

Budgeting for eDiscovery: Cost and Containment Considerations



When computer engineer Ray Tomlinson sent a text message to himself between two adjacent computers in 1971 the stage was set for an electronic revolution that would change the way the world communicates in both business and personal matters. Today it has been estimated by various sources that anywhere between 175 to 300 billion emails are sent on a daily basis. This has been one of the contributing factors over the last decade to legal teams facing complex challenges addressing electronically stored information (ESI) during the discovery process. Difficulties ranging from identifying sources of relevant information which may or may not be readily apparent or accessible, to the often daunting task of reviewing potentially millions of pages of electronic files, continue to be a factor in the escalating costs of eDiscovery.

By Kathleen Skapik* and Susan Rogers*

The 2006 amendments to the Federal Rules of Civil Procedure (FRCP) have seemingly helped to reduce some of the cost burden through required meet and confer planning conferences. And although technology is now in place to reduce duplicative and irrelevant files and enhance the speed of the review process, numerous cost challenges remain. In particular, how to accurately budget for eDiscovery and how to contain costs based on a myriad of “unknowns” and a seemingly exhaustive number of types of costs.

Types of Costs

The costs of eDiscovery can generally be classified as sanctions, lost opportunity, soft and hard costs. Although the cost components in each of the categories may overlap, general definitions for each can be established. Sanction costs, such as those levied by the courts for obvious or unintentional neglect of the “duty to preserve” and lost opportunity costs, including cost penalties for taking resources away from “business as usual” activities are all but impossible to quantify. Soft costs can range from efforts spent preparing for potential litigations, including the creation and implementation of effective eDiscovery record management and litigation hold policies and proactively documenting technology infrastructures, to identifying sources of electronic files of potential value. Some categories of soft costs, such as purchases of integrated software systems for case management and litigation holds, border the definition of hard costs.

Hard costs are distinguished by more definable parameters inherent to most eDiscovery, such as the costs of forensically collecting electronic data from multiple custodians. Hard costs typically involve expenditures for resources and services outside of a company’s internal operation. Hard costs are visible and definable, often simply because an invoice needs to be paid. They are not necessarily easy to budget but are of mandatory importance to contain.

Hard Cost Budgeting: Collection and Processing

Hard cost budgetary components are usually those that have a high-price ticket potential. Traditionally among these are the costs to collect and process electronically stored materials for review and production. Unlike hard copy files, the nature of electronic files makes accurate budgeting in these broad line items difficult, and almost impossible without underlying assumptions that may be more “guesstimates” than estimates. Electronic files can be identified from vari-

ous sources, exist in a variety of file types and formats and can be inaccessible or useless when separated from the system in which they were created. They are also dynamic in nature and can be easily altered (including hidden information, such as metadata that could be of substantial importance) and are greater in volume than hard copy files.

Sources can be domestic, as well as global for U.S. corporations with an international presence and may include multiple custodian hard drives, network servers, back-up tapes, Internet-based collections, PDAs, flash drives and home office PCs among others. Types of electronic files include emails, text messages, end-user files (e.g. spreadsheets, word processing documents, data files), application files, phone messages and audio/video recordings to name only a few. Electronic files are also moved among file types, such as voice messages that become emails. Sources of eDiscovery are no longer limited to a physical work environment. With virtual employees comes the need to address home offices that are supported by company provided or employee-owned equipment and devices.

Formats vary according to the system on which the files were created, such as Microsoft Exchange/Outlook and IBM Lotus Notes for emails. Accessibility can be hampered by source, such as backup tapes that may need to be restored to read, by type, including deleted, lost and password protected files and by format, such as files created by customized programs.

Volume depends on availability and accessibility as well as file type. Some file types result in more data to consider, such as spreadsheets or voice/video files that make spreadsheet volumes look small in comparison. Content considerations include file types that have little or no substantive value, for example system files, duplicates and files of little or no relevancy to the subject coverage. Although files are in electronic form it does not mean that they are readily available, immediately accessible, or contain content of value. All of the influencing factors noted above need to be considered in any budgeting for collecting and processing electronic files.

Unfortunately, trying to qualify these influencing factors is complicated by a lot of “unknowns” at the start of the eDiscovery process. Although the names of custodians who have pertinent files may be known, the volume of information, as well as file sources beyond evident hard drives and network servers, may be close to a mystery. Unlike hard copy collections which are visibly seen and can be physically counted for estimating purposes, few can envision what a “byte” of information looks like or how much volume it equates to.

Hard costs are distinguished by more definable parameters inherent to most eDiscovery, such as the costs of forensically collecting electronic data from multiple custodians.

Estimating the volume of electronic files that need to be collected and processed is not easy and is never accurate until the processes are essentially complete. For cost estimating and budgetary purposes most industry experts use a paper equation of 60,000 to 75,000 pages per gigabyte (GB) or approximately 30 boxes and a Terabyte equates to over a 100 million pages. It has also been conservatively estimated that a person generates one gigabyte of emails per year.

