



**Board of Inquiry**  
into the McCrae landslide

**ANNEXURE A TO**  
**PRACTICE DIRECTION NO. 4: PRODUCTION OF DOCUMENTS**  
**DOCUMENT MANAGEMENT PROTOCOL**

**FORMAT OF DOCUMENTS**

- 1 Material Providers should note the following standards for producing documents to the Inquiry:
  - 1.1 All documents should be provided with:
    - i. an accompanying itemised electronic index of documents in Microsoft Excel format (**Index**).
    - ii. provided in an electronic format in accordance with the Production Loadfile at **Schedule 1**.
  - 1.2 The Index and Load File should contain the following data for each document where available:
    - i. Document ID;
    - iii. Host Reference;
    - iv. Document Type;
    - v. Document Date;
    - vi. Document Title;
    - vii. Author (From);
    - viii. Recipient (To);
    - ix. Recipient (CC);
    - x. Recipient (BCC);
    - xi. Notice No.;
    - xii. Notice Tranche No.;
    - xiii. Notice Schedule Item;
    - xiv. Withheld;
    - xv. Withheld Reason;
    - xvi. Restriction requested;
    - xvii. Reason for restriction request.



## **Board of Inquiry** into the McCrae landslide

- 1.3 All documents should be provided electronically where possible, with their original filenames; where scanned, they should be named appropriately.
- 1.4 Documents should be provided in their native format (e.g. Microsoft Word, Microsoft Excel, email files (e.g. msg, eml) with their attachments).
  - i. 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.
  - ii. Hard copy documents should be digitised to multipage PDF format.
  - iii. Scanning of paper documents should be undertaken at a standard of 300dpi.
  - iv. 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.
  - v. All documents should be provided in unredacted format, unless agreed with the Inquiry beforehand.
- 2 The standards set out above do not preclude the Inquiry from accepting documents, at its discretion, in a hard copy format or as objects.
- 3 Material Providers should provide all metadata where practicable. It is accepted that complete document metadata might not be available for all electronic documents.
- 4 Original versions of all documents must be retained by the Material Provider.



## Board of Inquiry into the McCrae landslide

### IDENTIFICATION OF DOCUMENTS BY MATERIAL PROVIDER

- 5 Each document must be identified with a Document ID and page number which are unique to each page and will be the primary means for identification of documents.
- 6 All Document IDs and page numbers are to be stamped in the top right hand corner of each page.
- 7 A Document ID must be in the following format:  
**PPP.BBBB.FFFF.NNNN\_(XXXX)**, where:
  - 7.1 **PPP** is a three letter 'party code'. A Material Provider producing documents should contact the Board of Inquiry at [solicitorsassisting@mccraeinquiry.vic.gov.au](mailto:solicitorsassisting@mccraeinquiry.vic.gov.au) prior to production to confirm the party codes available for use. The relevant party code will ordinarily be provided to each Material Provider in the relevant Notice.
  - 7.2 **BBBB** is a 4-digit 'box' number identifying separate collections of documents (for example in relation to a particular Notice). The number is to be between 0001–9999.
  - 7.3 **FFFF** is a 4-digit 'container' number identifying further separate collections of documents. The number is to be between 0001–9999.
  - 7.4 **NNNN** is a 4-digit number used to differentiate individual pages and/or documents. In some cases, NNNN operates as a document number rather than a page number because individual pages are not numbered (i.e. non-standard native files not produced as searchable PDFs). This number is padded with zeros to consistently result in a 4-digit structure.
  - 7.5 **(XXXX)** is an optional 4-digit number used to identify suffix rendered PDF pages. It is only required where a Material Provider chooses to review documents in native format in their document review platforms and render documents to PDF for the purpose of production. The suffix must be preceded by an underscore and padded with zeros to consistently result in a 4-digit number structure.



