



Designing Optimized Use of Business Information with the IBM Global Archive Solutions Center

Analyst: Anne MacFarland

Management Summary

Retaining business information beyond its immediate use is insurance should things go wrong – but it is using and leveraging that information that assures that things go right. It is optimizing that use of information that builds the competitive differentiation upon which business success depends. No one else has the business knowledge that you have – if you use your business information well. Optimizing business use of information is not the same thing as assuring its availability within an application process. The insurance-coverage paradigm of *backup* will not do the trick. Instead, **one must turn to an older paradigm – the organization of information for use and tactical reuse known as *archiving*.**

The rationale for archiving is that *information*, as documentation of a situation at a point in time, does not and should not change. When the immediate purpose for which it was captured has passed, it still has value as historic information for the trending and analysis that underlie business strategy. Cyclical events, such as quarterly closes, and sporadic events, such as hospital visits, make certain information once again critically important. Back when capacities and connectivity were limited, the only choices for information were *keep* or *delete*. Technology now gives organizations the tools to support a potent middle-ground environment, where a wide variety of information can be offloaded yet still accessible for fairly immediate use. The offload gives the benefits of unburdening overstuffed information processes (be they databases, file systems, or email), which improves performance and may ease licensing costs. It can ease application upgrades and, more particularly, application retirement. Creating a digital archive provides a single and unified point of reference (avoiding many variant or corrupt versions of the truth). It creates a separate zone of management, where rules not needed within the generating or capture process can be defined for long-term retention, security, use, and, where appropriate, eventual deletion. Often, an archive can reduce operational and energy costs. **Properly done, it always helps the organization work better.**

Devising an archival strategy is not just a matter of storage hardware and information management software, though they are certainly essential parts of the solution. **What is needed to keep business information usable and findable, while protecting customer privacy and business confidentiality, will depend on the nature of both business operations and the nature of market competition.** The devil is in the details, and a completeness of information is important. So, **designing an archival strategy quickly can become overwhelming.**

IBM has been amassing relevant assets and partnerships, and recently opened a *Global Archive Solutions Center* in Guadalajara, Mexico, where customers and potential customers can come to meet with experts that will help them craft, test, and validate the retention aspects of their Information Infrastructure. For more about the process of constructing this archive tier, and how IBM can help, please read on.

IN THIS ISSUE

➤ The Evolution of Information Infrastructure.....	2
➤ Getting Ready to Leverage Information by Digital Archiving	3
➤ The IBM Global Archive Solutions Center.....	4
➤ Customer Use Cases	5
➤ Conclusion	6

The Evolution of Information Infrastructure

Businesses have always been a matter of optimized use of *information* and *process*. In the past few decades, much attention has been paid to the optimization of business process, leading to many tools focused on making process work better and, particularly, faster.

Business information has also received its share of attention, in database and data warehouse development, business intelligence analytics, records management, content management, and other initiatives. More recently, regulatory compliance has extended information retention requirements. Information has also become a competitive weapon, for as the Internet has increased competition and pruned customer loyalty along with profit margins, repeat business must be earned, usually with the assistance of the right information at the right time in the right context.

Thanks to the Internet and telecommunications networks, business process has taken well to distributed operations. Business information has not been so lucky. **The coordination, synchronization, and security of information across the increasingly disconnected enterprise are a challenge.** That challenge grows more difficult as the growth of structured and unstructured data, on which the business depends, accelerates.

There are many reasons for that growth.

In many industries, *businesses have consolidated*, either organizationally or by partnering, to gain efficiencies from the economies of scale.

Government regulations now mandate that more kinds of businesses document their businesses more extensively and keep that documentation longer.

New sources of information about the business operations are now available – from IT systems, RFID sensors, etc. Most of this information comes in masses of detail. Analysis can derive useful information that can optimize operations.

There are *more sources of information about markets* – not all of them trustworthy – that can form the basis of a much richer characterization of markets and threats.

Information itself has gotten bigger. Infor-

mation from medical equipment, now digitized, requires more and more storage per image as higher resolutions that can improve diagnoses become available.

This wealth of information presents opportunities to optimize the business in many ways – and offers many ways to do it badly. If customer information is not kept private, corporate brand value will erode. If company information is insecure and altered or corrupted, both customers and partners will have trouble doing business. If information is not findable, people will do the best with the information they have – often with unfortunate results. Cindy Grossman, IBM Vice President, Tape and Archive Storage Systems, notes, “In the last two years, retrieval of information, not just its retention, has become top of mind.”

Keeping more information on production systems, particularly in a time of escalating energy costs, is neither prudent nor useful. The larger the production database, the longer queries take, slowing exactly those transactions that must be optimized for speed. The larger and more disorganized the file systems, the harder it is to find information, particularly if no effort has been made to keep information findable. Often, impatient users resort to creating short-term data marts and stashes of business information that represent a security risk (if they are unknown to IT and uncontrolled), an operational impediment (if duplicate data is itself duplicated), and a business risk (if the information used is out of date – or incomplete).