With numerous electronic file types and formats, some form of additional processing is needed for standardized retrieval and legal team review. Cost differentials linked to desired delivery formats, such as native files and TIFF/PDFs should also be determined as part of the budgetary process.

At a minimum, the following information should be compiled before an initial budget for collecting and processing electronic files can be prepared.

- Custodian list.
- Sources and location of files.
- Email and network system descriptions.
- Dominant file types and formats.
- Estimated volume.
- Desired delivery format (e.g., native file, TIFF/PDF).

Budgeting of eDiscovery simultaneously mandates cost containment. With multiple sources and increased volumes, limiting the scope and universe of files to be addressed in a legally defensible manner can significantly reduce costs.

Cost Containment

Cost containment starts with an overall eDiscovery Management Plan collaboratively developed by corporate counsel, corporate and law-firm litigation support and technical staffs, outside legal counsel and any needed third party resources. Additionally experienced eDiscovery experts can provide guidance regarding cost considerations and estimating guidelines, as well as offer valuable cost saving suggestions. The plan should also include issues to be discussed at the meet and confer that encourage cost savings. Using technology to reduce or limit expenditures is vital to cost reduction and should be a major portion of any eDiscovery Management Plan. Examples of best practice strategies that have been successfully shown to save and contain costs of ESI collection and processing include the following.

Meet and Confer Agreements

Amendment 26 (f) to the FRCP adds the requirement that all parties discuss the discovery and production of ESI during a discovery planning conference. Several decisions that may be mutually agreed to during this conference may result in cost containment.

- Limits on the discovery scope using criteria, such as date ranges and keywords.
- Restrictions on custodian lists, for example key players only.
- Priority of addressing custodians and other staging matters.
- Agreements on processing files that may not be not “reasonably accessible”.
- Strategy for handling backup tapes.
- Use of production formats that are less cost costly to implement.

Using outside resources, including seasoned eDiscovery experts, to prepare for a meet and confer conference may also provide significant strategic and cost saving advantages if their input is included in the early stages of developing the eDiscovery Management Plan.

Outside Experts

Qualified and experienced computer forensic and eDiscovery specialists provide more than data collection and processing. Unlike most information technology departments that focus on keeping end-users productive and systems operational, computer forensics experts are investigators. Using special collection tools and recovery techniques such experts can often restore data that may be initially seen as lost, deleted, or hidden, either by accident or deliberately. They can help reduce the risk of spoliation and sanctions by using sound preservation techniques and chain of custody data authenticating. If need be they can also prepare affidavits and provide expert witness testimony that the data was preserved and collected in a legally defensible manner.

eDiscovery specialists working with the IT department and legal teams can develop collection and processing methodology that are cost effective, time efficient and also legally defensible. They can assist the eDiscovery management team in identifying key custodians and where files of potential value may reside, including those being worked on at home and on hand-held devices. They can design processing approaches that maximize cost and time savings in con-

With multiple sources and increased volumes, limiting the scope and universe of files to be addressed in a legally defensible manner can significantly reduce costs.

Qualified outside experts maximize the use of technology to meet clients' collection and processing needs and proactively enhance their service offerings as new technology is unveiled.

verting files to a standardized format for legal team review. Qualified providers have the technology and up front resources to remove system and application files that are not candidates for further processing and review, as well as identify and remove duplicates that if left as is would incur redundant time and costs to process and review. They also provide the ability to further remove files of little or no relevancy based on date ranges, keywords, or other client specified criteria. Some also offer near-duplicate detection services that can further expedite the review process by grouping documents of similar content.

Most full-service eDiscovery providers offer a web-hosted review tool that that can be used by the legal teams for enhanced savings and efficiency and can provide the data in various formats for use on in-house retrieval systems. Qualified eDiscovery providers have experienced project managers that approach the processing phase from a litigation perspective, along with extensive technical resources to meet capacity and scheduling requirements. Such providers can also suggest additional cost savings and share best practice strategies based on experience working on multiple projects with multiple clients and handling enormous volumes and types of electronic files. Although it is almost impossible to accurately predict how much a processing project might cost until it is implemented, experienced eDiscovery experts can provide fairly realistic estimates based on similar scope of work parameters and collection descriptions. Such outsourced services add another piece to the intricate eDiscovery budget.

Qualified outside experts maximize the use of technology to meet clients' collection and processing needs and proactively enhance their service offerings as new technology is unveiled. Such third party expertise is often perceived as being more objective than in-house resources conducting similar activities.

Technology

A third major consideration is using technology to contain costs. Technology supports cost containment of eDiscovery from preservation, collection and processing to review and production. Technology can be used in cost reduction techniques for handling backup tapes and various technologies applied to reduce initial collections to unique, highly relevant documents.

Backup Tapes

Backup tapes most often contain copies of archived electronic files that no longer warrant immediate accessibility, or contain active electronic data that has

been periodically copied and maintained for disaster recovery purposes. Backup tapes use data compression to increase storage capacity. Data is usually stored in a non-standard format that cannot be directly read without special restoration processing, which can be extremely expensive. Duplication is also inherently high among the large number of recovery backup tapes that may exist due to frequent creation schedules such as daily, weekly, or monthly. Backup tapes need to be preserved, but the likelihood that they will be used is less than most anticipate. The majority of the information that meets production requests comes from active files.