## Board of Inquiry into the McCrae landslide

- 8 An example of the Document ID structure is as follows: **ABC.0001.0001.0020**, where:

ABC	Party Code
0001	The unique 'box' number allocated by the Material Provider
0001	The unique 'container' number allocated by the Material Provider
0020	The unique document number within the 'container'

- 9 Assigned Document IDs must be unique to each document and must not be reassigned to subsequent documents produced.
- 10 It is understood and accepted that Document IDs may not be consecutive as a result of the removal of irrelevant documents during review. A Material Provider must however identify host and attachment documents with consecutive Document IDs.
- 11 If alternate numbering is required, please contact the Solicitors Assisting the Board of Inquiry to discuss.

### DOCUMENT HOSTS AND ATTACHMENTS

- 12 Every electronic document that is attached to or embedded within another document will be treated as an Attached Document. A document that contains at least one Attached Document will be called a Host Document. A document that is neither a Host Document nor an Attached Document will be called a Standalone Document.
- 13 Examples of Host Documents and Attached Documents include:
- 13.1 an email, letter or fax (Host Document) and its attachments (Attached Documents); and
- 13.2 an electronic file (Host Document) that has other files embedded within it (Attached Documents).
- 14 If an Attached Document also contains attachments, those attachments will be treated as attachments to the Host Document.



## Board of Inquiry into the McCrae landslide

- 15 Material Providers must ensure that false or unnecessary relationships between Host Documents and Attached Documents are not created by:
  - 15.1 taking reasonable steps to ensure that email footers, logos, and other repeated content are not separated as Attached Documents; and
  - 15.2 ensuring that physical or digital document containers, such as hard copy folders or electronic ZIP container files, are not identified as Host Documents, unless the identification of the container as a Host Document is necessary to the understanding of the documents within that container.
- 16 Unless required to provide documents in their native structure for technical reasons, documents should be extracted from their containers and the container itself should not be produced.
- 17 For hard copy documents, document delimiting is to be done logically. This means that physical delimiters such as staples, clips, dividers, may be used as a guide only and the start and end page of a document should be logically determined. All documents comprising a brief, file, or similar are separate documents.

### DOCUMENT METADATA

- 18 Wherever possible, a Material Provider is to rely on the automatically identified metadata of electronic documents. Automatically identified metadata should be used when:
  - 18.1 searching for documents;
  - 18.2 itemising documents in a list; and
  - 18.3 preparing a production of documents in accordance with the Production Specification at **Schedule 1**.
- 19 Material Providers should take reasonable steps to ensure that all appropriate document metadata is not modified or corrupted during collection and preparation of electronic documents for review and production.
- 20 Document metadata is to be automatically extracted using UTC + 10 (Sydney, Melbourne, Canberra) as the time zone in the processing application.



## Board of Inquiry into the McCrae landslide

- 21 It is accepted that complete document metadata may not be available for all electronic documents. Material Providers should attempt to provide complete metadata where practicable.
- 22 Hard copy documents must be produced as PDFs, together with extracted text files together with a load file as per **Schedule 1** where possible.
- 23 Material Providers must provide information regarding the software and procedure used to automatically identify the metadata of their electronic documents if requested by the Board of Inquiry.
- 24 Original versions of all documents must be retained by the Material Provider producing them.

### DE-DUPLICATION OF DOCUMENTS

- 25 Material Providers must take reasonable steps to ensure that duplicate documents are removed from the produced material (De-Duplication).
- 26 The Board of Inquiry acknowledges that there may be circumstances where duplicates need to be identified and produced for evidentiary purposes.
- 27 Duplication must be considered at a document group level. That is, all documents within a group comprising a Host Document and its attachments, will be treated as duplicates only if the entire group of documents is duplicated elsewhere. An Attached Document must not be treated as a duplicate if it is merely duplicated elsewhere as an individual standalone document that is not associated with another group of documents.
- 28 Material Providers must apply electronic De-Duplication using a MD5 algorithm.
- 29 Material Providers will take reasonable steps to remove irrelevant system files and immaterial content including temporary internet files, 'thumbs.db' files, and cookies.
- 30 Temporary internet files and cookies are to be excluded from the production process.



