

## e-discovery: new challenges, new strategies

Two-thirds of large companies and almost as many small ones have faced costly legal discovery requests. Is your organization prepared for the new world of e-discovery?



## E-Discovery: New Challenges, New Strategies

### Table of Contents

Introduction.....	1
The New World of Litigation.....	2
What the Revised FRCP Requires.....	3
Going Backward to Go Forward.....	6
Of Precedents and Emerging e-Discovery Solutions.....	8
Technologies of e-Discovery: The Short List.....	9
e-Discovery and Data Retention Best Practices.....	13
ThinkerNet Articles:	
> Data Retention: Why Internet Companies Need to Develop a Strategy.....	16
> Why Companies Are Losing the Battle to Control Email.....	19
Notes.....	21

As more of the information we use to do business is created and kept in electronic form, we bump into unexpected (electronic) consequences. If your company should ever be subjected to a court order to ante up all the emails and other electronic data relevant to a pesky lawsuit, e-discovery (electronic discovery) will be critical to producing this information.

E-discovery refers to searching, locating, and securing relevant evidence in a court case that resides in electronic form. It includes all types of electronic files, including Web pages and email correspondence, as well as database, word processing, and spreadsheet files.

At first glance, for the uninitiated, this might seem like no big deal. Until you get a look at the size of the email archive, not to mention the assorted corporate databases in which pertinent data may be lurking, and try to figure out how to extract that "responsive data" – *all of it* – in time to meet a judge's looming deadline. Imagine, for instance, being ordered to produce 1,000 emails in 30 days – culled from an average email load of 3 million messages *per day*.<sup>(1)</sup>

This is when you'll start hearing the phrase "data retention policy" in your sleep.

Let's begin with archiving information.

According to Enterprise Strategy Group, private-sector archive capacity will reach 27,000 petabytes — that's 27 million gigabytes — by 2010. Much of this archive space is needed to house burgeoning volumes of email.<sup>(2)</sup>

If you think finding a way to cost-effectively store all that data is a problem, imagine for a moment the challenge of having to pull out a very narrowly defined subset of it — fast and without missing anything relevant. This is what e-discovery is all about.

## The New World of Litigation

Changes to the Federal Rules of Civil Procedure (FRCP) in December 2006, which was the first time electronically stored information (ESI) was addressed, made e-discovery the law of the land. It also has made courts much more demanding about how organizations handle information during litigation and imposes potentially costly penalties on those whose efforts are deemed inadequate.

Although the changes to FRCP are U.S.-specific, they have an impact on any organization facing legal action in the United States. In addition, judiciaries in other nations are addressing e-discovery; the most notable of these is Part 31 of the Civil Procedure Rules in the U.K.

## What the Revised FRCP Requires

Have you had your Morgan Stanley moment?

In 2005 litigation, Morgan Stanley failed to produce relevant data within an adequate timeframe, provoking the judge to issue adverse inference instructions to the jury (i.e., assume relevant data was *intentionally* not disclosed rather than missed accidentally). Morgan Stanley lost the case — and \$1.45 billion.

This case, and others like it, foreshadowed a future that's now upon us.

In December 2006, the Judicial Conference of the U.S. amended the Federal Rules of Civil Procedure (FRCP), clarifying the discovery obligations of parties to any litigation and, for the first time, making specific reference to electronically stored information (ESI), notably email and instant messages. The upshot: Any organization facing litigation in U.S. federal courts must be able to respond in timely fashion to an e-discovery request.

One amendment — 26(f) — calls for parties to meet and confer on the disclosure of electronically stored information and agree to the form in which it will be produced. Rule 16(b) requires that, within 99 days of a lawsuit's initiation, legal teams report to opposing counsel about who owns data relevant to the discovery request, where that data resides, and how much data there is.

Amendment 37(f) creates a safe harbor: Sanctions cannot be imposed for failure to provide information when this is a result of the *consistent* application of data retention rules that include appropriate destruction of unneeded data. Inconsistent application of data retention policies obviates protection under this amendment.

