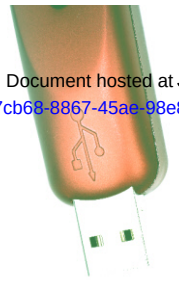


A SPECIAL REPORT ON THE IMPLICATIONS OF

Electronic File Metadata

by

Richard E. Davis, J.D.

HOW IT FACILITATES

- Litigation Discovery
- Records Retention Program Implementation
- Reporting for Compliance & Information Security Auditing
- Dormant Data Liability Risk Mitigation

- A COMPREHENSIVE OVERVIEW OF ELECTRONIC FILE METADATA, IT'S ORIGINS AND TYPES.
- USE CASES FOR CORPORATE DATA MANAGEMENT.
- HOW COUNSEL CAN LEVERAGE METADATA FOR DISCOVERY.

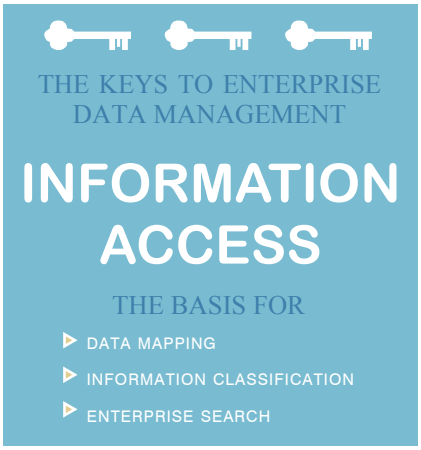
Introduction

"Metadata is to electronic files what a candy wrapper is to candy. It can tell you a lot about a file's flavor, brand, quality and ingredients, well before you even open it."

Corporate information management can be a very complex proposition. It often involves different corporate stakeholders who view and use information in different ways based on a variety of business needs, legal or regulatory requirements. In many instances the way in which people use corporate information can cause conflicts related to organizational policy, technology capacity and profit center activity. For example, a data pack rat (a user common to many organizations), may be a stellar sales person, new drug innovator or rainmaker. Moreover, there may even be a direct correlation between their hoarding copious amounts of information and their success.

To an IT manager, whose focus is data protection and availability, much of the packrat / rainmaker's data may appear to be duplicative and difficult to manage. IT Managers whose responsibilities include supporting users, enhancing network performance, managing data storage and disaster recovery are increasingly challenged by "unmanaged" data that accumulates in inaccessible data storage points throughout the enterprise. This situation is played out in virtually every corporation in America. Not only is it an obstacle to efficiency, it is one of the most significant information risk management hurdles corporate entities face today. Given the demands of managing information in today's corporate environment, some organizations have mandated greater interdepartmental coordination of data management efforts. The result is that corporate departments such as legal, HR, finance, research and records management, are now working together with unprecedented synergy on risk management initiatives that include stemming unmanaged data proliferation, reduction of dormant data liability and the protection of intellectual property.

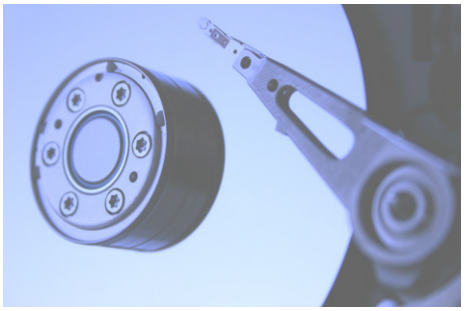
Recognizing and defining the problem is critical to reigning in risk, but it is only the first step. Coming to terms with the fact that ones organization lacks the ability to access or discover the existence of disparate information content that may or may not be impact the organization's viability is difficult enough to deal with. However from the perspective of IT and associated personnel, the ability to deal with the situation only becomes more complex when the politics of information management rears its head.



Despite consensus among stakeholders that something must be done to address information risk, questions such as who in the organization will pick up the gauntlet, own the initiative, champion the solution and educate the masses often creates what the author refers to as corporate "administrative initiative inertia" (AII). Like

ENTERPRISE INFORMATION GOVERNANCE BEGINS WITH BASIC INSIGHT...

THAT YIELDS BENEFITS FOR ALL CORPORATE STAKEHOLDERS



many conditions characterized by three letter acronyms, All generally continues until within an organization until the occurrence of some non-discretionary compelling event that requires the said organization to immediately access germane information within the confines of its firewalls.

As many house-legal and IT departments that I have worked with can attest to, reactive, event driven occurrences are the impetus for most enterprise information management and control initiatives. When an organization is in the red zone of reactive discovery, it really doesn't matter who owns the initiative anymore. At this point, All has been supplanted with replaced with CR-COCFD (costly, risk compounding, operationally disruptive corporate fire drill), a condition that requires seemingly bottomless budget.

Regardless of whether an organization suffers from administrative inertia or it finds itself in a state of reactive discovery the practical reality is that they can avail themselves of powerful, cost effective technologies that will help them address all their reactive discovery workflow, proactive storage management, records and data risk management concerns that are at the heart of the enterprise information management quandary.

Early Stage. Discovery Cost Containment Tip

- Get your client to buy in to their data preservation obligations at the outset of the litigation event.
- Leverage cost effective technology to give you a sense of the size, scope and overall discovery cost - you'll reduce client sticker shock.
- Seek stipulations with other parties as to the form of responsive data production early in the process.

Cost Containment Tip # 1

The Key to Managing Electronic Information - Harnessing Metadata

The key to developing the solution to the quandary lies in the understanding and management of metadata across the enterprise.

Metadata has been defined as:

> "Information about data."
>
> "Information about information."
>
> "Structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities."

