

Technical Guidance for Material Providers

1. Material Providers should note the following standards for submitting evidence to the Inquiry:
 - a. All documents should be provided with an accompanying index listing them, along with their locations in any systems or folders.
 - b. All documents should be provided electronically where possible , with their original filenames and, where scanned, they should be named appropriately.
 - c. Digital documents should be provided in their native format (e.g. Microsoft Word, Microsoft Excel, email files (e.g. msg, eml - with their attachments), together with any underlying metadata.
 - d. Documents should be submitted at document level, i.e. multiple documents should not be grouped into one large PDF. Similarly, a document should not be scanned as individual pages, but as a single PDF.
 - e. Hard copy documents should be digitised to multipage PDF format.
 - f. Scanning of paper documents should be undertaken at a standard of 300dpi.
 - g. Scanned (PDF) images of paper should show all information on the paper, including all corners and both sides of the paper (where information is contained on two sides) and any annotations. No data should be cut off during the scanning process. If the paper that is being scanned has any attachment (e.g. a sticky paper note) or cover sheet, it should be scanned twice, both with and without the attached note.

- h. Documents should not be password protected, encrypted or otherwise restricted at file level unless agreed with the Inquiry in advance. Where passwords are required, they should be supplied separately to the Inquiry.
 - i. All documents should be provided in unredacted format, unless agreed with the Inquiry beforehand.
- 2. In so far as is possible, when compressing material for upload, please do keep any original folder/file structure/categorisation you have applied to the material as part of your records management. It will be helpful to the Inquiry if any work on categorisation of the material has already been applied.
- 3. The Inquiry will arrange for the secure transfer of documents with Material Providers. Documents will be held securely when in the custody of the Inquiry. Material Providers will need to consider their own obligations under the Public Records Act 1958 regarding documents provided to the Inquiry and other potentially relevant documents they hold.
- 4. As outlined above, each disclosure should also be made with an accompanying index (.csv/excel file). The following index should include any unique identification number that a Material Provider wishes to be applied to any documents supplied upon ingestion of such material onto the Inquiry's document management system ("**DMS**").
- 5. The Inquiry will refer to the following unique reference as the 'Control Number'. Please ensure you provide a .csv/excel/index file that aligns exactly with the names of the files (including their formats), along with the Control Number, to those files which you have uploaded (please see example below).

Control Number	File Name
MSTR0000001	Letter from x to y - Dec 2013.doc
MSTR0000002	2005 - Spreadsheet Audit.xls

...	...
MSTR0006345	2015 Internal Guidance.xls

6. The Inquiry will also supply all ingested files on its DMS with a Unique Reference Number ("**URN**") for all internal Inquiry purposes (including any reference within a final and/or interim report). If no Control Number is provided by the Material Provider, the Inquiry will apply the same reference as the URN.
7. If a Material Provider stores material on a DMS, we ask that the material is provided to the Inquiry as a formal production export from that system. Your DMS supplier or eDisclosure team should be able to assist with preparing the export. The **Annex** to this guidance sets out the technical requirements for production data and is intended primarily for eDisclosure and IT professionals.

V2.0 August 2026

Annex

Production Specification for Document Management Systems

1. The purpose of this Annex is to set out the standard requirements for the exchange of production data between Material Providers and the Inquiry. It is intended for Material Providers who operate their own Document Management System (**DMS**) and can provide material to the Inquiry as a formal production export. Material Providers who are not using a DMS in this way do not need to follow this Annex.

Electronic documents: (i) Native format; and (ii) Image format

2. Native format: where possible, documents are exchanged in their native format. Unless previously agreed between parties, or in the case where documents have been redacted or withheld for privilege, documents will be provided in native format. For the avoidance of doubt, Material Providers are not required to produce documents solely because they form part of a family containing relevant material. Documents may be withheld where they are not relevant to the Inquiry's **Terms of Reference** and placeholders should **not** be provided solely on that basis. However, where a relevant child document is produced, its parent document should also be produced, notwithstanding that the parent document may not itself be relevant. Similarly, where a relevant parent document is produced, any child documents that are themselves relevant to the Terms of Reference should also be produced.
3. Image format: documents will be provided in image format to maintain redactions or include placeholders. Placeholders allow documents considered wholly privileged to be withheld from the production. Additionally, documents will be produced in image format if they have any attachments that contain redactions or are wholly privileged. This is to prevent inadvertent disclosure of withheld content which will occur if disclosed in native format.

4. Text files: A text (TXT) file containing extracted text must be provided for each document. For redacted or placeholder documents, post-production OCR text must be provided. OCR text must reflect the removal of redacted content, or the replacement of all text content with placeholder specific text.

Load files: (i) Concordance data file (DAT), and (ii) Opticon file (OPT)

5. Concordance data file (DAT): this file is used to load documents in the production volume to a supported review platform. It includes document identifiers with family relationship fields, associated metadata fields and file path links to the associated native and text files (extracted OCR text).
6. Opticon file (OPT): this file is used to load image files for documents in the production volume to a supported review platform. It includes relational links to each image and allows the review platform to associate multiple images (pages) with a single document.

Technical Format and Specification

7. Native format documents:
 - a. Native files must be stored in a separate folder from text and images.
 - b. The native file location must be provided in the DAT file.
 - c. Native files must be named after the production document identifier.
(With their original filenames preserved in metadata/index file/s).
 - d. Native files should not contain archive or container files (for example, ZIP, RAR – which should have been unpacked or processed).
 - e. Folder and file names must not contain commas and quotes.

