



Document production guidelines

Draft 30/06/15

Contents

PURPOSE OF THESE GUIDELINES	3
CONTENT OF THESE GUIDELINES	3
REQUEST FOR DOCUMENTS	3
‘DOCUMENT’ DEFINED	3
PRODUCING HARDCOPY DOCUMENTS IN RESPONSE TO A REQUEST.....	4
PRODUCING HARDCOPY DOCUMENTS IN PAPER	4
PRODUCING HARDCOPY DOCUMENTS IN ELECTRONIC FORMAT	4
PRODUCING ELECTRONIC DOCUMENTS IN RESPONSE TO A REQUEST	5
PRODUCING ELECTRONIC DOCUMENTS IN THEIR NATIVE FORMAT	5
PRODUCING ELECTRONIC DOCUMENTS IN OUR PREFERRED FORMAT	5
DELIVERING DOCUMENTS TO THE COMMISSION	5
WHAT TO DO IF YOU HAVE QUESTIONS OR NEED FURTHER INFORMATION	6
ATTACHMENT A – PRODUCING HARDCOPY DOCUMENTS IN ELECTRONIC FORMAT.....	7
SCANNING DOCUMENTS.....	7
NAMING IMAGE FILES.....	7
ORGANISING IMAGE FILES	7
REDACTING IMAGE FILES	7
ATTACHMENT B – PRODUCING ELECTRONIC DOCUMENTS IN THEIR NATIVE FORMAT.....	8
COPYING NATIVE FILES.....	8
NAMING NATIVE FILES	8
SECURITY	8
ORGANISING NATIVE FILES	8
REDACTING NATIVE FILES.....	8
ATTACHMENT C – PRODUCING HARDCOPY AND ELECTRONIC DOCUMENTS IN OUR PREFERRED FORMAT	9
CONVERSION TO IMAGE FILES	9
SEARCHABLE TEXT	9
CODING	10
DOCUMENT CODING FIELDS	10
PRODUCING ATTACHMENTS	11
PRODUCING TRANSLATIONS	11
LOAD FILES	11
DUPLICATE DOCUMENTS	12
SPECIAL CONSIDERATIONS FOR CERTAIN DOCUMENTS HELD IN ELECTRONIC FORMAT	12
REDACTIONS FOR PRIVILEGE.....	13
GLOSSARY	14

Purpose of these guidelines

1. These guidelines are designed to help you produce documents (electronic and hardcopy) in response to a Commerce Commission request for documents.

Content of these guidelines

2. These guidelines describe and explain several methods for the production of documents in response to a request.
3. Although we encourage you to produce hardcopy documents in electronic format, the production of hardcopy documents in hard copy is discussed at paragraphs 16 and 17 below.
4. Attachment A to these guidelines describes an acceptable method for producing hard copy documents in electronic format.
5. Attachment B to these guidelines describes an acceptable method for producing electronic documents in electronic format.
6. Attachment C to these guidelines describes our **preferred** method for the production of all documents, hardcopy and electronic.
7. We have included a glossary, at the end of these guidelines, to assist you to understand some of the technical terms used.

Request for documents

8. We request documents in three ways:
 - 8.1 voluntarily, by way of a verbal or written request;
 - 8.2 as part of a leniency or cooperation agreement; or
 - 8.3 compulsorily, by way of statutory notice under s 98 of the Commerce Act or s 47G of the Fair Trading Act.
9. In these guidelines we refer to these three ways of requesting documents collectively as 'requests'.

'Document' defined

10. A 'document' is defined in both the Commerce Act and the Fair Trading Act. The definitions are wide and include both:
 - 10.1 hardcopy documents, eg, paper files and folders; and
 - 10.2 documents that are stored in electronic form (electronic documents), eg, emails, Word, and Excel documents stored on computer systems, mobile devices, etc.

