

A CommVault® White Paper

Universal Search Yields Maximum Business Value

*Improving Access to Vital Information Delivers
Widespread Business Intelligence Benefits*

Contents

Executive Summary.....	3
Closing the Information Gap.....	4
What's Your Data Worth?.....	6
Universal Search and Discovery.....	7
Universal Truth about Universal Search.....	10
About CommVault.....	10

Executive Summary

Increasingly, companies of all types and sizes are relying on electronically stored information (ESI) to run their businesses and gain competitive ground through increased knowledge sharing, improved collaboration and more efficient communication with employees, partners and customers. As more and more organizations migrate toward paperless environments, this dependence on digital data deepens—as does the demand and necessity to find information quickly and easily, regardless of where it resides.

For a variety of reasons, the task of efficiently identifying, finding and obtaining all sorts of information has catapulted to the top of organizational priority lists across most market segments. Consider the common thread in each of the following real-world examples, which collectively reinforce the pervasive need for expedited information access:

- A local government office seeks a fast and efficient way to accommodate residents' requests for previously unreleased city-council meeting notes or changes to local ordinances as part of efforts to comply with the Freedom of Information Act.
- A manufacturer wants to speed technical support response times by equipping customer service reps with tools enabling them to quickly locate photos and parts descriptions in product manuals.
- The HR department of a global organization is conducting an internal investigation to get to the bottom of a “he said, she said” employee conflict.
- A hospital billing department needs to collect all data associated with specific patient ID numbers to reconcile outstanding invoices for services rendered.

The underlying theme in the aforementioned scenarios is the pressing need to shorten information retrieval times through the efficient mining of ESI. Equally important, but less obvious, is the goal to discover relevant data, in relationship to other bits of information, in order to extract real business value. Intelligent, universal search capabilities not only facilitate faster identification, collection and access to relevant information, but this functionality empowers organizations to draw correlations between the data and business process that created it in the first place.

Organizations responding to litigation and legal discovery actions were among the first to embrace the concept of intelligent search on unstructured information (e.g., e-mails, files and documents) as a way to lower eDiscovery costs and complexities. Countless examples exist of how companies have reduced or eliminated hefty eDiscovery outsourcing fees while boosting lawyer productivity and operational efficiencies with smarter, automated search functionality. Moreover, the ability to produce and review all discoverable data on demand has enabled companies to develop proactive litigation response plans and reach case assessments much earlier.



While legal teams led the way in embracing and taking advantage of newer, more advanced search technology, different departments (e.g., human resources, product development, manufacturing, customer support, etc.) across a wide variety of organizations now are catching on to the positive and far-reaching business benefits. The bottom line: organizations that can find and utilize the most relevant information quickly and efficiently are in the best position to make better, more informed business decisions.

Of equal importance in today's tough economic times is the ability to reduce costs while increasing organizational agility. The result can make the difference between success and failure. For savvy business leaders, intelligent search capabilities hold the key to unlocking the true value of information, while also uncovering untapped business potential.

Closing the Information Gap

According to Gartner, information is growing at rates of more than 70 percent per year. While this fact in itself is staggering, also consider that most of this information remains unmanaged and scattered across enterprises in many forms and formats, including financial reports, personnel files, customer records, e-mails, research results, etc. In many cases, there is little to no correlation with underlying business processes, which has led to the costly, cumbersome and risky practice of "saving everything." Unfortunately, this laborious approach makes it exceedingly difficult to locate relevant information when needed, creating an unwanted gap between users and their data, while also making unnecessary compliance risks and governance concerns a reality.

Instead of engaging in a frustrating and often unproductive game of "corporate hide-and-seek" when trying to locate desired data, it's possible to apply automated policies that classify, organize and retain information as it moves through the enterprise. Most importantly, applicable business values can be assigned along the way while redundant data with no real business value can be identified and eliminated.

For a shining example of the many benefits of streamlined information access, look no further than eDiscovery. According to Osterman Research, eDiscovery represents approximately 35 percent of most litigation costs. For that reason, companies are making large investments to improve and accelerate the process of collecting and retrieving multiple types of data with the goal of combining the results in a single, searchable repository. In doing so, they can drastically reduce the amount of data collected, which significantly reduces expensive document culling and review efforts.

As Diagram One illustrates, a plethora of eDiscovery process improvements can be realized by leveraging a consistent, repeatable and automated strategy for identifying and reducing the amount of data required for a legal review.

