



**ATTORNEY GENERAL OF MISSOURI**

**JOSHUA D. HAWLEY**  
ATTORNEY GENERAL

JEFFERSON CITY  
65102

P.O. Box 899  
(573) 751-3321

IN THE MATTER OF:

*Via Hand Delivery*

Google, Inc.

CID No. JC-63-17  
November 13, 2017

**CIVIL INVESTIGATIVE DEMAND**

TO: Google, Inc.  
1600 Amphitheatre Parkway  
Mountain View, California 94043

The Attorney General of the State of Missouri believes it to be in the public interest that an investigation be made to ascertain whether Google, Inc. ("Subject") has engaged in or is engaging in any practices declared to be unlawful by § 407.020, RSMo. This investigation will inquire into the activities and representations of Subject in connection with (a) the collection, use, retention, storing, sharing, sale, and dissemination of information and data relating to Google users and the use of Google websites and products; (b) the appropriation of data, information, and images from websites that compete with Google in non-search-engine markets; and (c) non-relevance-based preferencing of Google-affiliated websites and demotion of the websites of Google competitors in Google's search-engine results page. The Attorney General has reason to believe that Subject's conduct in the aforementioned areas and others involves deception, fraud, false promise, misrepresentation, unfair practices, and/or the concealment, suppression, or omission of material facts within the scope of the Missouri Merchandising Practices Act.

Please note that materials and information produced pursuant to this civil investigative demand may be disclosed to other state and/or federal law-enforcement agencies pursuant to § 407.060.1, RSMo.

The Attorney General believes that you have information, documentary material, and/or physical evidence relevant to the investigation described above.

## DEFINITIONS

In this Civil Investigative Demand, the following terms shall have the meanings set forth herein:

1. “You” and “Your” mean Google, Inc.; Google, Inc.’s subsidiaries, parent companies, and sister companies; and all agents, representatives, employees, independent contractors, attorneys, and other persons acting or purporting to act on behalf of Google, Inc. and/or its subsidiaries, parent companies, or sister companies.

2. “Document” includes every “writing,” “recording,” and “photograph” as Federal Rule of Evidence 1001 defines those terms, as well as any “duplicate” of any writing, recording, or photograph. “Document” includes but is not limited to electronic documents, files, databases, and records, including but not limited to emails, voicemails, text messages, calendar appointments, instant messages, MMS messages, SMS messages, iMessages, computer files, spreadsheets, and metadata. The term Document includes every draft of any other material that falls within the definition of Document.

3. “Communication” means any expression, statement, conveyance, or dissemination of any words, thoughts, statements, ideas, or information, regardless of form, format, or kind. “Communication” includes but is not limited to oral or written communications of any kind, such as telephone conversations, discussions, meetings, notes, letters, agreements, emails or other electronic communications, facsimiles, and other forms of written or oral exchange that are recorded in any way, including video recordings, audio recordings, written notes, or otherwise. Any Communication that also falls within the definition of “Document” shall constitute both a Document and a Communication for purposes of this civil investigative demand.

4. With regard to a person, “Identify” means to state with specificity the person’s legal name, aliases, last-known home address, last-known business address, current employer, current job title, all known telephone numbers, and all known email addresses.

5. With regard to a Communication, “Identify” means to state with specificity the date of the Communication; the medium of communication; the location of the Communication; the name(s) and alias(es) of the person(s) who made the Communication; and the name(s) and alias(es) of all persons who were present when the statement was made, who received the Communication, who heard the Communication, or who came to know of the content of the Communication at a later time.

6. “All” and “any” shall each be construed to encompass the meanings of the words “all” and “any.”

7. “Person” means any natural person, corporation, proprietorship, partnership, association, firm, or entity of any kind.

8. “Third Party” means any Person (as defined herein) that does not fall within the scope of the definition of “Google” (as defined herein), except that “Third Party” shall not include any federal, state, or local governmental entity, nor shall “Third Party” include any natural person acting on behalf of any federal, state, or local governmental entity.

9. “FTC Agreement” means the December 27, 2012 correspondence from David Drummond, Chief Legal Officer of Google, to Jon Leibowitz, then-Chairman of the Federal Trade Commission, available at [https://www.ftc.gov/system/files/documents/closing\\_letters/google-inc./130103googleletterchairmanleibowitz.pdf](https://www.ftc.gov/system/files/documents/closing_letters/google-inc./130103googleletterchairmanleibowitz.pdf).

