



Creating an ROI for Information Governance, with No Budget

Information Governance





CHEAT SHEET

- **Maximize employee productivity.** Data management can build small efficiencies into serious cost savings.
- **Save on data storage.** Eliminate redundant and inefficient data storage protocols, de-duplicate identical copies, know when to delete outdate material, and remove non-business content.
- **Streamline eDiscovery.** If your eDiscovery process is aligned with your overall Information Governance process, collection, processing, early case assessment process and review should be achievable within the enterprise environment, without creating an additional copy.
- **Think future value.** Analytics such as natural language processing, semantic analysis, and concept mapping present a rapidly evolving area of potential ROI for information governance efforts.

There are several pressing questions that arise when commencing an information governance initiative:

1. Who should be on the committee?
2. How widespread do we want to deploy?
3. Can we use any of our existing solutions?
4. How are we going to pay for this?

After many years in the information governance (IG) space and having talked to literally hundreds — if not thousands — of corporate employees tasked with overseeing or participating in their organizations' initiatives, the resounding reply is, "There is no budget!" Rarely is this the absolute truth. While there may not be resources to add additional sums to the existing budgets, the reality is that there's almost always budget available. It just takes revisiting what you are currently doing and identifying areas that need to be updated to more efficient processes. Creativity with accounting isn't required to "create" a budget for a new acquisition. Instead, it's creativity in evaluating the potential impact of new solutions that's called for. Often, calculating a return on investment (ROI) that isn't initially obvious is vastly underestimated.

Information governance, in particular, is prone to roughshod calculations of ROI that fall short of reality. And it's not difficult to understand why: anyone who has ever held a place on an IG committee is well aware of the vastness of the ocean that needs to be boiled. The overwhelming scope of information governance itself, combined with the overwhelming number of information governance products on the market, makes for a likely case of decision paralysis ... there's just too much to decide. Thus a common approach is to stick to simplistic metrics for ROI analysis — metrics that are prone to underestimate the wide-ranging benefits of a well-planned information governance strategy.

When you think of information governance you have to think about it as any other process within your organization. Are you still running the old IBM Selectric typewriters or steno machines in your environment? Of course not. Why? Because there are better, more effective ways to manage information. Information governance, whether applied across an enterprise or merely deployed for a

targeted business unit, can generate an ROI quickly and easily by utilizing dollars that are currently being spent to accomplish many of the same tasks, but with less efficient utilities.

Let's revisit the typewriter's replacement with the computer. At the risk of dating myself, I remember earlier days in the legal profession when the drafting of an agreement, discovery device, or any number of typed documents necessitated the retyping of an entire page because the attorney didn't like the language on one single line. Alternatively, there was a nearly as tedious option of whiting out the affected line, and then placing the page back into the typewriter and trying to line it up with the typing head to attempt to correct the errors to insert the new language. Today, that same process can be corrected in seconds by merely pulling the file up on the computer, correcting the language and reprinting the document. Properly calculating the applicable ROI is thus often a process of realigning these dollars toward more effective processes and systems. The man hours previously spent making small corrections are now free to be allocated elsewhere, and these hours of productive time make a big difference in the long run. The calculation of a potential ROI can be as detailed as the business need requires, but some readily available metrics will typically justify the associated costs. If done well, an information governance investment will not only pay for itself over a short window of time, but actually save the organization potentially tens of millions of dollars a year moving forward.

When you look at offsetting existing overhead, keep in mind that you may not need to dig very deeply. Some of the areas may render enough cost savings to warrant an enterprise-wide solution with a relatively quick payback period, while other areas may not apply to your organization.

Some areas to consider:

Employee productivity

We all struggle with implementing records retention/disposition policies and having them accurately carried out by the employees. What is the cost to the organization?