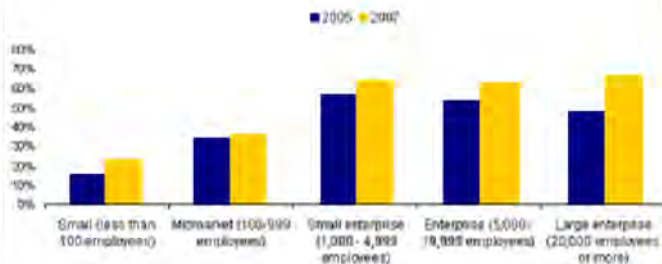


Not surprisingly, e-discovery activity is on the rise across the business spectrum, but it's an expensive procedure. Having even a junior attorney review data to determine its relevance to the litigation at hand can exceed \$30,000 for a gigabyte of data.<sup>(3)</sup>

### Increase in e-Discovery Activity 2005-2007, By Organization Size

*Mousover chart for larger view*

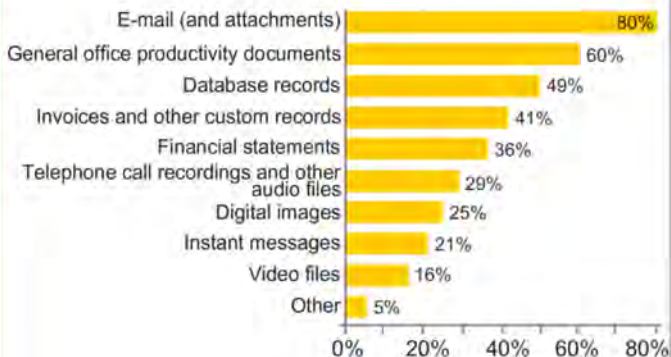
To your knowledge, has your organization been involved in a legal proceeding or regulatory inquiry that necessitated a search for and/or retrieval of electronic records?  
(Percent of respondents)



Source: *The Enterprise Strategy Group, Inc., 2007*

### Most Frequently Requested Record Types

To the best of your knowledge, which of the following record types has your organization been asked to produce in a legal proceeding or regulatory inquiry? (Percent of respondents, N = 107, multiple responses accepted)



Source: *The Enterprise Strategy Group, Inc., 2007*

To reduce these reviewing costs, organizations involved in e-discovery must find ways to limit the amount of data that needs to be reviewed and boost review process efficiency. Such goals can be achieved only by:

- > **Creating and implementing an organization-wide data retention policy** that takes into account all regulations that your organization must comply with, such as Sarbanes-Oxley and HIPAA (the Health Insurance Portability and Accountability Act, which includes data retention requirements).
- > **Using the right technology tools** to (1) store and secure data according to its place in your organization's data lifecycle; and (2) retrieve only the data you really need on demand.

### Going Backward to Go Forward

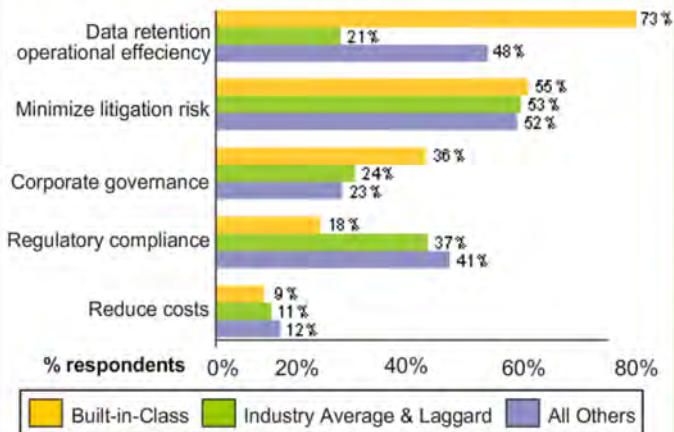
Retrieving the right information quickly from huge and diverse data stores is hard work – and beyond the reach of standard keyword searches. Confounding factors include the ambiguity of language (does "diamond" refer to baseball or jewelry?); lack of sufficient metadata; errors (human and otherwise); and new media (blogs, podcasts, .wav, .mpg). What's more, much of this data is exact or near-exact duplication – as much as 60 percent by some accounts.<sup>(3)</sup>

"If organizations want to cut costs, they should move backwards," says Brian Babineau, senior analyst at Enterprise Strategy Group, an industry analyst firm specializing in storage and information management. "The first step should be to invest in technology that makes it easier to identify responsive data – such as search solutions (for data in place), online archives (no more going to tape), and other data retention solutions. Next, using solutions that help cull data before it's sent to a legal service provider can cut costs as well."