10. “FTC” means the Federal Trade Commission, as well as any member, employee, or agent of the Federal Trade Commission.

11. “European Commission” means the European Commission, as well as any agency, employee, officer, or agent of the European Commission.

12. “Lowe Letter” means the September 11, 2017 correspondence from Luther Lowe, Vice President of Global Public Policy for Yelp, to Acting FTC Chairman Maureen Ohlhausen.

13. “Google Affiliate” means any website, interface, functionality, product, service, or program operated, owned, and/or controlled by Google (as defined herein). The term “Google Affiliates” includes, but is not necessarily limited to, those websites, interfaces, functionalities, products, services, and programs identified at the website [www.google.com/about/products/](http://www.google.com/about/products/).

14. “SERP” means search engine results page.

### **DEMAND FOR DOCUMENTS AND INFORMATION**

Pursuant to § 407.030, RSMo, the Attorney General demands that—no later than 10:00 a.m. (Central) on January 22, 2018—You produce the following documents and information, to the extent that they are within your possession, custody, and/or control.

Your document production must comply with the Missouri Office of the Attorney General Production Specifications and Data Delivery Standards, a copy of which is attached hereto.

In responding to each Request contained in this civil investigative demand, You should identify—by Bates range, or by file names and locations—which Documents are responsive to each Request.

If You withhold any responsive materials based on an assertion of privilege and/or the work-product doctrine, you must produce a privilege log that provides—for each Document or Communication withheld—sufficient information to permit the Attorney General’s Office to assess the applicability of the privilege and/or the work-product doctrine.

1. Produce every version of the Google Privacy Policy that has been in effect at any time from October 1, 2012 to the present.

2. Produce all drafts of each version of the Google Privacy Policy produced pursuant to Request 1 above.

3. Identify all Google current and former employees, officers, contractors, and agents who participated in the drafting, writing, editing, or revision of each version of the Google Privacy Policy produced pursuant to Request 1 above.

4. Produce all Communications sent or received by any person identified pursuant to Request 3 above between January 1, 2012 and the present that relate to the Google Privacy Policy.

5. Identify the search terms and parameters used to identify materials responsive to Request 4 above.

6. Produce all contracts and agreements pursuant to which Google discloses any information about Google users to any Third Party.

7. Identify all Google current and former employees, officers, contractors, and agents who raised any concern or filed any complaint regarding the Google Privacy Policy at any time from January 1, 2012 to the present.

8. Produce all Communications relating to any concern or complaint responsive to Request 7 above.

9. Identify the search terms and parameters used to identify materials responsive to Request 8 above.

10. Produce all organizational charts for personnel working on Google Attribution.

11. Produce all organizational charts for personnel working on Store Sales Management.

12. Identify every Person that participates in any of “Google’s third-party partnerships, which capture approximately 70% of credit and debit card transactions in the United States,” as described in the Google post *Powering Ads and Analytics Innovation with Machine Learning*, <https://adwords.googleblog.com/2017/05/powering-ads-and-analytics-innovations.html> (dated May 23, 2017), including each of “Google’s third-party partner[s].” In responding to this Request, You need not Identify any Google employees or officers.

13. Produce all contracts and agreements between Google, on the one hand, and any Person identified pursuant to Request 12 above, on the other hand.

14. Produce all contracts and agreements to which Google is a party pursuant to which any information about any credit-card and/or debit-card transactions is disclosed to Google. This Request seeks contracts and agreements executed on or after January 1, 2012.

15. Identify every Person that has provided any consideration to Google in order to participate in the Store Sales Management program and/or the Store Sales Visit program.

16. Produce all contracts and agreements between Google, on the one hand, and any Person identified pursuant to Request 15, on the other hand.

17. Produce all Documents—including all Communications—relating to the aggregation, anonymization, and protection of consumer or user information obtained or exchanged pursuant to the Store Sales Management program and/or the Store Sales Visit program. This request seeks all responsive materials that were either created and/or transmitted on or after January 1, 2012.