Many applications currently being used by corporations today to manage data require employee time and effort to determine where each document should be classified within a huge records retention policy. Far too frequently, these processes have proven to not be effective. Why? As a result of the most recent downturn in the economy, employees have been tasked with wearing several different hats. They are completely overwhelmed just trying to complete their day-to-day workload. Then, we have placed on them the additional requirement of classifying information to meet a record classification schema for all of the work that they create. If given the option, we have seen that the employee often merely selects the longest retention period and classifies large groups of information under that time period regardless of whether it is appropriate to do so. These shortcuts diminish the benefits and undermine the intent of classifying information. What if the lion's share of the effort could be done in an automated process? Today, there is technology that can aid in these processes, thus ensuring integrity in the retention policies around the organizations' data stores and disposition processes.

Given the number of mergers, the downsizing of organizations, and the retirement of our baby-boomers, finding information has become a huge problem. If data isn't properly managed in a fashion that allows future access, you might as well not even have it. The inability to quickly and accurately locate past work product forces employees to needlessly and unproductively re-invent the wheel, at great toll to the organization. How can we repurpose and reuse information, or utilize data to defend a position, or just find something that you did a few years ago that you don't remember where you stored it? What does this cost any given company? You can plug in your company's

specific numbers, but a conservative, back-of-the-envelope metric, such as the following, is

often a good (and eye-opening) start. Using an example firm with 10,000 employees, the ability to save just 15 minutes per day per person on managing/searching for electronic data can literally total millions in savings.

Estimated, example data management costs:

- 10,000 employees
- 15 minutes saved per day
- 5 days per week, 49 weeks per year
- Average pay of \$20 per hour

10,000 employees x 15 minutes x 5 days per week x 49 weeks per year = 36,750,000 minutes

This totals 612,500 hours saved per year.

When a rough wage estimate of \$20/hour is taken into consideration, this totals: \$12,250,000 per year.

Thus, if data management is able to save each person a modest 15 minutes of time each day at a large company, it can easily add up to \$12+ million quite quickly, even with conservative estimates.

Data storage costs

Every company is different; however, there is one thing that we do know... companies tend to keep way too much information that has nothing to do with the nature of the organization, and often multiple copies of it.

A simple example sheds light on this problem. Assume that an employee creates a document in Word, stores a copy to their hard drive, and then attaches it to an email sent out to 50 members of the team working on the project for their edification. Each of the 50 recipients then store it on their individual hard drives, file shares or SharePoint sites. That single document now exists in the environment potentially hundreds of times. Now multiply that across all of the work being conducted throughout your organization. We compound this problem with the employees' personal chit-chat and junk that is also stored in combination with the business data. Implementing a viable information governance initiative will assist your organization in not only doing a better job with the data that you need to keep, but aid you in identifying what you can eliminate.

Despite today's prevailing "storage is cheap" mentality, enterprise-scale data environments still remain at risk for unnecessary storage bloat and associated storage costs, not to mention the associated risks of preservation requirements in the event eDiscovery issues arise. Millions of tiny increments in expense can add up to significant sums over time — a phenomenon that has been leveraged by criminals and movie plots alike. In a very large business, the static storage cost alone can impact the bottom line; however, the real long-term expense accumulates with the need to sort, search, and potentially move that same data when it clutters the environment where other — more relevant — information is also stored. Technology today provides opportunities for organizations to search their data store, much like a natural-language Yahoo!, Bing or Google search. However, do you want the same situation that we find in these environments, where your search yields over a million hits and only a small subset (if you can even find it) has any relevance to what you were really

looking for? Maintaining these volumes of nonbusiness information, or business information that has passed its useful life, will cost organizations dearly in not only storage costs, but in loss of the benefits that would have been yielded in providing employees with access to all of that information. Ultimately, what you are really looking for is access to viable, timely information.