## **Board of Inquiry** into the McCrae landslide

### **EXCLUSION OF UNUSABLE FILE TYPES**

- 31 An NIST filter is to be applied to a Material Providers' electronic documents to remove files with no user-generated content, such as system files and executable files, so that these are excluded from searches and production (to the extent possible).
- 32 Temporary internet files and cookies are to be excluded from the production process.

### **TREATMENT OF EMAIL CHAIN CORRESPONDENCE**

- 33 Where an email is identified as relevant and it forms part of an email chain, the Material Provider must produce the entire email chain.

### **USE OF ADVANCED ANALYTICS TECHNOLOGY**

- 34 Material Providers may use advanced analytics technologies at their own discretion, but they must maintain the integrity and context of the documents, and produce entire document groups including all attachments.
- 35 Material Providers may use 'email threading' technology to minimise document review. Where this technology has been used, Material Providers may provide only the relevant end point email with its attachments.
- 36 Material Providers may use technology commonly referred to as 'TAR / Assisted Review / Predictive Coding' for document review at their discretion. Material Providers do not need to seek agreement to use such technology, but must disclose to the Inquiry that it has been used and implement processes to ensure that they are meeting their obligations under a Notice or otherwise by providing only material identified as relevant to the Notices issued, along with their document group.



## **Board of Inquiry** into the McCrae landslide

### **DATA SECURITY**

- 37 Material Providers must take all reasonable steps to ensure that the data is useable and is not infected by malicious software.

### **ERRORS IN PRODUCED DOCUMENTS**

- 38 If errors are found in any produced documents, the Material Provider producing must provide a corrected version of the document to the Board of Inquiry as soon as possible once that error is identified.
- 39 If errors are found in more than 25% of the produced documents in any one tranche, the Material Provider must, if requested by the Board of Inquiry, provide a correct version of all documents within the tranche.
- 40 A written explanation setting out the reasons for the errors in the documents and describing the data affected must be provided by the Material Provider if errors are found in any produced documents.

### **ELECTRONIC PROVISION OF DATA FOR PRODUCTIONS**

- 41 Unless otherwise agreed or ordered by the Board of Inquiry, the information provided and delivered to the Board of Inquiry must be contained on agreed electronic media, being either:
- 41.1 SFTP services of the Material Provider providing the documents; or
- 41.2 USB media.
- 42 In all cases parties must apply encryption to the zip file uploaded to the SFTP or the USB media provided and the password must be shared with the Board of Inquiry via a separate email, at the time of confirming the delivery.





## **Board of Inquiry** into the McCrae landslide

### **Schedule 1**

#### **PRODUCTION SPECIFICATION FOR LOAD FILE (RELATIVITY COMPLIANT) AND DOCUMENTS**

##### **PRODUCTION FORMAT**

- 1 Documents must be provided electronically, using a .DAT/OPT data file format and in Microsoft Excel format.
  - 1.1 The first line of the .DAT file must be a header row identifying the field names.
  - 1.2 The .DAT file must use the following Concordance® default de-limiters:
    - 1.3 Pilcrow ¶ ASCII character
    - 1.4 Quote ¨ ASCII character.
- 2 Date fields should be provided in the format: DD-MM-YYYY.
- 3 If the production includes emails and attachments, the attachment fields must be included to preserve the parent/child relationship between an email and its attachments.
- 4 Productions must include an extracted text file for each document. A TEXTPATH field must be included to provide the file path and name of the extracted text file on the produced storage media. The text file must be named after the Document ID. Do not include the text in the .DAT file.
- 5 For productions that contain PDF or Native documents, a LINK field must be included to provide the file path and name of the native file on the produced storage media. The native file must be named after the Document ID.

##### **DOCUMENT FOLDER STRUCTURE**

- 6 The file name of each document must include the relevant file extension, e.g. 'DocumentID.xxx' where 'xxx' is the file extension.
- 7 The top level folder containing every document must be named \Documents\.
- 8 The documents folder must be structured in accordance with the Document ID hierarchy, e.g. Documents\ABC\.



