IRON MOUNTAIN®

# DARK DATA TASK FORCE REPORT

## Identification and Remediation of Dark Data in Law Firms

JULY 2015

LAW FIRM INFORMATION GOVERNANCE SYMPOSIUM JULY 2015

# CONTENTS


BACKGROUND . . . . . 1

EXECUTIVE SUMMARY . . . . . 5

INTRODUCTION . . . . . 5

WHAT IS DARK DATA? . . . . . 5

DARK DATA VS. BIG DATA . . . . . 7

WHY FOCUS ON DARK DATA? . . . . . 7

LANDMINES: WHERE IS DARK DATA HIDING? . . . . . 8

CONSIDERATIONS . . . . . 9

CLIENT DATA CONSIDERATIONS . . . . . 10

POLICY AND RETENTION/DISPOSITION CONSIDERATIONS . . . . . 13

COST SAVINGS ACHIEVED FROM MANAGING DARK DATA . . . . . 13

APPROACHES TO ELIMINATING DARK DATA . . . . . 15

WHAT HAS NOT WORKED? . . . . . 16

SEVEN STEPS TO IMPLEMENTING MORE AGGRESSIVE FIRM POLICIES . . . . . 17

EMERGING TECHNOLOGIES . . . . . 19

DATA LOSS PREVENTION VS. FILE ANALYSIS SOFTWARE . . . . . 20

LEGACY DATA CLEANUP . . . . . 22

MANAGING DATA GOING FORWARD . . . . . 22

AUTOMATE POLICY-DRIVEN CLASSIFCATION . . . . . 23

MANAGE IN-PLACE . . . . . 24

CLIENT CHARGE BACKS AND WHOLLY-OWNED SUBSIDIARIES . . . . 24

SUMMARY . . . . . 26

ACHIEVING BALANCE . . . . . 26

APPENDIX A: GLOSSARY OF TERMS . . . . . 28

APPENDIX B: SAMPLE CHECKLIST FOR THE COLLECTION OF DARK DATA . . . . . 31

APPENDIX C: SAMPLE DATA MAP . . . . . 33

APPENDIX D: SAMPLE DATA FLOW DIAGRAM . . . . . 34

APPENDIX E: SAMPLE ELEVATOR PITCH . . . 35

REFERENCES . . . . . 36

BIBLIOGRAPHY . . . . . 36

# BACKGROUND

Since 2012, the Law Firm Information Governance Symposium has served as a platform for the legal industry to collaborate on information governance (IG) best practices in the unique setting of law firms. The Symposium publications offer definitions, processes and best practices for law firm IG. In 2014, four task forces were assembled by the Symposium Steering Committee to work on specific, current law firm IG topics. This Dark Data Task Force Report explains the pressing problem of ever-increasing amounts dark data in law firms, describes how to bring this rogue information under control and provides best practices to keep this issue from intensifying into the future.

## SYMPOSIUM STEERING COMMITTEE

**BRIANNE AUL, CRM**
Firmwide Records Senior Manager
Reed Smith, LLP

**LEIGH ISAACS, IGP, CIP**
Director, Records & Information Governance
White & Case LLP

**RUDY MOLIERE**
Firm Director Records and Information
Morgan, Lewis & Bockius LLP

**STEVEN SHOCK**
Lead Consultant / Interim Director, Network
Information Management Systems
eSentio Technologies

**CHARLENE WACENSKE**
Senior Manager Firm Wide Records
Morrison & Foerster LLP

## DARK DATA TASK FORCE

**ANGELA AKPAPUNAM**
Director of Document Lifecycle Services
WilmerHale

**SCOTT CHRISTENSEN**
CIO at Large

**PATRICIA A. FITZPATRICK***
Director of Information
Governance & Compliance
Katten Muchin Rosenman LLP

**GRANT W. JAMES, CRM**
Senior Manager Information Governance
Troutman Sanders LLP

**SAMANTHA LOFTON**
Chief Risk and Information Governance Officer
Ice Miller LLP

**FARON LYONS**
Enterprise Account Executive
Alfresco Software

**DANA C. MOORE, IGP**
Manager of Records
and Information Compliance
Vedder Price PC

**CHARLENE WACENSKE**
Senior Manager Firm Wide Records
Morrison & Foerster LLP

***Task Force Leader**

## SYMPOSIUM PARTICIPANTS

Iron Mountain would like to thank the following individuals for participating in the peer review sessions of the 2015 Symposium event and for sharing their perspectives and expertise during the creation of this task force report.

**ANGELA AKPAPUNAM**
Director of Document Lifecycle Services
WilmerHale

**KAREN ALLEN**
Manager, Information Governance Technologies
Morgan Lewis & Bockius LLP

**DERICK ARTHUR**
IG Operations Manager
Cooley LLP

**BRIANNE AUL, CRM**
Firmwide Records Sr. Manager
Reed Smith LLP

**BRYN BOWEN, CRM**
Principal
Greenheart Consulting Partners

**BETH CHIAIESE, CRM, MLIS**
Director, Professional Responsiblity & Compliance
Foley & Lardner LLP

**SCOTT CHRISTENSEN**
CIO at Large

**TERRENCE COAN, CRM**
Senior Director
HBR Consulting

**JULIE COLGAN, IGP, CRM**
Head of Information Governance Solutions
Nuix

**GALINA DATSKOVSKY**
CEO
Vaporstream

**BRIAN DONATO**
CIO
Vorys, Sater, Seymour and Pease LLP

**BETH FAIRCLOTH**
Director of Risk Management
Seyfarth Shaw LLP

**STACEY FIORILLO**
Director of Records Management
and Information Governance
eSentio Technologies

**PATRICIA FITZPATRICK**
Director of Information Governance & Compliance
Katten Muchin Rosenman LLP

**JAMES FLYNN, CRM**
Director of Records and Docket
Winston & Strawn LLP

**GRANT JAMES, CRM**
Senior Manager Information Governance
Troutman Sanders LLP

**SHARON KECK**
Director of Risk & Records Info. Management
Polsinelli, PC

**CHARLES KENNEDY**
Firm Director of Records and Docket
Jones Day

**SAMANTHA LOFTON**
Chief Risk and Information Governance Officer
Ice Miller LLP

**FARON LYONS**
Enterprise Account Executive
Alfresco Software

**RUDY MOLIERE**
Firm Director Records and Information
Morgan Lewis & Bockius LLP

**DANA MOORE, IGP**
Manager of Records and Information Compliance
Vedder Price PC

**DERA NEVIN**
Director, eDiscovery
Proskauer Rose LLP

**RANDY OPPENBORN**
Director, Information Governance
Foley & Lardner LLP

**ALEXANDRA PROPHETE**
KM Operations Manager
Cleary Gottlieb Steen & Hamilton LLP

**DEB RIFENBARK, IGP, CRM**
Director of Records and Compliance
Stinson Leonard Street LLP

**STEVEN SHOCK**
Lead Consultant / Interim Director
Network Information Management Systems
eSentio Technologies

**SCOTT TAYLOR**
Manager of Records, Conflicts
& New Business Intake
Smith, Gambrell & Russell LLP

**CHARLENE WACENSKE**
Senior Manager Firm Wide Records
Morrison & Foerster LLP

**JOHAN WIDJAJA**
Assistant Director Records & Information
Morgan Lewis & Bockius LLP

**JOEL WUESTHOFF**
Senior Director
Robert Half Legal

# EXECUTIVE SUMMARY

This report was created for law firm information governance (IG) management and law firm executives who are beginning to recognize, or are in fact currently dealing with the problems associated with dark data. The report logically begins with an industry (and Symposium) definition of dark data and explains where dark data can be found in law firms.

The report then moves into a detailed explanation of the problems/issues/challenges law firms face with the unmanaged accumulation of dark data, including rising storage costs, reduced employee productivity and increasing risk around client data leakage.

The report then explores the various tools that can help bring this dark data phenomenon under control. The tools discussed include the creation of polices and workflows to address dark data now and into the future, as well as emerging technologies such as File Analysis Software (FAS) and Data Loss Prevention (DLP) software that can automate the identification, classification, management and disposition of dark data.

Finally this report offers a three-pronged approach for managing dark data in law firms.

## INTRODUCTION

"Olly olly oxen free! Come out, come out wherever you are." Finding dark data is often like playing the childhood game of hide and seek. The information governance (IG) professional, playing the role of the "seeker" or "it," is on a mission to uncover data that has been hiding in various firm repositories for many years. The seeker "tags" objects as they are found and continues with this strategy until all dark data has been flushed from its hiding places.

"Ten, nine, eight, seven, six, five, four, three, two, one! Ready or not, here I come!" cries the IG professional as she or he begins the hunt for dark data. This dark data report provides practical guidance for law firms who may be searching for dark data, now or in the near future. The report is a compilation of information obtained through a survey of Symposium participants, industry research and group member work experience. A framework is provided for evaluating the law firm dark data posture. In addition, recommended strategies are offered as a way to justify an evaluation and eventual analysis of dark data. The strategies considered include uncovering dark data, continuing management and cost-benefit and risk assessment. Law firm IG professionals may find this information useful as they explore and settle on an appropriate methodology for their firms.

## WHAT IS DARK DATA?

There are many characterizations of dark data. In July 2014, the LegalTech® West Coast panel "Recover or Delete Dark Data" defined dark data as enterprise data that is predominately uncategorized, has limited visibility to the organization (if not completely obscured) and because of its obscurity, serves no apparent business purpose.[1] Law firms are experiencing growing volumes of dark data across their technology platform(s). In fact, dark data is lurking in many data and content systems including mobile devices, local computer drives, email, network file shares, legacy paper files, cloud file sharing services and even structured databases, such as the document management system. Dark data is largely unstructured, such as real-time communications and documents, but can also be semi-structured, for example XML code, or structured, as in a database. Few firms today have insight into their dark data and many are fearful of exposing content that could be highly confidential or contain information subject to legal discovery.

The Dark Data Task Force surveyed a small sample of law firm records managers and IG professionals. The Dark Data Survey results showed that most firms report they are either currently addressing dark data or plan to do so within the next year (see Chart 1). Firms with no plans to address this topic are in the minority.