Another major step in closing the information gap is giving users direct access to search technologies through the use of easy-to-use, web-based tools that let authorized personnel search for content themselves within files, documents and e-mails, including attachments, across file systems, archives and backup copies.

The opportunity to empower end-users and knowledge workers with self-search capabilities produces powerful, tangible business benefits. First, it removes the bottleneck often experienced between end-users seeking information in a timely manner and resource-constrained IT departments lacking the time and staff to perform laborious search and retrieval efforts. Secondly, this self-serve approach also yields significant productivity benefits as end-users can utilize the latest search technologies to access information resources in seconds rather than minutes or hours.

In the legal world, having the ability to leverage “one click” legal hold features is another powerful tool for preserving data as part of an overarching evidence management strategy. This feature is particularly helpful when searching data from multiple retention policies that need to be retained and managed as one distinct set. A legal hold policy that spans both backup and archive also eliminates the unfortunate disposal of evidence, such as the deletion of e-mails, based on regular retention policies.

While reducing the threat of sanctions and fines associated with tardy or incomplete production of relevant ESI is one of the most visible and widely discussed benefits of universal search, it's important to consider other applications beyond the realms of legal and compliance. By taking advantage of content-inspired information management, organizations of all sorts can determine what their data is really worth while gaining the freedom to create game-changing business models and new revenue opportunities.

eDiscovery Process Improvements

- Find & collect data from offline media for litigation
- Move to proactive litigation readiness
- Discover information on laptops & desktops
- Early case assessment: go / no-go
- Execute a global legal hold policy
- Cull data prior to GC/Attorney review
- Reduce the time/risk in providing information
- Identify and classify client-attorney privileged info
- Increase lawyer efficiency in review

Firms can be better prepared by deploying automated, efficient electronically stored information (ESI) identification of all ESI across backups, e-mails and archives.

Such was the case with a major retailer that bought thousands of different products from thousands of manufacturers. By paying off all their invoices in 30 days, the retailer reaped a 10 percent savings, but keeping track of all the discounts was an onerous and time-consuming task. As a result, the company resorted to paying an outside company a six-figure fee plus a percentage to ensure all appropriate discounts were leveraged. After deploying an intelligent search capability internally, the retailer was able to elevate data management functionality to where the corporate databases could be easily identified, tracked and searched for the discounts, thereby eliminating any third-party services and turning a previous cost center into a profit center.

What's Your Data Worth?

To accurately assess the value of corporate information, enterprises need to embark on a fact-finding mission to determine what data is needed and why, while also discovering where it resides, how often it needs to be accessed as well as how long it needs to be kept for retention purposes.

To assist in this preliminary data assessment phase, it's helpful to answer a series of basic questions about how information flows throughout an organization and how it's used for various business purposes. Diagram Two identifies questions that should be asked to help gauge the importance of corporate data and then provide guidelines for building a roadmap for where the data should reside for more efficient data mining.

In most organizations, it's particularly challenging to find and classify unstructured data, such as e-mail residing in file shares or moving through web-based channels. For that reason, it's crucial for a content organization solution to review and classify all data, regardless of location or format, and then apply intelligent tags to organize everything by business need. By managing data through classification tags, even unstructured data becomes more meaningful as an organizational asset.

For an example of how classification tags work, consider the case of a product manufacturer trying to curtail a continuing leak of competitive information. It seemed that every time the company announced a new product with market-leading, unique functionality, their closest competitor responded far too quickly with an offering of its own that included copycat capabilities.

The company suspected that a member of the product development team was leaking confidential product plans to the rival, but initial internal investigations failed to produce any incriminating evidence. When the company dug deeper with the benefits of classification tags, however, e-mails were unearthed that had been sent to the competitor linking details on the latest development project with testing results on an advanced feature set. By assimilating these two disparate e-mail strings, the company was able to identify the individual responsible for disclosing the top-secret information.

Questions to Help Assess Data Value

- 1.** What data is pertinent to a pressing business requirement and can it be found easily?
 - Not just in an archive, but backups and live content as well?
 - Not just an e-mail, but documents, SharePoint, etc.?
- 2.** What is the company's data/document retention policy and how well is it enforced?
 - Where and how long are e-mails stored?
 - Users desktops, laptops, file shares in PST or offline folders?
- 3.** What is the company's policy for disposing data once the retention period has expired?
 - Not just within an archive, but across backup media as well?
- 4.** How long would it take to perform these actions against:
 - Backup data?
 - Archive data?
 - Live data?

By answering a series of questions, companies can identify the "right" information that needs to be organized and classified based on business and governance needs.

