

Email Archiving for Compliance and Discovery

Legacy Email Ingestion - Ensuring Maximum ROI on Your Investment

Michael Daniec, President, National Data Conversion

Table Of Contents

Executive Summary	3
Compliance Strategies	4
Obstacles	9
What it Takes	12
Why it is Not Easy.....	13
In-House or Outsource?	15
How to Select a Vendor Partner	18
What NDC Brings to the Table	19
About the Author	20

Executive Summary

- **Changing risk profile**
- **Emails at center stage**
- **Maximum ROI**
- **Its all about the data**

Litigation and regulation - these two challenges are rapidly changing the face of corporate governance, risk management and compliance (GRC), across a wide swath of industries, market sectors and scale of enterprises. One aspect of this trend is that it is revolutionizing the way enterprises organize, manage and discover information captured in emails.

The risk profile of doing business in America, and worldwide, has changed irrevocably. No organization is immune from either the probing eyes of regulators or the unexpected strike of litigation. Overnight, normal operations can become the object of intense scrutiny and the most trivial gaps in operational controls can cause monumental headaches and risks. Compounding the challenge is the phenomenal growth of electronic data. Its ubiquity and pervasiveness have established ESI (Electronically Stored Information) as the new baseline for records keeping. Effective electronic records management and discovery response have been quickly thrust into the spotlight and imbued with unprecedented urgency.

To defend itself against this onslaught of regulatory and legal pressures, corporate America has awakened to the importance of effective management of its electronic information assets. Key components of this renewed focus are the development of defensible, rules driven records management repositories and effective electronic discovery capabilities. Email archiving has quickly percolated to the top of the priorities in the needs and risks pyramid. Given the paramount significance of emails in litigation and in regulatory mandates and investigations, it is no wonder that email archiving is a top priority for most organizations, even in unregulated, or should we say, less regulated, industries (it has been estimated that an enterprise may be subject to any of over 18,000 regulations).

Even as email archiving has become the de-factor standard across all industries, best practices and solution sets are still evolving. Despite a variety of views regarding specific policies, extent and urgency, the momentum seems clearly to be shifting towards a recognition that these solutions need to be comprehensive in terms of both the universe of records to be managed, and the needs they must address. Whereas, some time ago, good intentions and a partial response could be contemplated the courts and regulatory bodies have become less tolerant of anything but the real thing - a full blown solution. This is particularly clear with, of course, the attention grabbing headline news of court rulings and regulatory fines.

However, implementing a technology solution is only the first step in the odyssey of effective electronics records management. Policies, procedures, training and operational controls all play vital roles. Above all, organizations are discovering that, in order to derive the expected benefits of their solutions, and full ROI on their investment, they need to data enable them with legacy data.

The proposition is straightforward on the surface of it. In order for the solution to play the expected role, it needs the appropriate data to act upon. An empty repository is like a car without gas, it may look nice sitting in the garage but is of little use. Data enabling the solutions, then, becomes a key component of effective response. To provision a repository with the full data set that will make it useful in any practical context, requires the loading up of multiple years worth of data.

NDC enables enterprises to better respond to regulatory compliance issues, lower e-discovery costs, and reduce corporate risks by provisioning their email archives and other records repositories with legacy content from backup tapes and optical disks in a timely and cost effective manner.

Compliance Strategies

- **Regulation Here to Stay**
- **Email Archiving to the Rescue**
- **Where is the Data?**
- **Raising the Bar**
- **Risk Factors**
- **New Solutions**
- **Summary**

It is no surprise, in our information driven economy, that management of electronic data and emails in particular, has come to the forefront of the corporate governance, risk management and compliance, GRC, debate. In conjunction with this, increasingly effective solutions to manage the full gamut of enterprise-wide electronic data content and needs are rapidly appearing on the scene. Already, the crude, first generation solutions are being replaced by more sophisticated tools and methodologies. Slowly, organizations are making the commitment to more ambitious and wider ranging initiatives.

Not that this progress is not marked by varying points of view regarding best strategies and methodologies. However, as continuing debate rages over tactics of implementation there is a fundamental recognition that the burdens of responsibility for managing electronic data is not a temporary phenomenon and requires long term strategic consideration and investment.

Part of this awakening can be seen in the development of new senior level corporate positions related to compliance and governance, the vesting of increased authority in previously sidelined records management departments and opening up of invigorated channels of reporting to the top.

Another aspect of this can be seen in the increased recognition that, after all, its all about the data and that no solution in this arena can be effective until properly provisioned with data reflecting a reasonable time continuum of the past. In the email archiving arena this means ingesting multiple years of legacy emails into the repository. The barriers to such provisioning are quickly falling as well: cost, time and organizational disruptions.

