THE GOVERNANCE OF THINGS -
eDiscovery in Australia and New Zealand

eDiscovery is the production of relevant documents that parties to litigation or an inquiry are required to produce to either a Court, Royal Commission, Commission of Inquiry or to a government regulator. The eDiscovery industry is a global industry reflecting the enormous growth in information and the specialised technology which has developed to meet that challenge. While eDiscovery is a specialist area of legal technology the challenge of document production extends well beyond in-house legal departments and requires the assistance of IT departments and records management.

eDiscovery Development

In the late 1990s, a large matter would have been hundreds of boxes of folders containing documents. Typically a Supreme or Federal Court commercial litigation matter in Austraia will have at least 1-2 million documents, where about 80% of the documents are emails or attachments, and the remainder are electronic files. While there are still matters involving hardcopy documents, this is quickly becoming the exception rather than the norm.

Year on year, since the early 2000s the size of the eDiscovery market has increased, and electronic sourced data (as opposed to hardcopy) started exceeding hardcopy in the early 2000s.

In order to manage these kinds of volumes of documents/data, an eDiscovery industry has evolved, and today the global eDiscovery software industry is valued at about USD$7 billion.

The 2016 internet minute infographic show what happens in an internet minute and highlights the massive amount of information being created each minute.

This infographic goes some way to explaining why there is an incredible proliferation of data leading to a greatly increased pool of data in litigation matters, which has resulted in the development of specialised technology as well as specialised legal technology personnel.

What volumes of data/documents are produced in litigation?

A rule of thumb is that for each person who is a potential witness (which is called a custodian) will typically have 100 emails a day, and each email will have a single attachment (so 200 documents). Assuming there are 20 working days a month, there is 4,000 documents (2,000 emails) per person per month.

A dispute of inquiring which covers a period of 2-3 years, and includes 10-15 witnesses or custodians, means the raw volume of documents will be between:

	Small	Large
Months	24	36
Custodians	10	15
Emails	**480,000** (2000 x 24 x 10)	**1,080,000** (2000 x 36 x 15)
Documents	**960,000** (4000 x 24 x 10)	**2,160,00** (4000 x 36 x 15)

Another way of measuring volume is by the file size. There can be significant variance in the size of emails and documents by industry. For example in telecommunications and media, there is a significant volume of presentations / video / images which tend to be large files, whereas a finance company would have significant volumes of spreadsheets. Another rule of thumb that we use is that an email and its attachment will be on average about half a MB (500KB) in size.

Therefore 10 GB of emails could contain about 40,000 documents and 100GB could contain 400,000 documents. Therefore in the preceding table, you could expect that there would be about 200GB and 540GB of data.

What are the steps involved in producing data/documents?

Once the relevant custodians have been identified and any other repositories of relevant information, the raw volume then needs to be reviewed, which is now done with the use of eDiscovery technology.

The eDiscovery processing tools are utilised to index and filter the documents. Typical filters that are applied include: date, custodian, deduplication, as well as keyword searches*.

Deduplication is an automated process where each document has a unique value (called an MD5), and then family groups (emails and attachments) have their unique values compared and the second instance of a duplicate is marked to be excluded.

You can generally exclude 30% of the dataset by deduplication, and up to 65-80% using combinations of the other filtering techniques. Therefore you can reduce the dataset:

	Small	Large
65%	336,000	756,000
80%	192,000	432,000

Once the basic filtering is done, further more precise filtering techniques can be carried out, including a series of processes where metadata is extracted, documents are converted to PDF**, each page of each document is digitally stamped with a unique identifier (ie a document number). This is referred to as eDiscovery processing.

*Keyword searches - keyword searches are a blunt tool, and are known to have flaws. They are either applied as inclusionary (all documents with a hit are included for further review), or as exclusionary (all documents with a hit are excluded from further review). There is a slow trend away from keyword searching as a method

of filtering, instead TAR is being used.

**PDF - most Courts in Australia require that standard documents be provided in PDF format, or for non-standard documents (spreadsheets) in native file format. At times documents will be reviewed in native format and then the responsive documents will be converted to PDF and stamped.

Legal Review

Although some systems have the eDiscovery processing and legal review integrated, typically a separate system is used for legal review. Therefore, once the processing of the data is completed, it is loaded or imported into the legal review system. The next step is sampling as well as conducting many iterations of keyword searches in order to prioritise the documents for review by lawyers (we call this legal review).

