capax | discovery

The 10 Steps to High Performance Ediscovery

By Steven Williams, Executive VP &
Allen Gurney, Sr. Director

The 10 Steps to High Performance Ediscovery

By Steven Williams, Executive VP and Allen Gurney, Sr. Director
Capax Discovery

TABLE OF CONTENTS

INTRODUCTION

In 2015, an estimated 560 million emails were sent EVERY MINUTE.[1] The World Economic Forum has classified data as a new asset class, and further maintains that personal data is becoming "the new oil." IDC predicts that by 2020, business transactions on the internet will reach 450 billion each day.[2]

> " There were 5 Exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days.
> *- Eric Schmidt, Google CEO*

With such an explosion of data, wading through vast volumes of digital information poses an ever-increasing challenge for electronic discovery (also called ediscovery). Ediscovery is the process of identifying, preserving, collecting, processing, searching, reviewing, and producing Electronically Stored Information (ESI) that may be relevant to a civil, criminal, or regulatory matter.[3]

These vast data volumes compound the struggles many organizations face with ediscovery. For many organizations, ediscovery is an expensive proposition. Estimates vary but ediscovery can cost more than a quarter million USD for a medium-sized matter involving 10 custodians.[4] Due to a lack of technology and/or technical resources needed to search and uncover information critical to litigation, many organizations struggle with finding the metaphorical "needle in a haystack." Compounding this problem, often organizations don't even know which haystacks to search, given the thousands of places data lives within organizations.

As a result of these and other challenges, ediscovery is often plagued with inefficiencies and inaccuracies, significantly increasing corporate risk and cost.

This whitepaper examines the ediscovery process and how to achieve high performance discovery, including:

- Proactive information governance approaches and technology
- Defensible disposition of dark data
- Enterprise-wide data management and archiving
- Machine learning and automated search technologies

LEGAL DISCOVERY

In the U.S. and many countries around the world, legal discovery is the process in which parties in a lawsuit exchange evidence. In the U.S., non-criminal litigation discovery is governed by the Federal Rules of Civil Procedure (FRCP). The FRCP were originally written to accommodate the sharing of paper records. In recent decades, however, these rules have been amended to accommodate the unique challenges of electronic data. In addition to the FRCP, State courts have their own rules. Both State and Federal rules are constantly evolving as judges interpret these rules creating case law. Thus, the "rules of the road" for discovery and ediscovery are constantly evolving, are not purely black and white, and vary across different jurisdictions, judges, and types of cases.

THE ELECTRONIC DISCOVERY PROCESS

Ten years ago, the Electronic Discovery Reference Model (EDRM)[5] was developed to assist practitioners of ediscovery with a conceptual best practice model of the electronic discovery process.   This framework is not linear and accommodates many of the process variabilities encountered in ediscovery.

The left half of the EDRM (left side) are activities typically managed in-house within a corporation or organization. The right half of the EDRM (right side) are activities typically outsourced to outside legal counsel and, in some cases, service providers. However, as with many business activities, there is significant variability in the degree of insourcing or outsourcing of these tasks. Thus, the point of transition from the insourced left side to the outsourced right side of the EDRM varies from organization to organization.