## TIMEFRAME IN WHICH DARK DATA WILL BE ADDRESSED

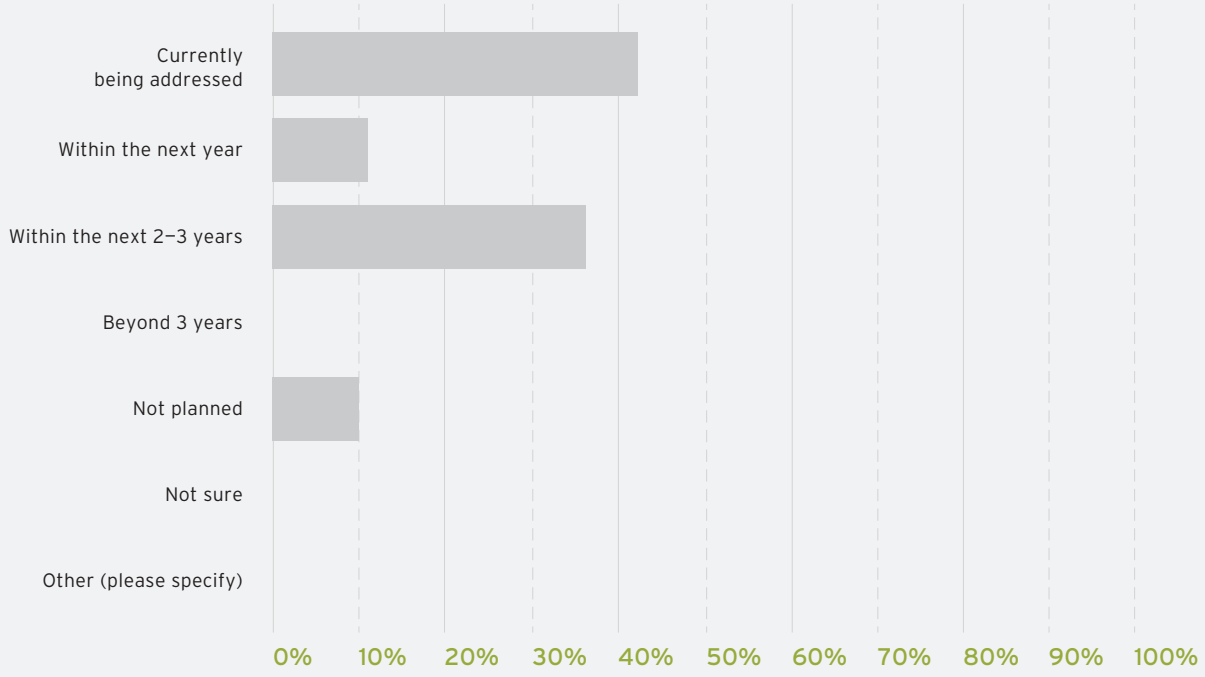



CHART 1

The survey also indicated that the majority of firms are either in the process of developing policies on dark data or already have these policies in place (see Chart 2).

## DOES YOUR FIRM HAVE FORMAL PROCEDURES/POLICIES FOR DARK DATA?

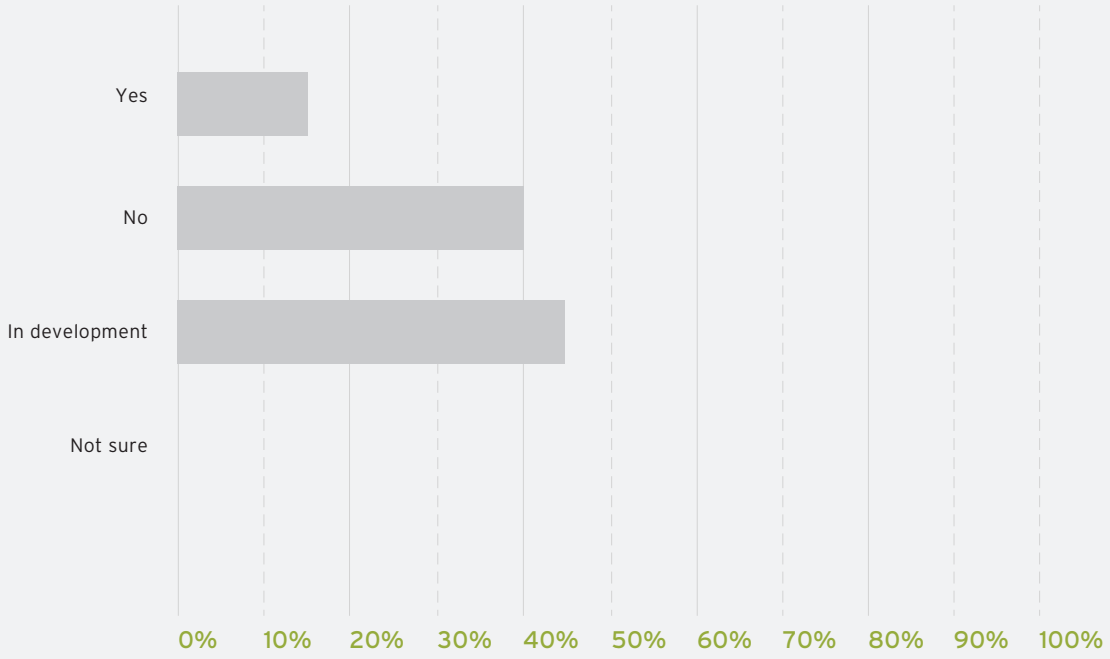



CHART 2

## DARK DATA VS. BIG DATA

We cannot adequately discuss dark data without also mentioning big data. Big data is specific large pools of data against which analytics are run. Big data has, in recent years, compelled law firms to consider new technologies such as enterprise search and predictive coding to help make content more accessible and meaningful. Many organizations, including law firms, find themselves becoming paralyzed by the growing mountain of big data, of which the vast majority is considered dark.

## WHY FOCUS ON DARK DATA?

Dark data is costly to analyze due to data quality issues, lack of resources to perform analysis and cleanup or because it is just not yet recognized as a problem by the organization. Many law firms take the stance of "what you don't know won't hurt you." If this is true, then shouldn't dark data be left alone? We know that enterprise data capacity is growing at a rate of up to 60 percent on average annually, making the ability to manage the growth of big data even more challenging and potentially prohibitive. The key challenge with dark data is in determining if there is any real value to justify the management of it. The main concern for most is that managing dark data would be more expensive than any realized value gained from its management.

To determine if dark data is even worth further analysis, law firms need a means of cost-effectively finding, reviewing, organizing and visualizing dark data on an ongoing basis. Some law firms are beginning to use file analysis software to spot trends in legal project costs or client behaviors for business development. This information can expose patterns that can improve law firm business planning. Firms should also employ file analysis software to archive valuable business data as well as defensibly destroy valueless dark data. Enterprise search and workflow applications can be leveraged to migrate high-value content to a structured repository based on end-user activity. Firms should work towards prohibiting the use of offending storage locations and provide functionally equivalent solutions for day-forward activities (i.e., file system emulation while cleanup activities are ongoing).

## LANDMINES: WHERE IS DARK DATA HIDING?

Dark data is alive and well in law firms. It lives in both paper and electronic format, evidenced by the number of unmanaged paper records located in abandoned practice group work rooms, storage areas, attorney offices, abandoned file cabinets, offsite storage and attorney home offices. Some users (you know who they are!) have secret file locations. Others refuse to close a file, saving space for the next big case. And then there are the secretive individuals that simply reply, "Only I need to know where those files are located," when asked about files.

Valuable information is often lost in unmanaged and ungoverned repositories. This dark data lives in dormant servers, legacy applications, unclassified email messages, departed attorney mailboxes and network share drives, as well as countless other repositories. The issue spans the entire firm and has no boundaries. It even expands to third-party vendors who are housing data on behalf of the firm in databases, such as systems that house administrative HR and payroll data, as well as third-party vendors that host discovery data for litigation cases and corporate deal rooms.

Generally speaking, most users want to be efficient and will follow processes that make sense. The firm's information governance team can help by seeking input from subject matter experts and designing workflows that make sense, save time, eliminate duplicate data and facilitate collaboration among teams, while saving server space and better managing the firm budget. However, each organization is unique. This report will attempt to give you some suggested places to begin your search for dark data and provide methods for bringing it under control on an ongoing basis. See Appendix B for a sample checklist for the collection of dark data.

In order to manage the madness, start with the creation of a data map by asking questions such as:

**» Where are people storing documents?**

**» What types of documents are stored there?**

**» What are the date ranges?**

**» What business units are affected?**

See Appendix C for a sample data map and places to look for ESI and hard copy documents.

Be sure to evaluate approved repositories for the firm by administrative department(s) and practice group(s). Keep in mind that document management needs will vary by group. Also be aware that a data map is a living document that should be updated on a regular basis. You will need to map the various ways that data comes into the organization. Established firm policy should be considered when mapping out standard data flows because revisions or updates may be necessary in terms of how users are taking on risk for the organization. Successful data flows are created in a collaborative environment and endorsed by senior management. Understand how dark data is created so it

can be identified and managed. For example, some users may maintain their own "just in case" copy of a data set on the shared drive or other users may fail to use established client matter numbers to store data in the document management system.

The Sample Data Flow Diagram in Appendix D tracks the movement of data into and out of the organization and defines the path for mitigating the addition of content that doesn't belong. Failure to establish these boundaries adds risk to the organization and makes it difficult to comply with litigation holds, court ordered destruction, client directives via outside counsel guidelines (OCG) and routine disposition as defined in the firm's retention policy.

## CONSIDERATIONS

Firms should determine which functional area within the organization will lead the charge in managing dark data. Most often this responsibility falls to the Records Management Group (Chart 3).