Philosophical arguments aside, most would agree that the average business wastes a lot of time, effort, and money on data that is essentially “junk.” To minimize the impact that such detritus has on the enterprise, three major — but complementary — approaches have emerged:

- **Identifying and eliminating nonbusiness content.** As a result of the eDiscovery process, it has become fairly common knowledge that 50 percent or more of the information that corporations maintain today has nothing to do with the nature of the business itself. By utilizing technology to quickly identify and weed out irrelevant and former-employee-generated personal data, such as those “Honey, don’t forget to pick up milk” emails, you can substantially reduce storage costs over time. Increasingly, some of this can be done by utilizing technology that routinely and automatically grabs information based on domain, content, or other identifiers and sets an appropriately short retention period for that information’s disposition.
- **De-duplication of identical copies.** Another area of potential saving on storage costs is TRULY single instancing information within your archive environment. There are many flavors of archive systems available; however, you need to ensure that your solution truly single instances your information. Do you really need 10,000 copies of the president’s email to all of the company employees wishing them a wonderful holiday? Or do you just need to save one unique copy with the accompanying metadata designating who received it and when?
- **Obsolescence and removal of outdated material.** A third area is eliminating business information that has passed its useful life. Back to the president’s email: just how long do you really need to keep that type of information? Who should have access? Many companies have gone through the process of creating records retention and disposition policies, but are they actually being carried out? What does it cost to keep this information around? By automating this process based on data creation date, you can rest assured that your data is managed in a cohesive and defensible fashion.

When you are looking at reducing storage costs, remember: it isn’t just about the costs associated with the hardware and software. It’s an entire ongoing process involving people and effort. You need to look at the costs associated with the software that supports the hardware, and the manpower for maintaining the size of the environment. If you reduce your storage footprint by 50 percent — or even a more modest percentage — what are those overall cost savings? Even in an era of relatively “cheap” storage hardware, these savings can be significant, especially to a large-scale enterprise.

eDiscovery

You can do as thorough an analysis as you’d like when it comes to an eDiscovery ROI. There are probably hundreds of eDiscovery vendors out there, and they’re all likely to tell you how they are going to save you money. However, what they won’t tell you is that as long as you continue to duplicate data and shuffle it between systems, there are going to be significant costs and risks associated with that endeavor. Instead, if the data is managed in a coherent and single-instanced fashion within your environment, there are several easily identifiable efficiencies that can be leveraged for corporate benefit:

- **Blanket holds.** Many times I hear the rationale: “we do custodian-based holds.” That’s

great, but do you really need to place a hold on everything that the individual ever did? Or would it be more effective to just place a hold on potentially relevant data? If personal data, non-business data, and outdated business data are being automatically removed from the system on a timely basis, the costs associated with placing holds, reviewing, and managing it for eDiscovery purposes is going to plummet. What might those savings add up to?

- **Duplicative copies.** How many copies do you really need? So much of what we do today is managed in many different places, and delivered to multiple individuals in the organization. So how many copies are you going to replicate and place on hold as well as ultimately process, review, and host? You need to track who had a copy and where it came from, but not every actual version. As stated earlier, there is duplicative data across the enterprise that is being created, every day, during the normal course of business. Is there any reason why all of these need to be individually maintained as copies rather than just appropriate metadata (links, shortcuts, etc.) “pointing” to one true copy?
- **eDiscovery process management.** Technology has reached a turning point and data for eDiscovery can now be managed as part of your overall data management initiative. This means more power to truly control eDiscovery costs. If your eDiscovery process is aligned with your overall information governance process, everything from collection, processing, early case assessment process (ECA), to review should be achievable within the enterprise environment, without creating an additional copy. What does this mean to the organization?
 - No collection fees
 - No processing fees
 - No hosting fees
 - Reduction of the volume of data to be reviewed by 50 percent or more
 - Most importantly, immediate access to your data for early analysis, thus allowing for a more timely decision as to potential settlement options, and saving the entire cost of the traditional litigation process: not just eDiscovery costs.

If everything is managed as part of a comprehensive information governance strategy, eDiscovery and records are fully integrated. This means that as data is released from legal hold (and all holds are cleared on the document) the data will automatically revert back to its prelitigation records retention policies and disposition schedule, without the need for manual intervention. For companies that have any eDiscovery overhead, these savings can warrant the costs of an effective information governance initiative, but for serial litigants, the savings are staggering. Take a look at what you spent on outside service providers over the past couple of years for collection, processing, review and hosting fees. If you are utilizing internal IT resources for any of these steps, what are the associated costs to your division for that work?