Electronic Discovery Reference Model / © 2014 / v3.0 / edrm.net
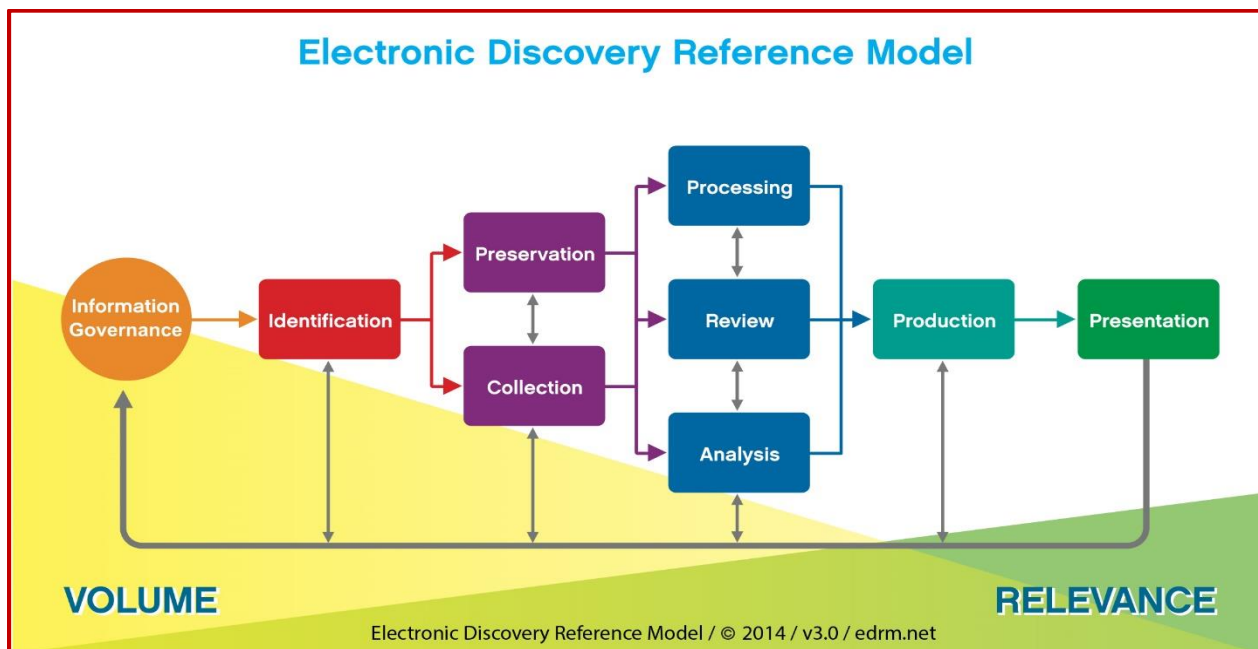
The ediscovery process involves a number of stakeholders and participants:

- In-house IT
- In-house Legal
- Organization's employees (custodians)
- Outside counsel (law firm)
- Law firm litigation support
- Vendors and service providers

Being a highly matrixed activity across several disciplines and departments, ediscovery often creates friction and complications for the stakeholders. For example, generally in-house legal drives ediscovery; yet, the bulk of the identification, preservation, and collection are technical tasks executed by in-house IT. Adding to the mix, often outside counsel will serve an advisory role and other vendors may assist with the technical work. As a legal process that involves technical execution, ediscovery is particularly challenging for the many parties involved. Effective communication and understanding of both law and technology can be difficult to maintain for everyone involved, since legal personnel may not be adept at technical jargon, and technical staff may not be fluent in legalese.

The early left side phases of ediscovery generally deal with large, unfiltered volumes of data. As the discovery process progresses, each phase depicted in the EDRM contributes to a refinement and reduction of data volume, focusing the relevancy of the dataset. This filtering process can be depicted as a funnel.
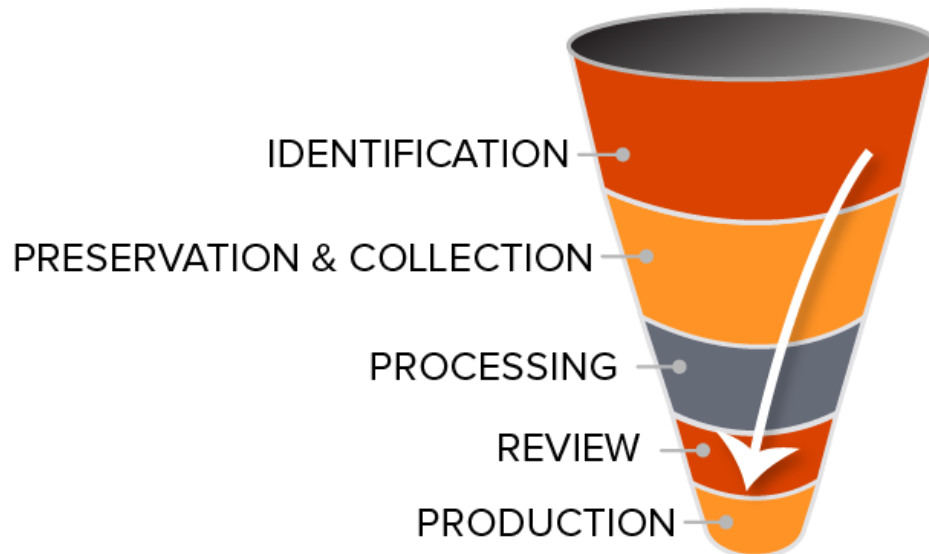
The cost of ediscovery is the principle driver for this funneling process. The review phase is one of the most expensive parts of ediscovery, and its costs are directly proportional to the volume of data. Thus, the more effective the funnel is prior to review, the greater realization of cost savings. Other benefits of a sound funnel process include faster review timelines, greater efficiency, and less risk of discovery missteps.