Another plus with intelligent content tagging is the ability to automatically group certain pieces of information in a more logical way to ease determination of its relevancy and determine records of business for further analysis. While this feature is commonly used for preserving relevant items in case files or legal hold sets, it can also be used to consolidate information in one central location, such as a company's document management or collaboration environment, including Microsoft SharePoint.

In this fashion, virtual workgroups can be created to expedite the review of searched, classified and tagged content without providing full access to the underlying search capabilities. For instance, a paralegal can initiate a search and then post business records in SharePoint for review by HR and business-line managers without them seeing everything produced during the initial inquiry. This level of functionality and confidentiality has broad game-changing applicability as it enables a company to take advantage of an existing asset to process information quickly without compromising data security.

Universal Search and Discovery

Once corporate content, encompassing multiple, different types of data, is classified, organized and combined in a single, searchable archive, it's now possible to find it easily and efficiently using a single index. As depicted in Diagram Three, CommVault® Simpana® software facilitates searches from a virtual, universal pool of discoverable information, including:

- Online and offline data,
- Archive copies,
- Backup copies,
- E-mail and attachments,
- SharePoint documents and objects,
- File system data, and
- Data stored on disk and tapes.



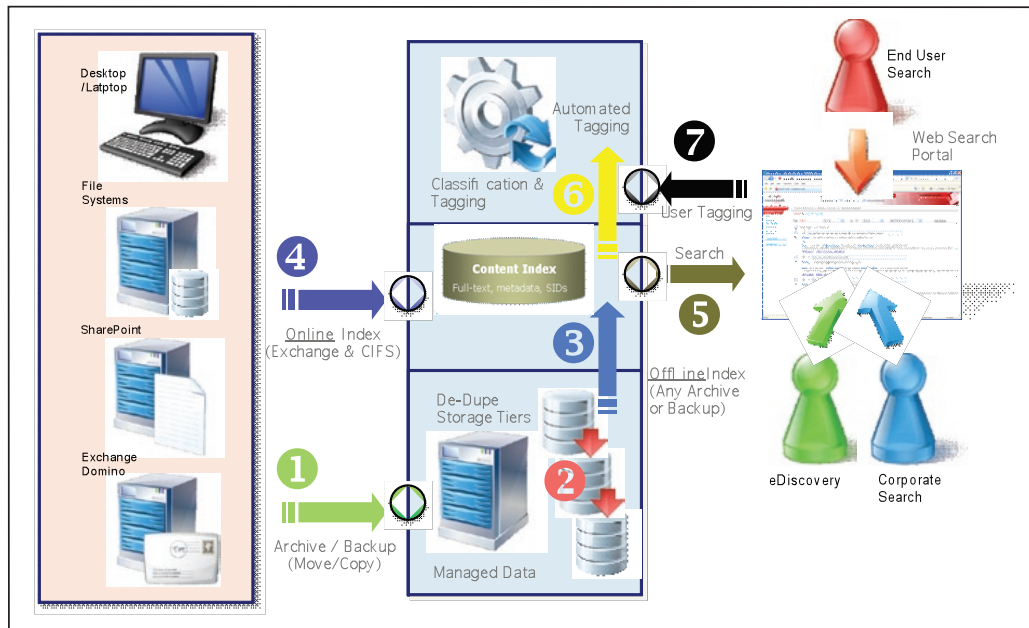
With a single search query, Simpana® software searches for files and e-mails using simple and advanced criteria, including search by file/message attributes, content keywords and phrases along with data policy properties including timestamps and job ID numbers. For a Midwestern hospital with a strong multi-national employee base, the fact that Simpana® data searches span 77 languages and 370 data types enabled transcending language barriers to locate relevant information.

Additional benefits can be derived from the ability to ingest data and recognize it in its native format, such as invoices, social security numbers, patient ID numbers, etc. A large university in the southeast, for instance, took advantage of this functionality to improve communications with its dispersed student population. After conducting a series of searches based on student ID numbers, the university was able to group students based on previous coursework so tailored e-mails on new classes could be disseminated according to previous areas of study. As a result, the university was able to increase enrollment in new classes while reducing overall marketing communications costs.

The overarching goal of universal search is to slice and dice data in as many ways as possible to improve the quality of search results. Adding further refinements to search criteria, through the use of Wildcard and Boolean search methods as well as alternative spelling suggestions, also can boost search-result accuracy. In some cases, sophisticated capabilities, such as search-within-a-search, search navigation or data mining and proximity-search, are available to deliver higher levels of search efficiencies.