ROI: We know what we know (and that’s the problem!)

It was Donald Rumsfeld who famously — or perhaps infamously — grouped military intelligence into three categories: the “known knowns,” the “known unknowns” and “unknown unknowns.” The same categories of uncertainty are all too common in the business world, and we tend to spend a disproportionate amount of time solving for the “known knowns:” the problems that we can most elegantly define, quantify and describe.

With information governance and calculation of ROI, we have a reasonable yet stubborn tendency to stick to the “known knowns” of existing data management requirements. Concrete legal rules, regulatory requirements, industry standards, and established policies are all easy to pinpoint and consolidate into a “checklist” of necessary governance tasks. The problem is that this initial checklist alone can seem insurmountable, especially given the complexity of existing data environments. Thus

IG committees are often left understaffed and overwhelmed, unable to tackle many of the “knowns,” let alone move on to the more nebulous “unknowns” that remain unanswered.

The result of the “known-knowns” bias is that the long-term ROI of information governance is often vastly underestimated, largely due to the factors that can’t fully be predicted. The paradox here is that the prevention for uncertainty can also be its cure: analytics. Data analysis allows predictive models to be built, approximating unknowns based on both past events and accessible data. But for useful analytics to be applied for more certain calculation of ROI, you first need data: data that is accurate and representative of the business questions being asked. With most enterprise data management efforts today being haphazard or incomplete, the problem is likely obvious. We have a Catch-22 or paradoxical situation; information governance is the necessary foundation of business analytics efforts, including the very analytics efforts that would likely help calculate its own ROI.

Future value: Analytics

The field of analytics presents a rapidly evolving area of potential ROI for information governance efforts, especially for unstructured content that has traditionally been considered to have little quantitative potential. This is because technology for harnessing and parsing the “messy” data of human-to-human communication is beginning to mature. Natural language processing (NLP), semantic analysis, concept mapping, and other assessments of human meaning are becoming more readily available. While many of these capabilities are currently confined to point solutions or specialty tools that work with limited amounts of data (predictive review, anyone?), we will soon enter an era where they will be applicable to larger data sets. This is where holistic enterprise information governance becomes immensely valuable. Sure, today you can gauge the average emotional response to public tweets that mention your brand. But what if you could leverage the same capabilities for your entire corpus of internal business communications? What if you could accurately determine trending topics amongst workers, identify and consolidate overlapping efforts, and identify the right resources and experts for projects?

As we look at analytics, we need to understand that most analytic “tools” in the market today rely on a small subset of the data at hand, upon which analysis is then performed. Often analysis requires sampling, which is subject to numerous biases and necessitates removal of data from its native environment, meaning that the analysis does not reflect a “real-time” assessment. It is a given that so much of the information that is being maintained has little or no business value, then we utilize that same data to conduct data analytics. The familiar statistics axiom once again rears its ugly head: “garbage-in, garbage-out.” If that data were excised of non-business information and fastidiously maintained in an ongoing, consistent manner, data analytics efforts would yield far more relevant, reliable, robust results.

Thus, when estimating the potential ROI of an information governance program, it is critical to also estimate the potential benefits of analytics initiatives within the enterprise. With huge sums of money (and vendor marketing dollars!) being poured into purported “Big Data” initiatives and products, it’s easy to forget that analytics tools are rendered useless without a foundation of relevant data. Ideally, a comprehensive information governance initiative provides the master repository of all relevant business content: content that can subsequently be used for value-driving analytics projects.

Stepping back: Assessing the existing data landscape

The first step in calculating the ROI of an information governance program isn’t a calculator or

spreadsheet; it's taking a good look at what you already have. Sometimes you need to take a step back to get a higher-level perspective before you can start moving forward.

