WHITE PAPER

Best Practices for Managing Archive Migrations

An Osterman Research White Paper

Published June 2014

Osterman Research, Inc.
P.O. Box 1058 • Black Diamond, Washington • 98010-1058 • USA
Tel: +1 253 630 5839 • Fax: +1 253 458 0934 • info@ostermanresearch.com
www.ostermanresearch.com • twitter.com/mosterman

EXECUTIVE SUMMARY

Many IT organizations are faced with shrinking or static IT budgets while their legacy archives continue to be expensive to maintain, difficult or impossible to upgrade, and falling short of needed functionality. New data responsibilities have arisen that many legacy systems are not capable of addressing, including retention laws, increasing security/privacy requirements, and more inclusive eDiscovery responsibilities and expectations. Over the years, organizations have tended to address these problems with short-term fixes. This strategy has produced numerous siloed data repositories spread across the enterprise that are not compatible with other systems, causing reduced productivity, higher costs, ineffective information management and increased risks.

To address these mounting problems, many organizations are now planning infrastructure upgrades. These upgrades drive the need to migrate huge amounts of data, a project many IT organizations are currently facing. In fact, IDC reported that at the end of 2013, 60% of large enterprise IT projects consisted of data migrations[i]. Though necessary, data migration projects are notoriously difficult, time consuming and costly when things go wrong.

The three biggest concerns organizations face during data migrations are:

- The risk of downtime or extended downtime with the ensuing impact to the business.
- The migration process itself and its impact on people and resources.
- Large budget overruns of the migration project.

Keys to a successful archive migration project include creating the most appropriate migration strategy and detailed migration plan, fully understanding the capabilities and limitations of the current legacy archive, choosing the best migration software to match operational and legal requirements, and communicating with affected employees regularly.

ABOUT THIS WHITE PAPER

This white paper discusses a variety of challenges around migrating legacy archives and also offers a selection of choices and recommendations for improving the archive migration process. The paper also provides a brief overview of its sponsors – Archive360, GWAVA, OpenText, QUADROTech, Smarsh and TransVault – and their relevant solutions.

Many IT organizations are faced with shrinking or static IT budgets while their legacy archives continue to be expensive to maintain, difficult or impossible to upgrade, and falling short of needed functionality.

OVERVIEW AND INTRODUCTION

As the majority of corporate information transitioned from mostly paper to almost completely digital, the management of this tidal wave of digital information became an opportunity and a liability for organizations. Many simply ignored this huge influx of digital data, instead letting their individual employees "manage" it – that is until regulatory agencies and the courts took notice.

REGULATORY DRIVERS

Beginning in the United States in 1999, the SEC and NASD (now FINRA) began paying much closer attention to the financial services industry, specifically broker-dealers. The articulation, application and extension of existing regulations required broker/traders to have their communications (emails) with customers archived and made available to regulators, as well as audited via supervision procedures. Other regulatory agencies followed suit and placed archiving, retention and governance requirements on many other industries, including healthcare, energy, transportation and life sciences. Additional, broader federal regulations like Sarbanes-Oxley (SOX)

and numerous employment-related data retention requirements widened the scope of archiving requirements to include most companies.

LEGAL DRIVERS

While this new regulatory focus compelled many organizations to begin actively managing and archiving their electronically stored information (ESI), additional pressure surfaced from another area, the legal system. The 2006 amended Federal Rules of Civil Procedure (FRCP) officially categorized ESI as "discoverable", placing new and costly requirements on civil lawsuit participants. Organizations quickly found that searching for and finding relevant electronic documents in the terabytes of mostly unmanaged unstructured data was very difficult, time-consuming and costly.

In addition to US requirements, there are a wide variety of discovery-related requirements outside of the United States, as well. For example, courts in England and Wales can require some type of standard disclosure – namely, the disclosure that a document "exists or has existed". The recipient of the disclosure has a right to inspection of the documents, albeit subject to a variety of restrictions[ii]. However, in April 2013 the UK Civil Procedure Rule 31.5 went into effect, permitting courts more discretion when ordering disclosure. Some of the rules in England and Wales are similar to the FRCP in the United States, such as the requirement to disclose relevant documents and the applicability of the Rule to electronic content[iii]. In most European nations litigants are not required to produce content that runs counter to the claims they make in a legal action. Requirements in the United Kingdom, however, can compel organizations to produce damaging content, but only after a court order[iv].

Driven by the new regulatory and eDiscovery requirements, many organizations began adopting third party email and document archiving solutions. These early archiving systems have now reached the end of their useful life and are becoming expensive and risky to keep active.

FUNCTIONAL CONSIDERATIONS

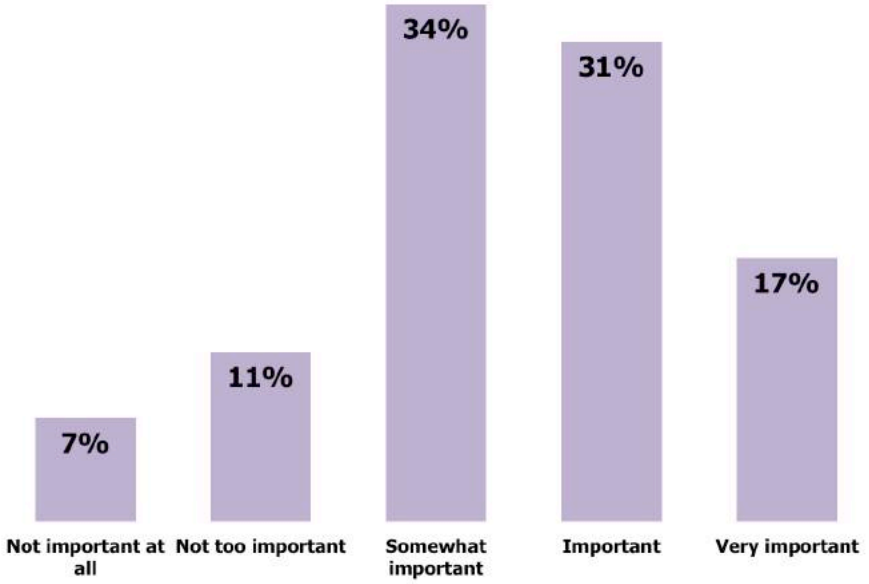
In addition to the regulatory and legal considerations that motivated organizations to archive is the functional rationale, most notably the need to reduce bloat on email servers while not capping storage for end users. Because many users employ email folders as long-term data storage for the documents and other information needed to do their day-to-day work, archiving offers a reliable means of migrating this content away from "live" storage – and thereby reducing email bloat – while giving users continued access to their older content.