Basic examples of email and document metadata field information include the following:

THE KEYS TO ENTERPRISE DATA MANAGEMENT

DATA MAPPING

ESTABLISHES:

- KEY DISCLOSURE INFORMATION FOR LITIGATION DISCOVERY
- INDISPENSABLE INFORMATION FOR RECORDS MANAGEMENT PROGRAMS
- CRITERIA FOR IT INFORMATION MANAGEMENT POLICIES
- INSIGHT FOR INFORMATION SECURITY & INTERNAL INVESTIGATION STRATEGY
- PROVIDES INFORMATION FOR ENTERPRISE INFORMATION TAXONOMY

EMAIL METADATA:
TO:
FROM:
CC:
BCC:
MESSAGE ID:
SUBJECT:
SEND TIME:
RECEIVE TIME:

FILE METADATA:
AUTHOR:
FILE NAME:
CREATE DATE:
MODIFY DATE:
FILE OWNER:

In the early days of computer based electronic file (e-file) management, e-file metadata was simply been a way to look at information attributes in ways that allowed systems engineers to validate data migrations, audit data flow and provide recovery checkpoints for system availability and accountability. A few short years ago, "electronic file metadata" was a concept used almost exclusively by individuals who took pride in wearing the latest pocket protectors to cocktail parties. Today however, the concept of electronic file metadata has become so significant, that it has now been codified as part of the legal lexicon. Metadata has many uses. In the context of litigation, it provides lawyers with the basis to substantiate or challenge an evidentiary chain of custody and the authenticity of data that has been identified or produced pursuant to discovery requests. It can be used to facilitate records retention management by allowing files to be classified by ownership (on individual or departmental levels), age or any other available characteristics for archival or other disposition. Metadata attributes allow IT departments to gather critical data proliferation statistics, groom and de-duplicate data for storage management purposes. In short, there are a host of organizational stakeholders that can benefit from using metadata reports. In the context of litigation, electronic file metadata or attributes are commonly used to create timelines detailing se-

[1] Often referred to as system, application and user metadata.

quences of relevant events or establish the identities of the parties who created, edited or accessed responsive documents. It can also be used to provide information about a file's contents, version or edit history.

In the context of the ESI (electronically stored information) disclosure requirements imposed by the Federal Rules of Civil Procedure, and the increasingly wide range of electronic file types and associated metadata attributes that are available for production in litigation, it is of paramount importance for attorney's and their clients to understand the key distinctions between operating system, application software and user created metadata[1]. From a discovery perspective, the implications of metadata's existence or its non-existence can be extremely significant. For example, if an electronic file production includes information about file creation dates, but nothing about the last modification date, a red flag should go up leading to an inquiry as to why the latter date does not exist, unless of course it is not relevant or the parties have stipulated that this information would not be included in the production.

The bottom line is, counsel with an appreciation of a client's information management infrastructure, file types and metadata subtleties will be adequately prepared for negotiating the appropriate protective order terms, discovery requests or selecting document processing or file review methods. They may also gain a significant advantage over adversaries that lack the requisite understanding of metadata or appreciation of its implications.

Corporations that have effective electronic file management systems or processes, generally have an implicit understanding of how to use metadata. They generally leverage metadata for business purposes (profit center activities) by using it in business intelligence activities and functions like data analytics, data mining, sales campaigns, etc. They generally tend to be more effective at managing and facilitating discovery for initiatives like litigation, internal investigations, compliance, records retention with greater cost effectiveness and efficiency than those that don't.

Early Stage. Discovery Cost Containment Tip

- Develop a data map as early as possible to help you identify places where potentially responsive information may exist...
- This will result in litigation hold procedures that are more streamlined and defensible.
- It will help minimize the risk of spoliation.

Cost Containment Tip # 2

Data - Mapping; Visualizing Enterprise Information With Metadata

When it comes to the requirements for technology and systems that can be deployed for information management initiatives such as compliance auditing, discretionary internal investigations, records retention and litigation discovery, the similarities from a requirements perspective are astonishing. For example, compliance driven records management and complex litigation both require the ability to access disparate content, report on it, search and classify the information. In the context of litigation, one of the first orders of business is to develop a strategy for **litigation holds** and **information gathering**. These precursor stages to the **analysis**, **review** and **production** of responsive information, require one to know about the existence of information whose content meets the hold requirement. Once this step has been accomplished, it one must have a method to **aggregate** and **defensibly secure** it. Hence the need, particularly when dealing with large networks on which data is dispersed, to develop a "data map[2]." The data mapping process serves many purposes. It allows counsel to become familiar with their client's IT infrastructure as well as provides them with the knowledge about potentially responsive data content and its location within the organization. It also helps with the formulation of collection strategies in that it provides them with the requisite information choose the appropriate collection methodology.

From a Federal Rules perspective, the data mapping process provides counsel with the informa-

""Metadata" - Never has a single word struck such fear or brought such giddiness into the hearts of attorneys everywhere. To some it means enormous amounts of additional review time, thus driving up the cost of the litigation exponentially. To others it represents a magic bullet, where a single, hidden date or detail will miraculously appear to spectacularly win a case. It depends on one's point of view."

"From a lawyer's perspective, there is no better indicia for the purposes of evidentiary authentication than electronic file metadata. "

- Scott Fischer of Array Technologies, on metadata:

[2] There is not set approach to data mapping. It can consist of a network topology with references to data locations, types and volumes or it can be as simple as a spreadsheet with columns indicating the same.

[3] The LITIGATION LOGISTICS Litigation Data Lifecycle model consists of several steps that relate to the efficient management the electronic discovery process. For more information contact the author.

tion necessary to comply with the FRCP 26(a) initial disclosure requirements. These steps can also allow counsel to vet and assess their adversary's disclosures. From a consulting perspective, Litigation Logistics uses a process of infrastructure analysis (IA) which results in a matrix describing:

1. The technical infrastructure of the organization.
2. Personnel hierarchy.
3. Data management polices (discretionary & non-discretionary).
4. Organizational culture.

Loosely defined, IA is a survey that gathers information about a target organization's information systems, retention policies, data storage methods etc., which includes a metadata driven graphical representation of physical and logical devices on which data is stored. This initial phase of analysis sets the stage for every subsequent step in the Litigation Data Lifecycle[3]. This is information that can help counsel with negotiating discovery scope pursuant to FRCP 26(b) at the meet and confer stage, in addition to facilitating the mandated disclosures and negotiations. Early IA helps shape discovery strategy and can have a significant impact on litigation budgets and cost containment.

Other concerns that can be addressed by early IA are privileged document criteria. Getting a handle on the identification and classification of privileged documents early, helps mitigate the risk of inadvertent production of privileged data. It is important to note that assessing the infrastructure of organizations with complex information management platforms will be extremely difficult at best without the involvement and assistance of the corporate IT department or an expert consultant in the field.