### ADMINISTRATIVE FUNCTIONAL AREA WITH PRIMARY RESPONSIBILITY FOR IDENTIFICATION AND ONGOING MANAGEMENT OF DARK DATA

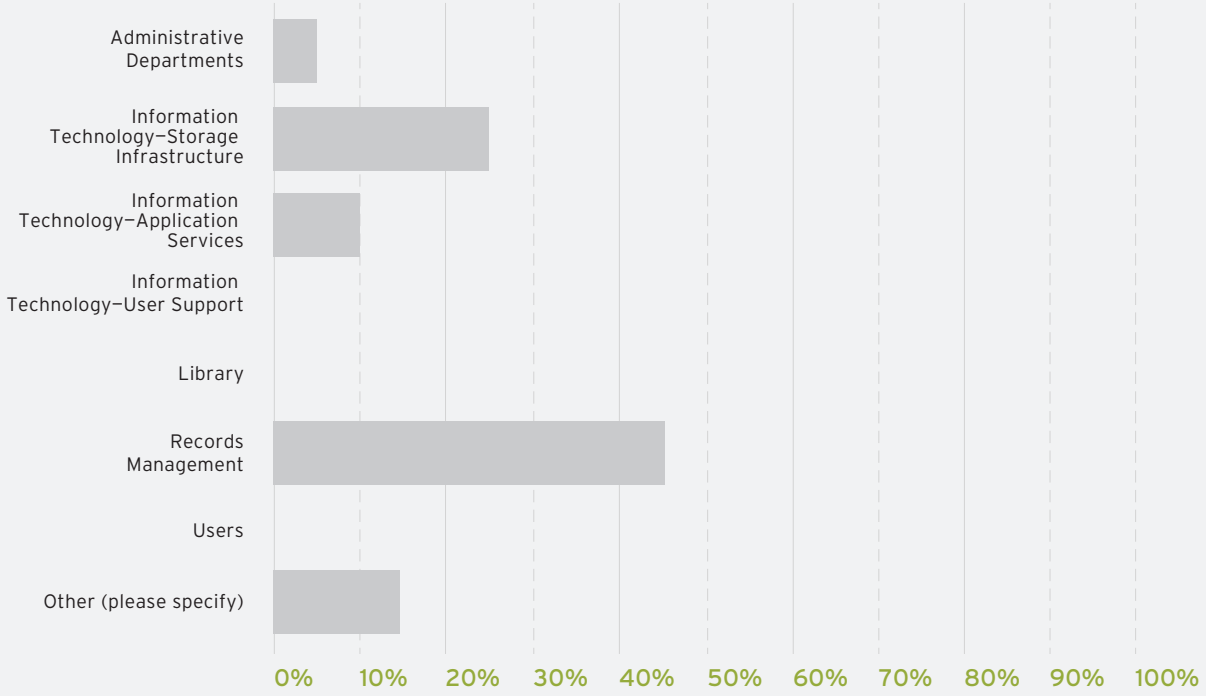



CHART 3

Another consideration is the frequency at which dark data will be evaluated. For electronically stored information (ESI), the most common frequency is annually or as part of a system or server migration (Chart 4). This approach will limit effectiveness unless some form of automation or system design can keep up with the growth of data.

## CLIENT DATA CONSIDERATIONS

The bulk of a firm's dark data is old client data. For long-term clients, this data is primarily paper, and as you progress in time, the data begins to transition to more electronic information. This can be a difficult problem to tackle because there is no obvious methodology to round up dark data that is right for every firm. However, there are several basic tenants to consider.

Risk-based decisions will need to be made which will require that you have conversations with your General Counsel. While dark data is a hot topic in the industry and an issue that almost every firm is dealing with, there is no bright-line rule or prevailing case law that provides specific guidance as to how to proceed. Most firms are forced to make risk-based decisions derived from the general topic surrounding their duty to safeguard client files. While the lawyer is clearly under an obligation to preserve and protect client records, it remains unclear for how long said records need to be retained (if no client directive exists). The conversation then turns into a question of: At what cost must the firm continue to safeguard client information and attempt to locate the client in order to obtain consent to destroy?

The life-to-date offsite storage cost of records relating to a particular matter can potentially exceed the revenue generated by that matter. Additional consideration should be given to potential costs that would be incurred if boxes were to be retrieved and examined. Typically, temporary staff needs to be brought in to facilitate the effort and firms report an 8-10 year return on this investment when this is done. There is a trend among firms subscribing to the position that these measures are cost prohibitive and create an undue hardship on the financial position of the firm at a time when there is significant pressure to provide the highest quality legal service at the lowest possible cost. These firms accept the risks associated with accelerated disposition and minimal client consent as a means of remaining competitive (and in business). Similar arguments are being made to dispose of legacy ESI, especially when the client matter (number) to which it belongs has not been linked to the record. Manual review and analysis of this data is equally burdensome and this problem is growing faster than ever.

If your firm decides to proceed along these lines, you must be able to document your process. That doesn't mean you must have a retention policy; many firms don't and can still safely destroy dark data. But you will need to demonstrate that you have consistently defined the scope of the destruction and that you have applied destruction consistently within that scope. Having a well-documented process that is consistently applied will help the firm defend against potential claims of spoliation in the future.

## FREQUENCY OF DARK DATA EVALUATION

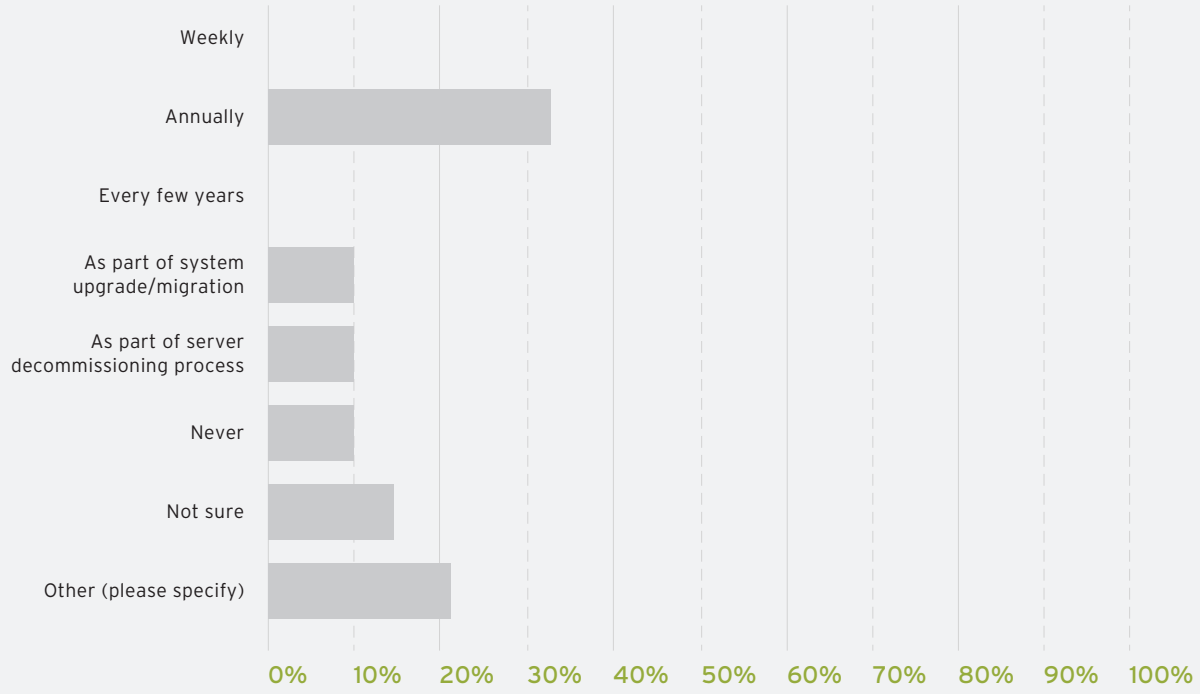



CHART 4

A big hurdle for firms is the idea that all data needs to be treated the same. This simply is not the case and frankly it is almost impossible to apply the same criteria to all legacy data—so don't. You should categorize your data into different buckets and apply a customized set of criteria to each bucket. Illustrated below are several rules that can be applied to different types of legacy data.

**Example 1:** All clients that have all matters closed for more than 30 years.

» **Review matters for legal holds.**

» **Notification of matter destruction to the responsible attorney if she/he is still at the firm. No internal notification where the responsible and billing attorney is no longer at the firm.**

» **No client notification.**

**Example 2:** Clients with all matters have been closed between 10 and 30 years.

» **Review matters for legal holds.**

» **Notification of matter destruction to the responsible attorney if she/he is still at the firm. No internal notification where the responsible and billing attorney is no longer at the firm.**

» **Compare client contact information from billing system or contact management system (if available) to information found by doing a web search to see if the client can be located. If located, send notification of destruction letter including matter list.**

**Example 3:** Clients with open matters as well as matters closed more than 10 years.

» **Review matters for legal holds.**

» **Review engagement letter and outside counsel guidelines (if provided).**

» **Notification of matter destruction to the responsible attorney if she/he is still at the firm. If the matter has no active billing and responsible matter notification should go to a partner of one of the other active matters.**

» **Use client contact information from billing system or contact management system to reach out to client for authorization to destroy data for closed matters. Also try and obtain instructions for the ongoing disposition of data relating to matters as they close.**

**Example 4:** Boxes sitting in offsite storage with unknown contents and the vendor box history confirms no activity in 10 or more years.

» **Check local rules of professional responsibility.**

» **Conduct conversation with General Counsel to confirm decision to proceed.**

» **Destroy boxes based upon lack of information and inactivity.**

» **Document decisions made and actions taken in permanent firm file.**

**Example 5:** Boxes sitting in offsite storage with known contents, but the vendor box history confirms no activity in 10 or more years.

» **Check local rules of professional responsibility.**

» **Conduct conversation with General Counsel to confirm decision to proceed.**

» **Destroy boxes based upon inactivity.**

» **Document decisions made and actions taken in permanent firm file.**

**Example 6:** Records brought into your firm by a lateral hire as "form files from their prior firm" who has since left your firm and left these records behind.

» **Check state bar listing for current contact information of attorney.**

» **Send notification to attorney asking to accept delivery of records and to confirm address.**

» **If no response in 60 days, conduct conversation with General Counsel to confirm decision to proceed with destruction of records.**

» **Document decisions made and actions taken in permanent firm file.**

**Example 7:** Records brought into your firm by a lateral hire as "form files from their prior firm" who is still at your firm.

- **Contact attorney and ask them to identify any that have since become active clients of your firm and link them to valid client matter numbers.**
- **Identify any records that need to be saved as a true "form file" and redact all confidential information and party name information. Save form to knowledge management system or departmental workspace.**
- **Delete all other ESI and paper records.**

## POLICY AND RETENTION/DISPOSITION CONSIDERATIONS

Dark data is inherently addressed in information governance policies that define standards for the creation, access, use, maintenance, storage and disposition of client and firm business information. The goal is to provide access to information while effectively managing risk, compliance with legal and regulatory requirements and controlling cost relating to information storage. Policies with the biggest impact on reducing the footprint of dark data include, but are not limited to:

**Approved Repositories:** Define approved records repositories while limiting or eliminating access to repositories for unclassified data. Processes that support this policy would include the intake of client representation data, eDiscovery and professional staff personal data.