Typically, to achieve high performance ediscovery, organizations focus on the left side of the EDRM—identification, preservation, and collection phases – given its importance in this funnel process.

E-DISCOVERY AND THE FEDERAL RULES OF CIVIL PROCEDURE

In the U.S., the FRCP is the primary legal guidance on conducting ediscovery in civil litigation.[6] The 2006 amendments to the FRCP were the first to specifically address ediscovery. These amendments were largely successful. However, additional amendments were passed in 2015, which further refine the FRCP rules on ediscovery.

The 2015 rule changes substantially impact ediscovery practices and technology in 2016 and beyond. Historically, electronic discovery was conducted en masse with large swaths of data being preserved and collected as part of the initial phases of ediscovery. For example, organizations frequently preserved all email indefinitely and collected entire user mailboxes in response to litigation. Now, under the updated FRCP, Rule 26 adds an emphasis on discovery proportionality where litigants are to conduct discovery in proportion to the size and characteristics of the case. The new rules enable many organizations to more selectively (granularly) preserve and collect individual content.

Another change involves the Rule 16(b) case management conference. In the past, some conferences were not actual meetings but, rather, a *pro forma* exchange of emails, correspondence, or phone calls. Today, the rules require face-to-

---

DISCOVERY IMPACT
HIGHLIGHTS OF 2015 FRCP REVISIONS

EFFICIENT, FASTER DISCOVERY
- Rule 1 – "…just, speedy, and inexpensive determination of every action and proceeding" by the Court and the parties
- Rule 4 & Rule 26 – Discovery conference and scheduling conferences occur 30 days earlier
- Rule 16(b) – Judicial influence on discovery early in litigation has increased

PROPORTIONALITY & TAILORED DISCOVERY SCOPE
- Rule 26(b)(1) – Proportionality limiting the scope of discovery
- Rule 26(c) – Cost shifting in discovery specifically authorized
- Rule 37(e) – Use of proportionality and reasonableness when determining spoliation sanctions

EARLY EVIDENCE ANALYSIS, COOPERATION & PLANNING
- Rule 16(b)(1) – Early in-person case mgt. conferences (mail and phone removed)
- Rule 26 – Advocating cooperation and planning
- Rule 34 – Discovery objections require more specificity

FOCUSED PRESERVATION
- Rule 37(e) – Clarifies spoliation sanctions where spoliating party "acted with the intent to deprive another party," which is a more forgiving standard than inadvertent or negligent preservation failures.

face "live" case management conferences, which include discovery planning. Similarly, substantive Rule 26(f) conferences on discovery planning are becoming more commonplace.

Combined with the new proportionality approach to discovery, these conferences provide organizations with new opportunities to more artfully and narrowly define discovery, substantially reducing ediscovery costs and risks. For example, an organization can test and analyze different discovery scenarios in real-time during these conferences to more effectively negotiate advantageous scope of discovery and to establish reasonable discovery expectations, such as appropriate timelines for discovery production. Prior to the new rules, negotiating discovery from a position of knowledge has not been a common practice, as legal representation would often make decisions absent of any substantive analysis or research into discovery scope. This was partially due to the older FRCP rules, which had parties over-preserving and producing out of an abundance of caution to avoid discovery missteps. The updated rules encourage more pragmatism and proportionality.