Barclay Blair, director of Kahn Consulting's IT compliance practice, believes organizations need:

- > Better information management programs and "source mapping" to help in understanding where to look;
- > Better compliance with retention and disposition programs so that business information is not unnecessarily retained in the first place;
- > Automated review tools to speed classification and review of content; and
- > Matter management systems to facilitate the management of content subject to multiple matters.

### Business Drivers for e-Discovery and Message



Source: Aberdeen Group, December 2007

### Of Precedents and Emerging e-Discovery Solutions

Rules of e-discovery are changing fast, since they're based on case law as well as on the FRCP. Recent cases have focused on ensuring the parties involved discuss keyword searches and devise a "search protocol." June 2007 saw the first case in which parties turned to conceptual search as an alternative search methodology.<sup>(4)</sup> This is significant in that it signals the judiciary's embrace of new technologies.

### Technologies of e-Discovery: The Short List

Analysts at Gartner Inc. have commented<sup>(5)</sup> that the ultimate answer to e-discovery challenges lies in achieving full proactive control over unstructured data. This, Gartner believes, will take between five and 10 years at most enterprises. Here's a glimpse of some of the technology solutions that are taking us toward that full proactive control over unstructured data:

#### Auto-categorization

Software that assigns file content into subject matter categories using a variety of techniques beyond the pre-existing category sets and seeding categories with keywords. These include clustering of sets of documents based on similarities; statistical Bayesian analysis of the patterns of words in the document; vector programs that represent every word and its frequency with a vector; neural networks; and more complex linguistic inferences.

#### Conceptual Search

Software able to retrieve relevant information without requiring the occurrence of the search terms in the retrieved documents so that information about a topic can be located by understanding what words mean in a given context.

*Statistical conceptual search* analyzes how terms are used within a document collection to be searched to determine underlying language structure and generally don't require any pre-constructed language models. *Linguistic conceptual search* methods, such as natural language processing and syntactic



approaches, are based on language models that must be created and maintained by people.

### ***Content Analysis***

Software designed to extract details from data to determine relationships and/or derive meaning. Because content analysis tools can also be serialized to deliver focused answers to complex questions, they are useful for e-discovery.

### ***Data de-Duplication***

This software eliminates redundant data by retaining one unique instance of data in storage media and replacing redundant instances with a pointer to the unique entry.

### ***E-discovery Support Solutions***

A diverse collection of products and services that identify, preserve, collect, process, review, produce, and manage electronically stored information in support of litigation requirements. These solutions use a variety of tools – including email archiving, indexing, repository management, electronic content management (ECM), content monitoring and filtering, and information access/search – adapted to litigation needs.

### ***Email Archiving***

This is a systematic, centralized, and policy-based approach to saving and protecting email message data so it can be accessed quickly at a later date. Capabilities usually include indexing and search capabilities, access logs, and lifecycle management

(enabling creation and imposition of policies concerning when email should be archived, where it should be stored, and when it should be deleted).

### ***The Electronic Discovery Reference Model***

An open standard created in 2005 to address the lack of standards and guidelines in electronic discovery. The completed reference model, now based on XML, provides a common, flexible, and extensible framework for the development, selection, evaluation, and use of electronic discovery products and services.

“I just came from the LegalTech tradeshow,” says Brian Babineau, senior analyst, Enterprise Strategy Group, “and I’ve never seen more momentum behind a set of standards in my career as analyst.”

### ***Visual Analytics***

This includes software that graphically presents information in terms of an analytic reasoning process so that relevant data can be more quickly recognized by human reviewers. These solutions can, for instance, graphically depict relationships between emails (like message threads) and files (e.g., documents created by the same user with the same four keywords).

Indeed, the right technology tools can provide key e-discovery efficiencies, including the ability to:

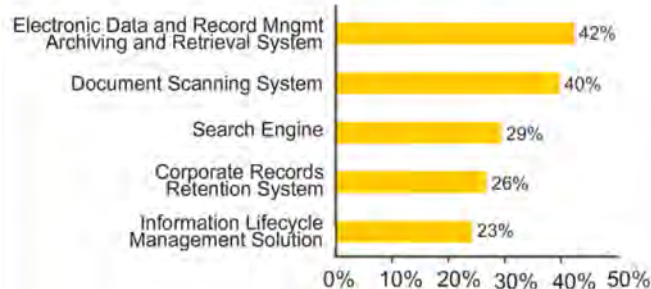
- > Maintain and use metadata about electronically stored information and the systems where they reside;
- > Cut down on the enormous volumes of data that must be reviewed for potential relevance by being able to eliminate any information contained in certain data stores;
- > Quickly analyze contents of multiple, disparate data sources that might contain responsive information;
- > Support decisions about relevance of data to an e-discovery request with intelligent sampling to understand the content of various data stores; and
- > "Crawl" intelligently through data stores to extract relevant information.