There are some general steps that can prove helpful to getting into the right frame of mind before discussing the potential ROI of governance. Knowing the “lay of the land” of the existing environment — no matter how messy or incomplete — will give a much more accurate view of potential value over time. An example “checklist” of 10 basic assessment items to help get started might look as follows:

1. Visually map out separate unstructured data governance systems and tools

- How many separate systems do you have for unstructured information?
- Count ECMs, email archives, eDiscovery point solutions, standalone analytics tools, etc.

2. Identify overlap in systems that may store copies (duplicates) of the same information

- Is it necessary to have the data copies in more than one system? How were they created?
- Are the copies managed equally for their appropriate lifespans?

3. Identify any unstructured data types that do not have a designated system for management

- Do you formally archive instant messages? Scanned images? Collaborative tools? Internal wikis? Etc.
- Are there any data types that are “slipping through the cracks?”

4. Map the unstructured data systems (from step #1) according to their roles in the Electronic Discovery Reference Model (EDRM)

- If there are gaps between EDRM steps, how are they bridged during the eDiscovery process?
- How does pricing of each individual system affect movement of data?

5. Identify and list potential points of failure in the preservation and legal hold process

- Does a legal hold applied to a piece of data automatically freeze its lifecycle?
- Does a legal hold require custodian confirmation or human action to take effect?
- Has the business ever faced sanctions for missing data or poor preservation?

6. Evaluate current search speeds, and know how long it takes for an “average” search

- Can the legal team easily search across all systems for data during early case assessment?
- Does search constitute a significant portion of time during the eDiscovery workflow?
- How much time do general business users spend searching for documents or data?

7. Identify the relevant jurisdiction which has the strictest or most complex

standards for data

- Consider national/state privacy laws, data encryption laws, industry-specific standards, etc.
- What costs are associated with meeting these requirements where the business operates?

8. List which of your data systems operate partially or fully in the cloud (versus only on-premise)

- Does access to these systems differ than systems that are entirely on-premise? Are you able to access the exact same data?
- What are the costs associated with securing, monitoring and accessing this data?

9. Pinpoint how data lifecycles are currently determined, applied and executed

- If data retention policies are not based on a specific period predetermined by relevant industry law, how are they determined?
- Do records procedures set useful lifespans for ALL data, not just traditional records?

10. Assess areas of concern for data security, as well as associated costs of a potential breach

- Is highly sensitive information (IP, personnel records, etc.) protected appropriately?
- Do legacy or End of Life (EOL) systems pose additional risk for security, or require excessive maintenance?
- What would be the cost of a major data breach to the company, including reputation?