## Board of Inquiry into the McCrae landslide

### OVERVIEW OF METADATA PROVIDED WITHIN THE DATA (.DAT) LOAD FILE

9 Required fields/metadata in a flat file format:

Field	Explanation	
Document ID	Document ID	
Host Reference	If the document is an attachment, this field contains the Document ID of its host document.  If a document does not have a host, this field is to be left blank\ null.	
Family ID	Each document will be assigned an ID which identifies the family group it appears in.  Standalone documents should get the same value as the Document_ID.  Host and Attachments in a family group should have the Host Document ID recorded.	
Document Date	DD/MM/YYYY HH:MM	
	Paper Documents	Determined on the basis of the date appearing on the face of the document.
	Undated Documents	Leave field blank\ null.
	Incomplete Date (Year Only)	For example, 01/01/YYYY 00:00
	Incomplete Date (Month and Year Only, or Day and Month Only)	For example, 01/MM/YYYY 00:00 DD/MM/1900 00:00
	Emails	Email Sent Date & Time
	Unsent Emails	Last Modified Date & Time
	Other Electronic Documents	Last Modified Date & Time; or Date appearing on the face of the document.



**Board of Inquiry**  
into the McCrae landslide

<b>Title</b>	Paper Documents	Determined on the basis of the title appearing on the face of the document.
	Email	Subject field from email metadata.
	Other Electronic Documents	Metadata file name or determined on the basis of the title appearing on the face of the document.
<b>From</b>	Format: Person[Organisation];	Paper Documents: Name of person to be determined on the basis of the face of the document [Name of organisation that produced the document as determined on the basis of the face of the document]
<b>To</b>		Emails:
<b>CC</b>		Electronic metadata – email addresses or email alias names.
<b>BCC</b>		
<b>Withheld</b>	Yes Part	Only required for documents being withheld in full or part. Single choice only.
<b>Withheld Reason</b>	Legal professional privilege  Parliamentary privilege  Public interest immunity	Basis on which a reasonable excuse is claimed:  Only required for documents marked as Withheld = Yes or Part.



**Board of Inquiry**  
into the McCrae landslide

<b>Restricted</b>	Yes Part	Only required for documents with restrictions in full or part. Single choice only.
<b>Restricted Reason</b>	Legal professional privilege  Parliamentary privilege  Public interest immunity	Basis on which a reasonable excuse is claimed:  Only required for documents marked as Withheld = Yes or Part.
<b>Notice to Produce No</b>		Request number as identified on the Notice
<b>Notice to Produce Tranche No</b>		Notice to Produce Tranche No in which the document is produced under
<b>Notice to Produce Schedule Item</b>		Notice to Produce Schedule item the document is relevant to
<b>MD5 Hash Value</b>		MD5 hash value used for de-duplication, if available.
<b>TEXTPATH</b>	FolderPath\Document_ID.TXT	Extracted text path.
<b>FILE Path</b>	FolderPath\Document_ID.EXT	Native path for documents produced in native format/PDF format.



## **Board of Inquiry** into the McCrae landslide

- 10 Material Providers should observe the following guidance in relation to information (To/From/CC/BCC) technical requirements:
- 10.1 These fields hold the names of parties associated with a particular document and their relationship to the document. It may also hold organisation information for these people.
- 10.2 Describing people
- i. A person's name may be referenced using:
    1. email addresses (for example, jcitizen@xyz.com.au); or
    2. Surname [space] first name initial (for example, Citizen J) where email addresses are not available; or
    3. by reference to a position (for example, Psychologist) where email addresses or surname and first name initial are not available; or
    4. by reference to an organisation associated with the person where email address, surname and first name initial and position are not available.
- 10.3 Multiple recipients must be separated by a semicolon.
- 10.4 Organisations must be placed into square brackets.