- f. The full path length for the native file location must be less than 250 characters.

8. Image format documents:

- a. Images must be stored in a separate folder from natives and texts.
- b. Images must be named after the production document identifier with a suffix indicating the page number of the document.
- c. Black and white: single page TIFFs, 1-bit, Group IV compression.
- d. Colour: JPEG format.

9. Extracted Text / OCR

- a. A text file must be provided for each document.
- b. Text files must be stored in a separate folder from natives and images.
- c. The text file location must be provided in the DAT file.
- d. Text files must be named after the production document identifier.
- e. Text files must be in UTF-8 encoding to preserve most foreign characters.

10. Image load file (OPT):

- a. Single page TIFF image files are constructed from sets of pages. For documents containing more than one page, the OPT load file within the production volume is required to associate sets of TIFF images to single documents.

- b. Each line of the OPT file must contain: a unique identifier for a particular image; and a relational path to the image file within the production volume.
- c. The OPT load file should be page-level, with each line relating to a single image and page. A corresponding image file should be present within the document production volume for each line entry.
- d. On the first page of a document, a start flag of 'Y' must be populated, and the overall number of pages within the document must be provided at the end. This field must be blank for all other pages within a document.

11. Metadata load file (DAT)

Load File Field Header	Description	Format	Limited Metadata
Disclosure ID	Unique Document level reference. Please use a unique production ID in the format of 4 letters followed by numbering padded to 7 zeros. Please use a 4 letter prefix that derives from the name of your organisation for ease of identification and reference (not just Prod or DOC) e.g. EPIQ0000001 for an Epiq disclosure.	Sequential, document-level	YES
Family ID / Group Identifier	Field to identify family grouping. Please populate this field with the parent Disclosure ID for every document within a family group.	Sequential, document-level	YES
Parent Date	Merged date value uniform across an entire family. This is a hierarchical date value that may be generated using different logic per processing tool or in 3rd-party data received.	dd/mm/yyyy hh:mm	NO
Date Created	The date/time a file was originally created.	dd/mm/yyyy hh:mm	NO
Date Last Modified	The date/time a file was last changed and saved.	dd/mm/yyyy hh:mm	NO
Date Sent	The date a file was sent through an e-mail transaction.	dd/mm/yyyy hh:mm	NO
Date Received	The date a file was received through an e-mail transaction.	dd/mm/yyyy hh:mm	NO

File Name	The name of a file, excluding root drive and directory path information. This will be populated only for non-email file types.		NO
File Extension	Extension following the proper filename used to identify file formats.		NO
File Type	The format of a file that determines with which software programs it can be used.		NO
Subject	Subject of an email. This will be populated only for email file types.		NO
Author	The author(s) of a document.		NO
From	The email address and/or name of the person who sent the email.		NO
To	List of e-mail addresses/names of people receiving the email.	Multi value delimiter is ";	NO
Cc	List of e-mail addresses/names that were carbon copied for the document.	Multi value delimiter is ";	NO
Bcc	List of e-mail addresses/names that were blind carbon copied for the document.	Multi value delimiter is ";	NO
Document Hash	The calculated hash value of the document. For purposes of consistency across providers please provide MD5 hash.		NO
Redaction	Y/N – Only when redactions are present.		YES
Placeholder	Y/N – Only when placeholders are present.		YES
Rule 9 Reference	Reference for the data supplied e.g. R9(1)		YES
Reference fields	Please provide any additional reference fields or document IDs that you think will be helpful to the Inquiry or that you are likely to rely on in correspondence with the Inquiry. For example, a field indicating if the document is relevant or non-relevant.		YES
Source Folder Path / Box ID (for scanned documents)	Location where the file was originally situated, provided as a folder path – where possible to do so. For scanned documents, please indicate which box/folder the document was located in.		NO
Native Path	Path of native files.		YES
Text Path	Path to text files.		YES

The .DAT file must be in UTF-8 encoding.

Standard delimiters:

Value	Character	ASCII Number
Column	¶	020
Quote	”	254
Newline	®	174
Multi-Value	;	059
Nested Value	\	092

12. Method of exchange

The Inquiry can provide and accept production media via the following methods of delivery. Encryption or password-protection will be used, unless otherwise agreed between parties.

- Encrypted Hard Drive / USB
- SFTP transmission

13. Samples

Standard volume naming convention: VOL_####

DAT file (.DAT) – example header line:

¶Disclosure ID¶Family ID¶Parent_Date¶Author¶File Type¶Native
Path¶Text Path¶

Example data line:

¶EPIQ0000001¶EPIQ0000001¶05/02/2000 12:54:23¶Bloggs, Joe¶Outlook
Email¶b.\NATIVES\001\EPIQ0000001.msg¶b.\TEXT\001\EPIQ0000001.txt¶

Opticon File (.OPT) – example:

EPIQ0000001,VOL_0001,.\IMAGES\001\EPIQ0000001_0001.tif,Y,,,2
EPIQ0000001_0002,VOL_0001,.\IMAGES\001\EPIQ0000001_0002.tif,,,,
EPIQ0000002,VOL_0001,.\IMAGES\001\EPIQ0000002_0001.tif,Y,,,2
EPIQ0000002_0002,VOL_0001,.\IMAGES\001\EPIQ0000002_0002.tif,,,,
EPIQ0000003,VOL_0001,.\IMAGES\001\EPIQ0000003_0001.tif,Y,,,2
EPIQ0000003_0002,VOL_0001,.\IMAGES\001\EPIQ0000003_0002.tif,,,,