11. Different considerations apply when producing hardcopy documents compared with electronic documents. The reason for this is that the information contained in hardcopy and electronic documents differs.
12. A hardcopy document contains no more information than is visible on the face of the document.
13. In contrast, an electronic document contains more information than can be seen by simply viewing it on a computer screen, or by printing it out and looking at it in hardcopy. This additional information can include the document's authorship, ownership, and date of creation.
14. This additional information is known as 'metadata'. The definitions of 'document' in the Commerce Act and the Fair Trading Act are broad enough to include metadata associated with an electronic document.
15. As discussed further below, we require electronic documents to be produced in one of two electronic formats to ensure the documents include their relevant associated metadata. If you do not produce the metadata associated with an electronic document, we may not regard your response as being in compliance with the request.

Producing hardcopy documents in response to a request

Producing hardcopy documents in paper

16. Although you may produce paper copies of hardcopy documents, receiving large volumes of paper documents can be difficult for us to handle efficiently. We therefore prefer and encourage you to produce all hardcopy documents electronically.
17. If you choose to produce paper copies of hardcopy documents, staple individual hardcopy documents together to help us identify when one document stops and another starts. Maintain the original and intended structural relationships of the documents with paperclips (eg, a letter and its enclosure or a memorandum and its attachment).

Producing hardcopy documents in electronic format

18. Producing hardcopy documents in electronic format will help us efficiently progress our investigation.
19. Hardcopy documents can be produced in electronic format in one of two ways:
 - 19.1 scanned in accordance with the instructions set out in **Attachment A**; or

19.2 in our preferred format, scanned and provided with searchable text and coding¹ as set out in **Attachment C**.

20. **Attachment C** is based on and consistent with the High Court Rules governing the listing and exchange of documents in discovery. It provides instructions on how to format and manually code hardcopy documents produced in electronic format.

Producing electronic documents in response to a request

21. Our primary concern is to ensure that we receive electronic documents with their relevant associated metadata. To achieve this you can produce electronic documents in either native format or in our preferred format.

Producing electronic documents in their native format

22. Native format means the format in which the documents were originally created. For example, if a document was created as a Microsoft Word document, its native format is as a Microsoft Word document.
23. **Attachment B** explains in detail how to produce electronic documents in native format.
24. Do not simply print out or convert electronic documents to pdf or any other image file format. Doing so will lose the document's metadata. If you produce electronic documents in this way, we may not regard your response as being in compliance with the request.

Producing electronic documents in our preferred format

25. **Attachment C** describes our preferred format for the production of electronic documents. This production method is compatible with Summation,² preserves key metadata, and is based on and consistent with the High Court Rules governing the listing and exchange of documents in discovery.
26. As with the electronic production of hardcopy documents, producing documents in this way will assist us to efficiently progress the investigation.

Delivering documents to the Commission

27. Produce your documents with a cover letter and index detailing:
- 27.1 the documents being produced and their applicable date ranges;
 - 27.2 the person(s) who had possession, custody, or control of the documents (the custodian);³

¹ Coding is an automated or manual process by which key information about the document such as names, dates, document type, and other relevant data is recorded.

² Summation is the evidence management software currently used by the Commission to store and review electronic documents.

³ The custodian may also be a shared folder, directory, drive, or other shared location.

- 27.3 the request or requests in response to which each document or set of documents is being produced;
 - 27.4 any redactions made to the documents for privilege and the basis of those redactions;
 - 27.5 any claim of confidentiality over a document; and
 - 27.6 an explanation of any de-duplication⁴ process you have used.
28. If you are providing documents in electronic format, load the documents onto electronic media such as hard drives, flash drives, CD-ROMs, or DVD-ROMs. Clearly label what you are producing with the case name and number (where applicable), your name, and the date of production.
29. Take all reasonable steps to ensure that all electronic files, including the text, coding/metadata, and load files are useable and are not infected by viruses or other malicious software.

What to do if you have questions or need further information

30. If you have any questions regarding the contents of these guidelines or how they apply to you, contact us for further guidance as soon as possible and before producing documents.

⁴ A method through which exact duplicates of electronic documents are removed prior to production. For more information, see Attachment C.

Attachment A – Producing hardcopy documents in electronic format

1. This attachment describes an acceptable method for producing hardcopy documents in electronic format. We are willing to discuss variations to this method on a case by case basis. If you wish to assist us by producing hardcopy documents in our preferred format, please see Attachment C.