In another key change, Rule 37(e) was completely rewritten to reduce broad over-preservation of electronic content. The FRCP drafting committee notes, "This rule recognizes that 'reasonable steps' to preserve suffice; it does not call for perfection." The committee further states, "Another factor in evaluating the reasonableness of preservation efforts is proportionality." Also, the committee writes, "Because the rule calls only for reasonable steps to preserve, it [sanctions] is inapplicable when the loss of information occurs despite the party's reasonable steps to preserve."

As the authoring committee notes, the Court may take adverse measures, such as sanctions, "only on finding that the party that lost the information acted with the intent to deprive another party of the information's use in the litigation," a substantially stricter standard than inadvertent or gross negligence.

Thus, organizations seeking high-performance discovery can now implement effective, common sense retention approaches, rather than a preserve-everything-forever paradigm.

SOURCES OF ESI

Formerly, email was the principle data source for ediscovery. Now, with the evolution of electronic communications, including text messaging, instant messaging, social-application messaging, digital phone calls (VOIP), and numerous other platforms, organizations are obligated not merely to conduct ediscovery on the easiest or most common application, but on any format or system that has relevant content.

Increasingly, ediscovery involves not only electronic communications, but electronic data managed or stored in other systems, such as Microsoft SharePoint, corporate financial systems, inventory systems, and web applications.

An additional complexity involves the many physical form factors in which data may reside. In the past, ediscovery collection efforts often focused on desktop computers or laptops. Today's corporate tablets,

mobile phones, datacenter servers, and cloud (virtual) environments have extended where ediscovery-related data lives enormously.

Employees' use of publicly available internet and cloud services for corporate use create further complications. Examples include Dropbox for sharing files, LinkedIn for sales and marketing, Hotmail for personal email, Google Drive for cloud storage, and Adobe Creative Cloud for graphic design.

High performance ediscovery necessitates a managed process and technology to work with this broad and ever-growing set of data sources.

DATA LIFECYCLE MANAGEMENT

Information governance is frequently thought of as a framework for managing information at an enterprise level, which supports an organization's immediate and future regulatory, legal, risk, environmental, and operational requirements.[7] A large part of this is managing the entire lifecycle of content, from its initial creation until its final disposition (deletion).

Data lifecycle management aligns with high-performance ediscovery in that it requires effective management of, and access to, data that is critical for successful ediscovery. Otherwise, ediscovery frequently inherits the bad habits and practices of suboptimal lifecycle management.

Depicted below is one view of the lifecycle of data as it progresses from its creation on the left to its disposition/deletion on the right. Throughout this entire cycle, effective management and security is key. Not uncommonly, data may be party to regulatory compliance obligations. These demands are most common during the early parts of lifecycle but can extend until disposition. Supervision activities
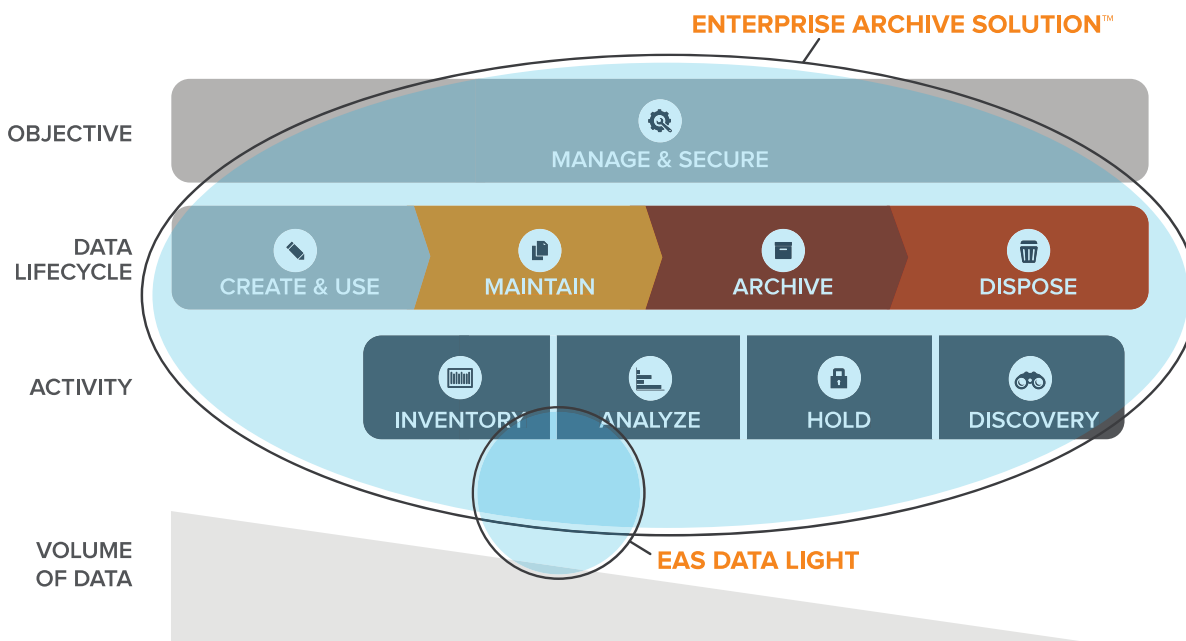
where content is actively monitored or reviewed for compliance is often a key part of supervision. Additionally, maintaining an inventory or catalog of such data may be part of a compliance program.