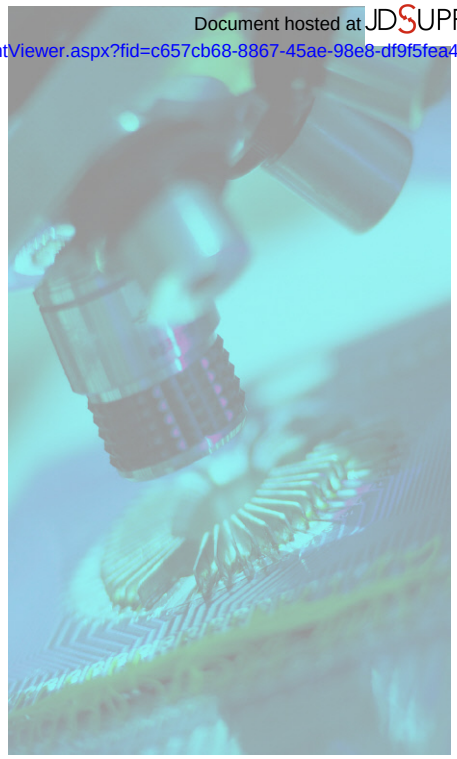
In summary, the goal of the IA is to:

1. Help build a "data map" consisting of data locations, types, ownership and other relevant attributes for the matter in question.
2. Help facilitate mandatory disclosures and provide ample information for meet and confers or other negotiations.
3. Facilitate a plan as to budgeting, collection strategy and methodology.

The Sources; How System Components Create Metadata

To understand when, how & where metadata originates, it is often helpful to think of metadata as being created by the interaction of computer system components in "layers." At the most basic level, most corporate information systems are comprised of the following common computer system components:

1. NETWORKED COMPUTER HARDWARE (ONE OR MORE PC'S / SERVERS/LAPTOPS) – these devices will provide "temporary" (RAM, or computer memory) and more permanent storage devices such as hard disk drives, CD/DVD-ROMS, USB devices, etc. where files reside. One will generally find a physical device on literally every desk in a corporation. The operating system, described below, controls and communicates directly with the computer hardware.
2. ONE OR MORE COMPUTER OPERATING SYSTEMS (SUCH AS WINDOWS XP) – common operating systems include MS-Windows XP, Windows 2000, Linux, etc. Many PDA's run scaled down versions of the Windows operating system. Application software runs in the operating system environment.
3. ONE OR MORE TYPES OF APPLICATION SOFTWARE (SUCH AS MS-WORD) – application software is used to create ESI. Examples include MS-Word, Outlook, Excel, etc. Computer operating systems take application software instructions and cause the hardware to perform the requisite actions, i.e. allocate space on a hard drive for a new file, write files to a CD ROM, etc.
4. USER OR PREPROGRAMMED INSTRUCTIONS – creating new files, deleting, editing or accessing existing files or running automated programs constitute giving instructions to a computer system.



How System Interaction Results in Different Metadata Types

Conceptually, the various system components or elements described in the preceding section interact with each other in layers as follows:

1. Hardware / Operating System Layer; Layer 1: this layer consists of the interaction between the computer hardware and the operating system. File and folder permissions, file ownership metadata are created here, for example, a domain administrator (IT person) uses the operating system administrative interface to give various users access to certain network locations (physical devices somewhere in the in the enterprise) to create or modify data
2. Operating System / Software Application Layer; Layer 2: this layer is comprised of the computer operating system interacting with the computer application software or proprietary program which runs in the operating system environment as application software. System metadata results from this layer. For example, an accountant in the finance department creates a new spreadsheet in MS-Excel. The metadata of the spreadsheet contains metadata fields of create time, modify time and last access

time and may contain author, title, subject, manager and company - the organization has control over the degree to which certain metadata is automatically captured. The three time fields are generated by Layer 1 and the latter five fields are generated by Layer 2.

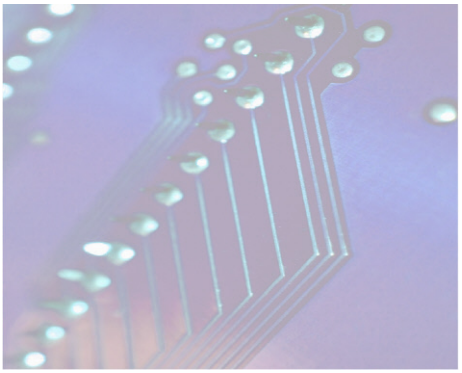
3. Application / Instruction Layer (User defined metadata) - Layer 3; this layer is comprised of the application software interaction (receiving instructions from) with an end user. The result is most often exemplified by a "data classification scheme." Processes in this layer are often characterized by the end user manually "tagging" documents with subjective criteria or creating a rule to auto-tag documents responsive to a particular set of criteria. For example, a pharmaceutical technician gathering clinical trial data may search multiple network shared directories and copy the data to a specific folder called "clinical trial data." The fact that the files are now in the target location creates de facto "source folder" metadata that will be captured at the time metadata extraction is conducted.

4. OLE Embedded Object Metadata: This metadata is derived from "embedded files." For example, if one inserted an MS-Excel or spreadsheet into a MS-Word document, the linked document possesses metadata information that is different than that of the parent document.

As the various computer systems layers interact, different electronic file metadata attributes are created. It is important to note that different application software or proprietary programs may generate file metadata attributes that are common to all electronic files, i.e. - create / save / access date and time, as well metadata attributes that are unique to the application software or proprietary program. Despite the fact that a significant amount of metadata can be generated from a collection of documents, only a relatively small portion of electronic file attributes are actually useful in litigation or investigations.

The distinctions between metadata types and origins described above are often melded into one overarching definition of "metadata, which causes much consternation to the metadata experts who appreciate metadata subtleties, distinctions and implications. It is important to note that an overly inclusive definition of metadata, particularly when discussing and negotiating document productions, can lead to problems between parties, especially when one party is more sophisticated than the other. As one might imagine, the result of this potential disconnect may preclude a meeting of counsel's minds, increase associated client disbursement costs and tax the resources of the courts which must now deal with a plethora motions related to defining what constitutes relevant metadata.

