Commercial Litigation Association of Ireland

Good Practice Discovery Guide

v2.0 – November 2015

Table of Contents

Foreword

This is the latest in a series of guidance designed to aid practitioners in the increasingly complex area of discovery. It is very much to be welcomed.

It has often been said in recent times that the growth of the burden of discovery obligations has come to represent a significant barrier to access to justice. It has been anecdotally reported that, in certain categories of cases, the cost of complying with discovery orders can approach 50% of the total cost of the litigation as a whole. Against that background it is hardly surprising that issues arising out of discovery have been the subject of significant debate. Furthermore, discovery covers a wide range of areas. At least those practitioners involved in the document based type of litigation in which the cost implications are most acute are likely to be very familiar with how the discovery system operates and, increasingly, with how good practice can at least reduce the cost without significantly reducing the benefit of discovery. Proper disclosure does, of course, play an important role in ensuring that any relevant documentary materials are available to the Court (and the parties) so as to enhance the prospects of a fair result to the litigation in question. However, even outside of the area of document driven litigation, some level of discovery obligations can arise in a whole range of cases and may involve, therefore, practitioners who may be less familiar with the area.

The intention of this guide is to attempt to cover the broad spectrum of discovery and also to enhance good practice in the area.

The earlier guidance given in respect of discovery has already found a measure of acceptance in the courts. Indeed it may well be appropriate to refer to such guidance as a form of "soft law" which does not bind parties in any formal sense but which represents a standard by reference to which the actions of parties may be judged. There are many benefits to such forms of soft law. First, precisely because they are not binding it will always be open to any party to suggest that the particular circumstances of an individual case require that things be dealt with differently than in the way that the guidance might suggest. Furthermore, most particularly in fast evolving areas, experience will inevitably lead to the need for regular changes. Guidance can easily be adjusted to meet new circumstances or to respond to perceived shortcomings. It is much easier to change guidance than to change formal rules of court or, less still, legislation.

There can be little doubt that discovery can properly be described as such a fast evolving area. Parties to many types of litigation, including for example personal injury litigation, have come to understand that discovery may provide for litigation advantage. But it is not always clear that the best way of going about seeking and providing discovery is fully understood by all involved in such areas. It may be hoped that this guide will provide assistance as to good practice even in the most routine of cases.

However, it is in the area of large volume discovery in cases with electronically stored information that the need to identify and enhance good practice has become most acute. To the extent that technology itself can be used to help provide an at least partial solution to problems which technology may have contributed to is at the heart

of current developments. Good practice will not necessarily solve all of the problems. There is a constant need to keep rules of court and other underlying aspects of procedural law in constant review in such fast evolving areas. However, there can be little doubt but that, even within the parameters of existing procedural law, the adoption of good practice on all sides may reduce the scale of the problem. Having a readily accessible account of what those most experienced in the area consider to be good practice can only act to the considerable benefit of everyone involved.

I am more than mindful of the very significant amount of work which all those involved in the production of this latest guide have put into the project. Our thanks are due to them in significant measure. I have, however, little doubt but that the fruits of their labour will be of significant benefit to litigators, whether involved in small scale cases or major document driven litigation and to the courts. It gives to judges an insight into how experienced practitioners feel that the process of discovery should ordinarily progress. It gives judges a potential standard by which to reference the way in which parties have conducted the discovery exercise. Of course judges may be persuaded, in individual cases, that there were good reasons for a party acting in the way in which they did. But knowing what the "industry standard" is can only be of benefit. Also this guide should provide comfort to practitioners (and in particular practitioners who may be venturing into areas with which they are not particularly familiar) with a basis on which to assess not only what they should do themselves but also whether what has been suggested from their opponent is reasonable and in accordance with normal practice. All of those advantages can only improve the process of disclosure and thereby act as an aid to what must always be the aim of any procedural law being to achieve the maximum likelihood of producing a fair and just result at the minimum cost to the parties. I recommend this guide to anyone who has even a tangential exposure to the discovery process.

The Hon. Mr Justice Frank Clarke

The Supreme Court

9th November, 2015

Chapter 1 Introduction

This Good Practice Discovery Guide was prepared by a Sub-Committee of the Commercial Litigation Association of Ireland to provide assistance to practitioners in the form of good practice recommendations and guidance aimed at managing the increasing challenges posed for the Courts, legal practitioners, and litigants in relation to discovery in disputes.

This Guide is drafted by reference to the usual life cycle of a matter; however it is sometimes necessary in practice to alter the stage at which particular steps should be taken in the discovery process. We have placed considerable emphasis on the practical issues in dealing with discovery and have made recommendations at certain points which are not currently prescribed by the existing Court rules. At all times our focus has been to streamline the discovery process with the particular imperative intention of maintaining proportionality and avoiding unnecessary costs in relation to discovery.

This Guide contains an overview of the typical steps undertaken in a discovery project. While the information and recommendations in this Guide stand independently, they are partially drawn from, and are consistent with, recognised international standards.

We hope that practitioners faced with the task of discovery, whether for a litigation, regulatory review, or investigation will find the content of this Guide helpful. At a high-level, this Guide will apply to all matters which involve the collation, searching and review of volumes of documents with some of the detailed sections only applying to complex litigation matters.

This is the second version of the Guide, the original having been published in April 2014. This second version combines both the legal and practical aspects of discovery and is therefore a lengthier document than the first version which referred to a separate technical guide. Updates to the Guide include a comprehensive set of appendices which may be used throughout the discovery process, which include detailed guidance and templates for many aspects of the process. Users of this second version may find some of the more detailed provisions are unnecessary for them to consider where discovery projects are smaller and more straight-forward. Users are therefore advised to use their discretion in that regard.

Acknowledging that good practice is an evolving concept, any observations or suggestions on the discovery process can be submitted to the Association through the website or to any of the Association's committee members.

Terms highlighted with an * have a detailed explanation available in the Glossary at Appendix N.

Chapter 2 Principles

Organisations and/or legal advisors may choose to follow these principles at an organisation/firm wide level for all matters which involve discovery. Ideally, parties to a matter will agree that these principles will apply to the matter and write to each other at the commencement of the process to confirm such agreement.

1. Perfection in the discovery process, reflecting the identification and disclosure of every possible document which ever existed, is typically an unreasonable and disproportionate expectation.

2. Hardcopy data and Electronically Stored Information (or ESI*), in all their forms, are discoverable and should be considered in every matter.

3. Parties should take all steps necessary to preserve sources of data, as soon as they become aware of a matter which is likely to require discovery.

4. Parties should, at a minimum, write to each other and ideally meet and confer on all aspects of a discovery project at the earliest possible time, but no later than when requests for discovery are exchanged. A documented discovery plan should be prepared in all matters.

5. The costs of discovery should be proportionate to the value under dispute in commercial matters. In matters where a financial value is not in dispute, the costs of discovery should be proportionate to the value which any documents discovered would bring to the matter. Proportionality should take into account the accessibility of data and the cost of retrieval, in addition to the cost of searching, reviewing, and production. Parties should not be required to produce deleted or residual data absent a demonstrated need and relevance.

6. Technology should be used to efficiently manage the process and data where at all possible in order to minimise the costs incurred. Printing electronic data or photocopying hardcopy data multiple times for the purpose of discovery will generally be a waste of time and money. Both should be avoided where technology provides an efficient solution and parties are encouraged to agree the use of a common review platform at the outset to facilitate as seamless an exchange as possible. The exchange of data electronically is to be encouraged.

7. ESI should be produced in a format which allows the receiving party the same ability to access, search, and review the data as the producing party. Basic metadata* should be maintained and produced, including document name, author, recipient, date created, and date last modified.

8. The integrity of data should be maintained, but production of irrelevant material should be kept to a minimum. As such, irrelevant portions of document families, such as irrelevant attachments, need not be discovered or produced.

9. Duplicate families of documents and/or duplicate portions of email threads need not be produced, however an audit trail of any such deduplication should be maintained, should the need to inspect duplicate documents be required later in the process.

10. Basic metadata fields are sufficient for scheduling purposes (save perhaps where some additional coding might be required) where the document is being produced. Where a document is listed in a schedule and not being produced (such as a privilege schedule), it may be necessary to provide an additional description beyond that of the document name. The description should enable the party to whom discovery is being made to understand the reason why the document is not being disclosed but should not be so detailed that it indirectly discloses the content of the document.

11. The solicitor owes duties to their client and the court to ensure that the discovery is thorough and properly made. The process must be conducted by appropriately qualified professionals with the necessary experience. Such professionals should be able to demonstrate that they possess the necessary skills and experience to carry out their duties in an efficient and effective manner.

    Note: It is important in this regard to ensure that a comprehensive collection of all likely relevant materials is made and conducted independently of the custodians and that the collection is forensically sound.

12. A discovery audit file should be maintained by all parties' legal representatives in order to record decisions taken and views in respect of relevance and privilege.

13. When a project is complete, protocols should be followed to close down the project and to ensure that it can be easily re-activated in the future if required.

Chapter 3 Outline of discovery phases

A discovery project will typically follow the phased approach set out in this guide. However, it is frequently the case that a discovery project will be iterative in nature, with some phases being repeated as more information comes to light. In addition, not all phases will be required on every project.

The phases involved in a typical discovery project include:

1. Identification Chapter 7	To identify custodians and sources of data which may contain information which is relevant to the matter.
2. Preservation Chapter 8	To take steps to preserve data where they exist, so that they may not be altered or destroyed in advance of collection. This includes the legal hold process.
3. Collection Chapter 9	To obtain a copy of the data sources identified, so that they can be processed and searched for data of relevance to the matter. It is important to acquire the copy in a manner which does not alter the original data.
4. Processing Chapter 10	To convert the data sources collected into a format which will facilitate their efficient searching and review. Early Case Assessment (or 'ECA') may be carried out to get a high-level view of the information and perform searches for key data. Data may then be filtered, if required, using filtering filters such as date range and keywords, which are used to identify documents which may be of relevance. Alternatively, the data sources may be prepared for the use of predictive coding.
5. Review Chapter 13	To perform a manual review and determine relevance and the privileged status of data highlighted as potentially relevant. This may be an entirely manual review, or utilise predictive coding.
6. Analysis Chapter 14	To take a deeper look at specific data, for example to determine its provenance.
7. Production Chapter 15	To produce a copy of the data which have been identified as relevant through during the review phase, in addition to a schedule of the relevant data.
8. Presentation Chapter 16	To prepare for, and to present, data in Court in an efficient manner.

While not always possible, you should aim to plan and execute a discovery project so that each phase needs to be completed only once. For example, collecting data for three rather than two custodians while onsite would typically have little impact on the amount of work required, whereas returning to collect the third custodian's data at a later date would likely double the effort. The same can be said for changes and/or addition of data at the review phase. Not only is effort required to add data at a later stage, but additional effort is typically required to integrate it into the existing

workflow. As with all projects, planning and getting it as close to right first time leads to significant efficiencies. This is especially so for ESI where repeating a process or phase at a later date can be disproportionately expensive.

A number of **appendices** have been included in this Guide. The primary focus of these appendices is to provide detailed guidance and template/sample documents which may assist in a discovery project.

A Discovery project checklist	This checklist may be used as an aide-memoire to ensure that all key points of a discovery project are addressed.
B Overview of discovery for parties	This provides a brief overview of discovery which may be provided to parties who have little or no prior experience in the process.
C Sample legal hold communications	This provides a set of sample emails/letters which may be used as a template for legal hold communications.
D Document identification questionnaires	This provides a set of questionnaires used at the identification phase to identify sources of potentially relevant documents.
E Managing audio and video data	This provides an overview of how audio and video data may be managed throughout the discovery process.
F Understanding deduplication, families, and threads	This provides an overview of what deduplication is and how it is used in discovery, what families of data are and how they impact discovery, and finally what email threads are and how they can be managed during discovery.
G Technology Assisted Review	This provides an overview of Technology Assisted Review ('TAR'), including Analytics and Predictive Coding. It includes common use cases for discovery.
H Sample discovery plan	This is a sample discovery plan which should be used to share information between the parties and the Court.
I Sample review plans	This is a sample review plan which should be used to document and plan the review by the producing party.
J Sample request for voluntary discovery	This is a template/sample letter of request for voluntary discovery.
K Sample affidavit of discovery with sample schedule	This is a template/sample affidavit of discovery with a sample schedule. This may be used as the basis for drafting an affidavit of discovery by parties.

L Consolidated version of current rules	This is a consolidated version of the current Irish Court rules which relate to discovery.
M Overview of legal privilege	This provides an overview of legal privilege and guidance of how it impacts the discovery process.

Chapter 4 Preparing for discovery

Parties will frequently complain about the disproportionately high costs, both measured financially and in management time, which are associated with litigation and other forms of dispute resolution and investigations. As this financial and time investment is contributed in a large part to dealing with discovery the best way to minimise such costs and time commitment is for such clients and/or their legal advisors to take proactive steps to prepare for discovery/disclosure at the earliest opportunity.

4.1 Discovery team

Quite a significant amount of the preparation for discovery can, and should be front loaded. Given the myriad of issues that may arise in the discovery process there is a clear need for factual, legal, and technical input at the start of the process. No one function within an organisation will have all of the skills to ensure the level of preparedness required to deal with a discovery request. Accordingly one of the first steps should be to establish a Discovery Project Team. At a minimum this team should include the main contact person from the client business/organisation charged with leading the input into the case, a Project Manager from the client business/organisation, a Legal Representative and an IT representative. In larger and/or time critical matters there will also be a need to employ an IT consultant who has access to eDiscovery technology and expertise in dealing with eDiscovery (if this expertise is not available in house). A person from the client organisation with knowledge of hard copy records and archive procedures should also be included.

In the context of certain matters it may also be useful to involve an expert (in the subject matter of the dispute) to assist framing and responding to a discovery request. He/she can: (1) advise what they he/ she requires from the opposition to enable him/her form a view as to the merits or otherwise of the case; (2) advise on the potential impact of particular documents on the case held by the party (which may not be immediately apparent); and (3) advise the team as to what technical documents should be collated.

4.2 Document types

It is important for those involved in a discovery project to have a high-level understanding of data types. At a high-level, data may be grouped into two categories:

- **Structured** – This is data (or records) which is stored in a structured format, such as a database. Examples include financial records, HR records, and customer databases, amongst many others. It is typically straight-forward to identify and locate such data as it is stored in a manner designed to facilitate easy retrieval.
- **Unstructured** – This is data (typically not formal records) which is stored in an unstructured format, such as email and folders containing loose files. The manner in which they are stored does not easily lend itself to identifying and locating data of potential interest. It is these unstructured data sources which pose the most challenge in discovery.

Data may also be classified as active (readily accessible), inactive, residual, or legacy (not-readily accessible):

- **Active data** is actively in use and can generally be accessed in the system in which it was created, such as the custodian's computer. This data can be accessed immediately, without any need for restoration or reconstruction. This is the primary source of data in a discovery. While active data may be relatively easy to access and collect, it can also be easily deleted or altered, thus early preservation is vital.
- **Inactive data** is not immediately accessible, such as data relating to completed activities which have been archived. Such data is often stored in a different format (such as compressed) and in an off-line location (such as backup tapes). The services of an IT professional are typically required to access this data.
- **Residual data** cannot be easily viewed using standard computer systems, such as hidden or deleted data, or data which has been erased, fragmented, or damaged. This typically requires forensic expertise to collect, recover, and interpret.
- **Legacy data** is created by software or hardware which is outdated and/or has become obsolete (legacy systems) and/or has been decommissioned. This data can be difficult to restore without the systems originally used to create and/or store it.

As with all document production requests, a party requesting inactive, residual, or legacy data must demonstrate that the relevance and materiality of such data justifies the (sometimes substantial) cost and effort of including it in the process. Consideration should also be given to whether data is available from a more accessible active source, such as substantially duplicative backup.

Documents may be stored in a number of formats:

- **Native documents** are those in the electronic or hardcopy format in which they were created and maintained.
- **Near-native documents** are those which need to be converted to a different electronic format to allow them to be managed as individual documents. For example, emails stored within a database or mailbox are typically extracted and converted into individual documents for each message (e.g. MSG or MHTML files).
- **Near-paper or image format** whereby a native document is rendered into a picture of itself in a non-editable electronic file. Essentially a 'picture' of the document is taken as it would exist if it were printed to paper. Based on the print settings of the document or computer, information can be lost or altered through the process.
- **Paper documents** are those originally created (e.g. handwritten) on paper, or those electronic documents which have been printed to paper.

The location where documents may be stored might include:

- **On-site** at a client's premises.
- Stored in a **client-controlled data centre**.
- Stored in a **shared data centre**, or in **the cloud**.
- Stored in publically accessible **online systems**, such as Dropbox, LinkedIn or Facebook.

It is important to note that the variety of data types and potential locations is constantly changing. Current common data types include hardcopy documents and ESI which incorporate not just email and loose files, but now include audio and video data, instant messages such as SMS text messages and a variety of online messaging systems (LinkedIn, Facebook, WhatsApp, etc.). Further, social and professional media websites also include combinations of messages, emails, and documents in various different formats. Appendix D contains sample questionnaires which may be used to identify potentially relevant data types and sources.

4.3 Good document management

Organisations that manage data in an organised and efficient manner will likely find it easier to deal with the burden of making discovery, particularly so in the early stages of identification, preservation, and collection. In summary, if an organisation has a good understanding of what data it holds and where it is held, this will make it easier to identify and retrieve it. This is the essence of good information governance and the foundation of efficient data management.

The topic of data management is far too broad to be covered in this Guide. There are however a number of areas where an organisation may look to improve its data management in relation to discovery:

- Establish and maintain a data classification policy and process, whereby different types of data are managed based on their priority to the organisation.

- Establish and maintain a data retention policy and process, whereby data is only retained for as long as it is required by the organisation and any regulations, and is then disposed of at the end of its useful life.

- Establish and maintain a data map which details the different types of data managed by the organisation. This should include lists of all systems, the type and classification of the data they hold, ownership, and access requirements.

- Establish and maintain a process for retaining and accessing data from historical systems and previous employees, including access to encrypted data.

- Implement a procedure for labelling potentially privileged communications when generated. This will greatly assist with the identification and ring-fencing of potentially privileged data.

There are a number of stakeholders that are key to achieving efficient data management and information governance in every organisation. They will typically include the Head of IT and Chief Information Officer or members of their teams. Where no formal information governance policy is in place within an organisation the completion of a discovery exercise can be a useful starting point in establishing one.
When considering good document management, parties should be conscious that additional costs incurred during discovery due to prior or existing poor document management may not be recoverable as costs in the matter.

4.4 Discovery response plan

For parties and legal advisors who have frequent requirements to undertake discovery, it is prudent to prepare a response plan. The following areas should be addressed in a tailored discovery response plan:

- **Notification and initial assessment** – This details how a new matter might be notified to a party and what information needs to be gathered in order to make an initial assessment. It also includes details of who will make the initial assessment of the matter and sets out reporting lines.
- **Approval process** – This details the approval process required within the organisation to decide upon and approve next steps, which will usually include approving the litigation hold.
- **Tailored litigation hold process** – This should contain a tailored version of the litigation hold process outlined in Chapter 7 below, including common technical measures used by the organisation.
- **Information map** – An up-to-date information map for the organisation should also be maintained in the discovery response plan.

4.5 Document security

Parties taking part in the discovery process, and especially legal advisors and the outside service providers who have been engaged to manage data on behalf of their clients, should ensure that appropriate security controls are in place at all times to protect data. This should include, as appropriate, access controls, encryption while in transit (and sometimes when at rest), and the secure disposition of data as soon as it is no longer required.

Chapter 5 On instruction

5.1 Briefing the party on the nature and extent of obligations

A solicitor should advise their client about its discovery obligations and the impact of the discovery process when first consulted in relation to a potentially contentious matter. In practice (unless the matter has been admitted to the Commercial List) there may be a significant number of months before requests for discovery are exchanged by the parties. However once a party becomes aware of a potential dispute it is obliged to take steps to identify and preserve all documentation which it holds which may be potentially relevant to the matters in dispute (see Chapters 7 and 8 and Appendix C).

Appendix B includes an overview of the discovery process which may be sent by a solicitor to their client when first instructed in a contentious matter.

The obligation to retain all data which may be of relevance extends to all hard copy documents and all ESI including, but not limited to, emails, SMS text messages, instant messages, back up data, faxes, spreadsheets, Word documents, and audio/video data, as well as data which may be stored online or contained in social media. In circumstances where a client fails to identify and preserve any data which is ultimately deemed to be of relevance to the matter, a Court has discretion to impose costs sanctions against such a party and may direct that this party bears the additional costs of having to retrieve and/or restore 'lost' ESI. In circumstances where this is not possible a Court may draw an inference from the fact that such data no longer exists and cannot be produced at the hearing of the action.

5.2 Assembly of the discovery team

At this time, the discovery team outlined at 4.1 should be identified and an initial meeting held to discuss the process to be undertaken.

5.3 Commencing an audit file

It is recommended that a discovery audit file is opened as soon as instructions are received and is maintained and updated throughout the proceedings. All steps undertaken and decisions made in the discovery process should be recorded in this file, including:

- Track details such as how the list of potential custodians and data sources was complied.
- Capture all of the information regarding the legal hold, including when it was put in place and when and the reviews and reminders to ensure that potentially relevant documents have been retained were carried out.
- Record any 'considered' decisions made about excluding time periods, sources, or parties from the discovery exercise.

- Record any filters agreed and any subsequent changes.
- Set out the logic to any decisions made during the process in relating to privilege or relevance.
- Identify any particularly significant issues that arose at any stage during the process and set out the steps taken to deal with them.

Maintaining an audit file makes it much easier, long after such decisions were made, to recall why such a particular decision was taken. This will be very important if one of those decisions later becomes the subject of particular scrutiny in court, e.g. if the other party to the disputes challenges a claim of privilege over a particular document or questions why certain documents have not been discovered, but also to set out to the other party or to a court why such an approach was adopted. If the approach taken ultimately comes under scrutiny, reference to an audit file at least will assist in demonstrating to a court the reasoning behind decisions and the timing of them. Although a court may not ultimately endorse the decisions made, if it is convinced that there was a logical and reasoned motivation for the decisions made, it will be less likely to impose sanctions than might otherwise be the case. The audit file might also explain why certain documents (later shown to be relevant) were not included with the documents discovered. Finally, it will also serve as a record of the work completed should costs be disputed.

There is no entitlement for the other parties to the matter to have access to a solicitor's discovery audit file, as it will be protected by litigation privilege.

Chapter 6 Pre-Discovery request

6.1 Scope

Thus far, this guide has focused on how a party might prepare itself to respond to a discovery request. At this time, or in parallel, the party will also be considering the discovery it wishes to request of other parties. The guidance in this chapter however applies to both the producing party and the requests they may make of other parties.

Careful consideration should be given to defining the scope of the discovery request in order to focus it to appropriately identify data of relevance to the matter whilst balancing the costs of retrieval proportionately. A fundamental factor in this process is setting the parameters, or scope, of the search that the litigant will carry out in retrieving the documents which are to be discovered.

What constitutes a reasonable search is heavily dependent on the facts of each matter. Factors which may be taken into account when determining the extent of a reasonable search include:

- The nature and complexity of the matter.
- The significance of any document which might be located from the search.
- The availability of the data from other documents or sources.
- The complexity and cost of retrieval:
    - Whether the data is readily accessible or not readily accessible from a technical perspective.
    - The cost of recovering not readily accessible data.
    - The location of the data.
    - The costs of including data sources in the overall process, particularly those which are expensive to collect.

6.1.1 The time period likely to apply

The obligation on a party is to disclose all documents within their power, possession or procurement which fall within the categories which the party making discovery has agreed, or has been ordered, to discover. Parties are only entitled to request discovery of, and by extension are only obliged to disclose, documents which are relevant to the matters at issue in a case. Whether a document is relevant to a case will, to a large extent, be determined by the nature of the disputed issues in the case and by the time period during which the facts the subject of the dispute took place. Usually, a document which did not exist at the time that the facts the subject of the dispute occurred will not be relevant to that dispute. Consequently, an important factor in determining which documents might be relevant to a dispute is determining the

beginning and end of the time period during which the facts the subject of the dispute took place. Once this has been determined or agreed upon, documents which were not in existence during that period may generally be considered to be irrelevant.

In the initial stages following instruction, a cautious approach should be adopted in assessing the date range to be applied to the preservation and collection exercises. All documents touching upon the matter which is now the subject of the dispute should be preserved. It is important to remind a party that even if it is possible to maintain a claim of privilege over a document, this does not relieve it of the obligation to discover the document by listing it in the relevant schedule.

It will always be possible at a later stage to refine the date range, particularly following receipt of further particulars of the claim being provided and/or following an agreement with the other party in relation to the time period to be applied. However, it might also be possible that at a later stage in the discovery process, the time period may need to be expanded. This might occur in circumstances where it becomes clear from the discovery provided that there are documents of assistance and relevance which pre-date the date upon which disclosure was agreed. Therefore a party should be very slow to dispose of documents which could potentially fall within the scope of the litigation unless, and until, the litigation concludes.

It may be possible to agree an appropriate end date for the discovery where a large discovery exercise makes it impracticable to continue the review up to the date of making discovery. There is generally no requirement to continue to make discovery after the filing of an Affidavit of discovery, save in respect of documents which predate the Affidavit and fall within the agreed or ordered discovery timeframe but which were only located after the Affidavit of discovery was sworn. Such documents must be disclosed in a Supplemental Affidavit of discovery.

6.1.2 The volume of documentation involved

The quantity of documents which the litigant holds and which will need to be searched in order to identify relevant documents is an important factor in determining the scope of the search that is to be carried out. Generally, the more documents that need to be searched in order to identify relevant documents, the more expensive the searching process will be. Documents which fall within the categories which have been agreed or ordered must be included as part of the discovery exercise.

Notwithstanding the number of good electronic review platforms which exist and various different steps which can be taken to remove duplicates and/or apply filters, it will still be necessary to carry out manual reviews of certain document and, obviously, the more documents which require manual review, the longer it will take to complete the discovery process. Parties are, at an increasingly early stage, being asked to provide an indication of the length of time that it will take to comply with a discovery request/Court order. Each party needs to be in a position to give an informed response as costs sanctions can be applied for failing to comply with a court directed timetable. In an extreme situation undue delay could lead to a claim or defence being struck out. Furthermore if the matter has been admitted to the list in the Commercial Court, the solicitors with carriage of the case on both sides are required to provide a

personal undertaking that they will use their best endeavours to ensure that the Court's directions (to include those relating to discovery) will be complied with.

Accordingly it is very important that you ascertain as soon as possible with your client and, where relevant, your eDiscovery service provider, what the likely volume of documentation will be and what resources from your client are available to assist in the discovery process generally.

If a decision has been taken to consider using predictive coding (see Appendix G) it may be necessary to seek agreement or leave to defer providing a specific indication of the time required for review until the completion of the predictive coding exercise.

6.1.3 Specific issues including accessibility/retrieval (cross-border or technical)

In discovering relevant documents, a party is obliged in the context of discovery to provide all data, which is documentation in its power, possession or procurement, which is relevant to the discovery categories. In certain circumstances this could include data held either by companies affiliated to or associated to with the party making discovery which is a party to the matter. It might also require documents which are held by foreign or agents and/or representatives of that company who are based in different jurisdictions. Therefore although an Order for discovery does not have extraterritorial effect in that it is not legally enforceable outside the jurisdiction, if a party to litigation in Ireland is ordered to make discovery and, if an entity has a legal entitlement to require another entity, which is based outside the jurisdiction, to provide certain documents to it, these documents must be discovered.

An issue may arise where such information is held in a jurisdiction in which the data protection or national privacy laws intentionally or inadvertently constrain such disclosure. Accordingly, where data is held in other jurisdictions, care needs to be taken to ensure a party would not contravene local legislation by releasing such documents to comply with discovery obligations. This is not to say that this impediment can be relied upon by a litigant as a means of shielding the production of such data. Courts will look very carefully at any such claims and will seek to identify, at a minimum, that no data were transferred to such locations after the parties became aware of the potential dispute.

Secondly, unique and particular issues arise in relation to the technical collection of information which is stored "in the cloud" or in some other 'virtual' repository. Depending on the means by which the documents have been stored, there may be considerable technical challenges associated with the retrieval of this information and which could have an impact on the collection costs.

Thirdly, data may be archived by way of technology which is now obsolete (backup drives, servers, discs, tapes, etc. – see 4.2 above) and it can be expensive and difficult to retrieve such material. It may not be necessary to retrieve all of this information and ultimately it will come down to a cost benefit analysis as to how important this retrieval is to the proceedings. The parties will have to consider the importance of the information held in this way, relative to its importance in assisting the parties and the Court to determine the dispute.

Finally, if information is stored on the personal device of a current or former employee or agent of a party, it may be difficult to compel him/her to provide this documentation. It may be necessary to review his/her contract of employment or contract to provide services as well as any IT policy. If the person on whose personal device the documents are held refuses to produce the documents and cannot be contractually compelled to do so, and the information is of importance to the dispute, consideration can be given to seeking this information directly by way of non-party discovery (if the non-party is based in the jurisdiction). However this will result in incurring additional costs as the applicant will be responsible for the non-party's costs in making discovery on a full indemnity basis.

6.2 Plan

At this time a full discovery plan[1] should be drafted. A template/sample plan has been provided at Appendix H. It is important to note that this is a living document which will evolve throughout the project. At its inception, it will be likely to only contain the steps taken in the identification and preservation phases, and the remainder of the document will set out the intended steps for the remainder of the project. This first version (version 0.1) may be used as a basis for discussions with the other party to the dispute for the purposes of reaching agreement on custodians and document types at an early stage, whilst also informing the requesting party of the intended approach to the remainder of the project. As each phase of the project is complete, the plan should be updated to reflect what was actually completed. Typically a version 0.2 draft of the plan is shared with the requesting party close to the end of the processing phase, but always before the review phase begins. This is essential as it is at this point where a very good understanding of the project will have been achieved, filters selected and/or the use of predictive coding proposed. Parties may update the draft to version 0.3, 0.4, etc. until agreement is reached and the first non-draft version of the plan may be circulated as version 1.0. At the conclusion of the project, the plan may be updated to reflect all steps actually taken during the project.

In drafting the plan, it will be necessary to assess the time periods potentially required to complete the project, prepare a budget, and start to consider the use of technology to facilitate the process.

6.2.1 Assessing the time required for retrieval and review

It is essential that an accurate estimate of the time involved is established to inform the overall timetable for discovery and to seek a realistic period for completion from the Court. The time period which the process of retrieving and reviewing documents

[1] *There is necessarily some overlap between the lawyers discovery audit file and the discovery plan. The plan is usually shared with other parties and states what has or will be done, whereas the audit file should keep track of reasons for decisions, working papers, etc. and would not be shared with other parties.*

for the purposes of making discovery will be influenced by each of the factors discussed, including:

1. The date range within scope for the discovery request.
2. The number of custodians from whom information has to be collated and their current location.
3. The means by which information is stored by both the organisation and by the individuals.
4. If any issues of the sort referred to at 6.1.3 arise.
5. Whether the party making discovery is proposing the use of predictive coding.

