

# THE INS AND OUTS OF ARCHIVING

## WHY IT'S BETTER THAN BACKUP

Contoural's MARK DIAMOND and ProStor's STEVE GEORGIS talk about the benefits as well as the potential pitfalls of archiving, and how organizations can avoid these dangers by using emerging best practices.

### What is driving archiving?

**MD:** Many factors drive archiving. One of the biggest underlying trends is that organizations continue to accumulate significant amounts of semi-structured data, such as e-mail, and a lot of unstructured data in terms of files. Structured data is part of data bases and this information is accumulating. Companies have an excessive amount of data and an excessive number of relevant documents associated with that data – and this is creating problems for them. First of all, primary storage can be expensive, even if the medium is not expensive. By the time you replicate and back it up, it costs a great deal of money. In addition, one of the factors driving the need for archiving is that organizations are having difficulty understanding what they have and whether they can find it. In our opinion, one of the greatest benefits of archiving is control. Archiving allows you to know what you have, where you have it, if it can be moved somewhere less expensive and if it can be recovered at will. The idea of archiving has really been around for a number of years, but only within the last three or four years have many organizations made archiving a priority – not only within IT but also within records management and the legal department.

**SG:** Once companies really start to think about it, they realize that archiving has many drivers and applications. It starts with what Mark just said, which is that companies are saving everything today on disk, largely because they do not know what they should keep and what they can discard. Because of that, the storage costs keep multiplying. In most enterprises today, storage is growing by about fifty percent per year. But the total cost of storage does not involve only the cost of acquiring the equipment, it also involves the cost of continuing to manage and operate that equipment. There are two general themes that drive archiving today. The first is reducing costs. As the data keeps growing, it is not sustainable for companies to just keep adding disks and piling data onto those disk systems. Nor is it sustainable to keep having to back up those disk systems. Therefore, archiving reduces cost and permits what I call cost optimization; in other words, the data is in the appropriate tier of storage and the company only pays what is really necessary for storage.

The second thing driving archiving is compliance. In terms of industry-specific regulations or generic legal requirements for e-discovery, companies are very concerned about compliance. This is likely what is driving companies to worry about saving everything. But most businesses are wasting money today by saving everything

in primary storage. They spend an excessive amount on it because they do not know how to manage it any other way.

### How is archiving different from backup?

**SG:** This is the biggest area of confusion among users today. In the past, most people considered backup and archiving to really mean the same thing, because they both revolved around the same technology: tape and tape libraries. Everyone had tape systems for backup and they also used them for long-term data retention. They were generally using the same software applications which were designed to push data out to tape. Most people think that archiving involves taking your backup tapes, sending them to a vault and saving them for a long time just in case they are needed at some point. But when you actually look at these two applications, backup and archive are really quite different. Backup provides protection against possible disasters such as system crashes, natural disasters or malicious intent by an employee or outsider. Backup provides copies of your data sets every day or every hour to a secondary storage system that is not vulnerable to the same threats as the primary system in order to protect against the loss of data. Archiving, on the other hand, ensures the continuity of the business and the data that the business depends upon by moving data one time from expensive primary storage systems to an archival storage system – which is generally much less expensive.

An archival storage system is something that is designed to preserve data for long periods of time for very low costs and then intelligently manage that data so that it can be found and retrieved. The economics of archive is driving the desire to move data to a lower cost form of storage. In addition, the mandates of compliance regulations add a whole new level of complexity to what it means to archive. Suddenly, there are rules about data retention, deletion, the ability to discover and retrieve that data and ensure the security and authenticity of it. There is much more intelligence required by the archiving process than ever before. What that means for a business is that the old method of just dumping to a dumb tape drive and sending it to the vault for a long period of time is not satisfactory anymore. What we need today is intelligent data management combined with low-cost, long-term storage.