The cost and effort to maintain legacy archiving systems is out of proportion to what could be achieved with modern storage systems or cloud solutions. Consequently, organizations are faced with maintaining old archiving systems that were not designed for the performance levels and additional requirements organizations need to effectively manage the huge amounts of ESI they work with now. This growing gap in performance and capability is driving the necessity to migrate large archive repositories to newer, more powerful systems. Organizations with legacy archives are now facing the prospect of migrating terabytes of content to more capable repositories and applications.

It is also worth noting that for email in particular, the user's "total mailbox" is actually stored in a combination of the email platform (for example Microsoft Exchange), the email archive (for example Enterprise Vault) and PST files. As such, projects to migrate archive data should consider the impact on the "total mailbox" and the dependencies between the discrete components, as shown in Figure 1.

The cost and effort to maintain legacy archiving systems is out of proportion to what could be achieved with modern storage systems or cloud solutions.

Figure 1
Importance of Coordinating Archive Migration with Production Systems
Survey conducted May 2014



Source: Osterman Research, Inc.

ORGANIZATIONS' PLANS FOR ARCHIVE MIGRATION

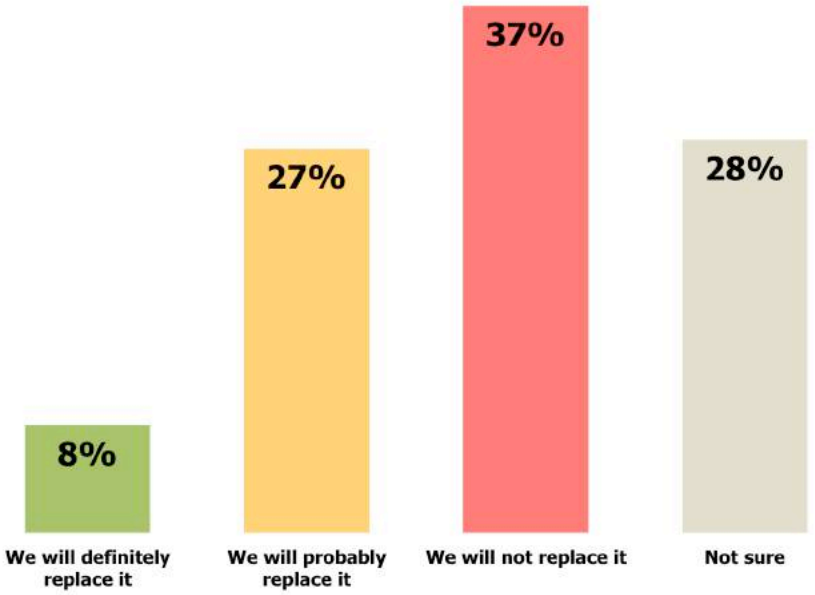
In late May 2014, Osterman Research conducted an in-depth survey of 164 organizations and their archiving system migration plans. We surveyed primarily mid-sized and large organizations across a wide range of industries. Key findings from the research, all of which will be published in a separate report, include the following:

- The typical archiving solution has been in place four years and eight months (median is 36 months).
- There is not a high level of satisfaction with current archiving solutions. For example, only 60% of organizations are "pleased" or "extremely pleased" with the current archiving solutions' ability to place legal holds on content, only 52% are this pleased with the speed of the solution's search performance, and only 44% are this pleased with the ability to delete content when necessary.
- Moreover, we found significant differences in the level of satisfaction with archiving solutions based on their age. For example, organizations with archiving systems that are more than three years old are nearly twice as likely "not to be pleased at all" with their ability to place legal holds on content (14.5% for older systems vs. 7.6% for more recent systems), the ability to establish different retention policies (16.7% vs. 11.0%), and the scalability of the system (15.2% vs. 11.2%).
- We also discovered a significant difference in the penetration of cloud-based archiving based on the age of the system: organizations with an archiving solution no more than three years old have placed 33.4% of their archived content in the cloud compared to only 13.2% for older solutions.
- Finally, we found that 7.6% of the organizations will "definitely" replace their archiving solution over the next 18 months while another 27.2% will "probably" do so, as shown in Figure 1. Not surprisingly, organizations with older archiving

Only 60% of organizations are "pleased" or "extremely pleased" with the current archiving solutions' ability to place legal holds on content, and only 52% are this pleased with the speed of [its] search performance.

solutions in place are much more likely to definitely or probably replace their archiving solutions during the next 18 months (39.8% vs. 30.1%).

Figure 2
Organizations' Plans for Archiving Migration Through November 2015
Survey conducted May 2014



Source: Osterman Research, Inc.

SHOULD YOU MIGRATE YOUR ARCHIVES?

Data migration is the process of transporting (and converting) data between computers, applications, storage devices or formats. Most don't look forward to the task of migrating a long-standing repository, especially since many data migrations are centered around legacy archives. When performing an archive migration, time should be taken to fully evaluate, cull, and "clean up" the current archived data, including defensibly disposing of content no longer needed.

There are many reasons that trigger the need for an archive migration, including a data center move or consolidation, an upgrade to a more efficient archive and storage technology, a move to the cloud, a need to improve litigation support capabilities, a requirement to address regulatory compliance issues, desktop refreshes, adoption of mobile or email infrastructure upgrades, or archive obsolescence. No matter the reason, an archive migration can be a complicated, risky and costly endeavor if not approached with the right tools and expertise.

DATA CONSOLIDATION

Application Retirement

Obsolete or unused applications that remain active in an organization's infrastructure can be a large resource drain, when one includes hardware and annual hardware support costs, annual software support contracts, and additional personnel needed to service the applications. This situation is even more challenging when underlying technology, such as Windows XP and previous versions of Microsoft Exchange, reach end-of-life.

When contemplating the retirement of an application, the fate of the existing data associated with it must be addressed. The most obvious issue is around regulatory

Eight percent of organizations will "definitely" replace their archiving solution over the next 18 months while another 27% will "probably" do so.

retention and litigation support requirements. If any data in the application repository is still within a regulatory retention period or is potentially relevant to an anticipated or current lawsuit, then that data must be kept, carefully migrated and secured. Once the regulatory and legal issues are addressed, the data's business value should be considered. In many situations, much of the data could still be of value to the organization and so should be migrated to another repository.

Storage Consolidation
As storage resources age and become less efficient, they are replaced with higher capacity and more technically advanced storage devices. Storage consolidation can reduce the number of devices under management, reduce hardware support costs, and can increase overall system performance. When starting a storage consolidation project, migrating existing data to the new storage resources requires a carefully considered and planned data migration process to ensure reliable file and metadata conversion, full data access, and performance. Once the data migration is complete and verified, old storage resources can be retired or repurposed.