18. Identify the search terms and parameters used to identify materials responsive to Request 17 above.

19. Produce all Documents—including all Communications—relating to CryptDB.

20. Identify each provision of each version of the Google Privacy Policy in effect at any time during 2017 that discloses to Google users that Google may (a) collect information regarding users' credit-card and/or debit-card transactions; (b) correlate information regarding users' credit-card and/or debit-card transactions with those users' online activity; and (c) correlate users' activity across multiple electronic devices.

21. Identify all categories of information and/or data that Google collects regarding users of Google products, websites, and/or interfaces.

22. Identify all categories of information and/or data that constitute "personal information" within the meaning of the Google Privacy Policy.

23. Identify all categories of information and/or data that constitute "sensitive personal information" within the meaning of the Google Privacy Policy.

24. Identify all categories of information and/or data collected and/or obtained by Google that constitute "non-personally identifiable information" within the meaning of the Google Privacy Policy.

25. Identify all categories of information and/or data regarding users of Google products, websites, and/or interfaces that Google discloses to or shares with any Third Party. For each category of information and/or data identified, Identify each Third Party to whom that category of information and/or data is disclosed or with whom it is shared.

26. Identify all categories of information and/or data that are disclosed to any Third Party pursuant to the Google Attribution program.

27. Identify all categories of information and/or data that are disclosed to any Third Party pursuant to the Store Sales Management program.

28. Beginning with October 2012 and continuing until the Present, for each month, Identify (a) the total number of viewings of the websites www.google.com and www.youtube.com; and (b) the total number of viewings of the Google Privacy Policy and any other privacy policy applicable to YouTube.

29. Produce all Documents—including all webpages—that have been available to Google users at any time between October 1, 2012 to the present that provide any information regarding ways in which users can opt out of any tracking or data collection by Google and/or any Third Party acting in concert with Google.

30. All Documents relating to any pattern, practice, policy, or algorithm using the occurrence of the webpage of a Google Affiliate's competitor in the Google SERP to affect in any way the placement of any Google Affiliate webpage within that SERP.

31. Identify the search terms and parameters used to identify materials responsive to Request 30 above.

32. All Documents relating to a 2008 presentation titled "Online Advertising Challenges: Rise of the Aggregators."

33. All Documents relating to a 2009 report titled "Product universal top promotion based on shopping comparison presence."

34. Produce all Documents that Google and/or its attorneys produced or provided to the FTC in connection with FTC Matter No. 111-0163.

35. Produce all Documents that Google and/or its attorneys have produced or provided to the European Commission in connection with European Commission Antitrust Matter No. 39740.

36. Produce all complaints received by Google and/or its attorneys relating to any actual or alleged use or display by Google (or any Google Affiliate) of images, data, and/or information crawled from any website. This Request seeks only those complaints received by Google and/or its attorneys on or after October 1, 2012.

37. Produce all Communications between Google and/or its attorneys, on the one hand, and any Person who sent a complaint responsive to Request 36 above, on the other hand. This Request seeks only those Communications that were sent or received on or after October 1, 2012.

38. Produce all Communications relating to any complaint responsive to Request 36 above. This Request seeks only those Communications that were sent or received on or after October 1, 2012.

39. Produce all cease-and-desist letters received by Google and/or its attorneys relating to any actual or alleged use or display by Google (or any Google Affiliate) of images, data, and/or information crawled from any website. This Request seeks only those cease-and-desist letters received by Google and/or its attorneys on or after October 1, 2012.

40. Produce all Communications between Google and/or its attorneys, on the one hand, and any Person who sent a cease-and-desist letter responsive to Request 39 above, on the other hand. This Request seeks only those Communications that were sent or received on or after October 1, 2012.

41. Produce all Communications relating to any cease-and-desist letter responsive to Request 39 above. This Request seeks only those Communications that were sent or received on or after October 1, 2012.

42. Produce all policies relating to the collection and/or obtaining of images, data, and information for use in Google Local OneBox and/or Google Local and/or G+ Local.

43. Produce all policies relating to the collection and/or obtaining of images, data, and information for use in Google Places, Google Flights, Google Hotels, Google Advisor, and/or Google Compare.