**Records Retention:** Define those records that should be maintained according to a predetermined retention schedule, where they should be stored and how they should be disposed of. Processes for the collection of dark data for review and classification at matter close will ensure the disposition of matter data at final disposition.

**Legal Holds:** Define the firm's obligations in the situation of pending or potential litigation, claim, investigation or subpoena. When the firm is required to preserve data, all repositories of dark data that could contain relevant records must be searched. When the firm reduces the amount of dark data, they can substantially reduce the risk of exposure to the firm and its clients.

**eDiscovery:** A policy that defines the management and lifecycle of data subject to an eDiscovery request could specify how and where load data and databases are stored, limit duplicative information and control access to data.

All information governance policies should address the management and protection of client and firm business information assets, including dark data.

## COST SAVINGS ACHIEVED FROM MANAGING DARK DATA

There are two types of return on investment (ROI) that can be linked to the destruction of dark data. The first is the minimization of risk associated with retaining dark data. These risks have already been discussed throughout this paper. The second is the direct cost to the firm of keeping both paper and electronic data past its useful life. It is important to quantify the savings to the firm that result from the destruction of the dark data. Once you have identified the dark data and you have determined a method of destruction, document the cost savings to the firm.

For archived boxes it is easy to calculate ROI. The following example provides a rough estimate of the cost to the firm. Let's say there are 200 boxes held in storage since 1950. The estimated cost of retaining these 200 boxes from the initial storage date to the present is approximately $31,200.

**1950 to 2015 = 65 years**
**65 x 12 = 780 months**
**780 x $0.20 (the average cost per month for storage) = $156 per box**
**$156 x 200 boxes = $31,200**

Every year the 200 boxes remain in storage, it is costing the firm another $480 (200 boxes at $0.20 each per month for 12 months). It is conceivable and even likely that the storage costs will have outpaced the profit realized from the case. If the boxes are administrative in nature, there was no profit, only a cost of $31,200. The $0.20 per month average storage cost will need to be adjusted to reflect your firm's average storage cost.

There may be additional costs associated with destroying the box such as perm-out fees, transportation fees and pallet fees. Usually the cost of destroying an archived box is recouped within 3 years (or less) if the contract doesn't have a perm-out penalty. Oftentimes vendors may negotiate lower rates when you undertake a large scale destruction project reducing your direct cost.

New electronic data comes into the firm much faster than data is destroyed. New incoming data sets tend to have a larger electronic footprint than data sets from just a couple of years ago which consistently drives up the need for more and more electronic storage. But that doesn't mean that data destruction is a useless endeavor. Most firms experience a positive trade-off when reducing the volume of electronic data because they are able to defer making the next storage purchase.

It is harder to calculate the cost of electronic data. With hardcopy storage you have a contract to reference and you know how many boxes are in storage. For electronic data it is not as simple; instead you need to work with your IT department to gather more information. The IT department is well aware of the fully loaded price of the various tiers of storage used and should be able to provide a per gigabyte (GB) cost for the storage resources used.

A good place to start calculating the cost of electronic storage is with the firm's premium storage severs. Examples of applications that utilize premium storage are email servers (Microsoft® Exchange), litigation support databases used for eDiscovery, SQL databases and other systems requiring higher performance. Premium storage is usually much more expensive than storage utilized for unstructured data, such as images used for litigation and general network file shares. The cost for premium storage (tier 1) can range from $500 to $1,000 per GB including overhead costs like disaster recovery, daily backups and ongoing maintenance. Lower tier storage costs can range between $2 and $25 per GB.

Calculating the estimated cost savings to the firm can be as easy as creating a simple spreadsheet. The cost will vary depending on what your IT department includes in the per GB charges. An example from one firm is below:

**The average email size (including attachments) is .413 MB.**
**The average document management (DM) system document size is .78 MB.**
**The cost of Exchange storage is $500 per GB.**
**The cost of unstructured data storage is $2 per GB.**

The example in Table 1 tracks the number of GB destroyed in Exchange, shared drives and the document management (DM) systems.

| CALCULATION OF COST SAVINGS FOR DESTROYED DATA | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| MONTH | # OF EMAILS DESTROYED | EMAIL DESTROYED (GB) | EXCHANGE TOTALS | DOCUMENTS DESTROYED IN DM | DM TOTALS (GB) | DESTROYED ON SHARED DRIVES (GB) | UNSTRUCTURED DATA TOTAL (GB) | UNSTRUCTURED DATA STORAGE COST TOTAL |
| October | 1,200,000 | 495.9 | $247,950 | 10,177 | 7.9 | 5.7 | 13.6 | $27.20 |
| November | 1,600,000 | 714.0 | $357,000 | 19,258 | 150.2 | 6.6 | 156.8 | $313.60 |
| December | 380,000 | 159.1 | $79,550 | 4,340 | 33.9 | 2.5 | 36.4 | $72.80 |
| Total to date | 3,180,000 | 1,369.0 | **$684,500** | 33,775 | 192.0 | 14.8 | 206.8 | **$413.60** |

TABLE 1

Ideally your dark data destruction projects will normalize into components of your retention program. When that happens it is still important to track cost savings. While this may not stop IT from having to purchase more storage, strong information governance practices will be a key driver in delaying that cost for your firm.

## APPROACHES TO ELIMINATING DARK DATA

When surveyed, firms represented in the Law Firm Information Governance Symposium reported they had implemented a number of procedures to eliminate the growth of dark data (see Chart 5).

## OPERATING PROCEDURES UTILIZED TO ELIMINATE DARK DATA GROWTH

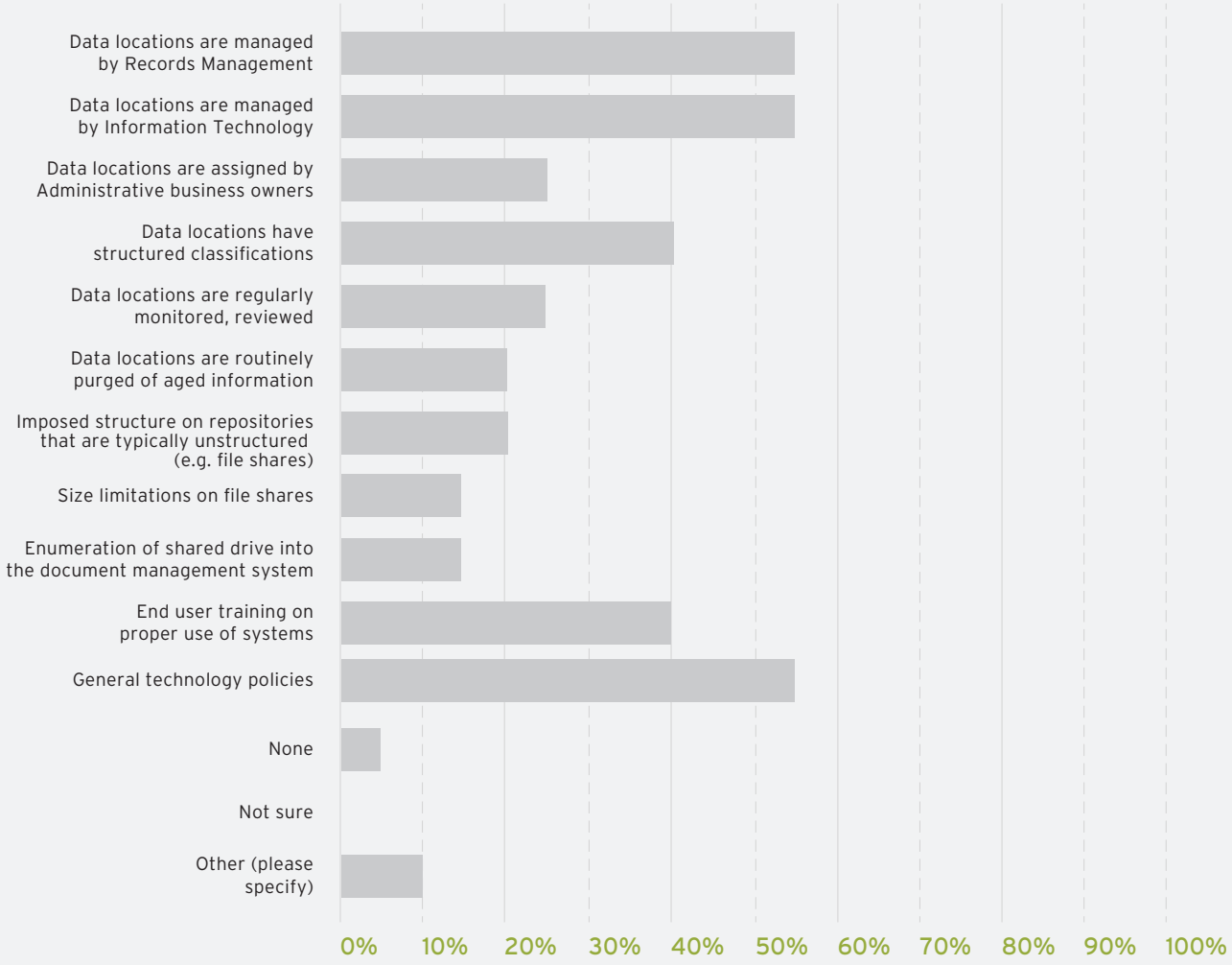



CHART 5

### WHAT HAS NOT WORKED?

It is important to recognize that dark data accumulation is a problem that will not go away by itself. In paper form, and especially in electronic form, dark data can lurk for years without being recognized as a cost or risk problem.

Unlike other projects that may offer their own incentives for compliance, dark data initiatives have different challenges. It can be difficult to entice lawyers to attend training for technology projects involving system upgrades or changes, but the lawyers are incentivized to ask for help or risk struggling with the new system. Conversely, a project to eliminate dark data across the enterprise supplies no real incentive for action for the lawyers. Simply telling people to "deal with it" usually will not work. Most individuals assume it's someone else's problem. Others incorrectly assume the problem will eventually just go away by itself.

## SEVEN STEPS TO IMPLEMENTING MORE AGGRESSIVE FIRM POLICIES

**1. Understand the Problem:** As with any 12-step program, it starts with admitting that you have a problem. The existence of dark data brings risks to the organization that must be understood and addressed. Like any important project, it is critical to get the right sponsorship support from the beginning. While minor projects might be announced by a Records Department or IT Department, those announcements are often overlooked as routine communications that can be easily ignored by the recipients, but C-level sponsorship is a different matter.