The time involved for review will depend on the amount of data identified as being potentially in scope. There are various technical or automated means by which the 'universe' of data can be reduced, however once these tools have been applied it will still be necessary to manually review the remaining data depending on the scale and size of the case. In some cases, it may be necessary to assemble a 'review team' or in smaller cases, it may be sufficient to have just one person conducting the review. The size and scale of the review team (and the technology used) will dictate the amount of time that will be involved in the review process. It is difficult, if not impossible, to accurately estimate how long a discovery project may take in advance of the identification phase being completed. Until the completion of the processing phase, it will be very difficult to give a precise estimate of the actual volume of data for review and/or predictive coding.

6.2.2 Budgets

Given the costs involved in a typical discovery project, preparing, monitoring and updating a detailed budget from the outset is essential to monitor costs and to ensure that they do not exceed that budget. A budget should be prepared by reference to each stage of the process and provision should be made for the diverse disciplines that may be involved in the discovery team and the fact that they can consist of internal and external parties as well as internal and external IT/eDiscovery consultants and experts.

A budget may be structured in a number of different ways. First there can be a global budget which would cover the entirety of the project. This is based on assumptions such as: number of custodians, amount of data, number of sources, location of documents, date range, etc. This is likely to change based on refinements to the discovery process as it progresses but it is of great assistance in managing expectations in relation to overall legal spend in the litigation, and secondly in determining if the proposed approach is proportionate to the matter at hand.

The preparation of a budget is also of essential importance as a means of controlling and limiting the breadth of discovery requests. Increasingly the courts are placing significant reliance on the cost of document review and production as a factor in assessing whether a particular timescale or category of discovery should be allowed.

More focused budgets can also be provided as the case advances and parties may choose to prepare budgets which break down the cost by reference to the different phases of the discovery process. Specific budgets setting out what proportions of the

overall budget will be spent on for example collection and review can be prepared with the party or eDiscovery service provider (if one has been engaged). Once the sources of the data becomes clearer it is usually possible to drill down and provide more specific detail of the various different tasks which could be involved in the discovery process. For example, if a budget is provided relating to the cost of recovering data from sources which are not readily accessible, such as backup tapes, this information can be given to the other party or to the Court for the purpose of seeking to agree either the omission of this source or the reduction in scope. This will allow parties to make an informed decision and will enable the Court to make nuanced directions as to whether it is proportionate to undertake such a recovery, all the while having regard to the circumstances of the case.

A budget can also be prepared based on the assumption of the review of a certain amount of data. If it subsequently transpires that, as a result of responding to the demands of either the Court or the other party in the context of the categories that are sought, the universe of data is substantially in excess of this original figure, this information can be provided both to the other side and to the Court as a means of convincing all parties that such a wide scale request may not be necessary for the fair disposal of the proceedings or for saving costs. It is also advisable to keep clients informed of escalating costs at every stage of the discovery process and to seek approval for additional costs in advance of incurring them.

When seeking tenders for the services of an eDiscovery services provider (scanning, collections, processing, hosting, review, etc.), it is recommended to get comparative quotes based on the same criteria. While it can be very difficult to estimate the scope and size of a project in advance, most experienced service providers will provide an accurate cost estimate for a sample project with defined estimates. These will typically include estimates for:

- The number and location of custodians. (e.g. 10 custodians in one location)
- The nature of the data and likely sources. (e.g. laptop/desktop computers, email servers etc.)
- The volume of data per custodian. (e.g. 10GB per custodian)
- The cost to collect the data sources. (e.g. it will take two people three days to complete)
- The time and cost in relation to any early case assessment / filtering that may be carried out
- The cost of system licences for reviewers.
- The time and cost to process the data sources into a searchable format, OCR, address problem files, and apply filtering criteria. (e.g. it will take four days to process 100GB of data at a cost of X per GB)

- The time and cost to bring data forward for review and provide access to the review team. (e.g. assuming 20GB of data is brought forward for review and 10 reviewers will need access for three months from one location)

- The time and cost to complete quality controls and a production (and any subsequent productions). (e.g. two days to complete quality checks and deliver a production of 10,000 documents)

- Any hourly support rates and the availability of support outside business hours, if required.

- Any ongoing cost of hosting the data.

- Any cost of archiving the data.

It is important to both identify any parts of a typical discovery project that are not covered in the cost estimate and to bear in mind that issues which are not covered in the original budget may subsequently arise and increase the cost of making discovery, but also to draw attention to the fact that costs could be ultimately incurred should that aspect need to be examined further. You should assume that all projects will be likely to have a number of problem files which will need to be managed, along with potential additions of custodians and data sources. Identifying what is included in estimates provided and what additional work may be required is important, as extra costs can cause the estimates contained in the original budget to quickly escalate to greater than the original budget if not properly managed.

When demonstrating issues such as proportionality of discovery to a court, it will be necessary to share the proposed budget with the requesting party and/or the court. This increases the importance of regularly monitoring the budget, and accordingly maintaining a budget as it relates to these issues will make any correspondence or motion easier to draft. However, save where the budget is deliberately disclosed, there will generally be no entitlement for other parties to have access to a solicitor's budget, as it is protected by litigation privilege.

6.2.3 Use of technology assisted review

An overview of Technology Assisted Review (or 'TAR') has been provided in Appendix G. At this early stage in the process, it will not be possible to determine if the use of TAR (either Analytics and/or Predictive Coding) will be applicable to the matter. It is however prudent to raise it as a prospect at this time with all involved should it be determined later that its use might bring significant efficiencies to the discovery.

6.2.4 Staged/phased discovery

As long as it can be managed within the overall timescale for discovery, it may be prudent to adopt a staged approach to discovery whereby certain custodians and/or document sources are included in the first stage and others (typically less likely to be relevant) are included in subsequent stages. This can be helpful in controlling the proportionality of the discovery and/or may be necessary due to time constraints (i.e.

one might want to cover the key custodian's documents first and then add other custodians later.)

It is important to be cognisant of the increased costs associated with staged discovery. Completing all aspects of a discovery project just once will be by far the most efficient and cost effective approach. Adding custodians and document sources will increase the cost as many aspects will need to be repeated for the new documents.

Re-reviewing the same document sources due to new categories, issues, and/or keywords is however very inefficient and should be avoided through good planning and engagement at the outset.

Chapter 7 Identification

The objective of the identification phase is to identify custodians and sources of data, which may contain information of relevance to the matter.

A party is obliged to discover any relevant document which is in its power, possession or procurement. This therefore includes documents:

- Which are or were in its physical possession.
- It has or had the right to possess it.
- It has or had the right to take copies of it or to inspect it.

There is no requirement to disclose duplicate copies of documents, but it is however important to collate all copies to determine the duplication. A copy of a document which differs at all from the original should however be treated as a separate document.

It is a prerequisite of the identification stage that the party has a clear understanding of the issues in the matter, the types of information of relevance to the matter, and what form they might take.

7.1 Identification of custodians

Clients should be asked at an early stage to identify and prepare a list of custodians (individuals) who may hold (or did hold) data relevant to the matters in dispute. Such a list can be created by speaking to key witnesses and others who may have involvement in the matter. It is also essential to determinate how and where these individuals would have stored this data e.g. on mobile phones, laptops, home computers, etc. A sample custodian-data source map is contained as Attachment One to Appendix H.

It is important to remind parties that they should consider existing and former employees when drafting a list of custodians. Furthermore, personal assistants and secretaries or other support staff are often copied on emails and should be included in the list of custodians. Even if an individual is no longer working for the organisation it may be that a laptop/desktop/other device used by him/her is still in use within the organisation. In such circumstances steps should be taken with the IT Department to identify and/or establish if any data of potential relevance to the dispute can be retrieved. Furthermore if a former employee holds documentation of relevance in their personal files, papers, devices or computers, he/she should be contacted and requested to preserve and to collate same and to furnish the documentation to the party. It is also important to clarify whether employees, even if contrary to company policy, use personal email addresses or computers to store documentation of relevance to the organisation and this should be clarified in early course.

Second, steps should be taken to identify whether data could have been stored under a username other than the custodians directly within the scope. For example, data could be located in the email mailbox or personal drive of an assistant or secretary employed by that person.

Third, anything which is held by the servants or agents of a party which could be of relevance to the dispute is within the power, possession and procurement of that

party. Accordingly the servants or agents should be requested at the earliest possible opportunity to retain all of the documentation held by them which could be of relevance to the subject matters in dispute. This would involve contacting professional advisors or service providers engaged by the party or may also include advising affiliated or connected companies (depending on the corporate structure of the group).

It may be helpful to prepare interview question templates with matter-specific questions when interviewing potential custodians to ensure all relevant issues are dealt with and addressed (this is particularly important if you are dealing with an ex-employee who may not appreciate multiple engagements).

It is likely that the identification of custodians, the identification of document sources (section 7.2), and the legal hold process (Chapter 8), may be carried out in parallel and/or iterate a number of times, as new custodians and data sources are identified, which need to be included in the process.

If the data identification questionnaire (Appendix D) is used and responses received, this would be of benefit in demonstrating the efforts made to properly identify all relevant documents and for audit purposes should an issue arise as to the completeness/integrity of the discovery made.

7.2 Identification of likely types and sources of data

Once custodians have been identified, identify the likely types and sources of data which may be tendered as evidence relevant to the facts in issue, and if agreed, the categories of discovery. Hardcopy and ESI data can be stored in any number of locations. To avoid disputes arising in relation to the completeness of a party's preservation efforts, it is critical that parties are instructed that they must understand how and where the data is located in their organisation.

In collating information, consideration should include hardcopy diaries, documents and files. Steps should be taken to ensure that ESI is identified in each potential source which is utilised by each custodian who could have data of relevance to a dispute. This will include mobile devices such as cell phones, PDA's, tablets, laptops and home computers.

Many organisations maintain a centralised computer server system that is 'synced' to users' mobile devices and makes it unnecessary and duplicative to collect ESI from other forms of media such as PDA's - however this is not universal. Therefore investigations should be undertaken with the IT function within the organisation as to what the position is before directing retrieval from individual document sources, though at collation stage it is advisable to adopt as comprehensive an approach as possible. An explanation as to the configuration of the system should be sought so that it can be averred to as necessary in an Affidavit of Discovery should it be appropriate or required.

Furthermore it may be necessary to restore data which has been archived either manually or by way of backup media. ESI also may now be stored in the cloud. It is extremely important that at an early stage in the process and prior to any data being

destroyed and/or deleted and/or overwritten that steps are taken to identify all potentially relevant data in dispute and to preserve same.

In some organisations, recordings of relevant custodians' audio/video calls may exist, in which case these also need to be scoped for potential relevance. (See Appendix E for suggested approaches to managing audio and video data through the discovery process.)

While most business information is stored in electronic format, most discovery projects will involve at least a small element of traditional hardcopy or paper documents. These will typically be lower in volume than ESI and, most importantly, the process of identifying, preserving, and collecting them will in itself result in a large part of the processing phase being complete up front.

If there is a large volume of hard copy data it may be more efficient to conduct a first pass hard copy review, to identify folders which contain relevant data for scanning. With a focused set of hardcopy data, it is typically only necessary to have them scanned into electronic format and then included in the review process alongside the ESI. This may include making the scanned data searchable, through a process known as OCR*, and may also involve having information regarding the contents of the data manually extracted, through a process known as coding*. When managing hardcopy data, it is important to ensure that the family relationship between data is retained, in addition to the ability to sort the data in its original order. This can be vital to the review process where data does not have a date.

The suggested identification checklist at Appendix D, with sample client letter, can be used as a guide to identifying potential sources of data with clients at an early stage in a matter.

Where it is unclear if a data source, such as backup tapes, may contain data of relevance to the matter, it may be prudent to undertake sampling of a statistically relevant portion of the data source in order to identify the volume (if any) of relevant documents contained therein. The results of such a sampling process can be used to determine the likelihood of uncovering further relevant documents should the full document source be included in the process.

Consideration should also be given at this early stage as to whether foreign languages will play a significant part in the documents subject to discovery as it may be necessary to include foreign language terms in key word searches for the purposes of filtering results. Parties may then plan accordingly in resourcing the process with appropriately trained personnel.

7.3 Communicating the output of the identification process

The completion of the data identification questionnaire and interviews generally enables a preliminary understanding of the likely volume of data potentially in scope; the costs and time to retrieve them; the technical issues that may arise in producing them and where there may be gaps in the data held. Although not provided for in the existing Court rules, it is recommended good practice for parties to share the headline

information gleaned from their respective data identification questionnaires long before discovery requests are exchanged.

Generally, it would be useful for each party to understand at this early stage the following information at a macro level, about each other's data:

- The date range under consideration.
- The number and identity of potential custodians and how accessible data is.
- The types of data held (i.e. emails, Word documents, Excel spreadsheets, etc.) and where they are held.
- Indicative volumes of data which may need to be reviewed and how it is proposed to filter this data to collect most relevant data.
- Any information available on likely cost and time to complete a review of this data.
- What data sources will not be included in a review exercise and why.

In complex matters, or where there may be disputes in relation to custodians at a later stage, it is highly recommended that initial discussions in the form of an early meet and confer session (see Chapter 12) take place in order to agree the number of custodians and document sources. Significant time and costs can be saved by agreeing custodians and data sources before any preservation and/or collection of data takes place.

The custodian-data source map should be reviewed and decisions made as to whether each document source identified will be included in the process. The decision not to include a data source should be appropriately documented in the audit file.

Where a data source is no longer available to the party, sufficient information regarding the reasons why it is no longer available should be recorded at this time so that an explanation can be included in Part Two of the schedule in the Affidavit as to Documents.

Chapter 8 Preservation

The objective of the preservation phase is to take steps to preserve data where it exists, so that it may not be altered or destroyed in advance of collection. This includes the legal hold process.

8.1 Legal hold process

One of the first steps in the discovery process is to inform relevant parties of their duty to preserve data which may be of relevance to the matter and to suspend routine/automatic data destruction processes. This is vital to helping ensure that relevant data is not lost or destroyed, whether deliberately or accidentally. This is best achieved by putting in place a 'legal hold', i.e. informing all of the relevant personnel, in writing, of their obligation to preserve all data that may be relevant to the actual or threatened proceedings. All actions taken to preserve data, and actions not taken, should be fully documented, along with the reasons why.

In order to be fully effective a legal hold should describe the nature of the proceedings and identify the types of data (both paper based and ESI) which may fall within the scope of a subsequent discovery request. The legal hold should be sufficiently wide to include the documents that would ultimately fall within the scope of the discovery request, any material that relates to the claims or defences associated with the proceedings and any data which might lead to a train of inquiry which would be of relevance to the proceedings.

It should be clear from the legal hold that it applies to all records, whether they are in paper or electronic form and that it covers copies of records and records which may be held at multiple locations. In particular a legal hold should specify that it includes data held in an employee's files, work spaces, computer hard drives, external hard drives, memory drives, USB sticks, voice mails, smart phones (including text messages), company servers and backup tapes, together with any data which may be stored on an employee's personal electronic devices (personal computer, iPhone, iPad, etc.) or in an employee's social media, or online storage facility such as Google Docs or Dropbox.

The legal hold should be addressed to personnel involved in the activities that are relevant to the dispute and also to the IT personnel or service providers of an organisation. Each party should be directed to suspend the destruction of and hold related data until such time as the legal hold has been lifted.

A record should be kept of the individuals to whom the legal hold has been sent and each party who has received the legal hold should be requested to send either an email or sign a document, confirming receipt and acknowledging that he/she has reviewed the legal hold, understands it and agrees to comply with it.

Given the likely duration of litigation it is advisable to issue periodic reminders of the legal hold and/or to modify the hold if it becomes apparent that the scope of the proceedings and/or all relevant information has expanded or indeed narrowed, (though any narrowing should be done with extreme caution). People will join or leave an organisation during the lifetime of the proceedings and you should ensure parties understand the need to advise new arrivals of the presence of the legal hold, as well as ensuring they have contact details for leavers to ensure enquiries can be made of them should the need arise.

Once proceedings have ended and/or copies of all documents have been secured and following consideration of whether the data may be required for any other purpose and/or similar proceedings, the legal hold can be released.

The format of a legal hold notice will be fact specific; however suggested draft legal hold notices are appended to this guide at Appendix C.

Note: In some matters, it may not be possible to immediately identify the relevant parties to include in the legal hold process. Where it is likely that the identification of specific parties will take longer than a week, and there is a risk that data could be lost in the intervening period, consideration should be given to issuing a broad legal hold notification to all personnel in the organisation alerting them to the need to preserve data until specific custodians have been identified.

8.2 Technical preservation

The legal hold process outlined above implements a 'soft' control in that it relies on custodians and other individuals not taking actions to alter or destroy data. It is therefore prudent to take additional steps to preserve data where they reside in the event that one or more custodians fail to act on the legal hold instruction. In the first instance, relevant systems which automatically destroy data should be suspended.

Technical preservation steps will be highly dependent on the data sources identified, their location, accessibility, and the capability of the available technology in place in the organisation. Steps may include:

- The most recent backups of email servers/file servers/application servers may be removed from backup rotation and stored securely.
- Technical controls may be implemented which prevent custodians from altering or deleting historical data.
- Access to hardcopy documents may be restricted to the discovery/litigation team only.

As with all steps taken in the discovery process, they should be fully documented in the audit file.

Note: Throughout the preservation phase parties and custodians must be instructed not to search for, access, or move any original data. Such access will likely alter the original data (and underlying metadata) and may cause the reliability of such data to be disputed at a later date. Any collection and searching of data should be carried out by appropriately trained personnel who will employ methods to protect the original data throughout the process.

Chapter 9 Collection

Once the sources of data have been identified, and preservation steps have been taken through the legal hold process to prevent accidental or intentional loss or destruction, the next step is to obtain a copy of the data sources, or selections of data from each source, so that they can be further processed and reviewed.

This typically involves working with the custodians, their IT provider, and any third parties who maintain custody of the data. As data sources are collected, the custodian-data map should be updated with the collection status.

9.1 Practical considerations

9.1.1 Where will the copying take place and by whom?

Will the copying take place onsite, or will the original data be taken away, copied and then returned? Alternatively, many sources may be copied remotely over computer networks. In some cases, such as data stored in remote or cloud computer systems, relying on a remote copy may be the only option (as well as the most efficient).

Who completes the collection is another important decision. Three key risks should be addressed. The first is whether the custodians themselves can be relied upon to complete an accurate unbiased collection. The second is whether the custodians and/or their IT team have the technical skills and tools to complete a collection without altering the original data. There is also the risk of varying opinions as to what is potentially relevant. Care should be exercised with self-collection by the custodians and/or those close to the matter. It may be more cost effective to engage the services of a data collections specialist and/or an IT team (who may be internal to the organisation or external specialists) with the necessary tools and skills to complete the collections independently of the custodians. Complex matters may wish to consider having the process completed or supervised by an independent specialist.

9.1.2 What is the scope of the collection?

In general terms it is more cost effective and less disruptive to a client's business if whole data source is collected and then filtered later for potentially relevant data within the source.

There are however a number of scenarios where it is more efficient to adopt a focused approach to some data sources and to only collect a focused sub-set of the source. For example, if there is a single project folder containing all data related to the disputed project, then it may be far more efficient to only take a copy of this folder rather than the whole data source (full computer server) which may hold folders for thousands of irrelevant projects. However, caution must be used when narrowing the focus of a collection at an early stage to take into consideration the probability that data may have been stored at other locations.

Generally, it is common practice for custodian-based data sources to be copied in full (such as email mailboxes, private network folders, and laptop/desktop computers),

whereas non-custodian-based data sources tend to undergo a focused collection (such as large server computers).

9.1.3 What type of collection is required?

One of the additional objectives of collecting ESI, in many cases, is to secure a forensically sound copy of certain ESI as it was stored on a particular date and time. This may be necessary if the admissibility or validity of the ESI is later questioned (and should generally be considered good practice in all cases).

There are a number of industry standard tools which are freely available which are capable of collecting data in a forensically sound manner. As with all tools however, there is a level of expertise required to use them effectively and help ensure that the original data and metadata is preserved throughout the copying process. Consideration should also be given to the use of encryption to secure copies of data. As outlined at 9.1.1 above, it is recommended that appropriately skilled personnel be engaged to complete this process. Such personnel will be adept in verifying the accuracy and completeness of documents which have been collected.

9.1.4 Hardcopy data

Hardcopy data should be scanned into electronic format, made searchable through an OCR process, and metadata extracted through a manual coding process. They should then be included alongside any electronic data throughout the discovery process. Photocopying hardcopy data is generally not as useful and cost effective as scanning into electronic format.

The collection of hardcopy data differs from that of ESI in one major aspect. The very physical effort of identifying and preserving hardcopy data usually requires that an element of review be performed in order to determine if the data source is likely to be of relevance to the matter. Because the moving of hardcopy data for scanning and coding requires significant effort, a level of effort is required to filter the documents prior to this. As such, hardcopy data collections tend to be focused and reduce the volume of irrelevant data up front. This is the opposite of how ESI is managed, which is typically more efficient to collect en-mass and then filter later.

It is important when managing hardcopy data that the ability to reconstruct their original order and family groupings is maintained throughout the process. In particular it is important to preserve the source information, including the custodian, the box description (if applicable), any box ID and any details as to the office from which the box or files came.

As with ESI, the process of identifying and collating hardcopy data should be supervised or led by a legal advisor.

9.1.5 Social/professional media

It is becoming more frequent that data and/or messages of relevance are stored on social media websites such as Facebook, or professional media such as LinkedIn. This data is typically accessible by viewing the website, but may require access provided by the owner of the content. In many cases it is difficult to obtain access to such data. There are three common approaches:

- Apply to the website owner to preserve and produce the data. This can be prohibitively time consuming and expensive (as a Court order may be required and/or website owner may reside in a different jurisdiction).

- Assuming access is public or otherwise granted, utilise the services of a specialist data collections provider who can acquire a copy of the content and verify its authenticity as a copy in Court if required.

- Again, assuming access is public or otherwise granted, simply take a picture of what is displayed, print and sign it in the presence of a witness (preferably a solicitor), and present it as a verified copy of what was viewed at a point in time.

Access to view and copy information from websites, unless publically available, should not be obtained through deception (such as posing as a colleague or friend in order to gain access).

9.1.6 Chain of custody

In advance of collections commencing, a decision should be made as to whether maintaining a written chain of custody record for all data sources will be necessary. A written chain of custody can be vital in helping to demonstrate that no unauthorised access was made to the original verified copies of data. In addition, written chain of custody can assist from a security perspective as well as an evidential perspective. Completing a chain of custody should add very little in terms of time and cost to the collections process and it is recommended to complete same.

Note: Once a copy of ESI has been made, it is typically safe to return the original for use by the custodian as the copy (particularly if a forensic copy has been made) can be later relied upon as an exact copy of the original.

Chapter 10 Processing

At this stage in the process, data will have been identified, preserved, collected, and will be collated in one central location. The objective of the processing phase is to firstly remove clearly irrelevant data types from the data set collected and convert the remaining data into a format which will facilitate efficient searching and review. Early Case Assessment (or 'ECA') may be carried out to get a high-level view of the information and perform searches for key documents. Documents may then be filtered, if required, using filters such as date range and keywords, to identify documents which may be of relevance. Alternatively, the document sources may be prepared for the use of predictive coding.

In many matters, the processing stage may also have been completed prior to agreeing the terms of discovery, as this is required in order to determine the number of documents for manual review, and therefore the cost and proportionality of the project.

The extent and nature of processing required in any given project will depend on the nature of the ESI collected, the technology being used, and the expected review process. The processing phase typically consists of the eight steps outlined in sections 10.1 through 10.8 below.

10.1 Remove irrelevant document types

In cases where a full forensic copy of a data source was collected, for example each custodian's laptop/desktop computer, these will contain large volumes of software code and other irrelevant document types. Only user-created data will likely need to be extracted from each full copy. These data types should be identified based firstly on discussing what data types the custodian used, but also by identifying the list of data types contained within the data set and verifying which ones require inclusion. The user-created data types included should be documented in Attachment Two of the discovery plan at Appendix H.

Where a focused collection of user-created data has been acquired, for example a single folder from a network share, a data type filter will typically not be required. Such focused collections will by their very nature likely only contain user-created data.

Note: Where not-readily accessible data types have been agreed for inclusion, such as deleted files, these would be recovered at this time and included in the process.

10.2 Convert into searchable format/load into database

The data set collected, less the irrelevant data types which were removed in step 10.1, are then loaded into an eDiscovery processing system, along with hardcopy data which has been scanned into electronic format. Such systems extract the data and metadata of ESI, including expanding any families of data, and store all the information in a searchable database. Initial statistics regarding the volume and types of data are then gathered.

10.3 Deduplicate

A family-level deduplication process is then run against all data in the set. This suppresses any duplicate families of data while leaving one copy of each unique family of data for further processing. While duplicate families are suppressed, the list of custodians who hold a duplicate family which was suppressed is recorded and included in the remainder of the process. This allows only one copy of the family to be considered, while also allowing the reviewer to quickly understand who held duplicates of the family. The result of this deduplication process is that the volume of data is typically reduced. Statistics as to the number and type of documents in this new data set are typically recorded. Appendix F contains a detailed overview of deduplication and families of data.

Note: It is common and also good practice to deduplicate data families and only produce one copy of each unique data family which is of relevance to the matter. Should a receiving party wish to inspect duplicate data, this may be requested after receipt of production and would typically require justification before the additional cost of inspecting/producing duplicate data would be considered.

10.4 Optical Character Recognition ("OCR")

The volume and type of non-searchable data is identified. An OCR process is then run against these documents in order to convert them to a searchable format.

10.5 Thread deduplication

Email thread deduplication is then run against all emails and their attachments. This process identifies the inclusive portions of each email thread along with the non-inclusive (or duplicative) portions of the email thread. The non-inclusive portions of the email threads are then suppressed from further processing. Email threads which are unable to be subjected to the email threading process should not be suppressed and should be included in further processing. Appendix F contains a detailed overview of email threads and thread deduplication.

Note: As with standard family-level deduplication, it is common and also good practice to deduplicate email threads and only produce one copy of each unique email thread which is of relevance to the matter. Should a receiving party wish to inspect duplicate email threads, this may be requested after receipt of production, and would typically require justification before the additional cost of inspecting/producing duplicate portions of email threads would be considered.

10.6 Manage problem documents

Quality controls at the processing phase must identify problem data which cannot be processed by whichever system is in use. These include corrupt files and encrypted files, amongst many others. Therefore, this data will not be accessible for keyword searching. (Note: their location and names, including metadata would however likely be searchable.)

When managing encrypted data, there are a number of options. The route to take will depend on the method which has been used to manage the encryption system in

place. Where an enterprise-wide encryption system has been used, it is typically possible for the IT manager of the system to provide a mechanism to unlock the data. Where individuals have set the encryption using a stand-alone system, such as simply applying a password to a spreadsheet, then it will be necessary to request the password from them.

In the event that the password/decryption key is not available, then a decision will have to be made on whether to attempt to remove the encryption by other technical means. This typically involves using specialist software. Such software can have a varying success rate, and while it is typically not a costly exercise to undertake, it can take a long time, with very little indication as to when, if ever, it will be successful.

It is therefore best to reduce the number of encrypted files to attempt to decrypt. One approach to this is only attempting to decrypt files which have been highlighted as potentially relevant either through their name, or through association with another file. For example, if an email is deemed to be relevant, however it has an attachment which is encrypted, then it may be useful to attempt to decrypt the attachment. However, if it is just an encrypted attachment to a non-relevant email, in a universe of many thousands of emails, then it may not be justified to incur the time and cost of attempting decryption.

The number of deduplicated encrypted/password protected data in the set should be recorded.

Where encrypted/password protected data are located in a data source which will not be subject to filtering, such as those in a shared project folder (i.e. likely to be relevant), all such data should have decryption attempted.

Throughout the process it is important to liaise with the opposing party to indicate that this methodology is being used and agree perhaps that if successful that the data might be the subject of a supplemental affidavit of discovery rather than delay substantive compliance within the Order of the Court.

Other classes of problem data are wide ranging. These may include very large spreadsheets which will not be viewable using the proposed review system, or complex technical drawings which may need to be converted to a different format to allow viewing. The approach to managing such data will be heavily dependent on the technology in use. Of importance is to identify the problem data and not inadvertently miss them, and then decide if they should be subject to filtering (which may not work) or brought straight for review.

Foreign language data, while not necessarily 'problem data' must be identified and a strategy for addressing them established at this time. This is due to the fact that filtering and other areas of the process, such as predictive coding and/or review, will be impacted by the presence of foreign language data (or data which has mixed languages). Keywords will need to be devised in the relevant language (having regard to considerations such as English language words which may have a small number of synonyms but which may have a multitude of synonyms in a foreign language), and reviewers with appropriate language skills will need to be engaged.

10.7 Apply filters and perform Early Case Assessment

It is usual that hardcopy data and other data sources, such as shared electronic project folders are not subjected to filtering. This is due to the fact that they will typically be data repositories for the project subject matter and therefore are likely to contain relevant data and should be reviewed in full.

The remaining ESI set will represent all data associated with each custodian and as such, the vast majority will likely have no relevance to the matter. In these circumstances, it would extremely rarely be proportionate or practical to manually review all such data for relevance.

It is usually necessary to apply filters to the data set in order to identify data which is likely to be of relevance to the issues in the matter. By way of assistance in the Discovery Plan at Appendix H, filters can be listed in Attachment Three and the total volume of data for initial manual review recorded.

It is vital to work with all involved (including the requesting party) to test the proposed filtering criteria for precision* (accuracy at identifying relevant documents) and recall* (actually finding as much relevant data as possible, while not returning too much irrelevant data). The objective of this testing is to ensure that the filtering criteria will highlight for review this data which is likely to be of relevance to the matter, whilst also managing the volume of irrelevant data for manual review. Testing should involve initial ECA using analytics tools such as clustering, categorisation, and themes, in addition to sampling the results of each filter. Timeline analysis should be used to identify any significant gaps in the data collated.

A filter report, or Search Term Report, is often produced which contains details of the combined number of documents which are responsive to one or more of the search terms. i.e. if data is responsive to two or more of the search terms, it is only necessary to review it once. It will also contain the number of combined families of data, which is the combined number above, plus their families, plus any other unique families which also contain a duplicate of the responsive data. This information is required in order to fully understand the volume of data for review and will heavily drive the decision as to how the review will be approached and managed, and if predictive coding is an appropriate technology to employ.

10.7.1 Consider predictive coding

When all of the data and its detailed statistics have been collated, the effectiveness of any proposed filters will be understood. It is very difficult to accurately decide on the use of predictive coding technology before this information is available.

Predictive coding is deployed at the review phase (Chapter 13 below), however at this stage the decision to be made will be which data sub-sets will be subject to predictive coding and if filters will be used to pre-cull this data sub-set before being sent forward for predictive coding.

Different predictive coding system vendors recommend different approaches as to whether data is filtered in advance of running the predictive coding process. i.e. some

insist that all deduplicated data/email threads are included in order to achieve the best statistical results, while others are happy for the data set to be pre-filtered before the predictive coding process. The decision should be made in conjunction with a predictive coding expert and will take into account the proportionality of the cost involved in bringing more data forward for predictive coding, versus the potential downside of pre-culling.

Generally, the data set is divided into sub-sets. Hardcopy data, shared project folders, and data which is not suitable for predictive coding (such as purely numerical spreadsheets in isolation or complex drawings or images) will usually be put into a sub-set which undergoes a traditional review, whereas emails and their attachments and other text-rich documents will be subject to predictive coding. Appendix G contains a detailed overview of Predictive Coding.