44. Produce all Documents that Google has produced to the FTC relating to Google's compliance with the FTC Agreement.

45. Identify the search terms and parameters used to identify materials responsive to Request 45 above.

46. Produce all internal audits relating to Google's compliance with the FTC Agreement.

47. Produce all Documents reflecting any steps that Google has taken to comply with the assurances made in the section of the FTC Agreement titled "I. Google's Display of Third-Party Content."

48. Identify the search terms and parameters used to identify materials responsive to Request 47 above.

49. Produce all Documents relating to whether Google will continue to comply with the assurances made in the section of the FTC Agreement titled "I. Google's Display of Third-Party Content" after December 27, 2017.

50. Identify the search terms and parameters used to identify materials responsive to Request 49 above.

51. Produce all Documents relating to the Lowe Letter.

52. Identify the search terms and parameters used to identify materials responsive to Request 51 above.

53. Produce all Documents reflecting any steps that Google takes to ensure that it complies with the terms and conditions (or terms of use) of websites that Google Crawls.

54. All Documents identified or referred to in your response to any Request above.

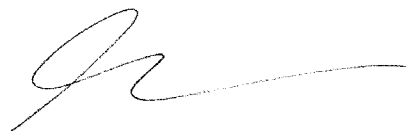
The Attorney General's Office may serve additional or subsequent civil investigative demands on you.

Please note that pursuant to § 407.080, RSMo, certain acts done with the intent to avoid, evade, or prevent compliance in whole or in part with any civil investigative constitute a Class A misdemeanor, which is punishable by a fine not to exceed \$1,000 for individuals and \$5,000 for corporations, or by imprisonment for a term of not more than one year, or both a fine and imprisonment.

No extension of the deadline for compliance with this civil investigative demand shall be effective unless it is reflected in a writing executed by an authorized representative of the Attorney General.

Submit the following Certification of Compliance and all responsive documents and information to:

Michael Martinich-Sauter  
Missouri Attorney General's Office  
P.O. Box 899  
Jefferson City, Missouri 65102  
Michael.martinich-sauter@ago.mo.gov



---

Michael Martinich-Sauter, Mo. Bar No. 66065  
Deputy Attorney General for Legal Policy and  
Special Litigation  
Missouri Attorney General's Office  
Supreme Court Building  
207 W. High Street  
P.O. Box 899  
Jefferson City, Missouri 65102  
(573) 751-8145  
michael.martinich-sauter@ago.mo.gov



In the Matter of:

Google, Inc.

CID NO. JC-63-17

**CERTIFICATION OF COMPLIANCE**

I/We hereby certify that all documents and information required by Civil Investigative Demand No. JC-63-17 which is in the possession, custody, control, or knowledge of, \_\_\_\_\_ has been submitted to the Missouri Attorney General as directed herein.

Signature \_\_\_\_\_

Title \_\_\_\_\_

Sworn to before me this \_\_\_ day of \_\_\_\_\_, 20\_\_.

\_\_\_\_\_  
Notary Public

My Commission Expires:

IN THE STATE OF \_\_\_\_\_ )  
 ) SS  
COUNTY OF \_\_\_\_\_ )

**AFFIDAVIT**

Before me, the undersigned authority, personally appeared, \_\_\_\_\_  
who, being by me duly sworn, deposed as follows:

My name is \_\_\_\_\_, I am of sound mind,  
capable of making this affidavit, and personally acquainted with the facts herein stated:

I am the custodian of the records of \_\_\_\_\_.  
Attached hereto are \_\_\_\_\_ pages of records from \_\_\_\_\_.  
These \_\_\_\_\_ pages of records are kept by \_\_\_\_\_ in the  
regular course of business, and it was the regular course of business of \_\_\_\_\_  
for an employee or representative of \_\_\_\_\_ with knowledge of  
the act, event, condition, opinion, or diagnosis recorded to make the record or to transmit  
information thereof to be included in such record; and the record was made at or near the  
time of the act, event, condition, opinion or diagnosis. The records attached hereto are the  
original or exact duplicates of the original.

Affiant

In witness whereof, I have hereunto subscribed my name and affixed my official  
seal this \_\_\_\_\_ day of \_\_\_\_\_, 20\_\_.

[SEAL]

Notary Public

My Commission Expires: \_\_\_\_\_

## Missouri Office of the Attorney General Production Specifications and Data Delivery Standards

### A. Document Categories

#### 1. Email, Attachments, and Other Electronic Messages

Email and other electronic messages (e.g., instant messages (IMs)) should be produced either natively or 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 below, 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.