Data may also be responsive to discovery obligations. Because it can take months and, in many cases, years for litigation to commence, much data involved in discovery is past its adolescence and is in mid-life or near retirement. As such, data involved in discovery more frequently occurs on the right side of the data lifecycle, but can occur throughout.

LEGAL HOLD

High performing organizations adopt strong information governance and data lifecycle policies, including aging and deleting content once it is no longer needed by the organization. Legal holds complicate this process. Legal holds require organizations to preserve data when that data is believed to be relevant to reasonably anticipated litigation. This preservation process typically involves suspending normally scheduled deletion policies and timelines in order to preserve the content for litigation. If such process is not suspended and relevant data is deleted, "spoliation" occurs. Spoliation is the leading cause of discovery sanctions.

Thus high performance discovery obligates organizations to implement effective and defensive legal hold processes. One critical activity involves submitting a notice to those individuals in possession of relevant data. Such individuals, often called custodians, are advised of the organization's duty to preserve content and are provided instructions regarding their obligations. In addition to, or in combination with, this notice process, automated processes and other technical steps are implemented to electronically lock down relevant data. Organizations often utilize archive systems as a component of this legal hold process.

THE GROWTH OF BIG DATA

The term "Big Data" was coined by NASA in 1997.[8] Since then, the concept of big data has become mainstream, and has a direct intersection with electronic discovery.

Big data is often characterized by five V's:[9]

- **Volume** – Quantity of data exceeding traditional methods or technology
- **Velocity** – High-speed requirements for intaking or outputting data
- **Variety** – Broad mix of content types and sources
- **Veracity** – Possible variation of quality or reliability of individual data segments
- **Variability** – Inconsistent or non-normalized data populations

Electronic discovery practitioners will recognize the similarities between big data and information frequently encountered in electronic discovery. Since traditional software and database technologies struggle with big data, new big data-specific solutions and technologies have been developed over the past few years. These innovations more effectively support the five V's of big data. Organizations seeking high performance ediscovery are adopting technology solutions that parallel and embrace these same five "V" characteristics.

In the past, electronic discovery software was built on traditionally structured databases and not architected with a distributed framework for scalability. Also, ediscovery software frequently struggled with handling exceptions common to both ediscovery and big data (veracity and variability). As a result of these deficiencies, traditional ediscovery software not modeled for the characteristics of big data are not as effective in culling and funneling out data during the initial phases of ediscovery. Thus achieving high performance ediscovery benefits from ediscovery software that is built with a big data paradigm.

DARK DATA AND DATA ROT

Unmanaged data residing in corporate repositories has become an increasingly significant problem. The exponential growth of data is forcing forward-thinking organizations to deal with the issue of managing dark data – operational data that is not being used. All organizations can benefit from managing and applying retention and disposition policies to this redundant, obsolete, or trivial (ROT) data.

> " 69 percent of information in most companies has no business, legal, or regulatory value.[10]

Departmental and multi-user network file share repositories are frequent ROT offenders. As shared resources, large quantities of content are deposited, often with no long-term ownership and no motivation to remove content after its usefulness, if any, has ended. Compounding matters, network file shares often have little organizational structure. If subdirectories are used, they may be poorly or inconsistently organized or may be organized with obtuse naming that is indecipherable to anyone except the author. File names like "update.xls" further obfuscate whether the content in question is valuable content or ROT.