**2. Acquire Sponsorship:** As with any major project, you should first identify those individuals in your organization that can serve as project owners and sponsors. In small firms this may be a combination of the General Counsel, Managing Partner, Executive Director, Records Manager and IT Manager. In larger firms it may also include members of the Risk Committee. Be sure your sponsors have the proper authority to set and enforce policy. The last thing you want is to spend valuable time working out the mechanics of a project only to find you don't have the proper authority or management backing to implement it. It may be desirable to form a subcommittee of individuals who have both a vested interest and understanding of the problem, as well as the requisite authority to recommend or implement policy.

**3. Educate:** Start by educating those same key individuals that will serve as sponsors for the project. It is critical that they are educated as to the risk factors involved with dark data, and the necessity of having a program to deal with those risks. Set detailed agendas for meetings, emphasizing the importance of their attendance and participation. Document the education process, and the decisions made by those individuals with written minutes of the meetings. Securing the history of decisions made, as well as when and why they were made will be important for future projects. This education and documentation process builds and maintains the necessary momentum so that you don't have to hear the question, "Why are we here again?" at future meetings.

**4. Communicate:** Establish a communication plan for the project. This should start with a "branding" of the project in a formal template, so that when communications are sent, individuals understand immediately the context of the communication. Establish from the onset that this branded program is being delivered on behalf of the firm with the support of the project sponsors. Remember that initiatives communicated by Records or IT Departments can sometimes be viewed as non-essential communications and ignored. It is imperative that recipients see this initiative as coming from the top.

Secure time in key department meetings to explain the purpose and objectives of the initiative, especially to the lawyers. See that the General Counsel or other Risk Counsel are delivering the project content, with Records or IT personnel available to answer questions as necessary. Be sure to convey that this project is a risk driven initiative and not something dreamed up by "the IT folks" to simply reclaim disk space on network shares.

A regularly published newsletter should be planned and executed. The newsletters should be sent from a distinguished mailbox of the sponsoring group (not Records or IT). This practice will ensure the program is viewed as initiated and supported by the project sponsors from its inception. The initial newsletter should focus on the business drivers for the program as well as establishing the executive stakeholders as sponsors for the project. While cost reduction may be a consideration, other factors should be emphasized including risk reduction and improved client service and relationships. Future issues of the newsletter should detail new policies that may be needed to encourage attorney and staff participation in the project. Those messages should clearly answer the question "How does this impact me?" It is important to first build a broad understanding of why the firm is undertaking the program before asking for a change in the normal process your attorneys and staff are used to.

Also, it is important to ask your committee to create an elevator speech which allows committee members to explain the reasoning and details of the project in 30 seconds or less. This will demonstrate that you are all "singing from the same hymnbook" with respect to why the program is important to the firm. Keep in mind that you will need to create and execute a true communication plan. The plan should not simply consist of one or two emails sent firmwide. Rather it should be a broad program of targeted communications over time across different mediums (email, paper handouts, posters, and presentations at department meetings). A fully implemented communications plan is the only way to generate attention to the program over the long haul.

**5. Get Creative and Provide Incentives:** As with any law firm program, care should be taken to view the project from the perspective of the attorney or staff member you are trying to win over. Try to answer the question, "What's in it for me?" Like any change management initiative, focus on the importance of getting key stakeholders involved at the right level and in the proper roles (sponsor, early adopter, communicator, etc.). Ask your Training or Marketing departments for ideas on how to get the word out and generate enthusiasm and participation. You could offer prizes or other incentives tied to finding and eliminating dark data. The department with the highest percentage of participation in the program could win a pizza party for lunch. Reward those attorneys willing to speak at department meetings in favor of the program with a nice dinner out for themselves and their partner. And finally consider other "gamification" techniques that have demonstrated success for similar projects that otherwise might have elicited a yawn from participants.

**6. Don't Keep Adding to the Problem:** Convince the firm to take early steps to avoid compounding the dark data problem going forward. This is a critical starting point where you need to gain support from your committee to create new policy and procedures to support this initial goal. Hold off on communicating new policy directives until after the early newsletters and other communication that describe the program and goals are distributed. Recipients will understand the need to stop adding to the problem as an important first step and will be prepared for further process changes. Examples of steps to avoid exacerbating the dark data problem include:

- **Add control mechanisms to disallow most individuals from creating new folders on share drives. This simple mechanism is critical as it serves as a gate to avoid future difficulty in establishing who owns data that you wish to destroy.**
- **Establish a control and logging mechanism for all new use of file shares, attributing them to specific client and matter numbers. Establish the owner of each of these file shares, with policies backing up the terms of their data ownership.**
- **Once the data contained in the share drives is linked to client matter numbers, the data can be passed to the document management system and stored there. A view can be created on the share drive to emulate the expected end-user experience.**
- **Adopt new processes to ensure that as individuals leave the firm, their data is assigned to someone else.**
- **IT departments are notorious for "belt and suspenders" approaches to making additional backups of systems, especially just prior to a planned system upgrade. These backups may pose a risk to the firm because they are additional copies, outside of the normal backup, rotation and overwrite/deletion schedule. They may be simply placed on another network share drive or other storage medium and quickly forgotten about. While the reasons for these backups may temporarily justify their existence, it is important that the IT staff understand the risk implications when those copies are not deleted in a timely manner. Work with IT staff to establish protocol for the elimination of these temporary copies as soon as possible.**

» **Integrate Microsoft® SharePoint such that all content resides in or is managed by the document management system.**

**7. Monitor and Adjust:** Rarely is a project plan perfect, so expect to make some adjustments to achieve the desired success over time. Make the required adjustments to the plan as you move through the project with the sponsoring group. As with any initiative, establish a mechanism to monitor progress and report back to your sponsoring committee on a regular basis so that they see the fruits of their labor. Consider including regular nuggets of progress in your newsletters and other communication to end users. Being told of project successes in locating and destroying dark data may serve as a motivating factor for others. Remember to thank and celebrate the accomplishments of those participating in the project in each of their respective roles.

## EMERGING TECHNOLOGIES

At some point, your firm will be faced with the daunting prospect of cleaning up dark data residing in various repositories across your firm. This cleanup process doesn't have to be a manual process. There are sophisticated emerging technologies that can be leveraged to assist with not only cleaning up legacy data, but also managing it going forward.

Gartner characterizes this software category as File Analysis Software (FAS) and has identified twenty plus stand-alone solutions available by both large, well-established vendors, and small vendors. All are designed to analyze a variety of data repositories, including file shares, Microsoft® Exchange, SharePoint, Box®, etc.

Practical uses of FAS include assisting with legacy data cleanup efforts, designing storage management strategies, supplementing system upgrades, migration and retirement projects, augmenting data security and access governance initiatives and identifying and remediating personally identifiable information (PII) and protected health information (PHI). Implementation of FAS can be a critical component of a defensible deletion cleanup strategy.

FAS can be used to analyze, index, search, track and report on electronically stored information (ESI) allowing for action to be taken on the files identified. Detailed file metadata (e.g., creator, last access date, etc.) is used by FAS to help analyze data in ways that simply aren't available in the source systems. This metadata can also be leveraged to assist with making information governance decisions.

A notable feature of most FAS solutions is the ability to quickly and accurately identify duplicate file copies across repositories, which in a typical organization can account for up to 10% of volume. Over the coming years, Gartner predicts these solutions will be incorporated into more comprehensive information governance solutions.

| FAS CAPABILITIES | |
|---|---|
| **FEATURE** | **BENEFIT** |
| Metadata Analysis | File attributes are available that are typically not available with storage management systems, such as creator (e.g. internal user ID), last access date, modified date, etc. |
| Content Awareness | Templates can be created to search (manually or automatically) for specific information such as personally identifiable information (PII) and personal health information (PHI). |
| Tagging & Classification | To aid with legacy file cleanup, files can be tagged individually (or in bulk), allowing for specific action to be taken (delete, declare as record, etc.). |
| Reporting | Valuable reports containing file metadata can be exported to aid the review by others for decision-making. |
| Policy Management | Policies can be created to take specific actions on documents tagged or classified in a specific manner. |
| Remediation | Many systems provide connectors to structured systems allowing for the movement of data from the unstructured repository to a structured repository. |

TABLE 2

## DATA LOSS PREVENTION VS. FILE ANALYSIS SOFTWARE

There is another emerging technology that can be useful in the fight against dark data as well: Data Loss Prevention (DLP) software. At this point in the emerging technologies discussion it makes sense to review the definition for both DLP and FAS.

**Data Loss Prevention** is a mature technology category of solutions designed to help protect information. DLP can detect unauthorized data flows or stop unauthorized flows of sensitive information, depending upon how it is configured. For example, if a user attempted to send an email containing sensitive data externally, or attempted to copy sensitive data to a USB drive, DLP software would stop the action and potentially report the violation to an authority within the company. It varies by the DLP technology used, but in general terms a method of identifying information is applied to an information transport or storage point, and if that information is detected, a policy enforcement capability is invoked. Some DLP solutions provide the ability to scan data at rest in order to identify sensitive content (PII, PHI, etc.) and apply tags.

**File Analysis Software** is a new technology category (Gartner published its first market guide in September 2014), used to provide detailed operational insight into large, unstructured data repositories. A majority of FAS solutions have the ability to scan data at rest to identify sensitive content (PII, PHI, etc.) and apply tags. Additionally, FAS

solutions provide data management controls to classify, alert, report, organize, collect and/or cleanse data. As a new category of software, many IT professionals are not yet familiar with FAS solutions or their usefulness in a security context.

As mentioned, DLP is considered somewhat of a mature market offering in the sense there are established products and a high degree of market adoption amongst organizations concerned with insider threats to data. However, numerous conversations with DLP users reveal areas of challenge where FAS solutions can be complementary to a DLP initiative. First, DLP deployments often suffer end-user backlash because they can slow down or prevent legitimate business activity. The operational insights into data and access rights provided by FAS solutions may provide a superior understanding of end users and the data landscape which could serve to refine DLP policies to more precisely target issues without impacting acceptable user behaviors. Secondly, a DLP solution could involve applying and/or reading classification tags on data. File analysis is complementary because it can audit tags applied and/or apply tags of its own that DLP can read. Item level metadata file analysis and DLP can be used in combination to combat insider threats.