10.7.2 Developing filtering criteria

It may be possible to identify and exclude wholly irrelevant data types, or to confine the discovery review to a specific date range and thus exclude data falling outside the date range. This filtering should be done at the outset. Care should be taken to ensure that document dates have not been corrupted in any way in the course of the processing, which can occur and may require specific IT input to correct.

The next step is usually to develop keywords. Developing effective keywords is complex and requires specialist input, from a legal professional and/or an IT specialist experienced in developing and applying keywords. It is an iterative process, so that the results should be monitored to identify and exclude any obvious false positives, while ensuring that care is taken to identify potential gaps. Overbroad keywords will result in a very large review set with a significant proportion of false positives, while overly narrow keywords risk missing relevant data. Keywords also need to anticipate potential spelling errors. The legal adviser must take care to ensure that keywords are not misspelt when entered.

One approach is to split keywords into those which identify parties and those which identify issues. 'Parties-based' keywords may be used to identify specific parties in a data set. For example, the producing party may search for all data which reference the requesting party. This would return all data between the parties and all data within the producing party which reference the requesting party. This is useful if the parties only ever had communication regarding the matter in dispute.

Where parties-based keywords return a large volume of irrelevant data (typically due to the parties conversing/transacting on a number of matters), then it is necessary to use 'issues-based' keywords to narrow the search. Issues-based keywords are typically combined with parties-based keywords and focus on the specific topics which are the subject of the dispute. For example, the name of the project, account numbers or project/property addresses/locations.

By applying parties-based keywords first and then narrowing using issues-based keywords, the producing party can refine keywords on a step-by-step basis, testing each iteration.

There are a variety of ways to combine parties and issues-based keywords (and also parties and parties keywords, and issues and issues keywords, or any combination

possible). These include Boolean operators such as AND and OR, and for excluding false positives, NOT. For example, searching for "John Smith" AND "Central Bank" may bring back a vast number of false positives whereby John Smith's email signature contains the text "Regulated by the Central Bank". Proximity searching may be employed to search for "John Smith" AND "Central Bank", but not where "Central Bank" is located within three words of "Regulated by".

In matters where it is unclear if spelling is correct, it is possible to use wildcards in place of letters. For example, "John Sm!th" might return both "Smith" and "Smyth". A concept referred to as fuzzy searching can also be employed to account for spelling variances. There are a wide variety of searching techniques available. It is recommended to engage with the technology provider and/or vendor for the systems in use as they will typically be able to provide expertise in this area.

Where a dispute regarding keywords arises, it can be helpful to outline the approach taken in developing the keywords in any correspondence and/or affidavit evidence submitted to the Court.

One key rule when searching documents is that it is generally not productive to search one's own data for one's own name, business name, or email address. Such keywords will almost certainly result in all data being responsive. It is also usually not productive to search for the word 'privileged' which tends to appear in footers to emails sent by organisations.

It can be very useful to sample the outcome of keyword searching to identify any false positives before commencing the review, as this can often dramatically reduce the volume of wholly irrelevant documents requiring review and may reduce costs.

Where a party has concerns that draft keywords submitted by the other party may not be accurate or sufficiently comprehensive, or may give rise to unnecessary costs because they are too broad, this should be highlighted. Parties are encouraged to meet to agree keywords and also to remain open to sharing the statistics which arise from the use of those keywords so as to demonstrate the results being achieved. Parties should take a reasonable and sensible approach to agreeing to exclude keywords and terms that create disproportionately irrelevant returns.

10.8 Bring forward for review

Data which is responsive to the filtering criteria applied, and those which are not subject to filtering, is then brought forward for review. Their families should also be brought forward (i.e. where an attachment is responsive, then its parent email should also be included). Where duplicates of responsive data exist within another unique family of data, this other unique family of data should also be brought forward for review (e.g. where the same attachment is attached to two different emails, both emails and two copies of the attachment should be included). This approach allows decisions regarding how duplicates and families of data are managed to be made throughout the review phase.

In the event that predictive coding, without prior filtering of data, is to be used at the initial review phase, then all unique families of data (after email thread deduplication) should be brought forward for review.

Chapter 11 Discovery request

Having pleaded your case, reviewed the case made by your opponent, engaged with your client's expert witnesses and identified what data you need to contest the litigation, you are now ready to request discovery from your opponent.

The Rules of the Superior Courts (No. 2) Discovery 1999 (SI No. 233 of 1999) (incorporated as Order 31 rule 12(6)(a)) introduced the requirement to specify precise documents or categories of documents of which discovery is sought and to set out the particular reasons why discovery of each document is required. (A consolidated version of the current court rules is attached at Appendix L.) In providing the reasons for seeking discovery, the discovery request should explain how each category of document is relevant to the material issues in the case and why discovery is necessary.

A template/sample of such a request can be seen at Appendix J. This request is generally served and agreed as a separate document to the discovery plan (which focuses more on the practical aspects of how the discovery will be completed and not the contents).

Save in exceptional cases, discovery is not requested until the pleadings have closed. Consequently, parties usually exchange requests for discovery after the delivery of a Defence or, if it is required, a Reply and Defence to Counterclaim. However, the need to consider discovery issues and to comply with preservation obligations arises far in advance of this, and it is recommended to engage with discovery related issues as soon as the parties become aware that a dispute is likely.

11.1 Focus of request

One of the most effective means of limiting the extent of discovery (and thus the costs involved) is to deliver a focused request which, on its face and in its content: (a) seeks only those documents which are relevant to the dispute; (b) articulates clearly in each case why the documents are relevant and necessary; and (c) does not present unreasonable difficulty or disproportionate expense for the recipient to comply with.

A balance needs to be drawn between seeking every possible type of document that might potentially be relevant and limiting the scope of the categories so narrowly as to overlook essential documents. A practice has in recent times developed which involves parties seeking a "catch all" category of documents such as "all documents on which the plaintiff intends to rely". This can lead to excessively broad categories of discovery which in turn increase the time and expense of making discovery. As such, parties should avoid such broad requests. Parties may come to an agreement whereby additional doucments which have not been produced, but which a party intends to rely, will be provided as soon as is practical in advance of trial.

Well drafted discovery requests generally have the following characteristics:

11.1.1 Clearly defined categories tailored to the pleadings

All documents sought must be relevant to a material issue in the case. As the issues in a case are defined by the pleadings. Poorly drafted pleadings tend to give rise to

poorly drafted discovery requests which are unfocused, imprecise, open-ended, generalised, difficult and expensive to comply with. Consider whether the issues in the case can be refined through the use of targeted and precise notices for particulars or interrogatories. This may be a far more effective and less costly exercise than seeking overly broad discovery in response to vague or imprecise pleadings. In addition, Practitioners should avoid using generic phrases such as "all documents relevant to…" and seek instead to define and be as prescriptive as possible as to the nature of the data that is being requested, e.g. "invoices" or "correspondence between A and B".

The request for discovery should explain why: (a) each document category which is sought is relevant to the matters at issue in the case; and (b) discovery of the document category is necessary for the fair disposal of the proceedings or to save costs.

The relevance of each category, must be separately explained by reference to specific paragraphs of the pleadings and the particulars. Parties should not seek discovery of documents where there is merely a possibility that they will be relevant to the issues in the case.[2] In order to be deemed relevant it must be reasonable to suppose that the category of documents, when discovered, will either directly or indirectly enable the party seeking discovery to advance its own case or to damage the case of the opposing party.

In demonstrating the necessity for discovery of each category, the request must explain how discovery of each category is necessary for the fair disposal of the case or to save costs. This is more than merely stating this in the request, and a statement that discovery is necessary, without any explanation as to why this is so, should be avoided. For example, it may be necessary to obtain discovery of a particular category of documents because the information, which it is anticipated will be revealed by those documents, is not available from any other source. Where a category is likely to contain documents which are confidential, the discovery request should explain why it is necessary that the documents be discovered notwithstanding their apparent confidentiality. Subject to arrangement between the parties, it may be permissible for the producing party to redact certain confidential or commercially sensitive data – but typically only where this data is not relevant. This point is further explained at section 13.3.

---

[2] In *Hannon v Commissioner of Public Works* (unreported, High Court, McCracken J., 4th April, 2011), McCracken J. held *"The court must decide as a matter of probability as to whether any particular document is relevant to the issues to be tried. It is not for the Court to order discovery simply because there is a possibility that documents may be relevant."* (pages 3 to 4).

In *Commerciale do Pacifique v Peruvian Guano Company* (1882) 11 QBD 55, Brett L.J. used the words *"contains information which may – not which must…"* (page 62). In *O'Callaghan v Mahon* [2008] 2 I.R. 514, Hardiman J. said that the *Peruvian Guano* test *"emphasises the reasonable possibility and not the certainty of usefulness…"* (page 619) and went to emphasise that documents which were required for purely speculative investigation should not be discovered.

The necessity for discovery will be considered having regard to all the relevant circumstances including the burden, scale and cost of the discovery sought. Categories of discovery sought should be confined to what is genuinely necessary for the fairness of the litigation. Practitioners should avoid merely asserting necessity without supporting facts.

While there is no explicit reference in Order 31 Rule 12 to the concept of proportionality, it is closely aligned to an assessment as to the requirement of "necessity on the facts of a particular case". Practitioners should ensure that there is proportionality between the extent or volume of the documents to be discovered and the degree to which the documents are likely to advance the case of the applicant or damage the case of his opponent, in addition to ensuring that no party is taken by surprise by the production of documents at trial.

11.1.2 Limit timescale involved

Practitioners should ensure that, as far as possible, the applicable timescale for every category of discovery is clearly defined so that only documents generated or coming into the deponent's possession over a particular period of time are captured by the request. The timescale should be limited so as only to encapsulate those documents which are both relevant and necessary. Timescales should be referenced specifically to the pleadings and fully explained. The exercise of limiting timescales may significantly reduce the amount of discovery captured by a particular category and thus the expense and burdensome nature of it, therefore avoiding an argument that it may be disproportionate or unduly oppressive. Different timescale limitations may be sought in respect of different categories of discovery, depending on the matters at issue.

11.1.3 Avoid duplication between categories

A well drafted discovery request should contain little or no duplication between categories of discovery requested. The duplication of categories of documents not only extends the scope of discovery to be made but it is also problematic insofar as the categorisation of discovered documentation is concerned.

11.1.4 Seek documentation in searchable format

Where the discovery sought includes ESI then the party seeking discovery should specify in their discovery request whether they seek the production of any documents in searchable form (or a format which allows the receiving party the same ability to access, search, and review the documents as the producing party) and, if so, whether that party seeks the provision of inspection and searching facilities using any IT system owned or operated by the party to whom the request is directed.

The only express restriction in the Rules on the entitlement to obtain such inspection and searching facilities is where the Court determines that "unreasonable expense" would be incurred for a party to search the data provided electronically. In that case, the court can order that the party providing the discovery should make available inspection and searching facilities using its own IT system, so as to allow a party seeking discovery to avail of any search functionality available to the party ordered to make discovery. In reality however it is generally more expensive (for all parties) to review discovery documents using the other party's IT system. In practice, an actual

request for interrogation of the other party's IT system is likely to arise only in exceptional circumstances, such as where the recipient cannot review and understand the data (where, for example, it exists on a bespoke software platform); where there is a concern about the integrity or completeness of the discovery; or where access to particular metadata is required for a specific reason. This Guide recommends this approach as a matter of good practice.

In order to comply with this aspect of the Rules you should seek detailed instructions as to the extent to which the relevant data is likely to be stored electronically and if there are any specific technology issues, before drafting the request for voluntary discovery.

11.1.5 Agreement to make voluntary discovery

An agreement by a party to make voluntary discovery has the same effect as if a court order in those terms had been made (Order 31 Rule 12(7)), provided that the party requested to make voluntary discovery was informed at the time of the request that:

a) Voluntary Discovery was being sought pursuant to Order 31 Rule 12.

b) An Agreement to make discovery would require it to be made in like manner and form and would have such effect as if directed by court order.

c) Failure to make discovery might result in an application to penalise the default.

Therefore, where an agreement to make discovery is reached, the party who has agreed to make discovery is obliged to produce an Affidavit as to Documents in the proper form and is liable to the same remedies for default in making discovery as apply for breach of a court order for discovery.

Every request for voluntary discovery should address points (a) to (c) above. The request should also confirm the time limit for response and for making discovery.

Practitioners receiving discovery requests should be thorough in examining the full extent of the discovery sought against their clients and be particularly live to the concepts of relevance, necessity and proportionality. Specific instructions should be sought from clients in relation to agreements to make discovery with the practitioner having clear instructions and an understanding of the practical issues which compliance with each category would raise.

11.2 Discovery requests against non-parties

Where it appears to the Court that any person or entity not a party to an action is likely to have or have had in their possession, custody or power any documents which are relevant to an issue arising or likely to arise in the action, the Court may order discovery or inspection of such documents, or may give leave to deliver interrogatories. The Rules regarding *inter partes* discovery apply equally to non-party discovery.

A party who seeks non-party discovery must request specific categories of documents and give reasons why each category is relevant and necessary. A request for discovery from a non-party must be proportionate and if it is oppressive may be resisted on

that basis. The Court must be satisfied that the documents are not available to the applicant from another source.

The party seeking discovery from the non-party must indemnify the non-party in respect of all costs reasonably incurred in making discovery. Discovery is made on oath in the usual way and the applicant is entitled to seek inspection of the discovered documents under Order 31 Rule 29. The obligation to provide the documents arises where the applicant undertakes to indemnify the non-party in respect of the costs of making discovery, although in practice a non-party may be reluctant to disclose the documents before its costs are discharged and the party seeking discovery may be content to discharge the costs prior to receipt of the discovery. This is, however, a matter for agreement between the non-party and the applicant, as there is no entitlement under the Rules to resist inspection pending payment. In order to avoid disputes in respect of non-parties' fees it is recommended that the non-party provide an estimate of cost to the applicant prior to commencing work with a discovery plan showing the methodology to be applied and any anticipated costs.

Chapter 12 Meet and confer

It is highly recommended and is good practice for the parties to meet and confer at the earliest possible stage with a view to agreeing the scope (categories of discovery) and approach (discovery plan) for the discovery. In complex matters, where it is expensive to repeat the preservation/collection/processing phases later, this should take place first at the identification phase (as outlined at 7.3). In all other matters, it should take place no later than the end of the processing phase (Chapter 10), when the volumes of documentation involved will be known and also when any search and filtering strategies have been tested.

The aim of the parties engaging with each other at an early stage is to encourage parties to reduce costs by focusing on the documents most likely to be of relevance to the matter in a proportionate manner. With the inclusion of ESI in the discovery process, it can be disproportionately expensive to repeat phases of the project should disputes arise late as to the scope or approach taken.

The information gathered by each side in compiling their discovery plan (Appendix H) should ideally be shared between the parties in advance of this meeting. Ideally this meeting would take place before discovery requests are exchanged. It is accepted however that in practice parties may not engage at this level until after requests have been exchanged. What is ideal however is that this engagement happens as soon as possible and certainly before parties finalise and agree categories and start completing their respective discoveries. Ideally, engagement would be ongoing and flexible enough to allow decisions to be made and revisited in light of further information coming to light during the discovery process. It is good practice to provide one's draft discovery plan to the other parties a number of days (preferably 7 days) in advance of meeting. This will allow parties to prepare any requests for further information or clarification in advance.

Any meetings should ideally be held in person where possible. Where appropriate, especially in discussing and agreeing the terms of the discovery plan (rather than the discovery categories), it is highly recommended that an IT or eDiscovery expert attend meetings where these sometimes technical topics may be discussed. Such meetings between specialists tend to be an iterative process and may or may not take place at the same time as meetings between legal advisors, where items such as discovery categories may be discussed.

12.1 Agree scope and approach to discovery

Parties should engage with each other with a view to agreeing all information identified in the template/sample discovery plan (Appendix H). In practice this may take several meetings in particular as the extent of the technical issues become more obvious to the parties. These types of information to be discussed include:

a) The identities of the custodians of potentially relevant documents within and outside of a party's organisation whose documents will be discovered.

b) The date range that the searches to be carried out by the party making discovery will cover.

c) The data sources which will be searched in making discovery, (e.g. hard copy, servers, back-up tapes, cloud storage, personal computers, smartphones and tablet devices) and the accessibility of the data sources.

d) Steps taken to preserve data and the method(s) used/proposed for collection.

e) The methods used to filter the collected data, including for ESI: removal of irrelevant data (such as computer code), the approach to deduplication and thread deduplication, metadata filtering (such as by custodian, sender/receiver, author, date range, etc.), keyword searching, and any technology assisted review/predictive coding methods.

f) The methods used to address exceptions, such as encrypted data and non-searchable data and whether data will be converted from their native format.

g) How the review will be conducted, including quality control processes and if a technology review platform will be utilised. If so, whether this will be a common discovery platform to which both parties can upload and exchange the data discovered by all parties.

h) How redactions will be managed.

i) The schedule and production format, including how families of data will be managed, whether data will be produced in native format, and whether irrelevant portions of data will be produced.

j) The format of the schedules to the Affidavit of Discovery and what fields will be provided by each party.

k) How inspections will be managed, if required.

l) How data will be managed and presented at trial.

It is highly recommended that parties load the data into a format where any proposed filtering (date ranges, keywords, etc.) can be applied easily. This will enable informed discussions regarding the suitability of the filtering criteria selected and the effort (and cost) required to complete any manual review. It may simply not be possible to decide upon or agree filtering criteria without having tested any such criteria.

At this stage the party making discovery should share the full discovery plan (including filters to be applied to the data) and it is good practice to have a dialogue about the sufficiency of the overall approach and proposed filters (and/or predictive coding). Where a party chooses not to engage constructively in relation to the discovery plan notified to them, a Court may have regard to this in any subsequent interlocutory applications concerning the adequacy of the discovery made.

12.2 Finalising terms of discovery

Following the delivery of requests for discovery, and the exchange of correspondence and meetings between the parties concerning discovery, the parties will either: (a) reach agreement on the totality of their respective discovery obligations; or (b) fail to

reach agreement on some material aspect of their discovery obligations and require their dispute to be resolved by a Court by way of motion for discovery.

12.3 Agreement on terms of discovery

Parties to litigation should make every effort to reach agreement on their discovery obligations, as bringing motions for discovery adds greatly to the costs of the action and increases the delays in resolving the case. An agreement to make discovery should be reduced to writing (in the form of documented categories) and be as precise as possible. Ideally, the agreement reached on the discovery plan should also be documented. The agreement should include terms that deal not only with the wording of the categories to be discovered, but also all other matters that are likely to arise when the party concerned is making discovery.

At a minimum, the agreement to make discovery should include terms which address the following:

- The wording of the categories to be discovered. The agreed wording should be as specific as possible and ensure that the party making discovery will be able to quickly identify whether any data falls within the wording of that category and ought therefore to be discovered. Parties should avoid using generic phrases when describing the categories.
- The deponent who will swear the affidavit of discovery.
- The period of time from the conclusion of the agreement to when the discovery will be delivered.
- Ideally it would also include the details contained within the discovery plan (Appendix H).

While Form 10 of Appendix C RSC requires that the party making discovery lists the documents in a manner corresponding with the categories in the agreement, documents may correspond to more than one category. Parties should attempt to reach agreement on whether the deponent should (i) list each document only once under the category to which it is corresponds most closely; or alternatively (ii) list each document under each category to which it corresponds. This guide recommends as a matter of good practice that parties list documents by reference to the most relevant category only. This approach can save significant time and costs during review.

12.4 Failure to agree – Motions for discovery

As with all interlocutory applications to a court, a party bringing a motion for discovery should issue a notice of motion and ground the application on affidavit.

The notice of motion will set out each separate form of relief which the applicant seeks from a court. This will usually involve listing the categories of documents for which an order for discovery is sought but directions on certain other aspects of discovery may also be included. This may include aspects of the discovery plan (which does not include the categories for discovery). For example, the court may be asked to make directions on the date range of the searches to be carried out by the party making discovery or the data sources that are to be the subject of that search.

The affidavit grounding the discovery motion should be sworn by the solicitor who acts for the party seeking discovery. It is generally preferable for affidavits used in interlocutory applications to be sworn by the parties themselves and not by the solicitors acting on their behalf. However in the case of applications for discovery, it is more usual for the solicitor to be the deponent as he or she will have more detailed knowledge of: (i) the material issues in the case; (ii) the reasons for which the categories of documents are relevant to those issues; and (iii) the necessity of those documents for the fair disposal of the case and/ or the saving of costs. The grounding affidavit sworn on behalf of the party seeking discovery should include averments dealing with the following:

- A brief background to the history of the dispute should be given and the material issues in the case summarised. In summarising the positions adopted by the parties on each material issue, the deponent should identify the relevant paragraphs of the pleadings in which the parties engage on the issue in question and exhibit any further particulars of pleading which have been delivered which illustrate the position adopted by either party on that issue.

- Reference should be made to any sworn replies to interrogatories which supplement the position taken by either party on the material issues.

- While it is not necessary to identify by name the witnesses whom the party seeking discovery intends to call at the trial, it is helpful to outline whether the party intends to adduce oral evidence from witnesses as to fact or expert witnesses on any of the material issues. This will assist the court in gauging the necessity for discovery.

- The categories of documents in dispute should be listed and the request for discovery by which they were originally sought should be exhibited. Order 31, Rule 12(1)(b) RSC requires that the affidavit grounding the motion furnish the reasons for which discovery of each category is sought. While it is common for parties to simply exhibit the request for discovery and allow the letter to speak for itself, it is better practice to recite the reasons for which each category is sought in the body of the affidavit itself. It is also very helpful to include the proposed discovery plan(s) and highlight the areas of disagreement from a practical perspective (separate to disagreements regarding categories).

- The positions respectively adopted by each party on the disputed categories and/or approach should be summarised with reference to the correspondence exchanged between them. All inter partes correspondence dealing with discovery should be exhibited to the affidavit. Additionally, the deponent should exhibit any other documentary material which illustrates the reasons for which discovery is sought. Alternatively, any pre-action correspondence which suggests that the respondent possesses, or has the right to possess, documents of the sort of which discovery is sought should be exhibited.

- Where several categories are in dispute, it may be clearer to separately address each category and to summarise the reasons given in correspondence by the respondent for its refusal to make discovery in the terms sought in respect of each category. Any revised wording suggested by the respondent should also be set out.

In some cases, the grounding affidavit may be accompanied by additional affidavits sworn to illustrate why discovery of some or all of the disputed categories is sought. For example, where an expert witness is required to inspect the categories of document sought in order to proffer an opinion in the case, and he or she cannot simply set out the reasons for this in correspondence, an affidavit from that expert may be delivered.

A respondent to a discovery motion may deliver a replying affidavit in order to set out the reasons for which he or she objects to making discovery in the terms sought by the applicant and/or why a more limited form of discovery is appropriate. A party usually declines to make discovery of a particular category because it believes that: (i) the category sought is not relevant to a material issue in the case; or (ii) discovery is not necessary for the fair disposal of the case or to save costs; or (iii) the scope of the discovery request is disproportionate and unreasonable in the circumstances. The replying affidavit sworn on behalf of the party resisting discovery should include averments dealing with the following:

- Where the respondent believes that averments made in the applicant's grounding affidavit have incorrectly described the material issues in the case or have misstated the position adopted by the respondent either on the pleadings or in correspondence, the position should be corrected by the deponent.

- Where the respondent claims that a particular category or approach is not relevant to the material issues in the case, the deponent should identify any particular part of the pleadings or concessions made in replies to interrogatories or correspondence which demonstrate either that the alleged issue is not actually a matter in dispute or if it is, that the applicant has not correctly described the parameters of that dispute.

- Where a respondent claims that discovery of a particular category of documents is not necessary for the fair disposal of the case or for the saving of costs, the court will generally balance the litigious advantage which the applicant claims discovery of the documents will confer against the prejudice that the respondent claims will be caused if an order for discovery of that category is made. Objections as to necessity frequently involve contentions that: (i) the applicant has alternative means of proof available to him other than the documents of which discovery is sought; (ii) the quantity of documents likely to fall within the category are so voluminous that an oppressive burden would be imposed on the respondent were an order for discovery to be made; or (iii) the documents likely to fall within the category to be discovered will contain irrelevant confidential information the disclosure of which will harm either the respondent or some other connected party.

- In certain circumstances, the court may be prepared to refuse to order discovery where it is satisfied that the applicant has alternative means of proof on the material issue in question available to him. Where a respondent objects to making discovery on this basis, the replying affidavit should exhibit documents or correspondence which the respondent received from the applicant, either prior to or in the course of the litigation, which show that the applicant has available to it an alternative source of evidence on this point. For example, correspondence from the applicant

which asserts that it has received expert advice on a particular issue which has come to a particular conclusion may be deployed in aid of a submission that the applicant has no need of the category now sought. Another example is where the applicant for discovery may already have received the requested documentation through a data access and/or freedom of information request or where the requested information is publicly available. While there could be merit in having the completeness of that documentation sworn on Affidavit, it would appear to be an unnecessary expense to require production of the same documentation again.

- Where the respondent asserts that an order for discovery of the particular category would require the review and collation of voluminous quantities of documents, details of the material to be reviewed should be given. This may include information as to the range of likely custodians, their geographical locations, and the range of document sources in which the documents are likely to be stored. This information may be supplemented by an affidavit from an eDiscovery/IT/document retrieval expert averring as to the number of documents that are likely to be stored within this document universe, the detail of the automated retrieval and searching techniques that will be required to collate the documents and the likely man hours and costs involved in such a process.

- Where the respondent asserts that an order for discovery of the particular category would be likely to lead to the disclosure of confidential information, the deponent should set out the basis on which a duty of confidentiality is said to arise, the identities of the persons to whom the duty is owed and why disclosure of the category sought is likely to impinge on that duty.

On the hearing of the discovery motion, the court will review the parties' affidavits and hear oral submissions on their behalf. Where an order for discovery is made, it is usual for the applicant to take a note of the wording of the order and then to email a draft of the order to the relevant court registrar so that the order may be perfected.

The court will also deal with the question of costs. Generally, where an applicant succeeds in obtaining an order for discovery for all or most of the categories in dispute, he or she will be entitled to an order for costs. Conversely, where the court has declined to make an order for discovery or has made only a limited order requiring a small portion of the categories in dispute to be discovered, costs will be awarded to the respondent. If it is not possible to determine which of the parties has been successful having regard to the scope of the discovery which has been ordered, the costs will generally be made costs in the cause.

A factor which will affect the exercise of the court's discretion is whether both parties were prepared to meet to discuss the parameters of discovery prior to issuing the discovery motion. Where one party was prepared to meet but the other was not, and the court is satisfied that had such a meeting been arranged it is likely that the discovery motion would not have been brought, the court may order that the party who declined to meet should bear the costs of the motion. Courts typically look unfavourably on parties who refuse or fail to engage constructively in the process and particularly those parties who refuse to meet and confer on discovery issues.

Parties should also be aware of the possibility under the RSC (see the consolidated rules at Appendix L) of applying for a variation of an order or agreement for discovery, if the discovery originally ordered or agreed proves to be unreasonable having regard to the costs or other burden of providing discovery – or conversely where further discovery is necessary.

Parties should be aware that a party seeking further and better discovery in respect of the discovery process undertaken should be able to show how the approach undertaken has resulted in documents of relevance to the categories being omitted from the discovery. This is best demonstrated by showing actual gaps in the discovery provided in terms of timeframes and/or the existence of documents (or portions of email conversations) which the responding party has not produced. A party suggesting that documents of relevance might have been missed due to the approach (discovery plan) undertaken by the opposing party will find it difficult to succeed with a motion for further and better discovery. It is highly recommended that evidence of omissions are raised between the parties before any such motion is considered.

Chapter 13 Review

The objective of the review phase is to perform a manual review and, by reference to the categories of discovery, determine relevance or potential relevance of data which remains following the application of various filters. This may be an entirely manual review. Alternatively, in suitable cases all unique families of data may be brought forward (with or without filters such as keywords previously applied) and predictive coding may be used to filter the documents to those likely to be relevant, after which a manual review may be performed.

Reviewers should be able to search, annotate, redact, flag, and bookmark individual documents or collections of data by reference to project specific issues (e.g. hot documents, privileged documents, or documents relevant to particular issues/categories in the matter). By review stage each party should:

a) Agree and document a review plan for a structured review, and a plan for ad-hoc searches. To include issues such as:
    i. How duplicates and families of documents will be managed.
    ii. How documents will be batched/divided in a structured review.
    iii. How marks, tags, annotations, redactions will be applied, and how privilege will be addressed.
    iv. How clustering of near duplicate documents and email threads will be utilised.
b) Identify documents requiring review by subject matter experts.
c) Avail of technical features in software to increase review speed including hits highlighting, email thread analysis, Boolean tagging, conditional tagging, and automated bulk tagging, etc.
d) Embed appropriate reviewer progress monitoring and quality control processes.

The review phase is overseen by a solicitor, who must be able to stand over the quality and completeness of the discovery when ultimately produced. A detailed sample review plan is attached at Appendix I. This contains a number of typical approaches to review and guidance of when each approach might be appropriate.

13.1 Categories

Each relevant document must be categorised in line with the categories of discovery as agreed and/or ordered. In practical terms, this can make for a lengthy and costly review process. A practice that has developed in recent years and which is growing in acceptance is that practitioners agree to list the documentation by reference to the most relevant category but include an averment in the Affidavit of Documents that this may not be the only category to which the document is relevant.

A sample averment in this regard is:

In making discovery [**INSERT NAME OF PARTY**] has for ease of reference listed each document being discovered under one of the [**INSERT NUMBER**] categories of documents within the Ordered Discovery. It [is/may be] the case, however, that there

is an overlap between various categories in the Ordered Discovery whereby a document might be relevant to a number of categories. [**INSERT NAME OF PARTY**] has been advised that it is not obliged, and could not reasonably be expected, to identify every category under which a particular document might be listed. Accordingly, while the documentation listed in the First Schedule under the [**INSERT NUMBER**] categories comprises the totality of documentation which the [**INSERT NAME OF PARTY**] is in a position to produce under the Ordered Discovery, the [**INSERT NAME OF PARTY**] does not thereby suggest nor is it the case that all the documents listed under any particular category are relevant to that category nor do they comprise all of the documents possibly relevant to that category out of the documentation being produced [but in listing the document as relevant to that category has considered that category to the most relevant.]

Where parties assign multiple categories to a document, the document should be listed and provided only once, with the categories it is responsive to listed together in the schedule.

13.2 Privilege

Privilege is the entitlement to refuse disclosure of the contents of a document, the existence of which is discoverable. Privilege may be claimed over a document and not the fact of its existence. All relevant documents over which privilege is claimed must be listed in Schedule 1 Part 2 of the Affidavit as to Documents. The documents should be individually listed and the type of privilege being claimed should be specified. A more detailed overview of the different types of legal privilege is set out in Appendix M. The assessment of whether documents are privileged are made by a solicitor familiar with the facts of the case and the discovery process.

13.3 Redactions

Circumstances may arise in the course of discovery which permit the redaction of portions of documents. In all cases where it is proposed to redact documentation it is considered best practice to advise the opposing party in advance of the swearing of the Affidavit as to Documents of the reasons why redactions are necessary and also to aver to the redactions and the basis for same in the Affidavit itself. As there is no express entitlement in the Rules to redact information, redaction should be used as sparingly as possible and only where it can be justified not least as the time involved in reviewing multiple chains of email containing data to be redacted can become an inordinately expensive exercise. The main reasons for redaction are:

- Privilege
- Relevance
- Commercially sensitive and/or confidential information
- Protecting personal data relating to other parties who are not party to the proceedings

Where data is relevant to the categories of discovery sought it must be discovered whether or not it is personal data, commercially sensitive and/or confidential, save where leave of the Court is obtained. In certain circumstances it may be permissible to redact commercially sensitive information where the information concerned is not relevant to the categories of discovery or does not advance the opposing party's case (but is part of a document which is – otherwise the document would simply be marked as not relevant). A party is not entitled to redact relevant commercially sensitive or confidential information as of right and must either seek agreement or leave from the Court prior to making such redactions.

