

**Managing**PARTNER

SPECIAL FOCUS SUPPLEMENT

# Supporting the changing litigation landscape



In association with

 <p><b>FIRST</b>Advantage</p>	<p>North America   Europe   Asia</p> <p><b>LITIGATION</b> consulting</p>
---	--

# Voyage of discovery

The route to electronic data discovery in six easy steps...

By **Reza Alexander**, **Lee Gluyas** and **Emma Hogwood**, DLA Piper UK LLP

**N**o matter what your level of experience working with electronic documents, balancing time, risk and cost in the conduct of electronic data discovery (EDD) continues to be one of the greatest challenges faced today in litigation or a regulatory investigation.

Technological advances, proliferation of electronic data, compliance regulations and current economic conditions all serve to increase both the importance and complexity of meeting this balance.

Although this balancing conundrum will never go away, it can be addressed effectively by following a robust readiness plan coupled with a number of systematic 'tried and tested' steps.

What is important to appreciate is that, although the storage media may have changed, the actual electronic discovery process need not be any different to the traditional methods used in paper disclosure.

Typically, an EDD project can be broken down into the following six distinct steps:

## Data preservation

At the first sign of anticipated litigation, a regulatory investigation or a request for production of documentation, immediate steps must be taken to preserve all relevant data and suspend all routine data-destruction policies. Steps must also be taken to ensure that the message is communicated organisation-wide – ensuring that no potentially responsive

data is destroyed or altered in any way. This is easier said than done, especially in larger organisations, but nevertheless it is important to impart, and be seen to have implemented, effectively.

A step-by-step process, guided by a structured questionnaire, should address issues including: why preservation is contemplated; what obligations may have automatically been imposed; what steps should be taken immediately to comply; which areas of the organisation have to comply; and, which key personnel, and what categories of data, may be affected.

## Data collection

For any electronic disclosure or regulatory investigation project to be effective, the data-collection process must be at the forefront of the project cycle, as it will ultimately define and determine the review strategies based on the amount and type of data gathered. The key issue is to find out what relevant data is available and where it is stored.

However, faced with a multitude of data-storage methodology, storage resources, levels of technical expertise and fluctuating IT resources to assist with the data collection, this first step can often be the most challenging.

Daunting as it may seem, the process of gathering the relevant data can be successful, as long as a systematic checklist is prepared. This can easily be completed by identifying and enlisting the assistance of key IT personnel within the organisation

whose data is being gathered, along with the acquisition or sight of the most up-to-date map of an organisation's IT network infrastructure.

The checklist should be as detailed as possible, allowing the user to obtain a clear understanding of the organisation's: IT infrastructure; applications; backup protocols; document-retention policies; mail-server data; file-server data; data about workstations; laptops; alternate forms of data storage (such as smart phones, personal home computers, memory sticks, portable hard drives, CDs and DVDs); traditional sources of data (such as hard-copy data); and, of course, the physical locations of the data.

However, no matter how similar the basic steps of data harvesting are for electronic data and hard-copy documentation, one must never underestimate the voluminous and more complex nature of electronic data. Although it is possible to enlist the assistance of your client's IT department and expertise to collect the relevant data in the majority of cases, it would generally be advisable to instruct neutral third-party electronic-evidence collection specialists to collect the data using the appropriate industry-standard procedures, especially in criminal and white-collar fraud investigations.

Above all, it is essential to maintain the integrity of the data, and ensure that a complete chain of custody for harvested data is maintained, including

a detailed log of every step undertaken to ensure an accurate and legally-defensible position.

### Data processing

Once the potentially relevant data has been identified, whether it be forensic images of hard drives, live data from mail and file servers, or even data from back-up tapes or CDs and DVDs from disparate locations, it may span Gigabytes, Terabytes or even Petabytes of data, meaning that reviewing the data may be impossible, or at best disproportionate, given the time and resources available.

Depending on how, and what type of data, may have been harvested, adhering to the structured data-culling steps outlined below should assist in reducing the harvested data by as much as 60 to 70 per cent, with consequential advantages of lower volumes of data to review, and thus lower overall costs.

1. **Eliminate operating system files.** A large percentage of files on a personal computer (sometimes up to 95 per cent) are program executable files, library files, compressed files used to install applications, and operating system files. The extraction of such operating system files is a common and widely-accepted first cut method of reducing the bulk of the collected data, and most experienced EDD vendors and processing bureaus should be able to assist and advise on this fairly routine and automated process;
2. **De-duplication.** The next most effective and popular objective method of data culling is the process of data de-duplication, which is a process where electronic algorithms are run on the data to identify exact

duplicates of documents by comparing the unique digital fingerprint of each document;