**MD:** Archiving ensures data protection. Using backup tapes as an archiving strategy is a very bad idea because backup tapes allow you to save the data, but make it extremely difficult to retrieve that information. Furthermore, some of the new requirements under the Federal Civil Procedure and Litigation are very stringent, and would require an excessive amount of time spent, for example, on reading backup tapes. Companies that depend on backup tapes as an archiving strategy are: a) finding themselves at risk of not being able to produce information in a short enough time period of time; or b) spending an excessive amount of money to be able to find the information. With our clients, we like to separate the storage protection strategy from archiving, because they are really different things.

**SG:** When we look at the technologies available today, there are really best-in-class technologies available for both the backup and the archive applications. Best practices dictate that customers look at the best technologies available for each, ultimately helping them optimize their costs.

### Do small and medium-sized enterprises (SMEs) face the same or different requirements for saving data as large enterprises?

**MD:** Many times SMEs believe that as private companies they are not subject to the regulatory requirements, but this is untrue. We have many clients who are both SMEs as well as clients who are large sized companies, and they all tend to accumulate data. Unfortunately, SMEs enter into lawsuits just as frequently as big companies, so they must have easy access to their information. SMEs face many of the same compliance requirements. They may not have Sarbanes-Oxley, but there are still thousands of regulations that require organizations to be able to preserve information. Sometimes it is even worse for SMEs because they are not as well prepared when a lawsuit strikes or when faced by regulators. Therefore, SMEs need to pay as much or more attention to the same drivers behind archiving as big companies; and increasingly, SMEs are indeed archiving as part of their business strategies.

**SG:** When ProStor first entered the market, we found that our customers had almost the same requirements as large enterprises but that they were daunted by the complexity and cost of implementing archiving. In particular, they were looking at some of the solutions that were available in the marketplace. They realized that they were far too expensive to purchase, maintain and manage, because smaller organizations tend not to have all of the IT specialists needed to manage some of those more complex systems. There was a very broad segment of the marketplace that had the exact same needs; they just needed something that was going to be less expensive and less complicated. After all, when we look at things like compliance regulations, they do not discriminate based on the size of the company. A small healthcare organization is subject to the same HIPAA regulations as a large organization. A small financial services firm is subject to the same SEC regulations as a large firm. When we started talking to customers, we found them pretending that the problem did not exist because they did not know how to solve it. Once they saw that there was an affordable and manageable solution, they became very enthusiastic about the prospect of being able to become compliant.

### What are those requirements?

**SG:** As I said, for SMEs, it has to be low cost. Companies typically do not have large budgets. Companies want both low entry costs – what it initially costs to purchase the solution – and low cost in terms of scaling and managing. Secondly, it must be easy to deploy and manage and thirdly, it must be flexible. Archiving serves many purposes. Enterprises typically find that they have multiple applications and that those applications may have conflicting requirements, so flexibility is key. Finally, it must be a complete



solution. Smaller enterprises need an all-in-one solution that provides data retention, data retrieval, as well as data protection and all the intelligence that they need for compliance.

**MD:** We actually see many drivers behind archiving decisions that organizations must evaluate, and there is some confusion concerning them. One of the drivers is compliance. There are new existing regulations that require companies to save information and actually, many organizations start archiving projects from a compliance perspective. Compliance is not always the primary driver. Some of the regulations, like SEC 1784 for financial services, broker dealers or HIPAA are drivers, but many of the regulations tend to be non-prescriptive. Driving your archiving strategy or policy principally from a regulatory point of view is a bit risky.

Another driver is data privacy. Organizations must be aware of what data they have so that the archiving strategy does not set them up for a privacy breach.

The biggest driver of all is litigation readiness. Increasingly, electronic documents and electronically stored information are becoming the focus for e-discovery. Almost every lawsuit has some e-discovery component to it. E-mails and files are a key focus in lawsuits. There are new requirements under the amended Federal Civil Procedure requiring that organizations are able to produce certain data. Often, only when companies face lawsuits and spend a great deal of money trying to produce this information do they realize the importance of archiving. Unfortunately, many times companies only want to put up the crossing guard after they have been hit by the freight train.