Archive Consolidation
Many times organizations find themselves with multiple incompatible legacy archives including an email archive, a file system archive, several SharePoint instances, and a content management solution.. This raises costs, increases legal hold and eDiscovery risk, and it reduces productivity. Consolidation into a single, higher performance archive would produce a positive return on investment (ROI) based on federated higher performance search capability and fewer hardware and software support costs.

Another common reason for archive consolidation is corporate mergers or acquisitions. When companies with existing archives are purchased by another company with the same or different archives, consolidating the separate archives into a single platform provides the same ROI mentioned above – fewer platforms to support pay license fees on, as well as reduced cost and risk associated with eDiscovery (fewer resources to search for legal hold and collection).

..an archive migration can be a complicated, risky and costly endeavor if not approached with the right tools and expertise.

Orphaned And Legacy Data Consolidation
An orphaned file is a data file that is no longer associated with any current application. Orphaned files can also include data items that belong to 'leavers', i.e. individuals that have since left the organization, but whose data is nevertheless important to retain.

Orphaned files normally occur when an old application is retired without addressing data files associated with the retired application. They can also occur when indexes to archive storage get corrupted. Over time, these unassociated inactive files can accumulate.

Legacy data includes all information considered inactive – data that is stored in various electronic formats and not currently used or managed. This can include huge amounts of data stored in email systems, files and data repositories originally retained for specific reasons such as disaster recovery, business needs, regulatory retention, or placed on legal hold and never released. Orphaned and legacy data files can still be subject to regulatory or legal hold/eDiscovery requirements and so should be reviewed before disposal. Data in question or determined to be subject to retention should be carefully migrated to a secure repository.

Consolidating And Migrating Data From Abandoned Or Obsolete Email Archives
Beginning in 1999, hundreds of email archiving vendors sprang up selling email archiving solutions. The problem now is that many/most of these email archiving vendors have gone out of business or have been bought by other companies that have discontinued support for these legacy email archives, leaving current customers with a tough choice:

- Abandon the email archive, a risky decision because of regulatory and litigation support requirements, or

- Migrate the archived email data and cull it over time. Migrating and filtering/culling the abandoned or obsolete archived content is the only way to absolutely determine what data can be defensibly disposed.

UPGRADING TO HIGHER PERFORMANCE, MORE EFFICIENT ARCHIVES

Many archiving systems have become outdated because of new security/privacy requirements, access control requirements, new file formats, and additional reporting requirements and eDiscovery capabilities, as well as advances in both hardware and software for improved capabilities. Legacy archiving systems are often a barrier to benefiting from new operating system, hardware performance and functional improvements. In today's business climate, organizations need to take advantage of all improvements to stay competitive and not run afoul of the courts and regulatory agencies.

Improved Search Performance

Older archiving systems relied on even older indexing and search technology. As archived data sets became larger, search response times became less reliable and much slower. For many companies, search performance became a major bottleneck and liability. The only sure way to improve the situation is to upgrade to a higher performance archiving solution. An integral part of this upgrade process is the migration of the existing archive data set in such a way that takes advantage of the new system capabilities with the existing archived data set.

Improved Scalability

Scalability is often overlooked or at least underestimated. Many legacy archiving customers have found out too late that the archiving vendor's promise of unlimited scalability and performance has fallen short. In many instances, these scalability shortcomings have arisen as the number of customers has grown at a rapid pace, thereby increasing the amount of content archived; or were faced with an especially large eDiscovery requirement and found that eDiscovery searches were taking days or weeks to finish. These situations can place the organization at risk and dramatically increase costs. The only option is to move to a higher performance archive with room to grow.

Improved Platform Support

Another area where legacy archiving solutions fall short of customer needs is with respect to staying current with platform support. Many organizations have found their planned infrastructure upgrades being held back by lack of archive vendor support for the latest operating systems and server and client applications. A general trend in the market for vendors to take over competing solutions and then lack focus on their development has led to the need to migrate to alternative solutions.

Legacy archiving systems are often a barrier to benefiting from new operating system, hardware performance and functional improvements.

ENHANCED LITIGATION PREPAREDNESS

As has already been discussed, due to architecture, performance levels, and/or functionality, older archiving platforms are often unable to execute specific, litigation-related tasks, causing additional expense and putting the organization at risk.

eDiscovery

eDiscovery is the process by which electronic data is searched, located, secured, reviewed and turned over to the other side in a legal action in response to an eDiscovery request. eDiscovery was further complicated in December 2006 with the issuance of the amended Federal Rules of Civil Procedure (FRCP). The amended FRCP documented two important points; 1) electronic data is discoverable and, 2) the duty to begin preserving potentially relevant information starts when litigation can be reasonably anticipated. The amended rules set the expectation that companies have a much higher level of control of their electronic information than previously

expected. The ability to respond to an anticipated or actual lawsuit with complete data collections, a fast and inclusive legal hold, and a detailed document review in the expected timeframe can mean the difference between winning or losing the lawsuit.

Legal Hold

One of the most important litigation preparedness requirements is that all potential parties to an actual or anticipated lawsuit find and secure all potentially relevant content with a secure legal hold, quickly – ensuring that evidence destruction (spoliation), including inadvertent deletions, cannot take place. In many situations, older archive solutions do not allow for granular content search and legal hold, requiring the entire archive to be put on hold until the discovery phase has passed. Most corporate attorneys are not comfortable putting an entire archive on legal hold simply because a small percentage of the archive *may* be responsive to a specific lawsuit.

Early Case Assessments

A key phase in the eDiscovery process is Early Case Assessment (ECA), the process of reviewing potentially relevant case data and evidence to estimate risk, cost and time requirements. ECA is used to set the appropriate go-forward strategy to prosecute or defend a legal case, addressing the issue of whether or not an organization should contest or settle a case. In older archives with basic indexing and slow search performance, early case assessment can be incomplete and time consuming, and because of the timeframes involved, may not leave decision makers with enough time to properly review evidence to create an appropriate case strategy. For organizations with normal or heavy litigation profiles, migrating archived data stores to higher performance archive repositories can quickly produce cost savings and reduce risk of inappropriate case strategies.

IMPROVED REGULATORY COMPLIANCE

For highly regulated companies like healthcare providers, energy companies, broker-dealers, investment advisors and pharmaceutical manufacturers, regulatory oversight has become a way of life. New, more prescriptive regulatory requirements centered on corporate data are becoming law every year. Many organizations employed archiving systems over the years to specifically address regulatory retention requirements. As those laws were updated and new ones adopted, data retention became just one of the requirements to be met. New laws require companies to actively protect and audit the access of customer personally identifiable information (PII) and protected healthcare information (PHI). Many older archives were not architected with these higher levels of security in mind. For those companies that cannot (or will not) protect that information, lawsuits, huge fines, and damaging PR can be the result. Migrating archived data to more secure repositories is the logical step to reduce that liability.

