



5 Things Records Management/Information Governance (RMIG) Professionals Can Learn From Their Corporate Litigation Counterparts¹

By Sandra Serkes
President & CEO of Valora Technologies, Inc.

Introduction

Lawyers may not always be commended for their strategic use of technology, but when it comes to managing large volumes of documents for a litigation matter, they do know a thing or two. They've had to learn the hard way what it means to ask for reams of document files and then manage and understand the disorganized contents that are ultimately provided. Fortunately Records Management and Information Governance professionals (aka **RMIG**) are in the position of lucky second, in that they are finally getting the attention and consideration they deserve, as well as the benefit of watching how their litigation management cousins fared first. RMIG professionals can learn a lot from the experiences of eDiscovery and litigation document management. Below are Valora's top 5 tips and techniques that RMIG professionals can learn from their corporate and outside counsel litigation counterparts.

1.0 Seeing the Content Beyond the Pictures

While most RMIG professionals are familiar with document conversion (that is, scanning paper records to digital image), not as many are familiar with what happens next. The simple thought is: I've converted the documents from physical to digital, I'm done, right? Wrong! A collection of digital images is about as useful as a box of old photographs. You know what they are, but what do they contain? In document management parlance, what do they *say*? It is the textual *content* that creates the meaning in most business documents and the reason for which they will ultimately be later retrieved, evaluated, analyzed, produced or disposed of.

In litigation, the users of converted content are, of course, attorneys - often with a voracious need to quickly get to specific documents for one purpose or another. One of the best tools available to attorney-users is intelligent **grouping** of documents. For example, documents are grouped by Document Type (ex: all the invoices), by Document Subject or Topic Area (ex: all the documents that discuss liquid chromatography), or by date range (ex: documents from June-Sept, 2007). Such groupings are typically created from the fielded or "bibliographic" information (metadata) of each document, as well as from the searched/indexed content of each document.

TABLE OF CONTENTS

Introduction – 5 Things RMIG can Learn From Litigation & eDiscovery

1.0 Seeing the Content Beyond the Pictures

2.0 Data & Document Distillation

3.0 Real Costs, ROI & Proportionality

4.0 Understanding Context

5.0 Storage is Not the Goal. Knowledge is.

¹ Earlier versions of this White Paper were originally presented at and published by ARMA International



There are also other types of easy groupings, such as documents that might have been created or stored together (Ex: employee personnel files), or documents that are similar in nature, what we call Near Duplicates. Finally, there are groupings of documents by intent, sentiment, parties, or conversation threads. All of these groupings are readily available and have proven their merits many times over in litigation. Crossing over to RMIG, intelligent groupings might include: NearDuplicates, Familial or Records-Storage Families, Topic or Subject Groupings, and Groupings By Customer, Product, or Service Type.

Such groupings are made possible by the addition of strategic information to the digital (scanned) picture, often called **indexing or coding**. Specifically, the application of **OCR text** adds searchability of content, as well as a whole field of analytic options.

2.0 Data & Document Distillation

It may seem counter-intuitive that it is wise to actually *remove* documents from a stored database. However, there are numerous reasons to do so. First, there is almost never a viable reason to store duplicate copies of the same document. There are extra costs in processing, storage and upstream expense of users having to sift through multiple instances of the document. Removing duplicates early is the best antidote to this type of needless expense. It is important to note that the actual removal of a document from processing and storage is not the same as removing all mention or indication of the document's original existence. It can easily be noted that a duplicate document once existed in the following location/custodian/format, without retaining the actual copy. This type of strategic removal of documents is often called **culling**.

Litigation document management teams have elevated document culling to an art form. There are methods of instantaneous duplicate/removal document detection (typically based on [Hash file comparisons](#)), and likely or statistical duplicate/removal document detection (typically based on predictive or [Rules-Based analytics](#)).

There are other reasons besides duplication that would suggest that a document be removed from queue. For example, documents might be obsolete (past a retention date, terms no longer valid or in effect, etc.). They might be irrelevant (wrong Document Type, Author, or contents), or available elsewhere (such as publicly available information). Or, they may simply be out of scope and not worth the expense to process, store and access over time.

Finally, consider that **documents that are retained are documents that are discoverable**. Many organizations have defined Retention Policies that dictate why, when, where and how long documents should be kept. However, having the policy and implementing or enforcing it are different matters. When retention schedules indicate different treatments for different types of documents (as most do), it is imperative that the organization be able to a) identify the Document Type accurately and b) audit the system on a regular basis to make sure the documents are being handled appropriately. Thus, systematic, planned, and monitored removal of documents may be very valuable, if only for ease in compliance with retention policies.

3.0 Real Costs, ROI & Proportionality

In eDiscovery, there is a notion where litigants attempt to avoid cost or effort **dis-proportionality** per FRCP [Rule 26\(b\)\(2\)\(C\)\(iii\)](#), in which “the burden or expense of the proposed discovery outweighs its likely benefit.” The same proportionality argument holds true for *all* document databases. If the cost to create, manage, maintain or use the database outweighs its usefulness, then the database has failed the proportionality test.



There are only 2 ways to boost proportionality: increase the benefits of the system or decrease its cost. These are not always mutually exclusive. As counterintuitive as it may seem, sometimes lower cost methods of creating and managing the database actually *increase* what the database yields. For example, in automated indexing, the cost to add one more field of information is often very low. That is, the cost to collect 2 fields vs. 3 fields is almost identical. But, by having that 1 additional field of information to search on, sort by, use for forecasts or other purposes, the usefulness of the system has increased significantly. (33%, actually.)