While it is not usually required or practical to redact portions of a document which are not relevant (rather the whole document is relevant or it is not disclosed at all), the issue of data protection should be considered in all matters. As outlined above, personal data from a data protection perspective will usually need to be discovered if it is relevant. However, should a document contain other personal data unrelated to the proceedings, then this may warrant redaction. For example, a list of financial interests where only one is relevant to the proceedings. Where personal data relates to other data subjects not party to the proceedings, the producing party must consider its obligations to protect these data subjects from onward transfer of their personal data. For example, a document which lists borrowers to a bank will likely need to have all of the borrowers not relevant to the proceedings redacted. A balance must be struck between discovery obligations and the clients' duty as a data controller/processor to protect data relating to other unrelated parties.

Given the proliferation of email as a form of communication and the resultant chains of email where such chains require redaction it is important to ensure that the text requiring redaction is consistently identified throughout the chains and to remember that the text may not occur in the same place in the chain of communication as it develops over multiple emails. This challenge can be largely overcome by utilising email thread deduplication, as duplicate portions of email threads which would need to be redacted identically would be suppressed.

When documents have been redacted it is important to verify that the document name and other metadata is also redacted from the schedule, in the case where the document name and/or other metadata might reveal the content which has been redacted. For example, a document named 'Impaired loans of John Smith and Mary Jones.docx', will need to have 'Mary Jones' redacted from its name in the schedule if Mary Jones' loan information has been redacted from the content of the document.

It should be borne in mind that the deponent of the Affidavit as to Documents may have to explain the basis for redactions at trial and it is recommended to note in the schedule the reason for redactions by way of acronym – e.g. 'RR' = redacted for relevance; 'RP' = redacted for privilege, etc.

As noted above, agreement to redact should be sought prior to carrying out a redaction exercise. In the event that a party is not agreeable to the redactions proposed, a recommended solution (prior to seeking the intervention of the Court) is to offer to the opposing legal team, an inspection of the material which it is proposed to be redacted. In practice, this tends to resolve matters.

13.4 Dealing with problem documents

Section 10.6 outlines some approaches to dealing with problem documents at the processing phase. However, many problem documents cannot be easily identified until manual review has been completed. This is due to the fact that it may be disproportionate to manually check every document in advance of review.

Typically reviewers will have the option to mark a document as 'Technical Issue' during the review and these documents can be segregated for resolution of any technical issues. This is typically in the case of large spreadsheets or drawings, or complex document types brought in as part of wider families of documents.

13.5 Client review

It is recommended that throughout the discovery process, the proposed deponent of the Affidavit as to Documents is kept apprised with progress. It is vital at the review stage that the deponent of the Affidavit as to Documents takes time to review the draft affidavit and schedule, and the documentation listed in the schedule.

Chapter 14 Analysis

The objective of the analysis phase is to take a deeper look at a document, for example, to determine its provenance. Structured data, such as accounting systems, can also be analysed to generate insights into specific transactions, or patterns of transactions.

The key difference between the review phase and the analysis phase is that the review phase is typically focused on determining if a document is relevant to a matter based on its contents, while the analysis phase is not so much focused on the content of a document, but its provenance, or its lifecycle, based on its metadata. This may include analysing duplicate versions of a document which have been suppressed through the deduplication process.

It is important to note, that the analysis phase is not always necessary. Analysis may be required for a single document, a subset of documents, or an entire category of documents. It may be useful to know the provenance of a document. The requesting party may also seek to dig deeper if certain questions arise in relation to the integrity of a document - however a specific court application will normally be involved before such background information is disclosed.

Chapter 15 Production

The output of the review phase will generally be a number of documents deemed relevant, some of which will also be marked privileged or partly-privileged and some may have redactions applied.

The objective of the production phase is to produce a copy of the documents identified as relevant through the review phase, in addition to a schedule of such documents. It is vital that parties engage early in the process in order to agree the production format, so that the receiving party is presented with a format which they can easily use. Requested changes to the format of the production after review may require elements of the review to be repeated and increase costs.

Adequate time should be allowed for the production phase to be completed. Modern productions can be complex as comprehensive quality checks are required as well as conversion of problem document types, and the production of electronic schedules, amongst many other tasks. While it may be possible to complete a small production in a number of hours, in a large discovery most productions take a number of days to complete to an appropriate level of quality. It is recommended that in large scale projects the parties ensure that an appropriate amount of time is incorporated into the timetable for discovery to cater for production and that as far as possible the data is exchanged in electronic format on an agreed platform.

15.1 Families of documents

Consideration for how families of documents (see Appendix F for a detailed description) are managed at production will have been included in the review planning and strategy. It may be helpful to include a schedule of irrelevant family members which have not been produced, and/or include a slip sheet for each document which has not been produced. This can assist is demonstrating that the document has been withheld intentionally, rather than due to a technical issue or oversight. It should not be the case that wholly irrelevant documents are produced just because they are associated with a family where only one member is relevant. Neither should it be the case that documents should be redacted in full in these circumstances. If a document is redacted in full, it would simply be more cost effective not to produce it in the first place. While it is prudent to only produce the relevant portions of document families, care should be given as to whether documents should be produced as orphans. For example, it may not be acceptable to produce an email attachment and not the parent email, simply because the parent email not be relevant.

15.2 Production format

There are a vast number of technical options available for production format, making it vital to engage with the receiving party early in the process in order to gain agreement. One common (and usually the most cost effective) option is that documents which have been marked as relevant are produced in their native format, along with a schedule listing their original metadata details, categories, etc. in a 'load file*'. Documents which have been redacted are produced alongside non-redacted documents, but in a redacted (e.g. PDF/TIFF) format. Documents which have been

marked as privileged would not be produced (save where redacted as part privileged), and a schedule of such documents is produced.

The RSC state that if requested production should be in a searchable format and in the format which they are held by the party making discovery. This is often referred to as native format. In most cases, this is the most efficient way to produce ESI, as it does not require the producing party to incur the cost of converting it to a different format. If a party decides not to produce documents in native format the reasons should be clearly explained and agreed before the documents are produced. Unless requested, documents should not be converted into a less accessible format (such as electronic images or to paper) for production purposes.

Where an OCR process has been completed to convert non-searchable documents to a searchable format, the results of this process may also be provided to the requesting party. Given the nature of OCR technology, such text should be provided on an 'as is' basis, with no assurances that the technology has rendered complete and accurate text.

In the event that a party converts ESI into a different format, steps should be taken to ensure that elements of the ESI, such as metadata, are not unintentionally lost or obscured in the process.

In the interests of efficiency, parties might consider agreeing a common production format, including schedule format and document numbering system, which will allow the output of all parties' productions to be merged electronically for trial. For example, parties could agree to provide a schedule in spreadsheet format with hyperlinks to underlying documents, with party A's document numbering starting with an A and party B's document numbering system starting with a B.

In larger discoveries where the documentation is contained on a discovery database the parties should agree to exchange the documentation online. To facilitate this, the parties should agree to use the same discovery technology platform as this will significantly reduce the time and cost at the point of exchange. Where this is not possible, the parties should endeavour to ensure that their systems are compatible.

Note: The actual documents produced are often renamed as their production number, with their original electronic file name being included in the schedule instead. This is helpful as often electronic file names are too long to be easily moved between disparate systems, so using the document production number as the name (typically a short alphanumeric string) avoids such compatibility issues between systems.

15.3 Schedule

A sample Affidavit of Discovery and schedule is attached at Appendix K. The sample schedule includes suggested standard fields and format. Further detail is included in Attachment Four of the Discovery Plan at Attachment H.

15.4 Inadvertent disclosure of privileged and/or other documents

Even with comprehensive quality controls in place, the complex nature of discovery projects can result in data which should not be disclosed, accidentally being disclosed. Parties may wish to come to agreement as to how data which is inadvertently disclosed may be handled. This is often referred to as a 'clawback agreement' and accounts for how a party might notify the other party of which data should not have been disclosed and what steps might be taken to remedy the situation. This may be in the case of privileged documents and/or documents subject to data protection restrictions. The absence of a claw back agreement does not dilute the obligation on a solicitor not knowingly to read or deploy an obviously privileged document belonging to another party and to notify the other solicitor of the receipt of the document promptly. Further information in this regard is set out at Appendix M.

15.5 Inspection

Once the Affidavit of Discovery has been served the opposing party is entitled to inspect the documentation listed in Schedule 1 Part I of the Affidavit.

It is most efficient if electronic copies of non-privileged documents being produced are provided at the same time as the schedule (typically on USB key or transferred online). This allows the documents to be electronically linked to the schedule.

Production of documents, whether copies or electronic versions, does not prevent a party seeking to inspect original documentation, which can be requested if necessary. Such inspection may be required in order to verify original signatures on handwritten documents, or more frequently when the documents produced cannot be rendered into a readable or understandable format without the use of specific technology which only the producing party has access to (and would be disproportionately expensive or impossible for the receiving party to obtain access to a similar system to read the documents). Examples of this include bespoke accounting and auditing systems, whereby the information is meaningless outside the original system used to generate and store it, or complex imaging or mapping systems whereby it is not possible to view the images outside the original system.

Where additional relevant documents are identified after production and inspection has been completed, a party has an obligation to produce these with a supplemental affidavit of discovery.

Chapter 16 Presentation in court

The objective of the presentation phase is to prepare for, and to present, documents in Court in a manner which facilitates their efficient presentation and the running of the matter.

Often, the most efficient method to present a document at a formal hearing is electronically on-screen in its native format. This saves considerable cost in printing bundles of documents, and time in leafing through large bundles of documents to find the one under discussion at a certain point in proceedings. However, in some cases, it may actually be more efficient or cost effective to print key documents for presentation.

Many venues, such as court rooms, are not yet equipped with technology which allows data to be presented to the courtroom on-screen. Therefore, adequate preparation and planning should be undertaken when deciding on the best method to present the data at a formal hearing and it is vital to liaise with the Court's Registrar in advance of trial to see what can be arranged, and to ensure that the trial judge is happy to review documents electronically rather than in hardcopy.

In cases involving large volumes of core documents, ideally the parties should employ trial management technology to avoid the proliferation of hardcopy files in court as far as possible. This requires early liaison between the parties, the Court and the Courts Service.

In some other jurisdictions courts have technology systems in place and simply instruct the parties as to the format of documents required to use the system. In such courts, the equivalent of the Courts Service manages and runs the systems. Currently in Ireland there are no courts that are equipped with such systems. As such it is incumbent upon the parties to agree themselves on the use of technology, select a single technology provider, and agree how costs will be managed (usually split between the parties based on the number of users each party has). There are a number of technology providers in this area who can work with the Irish Courts Service, should the parties and the Court think it helpful. General international experience shows that significant time and cost savings can be achieved at trial stage when such systems are deployed.

The court will expect technology to be used in a manner which ensures equality of arms as between the parties as regards access to the technology and training required for its use. Parties should therefore seek to agree a suitable platform for document management during the trial as well as an appropriate timetable for testing and for training for the court, counsel and parties to the litigation. For such technology to work effectively at trial it is important that the parties agree a common document identifier convention for documents. Counsel should be prepared to identify the relevant document identifier and page or paragraph in respect of each document being produced at trial so that it can be produced on screen promptly. Where possible parties should provide the technology service provider with the document identifiers required for the following day, confidentially if necessary, to ensure the smooth running of the trial, although in practice this may be difficult. Most

technology providers providing these services provide a person skilled in the use of the system for the duration of the trial.

16.1 Co-operation in relation to preparation of core books

In advance of the start of a trial, all parties to litigation which has been admitted to the Commercial List must agree on the content of the core books of documents which are likely to be frequently referred to by the Court, Counsel and witnesses in the course of the trial. It is also recommended best practice to try to do the same in all other divisions of the High Court.

Where electronic trial management technology is not being used, it may help if core books are produced in hardcopy format and all other information is kept in electronic format (assuming it can be easily accessed at trial).

16.2 Amalgamated books of discovery

A simple method of arranging core books is often to agree that the documentation on which each party proposes to rely is presented in strict chronological order rather than by category of discovery or theme. This way, the books can be easily supplemented at trial if parties wish to add additional documentation.

As a general rule of thumb, key documents referred to in the pleadings or replies to particulars should be included in the core books. Further, if there is a direction for the service of witness statements then any documents referred to in the witness statements should also be included and ideally list the document number or identifier ascribed to the document in the Schedule to the Affidavit as to Documents. Parties should wherever possible limit the amount of documentation contained in the core books to that documentation which is truly 'core'.

Appendix A Discovery project checklist

Preparing	1. Brief client 2. Assemble team 3. Commence audit file 4. Draft discovery plan 5. Draft budget
One – Identification	1. Identify custodians 1. Identify document sources 3. Early stage meet and confer
Two – Preservation	1. Complete legal hold process 2. Complete technical preservation steps
Three – Collection	1. Decide on scope and type of collection 2. Plan collection logistics 3. Collect and scan hardcopy documents 4. Collect ESI
Four – Processing	1. Remove irrelevant document types 2. Convert into searchable format and load into database 3. Deduplicate to unique families only 4. OCR 5. Thread deduplication 6. Manage problem documents 7. Apply filters and perform ECA (consider predictive coding) 8. Publish for review
Discovery request	1. Develop discovery request to other parties 2. Meet and confer to agree discovery plans
Five – Review	1. Establish review team 2. Develop approach to review and document review plan (two-pass, single-pass, predictive coding)
Six – Analysis	1. Identify the need for analysis and document analysis plan 2. Complete analysis and report
Seven – Production	1. Produce documents and schedules 2. Arrange inspection, if required
Eight – Presentation	1. Agree presentation format and use of technology

Appendix B Overview of discovery for parties

What is "Discovery"?

Discovery is a part of the litigation process when a party in the proceedings must disclose to the other, the existence of all relevant documents which are "*in your power, possession or procurement*". Discovery is made either voluntarily on request, or by Order of the Court, and is set out in the form of an Affidavit which includes a list/schedule of relevant documents.

What documents are deemed "relevant"?

Any documents[1] which relate to the issues raised in the pleadings and particulars and more specifically fall within the list of categories which have been agreed at the time of seeking voluntary discovery or directed by the Court.

What does "in your possession, power or procurement" mean?

You are obliged to disclose all documents which are not only in your physical possession but also documents which are in the hands of an agent, servant or related company even if it is only temporarily. If you have an enforceable right to obtain a document which is relevant to a category, this must also be discovered.

You will also have to account for documents which were once in your possession, power or procurement but no longer are as part of the discovery process.

How should you prepare?

You should retain and preserve all documents at all your locations, both active documents and those in storage. You should retain and preserve all copies of a document (duplication must also be discovered). Any meeting notes, handwritten or printed should also be retained and preserved.

In the likely event that you conduct internal communications via email, identify who within your organisation is likely to have been party to any of the emails circulated in relation to a relevant topic.

You should investigate how information is stored on your computers, to include account information, memos, correspondences or emails. You should then clarify if such storage plans are in fact followed by the individuals using them and if not find out how they archive/delete matters. Original files should not be re-arranged or altered in any way – the whole file should be provided for assessment even though you may feel only part is relevant. If information which might be of relevance is held by an external storage/archiving company – ensure that such files are not destroyed as part of a routine document destruction policy.

Immediate steps should be taken once litigation is threatened to properly preserve electronic data and/or documents that can reasonably be anticipated to be relevant to litigation. Emails in printed form, stored in memory or in back up files and even deleted, may be discoverable.

Document and data destruction policies within your organisation may need to be suspended, pending identification of all potentially relevant documents.

---

[1] A document has been deemed to be anything in which information of any description is recorded (and therefore includes electronically stored information such as emails, SMS text messages, instant messages, backup data, excel spreadsheets etc.

What should I do with the documents once I have collected them?

Take in and hold securely any original documents. In the case of electronic documents, these should involve an IT professional and may require the assistance of a data collections expert who will help ensure that originals are not accidentally altered or destroyed in the process. Prepare a duplicate record of any documents you are likely to require access to during the duration of the litigation. Original files should not be altered or re-arranged as part of the process.

What if a document may have been destroyed in accordance with a document retention policy?

Once litigation has been contemplated you must cease any routine destruction of documents or data which might be relevant. This includes any documentation which may feel might be damaging in your case. You must notify all those who may have documents relevant to the action to cease destruction immediately. You should also properly record all efforts made by you in this regard.

If you are aware of documentation which once existed but which you now cannot locate and you feel may have been destroyed, its existence must still be recorded in the Affidavit and an explanation given as to its likely whereabouts provided.

Do I have to discover commercially sensitive and/or confidential documents?

Yes. Confidentiality is not a bar to discovery and relevant commercially sensitive or confidential documents must be discovered. In certain instances, the Court may permit commercially sensitive information within the body of a document to be redacted (blacked out), or may direct special arrangements to protect the confidentiality of documentation. You must set out the reasons for redacting documents and ideally seek to agree a basis for redactions prior to discovery being made. Documents received for the discovery process can only be used for the purpose of the proceedings. If they are used for any other reason this could amount to contempt of court. All employees connected with the proceedings should be advised of this.

What if a document is privileged?

A document that is relevant to the categories sought/ordered will have to be discovered i.e. listed in the Affidavit of Discovery, even if it is privileged. However it will not have to be produced to the other party.

What documents are legally privileged?

There are two main types of legal professional privilege; Legal advice privilege and Litigation privilege.

Litigation Privilege **-** Any documentation prepared for the dominant purpose of contemplated or pending litigation is "privileged". Documents created for more than one purpose, of which one is litigation, may not be privileged.

Legal Advice Privilege **-** This privilege will arise in circumstances where litigation is not in contemplation or in being. The general rule is that it will only extend to communications between a lawyer and his/her client which concern the seeking or provision of legal advice. It does not apply to mere legal assistance. Further guidance in relation to Legal Privilege is at Appendix M.

How do I protect the privileged nature of such documents?

Ensure that any documents which are privileged are not sent to external parties. Such disclosure is likely to be considered to be a waiver of any claim to privilege.

What is the procedure once all relevant documents have been located?

Your legal team will then typically review all documents and all relevant documents will be delivered in their present state, reviewed and then scheduled if (a) relevant and (b) relevant but privileged.

All relevant documents will then be copied and the documents listings will be incorporated into a draft Affidavit of Discovery. The deponent of the affidavit should then review the schedules and discovery documents, where possible. Where there is a very large volume of discovery and this manual review is not possible for the deponent, a detailed process of briefing the deponent on the process and the discovery made will be required so that the deponent can properly swear the affidavit.

Who should be the Deponent for the Affidavit of Discovery?

The Deponent (i.e. the person swearing the Affidavit of Discovery) should be someone who has knowledge of the documents listed and an authority to represent you/your company. The Deponent may be cross examined on oath about the contents of the Discovery and the steps taken and searches made in the production of the Affidavit of Discovery. Furthermore as part of the Affidavit of Discovery the Deponent must swear that he/she is aware of their obligation to discover all documents (to include electronically held information) within a party's power, possession and procurement which are of relevance to the categories agreed/ordered and which may "enable the party receiving the discovery to advance its own case or to damage the case of the party giving discovery and which may fairly lead to a train of inquiry which may have either of those consequences".

Appendix C Sample legal hold communications

Sample text for legal hold emails can be seen below. There are a number of steps in the process which should be considered:

- The communication is typically sent by email, or in hardcopy
- It is distributed to all custodians, including IT custodians, and 3rd parties who may have documents relevant to the matter
- While it should be broad in nature at the outset, every effort should be made to focus the request so as to not to overburden custodians
- Responses must be tracked, including responses confirming agreement
- Periodic reminder notices should be sent for the duration of the legal hold
- There should be a mechanism for releasing the legal hold when it is no longer required

C.1 Template legal hold email

To: [Potential Custodians]
From: [External/General Counsel]
Subject: [Matter reference] – Legal hold – Preservation of relevant information

Dear [Custodian name],

As you may be aware, we have been notified of a potential [litigation/regulatory review/complaint] regarding [the work we completed for [insert client or project name]/the product we supplied to [insert client name]]. We intend to vigorously [pursue/defend] these [proceedings/review/complaint].

During the course of this [litigation/review/investigation] it is important that we are able to make our paper files and any electronically stored information (ESI), which could be of relevance, available to our legal team. Also, if discovery requests are made in the course of the [litigation/review/investigation] we may need to make them available to lawyers representing [complaint]. It is therefore essential that you take immediate and affirmative steps to preserve all paper documents and ESI which may be of relevance to this matter which are in your custody or control.

Please note that this will include all documentation and information stored on your work laptop, mobile phone, blackberry, home computer, and any other portable devices, such as USB keys, etc. in addition to information stored within our [project/engagement] files and our shared servers. It also includes all forms of documentation such as correspondence, diaries (electronic or hardcopy) and instant messages. The above list is intended to give examples of the types of information/records you should preserve, but is not exhaustive. If you have any queries as to whether you should be retaining something, please do not hesitate to raise them directly with [me] or [secondary contact].

Where you are unclear as to whether a document may be potentially relevant, you should preserve that document for a more detailed review by our legal advisors at a later stage. Please do not search for or attempt to access potentially relevant information at this time. Doing so may alter the documents unnecessarily. The only preservation steps required are to not access, alter, or delete, any potentially relevant information. If you fail to preserve these materials it could be detrimental to our position in the [litigation/review/investigation].

As you will be aware, we have a records retention policy in place. During the time that this legal hold is in place, you must suspend compliance with records retention policy in respect of those documents that may be relevant to the matter.

Our IT team has been notified of this legal hold. They will be working with us to help ensure that we implement the legal hold effectively. We will follow-up with more information as the [litigation/review/investigation] proceeds, including advising you as to when the legal hold is no longer required. In the interim, please respond to this email and confirm:

1. That you have reviewed this notice
2. That you understand the notice and agree to comply with it

If you have any questions please contact [me] or [secondary contact] at [insert contact details].

Regards, [External/General Counsel]

C.2 Template legal hold reminder email

To: [All custodians only]
From: [External/General Counsel]
Subject: [Matter reference] – Legal hold – Reminder

[Forward original full legal hold email]

Dear All,

Please be reminded that the legal hold is still in place until further notice. We will follow-up with more information as the [litigation/review/investigation] proceeds, including advising you as to when the legal hold is no longer required. In the interim, please respond to this email and confirm:

1. That you have reviewed this reminder and the original notice below
2. That you understand the notice and agree to comply with it

If you have any questions please contact [me] or [secondary contact] at [insert contact details].

Regards, [External/General Counsel]

C.3 Template legal hold release email

To: [Each custodian who data collection has been completed fully]
From: [External/General Counsel]
Subject: [Matter reference] – Legal hold – Release

[Forward original full legal hold email]

Dear [Custodian name],

The legal hold referred to in the email below no longer applies to you. If you believe that you still have potentially relevant information which has not been collected to date, then please contact me immediately.

Many thanks for your assistance in this matter.

Regards, [External/General Counsel]

Appendix D Document identification questionnaires

There are two sections to the document identification questionnaire. The first is the custodian questionnaire which is used to gather information as to the documentation of relevance to the matter which the custodian has knowledge of. The second is the IT questionnaire, which is used to gain a deeper understanding of the IT systems in place which may contain ESI of relevance to the matter. This additional technical information will unlikely be known by the custodians themselves, and typically require IT management's knowledge of the systems in place. A sample cover letter is included in this Appendix. It is followed by the two questionnaires.

D.1 Sample cover letter

Dear Sirs,

We refer to our recent instructions in respect of the dispute between [X] and [Y]. As you are aware through earlier advices and discussions, as these proceedings progress through the courts you will be required to provide discovery of all documents relating to the issues in dispute.

To prepare for this, it is our practice to request all clients to complete the attached identification questionnaires, the purpose of which is to help us to ensure that all relevant documentation is identified and preserved and to assist us in assessing the volume of documents which [client] holds regarding the issues in dispute.

We will firstly need to identify all potential custodians of documents and speak with them to identify all potential sources of documents which they hold in relation to the [(contemplated) litigation]. We will also need to confirm the date range of relevance to the matter. These consultations should take place as soon as possible and all relevant document sources should be identified so that the retrieval of relevant documents can be commenced. You should contact each of these individuals in order to clarify their individual document retention practices. We have enclosed a custodian document identification questionnaire to assist in this regard. Once the discovery process for these individuals has been completed we can together assess whether the process should be repeated for any other employees.

It may be necessary to obtain information regarding your IT systems. The IT document identification questionnaire enclosed is designed to stimulate discussion between you, your IT department, and ourselves. It does not present an exhaustive list of document sources that you must consider nor do we imply that all of its terms and/or sections apply to you. The purpose of the questionnaire is to identify the entire universe of documents which may have to be considered in the [(contemplated) litigation] in order to gain an understanding of the amount of documentation that might potentially have to be discovered as matters progress. Your IT department may be able to assist in clarifying the rules for storing information on computers at [Client] in order to identify the universe of documents that are potentially relevant to the issues in dispute.

When reviewing and completing the attached identification questionnaires we recommend that you consider who has access to documentation stored not only in your business premises but also those employees and service providers who use their personal devices for business purposes. It is extremely important that you fully understand your obligations to retain all relevant/potentially relevant documentation and we will discuss this issue with you to ensure you fully understand your obligations.

We recommend that once you have reviewed both questionnaires and considered them with your IT department, we arrange to meet to discuss your findings. In the meantime if you have any queries or comments regarding the above or the attached questionnaire please do not hesitate to contact me.

Yours faithfully, [Counsel]

D.2 Sample custodian document identification questionnaire

This questionnaire may be used to assist in the determination of the hardcopy document and ESI sources which a custodian has access to. The custodian may then be asked to provide information as to which of the sources may contain documents relevant to the matter, given the background to the matter and the relevant time periods, etc. Further, for all sections below, it may be necessary to determine what was in place during the time under review, and what has become of those systems and ESI if they are no longer in use.

A) General

1. How long have you been employed with [client], and what roles have you held for which periods? What physical locations have you been located at, and what address are you located at now?
2. Considering the issues in the matter, covering the time period between [X] and [Y], what documents, both hardcopy and electronic, might you have or had, which are relevant to the matter?
3. Where do you keep hardcopy documents which may be of relevance to this matter?
4. Are there any relevant hardcopy documents or ESI that once existed but are no longer held by you? If so please provide full details of the documents together with what became of them.
5. Do you have a policy of archiving your hard or soft copy documents? If so please give full details of same.

B) Custodian-based document sources

1. What desktop and laptop computers do you use?
2. What mobile devices do you use? e.g., Blackberry, iPhone, iPad, Tablet PC, Palm, GPS, and mobile phones.
3. What portable storage devices do you use? e.g. USB keys, floppy disks, CD/DVD's, ZIP disks.
4. What email accounts do you use? Do you have more than one account?
5. Do you have a private folder which only you have access to on a network server? If so, what is its name, and what drive letter do you use to access it?
6. Do you use instant messaging? e.g., SMS text messaging, Sametime, Office Communicator, Microsoft Lync, Yahoo IM, Google Talk, Skype, etc.
7. How do you remotely access your corporate IT systems? Do you use any personal computers/devices at home for your work?
8. Do use any externally hosted networking websites? e.g., Linked-In, Facebook, Twitter.
9. Do you have colleagues or an assistant who would have access to your documents?
10. Are there any other locations where ESI may be stored that you are aware of? e.g. voicemails, video conferencing systems.
11. Do you use any form of encryption and/or password protection on the devices you use and/or on individual documents?
12. Do you have ESI which would be considered personal data under the Data Protection Acts?

C) Non-custodian document sources

1. Do you use shared folders located on a server computer which others have access to? If so, what is its name, and what drive letter do you use to access it?
2. What transactional systems do you have access to? e.g. accounting, Payroll, HR, manufacturing, funds transfer, etc.
3. Do you have access to any internally hosted websites and/or collaboration sites? e.g., Internal file sharing websites, SharePoint, eRoom, etc.
4. Do you have ESI which is hosted on the Internet? e.g., externally hosted websites, file-sharing websites, Google Docs, etc.
5. Are there any other systems which you use to access and/or store ESI? e.g., fax, scanning, etc.

D.3 Sample IT document identification questionnaire

In parallel, or following the receipt of the responses from the individual custodian document questionnaires, this IT questionnaire may be used to further explore the document sources, and gain more detailed information as to the underlying IT systems in place. It may be necessary to determine what was in place during the time under review, and what has become of those systems and ESI if they are no longer in use.

A) General

1. Who manages the IT infrastructure in the organisation? Please provide contact details.
2. Who supports the IT infrastructure in the organisation? Please provide contact details.
3. Are there any 3rd parties that either manage or support the IT environment? Please provide contact details.
4. Are there any 3rd parties who process or host ESI on behalf of the organisation? Please provide contact details.
5. How many IT users are there in the organisation?
6. What are the primary technologies in use? i.e., Windows, Linux, desktops, laptops, etc.
7. What standard applications are in use? e.g., Word processing, spreadsheets, Internet access, etc.
8. What encryption technologies are deployed?
9. Are there any in-house or industry-specific software programmes deployed?
10. Is there a legal/regulatory requirement that requires the organisation to retain ESI?
11. What physical locations/addresses does [client] have employees and IT systems located?

B) Custodian-based document sources

1. What make and model of desktop and laptop computers are deployed? What is the hard disk type and size in these computers?
2. What portable storage devices are deployed?
3. What mobile devices are in use? e.g., Blackberry, iPhone, iPad, Palm, GPS, Tablet PCs, and mobile phones.
4. What email platform(s) are in use? e.g., Exchange, Lotus Notes, Novell GroupWise. Are hosted email services, such as Gmail, Hotmail, Yahoo Mail, etc. permitted for business use?
5. What mailbox size quotas are in place? Are these different for various custodians?
6. What happens when a user meets or exceeds their quota?
7. What automated processes are in place? i.e., deletions, sweeps, etc.
8. Where are the email servers located?
9. How and where is mail archived? Is it automated?
10. Is My Documents redirection in place? Is it standard, or on a user-by-user basis?
11. Do users have their own private network folders? i.e., only the user has access?
12. What instant messaging systems are in place? SMS text messaging, Sametime, Office Communicator, Microsoft Lync, Yahoo IM, Google Talk, Skype, etc.
13. How are remote access services provided to users? What functionality is provided remotely? i.e., full access to LAN, just email, Citrix, etc.