Over the last several years data on backup tapes has been requested more frequently simply because they contain a lot of data. Although some technology is available to index backup tapes for content analysis before restoration, resulting inventories are typically limited. With the changes in the FRCP, backup tapes are increasingly being called in question as "not readily accessible" due to the cost burdens to restore. This strategy has seen success, but widespread elimination of backup tapes for eDiscovery has not. If accessible backup tapes need to be included in the discovery net, a cost containment strategy that is gaining acceptance is using known information and samples to statistically extrapolate what type of tapes will yield an acceptable percentage of responsive data. This sampling strategy is being more frequently addressed during meet and confers as an effective cost containment measure.

Volume Reductions

As eDiscovery budgets are extremely volume sensitive, attention needs to be given to identifying and using best practice strategies and proven technology that can reduce what is collected to the non-duplicative, highly relevant review candidates. Any such tactics must be implemented with assurances of collection integrity and defensibility.

There are several demonstrated and commonly accepted technological approaches to reducing initial collection volumes, a few of which are as follows.

- Removing system and other applications that have no content value and are not candidates for further processing.
- Eliminating other files types, often by file extension, that would provide no value if processed.
- Identifying and removing duplicates either within and/or among custodians (i.e., de-duping).

- Culling data content down to candidates that meet client defined objective and subjective criteria, such as date ranges and keyword filters (i.e., filtering).

De-duping is often performed using a comparison of a digital fingerprint, often referred to as a hash value, which represents the binary content of a file and is unique to every file. Using keyword searches and other filters, such as date ranges, is one of the single most important developments in electronic discovery processing in recent years. Mutually agreed to concepts and keywords to limit production and review are becoming a major point of discussion during meet and confer conferences. Depending on the type of ESI being processed, de-duplication and filtering has been known to reduce electronic document collections by as much as 85%. Resulting cost savings of removing these files from further processing and review is exponentially high.

There are two new technologies that are gaining popularity in reducing volumes for processing and review.

- Concept searching.
- Near-duplicate detection.

Concept (or cognition) searching uses natural language and the meaning of words to perform semantic, pattern, statistical and Boolean searches against the full-text of files in order to further cull the population down to documents of the highest interest. In addition, significant amounts of electronic files may also be closely related, but not actual duplicates. Near-duplicate detection allows these similar files to be grouped together for further culling or more efficient document review.

All of the above volume reduction measures which result in cost savings should be implemented during a pre-processing phase before additional resources and money are spent on converting and reviewing duplicative and unwanted files. Using technology to reduce time and costs associated primarily with hard and sanction costs is a major component of any eDiscovery Management Plan.

Budgeting Tips

The challenges of creating a realistic and comprehensive eDiscovery budget with built-in cost containment measures that can be tracked and monitored against actual experience still remain. Following are several overall budget tips to lessen these challenges.

- Concentrate on budgeting hard costs which usually involve outside expenditures that may be considerable.
- Organize an eDiscovery management team to develop an eDiscovery Management Plan (including budget); include, as appropriate, representatives from corporate counsel, outside counsel, in-house and law firm litigation support, technical, and records management staffs, as well as outside experts.
- Involve eDiscovery experts in the budget planning process as early as possible; they can provide fairly realistic cost estimates based on a defined scope of work, as well as suggest cost saving best practices and methodologies to reduce budgetary commitments.
- Develop at least a preliminary budget that takes into account estimated volumes and anticipated scopes of work; eDiscovery budgets can potentially exceed the monetary values at stake in a matter.
- Understand the nature of electronic files and where and what cost containment measures can be employed to reduce the amount of money needed.
- Develop strategies to reduce costs that can be discussed during meet and confer conferences, including backup tape processing and keyword filtering.

Although the electronic data revolution has less of an environmental impact it has created a cascade of challenges to legal teams, including but not limited to the difficulties in budgeting for eDiscovery. The areas discussed provide some information, organization and guidelines to lessen this challenge, as well as set forth cost containment considerations that can impact the monetary resources that may ultimately be required.

*Kathleen Skapik is a Regional Director for TechLaw Solutions. She has over 28 years of experience in litigation support, including 12 years directing litigation support operations. She has worked with numerous Amlaw 100 and Fortune 500 companies to devise strategies for the organization, management, and processing of electronic and hard copy files for litigation preparedness and as evidentiary materials.

*Susan Rogers is a Regional Director for TechLaw Solutions with nearly 24 years of experience in litigation support. A former Manager of Litigation Support Systems at a large law firm, Ms. Rogers has implemented document repositories and participated in the legally defensible collection and harvesting of client eDiscovery data.

Using technology to reduce time and costs associated primarily with hard and sanction costs is a major component of any eDiscovery Management Plan.