The financial services industry in many countries has specific regulatory requirements targeted at auditing employees in particular roles to ensure customer and other interactions meet professional and regulatory guidelines. In the United States, for example, FINRA recently adopted rules (FINRA 3010 and 3120 replacing NASD Rules 3010 and 3120) that, in part, require companies to setup systems to actively monitor and audit (supervise) the communications of selected employees to protect against market abuse, as well as to ensure professional conduct with customers. This particular audit and supervision requirement demands efficient and demanding workflows.

This supervisory workflow has been adopted by other non-financially regulated organizations for the purposes of auditing employee adherence to corporate Internet, email and instant messaging use policies. Without more advanced archiving systems built for these types of advanced workflows, supervisory functionality is impossible.

Another area in which legacy archiving solutions fall short of customer needs is with respect to staying current with platform support.

MIGRATING TO A CLOUD ARCHIVE

Cloud archiving has become a viable alternative to siloed on-premise archiving platforms. A cloud archive can offer an optimized storage solution for long term retention along with additional cloud managed services, such as more advanced access security, regulatory policies and data management, and more efficient litigation support services. The cost savings of cloud archiving versus on-premise archiving can be obvious when the upfront costs of hardware, software, additional experienced personnel, and annual support are taken into consideration.

Because of these differentiators, many organizations are moving their archiving requirements to the cloud, requiring current archive on-premise repositories be migrated as well to realize the benefits from cloud archiving adoption.

ARCHIVING OF NEW DATA TYPES

Since many of the early archiving systems were introduced, many new data types have been introduced that also should be archived and made searchable for the regulatory and eDiscovery reasons already covered. These new data types include various social media platforms, collaboration applications, work files, share drives, instant messaging, voice calls, video conferencing, and unified communications. Many of these different content types can be related and benefit from systematic grouping, much like being able to automatically construct an email conversation thread in the past. Obviously, older archiving solutions were not able to address different and unknown data formats and usually were targeted at one specific platform, such as email archiving. Organizations are moving to more format-inclusive archiving systems and so are considering the migration of older archives into the new solutions to enable more expansive information capabilities.

Cloud archiving has become a viable alternative to on-premise siloed archiving platforms.

MANAGING THE MIGRATION PROCESS

Archived content can be complex. Archive date, folder structures, assigned retention periods, legal status, content metadata, and content access logs can be valuable to the organization, as well as part of a responsive legal record. If legal considerations will play a part in a migration, it is considered a best practice that all aspects of an archived record remain unchanged during the migration process. This section discusses industry best practices when migrating archived data.

PLANNING THE MIGRATION

Regardless of the reason for the archive migration, it is always a good idea (and best practice) to develop a detailed migration plan before software and equipment are purchased or the migration is begun. The migration plan should include:

- **The Migration Strategy**
  As detailed in the next section, there are various strategies that organizations have followed to successfully perform data migrations. *The Phased Migration*, *Date Forward Migration*, *Immediate Cutover Migration and Sync and Switch* all define high level approaches to the process. Once an appropriate strategy is chosen, detailed planning can begin.

- **Objectives**
  The purpose and goals for the migration – for example: *the purpose of this archive migration is to reduce overall IT support costs while enabling employees faster, more accurate access to their archived data by migrating the entirety of the legacy email archive into our cloud-based Office 365 instance no later than August 17th with less than a 3% loss of data.*

- **Migration Project Timeline**
  A list of start and end times for major processes involved in the migration program. A high level timeline can be created and communicated to those

effected by the migration, and a more detailed timeline can be used by those managing the actual migration.

- **Risk Analysis**
  A process for defining and analyzing the dangers to individuals, businesses and customers posed by the migration process and creating alternate action plans if disasters do occur.

- **Resource Plan**
  An analysis of the resources (people, equipment and materials) that will be needed and in what quantities to successfully perform the migration.

- **Project Budget**
  The expected cost of the migration project including any additional administration, hardware, network and storage resources.

- **Communications Plan**
  The communications plan describes the migration project, timeline, projected levels of impact (downtimes, etc.), and contact information when questions or unplanned issues arise. The target of the communications plan are the project stakeholders, including management and all employees that could be affected by the project.

- **Technical Considerations**
  As detailed in the next section, the technical details that should be considered beforehand so that delays are not a factor in the migration. These include data format, bandwidth, legal requirements, etc.

DATA MIGRATION STRATEGIES

There are several possible migration strategies that organizations can follow depending on their budgets and willingness to accept higher risk. These migration strategies include:

There are several possible migration strategies that organizations can follow depending on their budgets and willingness to accept higher risk.

Phased Migration
In the phased migration, smaller groups of employee data is migrated and uploaded to the new system. This approach has the benefit of affecting end-users the least while providing them access to their archived data as soon as possible. An added benefit is that if problems occur during migration, reverting to the legacy system can be accomplished quickly. The downside of this approach is that it lengthens the entire migration process since both the old archive and the new archive reside in the enterprise at the same time.

Date Forward Migration Strategy
This migration strategy employs a parallel process that keeps both the old system and new archive system running for longer periods of time. In this strategy, the new archive system is kicked off and begins archiving on day one moving forward. Employees must access the new archive for data archived after day one and the old archive for files archived before day one. The intention is to eventually migrate the old archived data to the new archive at some point in the future. The main benefit of this strategy is to bring the new archive up and start using it very quickly. The other benefit is that the old archive is still available and accessible so that if problems occur with the new system, reverting to the old system can be done quickly. The downside is that end-users have two systems to access and search for content, as well as the increased strain on the infrastructure. Wherever possible, the time period during which users must access both archives should be kept to a minimum.

Immediate Cutover Migration Strategy
The next migration strategy is one in which all data in the old archive is migrated into the new system, verified and then the old system is shut down. The benefit of the

immediate cutover strategy is to complete the migration as quickly as possible. The downside is the higher risk involved. If something goes wrong in any part of the process it can be difficult and expensive to go back.

Sync and Switch Strategy
The last strategy is designed to allow rapid migration of users who can continue to access the old archive while their own data is migrated. There is then a virtually instantaneous switch where the user is switched to the new archive, short cuts are updated in their mailbox, any "final" archive items are moved and the old content is (optionally) removed. This approach has the advantage of rapid migration of users and unlike other strategies it does not require users to access multiple archives, nor does it place the entire archive at risk.

A best practice for any migration process is to backup all data to be migrated and keep it safe until the migration has been determined to be successful.