Email Archiving to the Rescue

It is indisputable that email archiving has become the fastest growing of all the Records Management and enterprise-wide ECM/ILM (Electronic Content Management/Information Life Cycle Management) components. Since emails have become a standard part of doing business and some estimates indicate that nearly half of all business critical information may be stored in corporate messaging systems, email archiving has quickly assumed a preeminent role in governance, risk management and regulatory compliance.

In addition, of all the available venues for implementing a records management discipline within the electronic records universe, email archiving is the best understood, the most mature and offers the quickest and most certain return on investment. Most companies have had a greater level of comfort with this component of the RIM (Records Information Management) world. This is reflected in a much greater level of adoption. Email archiving has, in effect, been accepted as a best practices response to various operational, legal and compliance challenges and has become the typical leading point for organizations embarking on the records management odyssey.

Traditional document management solution sets, including ones that preceded the email phenomenon, have not evolved into the forefront of the records management adoption curve as email archiving has. To a large extent, EA has bypassed and outpaced the general document management disciplines to quickly become the most commonly implemented component of such approaches.

However, according to a Forrester Research study, the ECM market is beginning to absorb some of the email archiving market and by 2008 over 40% of the message archiving market may be owned by ECM vendors that offer more comprehensive Records Management solutions of which email archiving will be only a component. Whatever the longer term trends, email archiving will clearly be the primary focus of records management for some time.

The three legs of the email repository rationale are cost reduction and risk mitigation in the compliance and litigation arenas and operational efficiency. Firms implementing email archives have recognized these value drivers behind their investments. Regulatory requirements, of course, have driven this adoption in regulated industries, but email archiving has now become a de-facto standard in most industries.

Where is the Data?

At the same time, as organizations move rapidly to implement and extend email archiving, they are already looking over their shoulder and realizing that the next logical step is to data enable these solutions with legacy data. Many of these organizations are discovering that backfilling is, in fact, a critical component of the ROI proposition associated with email archives. Only when they fill the archive with legacy data can they expect to derive the full anticipated return on their investment.

For one thing, having a single point of reference for all instances of this type of data streamlines any possible uses of the data for litigation response, compliance or even standard operational purposes. One of the more powerful arguments for implementing such repositories, risk mitigation, can only be fully realized when the repository is loaded up with appropriate universe of historical data.

However, until recently, even progressive organizations were reluctant to embark on a backfilling process to data enable their repositories. Initially, there was no clear definition of whether this was really necessary or how much data needed to be loaded. In many cases, the implementation of an email archive was the minimum necessary response to meet an immediate compliance mandate and there was little management commitment to venture beyond that.

Many stakeholders even debated whether data enabling an email archive lowered or raised the risk profile of an organization. According to one camp, having the emails readily available only increased the likelihood of their being involved in discovery. Others insisted, that this is only an illusion, it will be demanded anyway, particularly as the courts, regulatory agencies and litigants get increasingly sophisticated in their understanding of the issues. In fact, developments such as the new FRCP rules (Federal Rules of Civil Procedure) clearly bring to the forefront issues of electronic discovery and make it less likely that any element of electronic records can be swept under the table.

In the meantime, a critical weight of regulatory pressures, including record fines, high profile legal cases and the rising costs of discovery in litigation and compliance have reached a threshold where the consensus has shifted to acknowledge that loading up email archives with legacy emails is inevitable and even makes clear business sense. The question has thus become when and how to undertake this and how far back in time to go.

Going All the Way

In resisting more comprehensive email archiving initiatives, some organizations have evolved closely reasoned arguments for partial solutions involving various policy driven regimens such as mailbox quotas, time based deletion policies (such as saving only 90 days worth of emails), etc. However, in practice these partial regimes have proven very problematic in implementation and enforcement and questionable in their organizational impact.

For instance, email box size quotas can seriously impact employee productivity and can have other more serious implications such as stimulating employees to delete important messages to make room, create personal archives to offload emails from their mailbox, etc. This can create a major headache for IT and expand, not diminish organizational risk.

As a consequence, the cards seem to be quickly falling on the side of comprehensive, fully automated email archives, loaded up with full legacy content. Since emails are now widely regarded as an essential component of corporate records keeping and recognized as such by the legal and regulatory communities, partial archiving may offer dubious legal protections and the validity of incomplete archives is increasingly under attack. From a purely practical perspective, partial archiving, with its associated challenges may actually be more difficult to implement and maintain successfully than full archiving, as it may require additional operational discipline and burdens.