Automated data mining includes classification and intelligent tagging capabilities along with automated analysis and trending analytics in order to group information into specific categories so appropriate actions can be taken. Applying automated policies improves both search accuracy and efficiency since the burden of making judgment calls on where and how to group certain types of content is removed from the user. Instead, classification decisions are reached once sophisticated analytics and trending information is applied to the process.

Once data has been classified effectively, it's possible to use a variety of advanced search functionality, such as "search-within-a-search" to efficiently address search requirements that grow both in scope and complexity as a lawsuit unfolds. For a New York-based company involved in a lawsuit, an initial search involving keywords expanded dramatically during the course of the eDiscovery process to encompass more than 90 search terms. Since a third-party had been retained for the first set of searches, the company realized it would face undue costs and time delays if it approved redoing all the searches for the expanded terms.



End-users, legal teams and compliance groups benefit from front-to-back search workflow that enables them to pull data from multiple online and offline sources for rapid indexing, searching, retrieval and review.

Luckily, the company had just completed a Simpana® Search deployment, so instead it leveraged the software's innovative "search within a search" and "search navigation" capabilities to quickly refine and narrow searches by applying more targeted criteria and mining into search results using 5 different content inspired suggestions. In doing so, the company was able to broaden its search, identify the most relevant results using intelligent navigation while controlling costs and avoiding unnecessary delays.

Proximity searches are equally effective in helping organizations find words that appear close to one another in documents and e-mails so they can be reviewed together in order to better understand the full context of the communication. For a government defense contractor dealing with highly sensitive, classified data, this feature provides the ability to quickly identify potentially sensitive communications that could spark national security concerns.

Apparently, certain typical phrases used by the contractor to describe project status are problematic when paired with language of a more sensitive nature. By performing proximity searches on a regular basis, the contractor can isolate these e-mails and review communications with its full, intended meaning in order to maintain a firmer grip on any security concerns or exposure risks.

Universal Truth about Universal Search

In today's data-driven world, the escalating demand to find and leverage information of all kinds transcends industries and geographical boundaries. More than ever, companies want and need to get more out of their information assets to drive smarter business decisions and improve knowledge sharing. The hardest part remains figuring out how to control and manage the data flow from a multitude of sources so it can be tamed and streamlined to enhance decision making and increase productivity as well as protect and improve the bottom line.

The advent of automated, smart content classification and intelligent, universal search and mining technologies opens the door to efficiencies and capabilities that could only be dreamed of just a few years ago. But as with many new concepts, the first step is to understand how it works and then determine how this can bolster the bottom-line business. One thing is for certain: data continues to proliferate at almost unfathomable rates, and the companies that will be best equipped to deal with this reality are the ones that are transforming their business practices today.

About CommVault

A singular vision—a belief in a better way to address current and future data management needs—guides CommVault in the development of Singular Information Management® solutions for high-performance data protection, universal availability and simplified management of data on complex storage networks. CommVault's exclusive single-platform architecture gives companies unprecedented control over data growth, costs and risk. CommVault's Simpana® software suite of products was designed to work together seamlessly from the ground up, sharing a single code and common function set, to deliver superlative backup and recovery, archive, replication, search and resource management capabilities. More companies every day join those who have discovered the unparalleled efficiency, performance, reliability, and control only CommVault can offer. Information about CommVault is available at www.commvault.com. CommVault's corporate headquarters is located in Oceanport, New Jersey, in the United States.



www.commvault.com ■ 888.746.3849 ■ E-mail: info@commvault.com

CommVault Worldwide Headquarters ■ 2 Crescent Place ■ Oceanport, NJ 07757 ■ 888-746-3849 ■ Fax: 732-870-4525

CommVault Regional Offices: United States ■ Europe ■ Middle East & Africa ■ Asia-Pacific ■ Latin America & Caribbean ■ Canada ■ India ■ Oceania

©1999-2009 CommVault Systems, Inc. All rights reserved. CommVault, CommVault and logo, the "CV" logo, CommVault Systems, Solving Forward, SIM, Singular Information Management, Simpana, CommVault Galaxy, Unified Data Management, QiNetix, Quick Recovery, QR, CommNet, GridStor, Vault Tracker, InnerVault, QuickSnap, QSnap, Recovery Director, CommServe, CommCell and ROMS are trademarks or registered trademarks of CommVault Systems, Inc. All other third party brands, products, service names, trademarks, or registered service marks are the property of and used to identify the products or services of their respective owners. All specifications are subject to change without notice.