Thinking of metadata in the categories described in the next section will help crystallize the concept of metadata types in terms of "neutral metadata," which is non substantive in nature and "non-neutral metadata" which is content substantive. The distinctions are as follows; non substantive metadata won't yield information about a document's privilege status or responsiveness to a particular issue. Substantive metadata can provide information that can result in a document being deemed privileged or responsive to a particular issue.



These categories help provide counsel with a conceptual framework of common metadata types in a way that helps the unfamiliar become conversant fairly quickly.

Characteristics of Layer 1 Type Metadata

In most modern computer environments, systems administrators can control user permissions that give them access to files, processes, programs and hardware. Users are defined to their computer environments (authenticated) and their systems rights are maintained in access control entries or access control lists (ACE or ACL). These are simple tables that tell the system who you are and what rights you have to create files, where (which physical & logical hardware device) you can store them and who can access them.

This security framework can help provide file or folder ownership as well as a frequency of access audit trail. This type of Layer 1 metadata can provide counsel with important custodial information about who in an organization has or had access to certain files or the contents of certain folders. File and folder security metadata is extrinsic to most electronic files and is often overlooked in discovery. It nevertheless can provide very valuable information about work flow as it relates to individual custodians or groups in an organization. Depending on the nature of a controversy, this type of file and directory security information may fall into the category of either neutral metadata (the kind that is generally not privileged) or non-neutral metadata (the kind that could potentially be privileged). Neutral metadata might consist of information such as, create

[4] There are many ways to change this metadata using MS-DOS commands or specialized software programs.

[3] The LITIGATION LOGISTICS Litigation Data Lifecycle model consists of several steps that relate to the efficient management the electronic discovery process. For more information contact the author.

ELECTRONIC FILE METADATA PARADIGMS CAN BE EXTENDED TO ANY KIND OF OBJECT

IF AN "OBJECT" HAS CHARACTERISTICS THAT CAN BE REPRESENTED AS DATA ELEMENTS

data or last modify date. Non neutral metadata could be considered to be a comment embedded in a file that says "I sent this document to counsel and they will get back to us." The latter is an example of non-neutral metadata.

Characteristics of Layer 2 Type Metadata

When a user creates a new blank MS-Word document, the OS (MS Windows XP, Linux, Unix) provides the file certain metadata attributes such as create, modify and access date and time. This type of information falls into the category of neutral metadata. This kind of external metadata, the kind that is there for the world to see, can also be looked at from the perspective that it is a "candy wrapper", (the candy being the substantive data in the native electronic file which is created by the application software). Some key points to consider about operating system metadata are the following:

1. Barring user intervention, operating system generated metadata, is created and managed by the OS. This type of metadata is related to electronic file and directory attributes such as: create date/time, access date/time, last written date/access time. With computer systems that are "on-line", this type of file and directory level metadata information is very easy to obtain.

2. Most computer users do not have the ability to change this information (nor as a matter of best practices should they be able to), although it is possible for sophisticated users to do so using a program like Attribute Magic™.

This type of metadata is generally objective in nature, and is generally not considered to be privileged in the common legal context of the word. Even so, however, there are situations where it could be deemed as substantive.

Characteristics of Layer 3 Type Metadata

Application software generated metadata (MS-Word, Excel, etc.) is much more complex and substantively rich. In addition to incorporating various and sundry Layer 1 & 2 type metadata attributes, such as "auto-date" information or network card MAC addresses, Layer 3 type metadata may also include user specified information such as author, organization, title or business unit. The distinguishing characteristic of this type of metadata is that it results from the interaction between application software, the user environment (network and technical infrastructure) and the user. This type of metadata presents the most concern in terms of inadvertent privileged information disclosure. User created substantive items may be included in a file such as tracked edits, comments and other information. This type of information is non-neutral metadata. In keeping with the prior analogy for operating system metadata, this type of metadata is akin to goods in a non-see through container, it's not there for the world to see. In some circumstances disclosure of this information might be intended, but in others it might not be intended. The obvious issue is that the type of information that it contains may be damaging, but if counsel is aware of the fact that this type of information can exist in their client's environment, contingencies and safeguards can be implemented to avoid unintended production.

Q: Is it possible to extract metadata from voice recordings captured by phone systems in a defensible fashion and in way that an attorney can use easily?

A: Yes. Not only can one capture metadata, such as the phone numbers of the parties, time and date of the conversations, one can "auto transcribe" the recordings into searchable, annotatable text files. The metadata captured from voice recordings, especially from VOIP systems can help with communications link analysis.

What the Metadata Says About Electronic Files

While metadata is just "information about electronic files," its evidentiary value gives credence to the fact that counsel needs to have thorough understanding of the target organization's electronic document creation, management and disposition environment[6] and relevant procedures. In addition to understanding the target organizations work product creation methods, counsel should have familiarity with the electronic information collection, review and production processes employed by vendors and consultants, as these processes have a bearing on the issues of

[6] See amendments to Federal Rule of Civil Procedure 26(a)(1)(B) concerning initial disclosures of "…electronically stored information…" and 26(f) concerning the affirmative duty of disclosure.

chain of custody and spoliation. This understanding should cover the information lifecycle management procedures as they exist in the "ordinary course of business" as well as incident based protocols. For example, many vendors use electronic file processing methods designed to render electronic files into formats that can be reviewed rapidly by counsel. When these processes are applied without oversight, they can result in significant excess expenditures and create situations where potentially privileged information may be overlooked by counsel.

Electronic discovery involving state of the art or legacy information systems will often require 3rd party e-discovery vendors to extract, convert and prepare electronic documents for review. This is often the case because responsive information may be cumbersome or impossible to review in its native state. Reviewing the contents of the backup tapes that have responsive information (because one's client did not have a proper litigation hold in place), can be a very difficult and costly process. Under certain circumstances, counsel may request that a vendor convert electronic files as hardcopy for attorney review or as well as generate static document images in TIF or PDF format that are devoid of file metadata. Where responsive documents are reviewed natively, the files can be subjected to a process that "scrubs" them of metadata. This is very important because when confronted with large document collections, the parties may opt for an on-line review of the discovery documents and share the costs. Scrubbing the native files of their metadata allows counsel to pick and choose what metadata will be available for review both by their own review teams or their adversary. A commonly

used electronic file review method involves converting the files to static images and extracting their metadata. This process results in the counsel having access to:

1. The original native files.
2. Static images of the native files (usually as multi-page or single page TIF or PDF files).
3. The metadata from the native files.