ROT and dark data dramatically increase costs and risks associated with ediscovery. Therefore, managing these with strong information governance practices is necessary to achieve high performance ediscovery. Many organizations exist "in a state of ROT" and need a method back to compliance and good data hygiene. These defensible disposition projects generally involve analysis of large quantities of data to make disposition decisions and to bring the remaining content under more effective management control. Some of this work can be done manually, but for large volumes an automated, rule-based categorization and tagging process is recommended. Such processes and technologies need mechanisms to provide defensibility to ensure the disposition process meets legal, investigative, and/or compliance obligations and expectations.

Some software employ machine-learning technologies such as trainable clustering to speed the identification of similar documents for defensibly applying retention and deletion policies en masse. These technologies should support analysis of a document's metadata and its content while supporting the hundreds of file types found within the organization. Ideally, the processes and technologies should leverage author and access information maintained in Active Directory and the file system.

MACHINE LEARNING

Machine learning is a broad category of technology that uses sophisticated algorithms and statistical probabilities to intelligently calculate outcomes. These technologies mimic human decision-making and often are classified as artificial intelligence. In the context of ediscovery, conceptual clustering is one of the earliest implementations of this type of technology. Often presented graphically as bubbles or heat-maps, similar documents are grouped together under the assumption humans will likely take similar actions with documents that contain similar content. Complex mathematical processes, usually based on word associations and patterns, help determine similarities between documents. This approach has been commercially available in ediscovery for about 15 years, and remains one of the most common types of ediscovery machine learning.
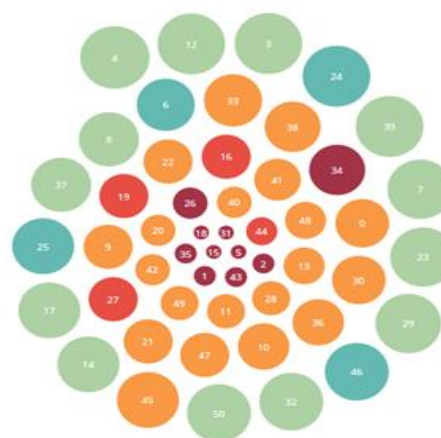
Over the last decade, new machine learning technologies have developed, most commonly under the category of technology assisted review (TAR). Since manual review is expensive and time-consuming, TAR's objective is to automate the review of documents for privilege and responsive determination. The biggest challenge with TAR is whether predictive coding is as good as what humans could achieve in a manual review. The growing consensus is that TAR can be effective, particularly when a statistically suitable set of training documents is used to demonstrate the defensibility of an automated review.

BUT HOW TO JUSTIFY THE INVESTMENT?

Organizations of all sizes struggle with developing the business case for dark data-related projects. When competing for budgets, cleaning up ROT might seem like a low priority. But it is possible to develop a persuasive business case by focusing on these benefits:

- Mitigation of risk
- Increased productivity
- Enhanced data mining
- Reduced ediscovery costs
- Improved regulatory compliance
- Reduced storage costs

Risk mitigation is often hard to quantify, but risk associated with dark data is real and substantive in business operations and, particularly, as it relates to litigation. Productivity improvements are more obvious, given the large quantity of time employees spend hunting for lost and misfiled information. When content is under effective information management, organizations have more opportunity to mine for data benefiting the operation of the organization.

Logically, reducing the volume of data reduces ediscovery costs. Defensible disposition projects can provide significant benefits in fulfilling an organization's regulatory compliance. One of the most obvious benefits is reduced costs for content storage, including reducing infrastructure costs associated with multiple copies of that content duplicated for business continuity and backup.

DEFENSIBILTY WITH REASONABLENESS, INTENT, AND PROPORTIONALITY

In the past decade, over 1,000 cases have involved possible sanctions for significant ediscovery missteps. Cases have been dismissed, juries instructed with adverse inference, and opposing parties awarded significant monetary awards. For example, last year a corporation was sanctioned over $7.4 million for improper discovery.[1]

This scenario, however, is not inevitable. Options to mitigate the risks and consequences of discovery missteps are available to organizations with high performing ediscovery practices. When dealing with large volumes of data, mistakes are inevitable. Therefore, being prepared for that eventuality, by demonstrating to a court reasonableness, intent, and proportionality, is of paramount importance.

