



IBM's Smart Archive Strategy Supports the Maturing Role of the Electronic Archive

Analyst: Anne MacFarland

Management Summary

Archiving, in the more long-standing definition of the term, is about *assuring the retention and authenticity of information that will be needed at some indefinite point in the future*. This is not just a matter of quickness of access, but also of completeness of finding. It is not just a matter of guaranteed authenticity, but also of lineage and of context elements that guide appropriate use. **Archiving is, basically, the ability to have a wide range of “more” information available when needed. This more is not a matter of capacities (though that is often involved). It is a matter of better - better leveraging all the information that an organization generates and collects.**

Business has always had an insatiable appetite for information, but its consumption patterns are far different from those of a casual surfer who only grazes for immediate gratification. Businesses worry about what information should surface on a business dashboard for different roles. Supporting empowered employees and reacting promptly to events (good and bad) is very important, as is keeping an eye on the complex interactions of organizations that they serve. **An archive allows businesses to leverage a breadth of information while retaining fewer copies of it.**

In the past few years, several trends have both accelerated the archiving of information and challenged some qualities of first generation electronic archives. These four new imperatives will shape archives of the future.

- **More!** Data sources have proliferated. Much of the information used by business is from external sources. Some information is not highly trusted. Some comes from partners with separate agendas. Some is of short-term value. This drives new kinds of policies about relevance, quality, and business value.
- **Other!** RFID and other new information generators produce streams of information that do not require the high availability of transactional information. Kept in their entirety, they would be outrageously costly – but a sampling may have business value. This places new business demands and requires a rethink of information lifecycle management.
- **Now!** Immediacy of use has high business value. This applies to information in an archive as well as to real-time data feeds. Therefore, a presentation strategy must be part of tomorrow's electronic archive.
- **Smart!** Analytics of operational data has become a critical part of business strategy. This is best done with an offload of information. An archive can be an unobtrusive, supportive source.

With its *Smart Archive Strategy*, IBM has addressed all of these imperatives. For more details, please read on.

IN THIS ISSUE

➤ Archiving for Tomorrow's Information Environment.....	2
➤ How IBM's Smart Archive Strategy Addresses the Imperatives	2
➤ Conclusion	5

Archiving for Tomorrow's Information Environment

Electronic archival repositories have been around for several years now. **The concept of archiving has successfully changed from meaning “hopefully permanent retirement of data” to “offloading data for continued use”.** This change is complete and irrevocable and comes just in time for large organizations.

As more of business process becomes digitized and automated, and as the human part of organizations becomes more mobile, outsourced, and frequently repurposed, institutional memory is in peril. Technology-assisted collaboration, which has allowed individuals to leverage a common knowledge base, does not help when you need trustable, non-current information.

Team collaboration environments, such as *Lotus* or Microsoft *SharePoint* aggregate content, but they are not designed to support long-term archival capabilities and controls. Hence, a separate archive is needed. This archive should be broadly imagined. For some organizations, archiving may come as a single centralized box, particularly if process latency is an issue. For others, a gateway fronting a variety of digital assets is more useful. Other digital archives may come as a federated infrastructure, as SaaS¹, or as a cloud service. **The imperative for archiving is strategic, but its actual implementation is situational.**

An archive has IT as well as business value. For IT, the data offload of archiving keeps real-time application performance from diminishing. When archiving is more than migration with a stub left behind, the ingestion process harvests enough application metadata² so that data can be used independently of the application that generated it. This *full retirement* archiving can make application upgrades less painful and application retirement more feasible. In either the stub-facilitated or the full-retirement case, archiving can reduce capacity-based software license charges, can allow use of lower-cost storage tiers, and can lower energy costs.

Archives have been around for long enough

¹ Philosophically, SaaS (Software as a Service) is a precursor to cloud service offerings. SaaS is application specific. Cloud services may include infrastructure services that may be independent of a single application (e.g., file storage or testing environments).

² The application-specific metadata that is harvested or created as part of the archiving process becomes the basis for classification, and a guide to proper reuse.

that many large businesses have dozens, even hundreds of them.

- Organizations archive information from their massive applications (SAP, databases, etc.) to improve performance and responsiveness.
- Archiving is popular for email, file shares, or desktops driven by litigation realities and the need to control the expense of eDiscovery.
- Archiving is a good and increasingly common strategy for medical imaging, driven by the need to get all the relevant information about a patient in a timely way, while preserving patient confidentiality.

With the popularization of predictive analytics, many other kinds of data feeds (operations surveillance, market baskets, logistics, etc.) become analytic targets that must be trustable and well organized. The archive approach offers one method of aggregating (if only logically) and ensuring such targets in a sustainable way. If the data guides strategy, it is worth preserving, if only for the Monday-morning quarterbacking.