In addition to virtually eliminating the presence of potentially privileged metadata, the process of native file to image conversion has many benefits, some of which are listed below:

1. In many instances, on-line image review can be conducted with more speed than that of native document review (because they are "normalized" to a common image format).
2. On-line images can be redacted to the extent a portion of the document is privileged but the remainder is responsive.
3. Images can be endorsed with confidentiality designations and bates numbers.

In some instances, there may be disadvantages to reviewing document images if the production method will be in native file format. For example, if a hardcopy or review of document images is conducted and the adverse parties have agreed to exchange native files as opposed to the "imaged" files reviewed by counsel, there exists the possibility that privileged metadata could be produced (if the image conversion and metadata extraction process did not capture or alert counsel to the presence of tracked changes or comments within the document). Consider the following scenario:

Producing counsel receives MS-Excel spreadsheets from their client. Producing counsel then has the spreadsheets converted to images and the metadata extracted. The spreadsheet images are reviewed in conjunction with the extracted metadata. Based on having reviewed the document contents, producing counsel determines that spreadsheets contain no privileged information and produces the native files to the requesting party. Unbeknownst to producing counsel, there are password protected hidden columns and rows with privileged information, formulas, and references to other supporting documents that have not been produced that counsel is unaware of. By prior agreement, receiving counsel gets the spreadsheets in native form. Upon the examining the files, she becomes aware of the fact that there is hidden information and un-hides it.

Consider the above scenario in light of ABA ethics opinion:

"Lawyers who receive electronic documents are free to look for and use information hidden in metadata – information embedded in electronically produced documents – even if the documents were provided by an opposing lawyer…"

One can never underestimate the importance of considering review and production strategy as early as possible. In the above scenario, if the documents were reviewed as static images, and they were produced as such, the inadvertently produced metadata would not have been produced. As the example illustrates, the image review and subsequent native file production to the receiving party led to information that the producing party:

1. Did not know existed.
2. Did not review for privilege.
3. Did not produce but should have.

Avoidance of Privileged Metadata Disclosures – Metadata Awareness Means Control of The E-Discovery Process

In litigation, metadata is traditionally viewed as being associated with electronic "documents". As we are already aware, information proliferation has resulted in wide array of electronic file formats that possess none of the physical characteristics of documents in the traditional sense of the word, but are responsive nonetheless. The amendments to the FRCP take this into consideration. Responsive electronically stored information (ESI) can include stored recordings of conversations, video conferences, recorded WebEx™, Netmeeting™ type online meetings. These multimedia files, by virtue of their rich metadata content and electronic nature, are subject the same electronic document based metadata framework discussed earlier, consequently they fall into similar litigation (or compliance) related information classification schemes.

Practitioners involved in large, information intensive corporate litigation discovery, should expect to see an up tick in requests for responsive multimedia file types. This is particularly going to be the case in corporations where their information systems that employ a "unified messaging" information management model. The unified

messaging paradigm is an approach that centralizes the creation, management and storage of email, instant messaging, document as well as streaming audio-visual forms of corporate work product. It rolls up this additional functionality with that of traditional voice telephony systems. The ESI from UMS' that facilitate video conferencing, internet meetings and the like fall under non-discretionary regulatory frameworks such as Sarbanes-Oxley, or under other discretionary document retention frameworks. Counsel should consider this when conducting an IA of target organization infrastructure or crafting requests for production. Many of the new UMS use VOIP (voice over internet protocol) technology can store objective (system generated) metadata that includes information about phone calls such as:

1. Inbound / outbound phone numbers.
2. Origination / destination phone numbers.
3. Phone call date, time & duration.
4. Call transfer routing & party conferencing information.

Newer VOIP phone systems captures this metadata information about communication exchanges as well as any associated electronic multi-media files (with associated metadata layer information) which are frequently stored on common hard drives. E-discovery processing can extract this type of metadata for review and analysis independent of or in conjunction with metadata from traditional document based sources. What's more, is that latest generation of voice recognition technology can be applied to these multimedia audio files to auto transcribe data such as voice messages left on voicemail systems. The transcribed results can be stored as searchable and printable text files – a process similar to applying OCR (optical character recognition) technology to scanned document images.

Challenges presented by the New Metadata Sources

While it's crucial to develop an awareness of emerging metadata sources, many firms and organizations are sufficiently challenged by today's discovery requirements. All too often, discovery is characterized by reactive fire drills style document and ESI collection exercises. With electronic files, reactive approaches create serious risks of metadata corruption (spoliation). In addition, there is often a high cost of conducting discovery in reactive mode which often forces parties to make strategic decisions that have nothing to do with the merits of a case.

Many times organizations that have responsive electronic files will try to contain up front discovery costs by having their internal IT people do document collection from servers and computers in ways that are not defensible. Frequently, this results in immensely costly recollection initiatives much later on in the discovery lifecycle. There is also the opportunity cost to consider; this accrues when corporate profit center operations slow down or suffer disruptions as their internal IT people shift their focus from budgeted operations to production-related tasks. Situations like this have a ripple effect on outside counsel as well. As production deadlines loom and the volume of information that needs to be reviewed grows, more needs to be done faster—this creates the statistical risk of inadvertent production of privileged information as well as blows legal budgets out of the water.

What Corporations Can Do to Control Discovery Costs & Metadata Risks

The level of complexity that the newer metadata sources and forms present is fairly significant. Multimedia electronic file types contain copious amounts of metadata that can be deemed responsive to discovery requests and thus potentially evidentiary in nature. Corporate document and information retention policies, to the extent that they exist and are implemented, must now subsume these new forms of ESI. Litigc compliance hold methods must now also account for them as well. The simple truth is, that many highly successful organizations still will resort to fire drill discovery initiatives because they don't have a standing information discovery framework that will help them cost and time effectively identify, protect, gather and produce discovery information on an enterprise-wide basis.

What's more, virtually every outside counsel employs different methods, different vendors and different processes, with different cost structures to achieve the same end: the acquisition and analysis of evidentiary materials. There is absolutely no question that most CIO's and GC's recognize that there is a better way to manage litigation risk management and processes. Many are well on the way to addressing this situation in their enterprises; others are simply losing sleep at night.