After the sampling / keyword searches, a review will be carried out on about 10-15% of the documents in the legal review system:

	Small	Large
15%	28,800	64,800
10%	19,200	43,200

In most matters, the legal review team will be making 10-15 decisions per document and includes whether or not each document is:

- relevant – that is, whether it falls within a category ordered to be discovered by the court or within the terms of a subpoena;
- subject to legal professional privilege, which means it will be included in the Discovered List of Documents, but is not actually produced for inspection by the other party or parties;
- confidential and subject to restricted inspection – such as a confidentiality undertaking by those who inspect the documents such as the other parties experts.
- a key document
- the categories or issues that the document is responsive to – the number of issues can vary from case to case and can be organised according to high level issues and then sub-categories.

Although the responsive rate can vary widely, it is not uncommon to have relevance rates of 5-10%:

	Small	Large
10%	1,920	4,320
5%	960	2,160

Latest Technology Developments - Technology Assisted Review(TAR)

An exciting development over the past 5 years, has been the addition of TAR to the legal review systems. In simple terms, a legal team is presented with small sample sets of documents that they review and code for responsiveness, and upon completion of each sample set, the TAR system conducts an audit of the coding consistency, and selects further sample sets for review. After several rounds of coding, the TAR system will reach a point where it has determined that enough documents have been reviewed that it can make predictions upon the responsiveness of the entire dataset. Typically this is a statistical table showing percentage bands of responsiveness ie 95-100% responsive, 90-95% responsive and so on. The TAR system will also show other statistical details on the precision and recall.

The next step is that the legal review team can shift priorities and focus on the document set that are predicted to be highly responsive, and/or utilise lower cost resources to review the documents that have the lower responsive rates.

Court Rules and Practice Notes

Everything that we do in the eDiscovery lifecycle is largely governed by Court Rules and Practice Notes, and when responding to Notices to Produce or Subpoenas in a Royal Commission, Inquiry or Regulatory authority in Australia, there

is an underlying assumption that the data will be handled according to the Federal Court Practice Note (GPN-TECH).

The Practice Notes provide an outline for the minimum standards for how data is exchanged between parties and the Court in a proceeding.

Each Australian state Supreme Court also has its own rules and practice notes for discovery. For example, the Victorian Supreme Court have released an updated practice note (SC Gen 5 - Technology in Civil Litigation) that commenced on 30 January 2017. What is interesting in this practice note is that it has a section (8.7-8.9 on pages 6-7) dealing with TAR, which sets out the following:

8.7 In larger cases, technology assisted review will ordinarily be an accepted method of conducting a reasonable search in accordance with the Rules of Court. It will often be an effective method of conducting discovery where there are a large number of Electronic Documents to be searched and the costs of manually searching the documents may not be reasonable and proportionate. In such cases, the Court may order discovery by technology assisted review, whether or not it is consented to by the parties.

8.8 As part of cooperating in the conduct of the proceeding, and to avoid later disputation, parties may agree on a protocol for technology assisted review as part of the discovery process and inclusion in a discovery plan.w

8.9 The protocol may include:

- The appointment of a Joint Operator; or
- The appointment by each party of a Party Operator;
- A general description of the system to be used, either by a Joint Operator or a Party Operator, in undertaking the technology assisted review process, including, but not limited to:
    - A continuous active learning protocol (using a constantly changing body of documents which are used to train the technology assisted review algorithm);
    - A simple active learning protocol (using statistical samples, including control sets, or random samples and the like);
    - A simple passive learning protocol (using other recognised statistical methods); andAny other appropriate system.
- A general description of the method to be used, including where relevant:
    - An outline of the steps to be undertaken as part of the protocol;
    - Where statistical measures are adopted for quality assurance purposes, details of the statistical measures;
    - Proposed members of the review team;
    - The management of non-text based documents;
    - The treatment of foreign language documents;
    - Any procedures proposed for high level culling and the elimination of repeated or duplicate content;
    - The method for determining the scale of relevance;
    - Any necessary manual review of the results produced by the system;
    - Any document or groups of documents a party proposes to exclude from the process; and
    - Any other matters relevant to the method adopted.
    - Arrangements for the clawback of privileged or confidential information which may have been inadvertently exchanged or disclosed as part of the technology assisted review process; and
    - Provision for the exchange of relevant documents that have not been disclosed as part of the technology assisted review process.

The revisions to both the Federal Court and Victorian Supreme Court Practice Notes make it a requirement for lawyers to be across legal technology. This is self-evident given the advances and availability of the legal technology now available, which enable the discovery and document production process in litigation and inquiries to be more efficient, accurate and cost effective notwithstanding the significant volumes of potentially relevant material.

MATTHEW GOLAB

Legal Technology Manager, Gilbert & Tobin