C) Non-custodian document sources

1. What server systems are in place and what business functions do they serve? e.g., file, print, email, application.
2. What operating systems are in use on the servers? e.g., Windows, Linux, Novell.
3. What access controls are in place? e.g., security groups, by folder/share.
4. What shared network folders are available to users? e.g., Finance Share.
5. How do users access these shares? Are there default mapped drive letters in use? e.g., the G drive.
6. How much data is stored in each share?
7. What transactional systems are in use? e.g., accounting, payroll, HR, etc.
8. What back-up systems are in place? What is the make and model of the system in place, and what back-up media does it use?
9. What is the media rotation/retention policy? Where is the media stored? What volume (GB/TB) of ESI is backed-up? How long does it to complete a full backup and a full restore?
10. What restoration capabilities are present on-site?
11. What happens to old backup tapes/systems? Is there a capability to restore old tapes?
12. Are internally hosted websites and/or collaboration sites in use? e.g., Internal file sharing websites, SharePoint, eRoom, etc.
13. Are there any externally hosted websites which hold potentially relevant ESI?
14. Are there any publicly available websites which host potentially relevant ESI? e.g., LinkedIn, Facebook, Twitter, My Space, etc.
15. Are there any systems in place which record voice and/or video recordings, including voicemails?
16. What other systems are in use by the organisation? e.g., Document management, Fax, scanning, etc.
17. Is there a data classification programme in place?
18. Is there a data retention programme in place?
19. Are there data management, disposal, and protection policies in place?
20. Is there a computer use policy in place? Have the data custodians in scope signed/acknowledged the computer use policy and does evidence of this exist?
21. Is there a remote access policy, portable media policy, or any other policies in place?
22. What is done when people leave the organisation? What happens to their ESI, computers, mobile devices, accounts, etc.?
23. Does the organisation store ESI which would be considered personal data under the Data Protection Act?

Consideration should be given to having both the custodian and the person completing the IT questionnaire acknowledge the completeness of the information provided by way of a signature.

Appendix E Managing audio and video data

E.1 Background

Audio and video data pose an additional burden on the discovery process as there are not the same technologies available to manage them through the discovery process as there are for non-multimedia document types.

Audio data refers to the recording of audio sound only. Video data refers to the recording of both audio sound and picture. As such, audio data has one element, the sound, which needs to be addressed, while video data has two elements, the sound and the picture, which need to be addressed.

E.2 Identification

Audio and/or video data will be identified as part of the standard identification phase. The work completed at the identification and collection phases is key to reducing the volume of potentially irrelevant audio and/or video data for processing and review. For example, using the original system's metadata to only extract audio/video data for a specified channel/phone number and between date ranges can significantly reduce the volume of data for processing and review.

The format in which audio/video data is stored will play a key role in determining the effort required to extract and search it. In fact, some formats may prove technically impossible or prohibitively expensive to extract and search. A second factor can be the quality of the recordings, which can impact the success of any automated or manual approaches to managing the data.

Note: Many older audio/video systems do not have built-in functionality to filter data at the extract stage, making it necessary to extract all the data and utilise specialist tools for searching and filtering later.

E.3 Approaches to managing audio and the audio portion of video data

There are four primary approaches to managing audio data during discovery. All have associated risks and benefits, and as in all cases the most appropriate approach should be chosen taking proportionality into account.

E.3.1 Manual review

The review team manually reviews (listens to) the audio data and identifies relevant content. This approach can become quite expensive when it takes and average of 3 hours to review each hour of audio, taking into account multiple review passes. This can be a time consuming and costly process for large volumes of audio data, however may work well for small volumes of audio data.

E.3.2 Manual transcribe and search, then review

The audio data is manually transcribed to a searchable text format. With this approach, teams of transcribers listen to the audio data and type out what they hear.

It can take up to 2 hours of time for each hour of audio as it often has to be re-reviewed due to complex multi-voice recordings. At the end of this process, the transcribed data still has to be searched and reviewed. This can be a time consuming and costly process for large volumes of audio data, however may work well for small volumes of audio data.

E.3.3 Direct computer index and phonetic search

Specialist computer systems are available which allow a computer to search audio data phonetically. This is referred to as 'phonetic searching', whereby the audio data is searched for sounds rather than text, and the results then reviewed for relevance. Keywords are converted to sounds, which the audio data is then searched against. Considerable effort is required to devise and refine the keywords through an iterative test process in order to manage error rates and achieve a proportionate degree of quality. These systems have been utilised extensively in other jurisdictions and have been reported to provide time and cost savings.

E.3.4 Computer transcribe and search, then review

A specialist computer system is used to automatically transcribe the audio data to text. At the end of this process, the transcribed data still has to be searched and reviewed. While this type of technology continues to improve, in isolation it may not yet be fully reliable to be accepted in legal proceedings. However, when coupled with manual quality controls it is emerging as a very useful tool to automate, at least partially, the process of transcribing audio to text.

E.4 Approaches to managing the picture portion of video data

While there are some automated technologies available for filtering and searching audio data, or the audio portion of video data, there are no currently widely used technologies for filtering and searching the picture portion of video data.

Automated technology can be used to identify movement in the picture portion of video data, for example CCTV recordings which record no movement for long periods of time may be filtered using this technology. More advanced automated video filtering technology may be available to automatically recognise patterns such as number plates or human faces, etc. Such technologies should be explored in the event that the video data is required to be searched for specific patterns.

E.5 Other data reduction options

In many cases, audio data will contain long periods of non-voice activity, or silence, and video data will contain long periods of no-activity. There are technologies available which can identify these silent portions of data and suppress them from further searching and/or review.

Appendix F Understanding deduplication, families, and threads

Electronically Stored Information (or 'ESI') has brought with it a number of additional relationships between documents which are not so prevalent in hardcopy documents. These concepts play an important factor in managing complex sets of documents through the discovery process.

F.1 Deduplication

One feature of ESI is the level of duplicate information generated and stored. For example, if Custodian One sends an email to Custodian Two and both their emails are included in the discovery process, then both duplicate emails will be present. Managing and/or reviewing duplicate documents will generally be a waste of time and money, therefore it is most efficient to suppress duplicates as early as possible in the process.

It is possible to calculate a digital fingerprint, or 'hash value', for any electronic document. Like a traditional fingerprint, this value (which is just a relatively short alphanumeric code) can be used to uniquely identify a document, and thus be used to identify two identical documents (where they have the same value).

One of the first steps (usually completed automatically) during the processing phase is to calculate the digital fingerprint or hash value for every document imported into the processing system. Most systems designate the first time it encounters a document as the primary copy and then designates each subsequent copy a duplicate.

An important feature of any system which performs deduplication for eDiscovery is that they will keep a record within the system of all those custodians who held a duplicate of the document which has been deduplicated. This 'duplicate custodian' list is presented with the document for review, allowing the reviewer to see which custodian had the document they are reviewing, as well as the list of other custodians who held an exact copy of it.

There are a number of different methods and algorithms available for calculating the document hash value. The most common algorithms are called MD5 and SHA. What is likely to be more important is what portion of the document is used to calculate the hash value. The most common approach employed by eDiscovery systems is to calculate the hash value based on the contents of the document only. For example, the hash value of a spreadsheet would be calculated from the contents within the document and not take into account the document metadata, which accompanies the document. This would ignore the fact that the duplicate of this document had a different name and was found in a different location. This is generally most efficient from a discovery perspective, as a detailed analysis of the document's metadata can be performed during the analysis phase if required. Another example is emails whereby deduplication systems frequently exclude the Bcc address from the hash value calculation. It will still be possible to know who had a copy of the email from the duplicate custodians list (assuming their documents were collected), however the

version of the email which actually showed them in the Bcc address field may have been suppressed through deduplication.

Once the hash value has been calculated it is then possible to deduplicate individual documents across the entire matter. However, unless the entire document set is comprised of individual documents, and no families, then this simple deduplication may not be appropriate. The section immediately below describes the more common approach of 'family-level deduplication'.

Note: While duplicate hardcopy documents might be scanned into electronic format, it is almost impossible to apply a traditional electronic deduplication to them. This is due to the fact that each scanned copy will have a different hash value. At best, some of the analytics technologies described in Appendix G might be used to identify similar documents.

F.2 Document families

F.2.1 What are document families?

A document family refers to a set of documents which have a relationship. An example of this would be the contents of the fax and its cover sheet. The most common example in eDiscovery is that of emails and attachments. The email is the parent and the attachment is the child. (To add complexity, sometimes the child attachment can be an email, which has its own attachment, referred to as the grandchild.)

By default, in discovery, documents should be considered in the context of their families. One of the key reasons for this is that a document alone may not reflect its true meaning, which may only be apparent when viewed in the context of its family. For example, a customer list attached to an email may have meaning when reviewed in isolation, however when reviewed with the email which states 'not to be disclosed' and on a certain date, the meaning of that customer list might change significantly.

It is important to note that when we refer to families in an electronic document perspective, there can only be one parent.

When filters, such as keywords, are applied to a document set, they will return a number of individual documents which are responsive to them. These individual documents may be single-document families, or may be part of a multiple-document family. If they are part of a multiple-document family, they can either be the parent, or be the child (or one of multiple children). Therefore, when an individual document is responsive to filtering criteria, it is good practice to consider it in the context of its family for review purposes. One approach is to review the individual document and determine its relevance before bringing the rest of its family for review. This is efficient whereby the individual document is deemed to be irrelevant, negating the requirement to review the rest of the family. Only if it is found to be relevant are the other family members brought into the review. An alternative approach is to review the whole family in one go where one or more of the family members are responsive to the filtering criteria. This can be efficient if the individual responsive family member is relevant, but not so efficient if it is not. Further details on approaches to review taking families into account can be seen in Appendix I.

F.2.2 Document families and deduplication

As outlined above, a family of documents can only have one unique parent. They can however have duplicates of children which are also attached to other parents. For example, the same spreadsheet may be attached to two different emails. These two emails represent two unique families, in that as a family unit each is unique, but do contain duplicate children. it is important to consider both families of documents for discovery. As such, both unique parents and two copies of the same attachment would be included for review. It may be that the same document is determined to be relevant in one instance and not relevant or privileged in another, due to its family relationship.

By way of illustration, figure 1 below shows three emails collected from two custodians. The first parent (P1) has two children (C1 and C2). The same parent (P1) is also found in the second custodian's mailbox, along with the same attachments. This is likely due to custodian one sending custodian two the email with attachments. As such, the copy of P1 in the second custodians' mailbox would be deduplicated as it is an exact duplicate for the family in the first custodians' mailbox. The second parent (P2) has two attachments, C1 and C3. C1 is a duplicate of the attachment in the first family; however it is attached to a different parent and has another new attachment C3 in the family as well. As such, the P1 and P2 unique families would be included for review, with C1 being included twice so that it could be evaluated in the context of each of its families.

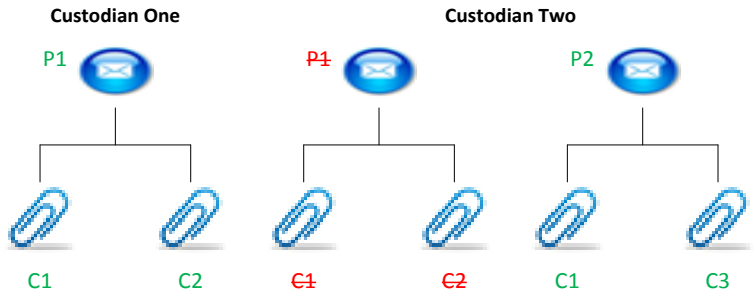

Figure 1 – Two unique families containing duplicate attachments and one duplicate family.

This is referred to as family-level deduplication. The hash value of each combined family is calculated and deduplication is completed at a family level.

F.2.3 Document families and review

When determining the approach to review it is important to decide how document families are going to be treated at production stage as this will determine how they are considered at review.

In the case of single document families, such as emails with no attachments and/or loose files, these are simply considered in isolation and marked as relevant or not relevant to the issues in the matter.

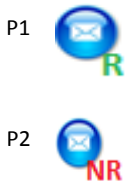

Figure 2 – Single document families marked as either relevant or not relevant.

It is good practice not to produce irrelevant documents during discovery. Doing so is generally a waste of time and money. As such, a balance has to be struck between presenting documents in the context of their families, but not producing irrelevant documents. A good compromise to this is to always produce the parent where the parent or any of its children are relevant; however irrelevant children should not be produced. As producing orphan children (producing just the child document and not its parent) leave it without context, this is not recommended.

Take for example (see figure 3 below) where we have an email (P1) with five sets of customer records attached as five spreadsheets (C1-C5). One of the attachments is responsive to the keywords (it refers to the customer in question), while the other four attachments refer to other customers and contain sensitive personal data. It is recommended in this scenario that the parent (P1) be produced (with references to other customers within the email redacted) along with the single relevant attachment (C1). The other four attachments (C2-C5) should not be produced. They could be redacted in full to protect the sensitive personal data of unrelated parties; however not producing them in the first place would be significantly more efficient.

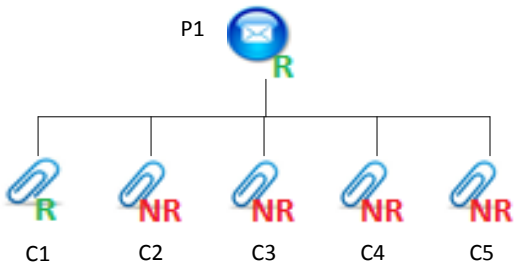

Figure 3 – Parent email and one child attachment are marked relevant and for production, whereas four other child attachments are marked as not relevant and will not be produced.

Where a parent is found to be relevant, but its children are found to be not relevant, then the parent may be produced in isolation.

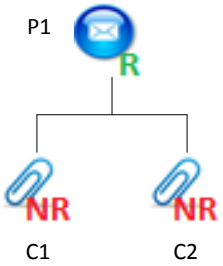

Figure 4 – Relevant parent with non-relevant attachments.

Where a parent is found to be relevant along with some of its attachments (i.e. mixed relevance attachments), then the parent and only the relevant children may be produced. The non-relevant children would not be produced.

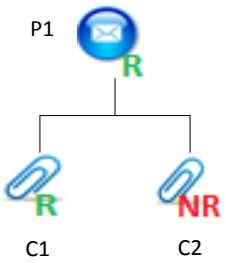

Figure 5 – Relevant parent with mixed-relevance children.

Where a parent is found to be not relevant, but one or more of its children are found to be relevant, then the parent should be marked as relevant and produced as well in order to show the context of the child.

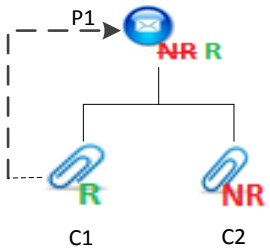

Figure 6 – Non-relevant parent changed to relevant due to association with child.

Note: It may be permissible to produce children in isolation where it is determined that the parent does not lend context to the child. However this is rare and the default position should be that, orphan children should not be produced.

It is important therefore, that when reviewing and marking documents in the context of their families that these principles be applied. For example, in figure 6 above, it is often the case that the parent will be reviewed first and marked as not relevant, then the first child will be marked as relevant and the second child as not relevant. It is then necessary for the reviewer to go back to the parent and change it's marking to relevant. This is a straight-forward step as the reviewer will be considering the document in the context of its family as part of the review.

Production of document families, such as emails and their attachments, may be summarised as follows:

- If a parent is relevant, but its children are not, then only the parent would be produced.
- If a child is relevant, then its parent would also be produced. Orphan children would not be produced in isolation.
- If there are multiple children, where only one is relevant, then only the relevant child and its parent would be produced. Irrelevant attachments would not be produced.

Further, checks should be performed before production to ensure that no irrelevant and/or orphan children are produced. It may be helpful to include a schedule of irrelevant family members which have not been produced, and/or include a slip sheet for each document which has not been produced. This can assist is demonstrating that the document has not been produced intentionally, rather than due to a technical issue or oversight.

F.3 Email threads

F.3.1 What are email threads?

When an email is sent (from Custodian One to Two) it is stored as an individual message (reference P1). Assuming it has an attachment (reference C1); this is sent attached to the message (P1/C1). The person who receives (Custodian Two) the message receives it as P1/C1. Both the original message and attachment are stored as two related documents on the senders (Custodian One's) computer and the same on the recipients (Custodian Two's) computer. This accounts for the basic duplication outlined above.

If Custodian Two replies to the message, this generates a new document (P2) with the reply, which also contains the original message (so P2+(P1)). The attachment is typically dropped and not included in the new message. Custodian One receives the reply from Custodian Two, so now Custodian One has P1/C1 in their sent items and P2 in their inbox, while Custodian Two has P1/C1 in their inbox and P2 in their sent items. In this simple exchange P1 needs to be reviewed because it contains the original attachment. P2 also needs to be reviewed because it contains the reply. It (P2) will contain the original text from P1, however will not contain the attachment.

If the example is extended so that Custodian One replies to P2, creating a new P3 in the process. P3 contains the reply, but also the original text of P2 and P1 (P3+(P2+P1)). Custodian Two receives P3 and replies to it again, which is another new message P4. P4 contains P3+P2+P1 and is received by Custodian One. They then pick up the phone and talk to each other and the emails stop.

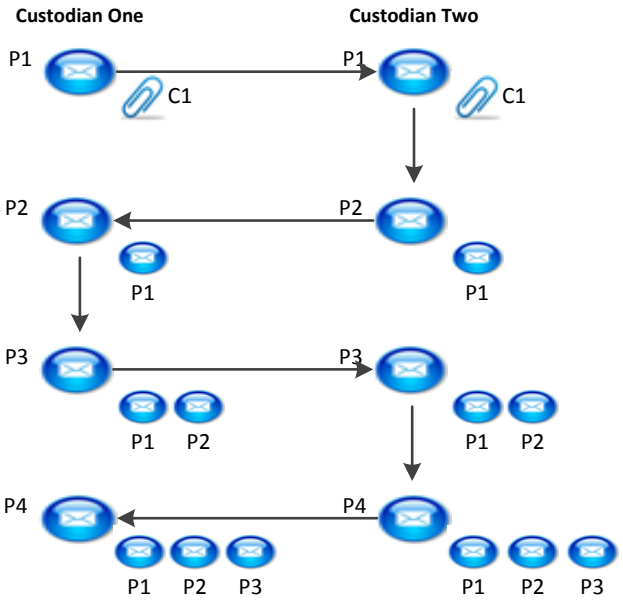

Figure 7 – Email thread between two custodians.

We now have a copy of the individual messages P1/C1, P2, P3, and P4, located as four individual messages with both custodians. If we collect both custodians email for discovery, our first step is to deduplicate both sets of messages. If Custodian One's email was processed first, then Custodian Two's will be suppressed through deduplication and we will be left with Custodian One's copy of P1/C1, P2, P3, and P4. This is simply using standard family-level deduplication. Figure 8 below shows that a 50% saving on review effort can be achieved using this simple family deduplication method.

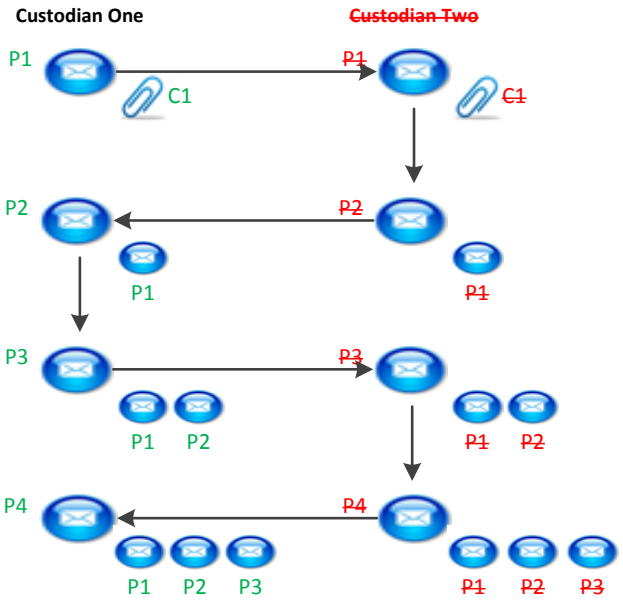

Figure 8 – Standard family-level deduplication on an email thread between two custodians.

The challenge is that we have the original P1/C1 message and attachment; we then have a thread of emails through to P4 which all contain the content of the previous emails, in addition to their new content. (i.e. P2 contains the text of P1 along with its own content/reply (less the attachment). P3 contains the text of P1 and P2 along with its own content/reply. P4 contains the text of P1, P2, and P3, along with its own content.)

Even with standard deduplication, there is a lot of duplicated text within the email thread. This is due to the fact that the email thread is typically not stored as a thread, but the individual messages, which are collected and included in the discovery process.

What further complicates the challenge is that each individual message will have a date associated with it, which will be the date it was sent or received. If the four emails which are outlined in the example above were sent days or weeks apart, then their dates would be days or weeks apart. If these emails were in a document set with 1,000 other emails, and that document set was sorted for review by date, then the emails would appear quite a distance apart, with many unrelated emails in between.

F.3.2 Email threads and sorting/deduplication

There is technology readily available which can identify the individual messages which comprise an email thread and link them together. This is known as email threading. [Such technology would identify the email thread as P1/C1, P2, P3, and P4, and present them for review in that order. Sorting by email thread is very useful as each message is shown linked to the next in the thread, which increases the speed and accuracy of the review. However, it is still necessary to review the duplicate text in P1, P2, and P3, rather than just reading the whole text which is contained in P4.

The solution to this challenge is referred to as 'email thread deduplication'. Such technology extends on threading identification by also designating each message as 'inclusive' or 'non-inclusive'. In the example above, P1/C1 and P4 would be designated as inclusive and would be reviewed, while P2 and P3 would be designated as non-inclusive and could be suppressed as duplicates from review. One might assume that P4 is the only inclusive email as it contains all of the text from P3, P2, and P1, however because the attachment was dropped after P1, it is necessary to include P1/C1 so the attachment is reviewed. The email thread would be sorted for review in the order P1/C1 and then P4, allowing for the more efficient and accurate review.

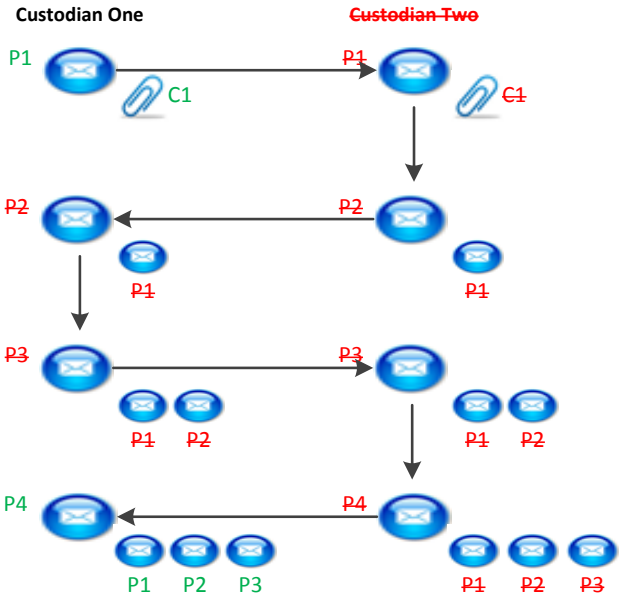

Figure 9 – The remaining email thread deduplicated within the thread.

As can be seen from figure 9 above, the number of documents for review is reduced from 11 to 6, some 45% reduction in review effort (in addition to the 50% already

achieved through standard deduplication). This can increase further with longer email threads. Where email thread deduplication has been utilised, there is generally no requirement to reintroduce duplicate/non-inclusive portions of the email thread at a later date. Note: It is not possible to sort by date and by thread as the two concepts are mutually exclusive.

Appendix G Technology Assisted Review

Technology has been used for some time to assist in the discovery process. In earlier years, document management systems were used only to manage documents which were known to be relevant to the proceedings and were used principally to view and mark such documents. In the past decade, however, it has become more common to employ more sophisticated document management systems in order to process and review documents to determine whether they might be relevant to a case. These systems are now frequently used to manage large volumes of potentially relevant documents through the review process, which is focused on identifying relevant documents. These 'review platforms' and their underlying processing technologies were the first form of Technology Assisted Review (or 'TAR') in that a computer was used to assist in the review process, rather than printing the documents for manual review. Computer Assisted Review (or 'CAR') is another term for TAR.

G.1 What is TAR?

There are many ways in which technology may be used to significantly increase the accuracy and speed of a review, thus reducing risk and cost. These can be split into two broad categories; those which arrange the documents in such a way as to make them easier for practitioners to (referred to as Analytics), and those in which the computer programme is "trained" to identify relevant documents (referred to as Predictive Coding). Both sets of technologies are often referred to under the umbrella term of TAR. It is important to note that different technology platforms may refer to these concepts using different terminology.

G.2 Analytics

Analytics technologies arrange documents in ways which make it easier to carry out a traditional review. They do this by automatically extracting relationships and patterns from documents without human intervention. These technologies are traditionally viewed as being low risk but leading to moderate reward. In eDiscovery, risk refers to the risk of omitting relevant documents, inconsistent review decisions, budget overruns, etc. Rewards refer to the reduction of time and cost, as well as the reduction of risk. It can be very useful to employ analytics techniques during the processing phase in order to identify groups of relevant or irrelevant documents and manage them accordingly. This can include the bulk tagging of irrelevant documents and/or prioritising the review.

G.2.1 Clustering

Documents identified as near duplicates may be grouped together in clusters, allowing, for example, multiple drafts of the same document to be reviewed together. While clustering can involve multiple document types (such as a Word document and a PDF document), it is most effective at identifying near duplicates of the same document type.

G.2.1 Categorisation

Documents which share common technical traits may be grouped in order to allow them to be managed in a specific way. This allows, for example, all large spreadsheets or drawings may be placed in the same category to allow them to be reviewed at the same time and also to allow for the consistent applicant of a redaction methodology.

G.2.1 Themes

Documents which share the same or a similar topic or subject may be grouped together for consideration. Theme-based analytics takes a heterogeneous view of a document, in that it works on the basis that a single document may contain a number of themes. Unlike clustering or categorisation, themes can occur across multiple document types and formats.

G.2.1 Email threading

A detailed description of email threads and their sorting and deduplication is set out in Appendix F. While email thread sorting and deduplication is a form of analytics, they are fast becoming standard practice in most matters.

G.3 Predictive Coding

The analytics technologies outlined above focus on making the human review process more efficient and do not involve the computer making decisions as to whether a document is relevant.

Predictive coding is the process by which computer learning and specialised software are used to allow the computer to be trained by expert reviewers to identify relevant documents. The software employs algorithms to learn on the basis of its interaction with reviewers which documents may be relevant. Similar methods may also be employed to determine if a document is privileged.

The expert reviewer trains the system on a small sub-set of the documents and from this the system learns to accurately predict the likelihood of the remaining documents in the set being relevant. Documents with a likelihood of being relevant above a certain probability can then be manually reviewed and their relevance verified.

Predictive coding has been accepted for use by the courts in various jurisdictions, including the US, UK, and Ireland. It has been shown through numerous studies to take significantly less time and cost than a traditional keyword-driven manual review. It has also been shown to typically provide a far greater level of recall* and precision* than the traditional approach. Predictive coding will not usually lead to the capture of 100% of all relevant documents, however no proportionate review method will achieve this statistical perfection. Predictive coding technologies typically identify more relevant documents than the traditional keyword and manual review approach.

Although there is no specific rule in Order 31 authorising its use, the Irish courts have taken the view that provided the process used is sufficiently transparent, TAR using predictive coding discharges a party's discovery obligations under Order 31, rule 12 RSC. The process must, however, contain appropriate checks and balances which render each stage capable of independent verification. A balance must be struck between the right of the party making discovery to determine the manner in which

discovery is provided and participation by the requesting party in ensuring that the methodology chosen is transparent and reliable. Ordinarily, this is a matter of agreement between the parties at the outset and it may be advisable to agree a protocol.

There are three main components to a predictive coding project; (a.) case expertise provided by subject matter experts (usually the legal team); (b.) a predictive coding engine; and (c.) a method for validation.

It has been widely accepted that it is not necessary for users of predictive coding technology to be experts in the underlying maths and statistics in order to effectively and safely use such systems. It is akin to suggesting that one does not have to be an expert in internal combustion engines to be a proficient driver.

As outlined throughout this guide, the use of predictive coding should be discussed with all parties at the earliest possible stage. From a practical perspective however, predictive coding will not be employed in the review process until all data has been collected, processed into a searchable format, and family-level and email thread deduplication have been completed. At that time, early case assessment will allow an analysis of the data to be completed and a determination as to the suitability of predictive coding for the matter.

The document set will be split into those sub-sets which will be suitable for predictive coding and those which are not (such as handwritten notes which have been scanned).

In a predictive coding project, it is recommended that the reviewer is a senior lawyer involved in the case and is supported by an expert in the use of the predictive coding system. Each step of their involvement should be documented appropriately on the discovery audit file so that it can be later referred to in court if necessary. Any decisions made in training the system will directly contribute to its learning and therefore the support it will provide in assisting review decisions.

Predictive coding comprises of four main steps:

- **ASSESSMENT** – The expert reviewer is presented with a sample of documents (referred to as the 'Initial Assessment Set' or 'Control Set' or 'Seed Set'), typically 500, which may be randomly selected by the system or using an agreed selection process. The expert reviewer marks these documents as relevant or not relevant to the issue(s) and also highlights if a document is to be withheld due to privilege (or other withholding requirement, such as confidentiality). This initial assessment set is used to develop a pool of documents which serve as the 'gold standard' against which the subsequent exercise is measured and tested. This gold standard is an accurate sample of the richness and percentage of relevant/not relevant documents in the overall set.

  Note: It is generally considered best to utilise a randomly selected set of documents from the entire corpus of documents for the assessment step. This typically gives a better statistical spread of documents for the system to learn from. i.e. the system learns itself and not just from documents you already know to be relevant. This may result in the randomly selected set

being of low richness (not enough relevant documents) which requires the set to be augmented with more randomly selected documents. Once a baseline has been achieved with the randomly selected documents, it is usually permissible to add known relevant and not relevant documents to the set to enhance richness. Alternatively, it may be permissible to use known relevant and not relevant documents to assist in achieving the baseline. Keyword searches may be used to identify likely relevant and not relevant documents from the wider document set.

Note: The predictive coding process can be used to identify documents which are relevant to any of the categories/issues in the matter or not. It can also be used on an issue/category basis, however it would then need to be repeated for each issue/category.

- **TRAINING** – With the initial assessment set used to start the training process, the system then provides the expert reviewer with small batches of documents to review and mark. The expert reviewer will mark these documents as relevant or not relevant to the issue(s) and will also highlight if a document is to be withheld due to privilege (or other withholding requirement). This will continue until the system itself determines that it has learnt enough and reaches a stable state. Alternatively, depending on the system in use, the expert reviewer might be presented with statistics at the end of each training set and will then work with the system provider's predictive coding expert to determine when the system has learnt enough to reach a stable state. This can typically be an initial relatively large batch, followed by a number of smaller batches, as is required to reach the desired level of reliability. These are referred to as 'Training Sets'.

  Another approach preferred by some technology providers is for the reviewer to complete a (usually larger) training batch and for the system to then decide the relevance of all remaining documents and provide the reviewer with another batch to 'correct'. This process continues until the overturn rate (or number of corrections made by the reviewer to the decisions made by the computer) reaches an acceptable level or threshold.

  Note: Documents should also be marked as Technical Issue during the Assessment and Training steps in the event that they cannot be opened or reviewed by the expert reviewer due to a technical issue. The producing party should then work with its technology provider to resolve such technical issues and allow the documents to be reviewed and marked as relevant or not relevant before the process continues.