The really big challenge they face is in institutionalizing the appropriate level of recognition and treatment of these new data forms as "document records" in their corporate culture. As organizations continue to create, capture and store new forms of information as well as content, they must continually adapt and change the mechanisms by which information is classified and managed within their organizations discretionary and regulatory document/record retention framework.

Fortunately the many CIO's and GC's that face these growing challenges, there are forward thinking consulting organizations that have the ability to help them create, implement and standardize best information retention and discovery practices. The right combination of cost-effective technology processes will address these seemingly daunting objectives and help organizations regain control of information management issues ancillary to legal processes.

As GC's and CIO's continue to take back responsibility by institutionalizing discovery processes that have spiraled out of their control, they will reap huge litigation cost control and risk In litigation, there has always been the possibility of inadvertent production of privileged documents or the incomplete production of responsive information. Given the diversity of electronic file types, the complexity of today's information management infrastructure and the increasing

Q&A

Q: Is metadata extraction expensive?

A: Not at all. There are many commercial file processing vendors that can extract e-File metadata and de-duplicate data collections quickly and cost effectively. In addition, hardware and software manufacturers are making it easier for corporations and their law firms to do this themselves, thereby saving tens of thousands of dollars on average for litigation discovery.

volumes of data to be analyzed and reviewed this risk increased significantly. For the technologically un-savvy, this can present a most vexing but surmountable obstacle. Under the proposed amendments to the discovery rules of the FCRP, counsel is charged with being thoroughly familiar with the information infrastructure of their clients and by extension that of their adversary. This means that, to the extent counsel does not already have the knowledge, she must gain an understanding the workflow nuances of the organization and characteristics of the files she is working with. Some pointers that may help counsel avoid some of the thornier issues related to the latter situation are described below:

1. Develop familiarity with the types of metadata that are generated by the target organization's information systems (as well as the user workflow that facilitates its creation).

2. Know which file types present the most significant hidden data challenges (i.e., excel spreadsheets often have hidden columns, rows; some files allow for password protection of selected data within a file with no file level password file).

3. Consider review and production strategies early on.

4. Sample the metadata in the responsive data set.

5. Stipulate what metadata fields will be produced early on (if any).

6. Stipulate that the form of production shall be as a "static representations of document images."

7. Include "claw back" provisions in the protective order to preserve privilege.

8. Use experts to help manage and streamline the technical aspects of discovery (maintains outside counsel focus in issues and merits).

Items 1, 2, 3 and 4 should be a standard practice in any discovery related matter. Items 5, 6 and 7 are significant in terms of their impact on safeguarding against production of privileged metadata. In 6, the phrase "static representation of document images," means that all metadata attributes possessed by the original document have been "scrubbed" from the responsive item. When "static representation of document images" is produced to a requesting party, it is merely a document image sans original metadata. While this approach precludes a "native file" document production, which in some instances is marginally less costly, it provides certainty that the only metadata that will be produced, if any, is the information agreed to by the parties in step 5. It is also important to note that, (while not discussed here) there are any number of highly effective and sophisticated methods that can whittle a native document collection down to a set that can be converted to images for production. This means that not all documents collected necessarily need to be converted to images. Using the right mix of electronic discovery techniques can have a significant impact on facilitating cost containment and help shift the counsel's focus back to their core strength, passion and competencies: lawyering.

The knowledge and technology needed to streamline and reduce the costs of discovery processes exist, but all too often the disjointed processes and divergent interests of the stakeholders involved lead to contention that adversely impacts the ultimate arbiter, who is the client. We at Litigation Logistics seek to return some semblance of practicality to technical facets of discovery. The reality is that the objectives have not changed in 50 years, only the playing field has.

SERVER METADATA GENERATED FOR USE IN CREATING A DATA MAP

Network Name	Device Description	IP Address	Location	Administrator	Operating System	CPU (MHz)	Memory (MB)	Manufacturer	Status	Files / Records	Data Size
sql-sales-01	database server	10.0.1.51	Row 1 Rack 2	Anna Misiec	Windows Server 2003, Datacenter Edition	3	2048	Contoso, Ltd.	OK	45000343	NA
sql-sales-02	database server	10.0.1.52	Row 1 Rack 2	Anna Misiec	Windows Server 2003, Datacenter Edition	3.2	2560	Contoso, Ltd.	OK	34583458	NA
sql-sales-03	database server	10.0.1.53	Row 1 Rack 2	Anna Misiec	Windows Server 2003, Datacenter Edition	3	2048	Contoso, Ltd.	OK	3435554	NA
web-sales-01	web server	10.0.1.12	Row 1 Rack 1	Don Hall	Windows Server 2003, Web Edition	2.4	1536	Contoso, Ltd.	OK	345344543	NA
web-sales-02	web server	10.0.1.13	Row 1 Rack 1	Don Hall	Windows Server 2003, Web Edition	2	1024	Contoso, Ltd.	Unavailable	43535	.007TB
web-sales-03	web server	10.0.1.15	Row 1 Rack 1	Don Hall	Windows Server 2003, Web Edition	2.4	1024	Contoso, Ltd.	OK	3344534	.05TB
ftp-sales-01	FTP server	10.0.1.14	Row 1 Rack 1	Anna Misiec	Windows Server 2003, Standard Edition	2	1536	Contoso, Ltd.	OK	55645656	.07TB
filestore-sales-01	file server	10.0.1.5	Row 1 Rack 2	Anna Misiec	Windows Storage Server 2003	1.4	1024	Contoso, Ltd.	Unknown	45346645456	27TB
filestore-sales-02	file server	10.0.1.6	Row 1 Rack 2	Anna Misiec	Windows Storage Server 2003	1.26	768	Contoso, Ltd.	OK	65546547654	34TB
rt-073-1000	1Gb router	10.0.1.27	Row 1 Rack 1	Don Hall			512	A. Datum Corporation	OK		
rt-077-1000	1Gb router	10.0.1.29	Row 1 Rack 2	Don Hall			512	A. Datum Corporation	OK		
ups-04-1500	1500 VA		Row 1 Rack 1	Don Hall				A. Datum Corporation	OK		
ups-06-1500	1500 VA		Row 1 Rack 2	Don Hall				A. Datum Corporation	OK		
tbu-01	5x30GB DLT		Row 1 Rack 1	Don Hall				A. Datum Corporation	OK		
corp-email-server	file server	10.0.0.56	Row 1 Rack 3	Don Hall	Exchange 2007	3.2	2048	Enron	OK	46568768678	22TB
marketing-email-server	file server	10.0.0.54	Row 1 Rack 4	Don Hall	Exchange 2007	3.2	2048	Enron	Legal Hold	68678678758	28TB
r&d-email-server	file server	10.0.0.55	Row 1 Rack 5	Don Hall	Exchange 2007	3.2	2048	Enron	Legal Hold	766875754	7TB

Table 1.