FAS solutions are able to generate huge amounts of information about the target repositories. An important capability of successful FAS systems is how the analyzed information is displayed to the user. If not done correctly via a graphical user interface (GUI), the data can easily be lost. For example, one solution presents file analysis results via a dashboard of pre-defined, easily configurable rule sets that classifies information into one of three categories: redundant, obsolete and trivial (ROT). Figure 1 shows an example of the default (i.e., out of the box) rule set:

**» A file is considered redundant if it is a duplicate of another file.**

**» A file is considered obsolete if it was last accessed or modified five or more years ago.**

**» A file is considered trivial if it is an image, audio, video or system file.**

---

## EXAMPLE OF A FAS SYSTEM USER INTERFACE

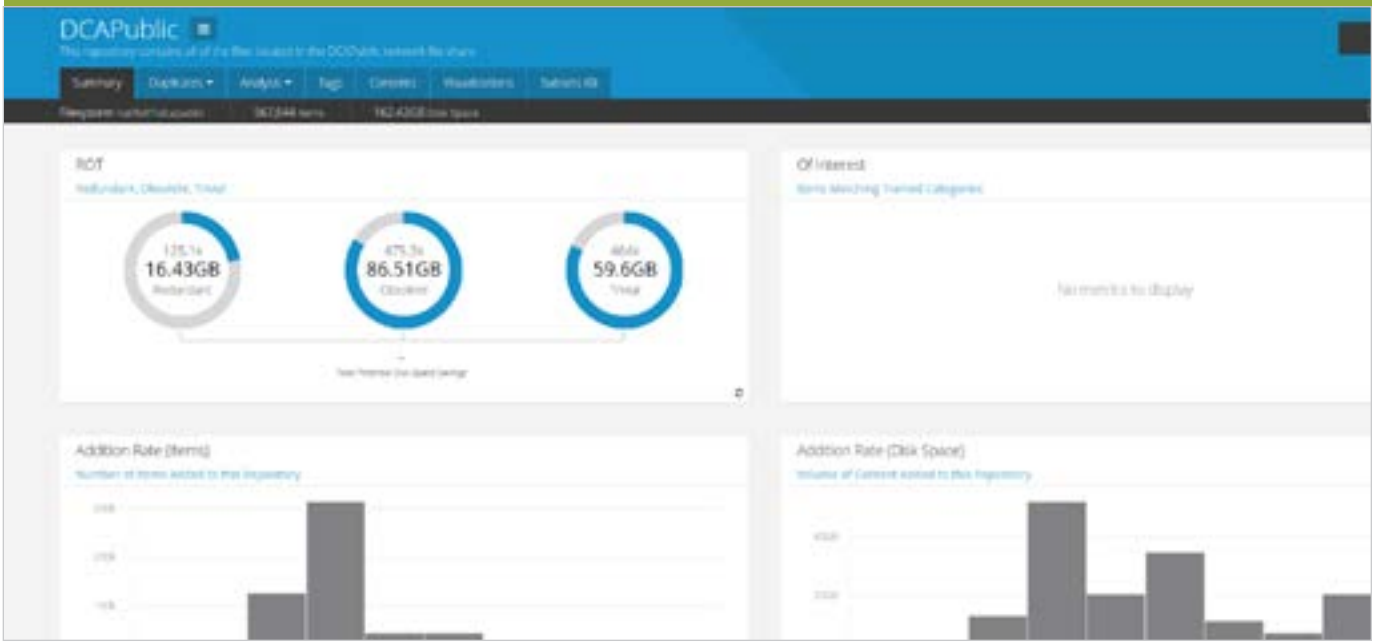



FIGURE 1

In addition to viewing redundant, obsolete and trivial content, additional summary data can be quickly viewed via other dashboard configurations including the 'Of Interest' tile which displays documents that may contain PII and/or PHI and the 'Addition Rate' tile which displays the amount of data added to the repository in each of the past ten years. This provides a sense of how quickly the repository is growing and how old the data is. Users have the ability to drill down further into the data.

## LEGACY DATA CLEANUP

Legacy data cleanup is generally a two-step process. The first step involves the identification and tagging of content for removal, preservation, protection or review. The second step includes an information governance professional creating policies that take action based on the just-applied tags. Below are examples of typical tags and associated configurable actions:

» **Remove: duplicate or old, irrelevant file types**

» **Preserve: business records, vital records, etc.**

» **Protect: confidential information, PII, PHI**

» **Review: HR, Legal, etc.**

Additionally, most FAS solutions can be configured in such a way that allows for a reason to be noted as to why a particular tag was applied. For example, if a file is tagged "remove," the reason may be that the file is a duplicate, convenience copy, superseded or decommissioned.

## MANAGING DATA GOING FORWARD

As the volume of electronic content continues to grow, it is critical for businesses to adopt an information governance program that does not rely exclusively on the manual efforts of users and administrators. Some FAS solutions can be configured to automate the consistent application of policy to content based on the conceptual understanding of information across various file formats. Additionally, various document metadata can be used for policy assignment as well.

It was discovered in the dark data survey that the vast majority of firms grant the Records Management department the primary authority to dispose of dark data (see Chart 6). However, these individuals often lack deep technical skills that would allow them to analyze data repositories using the native administrator tools that come with the solution. Making FAS technology available to Records professionals will allow them to play an active role in the review and analysis of this data.

## PRIMARY ADMINISTRATIVE FUNCTIONAL AREA WITH PRIMARY AUTHORITY FOR DARK DATA DISPOSITION

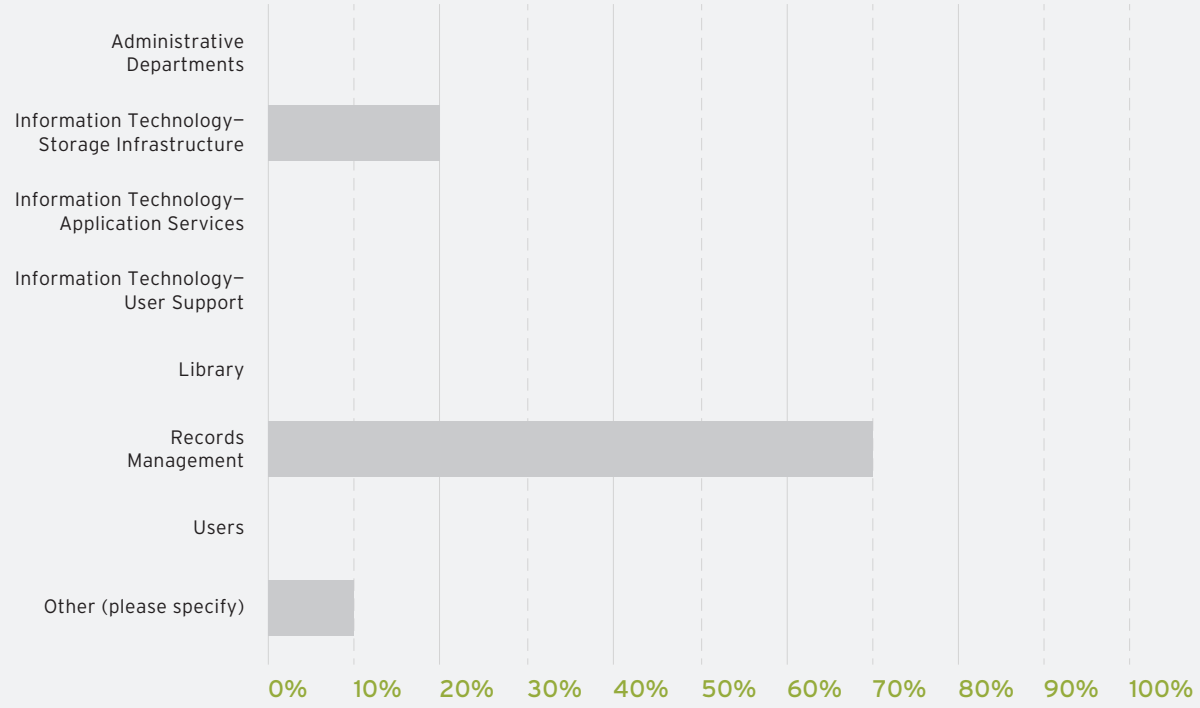



CHART 6

## AUTOMATE POLICY-DRIVEN CLASSIFCATION

Policies can be created using keywords, metadata or example documents using a simple web-based wizard, which is designed for non-technical users who are most familiar with the organizational content and well-versed in laws and regulations. Policy creation is intuitive and its enforcement is automatic.

Many FAS solutions provide automated policy application governing all aspects of the information lifecycle including:

**» Deletion prevention**

**» Storage management**

**» Disposition management**

**» Policy creation with web-based, non-technical dashboards**

**» De-duplication across repositories**

Data categorization is critical to the application of policies. While traditional enterprise content management (ECM) systems rely upon individual users to categorize and tag information, FAS solutions leverage an indexing solution to analyze document content and metadata to categorize and apply policy.

## MANAGE IN-PLACE

Many FAS solutions provide the flexibility to perform specific actions on content no matter where it resides (in-place). These in-place capabilities simplify the management of enterprise content according to business value and lifespan. FAS solution can leverage categories to apply policy into other systems via a connector. These policies can dictate several actions including:

» **Hold**

» **Release hold**

» **Copy**

» **Secure copy**

» **Move (between repositories)**

» **Apply tags**

» **Delete**

» **Declare record**

» **Initiate an automated workflow process**

Further, email that falls under an automatic cleanup rule can be checked to see if it matches a records category before deletion, minimizing the risk of important information being deleted inadvertently if the user has not actively declared it as a record. The efficient management of these information platforms throughout their lifecycle greatly reduces storage and infrastructure costs.

Tackling the growing dark data issue will be challenging to resolve without the capability FAS solutions provide. One theme that seems to be universal among authors that write about developing a defensible deletion strategy is that leveraging technology is an important part of the equation for reducing ESI.

## CLIENT CHARGE BACKS AND WHOLLY OWNED SUBSIDIARIES

More than likely, a majority of the dark data stored on law firm network file shares will be litigation support data that the law firm obtained directly from the client. This ESI will contain client-provided content such as PST files and images of documents produced.

There are two problematic characteristics of this type of ESI. First, these materials are typically extremely large, making the data difficult to manage. Second, holding this data beyond the client's own prescribed retention period can create additional risks for both the client and the law firm. In fact, it is not uncommon today for clients in specific industries, especially heavily regulated industries, to ask their law firms to sign agreements, oftentimes in the form of outside counsel guidelines (OCG), stating that client-supplied materials provided to the law firm in support of a matter will be deleted from the law firm systems within a specified period of time.