The FRCP as amended last year directs that sanctions are warranted when the party "acted with the intent to deprive another party." If organizations can demonstrate their actions were reasonable and the inappropriate discovery action was accidental or inadvertent, they likely could avoid sanctions.

Suggestions for demonstrating reasonableness, intent, and proportionality include:

- When determining if a legal hold is needed, document the decisions that are made and the rationale behind those decisions—such as why or why not this matter triggered a legal hold.
- When issuing legal holds, document the decisions made and the rationale behind those decisions—such as which custodians are included, or other parameters, including timeframes, sources, etc.
- When determining discovery parameters such as a keyword list, document why certain words and parameters were chosen and the rationale behind those choices.

---

[1] *In re Delta/Airtran Baggage Fee Antitrust Litigation*, 2012 U.S. Dist. LEXIS 13462 (N.D. Ga. February 3, 2012)

- Maintain chain of custody documentation for discovery-related content. For content managed by a technical system, verify that the system maintains and can easily report out the chain of custody.
- When proportionality is a factor in discovery, document the rationale used for determining proportionality. Screenshots or reports from early case assessment (early evidence assessment) software may be particularly compelling.

10 STEPS TO HIGH PERFORMANCE EDISCOVERY

1) **Inventory all enterprise data sources**	Maintain a continually updated inventory of all enterprise data sources, including those provided by third-party vendors and public cloud services.
2) **Maintain comprehensive information governance across sources**	Manage data proactively and across all sources to positively impact discovery performance, knowing ediscovery frequently inherits the bad habits and practices of suboptimal data lifecycle management.
3) **Implement pragmatic retention policies and enforcement**	Avoid over-/under-preservation of data. Both too much and too little data preservation increases your cost and risk.
4) **Dispose of redundant, obsolete, and trivial (ROT) content**	Reduce data volumes and corporate risks by eliminating ROT from unmanaged or unorganized file shares and other repositories.
5) **Leverage targeted, surgical legal holds**	Reduce discovery volumes and associated costs with focused legal holds that minimize the preservation of unnecessary content.
6) **Take advantage of early evidence analysis and negotiated, proportional discovery scope**	Employ proportional discovery as allowed by the recently revised FRCP, which is particularly compelling when combined with early evidence analysis techniques, to reduce discovery costs and risks.
7) **Employ fast, focused searches**	Use modern technology to quickly and iteratively search content and to narrow and focus those results through metadata and other filtering techniques. Avoid only using a basic list of keywords with no other parameters.
8) **Use machine learning to increase review efficiency and accuracy**	Use technology assisted review (TAR) and other machine learning technologies to dramatically enhance the document review process, resulting in fewer mistakes and reduced costs.
9) **Run defensible exports with chain of custody**	Increase defensibility and demonstrate ediscovery proficiency through proper chain of custody.
10) **Ensure legal defensibility by documenting reasonableness, intent, and proportionality**	Document the decision-making process to demonstrate defensibility, particularly with the new 2015 FRCP amendments, for inadvertent, but inevitable, mistakes.

HIGH PERFORMANCE EDISCOVERY AND ENTERPRISE ARCHIVE SOLUTION (EAS)™

The solution is simple. You can achieve high performance ediscovery and the 10 steps outlined in this whitepaper with Enterprise Archive Solution. EAS delivers industry-leading information governance for the entire enterprise. Data from across an organization can be rapidly searched, preserved, and managed through its entire lifecycle, using updated features designed to meet today's litigation and regulatory obligations. EAS takes archiving further by combining the power of a sophisticated granular disposition policy engine with flexible storage management.

EAS OVERVIEW

Archiving has evolved from a tactical IT need to optimize email storage into a corporate requirement for proactive information governance and risk management. EAS is an enterprise-grade, scalable, and comprehensive archive software solution, bringing order to information chaos by putting you in control of your data.