CONSIDERATIONS BEFORE STARTING A DATA MIGRATION

The success or failure of migrating large data archives can be influenced by many factors. Understanding and addressing these potential migration issues ahead of time will better ensure a successful migration project.

FUNCTIONAL CONSIDERATIONS

Archive Data Format
The data format of the archived content is an important factor when migrating data. A few archive solutions would convert the archived content to a format designed to ensure more efficient storage and recoverability over very long periods of time. For example, some older email archiving solutions would separate an email message from the attachment. Other email archiving solutions would "containerize" a large number of emails into groups for storage efficiency. A key when migrating archived data is to verify that the migration solution is able to "understand" and completely restore the archived file to as close to its original state as possible including, in the case of eDiscovery and regulatory response, folder structures and all metadata.

A best practice for any migration process is to backup all data to be migrated and keep it safe until the migration has been determined to be successful.

Archive Data Age
The age of the archived data will relate to the data format consideration mentioned above – can the migration application technically work with the particular data format? The other data age consideration centers on business or regulatory retention periods and eDiscovery requirements. Industry best practices suggest that expired data should be disposed of as soon as it ages beyond its retention period AND is not potentially relevant to any current or anticipated legal actions. Disposal of expired information reduces storage costs, storage management costs and eDiscovery costs.

Archived Data Location
Where the archived data currently resides becomes an issue under certain circumstances. These include country-specific legal restrictions if the archived data is stored within the borders of a country that has laws against moving data outside its borders. The other location-related issue is closely related to bandwidth capability as discussed below.

Available Bandwidth
Network bandwidth refers to the amount of data that can be transferred from one device to another in a set period of time dependent on device interfaces and network connection technology. For example, the time needed to migrate a 20 terabyte data set using a continuous bandwidth of 1.5 Gbits/sec is 33 hours and 22 minutes. Actual data transfer rate is only as fast as the slowest piece of the migration infrastructure, which includes device interfaces. Another issue with bandwidth and large file migration is the possibility of consuming too much infrastructure bandwidth and

thereby negatively affecting other enterprise applications. When planning for a large archive migration, a best practice includes calculating the available bandwidth that can be used without affecting other enterprise functionality or, alternatively, plan for the migration during non-business hours.

DATA INGESTION CONSIDERATIONS

The destination application/repository may have limitations on the speed of data ingestion affecting the total time to migrate the data. For example, there may be little limitation on data ingestion if an organization is migrating archives from one internal archive application to another of the same archive application. The destination ingestion performance will be subject to interface performance, as well as any file size limitations, metadata, and any data conversions that are required by the new destination application.

LEGAL/REGULATORY CONSIDERATIONS

Anytime large data migrations occur, all legal and regulatory aspects should be considered. The corporate legal department should be consulted on the status of any current or anticipated legal proceedings that would affect legal hold responsibilities, or eDiscovery collection and review processes for data potentially in the archive. It's a risky undertaking to move/migrate data that is involved in a lawsuit and should be well understood. The archived data could be inadvertently lost or changed during migration triggering spoliation (destruction of evidence) accusations. The movement of legally responsive content without strict audit and oversight could introduce "chain of custody" questions into the proceedings causing data beneficial to a legal defense to be ruled inadmissible.

Data subject to regulatory requirements can pose additional risks during migration without planning and oversight. Loss of regulated data can trigger investigations, agency lawsuits, fines, and negative publicity. When migrating regulated data, care should be taken to ensure the migration process does not put the data at risk.

END USER CONSIDERATIONS

Before the migration process begins, planning should include how the data migration will affect end-users. How long the migration process will take, access rights before and after the migration, end-user notification, and help desk support during the process will affect the perceived success or failure of the archive migration.

Before the migration process begins, planning should include how the data migration will affect end-users.

Access During Migration

The question of end-user access to archived data during migration can be addressed in several ways. Communicating with end-users before and during the migration can limit employee productivity issues as well as calls to the help desk. If the migration is appropriately planned and managed in stages, unavailability of archived data can be kept to a minimum. In certain circumstances, the archived data could remain available as it is being migrated. The key is in communicating with end-users about the dates and timeframes that data may not be available, as well as how they will connect to their archived data after the migration is complete.

Access And Use After Migration

A data migration project can trigger worries of data access after the migration is complete. Questions like:

- Will I still be able to find and use my archived data?
- Will I have to go through extra steps to get to my data?
- How long will I have access to it?
- Will my data be automatically deleted without my approval?

- How will data be archived in the future, by policies or by my choosing to archive it?
- Will others also have access to it?
- Will I be trained on how to use the new archive?

The answers to these and other questions will depend on the new archiving system an organization chooses to implement and the policies it implements. Archiving industry experts have suggested some basic "rules" to consider when evaluating post migration practices. First, don't drastically change end-user use models just for the sake of change. If possible, keep usability as close to the old use model as possible. For example, moving from a file system drag-and-drop use model to one that employs a "check-in" and "check-out" model can cause productivity issues and could drive some end-users to develop "work-arounds" to avoid using the new system. If, because of new legal or regulatory requirements, the use model must be radically changed, automate the new processes as much as possible.

Second, migration from on-premise archives to cloud-based archiving systems could highlight slight differences in end-user expectations. The speed of response (latency) to a mouse click, a drag and drop operation, or the time it takes to open a file completely might take slightly longer (microseconds versus milliseconds) than before. These small differences can taint end-user acceptance if not discussed ahead of time.

End-User Training

Communicating migration plans and progress to affected end-users is an important part of a migration process to cut down on end-user frustration and costly calls to the help desk. As important as keeping end-users informed of the migration process is training them ahead of time on any new processes, procedures, use policies and automation that will affect their access and use in the future.

Migration from on-premise archives to cloud-based archiving systems could highlight slight differences in end-user expectations.

OTHER CONSIDERATIONS

Migration Time Frame

Accurately estimating the total time of the migration process is important to set expectations for both end-users as well as management. Calculating migration time should be straightforward when all considerations are taken into account. Gating items such as available bandwidth, hours per day (or night) the allotted bandwidth is available, the amount of data to be migrated, any data transfer limitation from both the original archive as well as the destination repository, and potential management issues are all variables in the migration process. One possible migration process consideration is to divide the migration into stages so that access to the entire archive is not blocked for the entire archive migration. For example, with an email archive, migrating select mailboxes or departments on specific days would lessen the negative affect to only those mailboxes being migrated on that day.

A key requirement for a successful data migration project is to calculate an accurate (and conservative) migration time frame and communicate it to affected end-users. If realistic expectations can be set and then met or exceeded, the migration will be successful.

