's C: drive) against email attachment versions of those same documents. The integrity of any produced email and any related "document family" must be maintained except as limited by any claim of privilege. De-duplication should occur both vertically within each custodian and horizontally across custodians. Vertical de-duplication is crucial when a production includes electronic documents from back-up tapes. Horizontal de-duplication must be done in a way that preserves (and produces) information on blind copy (Bcc) recipients of emails and other custodians whose files contain the duplicates that will be eliminated from the production.

1. Custodian Append File

A Custodian Append file is to be produced when de-duplicating ACROSS custodians (*i.e.,* horizontal de-duplication) and data is produced on a rolling basis. The file must be provided on an incremental basis **starting with the second submission;** as more custodians are discovered for previously produced documents, this file is updated with **only the new** custodian information. The Custodian Append File is a two-field delimited file consisting of the DOCIDs of the previously delivered document and the **new** custodian names for the duplicates of those records that would otherwise have been produced in the subsequent (new) submissions.

These specifications do not allow for near de-duplication or email threading. These formats must be discussed separately with the Division and written consent obtained prior to the use of such techniques for production.

C. Document Numbering

Documents must be uniquely and sequentially Bates-numbered across the entire production, with an endorsement burned into each image. Each Bates number shall be of a consistent length, include leading zeros in the number, and unique for each produced page. Bates numbers should contain no more than three segments. For example, a company identifier, a middle segment identifying the custodian, and a sequential page counter with connecting hyphens. The number of digits in the numeric portion of the format should not change in subsequent productions, nor should spaces, hyphens, or other separators be added or deleted.   Under no circumstances should bates numbers contain embedded spaces , slashes (/), backslashes (\), carats (^), underscores (_), ampersands (&), hash marks (#), plus signs (+), percent signs (%), dollar signs ($), exclamation marks (!), pipes (|), any character used as a delimiter in the metadata load files, or any character not allowed in Windows file-naming convention (,\ / : * ? " < > | ~ @ ^). Bates numbers may contain hyphens (-).

D. Privilege Designations

Documents redacted pursuant to any claim of privilege will be designated "Redacted" in the EPROPERTIES field as described in the Metadata Table. Appropriately redacted searchable text (OCR of the redacted images is acceptable), metadata, and bibliographic information must also be provided. All documents that are part of a document family that includes a document withheld pursuant to any claim of privilege will be designated "Family Member of Privileged Doc" in the EPROPERTIES field as described in the Metadata Fields table for all other documents in its family. Placeholder images with BEGDOC#, FILENAME, FILEPATH and reason withheld (*e.g.,* "Privileged") should be provided in place of the document images of the privileged document.

E. Sample

Before beginning production, a sample production covering files of all types, including emails and attachments, loose files including spreadsheets and presentations, redacted documents, etc., must be provided to the Division. The sample size should be between 500 to 1000 records to be large enough to be representative and small enough to review quickly. The Division will take a few business days to evaluate the sample and provide feedback.   If there are any problems, corrected samples will need to be resubmitted until the Division can confirm the problems are resolved.

F. Load File Set/Volume Configuration

Each production must have a unique PHYSICALMEDIA name associated with it. This PHYSICALMEDIA name must also appear on the physical label. The PHYSICALMEDIA naming scheme should start with a 2 or 3 letter prefix (identifying your company) followed by a 3-digit counter (*e.g.,* ABC001). Each separate volume delivered on that media must also have a separate VOLUMENAME associated with it. On the root of the media, the top level folder(s) must be named for the volume(s). VOLUMENAME(s) should also be indicated on the physical label of the media. The volume naming scheme should be

based on the PHYSICALMEDIA name followed by a hyphen, followed by a 3-digit counter (*e.g.*, ABC001-001). Load file volumes should be as large as practical but not contain more than 100,000 records each. The VOLUMENAME should increase sequentially across all productions on the same PHYSICALMEDIA.

Under the VOLUMENAME folder, the production should be organized in 4 subfolders:
1. <u>DOCLINK</u> (contains linked native files , may contain subfolders, with no more than 5,000 files per folder)
2. <u>IMAGES</u> (may contain subfolders, with no more than 5,000 image files per folder)
3. <u>FULLTEXT</u> (may contain subfolders, with document-level text files)
4. <u>LOADFILES</u> (should contain the metadata, DII, OPT, LST, and custodian append files)

G. Deliverables

A cover letter spreadsheet must be delivered with each submission and should provide statistical information about the volume(s) and media produced. **Provide this in hard copy format and electronically on the deliverable media**. A sample is included in this PDF.