EAS is a complete information governance solution. EAS includes modules for electronic discovery, compliance, dark data analysis, and audio-video, with connectors to all core enterprise technologies and systems. Gone are the days when separate manual searches are needed to query multiple data repositories across an organization.

EAS benefits from HP's Intelligent Data Operating Layer™ (IDOL) search technology, the market-leading enterprise information processing platform that uses Meaning Based Computing technology to form a conceptual and contextual understanding of over 400 file formats.

LEGAL DISCOVERY WITH EAS DISCOVERY

Responding to investigations and lawsuits has never been easier with the enhanced discovery capabilities of EAS. EAS Discovery handles critical discovery capabilities including legal holds and early case assessments (ECA). EAS is one of the first archive solutions on the market to be fully compatible with the new proportionality and other obligations found in the 2015 FRCP amendments. Legal teams can develop smart discovery strategies by quickly assessing discovery search parameters with sophisticated analysis and reports on discovery volumes and associated costs.

SUPERVISION COMPLIANCE WITH EAS

EAS delivers sophisticated compliance and supervision capabilities for regulated firms, registered investment advisors, and individuals and organizations facing challenging regulatory obligations. EAS includes automated policy enforcement enabling regulated organizations to enforce supervision and related recordkeeping and data preservation obligations. The EAS Supervision module supports

streamlined compliance review with customized email supervision policies and audit trails for compliance with FINRA, SEC and other regulatory agencies. EAS Supervisor also features concept searching, automatic categorization, case management, and real-time and post-process monitoring. With EAS, organizations can proactively enforce and monitor users' adherence to a wide range of regulations and policies.

DATA LIFECYCLE MANAGEMENT WITH EAS

EAS provides end-to-end data lifecycle management, giving you granular control over the birth, life, and death of data governed by EAS. EAS eliminates data silos with connectors to all of the systems common in today's corporate environment, including traditional repositories and new cloud and social media technologies. Fulfill challenging EU privacy regulations and records management needs with the enhanced capabilities and flexibility found only with EAS.

The formula-based policy editor provides unsurpassed control over data's lifecycle. Use a wizard-driven policy editor option for ease-of-use or create highly advanced policies using a scripting option built directly into the policy editor. Policies can be as simple as a single universal policy governing all data or as complex as dozens or hundreds of discrete policies to support even the most complex data retention schedules.

DARK DATA WITH EAS DATA LIGHT™

With EAS Data Light, uncover and gain access and control over data collections that otherwise are unmanaged or inaccessible. Bring into the light and make searchable file shares, unorganized repositories, and corporate dumping grounds of electronic content. With the many technology connectors available with EAS, crawl and extend management control over those data collections. Enable defensible disposition of large quantities of data using automated, rule-based categorization and tagging.

END NOTES

[1] Email Statistics Report 2015-2019, The Radicati Group, Inc., A Technology Market Research Firm, http://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf (February 2015)

[2] John Gantz and David Reinsel, The Digital Universe Decade – Are You Ready?, IDC, http://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf (May 2010)

[3] Maura R. Grossman and Gordon V. Cormack, EDRM page & The Grossman-Cormack Glossary of Technology-Assisted Review, with Foreword by John M. Facciola, U.S. Magistrate Judge, 2013 Fed. Cts. L. Rev. 7 (January 2013)

[4] Kim J. Myrdahl, Peter H. Walsh, et. al. Arms Around the Tempest: E-Discovery, Cost Control and Legal Project Management. (November 2011)

[5] Electronic Discovery Reference Model (EDRM), www.edrm.net

[6] For more information on the FRCP, see https://www.law.cornell.edu/rules/frcp and http://www.uscourts.gov/rules-policies/current-rules-practice-procedure/federal-rules-civil-procedure

[7] For more information, Smallwood, R. F. (2014). Wiley CIO: Information Governance: Concepts, Strategies, and Best Practices. Somerset, NJ, USA: John Wiley & Sons, Incorporated.

[8] http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours

[9] The 3 V's were first introduced by Gartner (Meta Group) analyst Doug Laney and have been expanded upon by the industry. https://en.wikipedia.org/wiki/Big_data

[10] 2012 Compliance, Governance and Oversight Council (CGOC) survey