2. **Attachments:** The parent-child relationship must be maintained with any production and notated through the load file fields provided with the production.
3. **Electronic (Loose) Documents:** Electronic documents, including, but not limited to, word-processing documents, spreadsheets, presentations, and all other electronic documents not specifically discussed elsewhere should be produced either natively or as image files with related searchable text, metadata, and bibliographic information. All passwords and encryption must be removed from electronic documents prior to production. However:
  - a. **Spreadsheets:** Must be produced in native format 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 PRODBEG and FILENAME. The Bates range for a native spreadsheet should be a single number. The linked native file name should match the PRODBEG with the appropriate file extension.
  - b. **Presentations:** Must 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 may be produced as image files with related searchable text, metadata and bibliographic information. However, the AGO retains the right to request that any presentation be produced subsequently in native format. Additionally, the AGO retains the right to request that any presentation produced in black and white be produced in color.
  - 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 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. **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.

#### 4. **Hard-Copy (or Paper) Documents**

Hard-copy documents are to be produced as black-and-white image files, except where otherwise noted 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 in the ordinary course, 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. If documents originally in color are produced in black-and-white, the AGO retains the right to request that such documents be produced in color.

#### 5. **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. A list of the other custodians with access to the shared resource should be provided in the CUSTODIAN field.

#### 6. **Other Sources:** The following types of data/document productions should be discussed with the AGO prior to any production to determine the most appropriate production format.

- a. Proprietary File Types and Non-PC or Non-Windows Based Systems
- b. Database(s) and/or dynamic data
- c. Audio and/or video data
- d. Foreign-Language data or documents

- B. **De-duplication:** De-duplication, both horizontally and vertically, within and across custodians is highly encouraged for electronic documents based on the files' MD5 or SHA-1 hash values. 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.

**Note:** When de-duplicating horizontally (i.e. across custodians), the CUSTODIAN field must reflect all custodians that held the duplicate record(s), which would have been produced but not for the de-duplication.

C. **Email threading**

A producing party may produce the "most inclusive email threads" based on the following:

1. A "most inclusive email thread" is one that contains all of the prior or lesser-included emails, including attachments, for that branch of the e-mail thread. In an email thread, only the final-in-time document need be produced, assuming that all previous emails in the thread are contained within the final message and provided that the software used to identify these "most inclusive email threads" is able to identify any substantive differences to the thread such as changes in recipients (*e.g.*, side threads, subject line changes), selective deletion of previous thread content by sender, etc.
2. Where a prior email contains an attachment, that email and attachment shall not be removed as a "most inclusive email thread." When an email thread branches, each branch shall be treated a separate email, and should be produced separately. Each branch that is the "most inclusive email thread" may likewise be treated as the most inclusive thread of all prior or lesser included emails within that branch and only the final-in-time email for that thread need be produced.
3. The AGO retains the right to request the individual emails contained in the most inclusive email threads be produced as needed.

D. **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 other special characters other than hyphens (-).

E. **Privilege Designations**

Documents redacted pursuant to any claim of privilege should be designated "Redacted" in the PROPERTIES field. 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 PROPERTIES field as described in the Metadata Fields table for all other documents in its family. Placeholder images

with PRODBEG, FILENAME and reason withheld (e.g., "Privileged") should be provided in place of the document images of the privileged document

**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., ABC001001). 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. DOCLINK (contains linked native files, may contain subfolders, with no more than 5,000 files per folder)
2. IMAGES (may contain subfolders, with no more than 5,000 image files per folder)
3. FULLTEXT (may contain subfolders, with document-level text files)
4. LOADFILES (should contain the metadata, DII, OPT, LST, and custodian append files)

**G. Deliverables**

The AGO accepts electronic productions loaded onto hard drives, USB drive, CD-ROMs, or DVD-ROMs. Each piece of media a unique identifier (PHYSICALMEDIA) must be provided and should also be physically visible on the exterior of the physical item. Should you wish to make a production by electronic transfer, please discuss with the AGO in sufficient time prior to your production.