And finally, as Steve said, storage is a motivating factor. Companies spend a great deal of money on primary storage, thereby consuming IT budgets. They should spend money for more strategic initiatives instead of consuming it on storage. I like to offer the analogy of a balance scale. If you load three of these drivers on a balance scale, it is enough to tilt the scale down and have organizations make a decision. Also, it varies from each individual company in terms of how they look at it. There are various advantages for archiving among all the different stakeholders.

**SG:** There are many hidden benefits of archiving, as we see with our customers who have deployed the ProStor Systems InfiniVault appliances. One very good example is that a financial services customer of ours who operates a call center was required by SEC regulations to record phone calls and keep them for seven years. What the client realized is that this was actually an archive requirement. Archiving phone calls is not something people normally think about because there is no business value to keeping the phone calls, but it turned out that it related to a compliance concern. We see many similar discoveries. Among the drivers that Mark mentioned, there are many subcategories for which people realize that they need to have a solution.

### Who should be involved in these types of projects?

**MD:** Many organizations should be involved. Sometimes we see archiving projects driven just by IT. There is a lot of good cost-benefit

analysis and you can often justify these from an IT basis. But we also see a lot of archiving projects get stuck. When IT is ready to implement the project and asks the company's legal department how long they can save data, the legal department develops a large committee of people to look at the archiving policy. The trickiest part of these archiving projects is to build a consensus. We help many organizations develop enterprise-wide document retention policies and strategies, and also help execute them; it really helps them put together some strategies for building a consensus. The legal department, records management, compliance, security, audit and IT should work together to build a consensus in terms policy. Then, they should look at the technology in order to execute the project.

There is some great technology that can help execute the project, but the policy part of it is equally important. We work with many CIOs and general counsels of both large and small organizations to help them figure this out. We have found that if a good consensus is built early in the process, other groups will support the archiving initiatives as well. In some cases, other departments even help with funding. Sometimes IT people want to do archiving just to save on storage costs, which is a good reason. Yet, one should remember that other people can benefit from this as well. There are benefits to all the business units, including records management and compliance. If you sell it on those terms, the organization enterprise will likely fund and support the archiving initiative.

**SG:** Because we at ProStor tend to target SMEs, often the project starts with the IT department. More often than not, it is initiated for storage cost reasons. Then, IT goes to the legal department or a compliance officer to find out what those compliance requirements are. The applications expand within the organization as new archiving applications are added. IT departments, compliance officers, the legal department and document management can all become involved. One of the keys is having a solution that accommodates a variety of requirements. Because so many departments get involved, having a solution that provides the flexibility to meet the requirements of different applications is a key factor.

### What are some of the pitfalls for archiving projects?

**SG:** I started talking earlier about the difference between backup and archive, and the legacy role that tape played in those processes. What we see today is that many organizations, perhaps even most organizations, are trying to move away from tape technology and toward disk-based data protection and archive strategies. This eliminates the pain points of tapes, but it adds new complications that can become significant pitfalls. First, disk archive systems are always spinning disks, they consume power, they must be cooled and they require management and operating expense dollars forever. On the other hand, tape and removable media can be taken offline and the only storing cost is the cost of the media itself. Today's disk archive systems tend to have perpetual costs associated with them.

A second pitfall is that today's systems generally require a secondary system in order to back up the archive. Both of those problems can be solved by deploying a system that includes

removable media as part of the overall solution. This provides long-term data retention involving only the cost of the storage media, because it does not consume any power. Optimally, this system would be integrated into a single, seamless system that is easy to deploy and manage.

Another pitfall is not adequately planning for capacity growth. As we discussed earlier, data is growing quickly. We continue to have to save everything and many organizations underestimate how large their archives will become. Some of the third party industry reports on the growth of archival data show that archival data is growing at about sixty percent per year, which is a lot. Therefore, the archiving solution must be very scalable and the economics of that scalability should make sense.