USING METADATA FOR FEDERAL RULES OF CIVIL PROCEDURE INITIAL DISCLOSURES

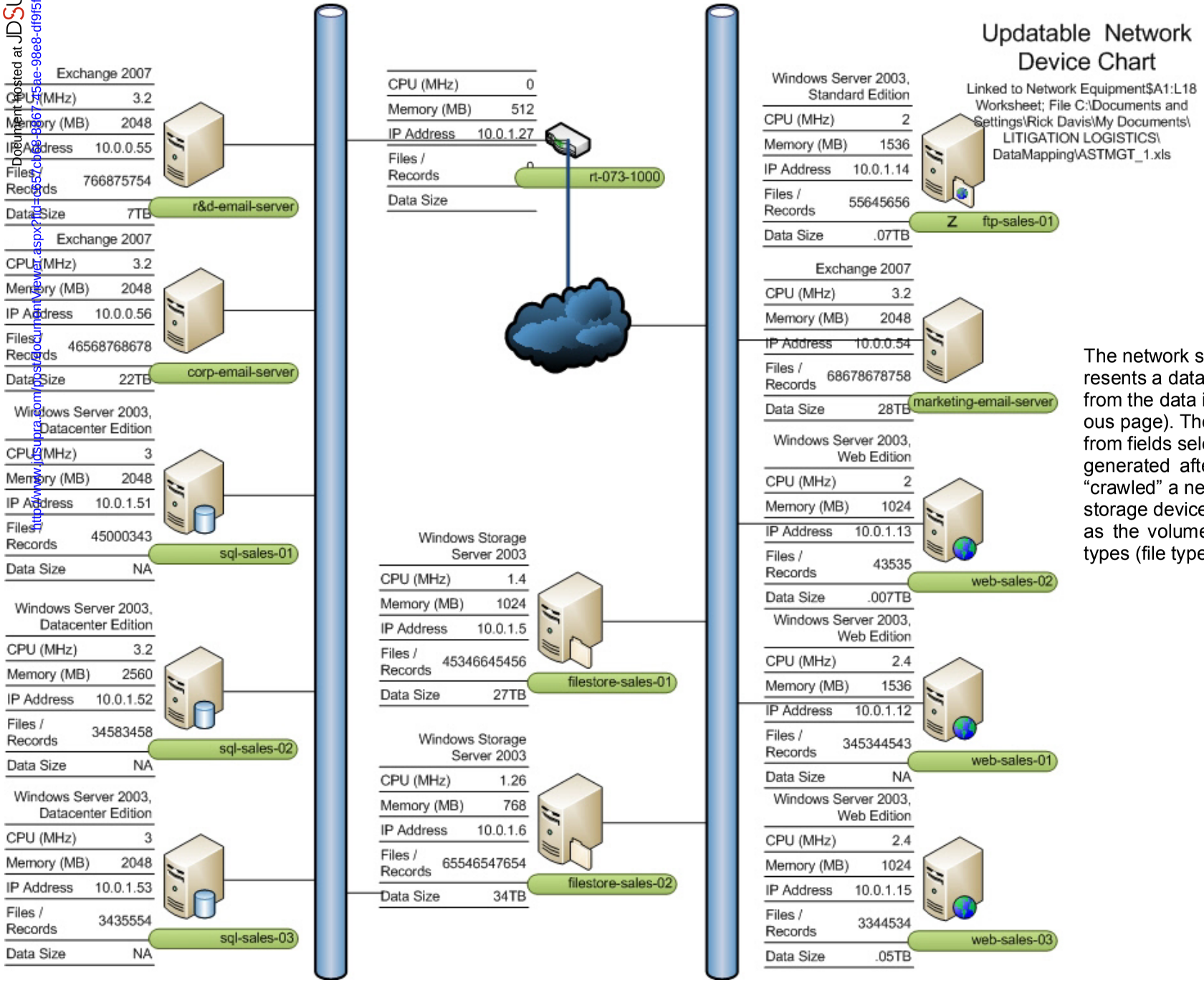


Figure 1.

The network schema in Figure 1. Represents a data map that was generated from the data in Table 1. (see the previous page). The data in Table 1. Comes from fields selected from the metadata generated after a Kazeon IS1200 “crawled” a network to identify various storage devices on the network as well as the volumes and associated file types (file types not shown).