One area in which RMIG teams consistently *undervalue* the cost of effort is when work is being performed by their own teams. If you or your team is performing active document management work (such as indexing/coding, scanning, sorting, shredding or inventorying documents), then these actions are a cost factor of your repository. When a temp worker actively sorts documents each day for a year, then those hours need to be added up and considered in the total cost of the system.

Litigation teams understand the need for such cost measurement and allocation because they are used to (and make their profits from) hourly billing. They are very aware of time spent on a project by all personnel involved, from the most senior partner to the most junior associate. All hours are accounted for and (presumably) billed to the project. RMIG teams would be wise to follow this model, at least for calculations of real costs and ROI of their document management/retention/knowledge systems. For example, a low-end scanner may cost only \$500. But, when an administrator, who costs \$60,000 (with fully loaded overhead) must spend all day, every day, scanning documents because it operates so slowly and the document count is high, then the total cost of that system is actually \$66,500. It is quite likely that outsourcing that work is a much better use of limited budget dollars. To best understand the value of the true efforts and contribution that RMIG personnel really contribute to the corporate bottom line, is it important to take into account the hours that such teams are actually performing document conversion and/or repository work.

4.0 Understanding Context

Context is a tricky notion, which has its roots in the same area as the start of this whole discussion. That is, context provides the *purpose* for collecting, organizing, controlling, retrieving and removing the documents in the first place. Litigation teams have learned that **context changes over time, across users and matters, and based on what is being asked at the moment.**

Litigation teams are highly structured. There are partners, associates, paralegals, clients, co-counsel and so on, each with their own database needs in the hierarchy. Smart documents databases account for all of these circumstances (the context) and more. RMIG professionals should ask who is going to be using the database, and for what purpose(s) over time? What “views” of the documents and the fielded data will be required? What reporting will be needed on processing status, addition/removal of documents or data trending over time? Do documents need to be translated to (or from) English?

Some users will need to add information to the documents. Others will be forbidden from this activity. Some users will want to see a “full” record that might include multiple documents in a family (*aka* an attachment range). Others will want only the immediate document in question. Some will want to see all documents “like this one,” while others will want only this exact one. Still others will want to see information trending across the document population as a whole. (Ex: distribution of documents over time, by Document Type or by Topic). Others will need only a pinpoint retrieval.

For databases with potentially sensitive information present, a whole host of access and presentation options may be at play. Is any of the information personal or private? Should any portion of documents be redacted? Should certain aspects of documents be highlighted or made more prominent? What will need manual review and what can be automated or “presumed?”



Nearly every one of these contextual presentation constraints has been addressed in the typical litigation document database. There are easy and low cost options for RMIG document databases to enable; but creators and users need to know to ask.

5.0 Storage Is Not the Goal, Knowledge Is.

Too often, RMIG projects are about “getting rid of the paper,” as opposed to learning (and preserving) something from the contents of that paper. In litigation matters, **the conversion of documents from physical to digital is just a byproduct** of the real goal - getting to the knowledge stored away inside those documents. In order to answer who knew what when, litigation teams must go beyond the simple conversion of documents to get to their real contents.

Simply moving from physical storage to digital storage misses most of the problem - and **a real opportunity for RMIG teams to master the institutional knowledge of their organizations**. By organizing, cataloguing and presenting the information locked away in stored documents, RMIG teams help address ever-increasing corporate concerns about liability, compliance, risk & exposure and business trends. Think of the conversion project as an opportunity to move to a more “enlightened” treatment of information assets. Evolve forward from shelf storage, DVD, or cloud storage into active, useful information governance.

Litigation teams have inadvertently led the way on this, as their goal has always been to get to the meaning (content) of the documents they “store.” I put store in quotes because they would not even think of their actions as storage. Indeed, they prefer the term “hosting,” which by definition has a temporary, transient nature to it. As if the documents were staying over in a sort of litigation hotel and will be released back to the wild once their useful stay (the trial) has ended.

Above are just 5 smart lessons that RMIG groups can learn from their litigation and eDiscovery counterparts when it comes to large-scale document databases. There are also **numerous lessons you can teach them**, and that is the topic of a companion White Paper to this one: 5 Things Litigation & eDiscovery Teams Can Learn from Records Management & Information Governance. For more information on these topics and many more, please visit Valora’s [website](#) and [blog](#), or contact us at: 781.229.2265.

About Valora Technologies, Inc.

[Valora](#) is a technology-based provider of automated document management, analysis & review services for the legal, record management & information governance industry. We offer outsourced services for paper and electronic populations to corporations & government agencies, as well as their inside & outside counsel organizations around the world.

Valora has developed a strong expertise in the processing, management & analysis of large and small matters with complex requirements, such as short deadlines, sensitive material & mixed languages. Our specialty is providing efficiency, organization and cost control.



About Sandra Serkes

Sandra Serkes is a dynamic leader with an extensive background spanning 25 years in software marketing, product management & corporate strategy, particularly in document processing & analytics, computer telephony & speech recognition. Today, Ms. Serkes oversees Sales & Marketing, Finance & Administration, Operations, Engineering and Corporate Strategy at [Valora Technologies, Inc.](#), where she is CEO, President & Co-Founder.

A graduate of Harvard Business School and MIT, Ms. Serkes is a frequent industry speaker & panelist. She is an active participant in the Women Presidents' Org., The Commonwealth Institute, the MIT Enterprise Forum, the MA Software Council & the Network of Harvard Alumnae. Ms. Serkes serves on the boards of several technology and service start-ups, and was named a 2006 “Woman to Watch” by Womens' Business magazine.