

Navigating the duplicate electronic document minefield

By Scott Gillard, CCH Workflow Solutions

De-duplication involves identifying and removing an exact copy of a file that is identical in every way.

If one little piece of metadata is different, for example if an email was opened and saved but with no changes, it is treated as two separate documents. While traditional technology can detect exact duplicates, it will not identify documents that can be classified as near-duplicates.

Near-duplicates are files with small differences which are not identified as being an "exact duplicate" but are very close in similarity. Examples of near-duplicate files include:

- Files with a few different words - this is the most common form of near-duplicates, and the most pressing business need;
- Files with the same content but different formatting - for example, the documents might have the same text, but use different fonts, bold type or italics;
- Files with the same content but different file type - for example, Microsoft Word and PDF copies of a file.

From the technological viewpoint, the first scenario - small differences in content - is the most challenging. The ability to handle these cases can be achieved with the application of near de-duplication technology.

Near-duplicates are especially common in email, business templates, such as proposals, customer letters, and contracts,



Scott Gillard

and forms, such as purchase or travel requests.

Being able to group near-duplicates enhances the document review process by:

- The user being presented with sets of near-duplicates. Instead of the random review of individual documents, near-duplicate sets enable a systematic, coherent review process;
- The user no longer needs to read each individual document. Instead, the user reads just one document from each near-duplicate set. To review the other documents in the set, the user simply compares the small differences;
- The near-duplicate grouping also ensures the consistent treatment of near-duplicate documents.

Customers using near de-duplica-

tion technology are finding 30-50% near-duplicates in typical enterprise environments. As a result, the near de-duplication process affords significant time and cost savings in the document review process, together with enhanced review quality. Near-duplicate percentages tend to vary widely between paper-sourced and electronic file discovery situations.

Results from recent cases, indicate:

- Electronic discovery case A: 58% near-duplicates;
- Electronic discovery case B: 35% near-duplicates;
- Electronic discovery case C: 40% near-duplicates;
- Electronic discovery case D: 67% near-duplicates;
- Large government agency: 53% near-duplicates;
- OCR case: 16% near-duplicates.

Near de-duplication can provide the following unique benefits to lawyers:

Less cost:

- Enables a more efficient, systematic review process - case studies show that near-duping can reduce a firm's review costs by 40%.
- For law firms, this facilitates a more competitive, reduced cost offering to your customers.

Less time:

- Prioritised review to cover



a Wolters Kluwer business

Continued page 15

**Navigating the
duplicate electronic
document minefield...
cont.**

more data in time window.

- Focus on high value-add versus low value-add activities.

Less risk:

- Coherent assignment enhances review quality.

- Focus on differences only which reduces errors and oversights.

- Allows consistent treatment of similar documents.

Imagine being able to take the documents provided to you as part of a disclosure, compare those documents back against your own documents, see the duplication between the two and be able to compare just the small differences quickly and easily. You could say that this is the technology that helps you find the "needle in the haystack".

The use of near de-duplication technology is fast becoming part of the standard suite of tools that lawyers use to build their case. Don't let yourself be "deduped" by your opposition. There are IT experts, forensic experts and, like CCH Workflow Solutions, eDiscovery experts in the market to assist you to keep ahead of the game.



Best practices for eDiscovery Workflow Solutions

By Scott Gillard, CCH Workflow Solutions

eDiscovery is the practice of collecting, restoring and storing electronic data in order to efficiently review, filter and retrieve those data sets from a litigation ready database. Following on from the growing trend in the US, eDiscovery is set to become a standard procedure for all litigious cases moving forward in Australia.

CCH Workflow Solutions discusses the fundamentals of eDiscovery via a six step process for lawyers engaging in eDiscovery practices. This process is split into two phases.

The first phase is the scope, planning and data collection phase. This involves building a team, planning your approach and collecting all relevant electronic data that may be required for your case. The second phase includes preparation and pre-processing, transforming the data into a required format such that it can be easily handled, reviewed, managed and analysed and then produced in Court or to the relevant parties through the litigation process.