Scanning documents

2. Individually scan each separate hardcopy document as its own image file (Tiff if possible, otherwise PDF).

Naming image files

3. Name each image file with a unique document ID⁵ (without spaces).

Organising image files

4. Organise the image files by custodian⁶ in correspondingly named folders.
5. Provide a list of the image files in text (.txt) or Excel (.xls/.csv) format including:
 - 5.1 the document ID; and
 - 5.2 the nature of any structural relationships between documents, eg, a letter and its enclosure or a memorandum and its attachment.

Redacting image files

6. Redact a document for privilege by:
 - 6.1 blacking out redactions on the image;
 - 6.2 stamping the image as “Redacted”; and
 - 6.3 retaining the original un-redacted and unaltered version of the document so that it remains available to be inspected by us or the Court if required.

⁵ Use an alphanumeric document ID to show the source of the document. For example, a document sourced from ABC Company might bear the document ID ABC.001.0001. In this case, ABC references the responding party, the first set of numbers references the volume number of the production, and the second set of numbers indicates the document number within that volume.

⁶ The person(s) who had possession, custody, or control of the documents. The custodian may also be a shared folder, directory, drive, or other shared location.

Attachment B – Producing electronic documents in their native format

1. This attachment describes how to produce electronic documents in their native format (native files). We are willing to discuss variations to this method on a case by case basis. If you wish to assist us by producing electronic documents in our preferred format, please see Attachment C.

Copying native files

2. When copying native files, preserve all relevant associated metadata.
3. Copy emails with all attachments (do not extract email attachments and produce them as separate documents).

Naming native files

4. Do not rename native files.

Security

5. Remove password protection and/or encryption from all native files.
6. Scan all native files to be produced for viruses and other malicious software.
7. Do not otherwise alter a native file in any way.

Organising native files

8. Organise the native files by custodian⁷ in correspondingly named folders.

Redacting native files

9. If you need to redact a native file due to privilege, please contact us to discuss the best method of production for that document.

⁷ The person(s) who had possession, custody, or control of the documents. The custodian may also be a shared folder, directory, drive, or other shared location.

Attachment C – Producing hardcopy and electronic documents in our preferred format

1. This attachment describes how to produce both hardcopy and electronic documents in our preferred format. We are willing to discuss variations to this method on a case by case basis.
2. Our preferred format is a production method that bundles together:
 - 2.1 image files generated from native electronic files or hardcopy documents;
 - 2.2 searchable text files;
 - 2.3 manual coding and/or extracted metadata; and
 - 2.4 a Summation or Summation-compatible load file that links together the images, text, and coding/metadata.

Conversion to image files

3. Except as set out in paragraphs 17-19 below, convert all documents, electronic or hardcopy, to multi-page Tiff image files.
 - 3.1 For electronic documents, computer generate the image files, rather than printing and scanning them. Hardcopy documents may be scanned.
 - 3.2 Name each image file with a unique alphanumeric document ID (without spaces) to show the source of the document.⁸
 - 3.3 Stamp the top right corner of each page of the image file with the document ID.

Searchable text

4. If text extraction is available, extract and store text without formatting in a generic format. If text extraction is unavailable, or in the case of hardcopy documents, use OCR.
 - 4.1 Create a single text file per document.
 - 4.2 Include page breaks that correspond to the 'pagination' of the image files.
 - 4.3 Name the text file using its document ID (without spaces).

⁸ For example, a document sourced from ABC Company might bear the document ID ABC.001.0001. In this case, ABC references the responding party, the first set of numbers references the volume number of the production, and the second set of numbers indicates the document number within that volume.

Coding

5. Enter manual coding and capture metadata extracted from electronic documents in fields as set out below and store that information in text (.txt), or Excel (.xls/.csv) format.