Culling And Disposal

Another consideration when planning a large data migration is to think about culling the archived data set either before or during the process. If the current archive system is capable, some basic culling of expired or especially old content (taking legal holds into account) can speed up the migration process. It is important to consider the legal and regulatory implications of such actions. Further discussion of culling and disposal can be found in the defensible disposal section later in this paper.

Old Archive Disposition

One last step in the migration process that should be addressed and planned for is what to do with the legacy archiving solution and data set that is "left behind". In US federal and many state courts, it is the responsibility of the party being discovered to search for all potentially relevant content in a relatively timely and complete manner. During the early periods of the litigation process (no later than the "meet and confer" meeting) all possible locations of case-relevant content must be disclosed to the other side in the lawsuit. If an old archive was not shut down or repurposed (making the data completely irretrievable), the opposing side could force its owner to search the old archive for relevant content, even if the judge is told that all archived data was migrated elsewhere. The odds are low (but not zero) of this being ordered if the judge cannot be convinced that there is no chance of unique relevant data being present on the old archive. The surest way to ensure this will not happen is to delete all data in the old archive after the migration is complete and verified, and repurpose all hardware used for archive storage as soon as possible.

MANAGING THE MIGRATION PROCESS

DEVELOP OR UPDATE AN INFORMATION MANAGEMENT POLICY OR SCHEDULE

When asked if they have an information management policy and schedule, many companies will respond by saying "of course" and proudly dig out a short, heavily copied document dated at least a decade ago with a small number of entries describing some key financial documents and their retention periods. An information management policy and schedule should address all information within the organization independent of format, both hardcopy and electronic. In today's business environment, more than 95% of all organizational information is created, consumed and disposed of in a digital format. Many organizations responded to this major shift in information by either ignoring it or by capturing and archiving all electronic content. A comprehensive information management policy and schedule will help decision makes to set migration policies for what data must be migrated for legal and regulatory reasons, and what can be disposed of if it is determined to have no business value.

DETERMINE ALL REGULATORY REQUIREMENTS AND HOW THEY RELATE TO THE ARCHIVED DATA

The information management policy and schedule mentioned above will help decision makers determine what, if any, content present in the archive must be retained and migrated and any special circumstances that should be considered. For example, PII and PHI can have specific security and privacy requirements that call for specialized retention and security procedures and automation. Another example of prescriptive obligations are the technical requirements spelled out for broker-dealers as to how their communications be retained (on WORM devices) and that they be serialized and systematically audited. Important questions about the specifics of the archive should be addressed before the migration begins, such as what calendar dates the archive covers, the types of content present, and the departments and specific employees that were archived, if not the entire organization.

DETERMINE ALL LEGAL REQUIREMENTS AND HOW THEY RELATE TO THE ARCHIVED DATA

Before settings are changed within the archive or any data is migrated, check with the corporate legal department to determine if any anticipated or current lawsuits affect the target archive. If there are no anticipated or current lawsuits or other legal actions that could potentially affect data in the archive, request a written approval from legal counsel that a data migration and or additional culling of the archive is approved. Once approval is received, the archive migration can proceed. It is always a good idea to continue with the archive action as soon as possible after receiving the legal approval because lawsuits can crop up quickly and cause the migration to be

Another consideration when planning a large data migration is to think about culling the archived data set either before or during the process.

put on hold. The point to keep in mind is migrating and/or deleting data is not illegal if there is no legal responsibility (legal hold and eDiscovery) to retain it.

If, after checking with legal counsel, it is determined that there may be potentially relevant data in the target archive, they should stop all migration actions that could impact the archived data and meet with the attorneys to determine what if anything can be done going forward. In US federal and most state courts, it is the responsibility of the data owner to protect all content in its current state, including all metadata that could be relevant in a legal action (legal hold) as soon as the legal action could be reasonably anticipated. Legal hold and discovery carry with them some very specific responsibilities that, if not followed completely, can cause the case to be lost. Relevant data must be found and secured as soon as possible and that data must remain in its native state as of the date it was placed on legal hold. This also means that any tools used during legal hold, collection and review should not change the data in anyway, including its metadata.

DATA INTEGRITY

As data is extracted from the legacy archives it is important to give due consideration of the extract process. While for some simple archiving platforms it may be feasible to extract data directly at a storage level. However for more advanced solutions, e.g. Symantec's Enterprise Vault, where messages are stored as chunks and attachments are shared between different mail objects supported APIs should be used. This is to ensure integrity and preserve the meta-data important for legal and regulatory requirements.

AUTOMATED WORKFLOW

During the migration process there will be many steps that can (and should) be automated. It is important not to lose sight of the fact that over the course of the migration users must be enabled and disabled, shortcuts should be cleaned, source data deleted, mailbox moves coordinated etc. If these steps cannot be automated in conjunction with the data move the overall process will be slow, prone to error and disruptive for end users.

DATA DUPLICATION

Earlier versions of archiving solutions were not able to distinguish between duplicate data, and so legacy archives can have many copies of the same files. Including data deduplication (single instance storage) as part of the migration process will save on storage resources at the destination repository. With migrations to the cloud, data deduplication can significantly reduce the amount of data to be transferred saving on both time to migrate as well as storage costs.

Earlier versions of archiving solutions were not able to distinguish between duplicate data, and so legacy archives can have many copies of the same files.

CONTENT CULLING

Many organizations are dealing with increasing annual storage budgets. There are several reasons for these annual increases; lack of an information management policy that permits employees to save everything forever, legal holds placed but never released, huge numbers of duplicate files, and a lack of file deletion after expiration. As part of the migration process, culling archive data based on age, keyword, legal status, file type, employee status, and size can help. The Compliance, Governance and Oversight Counsel (CGOC) published a survey[v] in 2012 that revealed that approximately 69% of the average enterprise data store was of no business value and could be disposed of without any effect on the business. Disposing of unneeded data before migration to the target repository will save time and money.

DEFENSIBLE DISPOSAL

Organizations are under no legal obligation to keep every piece of information generated or received. Consequently, the regular, systematic deletion of expired or valueless information (if no legal or regulatory obligations are present) should be a regular process in any information management program. But, in today's regulatory and legal climate, disposing of data can carry a great deal of risk if not performed

defensibly. During the culling step of the migration process, documenting the steps taken to find and categorize candidate files for disposition is advisable. To be a defensible process, an organization must show:

- A consistently followed, good faith effort to find and protect any content that is subject to regulatory and legal retention rules.
- Reasonable and reliable policies, procedures, processes and technology to differentiate and categorize data for disposition.
- The lack of a duty to preserve deleted data at the time of disposal.

Tested and documented defensible disposal processes can save money and time by reducing the amount of data that needs to be migrated while reducing the risk of disposing of information that should have been retained.