Many solutions vendors and, increasingly, industry experts, suggest that effective partial archiving is a myth. Emails, in particular, have a peculiar ability to survive any records deletion/destruction regimes, especially the most inconvenient emails. I am reminded of a recent records management and eDiscovery conference where a panelist, who suggested his organization had implemented a successful, and very disciplined, 90 day email retention policy, was asked if in-fact they didn't have backups of the emails going past that time frame. After some prodding he reluctantly admitted, that it was likely they did, have such backups for "disaster recovery" or similar purposes. Aside from any other consideration, this, of course, would be an embarrassing admission in any legal context and would completely undermine the justification, or supposed protection, of the 90 day retention policy.

Much can be said, therefore, regarding the advisability from a legal as well as a practical perspective, of biting the bullet and loading up an email archive with a full set of legacy emails, from backup tapes or optical media. Major international organizations have come to appreciate this and have undertaken increasingly wider ranging and backward reaching legacy email restorations and ingestions into their archiving solutions. The point of gravity has shifted so far that many of the top tier international financial powerhouses have determined that they need to capture all of their email traffic indefinitely going forward and as far back as their backups allow, at least until the air clears and clearer guidelines are developed by the courts and regulatory agencies.

And still, many organizations procrastinate in actually embarking on this path and all too often, block themselves into a costly fire drill based on a pending overdue litigation or compliance investigation. Time after time, we have seen top notch corporations sit on a decision, unable to commit the funding and overcome their organizational inertia, until an enabling event such as litigation or investigation breaks through the ice. Then, under incredible time pressures, facing huge penalties, they scramble to achieve, at whatever price the market will bear, what they could have performed at a much lower price point and minimal disruption just a short time previously.

Raising the Bar

In the compliance arena, regulatory agencies pursue ever expansive definitions of what types of information need to be retained and made readily accessible, from traditional customer statements to emails and now instant messaging traffic and such. It would be reasonable to expect that, as technologies evolve to support new forms of communications, these too, will fall under both regulatory and litigation scrutiny.

Not only has the bar been raised in terms of types of communications subject to regulation and their accessibility but also the risks associated with non-compliance have grown dramatically through much steeper and more personalized penalties.

Risk factors

Compliance and litigation response are not easy investment propositions. Some industry pundits, recognizing traditional management reluctance to invest in this arena since it is perceived as strictly a cost center, suggest using a TCF (Total Cost of Failure) calculation as basis for decision making.

Risk factors to consider:

- costs of regulatory and legal discovery
- consequences of inadequate response
- operational disruptions and burdens associated with discovery
- inability to perform effective early case/risk assessment
- impact on shareholder value

New Solutions

In the meantime, the world of compliance solutions is moving forward rapidly, in response to market pressures. Already, the first generation tools are being replaced with newer technologies enabled with more advanced features. Powerful review and electronic discovery tool sets are becoming standard components. Wonderful solutions have emerged in both arenas.

For instance, traditional keyword search tools that simply index and provide search capabilities are being replaced with new technologies that offer a deeper level of intelligence. By analyzing the unique properties of email (such as recipients, replies, forwards, subject header information, etc.), and combining this information with organizational data, these solutions dynamically derive email-specific

communications patterns. The products can apply various linguistic and statistical analysis algorithms, to extract intelligence from the streams of emails and attachments.

Likewise, various email policy enforcement and active monitoring toolsets have appeared. Even industry behemoths are paying close attentions, as evidenced by Microsoft's inclusion of a modicum of fundamental compliance related features in the upcoming Exchange 2007.

These market-leading capabilities can provide email analysis capabilities including a single view into enterprise wide email repositories via a unified interface, mapping of discussion threads by dynamically linking together all related messages and providing a chronological thread of the entire discussion. By walking the thread you can quickly identify all the participants, who knew what and when.

The one constant that remains is that all of these solutions require data to operate on.

Summary

Corporate America faces unprecedented challenges in attempting to lower the rising costs of litigation and compliance. The explosion of electronic documents in business and the corresponding and increasing mass of electronic discovery associated with litigation have placed the spotlight on finding the most effective means of reducing costs associated with the discovery phase of litigation and compliance. Implementing an email archive and enabling it with legacy emails is quickly becoming a universally recognized best practice response.