A solution that some law firms have identified is to develop a vendor-like wholly owned subsidiary staffed with their own lawyers (typically staff or contract attorneys) and support staff that can assist clients with document collection,

review and timely deletion instead of hiring an outside vendor to perform document collection and review. Below is a short list of benefits a law firm should realize when establishing a wholly owned subsidiary for the collection and review of ESI:

- **A more legally defensible and efficient collection/review strategy. ESI would be collected and handled by appropriately trained professionals that can leverage the most suitable technology, resulting in additional speed and cost savings.**
- **Appropriately trained personnel can be used to more effectively target data collections to ensure that only the most relevant ESI is collected, avoiding over-collection of documents that are irrelevant to the case.**
- **In many instances the wholly owned subsidiary is located on premise of the law firm itself, improving communication and knowledge transfer between review teams and case teams.**

In addition to the benefits described above, perhaps the most important benefit to the law firm information governance professional is that collected client ESI will be stored on the wholly owned subsidiary system and not on a law firm-owned system. This structure provides a solid basis to justify a charge back to the client for storage, thus incentivizing the return or deletion of the ESI as soon as possible.

## SUMMARY

### ACHIEVING BALANCE

Dark data is alive and growing in the form of unmanaged paper and abandoned electronic storage repositories. Given the fact that enterprise data stores are growing at a rate of 60% or more annually,[2] it is likely that some of that will end up as dark data.

How law firms manage the cleanup of dark data will vary from firm to firm. Each will need to determine if the cost of cleanup and ongoing management outweighs the potential risk of doing nothing today and addressing the struggle of managing that data down the road. Luckily, since managing dark data can yield a return on investment from the destruction of both paper and electronic records, the financial investment will pay off.

As dark data is identified and managed, the information governance team can help develop new workflows that make sense, save time, eliminate duplication and facilitate collaboration among teams.

Dark data management best practices involve a three-pronged approach:

- **Define Your Policy: Your information governance team can help in the development of policies that define repository concepts and address compliance concerns like records retention, legal holds, mandated production or destruction and eDiscovery.**
- **Create a Data Map: The creation of a data map will identify where and what document types are being stored. The process followed to create this document will reveal opportunities to reengineer existing workflows.**
- **Begin Data Cleanup: Utilize tools such as FAS to aid in the analysis of data for tagging, making disposition decisions and ongoing information management activities.**

Dark data has been found in some very obscure places. Firms anonymously report finding disconnected servers under the desks of IT personnel. Others report vendors that have approached them with unlabeled discs containing client data from more than 20 years ago in formats that are no longer supported or readable by today's technology. Some reported finding user data on file shares that were originally designated for other purposes. Regardless of where dark data is found it must be dealt with quickly. Many firms are beginning to segregate or delete dark data as they find it (see Chart 7).

## IS DARK DATA ROUTINELY SEGREGATED OR DELETED?

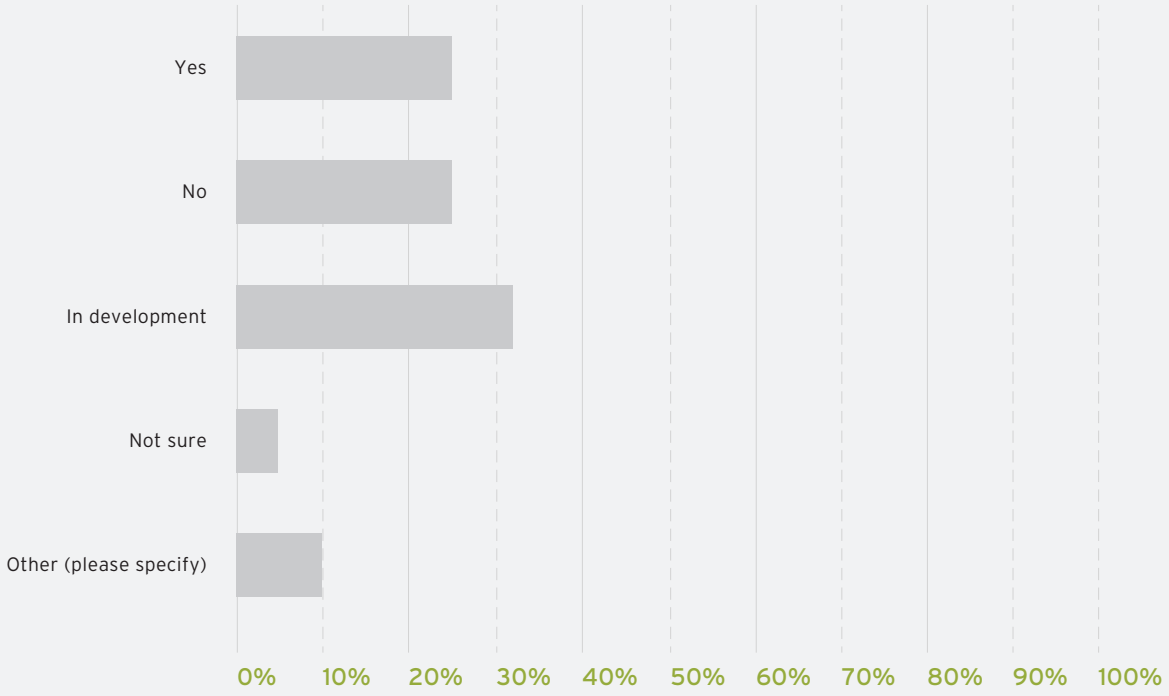



CHART 7

## APPENDIX A: GLOSSARY

| TERM | DEFINITION |
|---|---|
| Big Data | Big data is specific large pools of data against which analytics are run. |
| Classification | A group of records related by common characteristics. |
| Dark Data | Enterprise data that is predominately uncategorized, its content has very limited visibility to the organization if not completely obscured, and because of its obscurity, serves no *apparent* business purpose. |
| Data Map | Relevant detail information (typically location, type, size, responsible individual, retention/back-up schedule, etc.) for all data maintained by the firm. |
| Disposition | The final action taken during the life cycle of a record within the firm, including: destruction, transfer to the client, transfer to third-party (such as another attorney or law firm), temporary release and permanent retention. |
| Documents | In this report, the term is used in the broadest possible context, meaning all paper files, computer files and written, recorded or graphic materials of every kind. All forms of communication of any type, and all other preserved data, regardless of the storage media. |
| Document Management System (DMS) | The use of a computer system and software to store, manage and track electronic documents and electronic images of paper-based documents. |
| Electronically Stored Information (ESI) | Refers to all information stored in computers and storage devices. This includes any data found in electronic documents, email, voicemail, instant and text messages, databases, metadata, digital images and any other type of electronic files. |
| Engagement Letter | A letter sent by the firm to a client that provides the framework for the legal work to be performed as well as establishes the course of communication and interaction throughout the representation. |
| File | A group of documents in any format or media related by subject, activity or transaction, often handled as a unit. |

| | |
|---|---|
| File Analysis Software | Technology used to provide detailed operational insight into large, unstructured data repositories. |
| Form Files and Templates | Records created by practice groups to facilitate their law practices which are often modeled using precedent work product. See also *Knowledge Management*. |
| Information | See the definition for *documents* above. The terms *documents*, *information* and *materials* are used interchangeably in this report. |
| Information Governance Policy | The component of an information governance (IG) program which provides policies and procedures specifying the length of time that an organization's records must be retained. The policy provides for the systematic destruction of records that no longer serve any useful purpose, and is implemented by effecting the destruction of records on a scheduled basis as specified in the legal practice records retention schedule or the firm administrative records retention schedule. The IG policy is one of the firm's major tools for controlling the growth of its records, as well as minimizing legal risks that can be associated with maintaining and destroying firm files. |
| Knowledge Management | Knowledge management is a concept in which the firm consciously and comprehensively gathers, organizes, shares, and analyzes its knowledge in terms of resources, documents, and people skills. |
| Legal Hold | Information in all media formats that must be preserved due to pending or potential claims, litigation, subpoenas, investigation or other legal considerations. |
| Materials | See the definition for *documents* above. The terms *documents*, *information* and *materials* are used interchangeably in this report. |
| Metadata | Data describing the context, content and structure of records and their management through the passage of time. The preservation of the record with its associated metadata is necessary to maintain the integrity of the record. Types of metadata include technical/structural, administrative, descriptive, preservation and use. |
| Non-Premium Storage | Slower, less expensive storage often used for unstructured data such as loose files on network shares and litigation support images. |
| Offsite Storage | A business that provides off-premises records service to the firm through storage, retrieval and disposition of inactive business records and other related tasks through a contractual relationship. |

| | |
|---|---|
| Outside Counsel Guidelines (OCG) | An agreement written by the client defining the terms by which they will enter into a business relationship with a law firm. These often include references to data security, privacy, retention/ disposition, conflict clearance and representation of competitors. |
| Premium Storage | Faster, more expensive storage often used for databases and other systems requiring high performance. |
| Record | A record is information created, received and maintained as evidence by an organization or person in the transaction of business, or in the pursuance of legal obligations, whether paper or electronic. |
| Records Retention | The act of maintaining or holding records for future use, often under policies and procedures of a formally established information governance program. |
| Records Retention Schedule | A comprehensive list of records series, indicating for each the length of time it is to be maintained and its disposition. The firm maintains separate schedules for legal practice records and firm administrative records. |
| Safe | A fire-resistive and highly secure enclosure used for the protection of critical client or firm documents. Also known as a "vault." |
| Vault | A fire-resistive and highly secure enclosure used for the protection of critical client or firm documents. Also known as a "safe." |
| Vital Records | Any record that must receive the highest level of protection because of its necessity to protect the interests of the client, attorney or the firm. Vital records are always stored in a vault or safe. |
| Workflow | A workflow management system is a computer system that manages and defines a series of tasks within an organization to produce a final outcome or outcomes. At each stage in the workflow, one individual or group is responsible for a specific task. Once the task is complete, the workflow software ensures that the individuals responsible for the next task are notified and receive the data they need to execute their stage of the process. Workflow management systems also automate redundant tasks and ensure uncompleted tasks are followed up. Workflow management systems may control automated processes in addition to replacing paper forms and manual processes. |
| Workspace | A "virtual file" in the DMS where electronic records, including emails, are catalogued in relation to a single matter. |

## APPENDIX B: SAMPLE CHECKLIST FOR THE COLLECTION OF DARK DATA

(Collection could be related to a project such as routine retention/disposition, network cleanup, or a court ordered destruction notice, etc.)