The Division accepts electronic productions loaded onto hard drives, CD-ROMs, or DVD-ROMs; however, production on hard drives minimizes costs and delay and is preferable. Where the size of the production exceeds the capacity of a single DVD-ROM, hard drives should be used as the delivery medium. For each piece of media a unique identifier (PHYSICALMEDIA) must be provided and should also be physically visible *on the exterior* of the physical item.

If the media is encrypted, please supply the tool for decryption on the same media, and instructions for decryption. A separate email must be sent with the password to decrypt.

All documents produced in electronic format shall be scanned for, and free of, viruses. The Division will return any infected media for replacement, which may affect the timing of the company's compliance with the production request.

The Division does not accept load file productions via email or those that are posted on download sites (*e.g.,* FTP, secure server).

H. Zip File Table of Contents

The attached zip file contains this document and a sample load file set following the guidelines set forth above.

1. <u>Sample Cover Letter Spreadsheet:</u> Sample .xlsx file for providing statistics associated with each submission.
2. <u>Sample DII Load File:</u> Sample format for the image load file, Summation image link file.
3. <u>Sample OPT Load File:</u> Sample format for the image load file, Opticon image link file.
4. <u>Sample FullText LST File:</u> Sample control list for loading extracted/OCR text.
5. <u>Sample Custodian Append File:</u> This is to be produced only when de-duplicating ACROSS custodians (*i.e.,* horizontal deduplication) and rolling productions are being delivered. Provide on an incremental,

rolling basis **starting with the second submission.** As more custodians are discovered for previously produced documents, this file is populated with **only the new** custodian information.

6. <u>Sample Metadata Load File.</u> Sample delimited text metadata file.

IMAGE and TEXT FILE SPECIFICATIONS, & LOAD FILE CONFIGURATION

Please review carefully for revisions.

Image/Native File Specifications

- Black-and-white Group IV Single-Page TIFFs (300 DPI). Color images should be provided in .JPG format when color is necessary.
- Image file names should match the page identifier for that specific image and end with the .tif (or .jpg if needed) extension. Example: ACME-ABC-0003072.TIF
- File names cannot have embedded spaces, commas, underscores, ampersands, slashes, back slashes, hash marks, plus signs, percent signs, exclamation marks, any character used as a delimiter in the metadata load files, or any character not allowed in Windows file-naming convention. (, _ & \ / # + % ! : * ? " < > | ~ @ ^)
- Images for a given document must reside together in the same folder.
- The maximum number of image files should be limited to 5,000 per folder.
- Native file names should match the BEGDOC#/DOCID entry for that specific record and end with the appropriate file extension.
- The maximum number of native files in a subfolder should be limited to 5,000 per folder.
- Any encryption or password protection will be removed from all native format files produced.

Searchable Text File Specifications and Control List Configuration

- Extracted text should be provided with all records, except for documents that originated as hard copy or redacted documents.
  - For hard copy documents, please provide OCR text.
  - For redacted documents, provide OCR text for the redacted version.
- There should be a single extracted/OCR text file per document, in ASCII Text format only. This load file configuration is not for use with a Unicode based production.
- The name of the text file should be the same as the document's first page/Bates number, with a TXT extension: DOCID.TXT.
- There must be a carriage return and line feed (CRLF) no later than the 250th character of the **first** line of every text file.
- All soft and hard returns in the native electronic or image file should be replicated as a Carriage Return Line Feed (CRLF) in the text file (i.e. the lines of text in the file terminate with a CRLF in correlation with the appearance of the native electronic or rendered image file). Pay particular attention to not allow multi-line paragraphs of e-mails to be rendered as a single, extended line of text.
- Text files should include page breaks that correspond to the "pagination" of the image files.
- Place text files under a "FULLTEXT" folder and provide a Control List file for loading in the "LOADFILES" folder on the delivery media.