Obstacles

- **The Cost Thing**
- **Why the Pain**
- **Do You Feel Lucky**
- **Which Side of the Street**
- **Traditional Views**
- **Its Not Easy**

Legacy data ingestion still faces many hurdles to acceptance. Cost, of course, is the most common denominator in any of the associated decision making. For one thing, the costs of ingesting a number of years of data into an email repository can often be far higher than the initial cost of the archive itself. This conflicts with the traditional technology view that the hardware and software should represent the highest cost components of any enterprise solution investment. It took IT management some time to recognize the value of software (that's from the days when the computer equipment itself cost millions of dollars and hardware manufacturers were giving away or bundling the software to drive hardware sales) and they are only now, reluctantly, opening up to the full value proposition of associated services.

Welcome to the information age. In case you haven't noticed, its all about the data. It follows, in the new information age paradigm, that custody and management of that data increases in importance and value. In fact, increasingly, the real value of the solution lies in the data itself. Cost should follow value. The traditional view simply does not hold water anymore.

We have seen a shift in the perception of value to where it is acceptable that software may be higher cost (value) than the hardware. With the current trends, services, such as data enabling a solution, are quickly gaining a similar value recognition.

However, there is still much resistance, within the traditional mind set, to accepting the value proposition of any services, proportional to the value of the data itself. Even when the value of services is recognized this value is typically associated with services such as consulting or software development. Somehow, data is just supposed to be there and not need much servicing and therefore such common, foundational services as data cleaning, administration, or data enabling are often overlooked as value propositions.

Interestingly, value has a peculiar elasticity. What is perceived as an unacceptable discretionary cost, becomes an easily accommodated cost of business under a legal or regulatory threat. Not only that, a cost, that, with some foresight, could have been substantially controlled when undertaken without time and legal constraints, can balloon to multiples of its original size when the same undertaking is performed in a rush, with severe legal repercussions threatening.

Why the Pain

On a practical level, the process of restoring from backup tapes and ingesting such huge amount of data represents a major technical and operation challenge. Let's review some of the hurdles:

- costs
- huge volumes
- technical challenges
- required resources
- time
- operational disruptions
- lack of well understood best practices
- uncertain future

- moving targets

So how do organizations respond to the above challenges?

Do You Feel Lucky, Today

A not infrequent response to the challenge posed by legacy data loading is for organizations to throw the dice so to speak and count on not needing the legacy data in the time span that they may be legally required to store this data. Subsequently, they hope, they can destroy it, following legally accepted guidelines, and avoid the whole issue of dealing with it.

There are a number of problems with this approach. First of all, it's a pure gamble. In addition, electronic data, unlike paper based documents, has a persistence and ubiquity that is astonishing and can be very surprising, in a bad way, in a legal setting. This means that even when you think you have gotten rid of it, it pops up from the most surprising sources, such as forgotten backups, systems taken off line but still warehoused, and, of course, the multiple parties that were recipients of the data in the first place.

Which Side of the Street is More Risky?

Another, not infrequent position is that having legacy data easily available in a repository increases legal risk, as much as it reduces it, by putting it into play, where otherwise it may not be, if it were less accessible. In other words, having it available is an invitation for it to be discovered in litigation or regulatory investigations.

This is also a rather risky proposition since courts are increasingly less tolerant of the inaccessible argument. The regulators and the judiciary are quickly getting smarter about the technical issues associated with legacy data and its accessibility. A whole community of vendors serving the litigation and compliance markets has made incredible strides in ensuring accessibility and reducing the cost, complexity and time associated with such access.

In fact, the new FRCP (Federal Rules of Civil Procedure), expected to go into effect December 1, 2006 are anticipated to bring issues of accessibility into the forefront of any litigation and allow much less room for maneuver regarding this issue, among others. One expected impact of this is that it may become more difficult to employ the "don't ask don't tell" strategy and hide behind general ignorance.

In addition, the new rules change the risk profile not just for the litigants themselves but even for legal practitioners, since they will be clearly held accountable for full disclosure of relevant information, whether or not the other side specifically requests it. They are also clearly tasked with ensuring their clients' compliance with discovery requirements. It will be interesting how the legal community responds to this new burden and how this changes the legal game when it come to electronic discovery issues.

This could also have an interesting impact on the issue of legacy data loading into email repositories as it may create a more general appreciation that legacy data is discoverable anyway and, therefore, one might as well strive for maximum longer term cost efficiency by making it more accessible through an archive.

Traditional Views

In addition, there are some deeply embedded misconceptions, about tapes, among the various stakeholders in the process, such as legal and management:

- they are difficult to work with
- they are very expensive to process
- they take a very long time to process
- they do not last
- not everything on tapes is recoverable
- they are too much hassle

There are probably some simple explanations why tapes have developed such a bad reputation, mainly, that many of the above perceptions were true some time ago and were formed from bitter experience. However, the world of tapes has advanced, just like all aspects of technology, and these barriers are no longer the insurmountable obstacles they once were.