Document coding fields

6. Provide the following coding/metadata fields for each document:
 - 6.1 **Document ID:** Ensure the document ID used to name the document is in the document ID field.
 - 6.2 **Custodian:** The person who had possession, custody, or control of the document. Identify the person by name in this field in the format “first name surname”. If the document was sourced from a shared folder, directory, drive, or other shared location, adequately describe that location in this field.
 - 6.3 **Document Date:** The date appearing on the face of the document in “DD/MM/YYYY” format. For example, code 15 January 2010 as “15/01/2010”.
 - 6.3.1 If a document is partially dated or only partially legible, enter the information that can be determined from the document.
 - 6.3.2 If the date is estimated, state that in an additional field entitled “Estimated date”.
 - 6.3.3 For meeting minutes, enter the date of the meeting, not the date the minutes were generated.
 - 6.4 **Document Type:** A meaningful, plain language description of the type of document being produced. For example, enter “email” rather than “Microsoft Outlook”, “Lotus Notes,” or “.msg file.” Likewise, enter “letter”, “memorandum”, “meeting minutes”, or “report”, rather than “Microsoft Word file” or “Word Perfect file”. Enter “translation”, for all translations.
 - 6.5 **Author:** The name of the author of the document (not the typist, or ‘entered by’ operator), in the format, “first name surname”. If only part of either the individual or organisation can be determined, provide the information available. If only an email address is available to populate this field, provide the name of the individual or organisation using that email address.
 - 6.6 **Operator:** The name of the typist, word processor, or ‘entered by’ operator, in the format, “first name surname”.
 - 6.7 **Recipient:** The name of the recipient(s) of a document, in the format, “first name surname”. If only part of either the individual or organisation can be determined, provide the information available. Where a document is coded as “meeting minutes”, provide the attendees at the meeting (rather than the actual recipient of the document).

- 6.8 **Copied To:** The names of all persons or groups that were sent copies of the document (cc:'ed), in the format, "first name surname" or "group name". If only part of either the individual or organisation can be determined, provide the information available.
- 6.9 **Blind Copied To:** The names of all persons or groups that were sent blind copies of the document (bcc:'ed), in the format, "first name surname" or "group name". If only part of either the individual or organisation can be determined, provide the information available.
- 6.10 **Parent Document ID:** The document ID of the document to which the produced document is attached, enclosed with, or embedded within.
- 6.11 **Attachment Document ID:** The document ID or IDs of any documents attached, enclosed, or embedded within the produced document.
- 6.12 **Confidentiality Designation:** If you are claiming that any portion of a produced document is confidential, enter "Yes" in this field. If no claim of confidentiality is being made as to the document, enter "No" in this field. Do not rely solely on the language of a cover letter to identify confidential documents.
7. Use a semi-colon ";" to separate multiple entries in any field.

For example:

8.	Recipient	John Smith; Jill Smith
----	-----------	------------------------

Producing attachments

9. Identify any document that is attached to, enclosed with, or embedded within another document as an attachment or "child". List attachments as separate documents and give them the next sequential document ID after their "parent".⁹

Producing translations

10. Link translations (formal or otherwise) of foreign language documents to the original foreign language document by next sequential document ID as if it were an attachment to the foreign language document.

Load Files

11. Provide a Summation or Summation-compatible load file with all produced documents that links together the images, text, and coding/metadata.

⁹ The top level or primary document to which other documents are attached, enclosed, or embedded is typically referred to as the "parent". All documents attached, enclosed, or embedded within that document may sometimes be referred to as "children". For example, if a letter has an enclosure, the letter is the parent and the enclosure or attachment is the child. If an email has two attachments, the email is the parent and the two attachments are both children. Identifying these structural relationships when producing documents helps maintain the integrity of your documents.

Duplicate documents

12. If you have access to software that can highlight exact duplicate documents from the same custodian using MD5 hash values,¹⁰ take reasonable steps to remove such duplicates before production (de-duplication).
13. De-duplication must not take place across custodians (Cross-Custodian De-Duplication). In other words, if a document is held by multiple custodians (eg, an email received by multiple custodians), produce each custodian's copy of the document.
14. The following examples are not exact duplicates. Produce these documents as separate documents:
 - 14.1 a copy or version of a document containing a material modification, obliteration, or other marking;
 - 14.2 an email identifying blind copy (bcc:) recipients; and
 - 14.3 an attachment that appears as a standalone document elsewhere in the production.
15. Keep all documents removed during your de-duplication process so that the documents are available to be inspected at a later date if required.