- **DECISION** – The system can then go ahead and provisionally code the remaining documents in the set. It will code the documents with a score of between 0 and 100 indicating its likelihood of being relevant. The emphasis of likelihood here is very important as the system assigns a likelihood score, but does not determine likelihood, that decision is left to the expert reviewers. A high score does not guarantee that a document is relevant,

while a low score does not guarantee that a document is not relevant. Assuming predictive coding is being used in place of the first pass review, this allows the review team to determine the cut-off point based on risk and cost (see below for other use cases leveraging the output of predictive coding). A cut-off point will be decided based on precision* and recall* (and f-measure*) values, whereby documents above the cut-off point will be manually reviewed for relevance and documents below the cut-off point will be subject to statistical sampling, but otherwise not reviewed. For example, all documents (and their families) with a likelihood of being relevant above 85% may be reviewed. Typically the lower the cut-off point the higher the probability that irrelevant documents will be reviewed, with associated costs, whereas the higher the cut-off point, the higher the probability that relevant documents will be missed. Proportionality is a consideration in selecting the cut-off point.

Note: It is usually possible to group documents into bands by score as an effective way to quantify the cut-off point. For example, taking bands of 10% each, one might determine that the majority of the documents fall into the 0-10% and 10-20% bands (i.e. unlikely to be relevant), while a smaller set fall into the 80-90% and 90-100% bands (i.e. likely to be relevant). This may assist in determining which bands of documents to include for manual review and which ones to randomly sample as part of the verification process below.

- **VERIFICATION** – Once the cut-off point has been determined (assuming predictive coding is being used in place of the first pass review), then it is necessary to complete statistical testing of everything below the cut-off point. Such testing or sampling can be used to provide reasonable assurance that the process achieved an acceptable level of quality. This can include:
    - Taking a random sample of a number (typically 500) of the documents below the cut-off point and reviewing them in order to determine if there are any relevant documents which were not identified by the system.
    - Discrepancy analysis can be utilised to identify discrepancies between system and human decisions, with any results being fed back into the system for further training.
    - Theme-based and near-duplicate searching can also be used to complete further discrepancy analysis.

  There will of course be some documents identified as relevant in this set; however they should be few in number and of marginal relevance to the matter. If there are many, and they are important, then the system (or expert advisor) will suggest that it receives further training before repeating the process on all documents again. (Any documents identified as relevant through the verification process, regardless of whether assessment and training are to be repeated, should be added to the set of documents being brought forward for manual review.) For example, assuming a sample size of 500 documents and a desired confidence level of 95%, then the margin of

error will be 1.9% if 5% (25 documents) of the sampled documents are found to be relevant. If 1% (5 documents) of the sampled documents are found to be relevant, then the margin of error will be 0.8%.

- An additional Disclosure step may be added in the event that it has been agreed between the parties that the results of the predictive coding project will be disclosed before proceeding to utilise the results of the project in further review. This disclosure step might include the following:
    - In advance of proceeding to the second-pass, the producing party might inform the requesting party of the results of the predictive coding process including the precision*, recall*, and f-measure* reported by the system, the results of verification, and the proposed cut-off point.
    - The producing party might provide an independently appointed solicitor or counsel with access to all those documents marked during the assessment and training steps (both relevant and not relevant documents), save those marked for withholding. The appointed solicitor or counsel would not disclose any information relating to the documents to the requesting party, however may discuss queries regarding designations of documents as relevant or not relevant with the producing party's legal advisors.
    - A further step which may be taken might be to only provide the requesting party with a list of the documents marked as relevant and not relevant (less those being withheld). This list would include a unique identification number, the name/title of the document, the document type, author/sender, recipient, and created/last modified date/time. This would allow the requesting party to perform an initial review of this information.

  It is not possible to determine and therefore disclose or agree the recall, precision, or f-measure in advance of completing the verification step and understanding the number of documents for manual review, etc. Therefore, this information should not be guessed or estimated until the time where it can be based on the actual results of the process.

  If at the end of the disclosure step the parties cannot agree on the cut-off point and/or the decisions made on any disputed documents during the assessment and training steps, they should meet in person with the relevant experts to discuss. If agreement cannot be reached, then either party might apply to the court for directions.

The output of the above four steps are a set of documents which have been assigned a probability of relevance to a particular topic (usually relevance, but can also be privilege or category, amongst others). The most well-known use for predictive coding is to determine the cut-off point based on risk and cost, and review all documents (and their families) with a likelihood of being relevant above the cut-off point. This typical application is generally known as using predictive coding to replace the first-pass review for relevancy, but still complete a second-pass manual review to bring in family members, confirm relevancy, and assign categories and privilege (and

complete redactions, if required). The difference is that this second-pass review is likely to be covering all relevant documents, with few, if any, false positives.

There are however a number of other applications once a probability of relevance has been applied to a set of documents. These include:

- **Prioritising documents for review** – All documents may still be reviewed, however can be prioritised by those most likely to be of relevance to the matter. This can allow the most relevant documents to be reviewed first by a core review team and documents likely to be irrelevant to a secondary or outsourced review team.
- **Remove irrelevant documents** – As part of Early Case Assessment (or 'ECA'), predictive coding can be used to identify and separate documents which are clearly irrelevant to the matter.
- **Verification of keywords** – Predictive coding may be used to perform a discrepancy analysis between those documents identified as potentially relevant through keyword and other filters and those which the predictive coding system identifies as potentially relevant. This can be a useful tool at the processing phase when keywords and other filters are being devised.
- **Quality control during review** – As the review progresses and/or upon its completion, predictive coding may be used to analyse discrepancies between the human review decisions and the computer.

As with all technology, predictive coding is better suited to some matters and not others. It is best suited to matters where:

- There is a large volume of text-based data (typically more than 50,000 documents for review, however TAR can work well on document sets of 20,000).
- The parties have engaged and are following a reasonable approach, such as outlined in this guide.

Predictive coding is not well suited to matters where:

- There is a large amount of non-text-based data, such as pictures or spreadsheets. (Some predictive coding systems work well with numeric data, whereas others do not.)
- Handwritten documents which have been scanned to electronic format or typed documents where handwritten notes in the margins would be a factor in determining relevance.
- There is a large volume of hard copy documents.
- The parties have not engaged in the discovery process (which may result is extensive legal costs in disputes regarding the approach taken during discovery utilising advanced technologies).

In all matters where predictive coding is considered, expert advice from the provider of the predictive coding system should be sought in order to determine if the use of predictive coding is helpful given the circumstances of the matter and the data types involved. As outlined above, it is usually not possible to understand the data fully until it has been collected and processed into a searchable format, deduplicated at a family

level and by email thread. As such, it may not be possible to determine whether predictive coding will be helpful in advance of that stage in the process.

There are significant advantages to utilising predictive coding in place of the traditional keyword search and review approach. The foremost of these is the increased accuracy when compared to keywords, and the fact that the level of accuracy may be chosen based on a well-informed proportionality decision. The second advantage is the significantly reduced time and cost of manual review. There is a cost in having a senior expert complete the training of the system, and all documents which the system finds to be potentially relevant still have to be reviewed, however this is typically a lot less than large teams reviewing documents identified by keywords which turn out to be irrelevant.

It is important to note that as with analytics technologies, different predictive coding technology vendors use different terminology when describing their processes and some of them employ slightly different processes.

G.4 Additional cooperation

As standard analytics technologies are commonly accepted and used, it would not be necessary to notify the requesting party as to their use (although good practice would see their use included in the discovery plan). However, as predictive coding technologies are relatively new in this jurisdiction, it is recommended that early engagement and agreement on its use should be pursued with all parties involved.

Common areas of discussion between parties seeking to utilise predictive coding include:

- Whether filters (such as keywords and date ranges) or analytics will be used to reduce the data set prior to predictive coding being employed.
- Whether the decisions used to train the predictive coding system are shared (also referred to as the 'control/assessment/seed set' or 'training set').
- What the cut-off point (or 'threshold') will be for manual review.
- What verification and/or quality controls will be carried out.

The time required for the review will depend on the richness of the data, that is the preponderance of likely relevant data within the dataset, and it will likely be necessary to defer providing an estimate for the review and production until after the completion of the predictive coding exercise.

Key to the success of a predictive coding project is having a system which is transparent, allowing the user (and possible the requesting party and the court) to see and understand why the system made a decision on any particular document. Equally important is having a statistically valid process and validating the results. It is important to understand that the predictive coding process, just like the traditional keyword search and review approach, is not perfect. As such, it should be measured against this alternative and not against unattainable perfection.

The discovery plan at Appendix H contains sample text which may be used to provide information to the various parties involved in a matter regarding analytics and/or predictive coding.

Appendix H Sample discovery plan

[Matter reference]

Discovery Plan

[Version 0.1]

[Date]

H.1 Background

This discovery plan (the 'Plan) sets out the steps that [Producing party] has taken to date and will take in future for the purposes of making inter party discovery in [Matter reference].[2]

[This Plan is in draft 0.1 format, and is the initial draft proposed by [producing party]. It is for the information of [requesting party]. Any amendments requested by [requesting party] should be promptly communicated to [producing party].]

or

[This Plan is in draft 0.2 format, and contains amendments to draft 0.1 as discussed between the parties. Any further amendments requested by [requesting party] should be promptly communicated to [producing party].]

or

[This Plan is in final 1.0 format, and sets out the steps that [producing party] has taken to date and will take in future for the purposes of making inter party discovery in [Matter reference].

This Plan has been produced following a process of dialogue between [producing party] and [requesting party]. Where necessary, [producing party] has also consulted relevant technology and eDiscovery experts and service providers.

Should [producing party] need to derogate from the Plan for any reason they will notify [requesting party] and seek to reach agreement in relation to any such necessary derogation.

Should it not be possible to reach agreement on the proposed Plan or any derogations from it, the parties agree not to proceed without leave of the Court and to seek liberty to apply.

[[Producing party] has engaged the services of [service provider] to assist in this matter. [Service provider] intends to utilise [technology platform(s)] in carrying out the services for the purposes of [producing party] making discovery.]

Notwithstanding this plan it remains the obligation of [producing party] and its solicitors to identify relevant documents and make discovery of such documents in accordance with Order 31 Rule 12 of the Rules of the Superior Courts, 1986 (as amended).

---

[2] *Note: Version 0.1 is typically the first draft provided for review by the receiving party. Version 0.2 and subsequent drafts (0.3, 0.4, etc.) may be used to reflect changes as the drafts are agreed between the parties. Version 1.0 will be the final version agreed between the parties.*

H.2 Scope

This plan governs all document sources, both electronic and hardcopy, which will be included in the discovery process and provides for how documents will be managed throughout the process, from initial identification through to final presentation in Court.

H.3 Approach and progress to date[3]

A phased approach has been taken to this process, consisting of the following eight phases:

- Phase One – Identification – the identification of document sources which may contain documents of relevance to the matter.

- Phase Two – Preservation – notifying the custodians and other parties of their duty to preserve documents and taking steps to help ensure that documents may not be lost in advance of collection.

- Phase Three – Collection – working with the custodians, IT team and other parties to obtain a copy of document sources identified. [Phases One, Two, and Three have been completed. Details of the steps carried out in the first three phases can be seen below.]

- Phase Four – Processing – converting the document sources collected into a format to facilitate their efficient searching and review. Documents were also filtered, using filters such as date range and keywords. This phase is expected to be completed by [date], subject to our discussions with you.

- Phase Five – Review – documents responsive to the filtering criteria and any documents that do not require filtering to be brought forward for review by [using Predictive Coding to narrow the set of documents to those likely relevant and then] manual review to determine the relevance and privilege status of each responsive document.

- Phase Six – Analysis – quality checking and technical analysis as required, for example to determine the provenance of a document.

- Phase Seven – Production – at the conclusion of the review and after quality checks have been completed, generation of discovery schedules and export of documents for disclosure.

- Phase Eight – Presentation – preparation of documents for presentation in court in a manner which facilitates their efficient presentation and the running of the matter.

---

[3] *This sample plan assumes that the producing party is close to the end of Phase Four. i.e. data is collated and filters/searches run, but review not yet commenced. Therefore it provides what has been done to date up to Phase Four and what is proposed to be done in the remaining phases. This may need to be amended. For example, if an early meet and confer has been agreed the producing party may only have completed Phases One and Two, and will be setting out the proposed steps for the remainder of the project.*

H.4 Phase One – Identification

The objective of the identification phase is to identify sources of potentially relevant documents.

The issues in this matter relate to [insert summary]. These issues relate to activities undertaken between [date] and [date]. We identified and isolated documents created/sent, modified, or last accessed between these dates. Undated documents, or those which could not accurately be dated, were included.

The first step in this phase was to identify a list of custodians who may hold or have held relevant documents. Custodians included individuals and external organisations who may hold documents on [producing party]'s behalf. The list of custodians collated is at Attachment One below.

We worked with the custodians to identify potential document sources, which included:

- Understanding the likely document types and date ranges through discussions and interviews, using the CLAI custodian questionnaire.
- Interviewing custodians to understand how they typically utilised technology and managed documents.
- Interviewing the IT team [and outsourced IT providers] to understand how data is managed from a technical perspective, using the CLAI IT questionnaire.

We identified the following custodian document sources:

- [Live email from each custodian's email account.]
- [Archived email stored on each custodian's laptop or desktop computer.]
- [Archived email stored on servers.]
- [Loose files from each custodian's private network folder.]
- [Loose files from each custodian's laptop or desktop computer.]
- [Email from backup/archive systems.]
- [Loose files from backup/archive systems.]
- [Structured records from accounting/HR systems.]
- [Online data from social/professional networking websites.]
- [Hardcopy documents from personal filing systems.]
- [Text messages from custodians' personal devices]

We also identified the following non-custodian data sources:

- [Loose files from shared network/project folders.]
- [Structured records from a manufacturing/quality control system.]
- [Hardcopy documents from centralised filing systems.]

We prepared this discovery plan in addition to identifying and addressing data privacy and security concerns with [producing party]. A plan to preserve and acquire copies of the document sources identified was then formulated.

H.5 Phase Two – Preservation

The objective of the preservation phase is to take steps to preserve documents where they exist, so that they may not be altered or destroyed in advance of collection. Preservation took two primary forms; a legal hold notice and technical measures.

[Producing party] issued a legal hold notice to all potential custodians within [organisation/department name] and [outside service providers who may hold data] on [date]. This notice instructed custodians to retain and not alter or destroy any potentially relevant documents, including electronic data. This notice was also sent to [producing party]'s IT team so that routine destruction processes could be suspended. [Producing party] complied fully with the CLAI guidance and obtained explicit acknowledgement from all custodians. A reminder notice was issued [weekly/monthly/quarterly] until all document sources had been collected.

In addition to the legal hold process described above, [Producing party]'s IT team implemented the following technical preservation steps on [date]:

- [The most recent backups of [email server/file server/application server] were removed from the backup rotation and stored securely.]
- [Technical controls were implemented which prevented custodians from altering or deleting historical email.]
- [Technical controls were implemented which prevented custodians from altering or deleting historical files.]
- [Access to hardcopy documents were restricted to [litigation team].

All document sources identified during Phase One above were included in the preservation process.[5]

H.6 Phase Three – Collection

The objective of the collection phase is to copy potentially relevant documents from the sources identified so that they can be processed and searched for relevant documents.

[Producing party] obtained a [forensic] copy of the document sources identified during Phase One above. This was completed for all document sources which existed at the time of the collection exercise, details of which can be seen in Attachment One below.

The copying was completed using different approaches, depending on the document source involved:

- [Hardcopy documents were scanned into electronic format and made searchable through an OCR process. Metadata associated with the documents was compiled into an electronic format as well through a manual coding process. Note: Unlike electronic document sources, where the full document source was acquired, a focused collection of only potentially relevant documents was completed for hardcopy documents.]
- [Electronic document sources were copied using industry standard tools, such as [name of tools], which help ensure that the original documents and metadata was preserved throughout the copying process. Documents were acquired directly to encrypted media so that the documents remained protected while in transit and in storage.]

[5] *If all sources were not preserved, then an explanation as to what was excluded and why should be included here.*

- [A chain of custody has been maintained for all document sources acquired. A primary copy of the document sources acquired has been stored at [location].]

All document sources identified during Phase One above were included in the collection process.[6]

H.7 Phase Four – Processing

The objective of the processing phase is to remove obviously irrelevant document types and apply date range parameters, and to convert the remaining documents into a format which will facilitate their efficient searching and review. Documents may then be filtered, using filters such as date range, email addresses, keywords, and other analytics, before being brought forward for review.[7] The eight steps outlined below were undertaken during the processing phase.

H.7.1 Remove irrelevant document types

In the case of each custodian's laptop or desktop computer, where either existed, a full forensic copy of the computer was acquired. This contained large volumes of software code and other irrelevant document types. Only user-created documents were extracted from each computer. The user-created document types can be seen in Attachment Two below.

Where a full forensic copy of the computer was not acquired, a document type filter was not applied, as by their very nature they were a focused collection of user-created documents and will likely only contain user-created documents.

H.7.2 Convert into searchable format/load into database

The document set collected, less irrelevant document types removed, was loaded into an eDiscovery processing system, [name of system]. It was found to comprise [number] documents (emails, their attachments and other loose files). This document set represents all of the documents acquired for each custodian, as well as the non-custodian document sources and the hardcopy document sources.

H.7.3 Deduplicate

A family-level deduplication process was run against all documents, suppressing any duplicate families of documents while leaving one copy of each unique family of documents for further processing. The list of custodians who held a duplicate family which was suppressed has been recorded and included in the remainder of the process. This allows only one copy of the family to be considered, while also allowing reviewers to quickly understand who held duplicates of each family. The resulting document set consisted of [number] documents from unique families of documents.

H.7.4 OCR

[number] non-searchable documents were identified, including [document types such as PDF and TIFF]. An OCR process was run against these documents in order to convert them to a searchable format.

---

[6] *If all sources were not collected, then an explanation as to what was excluded and why should be included here.*

[7] *Such review may or may not include the use of Technology Assisted Review.*

H.7.5 Thread deduplication

Email thread deduplication[8] was run against all emails and their attachments to identify the inclusive portions of each email thread, along with the non-inclusive (or duplicative) portions of the email thread. The non-inclusive portions of the email threads were then suppressed from further processing. Email threads which were unable to be subjected to the email threading process have not been suppressed and have been included in further processing.

H.7.6 Manage problem documents

[number] [password protected/encrypted] documents have been identified in the remaining document set, the contents of which will not be accessible for keyword searching.[9] We will attempt to access these documents through decryption techniques, where their name, location, and/or associated document (such as parent email) are responsive to filtering criteria.

We will attempt to access any documents that are not subject to filtering [such as those in shared electronic project folders].

[Describe any other class of problem documents and how they will be managed.[3]]

H.7.7 Apply filters and perform ECA

Hardcopy documents [and other document sources, such as shared electronic project folders] have not been subject to filtering. Shared project folders are document repositories for this discovery exercise are likely to contain relevant documents.[4]

The resulting document set, numbering [number] documents, represents all documents associated with each custodian and as such, the vast majority will have no relevance to this matter. It would not be proportionate or practical to manually review each document for relevance.

[The Producing Party] has therefore applied the filters outlined in Attachment Three to the document set in order to identify potentially relevant documents, resulting in [number] of responsive documents for manual review.

[Producing party] has worked with their legal advisors to test the filtering criteria for precision and recall and to highlight for review documents likely to be of relevance to the matter, while seeking to reduce the volume of irrelevant documents requiring manual review. [This testing involved initial early case assessment (or 'ECA') using analytics tools such as clustering, categorisation, and themes, in addition to sampling the results of each filter.]

---

[8] *See Appendix F of the CLAI Good Practice Discovery Guide v2.0 for a detailed explanation of email thread deduplication.*

[9] *their location and names, including metadata would however likely be searchable.*

[3] *Documents that have become corrupted and are not accessible but are known to be relevant should be listed in the Second Schedule to the Affidavit as to Documents.*

[4] *If the parties have agreed an 'end date' for discovery to facilitate collection of data the shared project folders may fall outside the discovery to be made, and this may not apply.*

H.7.8 Publish for Review

[Responsive documents and those not susceptible to filtering will be published for manual review. Their families will also be uploaded (i.e. where an attachment is responsive, its parent email will also be uploaded). Where duplicates of responsive documents exist within another unique family of documents, the other unique family of documents will be published for review (i.e. where the same attachment is attached to two different emails, both emails and two copies of the attachment will be published). This allows decisions regarding how duplicates and families of documents are managed to be made throughout the review phase.]

Or
[It is proposed that predictive coding, without prior filtering of documents, be used at the initial review phase. all unique families of documents will be published for predictive coding.] [5]

H.8 Phase Five – Review

The objective of the review phase is to identify all relevant documents using manual review and mark documents as privileged, relevant to specific categories and requiring redaction, where appropriate. This will be completed on a document by document basis by [review team].

Or
[[If predictive coding is in use] The objective of the review phase is to utilise predictive coding to highlight documents of potential relevance and then perform a manual review of those documents. This will be completed by [review team]. Each document will have a system generated determination made as to its likely relevance to the issues in the matter and those identified as potentially relevant will be manually reviewed for privilege, relevance to categories and redactions, where appropriate.]

[Choose Model 1, 2, or 3 below and delete other content. See Appendix I of CLAI Good Practice Discovery Guide for guidance on different approaches to review.]

H.8.1 Two-pass review without predictive coding [Model 1]

As many responsive documents form part of wider families of documents, a two-pass review will be required consisting of:

- **First pass –** Assessing whether a document is relevant and suppressing clearly irrelevant documents from further review. This review pass will consider documents in isolation and only one unique copy of each document responsive to the filtering criteria will be reviewed.

- **Second pass –** Documents identified as relevant through the first pass review, their families (or related documents) and other unique families which contain duplicates, will be included in this review pass and considered for privilege and categorisation. Documents requiring redaction will be identified at this stage. [Emails and their attachments will be grouped by deduplicated thread, and loose files (not

[5] *Different predictive coding system vendors recommend different approaches as to whether documents are filtered in advance of running the predictive coding process. i.e. some insist that all deduplicated documents/email threads are included in order to achieve the best statistical results, while others are happy for the document set to be pre-filtered before predictive coding.*

part of email families) will be grouped by near duplicate, allowing for a more accurate and efficient review.]

To ensure consistency, daily review meetings will take place at which reviewers will flag and discuss query documents or those considered borderline for relevance or privilege with a supervising solicitor.

- **Redaction pass –** Documents requiring redactions will have redactions applied during this review pass.

At each pass of the review, at least [percentage]% of review decisions by each reviewer will undergo quality checks by [quality check team].

H.8.1 Single-pass review without predictive coding [Model 2]

A single-pass review will be conducted consisting of:

- **First pass –** consideration of relevance and privilege status of all responsive documents, their families (or related documents) and other unique families which contain duplicates. Documents requiring redaction will also be identified at this stage. [Emails and their attachments will be grouped by deduplicated thread, and loose files (not part of email families) will be grouped by near duplicate, allowing for a more accurate and efficient review.]

  To ensure consistency, daily review meetings will take place at which reviewers will flag and discuss query documents or those considered borderline for relevance or privilege with a supervising solicitor.

- **Redaction pass –** Documents requiring redactions will have redactions applied during this review pass.

At each pass of the review, at least [percentage]% of review decisions by each reviewer will undergo quality checks by [quality check team].

H.8.1 Two-pass review with predictive coding [Model 3]

A predictive coding review will be undertaken consisting of:

- **First pass –** [[If pre-filtering has been used] All responsive documents, their families (or related documents) and other unique families which contain duplicates, will be included in this review pass.]
  Or
  [[If pre-filtering has not been used] All families of documents which remain after family-level and email thread deduplication will be included in this review pass.] The predictive coding review process comprises four main steps:

  - **Assessment –** The expert reviewer will be presented with a sample of [number] documents (referred to as the 'Control Set'), which will be randomly selected by the system [or a seed set

selected using an agreed selection process]. The expert reviewer will mark these documents as relevant or not relevant to the discovery categories as a whole[6] and will highlight if a document is to be withheld due to privilege or other withholding requirement[7]. This initial assessment set is used to develop a pool of documents which serve as the 'gold standard' against which the subsequent exercise will be measured and tested. In the event that the richness of the sample set is low, then additional documents will be added until the richness is at an acceptable level to proceed.

- **Training –** the system will provide the expert reviewer on an iterative basis with small batches of documents to review and code as relevant or not relevant and highlight if a document is to be withheld due to privilege or other withholding requirement. This will continue until the system determines that it has learnt enough and reaches a stable state. [The expert reviewer will be presented with statistics at the end of each training set and will work with the system provider's predictive coding expert to determine when the system has learnt enough to reach a stable state.] This may be between [number] and [number] additional batches of [number] documents each and are referred to as 'Training Sets'.[8]

  [**Note:** Documents may also be marked as Technical Issue during the Assessment and Training steps in the event that they cannot be opened or reviewed by the expert reviewer due to a technical issue. [Producing party] will work with its technology provider to resolve such technical issues and allow the documents to be reviewed and marked as relevant or not relevant.]

- **Decision –** The system will provisionally code the remaining documents in the set with a score of between 0 and 100 indicating likelihood of being relevant. A cut-off point or threshold will be decided in conjunction with the system provider's predictive coding expert based on the risk of missing key documents and the cost of reviewing large volumes of irrelevant documents, taking into account the precision and recall [and f-measure] reported by the system and having regard to proportionality.
- **Verification –** Once the cut-off point has been determined, statistical testing of everything below the cut-off point will be carried out, including:
  - Taking a random sample of [number] documents below the cut-off point and reviewing them in order to determine if there are any relevant documents which were not identified by the system.
  - Discrepancy analysis to identify discrepancies between system and human decisions, with results being fed back into the system for further training.

---

[6] *Predictive coding in respect of each specific category requires a series of predictive coding processes and is likely to be extremely time consuming and involve very significant resources. It also may risk missing borderline documents. For this reason it is recommended that during the training of the system reviewers code documents for relevance to the discovery categories as a whole.*

[7] *It is essential that documents are marked for privilege and/or commercial sensitivity at this stage to avoid duplication. Note that predictive coding filters only for relevance, not for privilege or commercial sensitivity. Documents will only be withheld on the basis of commercial sensitivity if they are not relevant to the discovery categories.*

[8] *Documents may also be marked as Technical Issue during the Assessment and Training steps in the event that they cannot be opened or reviewed by the expert reviewer due to a technical issue. [Producing party] will work with its technology provider to resolve such technical issues and allow the documents to be reviewed and marked as relevant or not relevant.*

- Theme-based and near-duplicate searching to complete further discrepancy analysis. This may include searching for document types known to be relevant, or searching against specific custodians' documents.

Should the verification process identify more than [percentage]% documents which were incorrectly coded by the system, the assessment and training phase will be repeated. Assuming an overall sample size of [500] documents, and a desired confidence level of [95]%, then the margin of error will be [1.9]% if [5]% of the sampled documents are found to be relevant. If [1]% of the sampled documents are found to be relevant, then the margin of error will be [0.8]%.

- **Disclosure –** In advance of proceeding to the second-pass, [Producing party] will inform [requesting party] of the results of the predictive coding process including the precision and recall [and f-measure] reported by the system, the results of verification and the proposed cut-off point.

  [Producing party] will provide access to [requesting party]'s nominated counsel to a schedule of all relevant and non-relevant documents marked during the assessment and training steps, save those marked for withholding.

  Or

  [[Producing party] will provide an independently appointed [solicitor/counsel] with access to the relevant and not relevant documents marked during the assessment and training steps, save those marked for withholding.] The appointed [solicitor/counsel] will keep the documents strictly confidential and not disclose any information relating to the documents to [requesting party], but may raise queries regarding designations of documents as relevant or not relevant with [producing party]'s legal advisors and formally indicate to [requesting party] whether he or she is satisfied with the designation of documents as relevant or not relevant in training the system.]

  If the parties cannot agree on the cut-off point and/or the decisions made on any disputed documents during the assessment and training steps, they should meet in person with the relevant experts to discuss. If agreement cannot be reached, then either party may apply to the court for directions.

- **Second pass –** It is not possible to determine in advance what the cut-off point might be. Once the cut-off point has been decided, all documents with a relevance probability above the cut-off point, their families (or related documents) and other unique families which contain duplicates, will be manually reviewed. Documents will be considered in the context of their families and will also be considered for privilege and categorisation. Documents requiring redaction will be identified at this stage. [Emails and their attachments will be grouped by deduplicated thread, and loose files (not forming part of email families) will be grouped by near duplicate, thus allowing for a more accurate and efficient review.]

  To ensure consistency, daily review meetings will take place at which reviewers will flag and discuss query documents or those considered borderline for relevance or privilege with a supervising solicitor.

- **Redaction pass –** Documents requiring redactions will have redactions applied during this review pass.

At each pass of the review, at least [percentage]% of review decisions of each reviewer will undergo quality checks by [quality check team].

H.9 Phase Six – Analysis

The objective of the analysis phase is to take a deeper look at specific documents, for example, to determine their provenance. If necessary, [producing party] may perform a detailed analysis of a document or groups of documents.

H.10 Phase Seven – Production

At the completion of the review, relevant documents will be produced by generating an electronic schedule of the documents in the form of [a spreadsheet/electronic load file/other]. The information to be included in the schedule can be seen in Attachment Four below. Documents identified as fully privileged will not be produced but will be scheduled separately in accordance with the Rules of the Superior Courts.

Documents will be produced in their native format by default. Exceptions to this include:

- [Hardcopy documents which have been scanned will be produced in PDF format.]
- Corrupt, password protected or encrypted documents may be converted to a different format (such as PDF) which enables their use.
- Redacted documents will be produced in [format, such as redacted PDF or TIFF] and will also be identified in the schedule as having being redacted.
- On rare occasions modified versions of native documents may be produced where it is not practical, possible, or proportionate to image them before redaction. Examples might include very large spreadsheets. Any such documents will be separately identified in the schedule.
- [Documents which require redaction but cannot reasonably be redacted such as very large spreadsheets or databases may be provided for inspection only.]

Document families, such as emails and their attachments, will be produced as follows:

- If a parent email is relevant but its children are not, then only the parent will be produced.
- If a child attachment is relevant, then its parent email will also be produced for context. Orphan child attachments will not be produced in isolation.
- If there are multiple child attachments, where only one is relevant, then only the relevant attachment and its parent email will be produced. Irrelevant attachments will not be produced.

[A schedule of irrelevant family members which have not been produced, and/or a slip sheet for each document which has not been produced, will also be provided.]

[Producing Party]'s production will comprise of the electronic files and the electronic schedule. (It is not intended to produce extracted text and/or the results of any OCR process which has been run.) They will be transferred on a portable storage device, such as a USB key, or transferred through a secure online transfer system. Both the files and the schedule will be encrypted and the decryption password or key will be provided to [requesting party] separately.

H.11 Phase Eight – Presentation

Once a venue for the hearing of the matter has been finalised, [producing party] proposes that the following electronic system be utilised to facilitate the efficient management of documents throughout the hearing and to allow the documents to be shared with all parties during the hearing.

[Include detailed information as to what system is proposed, any external providers required, and any costs involved. Also include which documents will be presented in hardcopy (such as the core books), and those that will be presented electronically (such as everything else).]

H.12 Attachment One – Custodians and document sources

Custodian-based document sources

No.	Custodian name	Live email	Archived email	Laptop loose files	Private network folder
1	[Custodian One]	[Yes/No]	[Yes/No]	[Yes/No]	[Yes/No]
2	[Custodian Two]	[Yes/No]	[Yes/No]	[Yes/No]	[Yes/No]
3	[Custodian Three]	[Yes/No]	[Yes/No]	[Yes/No]	[Yes/No]

Non-custodian-based document sources

No.	Source name	Description
4	[Source name one]	[Loose files from shared network/project folder.]
5	[Source name two]	[Hardcopy documents from centralised filing system.]