Its Not Easy

Still, one must clearly understand that tape restoration in the compliance context differs from normal, operationally dictated restoration. Here are some of the distinguishing factors:

- additional technical challenges such as older backup media and formats represent, one needs to be able to deal with the full range of issues that may arise, from media problems to lack of antiquated software tools
- special resources are often required, including specialized tools to overcome software and hardware limitations such as are often associated with this type of work
- large volumes and, all too often, tight deadlines pose a unique urgency, with potential catastrophic consequences to failure, and can be highly disruptive of business operations
- higher standard of integrity, the processes utilized need to be able to withstand external scrutiny by various parties such as regulatory bodies and the courts
- defensibility issues, the courts and the other party are much more likely to scrutinize and challenge in-house execution vs work performed by experienced outside experts specializing in this type of work

What it Takes

- **Strategy**
- **Bandwidth**
- **De-duplication**
- **Cataloging**
- **Culling**
- **De-duplication**
- **Normalization**

Successful execution of legacy data restoration and ingestion projects requires a thorough understanding of all the issues involved and a reliable strategy to address them. This strategy must encompass the full cycle of processing phases, from collection through tape restoration, email processing and ingestion into the repository.

It must encompass the following issues:

- realistic time frames and cost estimates
- technical challenges
- resources required
- impact on operations
- risks, including operational, legal and regulatory
- assurance of transparent, consistent and auditable processes

The key element of a successful strategy is that it must be highly reflective of the characteristics of the source technology environment, including its history. A thorough understanding of the technology environment requires knowing:

- both the equipment and software architectures that were employed
- the key personnel, particularly key players responsible for system administration and backups who may
- varieties of emails that are available and needed for the archive
- the procedures and policies employed in creating backups
- physical locations of the various key technology components such as email servers and their backups

All of above details can impact the specifics of how to optimize a project and assure timely and reliable completion. For instance, effective de-duplication of emails prior to ingestion into a repository can dramatically reduce the scope and time required for the overall project. This, however, requires specialized expertise and tools that only experienced vendors may possess. It is critical to avoid the many tripping points that can endanger the integrity of the process. Even seemingly remote details such an upgrade of an Exchange email server versions at some time in the past can have a surprising (for the inexperienced) impact on the required de-duplication strategy. These are the kinds of issues one does not want to discovery in the midst of an ingestion project.

The typical processing steps associated with tapes are:

- cataloging (sometimes a critical step to ensure effective processing)
- sampling (often a good way to test assumptions and limit scope)
- tape restoration
- culling of contents for relevant sets
- de-duplication of content
- normalization into a target repository format
- ingestion of the content into the repository
- production of audit trail reports and verification

A specific project may involve one or more of the above steps. Sampling is often overlooked but can be a powerful weapon in reducing scope of work.

Why Its Not Easy

- **Outdated technologies**
- **Many formats**
- **Missing skill sets**
- **High volumes**
- **Time constraints**
- **Lack of information**
- **Lack of tools**
- **Old media**
- **Undefined best practices**
- **Legal & Regulatory Considerations**

Those who have had any experience in this arena, are fully appraised of the true scope and the incredible variety of issues and problems that can arise in legacy provisioning of a repository. Here is a brief rundown of the issues.

For starters, backup tapes often go back in time to technologies and skill sets that are no longer available. On top of this, given how dynamic and quickly changing technology is, the number of variables at play can be staggering. Consider, for example, the variety of tape technologies available on the market over the years, the many backup software solutions, different operating systems and email server environments.

Above all, high volumes change the game. Legacy tape restoration now often encompasses tens of thousand of tapes with many of terabytes or petabytes of data to be processed, including billions of emails. To be able to effectively scale to meet the requirements of such volumes, represents huge technical and operational challenges, that can be further compounded by frustratingly tight deadlines, such as often crop up if the effort is driven by a pending litigation or compliance need.

Furthermore, most organizations have little information regarding what is on their backup tapes, created some years ago. This is understandable since these tapes may go back in time far enough so that organizational knowledge about them could be seriously impacted by staff changes, technology upgrades and misplaced records, particularly, given the lax attention paid to tape backups historically.

Although this seems like an insurmountable obstacle, there is a simple response available, such as performing, quick, low cost catalogs of tape contents. It is surprising how often this option is overlooked and yet what significant impact it can have on reducing overall project costs and timeliness. Basic cataloging can normally be performed pretty much at maximum raw tape speed, with some backup types it may require even less time than that.