Special considerations for certain documents held in electronic format

16. Produce presentations in full slide image file format (including searchable text and metadata) with any speaker notes following the full images of the slides.
17. Produce digital copies of photographs, video, or audio recordings (including voicemail) that were either originally created using digital devices or have previously been digitised.
18. If technically feasible and cost-efficient, produce photographs, video, and audio recordings (including voicemail) that are not already in a digital format in their original analogue format.
19. Produce electronic documents in their native format if necessary to preserve their content, functionality, and/or integrity. For example, some Microsoft Excel spreadsheets may contain formulae which cannot be viewed unless produced in native format. Likewise, some types of presentations may contain animations, sound, or other embedded media that might be lost if produced in image file format.
 - 19.1 Remove password protection and/or encryption from all produced native files.
 - 19.2 Do not rename or otherwise alter a native file in any other way.

¹⁰ An MD5 hash value is a mathematical algorithm that creates a unique value for a document or data set, like a digital fingerprint of the document.

- 19.3 Preserve all associated metadata when copying native files.
- 20. Produce colour versions of documents if colour is necessary to convey the meaning and/or context of the document.
- 21. Extract or unpack archive file formats (eg, .zip files).
 - 21.1 If the archive file is an attachment (eg, to an email) and either the parent document or one or more files contained in the archive is determined to be relevant, produce the parent document and produce each individual file contained in the archive as a separate attachment to that that document.¹¹
 - 21.2 If however, the archive file is a standalone document (ie, not attached to another file), produce only relevant files contained in the archive.

Redactions for privilege

- 22. To redact an image file for privilege:
 - 22.1 black out the redacted portions on the image;
 - 22.2 stamp the image as “Redacted”;
 - 22.3 identify the type of privilege being claimed in a separate field (eg, a “Notes” or “Comments” field); and
 - 22.4 retain the original un-redacted and unaltered version of the document so that it remains available to be inspected by us or the Court if required.
- 23. If you need to redact a native file due to privilege, please contact us to discuss the best method of production for that document.

¹¹ Please contact us if you find that a specific document or set of documents makes this instruction disproportionate.

Glossary

Archive File: A collection of electronic files stored in a compressed format.

Attachment: A document or file associated with another document or file through its original and intended structural relationship, eg, a letter and an enclosure or an email and an attached file. There may be multiple attachments associated with a single “parent” document.

Author: The person, office, or designated position responsible for a document’s creation or transmission. In the case of a letter, the author is usually indicated on the letterhead or by signature. In some cases, the software application producing the document may capture the author’s identity and associate it with the document.

Coding: The process by which a document’s details and properties are recorded. Coding can be automated, manual, or a combination of the two.

Coding Fields/Fields: Database fields used to categorise and organise documents. These fields are used to record an individual piece of data about a document, such as the author, recipient, or date.

Computers and Mobile Devices: Electronic devices including but not limited to servers, desktops, laptops, tablets, and smartphones.

Cross-Custodian De-Duplication: A process that culls out multiple copies of an electronic file that reside with different custodians.

Custodian: The person(s) who had possession, custody, or control of the documents. The custodian may also be a shared folder, directory, drive, or other shared location.

Custodian De-Duplication: A process that culls out multiple copies of an electronic file that reside with the same custodian.

De-Duplication (“De-Duping”): The process of comparing electronic files based on their characteristics and removing or marking duplicate files.

Document: See definition of “document” in section 2(1) of the Commerce Act 1986 and section 2(1) of the Fair Trading Act 1986.

Document Date: The original creation date of a document.

Document ID: An alphanumeric code used to identify, track, and show the source of a document. For example, a document sourced from ABC Company might bear the document ID ABC.001.0001. In this case, ABC references the responding party, the first set of numbers references the volume number of the production, and the second set of numbers indicates the document number within that volume.

Document Type (Doc Type): A field used in coding. Examples include letter, memo, report, article, meeting minutes, etc.

Electronic File: A document, program, or data set stored, run, or maintained electronically on a computer, computer system, server, mobile device, or electronic media.