All deliverables should identify, at a minimum, the following:

1. Case number
2. Production date
3. Producing party
4. Bates Range

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

All documents produced in electronic format shall be scanned for, and free of, viruses. The AGO will return any infected media for replacement.

## **IMAGE and TEXT FILE SPECIFICATIONS, & LOAD FILE CONFIGURATION**

### **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.
- File names and folder names should not contain embedded spaces or special characters (including the comma).
- Images for a given document must reside together in the same folder.
- Native file names should match the PRODBEG for that specific record and end with the appropriate file extension.
- Native files should have a placeholder image numbered by the PRODBEG of the file and at a minimum contain the PRODBEG and FILENAME.
- All files must have a unique bates number.
- All images must be endorsed with sequential Bates numbers in the lower right corner of each image.
- 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.
- Text must be produced as separate text files, not as fields within the .DAT file. The full path to the text file (OCRPATH) should be included in the .DAT file. We require document level ANSI text files, named per the BATESBEG/Image Key. Please note in the cover letter if any non-ANSI text files are included in the production. Extracted text files must be in a separate folder. There should be no special characters (including commas in the folder names). For redacted documents, provide the full text for the redacted version.

### **Metadata Load File Delimiters and Configuration**

- Field Separator = Column (ASCII 020)
- String value delimiter = Quote (ASCII 254)
- Newline delimiter = (ASCII 174)
- Multi-value separator = (ASCII 059)
- Date format YYYYMMDD (date type fields only)
- Time format HH:MM:SS

**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 - EXP-0000001,\ABC001\Images\001\ ABC0000001.tif,Y,,,



The metadata of electronic document collection should be extracted and provided in a .DAT file using the field definition and formatting described below:

Field Name	Sample Data	Description
PRODUCING PARTY	Company XYZ; John Smith	Producing Party Name
PHYSICALMEDIA	ABC001	Unique identifier for that media
VOLUMENAME	ABC001-001	Production volume number
CUSTODIAN	Smith, John; XYZ Dept.	Custodian(s) and/or source information
HASHMD5 (or SHA-1)	d41d8cd98f00b204e9800998ecf8427e	MD5 (or SHA-1) hash value used for de-duplication or other processing
PRODBEG	EXP0000001	Starting Bates number per document
PRODEND	EXP0000001	Ending Bates number per document
PRODBEGATTACH	EXP0000001	First Bates number of a single attachment
PRODENDATTACH	EXP0000001	Last Bates number of a single attachment
ATTACHRANGE	EXP0000001-EXP0000009	Bates range from the first page of the parent document to the last page of the last child document
PARENTBATES	EXP0000001	First Bates number of a parent document/email; this field should be populated for each child document.
CHILDBATES	EXP0000002; EXP0000007	First Bates number of each child attachment; can be more than one Bates number if multiple attachments; this field should be populated for each parent document.
FROM	John Smith; john.smith@abcco.com	Email sender
TO	John Smith; john.smith@abcco.com	Email recipient(s); semi-colons should separate multiple entries
CC	John Smith; john.smith@abcco.com	Email carbon copy(s); semi-colons should separate multiple entries
BCC	John Smith; john.smith@abcco.com	Email blind carbon copy(s) semi-colons should separate multiple entries
SUBJECT	Your Subject Here	Email Subject Line
FILENAME	YourFilenameHere.doc	The original native file, including extension
DATESENT	YYYYMMDD	Date email sent
TIMESENT	HH:MM:SS	Time email sent
TIMEZONE	CST	The time zone in which emails were standardized during conversion
TIMERECEIVED	HH:MM:SS	Time email received
AUTHOR	John Smith	Author of a document
DATECREATED	YYYYMMDD	Date the document was created.
TIMECREATED	HH:MM:SS	Time
DATE LAST MODIFIED	YYYYMMDD	Date document was last modified
FILEEXTENSION	.msg; .doc; .xls; .ppt	File extension of native document
FILE SIZE	550 MB; 2GB	File size in bytes
PGCOUNT	2	Number of pages in native file or email
FILEPATH	P:\shared\smithj\yourfilenamehere.doc	Path where native file document was stored, including original filename
OCRPATH	Text/001/EXP0000001.txt	Path to extracted text
PROPERTIES	Redacted; Attorney-Client Privilege	Privilege notation, Redacted, Document withheld based on privilege