Tapes are not perfect. In fact, all magnetic and optical media, tapes, disk drives and optical platters, have manufacturer specified useful life spans. Normally this is expressed in terms of Meant Time Between Failures (MTBF) or the average shelf life under near ideal/reasonable storage environmental conditions.

Helpful as these generally optimistic figures may be in a generalized context of choosing forward looking technologies, the fact is that any specific piece of media can vary widely from the manufacturer's specifications. Not only that, in practice, for instance, we find that with a large enough sample of tapes it is not uncommon to get a 1-2% media failure rate even under the best circumstances. Higher error rates raise red flags and require more detailed examination to establish causal factors.

What does this mean in the context of a particular project? Be prepared for surprises. No two tape restoration and processing projects are ever exactly the same. Flexibility and expertise are essential to overcoming any obstacles uncovered in the course of a project.

Legacy data ingestion into a compliance related repository is often treated as a standard operational procedure, especially when executed in-house. This can be problematic from a number of perspectives, particularly if scrutinized by regulatory agencies or courts.

For one thing, most organizations have little experience with this type of work and typically encounter many issues along the way. These issues may span technical elements as well as operational considerations. Even if these issues do not halt the execution of the project, they may endanger the integrity of the results, sometimes without a recognition of the problems by staff responsible for the ingestion.

Thus tape processing for compliance engenders considerations above and beyond what "normal" tape restoration may encompass. Simply put, tape processing in a compliance or litigation context needs to be legally defensible in a court of law and withstand judicial scrutiny. This places additional burdens on top of the traditional technical challenges that tape restoration represents. In order to assure the integrity and defensibility of the process in court, tape processing needs to be transparent, consistent and fully documented. Since the traditional legal practice is to attack the process if you can't dispute the results, you need to ensure that the process will bear up under such examination.

In-house or Outsource?

- **Legal Defensibility**
- **Stress Points**
- **Penny Wise..**
- **Risk reduction**
- **Is there Really A Choice?**

There is wide, and ever increasing, industry acceptance that loading up an email repository with legacy emails simply makes sense and allows a fuller ROI on the solution investment. Alongside with this recognition, one of the very fundamental decision management faces is whether such effort should be undertaken in-house or outsourced. It is instructive to examine this key decision point organizations encounter when they do commit to a legacy data ingestion of their repository.

In-house versus outsource issues tend to spark passions and be debated rather fiercely. It is hard to avoid partisanship and it is important to lay out as clear and cogent rationales for either approach and let the logic guide the decision making. Having said that, I must be just as clear that this article lays out the underlying rationale for outsourcing, simply, because in this case the author cannot find a powerful enough argument for the alternative.

Even best run organizations that have recognized the need for backfilling often stumble with this issue. Whether this is because of a good faith, but misguided, from a business and legal perspectives, desire to cut perceived costs, or a more general organizational reluctance to outsource anything, companies have a very mixed track record in this arena.

However, anyone tasked with legacy data restoration, whether from tapes or optical disks, very quickly discovers just how challenging this can be when attempted in-house. There are many reasons for this. For one thing, few organizations have the experience or resources to undertake this type of work successfully. On the other hand, it simply does not make sense to make the extensive investments necessary to meet a one time, or, at best, an episodic need such as legacy data ingestion.

Legal Defensibility

It is no secret that many cases and regulatory sanctions have involved spoliation issues associated with in-house attempts to process discovery requests that have backfired. The consequences have often been catastrophic, ranging from heavy fines to the most severe sanctions.

Most often, in-house resources are simply not up to the task to assure timely, legally defensible production. The risks become too great as the courts become increasingly impatient with any delays and errors in discovery. Any mistakes become open fodder for the opposing party and find increasingly receptive audience in the judiciary.

It is imperative, therefore, to assure integrity, defensibility and timeliness. In practice, the typical obstacles to achieving this take many forms:

- technical hurdles - it is far more difficult than first appears
- resource availability - it requires far more resources and effort (read cost) than can be anticipated if you haven't done it already
- confidence - low reliability and integrity of any in-house process
- time - it takes far longer if you are not prepared and have the appropriate resources tested and available

- mistakes are costly - you don't want to learn from your mistakes
- hassles - it can be overly disruptive of your normal operations
- can you prove it - in-house efforts tend to lack the legal credibility of outsourced work performed by recognized experts in the field
- experience - in-house staff are at a disadvantage to apply the lessons that an experienced vendor brings to bear on such projects