Step 1 - Scope & planning

The first and possibly most important task, something that is often forgotten, is to preserve all available data. Almost immediately, emails and electronic files must be prevented from being deleted and an understanding gained of the current back-up procedures.

The next step is to build a team. The IT manager is a valuable asset to this team, as they should have a full understanding of where everything is located. It may seem

obvious to search the file server for documents, but you should also consider everything that exists locally on PC's, laptop computers, PDA's, external drives, CD's, DVD's, mobile phones and voice-mail. These devices may contain the valuable evidence you require. Bringing in a litigation support or computer forensics experts could prove valuable at this point, as they can help you to identify where the relevant data could possibly be stored.

Step 2 - Data collection & preservation

Now that the data has been preserved and a team assembled with a clear understanding of what needs to be done, the required data should be identified. Nominate the best people available who have the knowledge and expertise to carry out this task. Drawing on all resources, find all relevant material, the type of data, where it exists and the expected volume of data. Also consider deleted information. There may have been fraudulent activity where deleted information requires retrieving, deleted data exists on hard drives and servers for months or even years after it has been deleted. Intimately knowing the volume of data that you have will ensure that you can negotiate appropriate time frames for completion of your discovery or disclosure process with the court.

Step 3 - Preparation & pre-processing

An initial analysis of the data can now be performed to determine

Continued page 16

Best practices for eDiscovery Workflow Solutions... cont.

the types of files and the volume of data that you are dealing with. This also enables the team to identify whether there is any data missing. Knowing the volume of data that you have will also enable you to make an informed decision on what you will require in the next steps of processing. Depending on your requirements, processes can be undertaken to minimize the volume of data that is required to be reviewed by the legal team.

Step 4 - Data upload, review and analysis

The resulting document set produced can now be uploaded for initial review by the legal team using a database such as Ringtail, Summation or Concordance. Depending on the type of system that is being used, the associated metadata and the documents contents are normally full text searchable allowing quick retrieval of documents based on key search terms. The main advantage to reviewing documents natively, is that it allows a document to be reviewed in the way it was intended to be viewed. Electronic

documents are three dimensional. When printed you lose that three dimensional view. For example, metadata about the document, such as "track changes" in Word documents or the formulas within spreadsheets are lost once the document is printed to another medium. When the document is viewable in its native format, this additional information is still present in the file.

Step 5 - Validation, rendering and stamping of images

At any stage in the legal review process, sets of documents can be exported from the litigation support system and passed on for further processing. The usual steps that are required at this point are Document Rendering and Image Stamping. Document rendering involves "printing" the document to a specific image format, such as TIFF or PDF. The format that is used is usually agreed between the parties prior to the commencement of the case. Image stamping involves electronically stamping a unique identifier on each page.

Step 6 - Production & presentation

Once the data has been transformed into the required format, the data can be finalised in a litigation support database to allow for 'electronic masking' or 'redaction' to remove privileged or confidential information. After the redaction process is completed, the final dataset is exported from the database and prepared in accordance with data exchange protocols or disclosure/discovery orders.

Following this six-step approach ensures the eDiscovery process is comprehensive, efficient and will result in the best outcome. Remember, don't be afraid to ask for help from experts within and beyond your organisation. There are IT experts, forensic experts and, like CCH Workflow Solutions, eDiscovery experts in the market to draw on.

This feature is a summary of CCH Workflow Solutions 'Best Practices for eDiscovery Workflow & Management Whitepaper'. To download a free complete & comprehensive copy of go to www.cchworkflow.com.au.

CCH Workflow Solutions First to Market with Discovery Cracker in Asia Pacific

By CCH

Sydney, 13 September, 2007 - CCH Workflow Solutions, a Wolters Kluwer business and leading provider of litigation support services, is pleased to be the first in Asia Pacific to incorporate leading eDiscovery software 'Discovery Cracker' into its eDiscovery suite of services.

Discovery Cracker is an electronic discovery software program used to process a wide variety of electronic data for matter-based and systemic eDiscovery. A vast feature set; built-in software fault tolerances and a high demand data processing architecture make Discovery Cracker the most well

regarded eDiscovery software available today.