Metadata Load File Delimiters and Configuration

- Field Separator   ¶   (ASCII 020)
- Text Qualifier   þ   (ASCII 254)
- Substitute Carriage Return or New Line in data   ®   (ASCII 174)
- Multi-value separator **(Do Not Follow with Space)**   ;   (ASCII 059)
- Date format   YYYYMMDD (date type fields only)
- Time format   HH:MM:SS in 24-hour format (*e.g.*, 04:32 pm formatted to 16:32:00 – Do not include AM, PM, or Timezone indicators).
- There should be one line for every record in the load file. A carriage return and line feed (CRLF) must appear at the end of each record and ONLY at the end of each record.
- The first row of each metadata load file should be a header row containing the field names. Field names must match Division Metadata Table field names.
- All requested fields should be present in the metadata load file whether data exists or not. Field order must remain consistent in subsequent productions.

Summation Image Load File (.dii) Configuration

- @T ImageTag Token - MUST be first token listed for a document record and MUST be identical to BEGDOC# and DOCID entries in Metadata load file.
- @D @I Image Path Tokens - Path to image files as they appear on the delivery media.
- Image Files – Individual file names or iterated listing of image filenames comprising the document. Note that iterated filenames CANNOT contain leading zeros inside the braces {}.
- Example using iteration for document ABC-JD-00030005 to ABC-JD-00030352:
  - @T ABC-JD00030005
  - @D @I\ABC002\Images\001\
    - ABC-JD-0003000{5-9}.tif
    - ABC-JD-000300{10-99}.tif
    - ABC-JD-00030{100-352}.tif

Opticon Image Load File (.opt) Configuration – Page level comma-delimited file containing seven fields per line.
PageID,VolumeLabel,ImageFilePath,DocumentBreak,FolderBreak,BoxBreak,PageCount

- PageID – PageID of the item being loaded. MUST be identical to the image name (less the file extension).
- VolumeLabel – Optional. If used it is preferable that it match the VOLUMENAME assigned in the corresponding metadata load file.
- ImageFilePath – The path to the image from the root of the delivery media.
- DocumentBreak – The letter "Y" denotes the first page of a document. If this field is blank the page is not the first page of a document.
- FolderBreak – Leave empty
- BoxBreak – Leave empty
- PageCount – Optional
- Example - ABC-JD00030005,,\ABC002\Images\001\ ABC-JD-00030005.tif,Y,,,

METADATA TABLE OF REQUESTED FIELDS

Please review carefully as fields have been added or modified from ATR's last set of specifications.

A "**X**" indicates that the field should be populated in the load file produced. "Other ESI" includes non-email files, such as, but not limited to MS Office files, WordPerfect files, etc.