Electronic Media: An object or device, such as a disc, tape, hard drive, flash drive, or other device on which data is stored.

Embedded: To embed a file or document is to insert or incorporate it within the body of another file or document. An embedded file is typically a non-text or multimedia (image, audio, or video) file inserted into a web page, presentation, PDF, or other document. For example, a picture or set of pictures embedded into an email so that they appear as part of the message rather than as attachments.

Encryption: A procedure that makes the contents of a message or file scrambled or unintelligible to anyone not authorised to read it. Encryption is used to protect information as it moves from one computer to another.

File Extension: One or more characters following the name of an electronic file used to indicate file type or format. For example, image files may have the file extension .pdf, .bmp, .gif, .jpg, or .tiff. Audio files may have the file extension .aud or .wav. There are many different file extensions identifying different file formats.

Filename: The name of an electronic file including its file extension.

Flash Drive: A small electronic device containing flash memory that is used for storing and/or transferring electronic files. Also known as a thumb drive, memory stick, jump drive or USB flash drive.

Format: The way a document is stored and used (eg, hard copy, electronic). For electronic files, specific applications may define unique formats for their data (eg, “MS Word document file format”, Excel file format). Many files may only be viewed or printed using their originating application or an application designed to work with compatible formats.

Hard Drive: A high-capacity storage device for electronic files consisting of a spinning disc inside a sealed unit. Hard drives may be internal to a computer or external, for use as back-up storage or for transport.

Image File/Image File Format: An electronic or digital picture of a document (eg, TIFF, PDF, etc).

Load File: An electronic file used to import documents into evidence management software (eg, Summation), linking the image files to information about the documents (eg, metadata, coded data, and searchable text).

Malicious Software (or Malware): Any software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems or mobile devices.

MD5 Hash: A unique value for a given set of data (the product of a mathematical formula), similar to a digital fingerprint.

Metadata: Information about a particular electronic file or data set which describes how it is formatted and how, when, and by whom it was collected, created, accessed, and modified. It is data *about* an electronic file as opposed to the file's content.

Native Format/Native File: The format in which the documents were originally created. For example, if a document was created as a Microsoft Word document, its native format is as a Microsoft Word document. An image file format, such as TIFF or PDF, also known as a "converted file format", is designed to retain an image of the document as it would look if viewed in the original creating application but do not allow metadata to be viewed or the document's information/data to be searched or sorted.

OCR (Optical Character Recognition): A process that translates and converts printed matter on an image into a format that a computer can recognise as searchable text.

Parent Document: The top level or primary document to which other documents are attached, enclosed, or embedded. All documents attached, enclosed, or embedded within that document may sometimes be referred to as "children". For example, if a letter has an enclosure, the letter is the parent and the enclosure or attachment is the child. If an email has two attachments, the email is the parent and the two attachments are both children.

Password: A code used to gain access to a computer, application, site, or electronic file.

PDF (Portable Document Format): An image file format developed by Adobe Systems. PDF files may be text-searchable or image-only.

Production: The process of delivering to another party, or making available for that party's review, documents in response to a request.

Redaction: A portion of an image or document is intentionally concealed to prevent disclosure.

Searchable Text: Searchable text refers to the ability to run a full-text search over a document to find all instances of the queried text or numerals within the document. Searchable text may be part of the native file (ie, MS Word documents and emails are typically searchable in their native format) or as a separate text file created during the processing of native electronic files by way of text-extraction or OCR.

Server: Any central computer on a network that contains data or applications shared by multiple users of the network on their client PCs.

Shared Folder/Directory/Drive: A file folder, directory, or drive on a server shared by multiple users on a computer network.

Summation: The evidence management software currently used by the Commission to store and review electronic documents.

TIFF: One of the most widely used image file formats. File name has .tif or .tiff file extension. Tiff images can be single or multi-page but are not text-searchable.

Unpack: A term used to refer to the extraction of individual files from an archive file.

Virus: A self-replicating program that spreads by inserting copies of itself into other electronic files.

ZIP: A common archive file format. Electronic files that have been compressed using this format may be referred to as “zipped.”