3. **Date filtering.** This step allows data to be eliminated if it falls outside responsive date ranges;
4. **Types and sources of data.** In appropriate cases – and with caution – certain documents, file servers or back-up tapes can be safely eliminated if they have originated from a non-responsive department, region or division of an organisation;
5. **Key players.** Key custodians of data should be identified at the outset, agreed with the opposing party and, if possible, their respective data prioritised and gathered at an early stage. It would equally be advisable to agree a list of the non-responsive players with the opposing party, allowing more efficient and subjective culling of the data;
6. **Key word/key term searching.** Filtering data with the aid of key word and key term searching is an effective means of eliminating data, and enables both parties to concentrate on the relevant and key information quickly. However, care must be taken to select exacting key terms, to ensure that potentially-responsive data isn't inadvertently eliminated. Active involvement of both parties in the choice of universal key words and terms must be at the forefront of every project as early as possible. Recent authorities and protocols such as the widely reported UK High Court case of *Digicel v Cable & Wireless* [2008] EWHC 2522, and internationally promoted Cooperation Proclamation by the Sedona Conference encourage a culture of greater collaboration

between the parties, which will ultimately save on time and costs; and

7. **Document types.** Filtering data by identifying, eliminating or prioritising certain document types and files is another effective and subjective means of data culling. For example, AutoCAD drawings and communications between key personnel may be the most important aspects of an engineering claim, and these, as well as e-mails, their attachments, and MS Word or equivalent document types, should be the first line of attack.

Although the above list is not an exhaustive one, these steps provide flexibility and an efficient way of reducing and prioritising the data population at the production stage.

### Data production

The ultimate goal of any disclosure or document-production project is to identify and produce documents that are relevant to the matter in hand, subject to any issues of client privilege.

Copies of the culled data can be loaded, if required, onto a review team's servers, allowing it to open and review copies of native files on their desktops, using a combination of search engines and third-party software native-file viewers. This, in effect, would be true native-file review, but it can be a slow and labourious process and only workable with small data sets.

Another approach is to instruct specialist vendors, or imaging bureaux, to process the culled data, ready for upload to an appropriate electronic document-review platform.

A further cost-effective solution would be the adoption of a software-as-a-service (SaaS) review type platform, which will not only allow

legal teams or corporations of any size to have access to the most sophisticated electronic document-processing and review solutions, but will also provide control and certainty over costs with the added ability to scale up in case the volume of data expands or becomes more complex.

These specialist organisations and services will typically extract text and full metadata from these files to populate a multitude of fields automatically, such as dates, author, recipient and document title, and provide exact electronic images of the data, with the option to keep the original native files with the data. In addition, each document is also uniquely numbered electronically.

Needless to say, it is always good practice to meet and confer with the opposing side as early and as often as possible to clarify, determine and agree the following:

- **Production format.** Paper, electronic data or both. If electronic, decide whether they are to be native files and/or images and whether to use extracted text;
- **Shared repository.** It may sometimes be advantageous for parties to share a web-based repository for all parties' disclosure documentation, which will simplify the production format;
- **Metadata fields;**
- **Production nuances.** Agree, in exacting terms, details such as the required load-file formats for each disclosing party, document numbering conventions, field names to be disclosed and their exact order; and,
- **Clarify delivery deadlines.**

However, regardless of the format in which the data is produced or

agreed to be produced, it is essential that the original native files are preserved, allowing the option to produce further copies from the unaltered file.

### Data review

The data review is one of the most time-consuming and potentially expensive processes, as it requires the dedication and attention of a team of reviewers – to review, organise, annotate and cull the data for privilege and relevance in preparation for disclosure or in response to a regulatory investigation. When faced with tight disclosure deadlines, stringent document-review techniques and quality checks become even more significant.

As daunting as this may seem, especially with the voluminous nature of electronic data, the process can be streamlined by using data-sampling techniques, targeted initial reviews on key players and the use of sophisticated search and review tools.

The most effective method is to ensure that the data is in a shared web-based or LAN-based electronic repository, which would allow all reviewers, regardless of physical location, to undertake the review process and produce relevant documents for disclosure. This would not only ensure the integrity of the original data, but will also reduce the time and resources allocated to the search and review process.

There is a growing and fast-evolving industry of intelligent (often called concept and meaning-based) searching tools, which use advanced linguistic algorithms, clustering and bayesian technologies to allow a reviewer to find, review and produce electronic documents that are conceptually related to an initial search query, legal or technical term.

When used correctly, this type of technology allows the review team to focus on the information pertinent to the case while reducing irrelevant documents and information.

### Disclosure

Once the documents are reviewed, the relevant non-privileged data must be produced in the agreed form for delivery to the opposing party or requesting authority.

At this stage, it is common for there to be some level of manual intervention for the successful processing of the data.

A decision may also have to be made to supply the disclosed set of documents with sequential pagination, showing no gaps for the removed privileged or irrelevant documents.

Finally, depending on the volume of the data in question, the disclosure set is either burnt onto CDs/DVDs, loaded onto portable hard drives or even printed to paper for delivery to the opposing party.

In summary, while the general process has not changed, the volume, complexity, means of review, and myriad available tools, storage and production methods have created new, complex and challenging obligations. Managing this complexity requires an understanding of what is an acceptable risk in relation to the time and financial resources available. Managing PARTNER

Lee Gluyas is a partner in the litigation and regulatory group of DLA Piper's London office. Emma Hogwood is a solicitor in the same group, and Reza Alexander is the litigation and practice support manager for DLA Piper UK LLP and serves as a member of DLA Piper's Electronic Discovery Readiness and Response Group.