Field Name	Field Description	Field Type	Hard-Copy	Email	Other ESI	Calendar Items
COMPANIES	Company submitting data	Multi-Entry	X	X	X	X
PHYSICALMEDIA	The unique identifier on the physical piece of media (e.g., ABC001)	Note Text	X	X	X	X
VOLUMENAME	Production volume number (e.g., ABC001-001)	Note Text	X	X	X	X
CUSTODIAN	Custodian(s) / source(s) -- format: *Last, First* or *ABC Dept*	Multi-Entry	X	X	X	X
TIMEZONE	The TimeZone from which the native file was collected.	Note Text		X	X	X
SPEC#	Subpoena/request paragraph number to which the document is responsive	Multi-Entry	X	X	X	X
HASHMD5	Document MD5 hash value (used for deduplication or other processing)	Note Text		X	X	X
HASHSHA	Document SHA1 hash value (used for deduplication or other processing)	Note Text		X	X	X
SEARCHVALUES	List of search terms used to identify record as responsive (if used)	Multi-Entry	X	X	X	X
BEGDOC#	Start Bates (including prefix) -- No spaces or special characters	Note Text	X	X	X	X
ENDDOC#	End Bates (including prefix) -- No spaces or special characters	Note Text	X	X	X	X
DOCID	Must equal the value appearing in the BEGDOC# field and be UNIQUE	Note Text	X	X	X	X
NUMPAGES	Page count	Integer	X	X	X	X
PARENTID	Parent record's BEGDOC#, including prefix (populated ONLY in child records)	Note Text	X	X	X	X
ATTACHMENTIDS	Child document list: BEGDOC# of each child (populated ONLY in parent records)	Multi-Entry	X	X	X	X
FAMILYRANGE	Range of the BEGDOC# value of the parent record to the ENDDOC# value (including prefix) of the last child record (for example, ABC-JD-00001201 - ABC-JD-00001220); populated for all documents in the group. Empty if the record is NOT in family grouping	Note Text	X	X	X	X
EPROPERTIES	Indicate all that apply : Record Type: E-Doc, E-Doc Attachment, Email, Email Attachment, Hard Copy, Calendar Appt Other Notations: Translation of [DOCID of original], Translated as [DOCID of Translation] Privilege Notations: Redacted, Privileged, Family Member of Priv Doc	Multi-Entry	X	X	X	X
FOLDERLABEL	Email folder path (sample: Inbox\Active); or Hard Copy folder/binder title/label	Note Text	X	X		X
FROM	Author of the Email or Calendar item (as formatted on the original)	Note Text		X		X
TO	Recipients of the Email (as formatted on the original)	Multi-Entry		X		X
CC	Names of the individuals who were copied on the Email (as formatted on the original)	Multi-Entry		X		X
BCC	Names of the individuals who were blind-copied on the Email (as formatted on the original)	Multi-Entry		X		X
SUBJECT	Email or calendar subject	Note Text		X		X
DATE_HC	Date of hard copy documents, if coded. Format: YYYYMMDD.	Date	X			
DOCDATE	This is a multipurpose date field. Populate with: DATESAVED for E-Docs; DATESENT for Emails; DATEAPPTSTART for calendar appointments; DATE_HC for hard copy documents, if available. Format: YYYYMMDD.	Date	X	X	X	X
DATECREATED	Date electronic file was created. Format: YYYYMMDD.	Date			X	
DATESENT	Date the Email was sent. Format: YYYYMMDD.	Date		X		X
TIMESENT	Time Email was sent -- Format: HH:MM:SS (use 24 hour times, e.g., 13:32 for 1:32 pm; timezone indicators cannot be included)	Time		X		X
DATERECEIVED	Date Email was received. Format: YYYYMMDD.	Date		X		X
TIMERECEIVED	Time Email was received. Format: HH:MM:SS (use 24 hour times, e.g., 13:32 for 1:32 pm; timezone indicators cannot be included)	Time		X		X
HEADER	The internet header information for Email sent through the internet;	Note Text		X		
INTERNETMSGID	Globally unique identifier for a message which typically includes messageid and a domain name. Example: <0E6648D558F338179524D555@m1p.innovy.net	Note Text		X		X
MESSAGEID	Proprietary email database/mailstore/post office file associated with centrally managed enterprise email servers. Microsoft Outlook PST EntryID, the UniqueID (UNID) for Lotus Notes, equivalent value for other proprietary mailstore formats.	Note Text		X		X
INREPLYTOID	Internet message ID of the Email replied to	Note Text		X		
CONVERSATIONINDEX	Email Thread Identification	Note Text		X		X
IMPORTANCE	Email flag indicating priority level set for message	Note Text		X		X
DELIVRECEIPT	Delivery receipt request notification for Email messages	Note Text		X		X
READRECEIPT	Read Receipt request notification for Email messages	Note Text		X		X
SENSITIVITY	Sensitivity field from Email messages	Note Text		X		X
REVISION	Revision number extracted from metadata of native file	Note Text			X	

DATESAVED	Date native file was last modified. Format: YYYYMMDD.	Date			X	
DATEPRINTED	Date native file was printed (metadata derived from Word documents, etc.)	Date			X	
EORGANIZATION	Company field extracted from the metadata of a native file	Note Text			X	
EAUTHOR	Author field value extracted from the metadata of a native file	Note Text			X	
LAST AUTHOR	Last Saved By field value extracted from metadata of a native file	Note Text			X	
ESUBJECT	Subject field value extracted from metadata of a native file	Note Text			X	
FILESIZE	File size in Bytes (integer value only - do not include unit of measure or decimal places - e.g., 568)	Integer		X	X	X
FILENAME	File name of native file (E-Docs or attachments to Email)	Note Text		X	X	X
APPLICATION	Application used to create native file (e.g., Excel, Outlook, Word)	Note Text		X	X	X
FILEEXTENSION	File extension of native file	Fixed Length 5 chars		X	X	X
FILEPATH	File path to native file as it existed in original environment	Note Text		X	X	X
DOCLINK	File path location to the current native file location on the delivery medium	Note Text			X	
DATEAPPTSTART	Start date of calendar appointment. Format: YYYYMMDD.	Date				X
TIMEAPPTSTART	Start time of calendar appointment. Format: HH:MM:SS (use 24 hour times, e.g., 13:32 for 1:32 pm; timezone indicators cannot be included)	Time				X
DATEAPPTEND	End date of calendar appointment. Format: YYYYMMDD.	Date				X
TIMEAPPTEND	End time of calendar appointment. Format: HH:MM:SS (use 24 hour times, e.g., 13:32 for 1:32 pm; timezone indicators cannot be included)	Time				X