CCH Workflow Solutions parent company Wolters Kluwer recently announced the acquisition of the Discovery Cracker product line from DocuLex to complement its growing number of CT litiga-

tion support solutions. Discovery Cracker was the first eDiscovery solution on the market in the US and continues to be the leading software building on its years of development and growth in the legal industry.

"The legal industry in Australia is currently under tremendous pressure to reduce the time and cost associated with litigation, specifically the discovery process. Discovery Cracker provides us with an extremely secure online data preview tool giving us increased visibility into the electronic discovery process. This means real time access to extracted data, allowing CCH eDiscovery experts to filter, view natively and tag individual files for processing faster and more efficiently than

ever before." said Gary Kendrick, Managing Director of CCH Workflow Solutions.

"We are excited about the integration of Discovery Cracker into our suite of eDiscovery services. We know our clients will be extremely happy with the added functionality Discovery Cracker now provides them. It is what they have been asking for".

Key benefits of Discovery Cracker:

- High volume true distributed processing;
- Highly scaleable;
- Processes hundreds of file types;
- Embedded files processing;

- Robust Quality Control module;

- Automatic blank page detection;

- Browser based native file searching/tagging.

For more information on the benefits of Discovery Cracker as part of CCH Workflow Solutions eDiscovery services please phone 1300 137 225 or visit www.cchworkflow.com.au.

Contact

Kyla Aldrich

CCH Workflow Solutions, Level 10, 190 George Street, Sydney

Tel: +61 2 8226 4999 Fax: +61 2 8226 4900 Email: kaldrich@cch.com.au

Making bequests easy

By Include a Charity

Research shows many Australians wish to leave a bequest in their Will but are unsure how to start.

To assist discussing the subject with clients, Include a Charity is offering lawyers FREE bequest information packs for display in offices. These packs can be ordered by calling 1800 400 300.

Being a primary point of contact for those preparing Wills, Lawyers play a

vital role in educating and informing about bequest options. It is estimated that around eight percent of the population make charitable bequests each year, equating to around 10,560 people per year. The Include a Charity consortium, comprising of Australian Red Cross, Cancer Council, Heart Foundation and Mission Australia, is keen to see more bequests left overall, but are not prescriptive as to who

bequests should be left to or how big the bequest should be.

For access to information regarding leaving bequests in Wills, specific information about the supporting charities, or general enquires please call 1800 400 300, visit www.includeacharity.com.au or email info@includeacharity.com.au.

Law Society CPD Calendar

Monday 12 Nov and Monday 19 Nov	CPD seminar - Occupational Health and Safety Law Council of Australia seminar 5.30pm - 7.00pm each day. Teleconferenced to Darwin, Katherine and Alice Springs. For a registration form and further information, please visit http://www.lawsocnt.asn.au/fmi/xsl/lstnt/doc_download.xsl?docID=2005134 .
Tuesday 13 Nov	CPD Workshop - Risk Management - Taming Technology 4.00pm-6.00pm DARWIN ONLY Presenter: Jane LeMessurier, LeMessurier Harrington. Two (2) CPD points. Visit http://www.lawsocnt.asn.au/fmi/xsl/lstnt/doc_download.xsl?docID=2005113 for further information.
Tuesday 20 Nov	CPD - A Civil Lawyer's guide to the Criminal Property Forfeiture Act NT 5.30-6.30pm at Centrepoint Business Centre. Presenter: Lex Silvester, Barrister. One (1) CPD point.
Wednesday 21 Nov	CPD - Cost Provision Follow-up. Presenter: Barbara Bradshaw. Further details to follow.
Tuesday 27 Nov	CPD - Duties of Solicitors and Counsel when making ex parte Applications 5.30-6.30pm at Centrepoint Business Centre. Presenter: Lex Silvester, Barrister. One (1) CPD point. Visit http://www.lawsocnt.asn.au/fmi/xsl/lstnt/doc_download.xsl?docID=2005115 for further information.