H.13 Attachment Two – Document type filters

The table immediately below details the document types which have been included:

No.	File type	File extension
1	[Microsoft Office]	[doc, docx, xls, xlsx, ppt, pptx, etc.]
2	[Portable Document Format]	[pdf]

H.14 Attachment Three – Other filters

The following filters have been applied to the document set.

H.14.1 Date range

Documents with a date created, modified, or last accessed, between [date/time] and [date/time] will be included.

H.14.2 Keywords

No.	Search term	Deduplicated hits
1	[Search term one]	[number]
2	[Search term two]	[number]
3	[Search term three]	[number]
n/a	All above combined*	[number]
n/a	Combined including families**	[number]

* This is the combined number of documents which are responsive to one or more of the search terms. i.e. if a document is responsive to two or more of the search terms, it is only necessary to review it once.

** This is the combined number above, plus their families, plus any other unique families which also contain the responsive document.

H.14.3 Other filters

[Any additional filters applied should be detailed here.]

H.15 Attachment Four – Production format

The following metadata fields will be included in the production schedule:

- Electronic link to the document
- Unique document identifier
- Unique family/parent identifier
- Whether the document is a parent or child in the context of any family relationship
- The date and time the document was last modified, or email sent in the format DD/MM/YYYY HH:MM:SS
- The document type
- The document author, or email sender
- The document recipient, or email Cc and Bcc
- The document name, or email subject
- The category the document falls into (where categories have been assigned)
- If the document was redacted and why

[The information provided in the schedule will be sorted by descending parent date/time. i.e. parents will be listed with their children next, and then the next family.]

Note: For electronic documents, the metadata will be derived automatically from the metadata contained within the electronic document. No manual process will be employed to verify or correct such metadata. For hardcopy documents which have been scanned, the metadata fields will be manually populated through the coding process.

Note: Each document will have a unique document identifier and a unique family identifier. These will be included in the relevant schedules and the underlying documents will be named by their unique identifier (the contents of the documents themselves will not be altered to include the identifier). Documents produced will not employ any form of Bates-stamping; rather the unique document identifiers will be used to uniquely identify each document in the production set. These may be used as follows:

- If the document identifier equals the family identifier, then the document is the parent.
- If the document identifier does not equal the family identifier, then the document is a child.
- The child's parent and other siblings can be located by searching for all documents with the same family identifier.

Appendix I Sample review plans

The review plans below outline the detailed sample approaches to the review phase of the discovery process.

Which approach to take – There are three alternative approaches included in this guide (Models 1, 2 and 3). These are based on common approaches currently followed. They are not definitive and there may be further alternative approaches available and/or ones which develop in future. With the three alternative approaches outlined here, there are a number of factors to consider when choosing the best approach. The first consideration is whether predictive coding is suitable for the dataset, in which case Model 3 will be the most appropriate starting point. If keyword filters are to be used without predictive coding, then Model 1 or Model 2 may be appropriate.

The main consideration between Models 1 and 2 is the density of the families of documents in the document set.

- If the density is high (families comprising of lots of family members), then a Model 1 review might be the best option. For example, if the number of unique documents responsive to the filtering criteria is 10,000, but this increases to 40,000 when families (and other unique families containing duplicates) are brought in, then the family density is said to average 4-1. Assuming a two-pass review whereby the 10,000 documents are reviewed in isolation in the first-pass, and maybe half (5,000) of them are found to actually be relevant. The families of these 5,000 are brought in for the second-pass, making 20,000 for second-pass. The total across the two passes is now 30,000 having being reviewed. This is typically split, with the first-pass being completed by junior reviewers and the second-pass being completed by senior (and more costly) reviewers. The alternative would be to have all 40,000 documents reviewed in a single-pass review by senior reviewers. In this example scenario, this approach may take more time and cost, while having senior reviewers review large volumes of irrelevant documents.
- If the number of documents increases from 10,000 to 14,000 when families (and other unique families containing duplicates) are brought in (i.e. low density – either lots of loose files/one member families, or all children in more dense families are responsive) then a single-pass review may be more efficient.

Only after filters have been applied and tested can the relevant statistics be determined at the end of the processing phase. It is for this reason that the approach to review should not be decided until such information has been gathered and reviewed in order to determine the most efficient approach to review.

[Matter reference]

Review Plan

[Version 0.1]

[Date]

I.1 Background

This review plan outlines the detailed approach to the review phase of the discovery process.[9]

[Models 1 and 2] The objective of the review phase is to perform a manual review of documents highlighted as potentially relevant through the filtering applied at the processing phase. This will be completed on a document by document basis by [review team]. Each document will have a determination made as to its relevance to the issues in the matter. In addition, documents will be coded for privilege, categorisation and may be redacted as necessary and appropriate.

[[If predictive coding is in use – Model 3] The objective is to carry out a predictive coding exercise to identify potentially relevant documents and then manually review those documents. The predictive coding will be conducted by [expert review team] and the manual review will be completed by [review team]. Each document will have a determination made as to its relevance to the issues in the matter. In addition, documents will be coded for privilege, categorisation and may be redacted as necessary and appropriate.]

I.2 Review team and responsibilities

[Review manager] has been appointed as the review manager for this project. S/he will be responsible for all aspects of the review, including:

- Structuring the review.
- Resourcing the review team.
- Planning the review, including documenting this review protocol.
- Training the review team on which markings/annotations to use and why, to include how families of documents will be managed and marked.
- Forming and managing the Quality Check ('QC') team to perform quality checks on the review decisions made by [review team], and perform QC on the final production.[10]
- Managing the assignment of batches of documents for review.
- Liaising with other service providers.

[Review platform provider] has been appointed and will be responsible for:

- Providing access to [review team] from [law firm/producing party] to the review platform.
- Providing training on how to use the review platform.
- Providing support on the use of the review platform and dealing with technical issues.

---

[9] *While the discovery protocol is intended as a means for information sharing and agreement between the parties, this review protocol is typically not shared between the parties, but rather between the various stakeholders (producing party, internal and external legal advisors, review team, and any external service providers) responsible for conducting the review.*

[10] *The QC team is usually formed with one or more senior reviewers and they may also complete review batches themselves.*

Review platform provider should provide a list of prerequisites for the review team, *which may* include items such as a reliable internet connection and any software installations required to use the review platform. It is highly recommended that any service provider should test the use of their review platform from the review team's location a number of days in advance of the review commencing. This can save significant time should technical issues arise. Review platform provider should also provide a list of scheduled times/days when the review platform is unavailable, for example due to routine maintenance.

I.3 Approach to review

A [single-pass/two-pass/two-pass including predictive coding] review will be conducted [on documents responsive to the filtering criteria previously applied/all unique document families brought forward from the processing phase.]

[Review platform provider should provide a brief overview of the key features of the platform required to complete the review. This might include:

- ***Views –** the main view which contains the list of documents to be reviewed.*
- ***Batches –** Groups of documents are split into batches, which can be assigned to individual reviewers and helps ensure that the same document is not reviewed by different reviewers at the same time as well as helping track progress.*
- ***Marking layout –** This is the section of the system where the reviewer can view the document and make choices regarding markings and annotations, which may be different for each pass of the review.*

[Choose Model 1, 2 or 3 below and delete other content.][11]

Model 1 Review (Two-pass review without predictive coding)

As many of the documents which are responsive to the filtering criteria form part of wider families of documents, a two-pass review will be required to consist of:

First pass

The first pass will identify whether a document is relevant to the discovery categories and will suppress clearly irrelevant documents from further review. This review pass will consider documents in isolation and only one copy of each document responsive to the filtering criteria will be reviewed.

The documents will be split into batches of [200] documents[12] named with the prefix [1P_Emails_BatchXX] and [1P_LooseFilesAndAttachments_BatchXX]. 1P_Emails batches will contain all deduplicated email threads responsive to the filtering criteria, sorted by email thread. 1P_LooseFilesAndAttachments batches will contain email attachments and other loose files sorted by near duplicate.

---

[11] *Review platform-specific terminology and capabilities will be a significant factor in designing the detailed review protocol. The alternatives below contain high-level concepts; however the review platform provider should be consulted in drafting this protocol. The QC steps outlined in these sample approaches are basic in nature and as such an appropriate QC protocol should be devised based on a risk assessment of the overall review phase.*

[12] *200 documents is optimum as it allows sufficient turnover in batches to allow QC to be completed regularly.*

A schedule of the batches will be provided to the review manager, who will complete the first batch with the QC team and compare decisions before the full review commences.[13]

Each reviewer will check a batch out and review the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	Choices	Mandatory
Relevance (Single choice)	Relevant Not Relevant Query Technical Issue	Yes
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

When each first-pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Identify all documents marked as Query, review them, and mark as Relevant, Not Relevant or Technical Issue (i.e. no documents will be left marked as Query).
2. Randomly sample [percentage]% of the documents in a batch for each reviewer.
3. Mark the documents as 'QC Complete'. If changes are made to the first-pass reviewers' markings, then also mark the document as 'QC Changes made'.
4. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with guidance and supervision.

At the conclusion of the first-pass review and at regular intervals during the review [review platform provider] will assess and address any documents marked as Technical Issue. Once any technical issues are resolved, [review platform provider] will mark these documents as Query and direct the review manager on how to access and review them. The review manager will then mark them as relevant or not relevant.[14]

Second pass

Documents marked 'Relevant' through the first pass review, their families (or related documents) and other unique families which contain duplicates, will be included in this review pass. Documents will be considered in the context of their families and will also be considered for privilege and categorisation. Documents requiring redactions will also be identified at this stage.[15]

The documents for review will be split into batches of [200] documents. Batches will be named with the prefix [2P_EmailsAndAttachments_BatchXX] and [2P_LooseFiles_BatchXX]. 2P_EmailsAndAttachments batches will

---

[13] *This can be helpful in ensuring that the review manager and QC team have a consistent understanding of the issues.*

[14] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

[15] *It is often possible to bring documents which have not been subject to filtering criteria (such as scanned hardcopy documents or dedicated project folders) straight to second-pass review and the first-pass review outlined above may be skipped for these sub-sets of document.*

contain all relevant email threads and attachments, sorted by email thread with the parent first followed by its attachment(s). 2P_LooseFiles batches will contain loose files sorted by near duplicate. Note that it is now necessary to combine attachments with their parent emails as they are being considered as a family.

A schedule of the batches will be provided to the review manager. Each reviewer will check a batch out to themselves. They will review the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	**Choices**	**Mandatory**
Relevance (Single choice)	Relevant Not Relevant Query Technical Issue	Yes
Category (Multiple choice)	Category 1 Category 2 Category 3	Yes, but only if Relevant
Privilege and Data Protection (Multiple choice)	Privileged – Legal Advice (Withhold) Privileged – Litigation (Withhold) Part-privileged – Legal advice (For redaction) Part-privileged – Litigation (For redaction) Data protection (For redaction)*	No
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

*This refers to documents which contain personal data which is not relevant to the matter, the disclosure of which may interfere with data subject rights.

When each second-pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Identify all documents marked as Query, review them, and mark as Relevant, Not Relevant or Technical Issue (i.e. no documents will be left marked as Query).
2. Randomly sample [percentage]% of the documents in the batch and review the markings applied by the second-pass reviewer.
3. Mark the documents as 'QC Complete'. If changes are made to the second-pass reviewers markings, then also mark the document as 'QC Changes made'.
4. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with guidance and supervision.

At the conclusion of the second-pass review and at regular intervals during the review, [review platform provider] will assess and address any documents marked as Technical Issue. Once any technical issues are resolved, [review platform provider] will mark these documents as Query and direct the review manager on how to access and review them. The review manager will then mark them as relevant or not relevant.[16]

[16] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

Redaction pass

Individual documents marked for redaction during the second-pass review will be included in this review pass. i.e. all documents marked Part-privileged – Legal advice (For redaction), Part-privileged – Litigation (For redaction), or Data protection (For redaction).[17]

The documents for review will be split into batches of [200] documents named with the prefix [RP_EmailsAndDocuments_BatchXX], [RP_Spreadsheets_BatchXX], [RP_Other_BatchXX]. RP_EmailsAndDocuments batches will contain emails and documents for redaction. RP_Spreadsheets batches will contain spreadsheets for redaction, and RP_Other batches will contain other document types. The latter two sets of batches will typically contain documents which are difficult to image and redact.

A schedule of batches will be provided to the review manager and each reviewer will check a batch out for redaction. They will review and redact the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	Choices	Mandatory
Redaction (Single choice)	Complete No longer required Technical Issue	Yes
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

When each redaction-pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Randomly sample [percentage]% of the documents in the batch and review the redactions applied.
2. Mark the documents as 'QC Complete'. If changes are made, then also mark the document as 'QC Changes made'.
3. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with additional guidance and supervision.

At the conclusion of the redaction-pass review and at regular intervals during the review, [review platform provider] will assess and address documents marked as Technical Issue. Once any technical issues are resolved, [review platform provider] will direct the review manager on how to access and redact them. The review manager will then mark them as complete or no longer required.[18]

This approach facilitates the exclusion of irrelevant documents at an early stage, whilst maintaining the consistency of families of documents for those which are relevant. It is however premised on the use of more junior reviewers at the early stages, with more senior reviewers completing QC and making the final decisions.

---

[17] *It is prudent to perform the redactions in a separate pass due to the technical process used to image the documents prior to redactions and the comprehensive technical QC the review platform provider will need to undertake. It is not advisable for anyone other than expert users of such systems to 'image and redact on-the-fly'. It is also essential that redactions are consistent across different copies of relevant documents to avoid inadvertent waiver of privilege.*

[18] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

Model 2 (Single-pass review without predictive coding)

A single-pass review will be conducted consisting of:

First pass

All documents responsive to the filtering criteria, their families (or related documents) and other unique families which contain duplicates, will be reviewed. This review pass will consider documents in the context of their families and reviewers will code for privilege, categorisation and whether redaction is required.

The documents for review will be split into batches of [200] documents. Batches will be named with the prefix [1P_EmailsAndAttachments_BatchXX] and [1P_LooseFiles_BatchXX]. 1P_EmailsAndAttachments batches will contain all relevant email threads and attachments, sorted by email thread with the parent first followed by its attachment(s). 1P_LooseFiles batches will contain loose files sorted by near duplicate.

A schedule of batches will be provided to the review manager and each reviewer will check a batch out for review. They will review the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	Choices	Mandatory
Relevance (Single choice)	Relevant Not Relevant Query Technical Issue	Yes
Category (Multiple choice)	Category 1 Category 2 Category 3	Yes, but only if Relevant
Privilege and Data Protection (Multiple choice)	Privileged – Legal Advice (Withhold) Privileged – Litigation (Withhold) Part-privileged – Legal advice (For redaction) Part-privileged – Litigation (For redaction) Data protection (For redaction)*	No
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

*This refers to documents which contain personal data which is not relevant, the disclosure of which may interfere with data subject rights.

When each first-pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Identify all documents marked as Query, review them, and mark as either Relevant, Not Relevant, or Technical Issue (i.e. no documents will be left marked as Query).
2. Randomly sample [percentage]% of the documents in the batch and review the markings applied by the first-pass reviewer.
3. Mark the documents as 'QC Complete'. If changes are made to the first-pass reviewers markings, then also mark the document as 'QC Changes made'.
4. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with additional guidance and supervision.

At the conclusion of the first-pass review and at regular intervals during the review, [review platform provider] will assess and address any documents marked as Technical Issue. Once any technical issues are resolved, [review platform provider] will mark these documents as Query and direct the review manager on how to access and review them. The review manager will then mark them as relevant or not relevant.[19]

Redaction pass

Individual documents marked for redaction during the first-pass review will be included in this review pass. i.e. all documents marked Part-privileged – Legal advice (For redaction), Part-privileged – Litigation (For redaction), or Data protection (For redaction).[20]

The documents for redaction will be split into batches of [200] documents. Batches will be named with the prefix [RP_EmailsAndDocuments_BatchXX], [RP_Spreadsheets_BatchXX], [RP_Other_BatchXX]. RP_EmailsAndDocuments batches will contain emails and documents for redaction. RP_Spreadsheets batches will contain spreadsheets for redaction, and RP_Other batches will contain other document types. The latter two sets of batches will typically contain documents which are difficult to image and redact.

A schedule of batches will be provided to the review manager and each reviewer will check a batch out for redaction. They will review and redact the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	Choices	Mandatory
Redaction (Single choice)	Complete No longer required Technical Issue	Yes
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

When each redaction-pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Randomly sample [percentage]% of the documents in the batch and review the redactions applied.
2. Mark the documents as 'QC Complete'. If changes are made, then also mark the document as 'QC Changes made'.
3. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with additional guidance and supervision.

---

[19] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

[20] *Note that it is prudent to perform the redactions in a separate pass due to the technical process used to image the documents prior to redactions and the comprehensive technical QC the review platform provider will need to undertake. It is not advisable for anyone other than expert users of such systems to 'image and redact on-the-fly'. It is also essential that redactions are consistent across different copies of relevant documents to avoid inadvertent waiver of privilege.*

At the conclusion of the redaction-pass review and at regular intervals during the review, [review platform provider] will assess and address any documents marked as Technical Issue. Once any technical issues are resolved, [review platform provider] will direct the review manager on how to access and redact them. The review manager will then mark them as complete or no longer required.[21]

Model 3 (Two-pass review with predictive coding)

Following deduplication and the exclusion into a separate folder of documents not susceptible to predictive coding (for manual review under Model 1 or 2, or simply skip predictive coding and bring straight to second-pass review) a predictive coding review will be undertaken, consisting of:

Predictive coding pass

[[*If pre-filtering has been used*]All documents, their families (or related documents) and other unique families which contain duplicates, will be included in the predictive coding pass.]

[[*If pre-filtering has not been used*] All families of documents which remain after family-level and email thread deduplication will be included in the predictive coding pass.]

The predictive coding pass comprises four main steps:

- **Assessment –** The expert reviewer will be presented with a sample of [number] documents (referred to as the 'Initial Assessment Set' or 'Control Set 1') randomly selected by the system [or a seed set selected using an agreed selection process]. The expert reviewer will mark these documents as relevant or not relevant to the discovery categories as a whole and highlight in the notes field whether a document is to be withheld due to privilege or other withholding requirement. This initial assessment set produces a pool of documents to serve as the 'gold standard' against which the system will measure and test its performance. If the richness of the sample set is low[22], additional documents ('Control Set 2') will be added until the richness is sufficient to proceed.
- **Training –** The system will provide the expert reviewer with small batches of documents to review and mark as relevant or not relevant to the discovery categories and highlight if a document is to be withheld due to privilege or other withholding requirement. This will continue until the system determines that it has learnt enough and reaches a stable state. There may be a number of iterations before the system reaches this point. [The expert reviewer will be presented with statistics at the end of each training set and will work with the system provider's predictive coding expert to determine when the system has learnt enough to reach a stable state.] These batches are referred to as 'Training Sets'.

  Documents may also be marked as Technical Issue during the Assessment and Training steps if they cannot be opened or reviewed by the expert reviewer due to a technical issue. [Producing party] will work with its technology provider to resolve such technical issues and allow the documents to be reviewed and marked as relevant or not relevant.

---

[21] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

[22] *That is, if there is a low preponderance of relevant documents within the sample. This is less likely to occur where the dataset has been generated using filtering criteria such as key word searches developed in accordance with these guidelines.*

- **Decision –** The system will provisionally code the remaining documents in the set with a score of between 0 and 100 indicating each document's likelihood of being relevant. This grading does not indicate actual relevance, but likely relevance. A cut-off point (or threshold) will be decided having regard to the risk of missing relevant documents, the cost of reviewing large volumes of irrelevant documents and proportionality, taking into account the precision and recall [and f-measure] reported by the system.
- **Verification –** Once the cut-off point has been determined, statistical testing of everything below the cut-off point will be carried out to include:
    - Taking a random sample of [number] of the documents below the cut-off point and reviewing them in order to determine if there are any relevant documents which were not identified by the system.
    - Discrepancy analysis to identify discrepancies between system and human decisions, with any results being fed back into the system for further training.
    - Document type, theme-based and near-duplicate searching to complete further discrepancy analysis.

  Should the verification process identify more than [percentage]% documents incorrectly coded by the system, the assessment and training phase will be repeated. Assuming an overall sample size of [500] documents, and a desired confidence level of [95]%, then the margin of error will be [1.9]% if [5]% of the sampled documents are found to be relevant. If [1]% of the sampled documents are found to be relevant, then the margin of error will be [0.8]%.
- **Disclosure –** In advance of proceeding to the second-pass, [Producing party] will inform [requesting party] of the results of the predictive coding process including the precision and recall [and f-measure] reported by the system, the results of verification and the proposed cut-off point.

  [Producing party] will provide access to [requesting party]'s nominated [counsel] to a schedule of all documents marked as relevant or not relevant during the assessment and training steps, save those marked for withholding.[23]

  Or [[Producing party] will provide an independently appointed [solicitor/counsel] with access to all documents marked relevant and not relevant during the assessment and training steps, save those marked for withholding.] The appointed [solicitor/counsel] will not disclose any information relating to the documents to [requesting party], however may discuss queries regarding designations of documents as relevant or not relevant with [producing party]'s legal advisors.]

  [If the parties cannot agree on the cut-off point and/or the decisions made on any disputed documents during the assessment and training steps, they should meet in person with the relevant experts to discuss. If agreement cannot be reached, then either party may apply to the court for directions.]

---

[23] *This schedule would provide a unique identification number, the name/title of the document, the document type, author/sender, recipient, and created/last modified date/time. This would allow the requesting party to perform an initial review of this information before determining if it is necessary to inspect any specific underlying documents.*

Second pass (manual review pass)

It is not possible to determine in advance what the cut-off point might be. Once the cut-off point has been decided, all documents which have a relevancy probability above the cut-off point, their families (or related documents) and other unique families which contain duplicates, will be included in this manual review pass. Documents will be considered in the context of their families and will also be considered for privilege, categorisation and redaction as appropriate and necessary.

The documents for review will be split into batches of [200] documents. Batches will be named with the prefix [2P_EmailsAndAttachments_BatchXX] and [2P_LooseFiles_BatchXX]. 2P_EmailsAndAttachments batches will contain all relevant email threads and attachments, sorted by email thread with the parent first followed by its attachment(s). 2P_LooseFiles batches will contain loose files sorted by near duplicate.

A schedule of batches will be provided to the review manager and each reviewer will check a batch out for review. They will review the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	Choices	Mandatory
Relevance (Single choice)	Relevant Not Relevant Query Technical Issue	Yes
Category (Multiple choice)	Category 1 Category 2 Category 3	Yes, but only if Relevant
Privilege and Data Protection (Multiple choice)	Privileged – Legal Advice (Withhold) Privileged – Litigation (Withhold) Part-privileged – Legal advice (For redaction) Part-privileged – Litigation (For redaction) Data protection (For redaction)*	No
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

*This refers to documents which contain personal data which is not relevant, the disclosure of which may interfere with data subject rights.

When each manual review pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Identify all documents marked as Query, review them, and mark as either Relevant, Not Relevant, or Technical Issue (i.e. no documents will be left marked as Query).
2. Randomly sample [percentage]% of the documents in the batch and review the markings applied by the reviewer.
3. Mark the documents as 'QC Complete'. If changes are made to the reviewers' markings, then also mark the document as 'QC Changes made'.
4. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with guidance and supervision.

At the conclusion of the manual review pass review and at regular intervals during the review, [review platform provider] will assess and address any documents marked as Technical Issue. Once any technical issues are

resolved, [review platform provider] will mark these documents as Query and direct the review manager on how to access and review them. The review manager will then mark them as relevant or not relevant.[24]

Redaction pass

Individual documents marked for redaction will be included in this review pass. i.e. all documents marked Part-privileged – Legal advice (For redaction), Part-privileged – Litigation (For redaction), or Data protection (For redaction).[25]

The documents for review will be split into batches of [200] documents named with the prefix [RP_EmailsAndDocuments_BatchXX], [RP_Spreadsheets_BatchXX], [RP_Other_BatchXX]. RP_EmailsAndDocuments batches will contain emails and documents for redaction. RP_Spreadsheets batches will contain spreadsheets for redaction, and RP_Other batches will contain other document types. The latter two sets of batches will typically contain documents which are difficult to image and redact.

A schedule of batches will be provided to the review manager and each reviewer will check a batch out for redaction. They will review and redact the batch before checking it back in as completed. The reviewer will be presented with a marking layout which will contain the following choices:

Mark	Choices	Mandatory
Redaction (Single choice)	Complete No longer required Technical Issue	Yes
Comments (Free text field)	n/a	No
Quality Control (Multiple choice)	QC Complete QC Changes made	Only available to QC team

When each redaction-pass batch has been completed, the reviewer will notify the review manager. The review manager will then assign/notify the batch to a member of the QC team, who will complete the following steps:

1. Randomly sample [percentage]% of the documents in the batch and review the redactions applied.
2. Mark the documents as 'QC Complete'. If changes are made, then also mark the document as 'QC Changes made'.
3. If the number of incorrectly marked documents is greater than [percentage]%, then the batch will be assigned to be reviewed again, in addition to providing the reviewer with additional guidance and supervision.

At the conclusion of the redaction-pass review and at regular intervals during the review, [review platform provider] will assess and address any documents marked as Technical Issue. Once any technical issues are

[24] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

[25] *It is prudent to perform the redactions in a separate pass due to the technical process used to image the documents prior to redactions and the comprehensive technical QC with the review platform provider will need to undertake. It is not advisable for anyone other than expert users of such systems to 'image and redact on-the-fly'. It is also essential that redactions are consistent across different copies of relevant documents to avoid inadvertent waiver of privilege.*

resolved, [review platform provider] will direct the review manager on how to access and redact them. The review manager will then mark them as complete or no longer required.[26]

This approach facilitates the exclusion of irrelevant documents at an early stage, whilst maintaining the consistency of families of documents for those which are relevant. It is, however premised on the use of more junior reviewers at the early stages, with more senior reviewers completing QC and making the final decisions.

I.4 Production criteria

At the conclusion of the redaction pass, documents marked as 'Relevant' will be produced.

As outlined in the discovery protocol, documents will be produced in their native format by default. Exceptions to this include:

- [Hardcopy documents which have been scanned will be produced in PDF format.]
- Corrupt, password protected or encrypted documents may be converted to a different format (such as PDF) which enables their use.
- Redacted documents will be produced in [format, such as redacted PDF or TIFF] and will also be identified in the schedule as having being redacted.
- In rare occasions, modified versions of native documents may be produced. This may be the case where it is neither practical, possible, or proportionate to image before redaction. Examples might include very large spreadsheets. Any such documents will be separately identified in the schedule.
- [Documents which require redactions, but cannot reasonably be redacted, such as very large spreadsheets or databases, may be provided for by inspection only.]

Document families, such as emails and their attachments, will be produced as follows:

- If a parent email is relevant, but its children are not, then only the parent will be produced.
- If a child attachment is relevant, then its parent email will be produced for context. Orphan child attachments will not be produced in isolation.
- If there are multiple child attachments but only one is relevant, then only the relevant attachment and parent email will be produced. Irrelevant attachments will not be produced.

Note: It is important that families of documents are marked according to the agreed criteria during the review. This will help ensure that effort is not expended late in the process correcting the consistency of family markings.

The following QC steps will be completed prior to production [usually by the review platform provider]:

- Verify that documents containing hidden data have been marked for production in image format, if required.
- Verify that families of documents have been marked appropriately from a relevancy and privilege perspective.

---

[26] *These quality control steps are considered 'batch-by-batch' QC. Further statistical or targeted (by keywords for instance) samples might be prudent across multiple batches and/or the set of documents for review.*

- Verify that documents marked for redaction are redacted and that no documents have been redacted which were not marked for redaction.
- Verify that categories and other markings have been applied to documents where required.

Verify that there are no containers within the production set which could result in an embedded item being inadvertently disclosed (e.g. an email attachment which is also an email and is still in a format which contains and embedded copy of a spreadsheet which has not been reviewed).

I.5 Review phase timelines

The current deadline for production is [date/time]. In order to complete QC and production and make the necessary copies of the production (documents and schedules) available, the review phase would need to be complete by [date/time].

The following provisional timelines have been agreed for the review phase:

- First-pass to commence on [date/time] and finish by [date/time].
- Second-pass to commence on [date/time] and finish by [date/time].
- Redaction-pass to commence on [date/time] and finish by [date/time].

It is very difficult to estimate in advance the number of documents which will require second-pass review and/or redactions, therefore the times outlines are indicative only.

Appendix J Sample request for voluntary discovery

[From solicitor for requesting party]

[To solicitor for responding party]

[Date]

[Matter reference]

Dear Sirs,

We refer to the above matter and to previous correspondence in relation to these proceedings. This letter constitutes our formal request pursuant to the Rules of the Superior Court seeking voluntary discovery from the [Plaintiff/Defendant].

TAKE NOTICE that, in accordance with the terms of Order 31, rule 21, of the Rules of the Superior Courts (as amended), the [Plaintiffs/Defendants] hereby require the [Plaintiff/Defendant] to make voluntary discovery of all documents which are or have been within its possession, power, or procurement, within the following categories:

Category 1

[Describe in detail the category of document being requested.]

Reasons

[Describe in detail the reasons for the category being requested.]

Category 2

[Describe in detail the category of document being requested.]

Reasons

[Describe in detail the reasons for the category being requested.]

Category 3

[Describe in detail the category of document being requested.]

Reasons

[Describe in detail the reasons for the category being requested.]

Category 4

[Describe in detail the category of document being requested.]

Reasons

[Describe in detail the reasons for the category being requested.]

Other

[Insert any definitions, as required to provide clarity to the categories and reasons.]

And **TAKE NOTICE** that:

1. Voluntary discovery is requested pursuant to Order 31, rule 12.
2. Any agreement to make discovery would require it to be made on oath in a manner and form and will have such effect as if directed by order of the Court.
3. Discovery is required to be made with the documents listed in a manner which allows the categories which they respond to be clearly identified.
4. Where documents of which discovery is sought exist in electronic format, production of the same in searchable form is requested. [Plaintiff/Defendant] reserves its position as to whether it will be necessary to seek the provision of inspection and searching facilities using any information and communications technology system owned or operated by the [Plaintiff/Defendant].
5. Objection will be taken to any attempt to adduce in evidence a document which has not been discovered.

In circumstances where the [Plaintiff/Defendant] confirms that they will make voluntary discovery, we require discovery to be made by affidavit sworn by them and furnished to us, together with copies of all documentation, within a period of [14] weeks from the date hereof.

Kindly note that in circumstances where [Plaintiff/Defendant] do not confirm that they will make voluntary discovery of all documentation referred to above, or if such confirmation is not received within a period of [2] weeks hereof, we will have no option but to issue a motion seeking discovery of the categories of documents identified above without further notice to you. Furthermore the content of this letter will be used to seek to fix your client with the costs of any application necessitated by reason of your clients' failure to make discovery as requested.

Yours faithfully,

[Solicitor for Plaintiff/Defendant]

Appendix K Sample affidavit of discovery

THE HIGH COURT

[Commercial]

Record No. [INSERT YEAR] [INSERT NO.] [P/S]

BETWEEN:

[INSERT PARTY[ies]

Plaintiff[s]

-and-

[INSERT PARTY [ies]

Defendant[s]

DRAFT / [SUPPLEMENTAL][27] AFFIDAVIT OF DISCOVERY

I, **[INSERT NAME]**, [**INSERT PROFESSION**], of [**INSERT ADDRESS**], aged eighteen years and upwards, **MAKE OATH AND SAY** as follows:

1. I am the [**INSERT DETAILS OF DEPONENT**] of the [**Plaintiff/Defendant**] herein and I Make this Affidavit of Discovery on its behalf and with its authority from facts within my knowledge save where otherwise appears and whereso appearing I believe same to be true and accurate.