Cover Letter Spreadsheet for ABC Company Production to DOJ										
Matter: ABC's Acquisition of XYZ										
Custodian	Physical Media Name	Volume Name	Begin Bates Number	End Bates Number	Intentionally Left Blank	Number of Records	Number of Images	Number of Native Files	Number of Extracted/OCRed Text Files	Volume Size in GB	Date Produced
Doe, John	ABC001	ABC001-001	ABC-00000001	ABC-00005825		258	5,825	13	258		
Doe, Jane	ABC001	ABC001-001	ABC-00005826	ABC-00009536		365	3,711	52	362		
TOTALS FOR THIS VOLUME						**623**	**9,536**	**65**	**621**	**0.50**	mm/dd/yyyy
Daniels, Jack	ABC001	ABC001-002	ABC-00009537	ABC-00015263		1,150	5,727	156	1,149		
Weiser, Bud	ABC001	ABC001-002	ABC-00015264	ABC-00018273		600	3,010	20	598		
TOTALS FOR THIS VOLUME						**1,750**	**8,737**	**176**	**1,750**	**1.39**	mm/dd/yyyy
Brown, Charlie	ABC002	ABC002-001	ABC-00018274	ABC-00025225		1,315	6,952	68	1,310	1.02	mm/dd/yyyy
Flintstone, Fred	ABC002	ABC002-002	ABC-00025226	ABC-00035625	ABC-00025698 - ABC-00025982	2,023	10,115	85	2,023	2.00	mm/dd/yyyy
GRAND TOTALS						**5,711**	**35,340**	**394**	**8,071**	**4.91**	

* This spreadsheet can be provided on a current production-specific or a cumulative production basis, but the totals per *volume* must be provided

** This spreadsheet does not reflect the sample production provided

*** Volume size is actually space used not total available space on the delivery media

CHECKLIST TO ENSURE PROPER DELIVERY TO DOJ ANTITRUST

ACTION	CHECKED
Does the production format conform to DOJ Antitrust Image and Text File Specifications and Load File Configuration?	
DOCLINK folder – Does count of native files = cover letter spreadsheet and count of records where DOCLINK field is populated?	
IMAGES folder – Does count of image files = cover letter spreadsheet?	
Are all images single-page tiff files or jpeg if color is necessary?	
FULLTEXT folder – Does count of TXT files = cover letter spreadsheet and count of records in control list (LST) load file?	
LOADFILES folder – Are the metadata, DII, control list, and custodian append (if applicable) files all provided? Does count of records match across metadata, DII, and LST files for the same volume?	
Are images and native files of presentations and spreadsheets provided properly (per negotiation)?	
Are line returns in fulltext files implemented properly? Does each hard and soft return in the native files appear as a Carriage Return Line Feed (CRLF) in extracted text file?	
Are all fields listed in the Metadata Table formatted correctly?	
Are all requested fields present in the load file? Is any requested field empty for all records?	
Has the DocDate field been populated properly from Email, E-Doc and Calendar metadata dates.	
Is the metadata file properly formatted without extraneous delimiter or null characters? Have conflicts between delimiter characters also embedded in metadata been resolved?	
Are the PARENTID & ATTACHMENTIDS fields populated correctly? Is the PARENTID field only populated on child records? Is the ATTACHMENTIDS field only populated on parent records?	
Has your custodian append file been created on an "incremental" basis, so that only new custodians will be added to previously produced records? Custodians previously listed on records should NOT be included.	
Have all family members of documents withheld for privilege been properly flagged?	
Does your Bates-numbering scheme exclude the characters prohibited in the Document Numbering specifications?	
Does your file naming scheme exclude the characters prohibited in the Document Numbering specifications?	
Has each physical piece of media been labeled on the exterior with a unique identifier per the configuration specifications (PHYSICALMEDIA)?	
Have you specified the VOLUMENAME(s) as the root folder(s) on media?	
Have you submitted your cover letter spreadsheet electronically and in hard copy?	