Stress Points

Because of the above challenges, we have seen many in-house efforts encounter the following common disappointments:

- dramatically extended time frames
- unexpected cost overruns
- suspect results
- inefficient and duplicative efforts
- multiple reversals and re-dos
- risks increased rather than lessened
- inability to respond effectively to the original goal of the effort ie litigation support and compliance
- impact on other, mission critical, business priorities

Penny Wise and Pound Foolish

In most cases, the IT department is first tasked with the responsibility and costs. Many IT departments feel challenged by the deceptively simple, on the surface of it, proposition represented by tape restoration. After all, they created the backup media (tapes or disks) in the first place and are the likely custodians. How difficult, the argument, goes can it be to just restore the old backups and ingest the contents into the archive.

There is often the budget dilemma where management, sometimes unappreciative of the level of effort and cost, allocates, minimal, if any, additional funds for this undertaking, and, therefore, it may have to come out, at least partially, out of the overall IT budget. Given these pressures, the hapless IT folks undertake to do the organization another big favor and try to meet the challenge themselves.

Aside from everything else there is the traditional IT reluctance to outsourcing in general. Sometimes, this may be driven by a good faith effort to save cost. Unfortunately, often, this may reflect poor business judgment such as is associated with a lack of a larger view. Surprisingly, from a business perspective, many technology decisions are still driven by the "must not outsource anything, if I can help it" orientation.

Or, the in-house staff feels challenged by what, at first glance, seems like an deceptively simple proposition. This, "if they (an outside vendor) can do it, surely we can do it too", mentality has been a traditional bane of effective decision making in technical environments.

Risk Reduction - But Will it Play in Court?

The legal credibility of having hired an outside expert is often itself well worth the price of outsourcing. The simple fact is that in-house efforts can raise legal issues that may be hard to overcome and can undermine the whole value proposition of the solution being implemented. Once attacked in court, a poorly documented process can be a red flag for further litigation.

We have seen more than one organization outsource the process only after having already executed some part of it in-house, due to a lack of confidence in the original outcome. If nothing else, you may find your in-house IT staff needing to testify regarding the processes utilized.

By the way, as most legal professionals recognize, in-house IT folks are not necessarily your favorite witnesses.

Is There Really a Choice?

Legacy data restoration is commonly perceived as, at best, a necessary evil. No one gets excited about being saddled with the task. Among other things, it is episodic and therefore not worth substantial long term investment. It is hard to get a clear ROI on such in-house investments.

Outsourcing, on the other hand, allows an organization to concentrate on mission critical efforts. In these times when high caliber technical staff is always hard to come by, it seems like poor utilization to have them saddled with legacy data conversions, it certainly doesn't make them happy. Smart IT management understands these issues, as well as the fact that nobody wins medals for legacy data restoration, but many have lost their jobs because this was not done successfully.

Many organizations, unfamiliar with the challenges of backup tape restoration assume that they would be better off doing this in-house. After all, the logic goes, it is only a tape backup, if we created it what could possibly be so difficult about restoring content from it. As understandable as it may be do attempt to lower costs by doing this in-house, this approach is often full of little understood peril. The larger the collection of legacy data, the greater the risks and the greater the need for professional help.

The fact is that legacy data provisioning of an archive is one of the more easily justifiable activities to outsource that any organization can encounter. The business logic behind this is straightforward when laid out clearly. Given the increased risk profile associated with not executing such work reliably and in a timely manner, few organizations can afford to play with fire.

How to Select a Vendor Partner

- **plan and prepare**
- **choose vendors wisely**
- **avoid last minute fire drills**
- **evaluate options ahead of time**
- **sign an MSA**
- **work with the vendor**

The right vendor can make a world of difference in assuring success of the backfiling process. Experience, reliability and integrity are key elements to look for.

Typical selection criteria:

- experience and track record
- focus
- technical capabilities
- scalability
- customer service orientation
- financial stability

Check references carefully, visit the vendor site(s), get sample pricing, all in the luxury of unlimited time unlikely to be available in the midst of a pending litigation or regulatory investigation.

The first step in preparation for legacy data ingestion is to verify and formalize a relationship with appropriate providers ahead of time. After developing a short list based on an evaluation process such as issuing and evaluating responses to an RFI (Request for Information) you can be comfortable formalizing a relationship. A simple instrument of this is to sign an MSA (Master Services Agreement) with selected vendors(s) ahead of need, while you have the luxury to negotiate the best deal.

It is imperative to create a cooperative environment of mutual trust that is the basis of success. Open and frank discussions are an integral part of this. A vendor must be willing and able to discuss promptly any concerns and issues that may arise.