Business information about a given point in time should not change. If you want to be able to track the pace and cadence of business operations, you will want to preserve it as historic data. If you want to be able to access it for analysis in various business contexts, you will want to put it into a repository that facilitates and tracks this mutual access. You will want to know where it came from, in what context it was generated, and how it has been transformed and cleansed along the way. This involves not only storage but also infrastructure to support timely reuse, containers to organize the bulk and track its use, and software to keep the information findable. How much of what technology is needed depends on the nature of an organization’s information and how it wishes to retain and leverage it. Therefore, an information assessment is needed to determine the strategy.

Getting Ready to Leverage Information by Digital Archiving

Architecturally, there are a few familiar routes to take. The federation mode of *grid* is infinitely flexible and allows some *status quo* to be maintained, but it can be a less-good paradigm where security, privacy, and tracking of who used what are important. Highly-scalable file systems address the above limitations, but are a less good match when certain information elements must be called out and manipulated for analysis. Database management systems specialize in presenting information for analysis, but their data structures can be unfriendly to new sources of information and unexpected uses of information. The creation of information objects, managed by a registry and facilitated by search, is a newer paradigm that relies on the quality of the registry, the search engine, and vocabulary control.

In general, **the watchword for long-term archives is *less is more*, for all structures have to be evolved due to migrations to new media and media standards, and to new or significantly-revised applications.**¹ In most cases, a more generic repository, accessorized with a separate tier of tools (that can evolve over time) is a better approach than a “magic box” with proprietary elements. Keeping the tiers discrete is an essential discipline, because any dependencies between layers make information migration more troublesome. But, as you will see later in this bulletin, the variety of approaches taken by IBM customers who attended the opening of the new Global Archive Solutions Center, all depends on the nature of the organization and what it is trying to do.

To get started, it is often useful to think about *the information that you have to work with* rather than *what you need to buy*, whether it be applications or architectures.

1. Think about the topography of your information. This is best done on a business unit basis. This often surfaces cultural issues of who owns the information – issues which must not deter a comprehensive survey. Often the following bulleted categories are helpful.

¹ For more information on the challenges of long-term archiving, see the issue of *Clipper Notes* entitled *The Long-term Challenges of Digital Archives – What You Really Need to Know*, dated June 6, 2008 and available at <http://www.clipper.com/research/TCG2008030.pdf>.

The information that gets business done. This includes product data, customer data, transactional information, and operational training materials. For health care, this information includes both physicians’ notes and the primary sources of MRIs etc. For oil and gas, pharmaceutical and media and entertainment, it includes images (seismic studies, protein folding, and animation files, respectively) that are expensive to generate and beneficial to reuse. Much of this information evolves rapidly, and it is important that people use the most current information. Centralization of data in a single place is the most efficient way of ensuring currency – and with today’s information-feed mechanisms, this centralization does not have to be a burden on the mobile or detached worker. This kind of information is where access-layer situational applications can come into play.

Documentation of the business can include financial data, email, contacts, and reports. Many of these come armed with business logic and metadata tags. Traditionally, most of this is managed in a records management system. Over time, keeping access to a completeness of this information is crucial in order to avoiding unwanted risk and bad decisions.

Information about business operations (often granular) that you want to capture and analyze for intelligence about market trends. This can include external sources of information that should be documented as to their limitations and bias. While individual sources may not be entirely trustworthy, aggregated over time and analyzed with all the tools and visualization now available, they often give insights into situations beyond what is possible using only internally generated information.

Other kinds of information. Inevitably, there is information that does not fall into the categories above. Much of this information has no enduring business value. The rescheduling of a meeting is rarely indicative of business risk or malfeasance. It is prudent to make policies to destroy this kind of information, as well as multiple copies of information as soon as is reasonable.

2. Think, then, of the use and possible reuse of this topography of information. Think specifically of use requirements by particular roles

within your organization. Who will use the information? How? What is the retrieval time frame? Sometimes, as in the case of medical images, there are multiple sets of requirements. For predictable events, such as doctor's appointments, information may be staged to a local office for immediate invocation at the time of the appointment. In emergencies, however, the information must still be available in a timeframe to promote good health outcomes. These explorations probably are best done by those who guide business strategy, and by operations experts who can articulate how, exactly, the information is used.

3. Think of the lifecycle of usability for your various kinds of information. This, as well as the government regulations that apply to your business, will help you start thinking about retention scheduling.

4. Think about risk. Risk is often independent from value, and merits a separate assessment. Information is an opportunity to glean insight – but also can be a liability.