Stepping forward: Determining a more accurate ROI

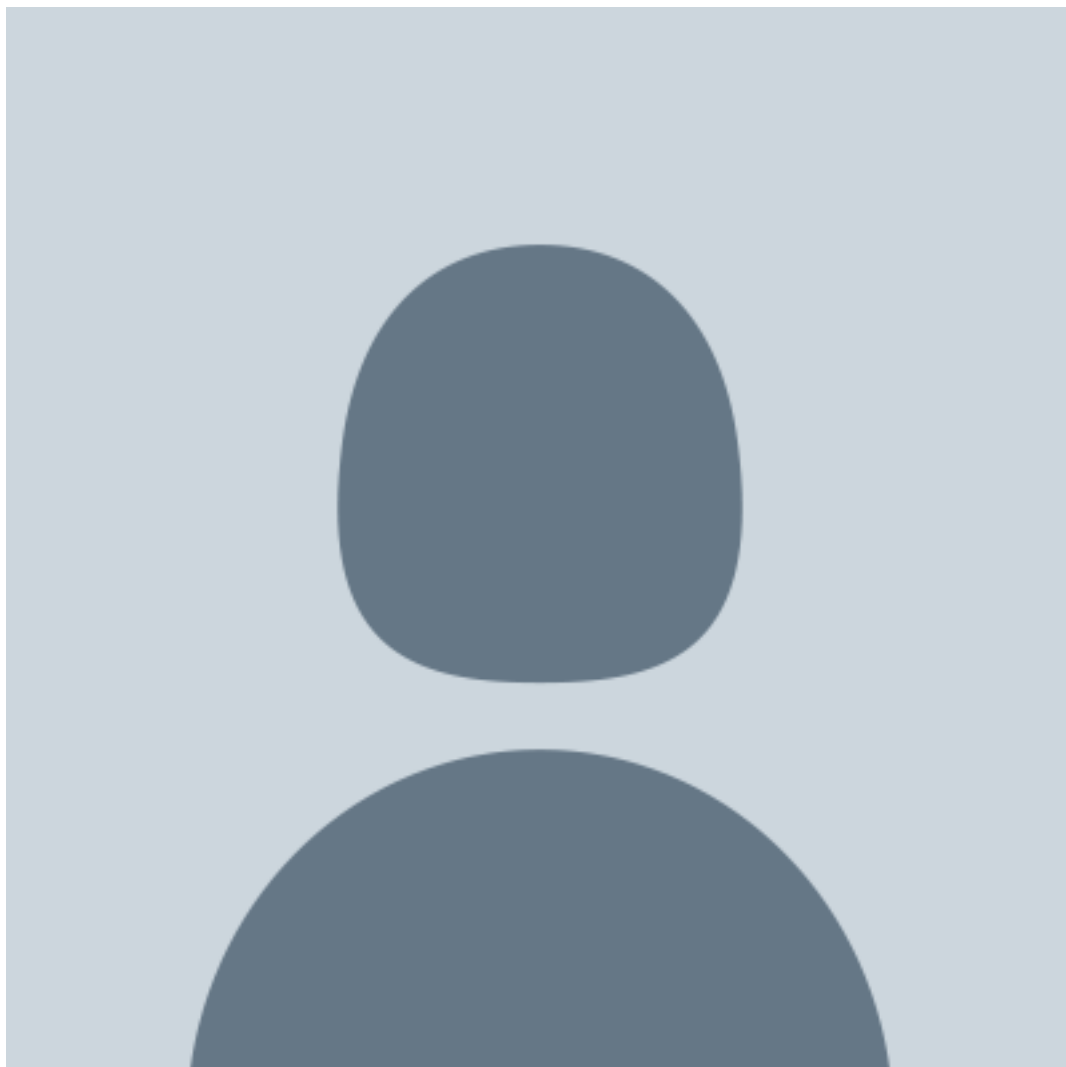
Even with a relatively short checklist like the one above, many “shadowy” areas of potential ROI begin to come to light. Redundancy and overlap of IT systems, inefficient storage practices, cost of moving data for processing and analytics, and waste of human work hours are all major factors that cannot be ignored in a thorough analysis of potential added value. But with more stakeholders involved in the IG discussion, even more cost factors should become evident, which is why any serious article on information governance needs to reiterate the need for good communication in the IG planning process. For each high-level person left out of the IG conversation, you’re likely also leaving out several areas of consideration that would have impact on the ROI, not to mention on planning of a comprehensive IG strategy.

So these are just some of the areas that you can quickly identify that may justify the ROI. The deeper you dive, the more benefits you will discover that justify the initiative. Think outside the box: it’s often some of the subtler benefits that are neglected in an ROI calculation and yet are quick to add up. The bottom line is that you cannot afford to sit idle. Maintaining the status quo rarely leads to high-performance business results. Time is money.

As the old saying goes, “You have to spend money to make money.” I propose an updated enterprise twist: “You have to spend some money to save tons of money.”

Much appreciation goes out to Paige Bartley of ZL Technologies for her contribution and editorial efforts.

[Linda G. Sharp](#)



Associate GC

ZL Technologies

Linda G. Sharp is an expert in the areas of information governance, management and eDiscovery. She has spent over three decades in the legal profession and over 15 years focusing on data management initiatives. Sharp has counseled many federal and state agencies, members of the bench, as well as Fortune 500 entities on the issues around eDiscovery, information governance, data privacy and security. She has partnered with large enterprises as they strive to resolve electronic data, privacy and security issues while balancing an ROI. Sharp is a member of the California State Bar and holds an MBA.