In few undertakings are details of such importance as in legacy data restoration and ingestion. They drive scope of work and all the associated elements such as price, turnaround and quality. Not all vendors are equal in their capabilities and price. In fact, tape processing is such a dynamic field both capabilities and pricing can vary widely from vendor to vendor and change on nearly daily basis. It pays to shop and check credentials.

Given all that, there may only be a handful of vendors that really specialize in tapes and are proficient at handling them. In brief:

- don't underestimate the complexity of the undertaking
- rest assured that vendor capabilities exist to deal with just about any challenges tapes represent
- the cost and timeliness may be surprisingly affordable given state of the art vendor capabilities

At the same time, the business case for utilizing specialists to process tapes has never been stronger. You simply cannot reproduce the resources and expertise necessary to do tapes quickly and reliably, on the required scale, that people who do this for a living have available. It's smart to piggyback off the prior efforts and lessons learned that a specialist in the field brings to bear on your behalf. A learning curve in this industry can be very expensive, even disastrous. Not to mention the various proprietary tools a viable vendor has behind its belt, to be able to overcome the common limitations of more traditional approaches.

What NDC Brings to the Table

- **20 years of experience**
- **40,000 projects**
- **5,000 clients**
- **clear focus**
- **track record**
- **unique tools**

For nearly 20 years NDC has specialized in a very narrow but critical aspect of the compliance services spectrum, tape restoration and processing for email archives and electronic records repositories. NDC has experience with the largest tape restoration projects ever undertaken in the history of email archiving. We know how to scale and meet the tightest deadlines. We also bring to the table a refreshing level of expertise, professionalism and dedication.

As pioneers in the industry, we have developed the many innovations that are indispensable to success, such as non-native file restoration. We offer:

- largest staff dedicated to data conversion
- unique, proprietary technologies and methodologies
- unparalleled track record spanning 20 years, 40,000 projects delivered to over 5,000 customers
- hundreds of thousands of tapes processed
- largest email restoration project ever undertaken
- specialized and focused on core competencies

The special value proposition that NDC brings to the table is our clear focus and unparalleled track record with this type of work. We restore electronic content from source media such as backup tapes and optical disks. We then cull it for content of interest, de-duplicate it and normalize it into customer selected target formats. This can cover both emails and other ESI (Electronically Stored Information) types such as loose documents, databases, document images and COLD streams and so on, including some very esoteric, leading edge formats such as voice logging tapes.

Our services encompass:

- state of the art tape and optical disk restoration
- email and other content processing (culling, de-duplication and normalization)
- voice logging/mail processing
- advanced legacy data conversions
- media duplication (tapes and disks)
- highly scalable, large volume processing capacity

We have worked, typically behind the scenes, on many of the most prominent cases in the country, supporting the various stakeholders. Our services have been utilized by the top tier firms in the nation, including the largest law firms, full service consultancies and most Fortune 100 corporations

NDC offers the experience, track record and time-tested methodologies, combined with a huge pool of dedicated resources to assure peace of mind.

Even though we have participated in some of the largest litigations and compliance efforts in the country, we are largely unknown in the industry since we have mostly worked behind the scenes supporting the efforts of the various stakeholders.

About the Author

- Founder & President

Michael Daniec is the President and CEO of National Data Conversion (NDC). Michael has over 30 years of dedicated experience in both technology consulting and management of computer-related business enterprises, culminating in the founding of NDC in 1987. Under his leadership, NDC has played an instrumental and pioneering role in the evolution of the tape restoration and data conversion industry, and he has grown the company to its current position as a nationally recognized leader of these services.

NDC has completed more than 40,000 successful conversions over the life of the company. As a testament to its success, NDC now represents many of the world's largest and most prominent businesses, financial institutions, law firms, branches of government, most prominent cultural and educational institutions.

Michael's earlier experience and thorough apprenticeship in the industry laid the foundation for NDC's success and has provided him with a comprehensive and intimate understanding of both the technical issues and business challenges involved with NDC's services. Prior to establishing NDC, Michael spent five years managing other computer technology-related companies. He started his first company in 1982, selling custom programming for business applications, including various business process automation applications. Michael's dedication to the industry resulted in his progressive mastery of the most advanced technical issues associated with the development and delivery of high tech professional services and the development of his credo "a proud dedication to service and excellence".

Under Michael's leadership, NDC became an early leader in the tape restoration and conversion industry and pioneered many of the methodologies currently in use in the industry, including non-native tape restoration. NDC currently offers the greatest range and depth of such services, with the most technically advanced methodologies.