5. Start a pilot digital archive strategy by focusing on a situation where there is great need or pain. Ideally, it should be a situation where the benefits of better information management are demonstrable and relatively easy to attain. Turn to IT to map what you have found to the information sources and applications involved.

Now, you are ready to leverage the IBM Global Archive Solutions Center at Guadalajara.

The IBM Global Archive Solutions Center

At the opening of the Global Archive Solutions Center, Andy Monshaw, General Manager of IBM System Storage, stated, "Storage alone is not an archive solution². In current technologies, storage of information for reuse is neither affordable nor sufficient. You need long term key management, reference architectures with the discipline to make tools like search effective at finding all the necessary information." He and IBM envision the facility at Guadalajara as a place where IBM most effectively can help customers build digital archive solutions for their specific situation, based on the IT assets they already have (not all of which will be from

IBM). It will leverage IBM and third-party technologies as a toolkit for business-focused innovation, and it has the layout and telepresence to call on expertise wherever it is located as the need comes up. This is far more than a briefing center.

The site in Guadalajara has been an IBM manufacturing facility since 1972. Since then, the area around Guadalajara has grown into a national center for high-tech companies and contract manufacturers. The existence of local universities, software expertise, and good telecommunications makes this a good site at which to design archival solutions to meet the imperatives of customers.

The Center features meeting rooms with extensive telepresence capabilities to leverage expertise from across the globe. It has a demo lab with servers and storage devices for customers to see running archival solutions to help them determine which best meet their needs.

IBM has enlisted every business unit to contribute its relevant expertise. Services from IBM Business and Consulting include an Archiving Assessment, Archive Strategy and Development, Archive Solution Architecture, Build and Certify Pilot, and Archive Architecture Deployment services – all tools customers may need to accelerate the archive project to provide business value sooner and more reliably. A broad range of software tools is relevant to archival endeavors, but it is worth calling out several products crafted particularly for archival projects.

IBM Information Lifecycle Management Services for Grid Access Manager is the brains behind IBM's GMAS³ Solution for PACS⁴ fixed-content data, featuring redundancy, multi-site support, and automated data migration to avoid hardware obsolescence.

IBM System Storage Archive Manager (SSAM) is Tivoli Storage Manager, tailored with software to disallow erasure or corruption.

IBM Scale-Out File Services (SoFS) takes IBM's GPFS (*General Parallel File System*) and marries it with SAMBA and *Clustered NFS*.

² Of course, the visibility of IBM's manufacturing process and full storage product line in the Guadalajara facility is not an accident.

³ Grid Medical Archive Solution.

⁴ Picture Archiving & Communications System, the standard for medical imaging.

All the Industry Models, Content Management, Database options, Master Data Management, Business Intelligence and Performance Management products from Cognos, and other elements that comprise IBM's *Information on Demand* strategies.

All the virtualization, de-duplication, compression, encryption, and other tools that comprise IBM's Storage Management.

A complete line of disk and tape offerings, including non-re-writeable and non-erasable storage, such as the *DR550* (a blended disk and tape compliance storage solution), *N series with SnapLock*, and the new high-density tape frame for long-term archive storage of LTO and 3592 tape cartridges⁵.

All the system monitoring, orchestration, service management, and service composition, security, system management and other tools that populate the Tivoli realm.

Open source elements, including tools like UIMA⁶, which IBM donated to open source, which facilitate the extraction of database structures and file attributes as metadata, enabling the persistence of more self-describing archive assets.⁷

IBM has many areas of archives-focused research. The broader initiatives include investigations of mega-scalability, pluggable architectures (leveraging POSIX), policy-driven management, and metadata specifications (industry standards). More-specific initiatives include work on data objects, scalability of archival ingestion, scalability of search, and dynamic transition to new infrastructures and access methods, as well as development of single-instance-infrastructure that can be addressed by multiple applications.

Customer Use Cases

At the opening of the facility on May 21, 2008, three customers described their archives. Each was very different.

Spectrum Health

At Spectrum Health of Grand Rapids,

Michigan, a system of 7 hospitals, 130 locations, and 2000 beds, uses IBM GMAS. The rate of growth of the use of the PACS system is 2% per month. At present, the growth of data requires two terabytes of additional storage each month. Spectrum's GMAS implementation has grown to 90 TBs in each of two redundant sites. To meet the data retrieval parameters of its radiologists, Spectrum uses 9-10 TBs of high-speed disk as a redundant cache. They had the last two years of images on disk, but this did not meet the need for a complete longitudinal history of each patient – ASAP.

The PACS system manages the retrieval and security of images, using digital fingerprints. Archive-level controls are less important – having PACS “in control” meets the use parameters, although Spectrum does see a need to create a separate tier for image management.

Improving image retrieval from fifteen minutes to five seconds turned the radiology teams into staunch supporters of the archives, and Spectrum is looking to add images from cardiology and other diagnostic areas.