2. Pursuant to correspondence between the parties [and the determination of [**INSERT NAME OF JUDGE** / The Master of this Honourable Court dated [**INSERT DATE**], I am advised and believe that the [**INSERT NAME OF PARTY**] is obliged to make discovery of [**INSERT NUMBER**] categories of documents as sought by the [**INSERT NAME OF PARTY**] (the "[Ordered] Discovery") as set out below:

[INSERT FULL DETAILS OF ALL CATEGORIES OF DISCOVERY HERE]

---

[27] Where Affidavit is to be an Affidavit Supplemental to an Original Affidavit of Discovery then wording in the following terms should be inserted into the Supplemental Affidavit of Discovery: "This Affidavit is supplemental to my Affidavit of Discovery sworn on [**INSERT DATE**] in these proceedings (the Original Affidavit of Discovery"). If required to provide reasons as to why a Supplemental Affidavit is being sworn then the following sample text may be of assistance: "This Supplemental Affidavit of Discovery is sworn principally in respect of the additional categories of discovery sought by [**INSERT NAME OF PARTY**] arising out of [matters pleaded in the Amended Statement of Claim. It also includes further documents located with respect to the Original Discovery]"

3. In order to comply with its discovery obligations, the [**INSERT NAME OF PARTY**], as advised by its solicitors, has conducted a wide-ranging and extensive search for documentation relevant to the Ordered Discovery. In particular, contact was made with all [insert appropriate description of staff etc. e.g. - details of officers and employees of the [**INSERT NAME OF PARTY**] [and all third parties] whom it was believed might have relevant documentation or be able to identify where such documents might be located. The documentation was then accumulated centrally and reviewed for the purposes of determining its relevance to the Ordered Discovery. As a result, the [**INSERT NAME OF PARTY**] has in its possession, power or procurement the documents which come within the terms of the Ordered Discovery set forth in the First Schedule hereto.

4. I wish to point out that the [**INSERT NAME OF PARTY**] in making discovery has for ease of reference listed each document being discovered under one of the [**INSERT NUMBER**] categories of documents within the Ordered Discovery. It is the case, however, that there is an overlap between various categories in the Ordered Discovery whereby a document might be considered to fall under a number of categories. The [**INSERT NAME OF PARTY**] has been advised that it is not obliged, and could not reasonably be expected, to identify every category under which a particular document might be listed. Accordingly, while the documentation listed in the First Schedule under the [**INSERT NUMBER**] categories comprises the totality of documentation which the [**INSERT NAME OF PARTY**] is in a position to produce under the Ordered Discovery, the [**INSERT NAME OF PARTY**] does not thereby suggest nor is it the case that all the documents listed under any particular category are relevant to that category nor do they comprise all of the documents possibly relevant to that category out of the documentation being produced.

5. The [**INSERT NAME OF PARTY**] has in its possession, power or procurement the documents[28] [and electronically stored information][29] relating to the matters in question in this suit and falling within the Discovery[30] as set forth in the First and Second parts of the First Schedule hereto.

6. The [**INSERT NAME OF PARTY**] objects to producing the documents [and electronically stored information] set out under [**SPECIFY PARAGRAPH NO.**] in the Second Part of the First Schedule hereto on the grounds that they are privileged and that they comprise communications of a confidential nature passing between the [**INSERT NAME OF PARTY**] and its legal advisers for the purposes of obtaining legal advice for or giving legal advice to the [**INSERT NAME OF PARTY**]. The [**INSERT NAME OF PARTY**] objects to producing the documents set out under [**SPECIFY PARAGRAPH NO.**] in the Second Part of the First Schedule on the grounds that they are privileged in that they comprise documents that came into existence after these proceedings were contemplated or commenced and were created with a view to defending such proceedings either for the purposes of giving or obtaining advice in relation to them or of obtaining and collecting evidence to be used or of obtaining information which might lead to the obtaining of such evidence or for the purposes of defending these proceedings.

---

[28] Documents of the same or a similar nature, when numerous, must so far as possible, be grouped together and numbered or otherwise sufficiently marked so as to be identifiable.

[29] The Rules Amend Order 31 Rule 12 of the Rules of the Superior Courts and in particular make provision for the discovery of electronically stored information.

[30] Parties providing discovery shall list documents or categories of information, and shall provide documents and information for inspection, in a manner corresponding with the categories in the agreement or order for discovery, or in a sequence corresponding with the manner in which the documents or information have been stored or kept in the usual course of business by the party making discovery.

7. The [**INSERT NAME OF PARTY**] has had, but does not now have, in its possession, power or procurement the documents [and electronically stored information] relating to the matters in question in this suit that are set forth in the Second Schedule hereto[31].

8. The last mentioned documents [and electronically stored information] were last in my possession, power or procurement on [**INSERT DATE**].

9. That [here state what has become of the last-mentioned documents or information, and in whose possession they now are].

9.A [Insert paragraph(s) dealing with redaction of documents as appropriate].

10. According to the best of my knowledge, information, and belief, [**INSET NAME OF PARTY**] has not now, and never had in its possession, power or procurement or in the possession, custody or power of its solicitors or agents, or in the possession, custody or power of any other persons, or person on its behalf, any document of any kind or any electronically stored information, or any copy of or extract from any such document or information relating to the matters in question in this suit, or any of them, or wherein any entry has been made relative to such matters, or any of them, and falling within the relevant categories of documents specified in the [[32]]other than and except the documents [and electronically stored information] set forth in the said First and Second Schedules hereto.

11. I understand that the obligation on a party giving discovery is to discover all documents and electronically stored information within [his/her/its] possession, power or procurement within the categories agreed or ordered to be delivered that contain information which may enable the party receiving the discovery to advance its own case or to damage the case of the party giving discovery or which may fairly lead to a train of inquiry which may have either of those consequences.

---

[31] Additional text that might be considered here, where the context permits is: "In addition to the documents set out at [SPECIFY PARAGRAPH] in the Second Schedule, it is possible, having regard to the scope of the Discovery and the time period to which it relates, that some documents as described at [SPECIFY PARAGRAPH] in the Second Schedule, within the scope of the Discovery, have not been retained or may have been overwritten in the ordinary course of business of the [INSERT PARTY] prior to the commencement of these proceedings."

[32] Specify whether letter requesting voluntary discovery of [Insert Date] or [order of [the Master of] this Honourable Court dated [Insert date]].

SWORN by the said [NAME OF DEPONENT] on the [DATE] day of [DATE] 20XX at [ADDRESS], before me a Commissioner for Oaths / Practicing Solicitor and [I know the Deponent] / [the Deponent has been identified to me by [NAME] who is personally known to me] / [prior to the swearing of this affidavit, the identity of the deponent has been established by me by reference to [insert particulars of photographic ID e.g. a passport (passport no. [*Passport number*] issued on [*date of issue*])

_____ _____

[NAME OF DEPONENT] Commissioner for Oaths/Practicing Solicitor

This Affidavit was filed by [Law firm name], Solicitors for the [**INSERT PARTY**], [Law firm address], on the [**INSERT DAY**] day of [**INSERT DATE**].

FIRST SCHEDULE

First Part

[INSERT INDEX OF DISCOVERABLE DOCUMENTATION]

Draft Schedule

Link	Family ID	Doc ID	Family	Date	Type	Author	Recipient	Cc	Bcc	Name /Subject	Category	Redaction

FIRST SCHEDULE

Second Part

1. **[Insert Schedule of Privileged Documentation or generic paragraph in relation to same]**

Family ID	Doc ID	Family	Date	Type	Author	Recipient	Cc	Bcc	Name /Subject	Category	Privilege

2.

(a) All documents including correspondence, notes and memoranda passing between the [**INSERT PARTY**] and [Legal advisors] seeking advice in relation to the various aspects of the matters the subject of the proceedings herein.

(b) All correspondence with and advices, draft pleadings and opinions of Counsel in relation to the various matters the subject of the proceedings herein.

(c) All correspondence and advices received from expert witnesses retained on behalf of the [**INSERT PARTY**].

(d) All documents including various memoranda, notes of meetings, correspondence, reports and drafts thereof produced by the [**INSERT PARTY**], [Legal advisors], Counsel [and experts] for the purposes of these proceedings.

SECOND SCHEDULE

The High Court

[Commercial]

Record No. [ ]

Between:

[Insert Party [ies]

Plaintiff[s]

- and -

[Insert Party [ies]

Defendant[s]

AFFIDAVIT

OF DISCOVERY

[Law firm name]

Ref: - [**INSERT REF**.]

Appendix L Consolidated version of current rules

RULES OF THE SUPERIOR COURTS (NO. 2) (DISCOVERY), 1999 AS AMENDED BY RULES OF THE SUPERIOR COURTS (DISCOVERY) 2009

1. The Rules of the Superior Court are hereby amended:

(i) By the substitution for rule 12 or Order 31 of the following:

12.(1) Any party may apply to the Court by way of notice of motion for an order directing any other party to any cause or matter to make discovery on oath of the documents which are or have been in his possession, power or procurement, relating to any matter in question therein. Every such notice of motion shall specify the precise categories of documents in respect of which discovery is sought and shall be grounded upon the affidavit of the party seeking such an order of discovery which shall:

(*a*) Verify that the discovery of documents sought is necessary for disposing fairly of the cause or matter or for saving costs;

(*b*) Furnish the reasons why each category of documents is required to be discovered; and

(c) Where the discovery sought includes electronically stored information, specify whether such party seeks the production of any documents in searchable form and if so, whether for that purpose the party seeking discovery seeks the provision of inspection and searching facilities using any information and communications technology system owned or operated by the party requested.

(2) On the hearing of such application the Court may:

(a) Either refuse or adjourn the same, if satisfied that such discovery is not necessary, or not necessary at that stage of the cause or matter, or by virtue of non-compliance with the provisions of sub-rule (6), or

(b) Make an order for discovery either in terms of some or all of the categories of documents sought or limited to certain documents or classes of documents within any or all of those categories, or otherwise as may be thought fit, and on terms as to security for the costs of discovery or otherwise, and for this purpose may adjourn the application in part;

(c) Where the discovery ordered includes electronically stored information and the Court is satisfied that such electronically stored information is held in searchable form and can be provided in the manner hereinafter referred to without significant cost to the party from whom discovery is requested:

(i) Further order that the documents or classes of documents specified in such order be provided electronically in the searchable form in which they are held by the party ordered to make discovery, or

(ii) Where the Court is satisfied that such documents or classes of documents, or any information within such documents, could not, if provided electronically, be subjected to a search by the party seeking discovery without incurring unreasonable expense, further order that the party ordered to make discovery make available inspection and searching facilities using its own information and communications technology system, so as to allow the party seeking discovery to avail of any search functionality available to the party ordered to make discovery.

(3)

(a) Any order made under sub-rule (2)(c) may include such provision or restriction and be subject to such undertakings from any party or person as the Court may consider necessary to ensure that documents discovery of which has not been ordered are not accessed or accessible, and otherwise to secure the information and communications technology system concerned.

(b) Such order may in particular include a provision that the inspection and searching of documents shall be undertaken by an independent expert or person agreed between the parties, or appointed by the Court in default of agreement (instead of being undertaken by the party seeking discovery), who may conduct such inspections and searches as may be required and report the results to the party seeking discovery.

(c) Where such order makes provision for inspection and searching of documents in the manner referred to in paragraph (b), the party seeking the order shall indemnify such independent expert or person in respect of all fees and expenses reasonably incurred by him, and the fees and expenses so indemnified shall form part of the costs of that party for the purposes of Order 99.

(4)

(a) Documents of the same or a similar nature and not in electronic form, when numerous, shall so far as possible be grouped together and numbered or otherwise sufficiently marked so as to be identifiable.

(b) Parties providing discovery shall list documents or categories of information, and shall provide documents and information for inspection, in a manner corresponding with the categories in the agreement or order for discovery and, subject to any such agreement or order, in a sequence corresponding with the manner in which the documents or information have been stored or kept in the usual course of business by the party making discovery.

(5) An order shall not be made under this rule if and so far as the Court shall be of the opinion that it is not necessary either for disposing fairly of the cause or matter or for saving costs.

(6) An order under sub-rule (2) directing any party or under rule 29 directing any other person to make discovery shall not be made unless:

(*a*) The applicant for same shall have previously applied by letter in writing requesting that discovery be made voluntarily-

(i) Specifying the precise categories of documents in respect of which discovery is sought,

(ii) Furnishing the reasons why each category of documents is required to be discovered,

(iii) Where the discovery sought includes electronically stored information, specifying whether the applicant seeks the production of any documents in searchable form and if so, whether for that purpose the applicant seeks the provision of inspection and searching facilities using any information and communications system owned or operated by the party requested, and

(*b*) A reasonable period of time for such discovery has been allowed; and

(*c*) The party or person requested has failed, refused or neglected to make such discovery or has ignored such request.

Provided that in any case where by reason of the urgency of the matter or consent of the parties, the nature of the case or any other circumstances which to the Court seem appropriate, the Court may make such order as appears proper, without the necessity for such prior application in writing.

(7) Any such discovery sought and agreed between parties or between parties and any other person shall, subject to sub-rule 4(9), be made in like manner and form and have such effect as if directed by order of the Court.

(8) In any case in which discovery has been sought and agreed and has not been made within the time agreed, the party who has sought same may make application pursuant to rule 21 provided that when seeking discovery the party requested was informed that:

(*a*) Such voluntary discovery was being sought pursuant to Order 31 rule 12;

(*b*) Agreement to make discovery would require it to be made in like manner and form and would have such effect as if directed by order;

(*c*) Failure to make discovery may result in an application pursuant to rule 21;

and the Court may, if satisfied that it is proper so to do, make such order under this rule, rule 19 or rule 21 as is appropriate or such other order as appears just in the circumstances.

(9) An application for discovery whether under sub-rule (1) or (46) shall be made not later than twenty-eight days after the action has been set down or in matters which are not set down, twenty-eight days after it has been listed for trial provided that the Court may order or the party requested may agree, to extend the time for the application for discovery in any case in which it appears just and reasonable so to do.

(10) The costs of an application to Court for discovery in any case in which prior written application has not been made or in which application has not been made within the time provided, shall be in the discretion of the Court.

(11) Any party concerned by the effect of an order or agreement for discovery may at any time, by motion on notice to each other party concerned, apply to the Court for an order varying the terms of the discovery order or agreement. The Court may vary the terms of such order or agreement where it is satisfied that-

(i) Further discovery is necessary for disposing fairly of the case or for saving costs, or

(ii) The discovery originally ordered or agreed is unreasonable having regard to the cost or other burden of providing discovery.

(12) An order under sub-rule (11) shall not be made unless:

(a) The applicant for same shall have previously applied by letter in writing to the other party specifying the variations sought to the order, furnishing the reasons why each variation is sought and requesting that party's agreement to the variations sought, and

(b) A reasonable period of time for agreement has been allowed, and

(c) The party or person requested has failed, refused or neglected to agree to such variation or has ignored such request.

(13) "documents", for the purposes of this rule and rule 29, includes all electronically stored information, and the reference to "business documents" in rule 20 shall be construed accordingly."

(ii) By the substitution of Form No.10 in Appendix C of the Form appended.

Other Referenced and Amended Rules of the Superior Court

Order 31 – Rule 19

19. If the party from whom discovery of any kind or inspection is sought objects to the same, or any part thereof, the Court may, if satisfied that the right to the discovery or inspection sought depends on the determination of any issue or question in dispute in the cause or matter, or that for any other reason it is desirable that any issue or question in dispute in the cause or matter should be determined before deciding upon the right to the discovery or inspection, order that such issue or question be determined first, and reserve the question as to the discovery or inspection.

Order 31 – Rule 20 (As amended)

20. (1) Where inspection of any business documents is applied for, the Court may, instead of ordering inspection of the original document, order a print or copy of any entries therein to be furnished and verified by the affidavit of some person who has examined the print or copy with the original entries, and such affidavit shall state whether or not there are in the original documents any and what erasures, interlineations, or alterations. Provided that, notwithstanding that such print or copy has been supplied, the Court may order inspection of the documents from which the print or copy was made.

(2) Where on an application for an order for inspection privilege is claimed for any document, the Court may inspect the document for the purpose of deciding as to the validity of the claim for privilege.

(3) The Court may, on the application of any party to a cause or matter at any time, and whether an affidavit or list of documents shall or shall not have already been ordered or made, make an order requiring any other party to state by affidavit whether any one or more specific documents, to be specified in the application, is or are, or has or have at any time been in his possession or power; and, if not then in his possession, when he parted with the same, and what has become thereof. Such application shall be made on an affidavit stating that in the belief of the deponent the party against whom the application is made has, or has at some time had, in his possession or power the document or documents specified in the application, and that they relate to the matters in question in the cause or matter, or to some of them.

Order 31 – Rule 21 (Remedies for Non-Compliance)

21. If any party fails to comply with any order to answer interrogatories, or for discovery or inspection of documents, he shall be liable to attachment. He shall also, if a plaintiff be liable to have his action dismissed for want of prosecution, and, if a defendant, to have his defence, if any, struck out, and to be placed in the same position as if he had not defended, and the party interrogating may apply to the Court for an order to that effect, and an order may be made accordingly.

Order 31 Rule 29 (Non Party Discovery)

29. Any person not a party to the cause or matter before the Court who appears to the Court to be likely to have or to have had in his possession custody or power any documents which are relevant to an issue arising or likely to arise out of the cause or matter or is or is likely to be in a position to give evidence relevant to any such issue may by leave of the Court upon the application of any party to the said cause or matter be directed by order of the Court to answer such interrogatories or to make discovery of such documents or to permit inspection of such documents. The provisions of this Order shall apply mutatis mutandis as if the said order of the Court had been directed to a party to the said cause or matter provided always that the party seeking such order shall indemnify such person in respect of all costs thereby reasonably incurred by such person and such costs borne by the said party shall be deemed to be costs of that party for the purposes of Order 99.

Appendix M Overview of legal privilege

Whether a particular document or category of documents might be considered privileged under Irish law requires an examination of its content and/or the context of the document and its creation. Only the Courts are competent to decide whether a claim of privilege is justified. A solicitor has a duty to assert privilege in respect of a document that appears to be privileged unless the client opts to waive privilege in respect of the document.

The main categories of privilege are:

1. Legal professional privilege (legal advice privilege and litigation privilege)
2. Without prejudice privilege
3. Common interest privilege
4. Public interest privilege
5. Journalistic privilege
6. Privilege against self-incrimination

M.1 Legal professional privilege

Legal professional privilege includes two distinct categories: legal advice privilege and litigation privilege.

Legal advice privilege

a) This exists over confidential communications between a lawyer and a client giving or obtaining legal advice. Legal advice is private between the parties and cannot be disclosed to another person without the consent of the client.

b) However, where a client communicates with his/her lawyer for the purpose of seeking "legal assistance" rather than legal advice there is not sufficient public interest to justify rendering such communications privileged. Legal advice is generally advice dealing with legal rights, obligations and remedies rather than administrative or transactional legal assistance.

c) Communications of fact and letters written on a client's instructions in relation to purely transactional matters as distinct from matters requiring confidential legal advice will not attract legal advice privilege.

When making discovery of documents over which legal advice privilege is asserted, where only a portion of the document contains legal advice this should be redacted (on agreement between the parties) and the document discovered as part privileged unless the parties expressly agree otherwise.

The professional relationship of lawyer and client must exist for legal advice privilege to apply. It is designed to encourage full and frank communication between a client and lawyer so that the lawyer is fully informed of the facts of the legal matter. Under Irish law, in-house lawyers and their employers are entitled to the same legal professional privilege as applies to external lawyers (noting however some differences in relation to certain EU investigations and civil law jurisdictions which may treat in-house and external lawyers differently in this regard). Advice given by an adviser who is not a qualified lawyer (or a trainee supervised by a qualified lawyer) does not attract legal advice privilege.

Legal advice privilege can extend to third parties but only where the third party is an agent for the purpose of communicating with the other party to give or obtain legal advice, not just an agent in the general sense.

Unlike litigation privilege (below), legal advice privilege does not protect communications between the client or lawyer and a third party such as a witness or an expert.

Litigation Privilege

a) This is the privilege that exists over confidential documents created because of an apprehension or contemplation of litigation or for the dominant purpose of prosecuting or defending litigation. This may include proceedings before a tribunal and/or regulatory or criminal proceedings.

b) Litigation privilege also exists over documents which come into existence after the commencement of litigation and for the dominant purpose of the litigation.

c) The test applied by the Courts in assessing litigation privilege is known as the "dominant purpose" test. The Court must be satisfied that the primary or dominant purpose of the client in creating the specific document was litigation either pending or threatened. If there are other equally important purposes and litigation is not the dominant purpose, the document will not attract litigation privilege.

d) Litigation privilege may be claimed over communications between clients and third parties if they are created for the dominant purpose of the litigation.

It is possible that part of a communication may be privileged notwithstanding that the document itself is not privileged. An example may be board meeting minutes which refer to legal advice received. Privilege may be claimed over that part of the document which contains or refers to privileged information. In these circumstances, and on agreement between the parties, the privileged information may be redacted and the document disclosed as part privileged.

M.2 Without prejudice privilege

Communications by parties to a dispute which are written or made for the purpose of settling that dispute and which are either expressed to be or are otherwise proved to have been made on a "without prejudice" basis are privileged. The purpose of the communication must be to try to settle the dispute/proceedings.

For a claim of without prejudice privilege to succeed the party claiming it must establish that the communication in question was made:

a) In a bona fide attempt to settle a dispute between the parties, and

b) With the intention that if the negotiations failed, the communication could not be disclosed without the consent of the parties.

The use of the words "without prejudice" are not sufficient in themselves to invoke privilege.

M.3 Common interest privilege

Common interest privilege preserves privilege in documents that are disclosed to third parties where a person voluntarily discloses a privileged document to a third party who has a common interest in the subject matter of the privileged document or in litigation in connection with which the document was brought into existence (e.g. a co-defendant). The common interest must exist at the time of the disclosure and it applies to both legal advice privilege and litigation privilege.

Examples of Relationships which are capable of giving rise to or supporting necessary common interest include: an insurer and the insured, the reinsurer and reinsured, the principle and agent, groups of companies, joint venture partners and co-defendants.

In examining whether a document is the subject of common interest privilege it is important to consider:

a) Whether the document would, in the hands of a single party, have had the benefit of privilege in the first place. If not, then no question of common interest privilege can arise.

b) If, however, the document passes the first test and has been released by one party to a second party it is necessary to ask whether the release was on foot of a common interest in either the litigation or advice.

c) If so, then the document remains privileged, notwithstanding its release by virtue of the doctrine of common interest privilege.

d) If not, then the release might be taken to be a waiver of any privilege which would otherwise have attached to the document.

M.4 Public interest privilege

Public interest privilege is not confined in its application to the executive functions of the State. It is also available where the balance of the public interest favours non-disclosure.

Where a claim of public interest privilege is made the Court is required to balance public interest in the proper administration of justice against the public interest put forward for non-disclosure in order to decide which interest is the superior public interest in the circumstances of the case.

The Executive cannot prevent the Courts from examining documents relevant to any issue in a civil trial for the purposes of deciding if they should be produced.

The categories of public interest in favour of non-disclosure include:

a) National security;

b) International relations;

c) The proper functioning of the public service; and

d) The prevention and detection of crime.

In order for a claim of public interest privilege to succeed, it is essential to show that the communication was brought into being in circumstances of confidentiality. In addition the courts will refuse to allow a claim in favour of non-disclosure of a class of documents. In order for the claim of privilege to succeed, it must be particularised and the damage identified to the public interest in question which will accrue from disclosure of each individual document.

M.5 Journalistic privilege

A journalist may be entitled to withhold from production documents which tend to reveal his or her confidential sources on grounds of journalistic privilege but such documents must be discovered by listing in the affidavit as to documents as with other privileged documents.

M.6 Privilege against self-incrimination

The privilege against self-incrimination provides a general immunity against any compulsion to produce information or documents which may incriminate the producing party. Where an order for discovery is made and the producing party wishes to assert this privilege in respect of a document or documents, the documents must be listed in the usual way in the first schedule second part and the fact that the privilege against self-incrimination is asserted identified expressly in the affidavit. It is important to note that the privilege must be asserted by the person claiming the privilege, or rather by the person who would be incriminated if the documents were disclosed.

M.7 Inadvertent waiver of privilege

In order to be privileged a document must be confidential, but confidentiality in itself does not give rise to privilege. It is possible that highly sensitive client documents will not be privileged and must be disclosed.

Disclosure to a third party or for a limited purpose does not always waive privilege and there is no universal rule that the disclosure of documents produced for the sole purpose of seeking legal advice or litigation to a stranger to that litigation constitutes a waiver of privilege in the document. It is advisable to record in writing any conditions regarding a limited disclosure.

It is open to a client to waive privilege and he may do so at any time in proceedings. However a client may not re-assert his right to privilege once it has been waived either expressly or by implication.

Where a client destroys the confidentiality of a document by choosing to disclose it to the opposing party or to the public generally, any entitlement to assert privilege will be waived.

the Courts have upheld the privilege attaching to documents disclosed in error.

In general terms, if a document over which privilege may be asserted is inadvertently disclosed without asserting privilege over it where:

a) not to do so was a clear mistake, and

b) privilege was asserted over another copy of the document within the discovery,

the Court may not allow the opposing party to rely on the document disclosed in error. A solicitor in receipt of a document which appears to be privileged should immediately contact the solicitor for the party whose document has been disclosed to confirm the status of the document and should not read or deploy the document until its status has been confirmed, unless a large volume of such documents has been discovered to the receiving party such that it appears that a conscious decision was taken to waive privilege.

Appendix N Glossary

ESI - ESI is simply information or 'records' stored in an electronic format. This can be in any electronic format on any type of device. The converse to ESI is information which is stored in hardcopy (or paper) format.

OCR – When Scanning a hardcopy document to a searchable electronic format, a process known as Optical Character Recognition or 'OCR' is required in order to make the electronic version of the hardcopy searchable. This is not an exact science, as it is reliant on the computer recognising text. Therefore it may not result in every piece of text in the hardcopy document being recognised and made searchable in the electronic copy (this is especially so in the case of handwritten text). However, once the accuracy of the process is understood, and adequate QC procedures are put in place, then it may be possible to rely on searching such documents using electronic tools.

Coding - Electronic documents have metadata built into them. For example, an email will have the author, the recipient, and the date, along with the fact that it is an email. The technology tools used in the eDiscovery process automatically recognise these metadata fields and present them to the reviewer as such (i.e. the computer recognises an email). A scanned copy of a hardcopy document is however akin to a photocopy of the document, and as such does not contain metadata as to who wrote the document, when it was written, or who it was posted to (although this information may be in the content of the document itself). While there are some technologies available which can identify such metadata from scanned hardcopy documents, it is a difficult task, as the information required is not often in the same location (as it would be in an email, thus allowing the computer to identify and present it automatically). The solution to this is to have a human 'code' the documents. This essentially involves the document being scanned and OCR'd, and it is then passed to a 'coder', who will review the document (to the extent necessary) and take note of who sent it, who it was addressed to, and what date it was posted. This requires a great deal of human interpretation, and as such is quite a manual exercise. The output of this process is that the 'coding' information can be used to facilitate the processing and review phases. For example, it may be necessary to have one reviewer review all letters between two parties in chronological order. This 'coding' information can be used to identify all letters and then sort them by date; a task which would not have been possible with just the scanned copies of the hardcopy documents alone.

Metadata – This is generally referred to as data about data. For example, when a document was created, last accessed, last printed, etc. It can be automatically created by the application used to create the document, by the operating system used to run the computer, or it can be manually created and/or modified by the user of the computer.

During the processing phase, three primary forms of metadata are typically extracted and/or generated and stored in the eDiscovery processing system. This allows all metadata to be located in one location, preserved securely, and to be produced from this secure location at production time:

1. **Document metadata** – This is the metadata embedded within the document itself. Examples include author, sender/recipient (for emails), and last printed.
2. **Operating system metadata** – This is the metadata which the operating system of the computer used to store the document holds about the document. Examples include the date/time that the document was created on or copied to the location on the computer, when it was last accessed, the location on the computer which is was stored, and who had access from a security perspective. This type of metadata stays with the operating system on the computer and not with the document itself, so must be copied as part of the collection phase. Often a forensic collection is required to record and maintain this type of metadata.
3. **eDiscovery process metadata** – This is metadata which is generated during the eDiscovery process. For example, the custodian name associated with the document source may be recorded at the collection phase, along with the make, model and serial number of the computer which it was retrieved from. Such metadata is helpful in tracing the source of an individual document.

As such, documents produced in native format will have the document-metadata embedded within them. This metadata and some operating system metadata (such as created/accessed/modified dates) will usually be included in the production schedule. Further, some eDiscovery process generated metadata (such as custodian names) will likely also be included in the production schedule.

When metadata is collated for a document is it typically stored in an eDiscovery processing database which stores a copy of the original document alongside a record of all the metadata associated with it. Each item of metadata is stored in a 'field' within the eDiscovery processing system. This allows searches on specific metadata fields to be carried out (such as all documents with a last printed date of X or Y). At production time, it is simply a case of deciding which metadata fields are produced in the schedule alongside the copy of the native document. As the original metadata is secured and produced in the schedule, it is only important to ensure that the content of the document is preserved. If the created date were to be accidentally changed during the production process, the original would still be in the schedule (and is therefore the only one which should be relied upon).

Converting native electronic documents to near-native images or to paper usually results in the loss of metadata and is generally not recommended.

A party requesting more detailed metadata and/or a forensic copy/image of the produced document (for example in a dispute as to the authenticity of a document) should demonstrate that the relevance and materiality of the requested metadata justifies the cost and effort in producing that metadata.

Load file – Is a file which contains a schedule of documents in a format which makes it possible to easily import the schedule and its accompanying documents into an electronic document management and/or review system. The load file also typically includes the original metadata associated with each document, along with other production related information, such as categories, redactions, etc.

Recall – Measures the percentage of responsive which have been identified. This is also known as a measure of completeness. In other terms, it answers the question: Of the relevant documents in the document set, how many were found?

Note: Without actually manually reviewing every document in the set, it is not possible to accurately determine the recall. Absent a full manual review, recall can only be calculated using a statistical sample, such as that generated during the initial assessment/control/seed set.

Precision – Measures the percentage of truly relevant documents identified within the document set. This is also known as a measure of exactness. In other terms, is answers the question: Of the documents thought to be relevant (by a predictive coding system or identified by filters), how many are in fact relevant?

F-Measure – This is the harmonic mean of recall and precision and is used to measure the effectiveness of a search (either filters or predictive coding). It is used as a target to achieve a high level of recall (find all the relevant documents) whilst also achieving a high precision (minimising the amount of manual review spent on reviewing irrelevant documents). The harmonic mean is preferred over a standard arithmetic mean as it falls closer to the lower of the two quantities. This helps avoid situations where a high arithmetic mean can be achieved with a high recall, but with low precision, or a high precision, but with low recall. There is no standard for recall, precision, and f-measure when determining the cut-off point in a predictive coding project. The predictive coding expert providing the system should always be consulted on a case by case basis when determining the cut-off point as it will vary on a project by project basis.