**MD:** We do find a number of pitfalls, but all of these pitfalls can be overcome. One is that there is a tendency to focus exclusively on the technical requirements. This is important, but by not having a consensus or really defining what the policy should be, many archiving projects get stuck.

Another issue is whether a consensus can be built across all the business drivers. At the end of the day, if you really look at issues like e-discovery and understand them, a good consensus can be built.

The third pitfall is that some organizations try to build manually dependent archiving strategies. They want people to pull documents manually, but the process should be automated as much as possible.

Next, archiving is largely about control – about knowing what data you have, where you have it and having it at the right place and the right time. Depending exclusively on humans to do this may result in error or lack of completion. Therefore, good tools and products should be used to automate the archiving as much as possible.

Finally, there is a tendency to not want to engage the business units. We always like to work with the business units to understand where their data is, what we can move and really explain what we aim to do. The pitfall is that people do not want to move things away from their desks. However, we explain to

companies how it can benefit the business units and that they will be able to save more data on a longer term basis.

These are fairly common pitfalls that organizations experience. On the other hand, they are also easily overcome. Archiving projects are hardest at the very beginning and they get easier. We urge our clients to keep at it, because there are many customers, both large and small, who have implemented archiving and benefited tremendously from it. They have more robust IT, litigation, compliance and records management strategies because of it.

**What are some of the emerging best practices?**

**MD:** One key best practice is to make it as automated as possible. Look at the cost of your disk storage and determine whether things can be moved to something a bit less expensive. Another of the best practices is to be able to control and know what you have and where things are. If something is in the archive, we know that it can be retrieved quickly and easily. Finally, the clearest best practice is to eliminate backup tapes as an archiving strategy. Using backup tapes for archiving is an extremely ineffective strategy. On a long term basis, it is okay to use backup tapes for data protection, but they should not be used as an archiving strategy.

**SG:** Automation is part of our mantra here at ProStor Systems. When the policies are automated, cost and human error is reduced. If the right tools and systems are combined, the system can be fully automated.

On the cost side of archiving, best practices involve combining data migration tools with an intelligent archive system. This allows you to automatically identify and classify the data that should be moved; then that data can be managed on the archival storage system. This reduces the management complexity as well as the cost.

On the compliance side, because we have different intelligence requirements for archival data, we can automate the retention, deletion, data classification, security and authenticity policies. This will ensure that the policies are uniformly administered, regardless of personnel.



STEVE GEORGIS, President and CEO of ProStor Systems, has 29 years of experience in data storage and telecom. He was a founding member of Exabyte and the Founder and CEO of Network Photonics. He has 24 years of data storage and data protection industry experience in general management, product development, marketing and business development at ProStor Systems, Exabyte, StorageTek and Data General. He holds 17 U.S. patents awarded in data storage and optical networking technologies.



MARK DIAMOND, President and CEO of Contoural Inc., is one of the industry thought leaders in litigation readiness, e-discovery, compliance, archiving, data protection and ILM strategies and practices. A frequent industry speaker, Mark addresses how organizations can better align legal and business requirements with IT and storage spending. He is also an expert in the business drivers around archival and the technical strategies for implementing them.



# Weathering the Storm of Regulatory Compliance?

Are the storms of information security, records management, e-discovery and lack of data privacy beating you down?

If so, do what many other customers have done – protect yourself and your business with the ProStor Systems InfiniVault™ archive appliance family.

InfiniVault addresses compliance issues by integrating easy-to-use compliance management with the lowest cost, long-term storage of any intelligent disk archive system. InfiniVault provides automated retention management, data immutability and security, along with online file access and offline data protection and archive in an all-in-one solution.

Find out how InfiniVault\* can secure your vulnerable business data and help you weather the storm of an audit, investigation or litigation. You'll be glad you did.

## InfiniVault™



Learn more: [www.prostorsystems.com/GRC](http://www.prostorsystems.com/GRC)

\*Bronze winner in the 2007 Products of the Year competition (Backup and Disaster Recovery Category) sponsored by SearchStorage.com.