Global Data Vaulting

SaaS vendor Global Data Vaulting offers archiving, content management, data protection and data recovery services to mid-market companies in the Americas and Europe. Leveraging several components of IBM's *Tivoli Storage Manager*, Global Data Vaulting can cut storage costs for these small businesses in half, while providing data recovery in minutes. It uses several redundant data centers at collocation service providers.

Customer preferences vary, as some want to free up on-premises storage, and some want to keep many copies of databases. All want instant availability of information on demand, and Global Data Vaulting assures this by mandatory quarterly recovery testing. It also is capable of escalating the service, in times of customer crisis, to more frequent saves of a more comprehensive domain of data. All information is locked down and secure.

Absa Group

Absa Group, a subsidiary of Barclay's, has acquired and operates many banks across sub-Saharan Africa, each with its own regulations and banking and insurance practices. Absa has a vision of an end-to-end visibility for all its diversity of operations, supported by a reusable

⁵ See [The Clipper Group Navigator](#) dated July 16, 2008, entitled *Lowering the TCO of the Data Center - IBM Innovates Tape Architecture ... Again!*, available at <http://www.clipper.com/research/TCG2008037.pdf>.

⁶ Unstructured Information Management Architecture.

⁷ The Storage Networking Industry Association, SNIA, is working on standards in this area.

framework, while maintaining compliance with all local regulations. Absa found that Enterprise Content Management could not provide a path to its goal. A transition to records management is the first step, but the company believes the long-term solution lies in an archival approach.

Absa has 50 TBs of production storage in 400 repositories, all with different data definitions. Rationalizing this environment, in the rigid timeline that Barclay's has set, is a complex effort. IBM had the breadth of data and content management tools and services to develop a solution that has kept the adoption of records management on schedule and successful. Barclays is looking to replicate this approach in all its geographies.

These customer examples typify the diversity of requirements, and their strategies show the variety of approaches that an archive can take.

Conclusion

If your business is suffering from an information glut, many IT vendors will be glad to offer a solution based on their discipline – but what you really need is a combination of multiple disciplines, enhanced with key processes from several others. Since this is, by its nature, a long-term investment, you will want to see these potential solutions running on real infrastructure in the company of experts who can advise and transfer knowledge to assure successful operations of a digital archive solution.

IBM and its partners offer the widest array of retention technologies and tools to bring to bear to create an archive solution to support a particular organization and its processes for the long term. The new IBM Global Archive Solutions Center, a first of its kind, offers a place to learn about and see industry-leading archival solutions that fit the organization and its budget. Leverage it!



About The Clipper Group, Inc.

The Clipper Group, Inc., is an independent consulting firm specializing in acquisition decisions and strategic advice regarding complex, enterprise-class information technologies. Our team of industry professionals averages more than 25 years of real-world experience. A team of staff consultants augments our capabilities, with significant experience across a broad spectrum of applications and environments.

- ***The Clipper Group can be reached at 781-235-0085 and found on the web at www.clipper.com.***

About the Author

Anne MacFarland is Director of Data Strategies and Information Solutions for The Clipper Group. Ms. MacFarland specializes in strategic business solutions offered by enterprise systems, software, and storage vendors, in trends in enterprise systems and networks, and in explaining these trends and the underlying technologies in simple business terms. She joined The Clipper Group after a long career in library systems, business archives, consulting, research, and freelance writing. Ms. MacFarland earned a Bachelor of Arts degree from Cornell University, where she was a College Scholar, and a Masters of Library Science from Southern Connecticut State University.

- ***Reach Anne MacFarland via e-mail at Anne.MacFarland@clipper.com or at 781-235-0085 Ext. 128. (Please dial “128” when you hear the automated attendant.)***

Regarding Trademarks and Service Marks

The Clipper Group Navigator, The Clipper Group Explorer, The Clipper Group Observer, The Clipper Group Captain's Log, The Clipper Group Voyager, Clipper Notes, and “*clipper.com*” are trademarks of The Clipper Group, Inc., and the clipper ship drawings, “*Navigating Information Technology Horizons*”, and “*teraproductivity*” are service marks of The Clipper Group, Inc. The Clipper Group, Inc., reserves all rights regarding its trademarks and service marks. All other trademarks, etc., belong to their respective owners.

Disclosure

Officers and/or employees of The Clipper Group may own as individuals, directly or indirectly, shares in one or more companies discussed in this bulletin. Company policy prohibits any officer or employee from holding more than one percent of the outstanding shares of any company covered by The Clipper Group. The Clipper Group, Inc., has no such equity holdings.

Regarding the Information in this Issue

The Clipper Group believes the information included in this report to be accurate. Data has been received from a variety of sources, which we believe to be reliable, including manufacturers, distributors, or users of the products discussed herein. The Clipper Group, Inc., cannot be held responsible for any consequential damages resulting from the application of information or opinions contained in this report.