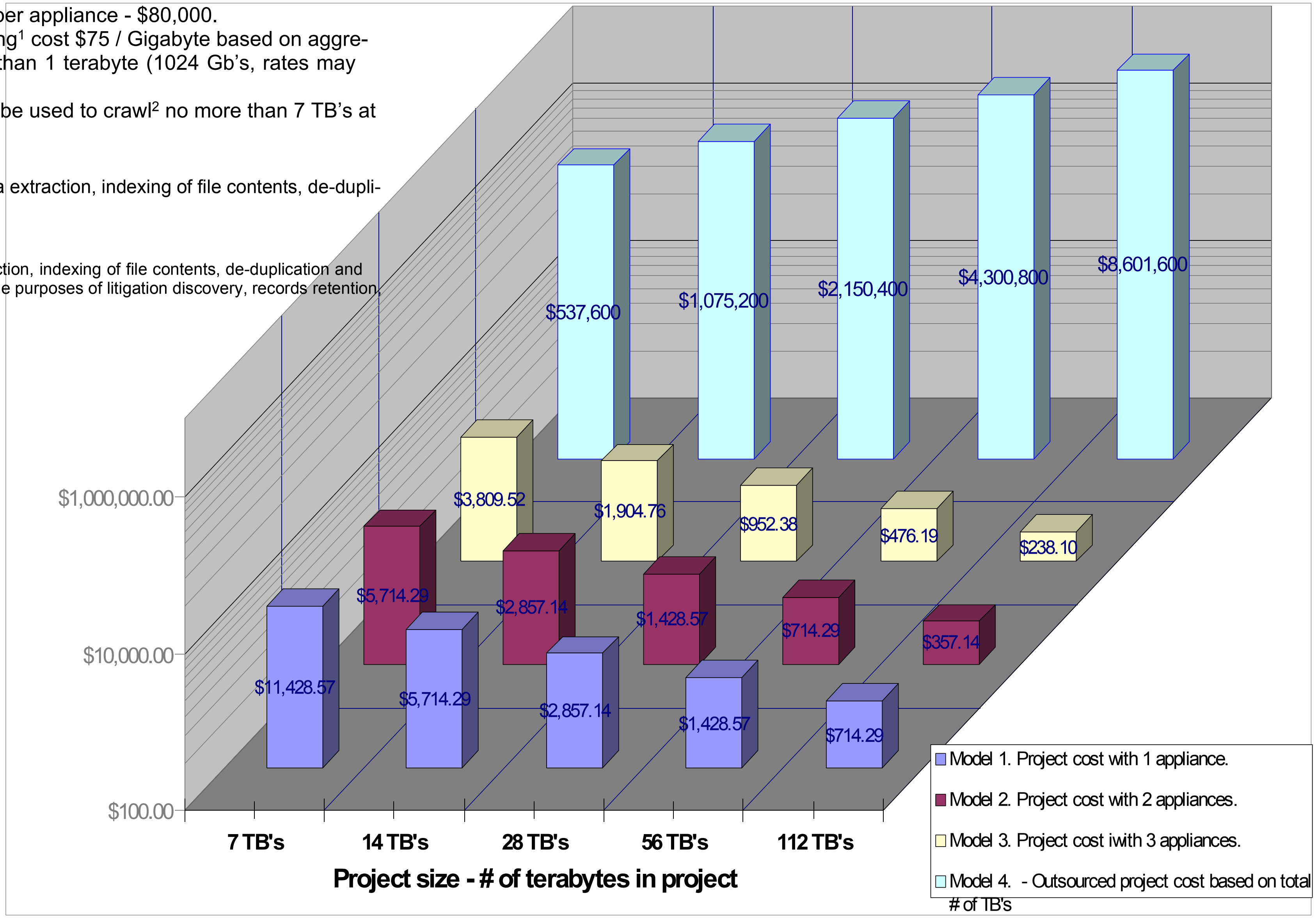
This model extrapolates data “processing” cost on a per terabyte basis using 1 - 3 appliances (Models 1 - 3 respectively) and outsourcing the processing of the same data using traditional vendor methods.

The assumptions are:

1. Base Kazeon IS1200 price per appliance - $80,000.
2. Negotiated vendor processing[1] cost $75 / Gigabyte based on aggregate data volumes of greater than 1 terabyte (1024 Gb’s, rates may vary).
3. Each Kazeon appliance will be used to crawl[2] no more than 7 TB’s at a time.

[1] Processing consists of metadata extraction, indexing of file contents, de-duplicate and keyword searching.

[2] Crawling defined as metadata extraction, indexing of file contents, de-duplication and rules based classification of data for the purposes of litigation discovery, records retention, compliance & internal investigations.



Litigation Logistics, LLC
Corporate Discovery
Solutions Architects
Litigation Discovery | Internal Investigations | Records Retention

Electronic Discovery services:
Litigation Discovery
Mergers & Acquisitions
Due Diligence
Records Retention
Corporate Litigation Readiness
Litigation Cost Control

Contact:
11 Mancuso
Ossining, New York 10562
Email: redavis@litlogic.com
Mobile: 646.306.3833
Fax: 914.923.3446

About Litigation Logistics, LLC

Litigation Logistics, LLC (LitLogic) is an electronic discovery consultancy that was formed in 2003. Today's litigation environment is fraught with electronic discovery challenges. Our mission is to help law firms and corporations contain litigation discovery costs and manage data liability risks by understanding their business as well as their operational and technical infrastructure.

As consultants, we believe in leveraging our domain expertise by combining it with practical technology to create relevant solutions that satisfy our clients ever evolving needs. Given the complexity of today's business environment, we work in conjunction with a consortium of international experts that allow us to provide strategic and tactical flexibility when it comes to all facets of litigation data lifecycle and records management.

The Founder

Richard E. Davis, JD - Managing Partner

Richard Davis is the former Director of Litigation Risk Management Services for Constantine & Aborn Advisory Services (CAAS). Prior his role at CAAS, he founded the Practice Management Department for Kenyon & Kenyon, LLP, an intellectual property law firm based in New York. His role at Kenyon involved advising selected Fortune 1000 clients of the firm with respect to data management policies. As early as the summer of 2005, Mr. Davis developed a standard data infrastructure assessment protocol (DIAP) to facilitate initial disclosures under what is now FRCP 26(a). This process involved the mapping of network infrastructure with organizational management information to identify areas that contain the highest likelihood of responsive information to litigation. Prior to his experience at Kenyon, Mr. Davis managed the litigation support and other IT groups at the venerable law firm of Cravath, Swaine & Moore.

As a consultant, he worked with National Data Conversion (NDC), one the most reputable tape restoration and legacy data conversion and media recovery specialists in the United States and has also consulted with investment banks on matters related to the acquisition of litigation support companies. He has taught numerous classes on the litigation data lifecycle and provided industry insight to companies such as Zantaz, CaseCentral, Iron Mountain, KVS & Ilumin (the latter 2 companies and associated products were acquired by Symantec) on the demands of litigation discovery.

He has consulted for members of the Federal Judiciary in the capacity of "Special Master" and has authored a number of CLE classes.

Mr. Davis holds a Bachelors of Business Administration degree in International Business Management as well as a Juris Doctor Degrees from Pace University. He has a Certificate in Corporate Incident Response Management and is a member of ARMA International. His articles on litigation technology management have appeared in The Corporate Compliance and Regulatory Newsletter, e-Discovery Law & Strategy, the New York Law Journal and he has been cited and quoted in numerous publications. Today, Mr. Davis works primarily with corporations and law firms to help them institute policy driven risk mitigation strategies for data management.