Among the most important of these capabilities are automated search techniques. Enterprise Strategy Group's Babineau sees several methods as important to e-discovery.

"Auto-categorization solutions like *Attivio*, [an active intelligent engine] that automatically groups large data sets by nouns, can help attorneys navigate a large corpus of data," Babineau notes. "In addition, conceptual search solutions that enable similar

types of groupings are emerging, along with visualization products – *Attenex* and *Clearwell* are examples. Still developing, but showing promise, are natural language processing solutions that help decipher different intent – by looking at positive versus negative statements – and terminology, such as knowing the difference between 'jeans' and 'genes.' "

### Top Five Implemented Technology Enablers



Source: Aberdeen Group, 2007

### e-Discovery and Data Retention Best Practices

"Information retention is about knowing what you have, what it is, where it is, and how long you should keep it," observes Blair of Kahn Consulting. "Quite simply, if you can answer these four questions, then doing business gets easier. So does responding to discovery requests."



Babineau offers some specifics:

- > Have a policy that extends to all content types – file, email, databases;
- > Remove duplicates from all repositories to reduce storage requirements, as well as the risk that a file or message may linger even though it's thought to be deleted; and
- > Develop a data deletion policy, making sure it meets all regulatory requirements – if you don't need to keep information for business, regulatory, or legal preservation reasons, then why not delete it?

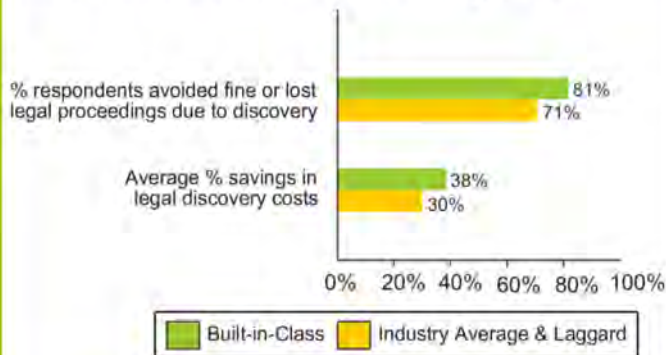
What is most important, asserts Ralph Rodriguez, senior vice president of research for AberdeenGroup's technology markets group, "is having a holistic plan with key cross-functional stakeholders from the CIO, CSO, and General Counsel's office. Everyone must agree on what's defined as success, know the plans, and execute to them."

Doing this pays off, according to Rodriguez's research. In its December 2007 report, *e-Discovery and Message Archiving*, Aberdeen identifies the actions of those organizations that have avoided fines or losing legal proceedings due to e-discovery problems, and have been most successful in reducing e-discovery costs. These firms tend to have:

- > A documented e-discovery and message archiving strategy, a designated department responsible for it, and quarterly e-discovery and message archiving audits;

- > A regular compliance review process in place;
- > A documented process and audit trail for data/records management; and
- > Several key technologies in place, including archiving and retrieval systems, policy-based automated tiered storage systems, document scanning systems, and automated data indexing and classification.

#### Business Benefits of Successful e-Discovery Practices



Source: Aberdeen Group, December 2007

"We're at an important time in the history of information management," says Kahn Consulting's Blair. "The sophisticated tools that are being developed to slay the e-discovery beast will increasingly find their way into the day-to-day management of business information."



## Notes:

- (1) *Surviving e-discovery*, 07 Sep 2007
- (2) *27 billion gigabytes to be archived by 2010*, Computerworld December 31, 2007
- (3) *Best Practices for E-Discovery*, KMWorld, February 2008
- (4) *E-Discovery and E-Recordkeeping: The Litigation Risk Factor As A Driver of Institutional Change*, National Science Foundation, July 2007
- (5) *MarketScope for E-Discovery and Litigation Support Vendors*, 2007, Gartner, December 2007