THE BUSINESS ADVANTAGES OF PROACTIVE DATA MIGRATION

There are several reasons to consider proactively migrating archive data before potential problems can crop up. These include:

- **Storage Savings**
  Older archives can contain a large percentage of duplicate files. These duplicates can amount to up to 75% of the archive. Another cause of excess storage use is all of the expired content resident in an archive. As "records" reached their compliance retention periods, they were not disposed of because many archives did not have retention policy management built into them so they retained expired content until manual deletion. An intelligent migration process can quickly recognize and dispose of duplicates, as well as expired content before moved to a new archive system, thereby saving large amounts of money in storage budgets.
- **Archive Scalability**
  Proactively migrating an older archive because of known scalability problems that will impact the organization at some point in the future can head-off last minute problems like system performance issues and increased risk due to regulatory retention and eDiscovery issues.
- **Ongoing Management And Maintenance**
  Having multiple archives across an organization places a strain on the administration team and introduces risk. For example, the backup of archives may not be synchronized, knowledge of multiple archiving applications may be required by administrators, system loads will be increased etc.
- **Performance**
  Many legacy archives relied on old technology indexing and search capabilities that cannot perform the types of complex and speedy searches needed today. Additionally, as the archive grows, these old technology search engines aren't up to the task and so searches can take hours or days to complete and sometimes are not consistent.
- **End of Life**
  Many older archive systems reach an EOL or "end of life" designation for reasons that include the archive vendor was purchased by another company and will not continue with the old solution, or the archive vendor has gone out of business. In the case of an EOL situation, there would be some period of time to migrate archive data to another solution, but in the case of a vendor going out of

An intelligent migration process can quickly recognize and dispose of duplicates, as well as expired content before moved to a new archive system.

business, an organization could find its archive and organization at a higher risk because of a rapid disappearance of support.

- **Litigation Preparedness**
  Many times organizations facing a lawsuit or anticipating legal action want to perform an early case assessment or ECA to determine case strategy, potential costs and relevant individuals with whom to meet. Many modern eDiscovery/ECA technologies will not work to their full potential or even at all with some older archives. In preparation for possible future lawsuits, the legal department may want parts or an entire archive migrated so that the data can be better analyzed, secured and processed. Proactively migrating archived data will enable faster ECA and eDiscovery.

SUMMARY

Legacy data archives continue to age and are becoming a major performance bottleneck for many organizations, driving costs up and productivity down. These aging archives are proving to be inadequate in meeting new regulatory and legal requirements, raising the risk of regulatory actions, fines, lawsuits, and huge monetary judgments.

This situation is pushing many organizations to upgrade their infrastructure, including replacing legacy archives. A necessary step in the upgrade process is the migration of legacy data from those existing archives. Legacy archives can become bloated and unstable, adding additional complexities and risks to the migration process. Furthermore, the migration of older data can add additional risk in the form of legal hold and eDiscovery responsibilities that requires archived content and metadata to remain unaltered during the migration process.

Legacy data archives continue to age and are becoming a major performance bottleneck for many organizations, driving costs up and productivity down.

Three of the biggest mistakes organizations make when starting an archive migration project are:

- Not fully understanding the legacy archive technology (format, metadata, etc.).
- Not fully understanding their legal and regulatory requirements for handling archived data.
- Not taking into full consideration the impact of migration on end users, both during and post-migration.

Not paying attention to these issues will significantly raise costs and risk during an archive migration.

The surest way to address these issues is to choose the right migration vendor and/or an archiving provider with a successful track record for the migration project. Vendors with documented experience in migrating legacy archives with legal and regulatory requirements in mind will ensure trouble-free and defensible outcomes.

SPONSORS OF THIS WHITE PAPER

ARCHIVE360

Archive360 develops Next Generation software tools for email archive migrations and handles the most demanding archive migration projects in the world. Archive360 is an Independent Software Vendor (ISV) working directly with our customers and a global network of migration specialists and archive solution vendors.

Archive360 focuses on solving email archive migration challenges in the most cost effective and efficient manner for our customers. We developed **Archive 2-Anywhere**, which is a message extraction and migration solution that customers can run independently with full support for maintenance and software support from Archive360. Archive 2-Anywhere, our email extraction and migration solution, includes remote setup, configuration, activation, and support and software maintenance during the lease period. Customers can be up and running in less than 30 minutes and enjoy extraction rates greater than 5 TB per day.

Archive 2-Anywhere enables rapid migration and cleanup of legacy email records between different archive platforms. It securely transports your legacy archive data into the new target environment, ensuring it remains seamlessly accessible - both for end-users and for compliance and eDiscovery purposes.

Archive360 Archive 2-Anywhere:

- Exceeds all industry performance expectations
- Enables the migration (and cleanup) of the entire email archive including messages, attachments, and message stubs (shortcuts) present on Exchange servers
- Enables full chain of custody reporting

Archive360 is committed to fully protecting your email records for every business, regulatory, legal and end-user requirement. Whether you're migrating your data to a new on-premise archive or into a cloud solution such as Office 365, our solutions are engineered from the ground up to successfully cope with the highly complex and massive volumes our clients see in their legacy email archive migrations. There is no other vendor in the world that can match Archive 2-Anywhere when it comes to helping your organization successfully migrate legacy archive data to a new supported platform.

www.archive360.com
@Archive360
info@archive360.com
+1 630 358 4448

GWAVA

For more than a decade GWAVA remains unsurpassed at providing and protecting messaging infrastructures with superior unified archiving and critical messaging security.

Organizations that want to reduce costs, manage complexity and mitigate risk, on-premises or in the cloud, rely on GWAVA for cross-platform and cross-application solutions. And with offices worldwide, GWAVA backs up its commitment by delivering service and support for thousands of customers globally—24/7.

GWAVA delivers best-of-class security software--to protect your email systems—as well as to archive massive amounts of messaging data. GWAVA solutions also include leading-edge social media and mobile message archiving. Consequently, organizations that demand secure and sophisticated messaging-protection solutions—organizations such as Harvard University, Dow Chemical and the U.S. Department of Justice—rely on GWAVA for their archive and messaging needs.

www.gwava.com
@GWAVA
info@gwava.com
+1 866 464 9282

The GWAVA social media archiving solution, Retain Social, securely archives social media communications. This archived data can be instantly retrieved and reviewed to help ensure compliance, protect organization reputation, enforce company policies and reduce the liability of social media usage.