Client Matter Name: _____

Client Matter Number: _____

**Relevant Custodian:** _____

*Relevant custodian/individual no longer with the firm:* ☐

### PHYSICAL FILES/HARD COPY RECORDS

Generate electronic file index of physical files.

1. Identify and send notice to records custodians to collect all relevant materials
2. Review workrooms and unmanaged file areas
3. Collect relevant loose materials and un-barcoded files
4. Migrate any loose filing to the proper client/matter file; barcode/track files
5. Update the electronic file index and location of relevant physical files/paper records
6. Review offsite storage:
   - **a)** Chronological files
   - **b)** Unidentified boxes
   - **c)** Files transferred to the firm with laterals
7. Attorney home offices

### ELECTRONICALLY STORED INFORMATION (ESI)

1. Data on IT system(s) - search for relevant information
2. Generate a data map or locate the existing data map
3. Identify key terms
4. Identify systems to be searched
   - **a)** Use FAS technology to search the network to identify dark data
   - **b)** Develop list from technology section of outline
5. Migrate electronic data to appropriate repository
   - **a)** Provide users instruction on searching for relevant data and how to migrate data to approved repositories
   - **b)** Email boxes
   - **c)** Hard drives
   - **d)** Network file shares
   - **e)** Document management system for items not profiled under the appropriate client matter
   - **f)** Departmental or special use applications
6. Collect external media: external drives, USB, CD/DVDs, etc.
7. Collect relevant LITSUP evidence: collected laptops, hard drives, etc.
   - **a)** Search litigation support repositories: review platforms, LITSUP network drives and hosted sites
8. Review case data maps and chain of custody logs
9. Review third party vendor relationships and manage relevant data
10. Manage departed user mail boxes and extract relevant data
    - **a)** Request IT to un-archive relevant email box or provide PST for departed users if applicable
    - **b)** Consider using LITSUP review tools to search departed user mail boxes

## COMPLIANCE TRACKING

1. Verify relevant information is not subject to any litigation holds
   - **a)** If data is being collected for a preservation copy related to a litigation hold:
     - **i.** Document the preservation process/retain information
     - **ii.** Update relevant tracking systems
   - **b)** If data is being collected as part of a client file release:
     - **i.** Delete data on firm servers 60 days after release to client
2. Confirm that the matter is closed and has reached the end of the retention period
3. Confirm request for destruction has been routed to risk management and practice group leaders for review and destruction approval
4. Confirm relevant ESI & physical records are destroyed
   - **a)** Verify relevant data is purged from IT systems
   - **b)** Verify custodian or records administrator has destroyed ESI and/or physical records in compliance with the destruction order or retention schedule
   - **c)** Maintain documentation relating to compliance with the retention schedule or destruction order
5. Update relevant tracking systems with final disposition
6. Save destruction certificate or other supporting details including this checklist in the matter engagement file

Destruction verified: ☐ Yes ☐ No ☐ N/A

Preservation collection completed: ☐ Yes ☐ No ☐ N/A

Date completed: _____ By: _____

Action taken/notes: _____

_____

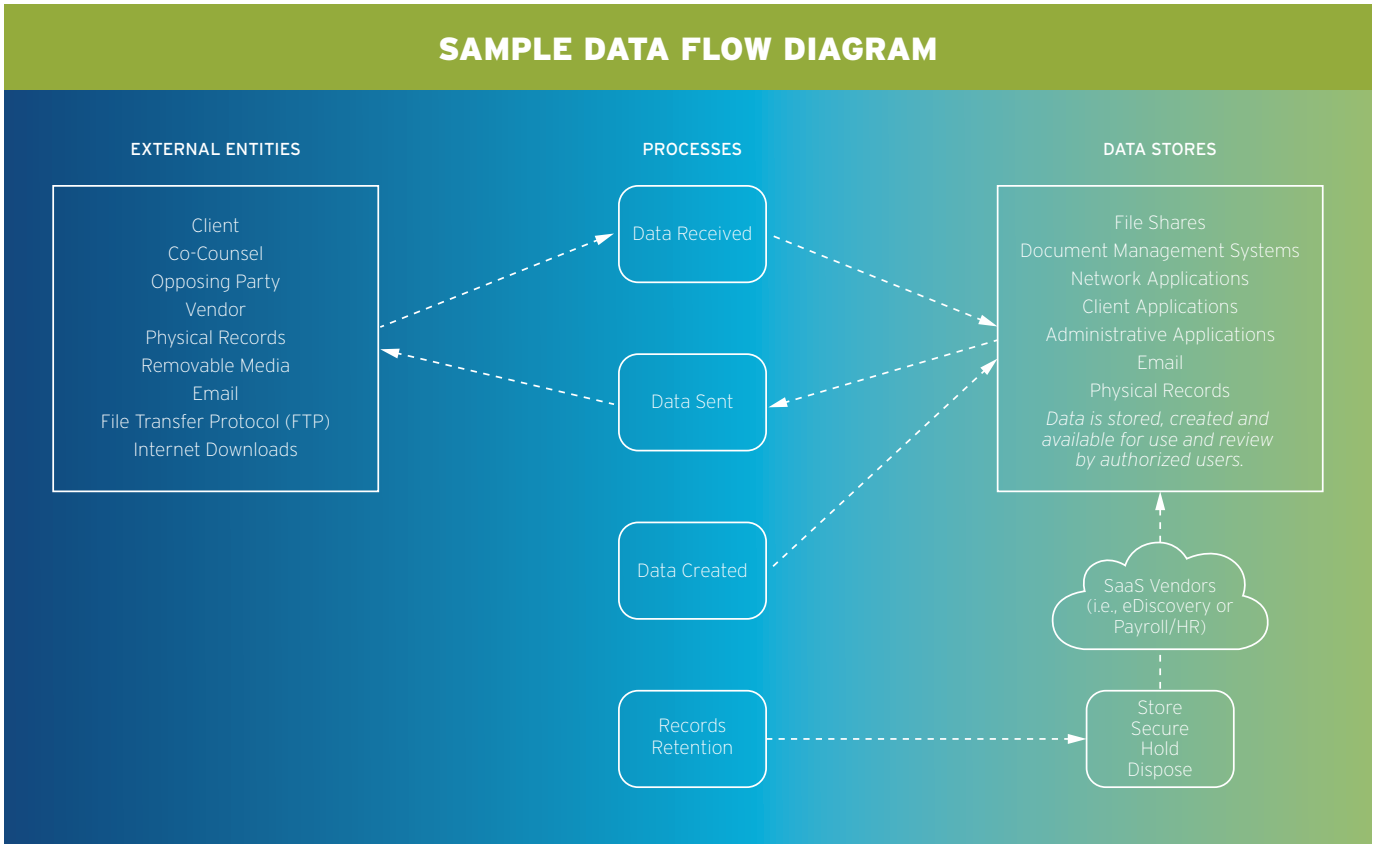
_____

## APPENDIX C: SAMPLE DATA MAP

### SYSTEMS AND PHYSICAL LOCATIONS TO LOOK FOR ESI AND HARD COPY DOCUMENTS

1. ESI stored in the firm's document management system.
2. ESI stored in the firm's email system (and messages downloaded to PST or OST files).
3. ESI stored in the firm's archive and/or backup systems.
4. ESI stored in the firm's training & development environments.
5. ESI stored on the firm's network drives.
6. ESI stored in departmental applications.
7. ESI stored in litigation support databases.
8. ESI stored on a custodian's personal computer.
9. ESI stored in a cloud environment, including drop boxes and deal rooms.
10. ESI stored on external storage devices.
11. ESI stored on cell phones, tablets, BlackBerry® and PalmPilot® devices and other similar portable digital devices.
12. Paper documents stored by the firm at offsite locations.
13. Paper documents stored onsite by the firm.
14. Paper documents in custodians' offices at the firm and at home.

## APPENDIX D: SAMPLE DATA FLOW DIAGRAM




As data is saved to the network in the appropriate firm repository, metadata and security should be defined. Data may be shared for collaboration by users who have a need to know with the appropriate security and encryption applied, as applicable, based on business and client requirements.

## APPENDIX E: SAMPLE ELEVATOR PITCH

**Question:** What is all this business about dark data I've been receiving notices about? What does all of it mean?

**Elevator speech answer:** Oh, yes the dark data project. I'm sure that you've seen some of the important communications from <insert name of important firm sponsorship>. Essentially the goal is to reduce risk to the firm related to data that we have in many places that we didn't know we had, or should be better categorized or deleted. As you can imagine, this could be a real problem for the firm as it relates to things like subpoenas, litigation holds and matter transfers, as well the associated cost for the firm to manage that data. There is also revenue-generating information such as business trends that could be buried in the information that we are not acting upon because no one knows that it is there. We appreciate your understanding and support of the project. Please keep reading the related newsletters and other communications on how you can help!

## REFERENCES


1 ARMA International. (2007). *Glossary of Records and Information Management Terms* (3rd ed.). Lenexa, KS: ARMA International.

2 IDC. (2012, March). Worldwide Big Data Technology and Services 2012 – 2015 Forecast. Retrieved from http://ec.europa.eu/digital-agenda/en/news/worldwide-big-data-technology-and-services-2012-2015-forecast


## BIBLIOGRAPHY


Michels, M. (2014, July 1). *LegalTech Quest Into Heart of Darkness: How to Balance Dark Data's Liability Risks and Information Profit.* Legaltech News. Retrieved from http://www.behblaw.com/Hidden-Pages/LegalTech-Quest-Into-Heart-of-Darkness-_-Law-Technology-News.pdf

Robek, M. F., Brown, M. F. & Stephens, D. O. (1998). *Information and Records Management.* New York, NY: McGraw Hill.

**ABOUT IRON MOUNTAIN**

Iron Mountain Incorporated (NYSE: IRM) provides information management services that help organizations lower the costs, risks and inefficiencies of managing their physical and digital data. Founded in 1951, Iron Mountain manages billions of information assets, including backup and archival data, electronic records, document imaging, business records, secure shredding, and more, for organizations around the world. Visit the company website at www.ironmountain.com for more information.



US-LAW-EXT-BR-080815-001