For more information about GWAVA and Retain Social, visit www.GWAVA.com or contact GWAVA at questions@GWAVA.com or 1-866-GO-GWAVA

OPENTEXT

OpenText Enterprise Information Management (EIM) technologies and business solutions allow organizations to take full advantage of enterprise information to gain better business insight, capitalize on opportunities to positively impact the business, improve process velocity, reduce risks related to information governance, and protect sensitive information and intellectual property from internal leaks and external threats. With growing volumes and a host of formats to manage and leverage, organizations need to bring structure to the unstructured. By doing so, they will be unleashing the power of information to drive faster decision making, improved agility, strong security policies, and an increased ability to both exploit the opportunities and control the risks of enterprise information. OpenText provides solutions across the entire range of core EIM capabilities – sophisticated, secure, high-value, and cost-effective – onsite, via mobile devices, private cloud, or in the cloud.
As the archiving backbone to many of these EIM capabilities, OpenText Enterprise Information Archiving (EIA) addresses a critical situation: skyrocketing volumes of data, global requirements for regulatory compliance, a growing need for litigation preparedness, and the reality of budget constraints.

OpenText EIA enables multi-faceted enterprise archiving in a single, scalable repository, which makes the solution easier to manage and less costly to operate. Tight integration with Records Management allows for optimized archiving capabilities like smart disposal of transitory items, auto-classification of records, and robust search for content. The result is a strong Information Governance framework that operates as a seamless extension of applications like email, ERP (Enterprise Resource Planning) and CRM (Customer Relationship Management).

OPENTEXT

www.opentext.com

@OpenText

info@opentext.com

+1 800 499 6544
(North America)

+800 4996 5440
(International)

QUADROTECH

QUADROtech provides scalable, trusted, automated and rapid solutions for the migration of email and user data between archives, Exchange, Office 365 and PST files.

Our products employ a proven methodology that automates processes and workflows to deliver archive migration projects on time and on budget. They minimize risk and ensure legal, regulatory and end users have continuous access to the information they require.

To date QUADROtech products have migrated in excess of 1 million users and approaching 2.5 PB of data.

Scalability throughout the migration process is achieved by enabling data transfer in the most efficient basis according to availability of resources and network bandwidth. Off-line data transfer is available without disruption and our "Synch & Switch" approach ensures data migration, regulatory and user access requirements can be met without compromise.

Our products and people are trusted to provide fully auditable, supported and compliant projects. In addition to being certified as both a Symantec STEP and Microsoft Gold Application Development Partner our engineering and support staff includes the largest number of Symantec Enterprise Vault Trusted Advisors from any

QUADRO tech

www.quadrotech-it.com

@QUADROtech_CH

info@quadrotech-it.com

+1 302 660 0166
(United States)

+44 1288 271 212
(United Kingdom)

+41 41 544 3800
(EU and Rest of World)

email archive migration vendor. We use published APIs where required to ensure complete supportability of migrated data.

Migration projects require automation to manage the workflows and processes for staged, partial and prioritized migrations according to business needs. Our solutions provide for data filtering to allow partial archive data migration and scheduling options to further meet operational requirements. Agility also means we provide the capability to execute storage driven migrations and remotely managed migrations.

For many migrations time is of the essence, especially when faced with looming deadlines such as platform end-of-life, mergers and acquisitions or rollout of new technology including mobile. Projects can be initiated immediately. This rapid approach eliminates costly delays. With data movement being on the critical path our platform is optimized for speed.

The QUADROtech solution suite includes ArchiveShuttle, MailboxShuttle and PST FlightDeck to address archive, Exchange, Office 365 and PST migration. As the only vendor to address the *total* mailbox our solution is unique in its ability to connect and coordinate migration across the three primary email content locations.

We provide direct export and import connectors for the major on-premise and cloud email archives platforms and maintain a global network of partners with the right level of expertise in email archive, Exchange and migration technology.

SMARSH

Smarsh delivers cloud-based archiving solutions for the information-driven enterprise. Its centralized platform provides a unified compliance and e-discovery workflow across the entire range of digital communications, including email, public and enterprise social media, Web, instant messaging and mobile messaging.

Founded in 2001, Smarsh helps more than 20,000 organizations meet regulatory compliance, e-discovery and record retention requirements. With its diverse and growing client base, the company has experience managing a variety of system migration scenarios. Smarsh is headquartered in Portland, Ore. with offices in New York City, Atlanta, Boston, Los Angeles and London.

www.smarsh.com
@smarshinc
sales@smarsh.com
+1 866 762 7741
+1 503 946 5980

TRANSVAULT

TransVault specializes in high volume, high speed, multi-vendor data migration solutions designed to preserve accessibility to an organization's business email records while maintaining chain-of-custody and integrity.

Customers rely on TransVault to insulate their valuable data from business change (such as mergers and acquisitions), as well as technological change (such as hosted archiving), media obsolescence and solution end-of-life.

Proven in over 1,000 migration projects, its flagship solution, TransVault Migrator, allows organizations to move their data selectively and defensibly in accordance with granular business and compliance needs, reducing overall data volumes and minimizing migration times.

The solution also pays close attention to 'the end-user experience', minimizing impact on users during the migration and ensuring their data remains fully and 'logically' accessible post-migration.

Supported platforms include Microsoft Exchange and Office 365, Symantec Enterprise Vault, EMC EmailXtender and EMC SourceOne, HP RISS and IAP, Autonomy ACA, ZANTAZ EAS, Mimosa and Message Manager, OpenText IXOS-eCONserver, Quest Archive Manager, Metalogix Archive Manager, iLumin Assentor, Mimecast, Proofpoint,

www.TransVault.com
twitter.com/TransVault
info@TransVault.com
+1 646 808 0407
(North America)
+44 33 33 40 44 33
(EMEA and Asia Pacific)

Google and many more. Additional archive platform support is being added all the time.

Its latest product, TransVault Sprint, features a powerful new UI that massively simplifies migrations between common paths, such as Symantec Enterprise Vault and Microsoft Office 365, resulting in a lower-cost, minimal-services approach to tackling migration.

TransVault Insight additionally offers advanced capability for managing and migrating the contents of ubiquitous personal archives (PST files), which includes the ability to selectively cull items in PST 'in place', or just migrate the items that are needed to a new platform, again, significantly reducing the migration task and overheads.

TransVault solutions are available through an international network of accredited partners, who uniquely for this application space provide services with engineers qualified to the TransVault Migration Specialist certification level.

REFERENCES

[i] *IDC Storage and Data Migration Services Overview*, November 2013
[ii] http://www.justice.gov.uk/courts/procedure-rules/civil/rules/part31#IDAALICC
[iii] http://www.clearwellsystems.com/e-discovery-blog/tag/practice-direction/
[iv] http://www.legaltechnology.com/the-orange-rag-blog/guest-article-the-ediscovery-passport/
[v] CIOInsight.com: *A Modern Governance Strategy for Data Disposal*