This is not always easy. Arvind Krishna, General Manager of Information Management in IBM's Software Group, stated at a recent event, “the nature of organizational information is that it is somewhat ordered but never neat. It is distributed, tiered, heterogeneous and addressed by an increasing variety of applications. Search on large volumes of data is only effective when the vocabulary involved is regulated, which often is not the case.” **He sees the new challenge as managing the burgeoning whole – a challenge that can be addressed with indices, reconciliations, metadata systems, intelligent data partitions, and taxonomy generation and evolution.**

How IBM's Smart Archive Strategy Addresses the Imperatives

IBM's *Smart Archive Strategy* unites the power of IBM software, systems and services capabilities to offer a comprehensive vision and common architecture for customers to maximize the value of their archiving strategies. **In any broad initiative, consistency of architecture and process support both a greater range of use and more scalability.** As well as all the deployment options mentioned earlier, IBM Smart Archive solutions can be delivered as a custom service, as SaaS, or via a cloud.³

³ These service options form a spectrum much like the infrastructure spectrum mentioned earlier, stretching from custom to contract to utility service, with economics that improve as local control diminishes.

IBM's Smart Archive Strategy now is accessorized with many new management pieces to enhance its business value. **At the heart of the IBM offering is a set of tools that, together with a system guaranteed to enforce retention and guarantee authenticity, gives organizations a way to achieve fuller use of what they retain and expansion of what they know.** The following will cover just some of the tools most relevant to the imperatives of *more, other, now,* and *smart.*

More! Meeting the Demands of Rapid Data Growth

With the growth of business data and the new variety of data useful to the business, more discipline about deciding what should be archived becomes necessary. While most archived data can be kept on less-costly storage media, the ingestion process is not without cost. These decisions are business decisions, not an infrastructure operation. Clarification and documentation of the relationships between different data sources will aid this process, and will limit improper use.

IBM InfoSphere Discovery

IBM InfoSphere Discovery automates analysis of data and data relationships. Deriving from IBM's *Exeros* acquisition, it works with both structured and unstructured data. Its discovery of data relationships documents context that can help define the shape of the collections in which archived data is held. **The more organized an archive is, the more useable its contents become.**

IBM Classification Module

The *Classification Module* supports full text analytics and gleans taxonomy elements. It supports a semi-automated *Assisted or fully automated Classification* process. This is useful for information that might formerly be known as "miscellaneous".

IBM Content Collector

IBM Content Collector is a modular policy and task routing engine that speeds information ingestion. It consists of *Source Connectors* (of which more are available now), *Task Connectors*, and *Target Connectors*. These work together to automate ingestion of information from source to target in a consistent manner.

The tasks handled by Content Collector include arcane litanies like *Records Declarations*. More custom tasks (perhaps those that are specific to a part of an industry) can be included

via an API and IBM Business Partner support. These capabilities help cross the *last mile* chasm between the templates of general business practices and the particulars of IBM's Industry Models⁴ and the particulars of a specific organization.

Archiving Solutions for eMail, File Systems, SharePoint, and SAP Applications

These solutions target many common pain points regarding business information management for specific environments. The integration of *IBM Content Collector* with *IBM Optim Data Growth Solution* for unified content and data archiving (for both structured and unstructured data) adds additional value.

This offering includes specialty ECM connectors for SAP archiving and document enabling, which handle both SAP-generated documents and data. IBM's *Optim Data Growth Solution* archives data from all other enterprise applications (such as those from Oracle, PeopleSoft, JD Edwards, and Siebel) and well as custom and legacy database applications. Taken together, this combination of IBM products not only addresses business information archived from inadequately secure situations (like SharePoint servers and file shares), but highly specialized situations as well (like email and SAP) in ways that work for both IT and the business.

IBM eDiscovery Manager and Analyzer

Another tightly targeted product, *IBM eDiscovery Management and Analyzer*, focuses on early case assessment optimization. Early case assessment feeds the settle-or-continue the litigation decision. This specialized tool saves a lot of money for organizations that are frequently in litigious situations.

These are a sampling of tools specific to archiving operations. Many other tools that address the challenges of data and capacity growth, such as deduplication, storage virtualization, etc., are found in IBM's Tivoli, System Software, and System and Technology Group portfolios.

"Other" – Embracing new Data Types and Sources

Many organizations are now making strategic use of information from external sources – information that often comes as *orphan information with no lineage*. Like crowd-sourced information, it can be of critical importance when used well.

⁴ The number of Industry Models IBM offers is expanding.

Many more data sources are now available to support operations and the decision making those operations require. Using a standard set of information sources is unlikely to provide business differentiation. Aggregates of detail information reveal many insights with business value.

Some of the new information sources come as data streams – event streams, transaction cadences, Web metrics, and the output of sensors and surveillance. Once deemed trivial in the larger scheme of things because of their limited context, in the aggregate they reveal patterns of activity and trends that have critical business value in an era where there are fewer eyeballs to watch what is going on in real time.

In these situations, sampling is appropriate and the classic transactional attributes of atomicity, isolation, consistency, and durability are not as important. Still, archiving samples from these streams, like archiving old medical images, has enduring value. One just must be able to winnow aggressively. A new tool, *IBM Content Insights for Assessment*, can help.

IBM Content Insights for Assessment

IBM Content Insights for Assessment uses content (text) analytics to make management, archiving, and trust decisions. It can also support the decommissioning of information that would otherwise be ingested – information that is duplicated or outdated, for instance. This tool also can decommission file shares and social software structures and information that were retained, while attaching provenance and history-of-use information directly to the content, as is desired.

This tool can also explore and analyze data in place, before it becomes a target of an archiving initiative. It can search, mine, and understand entity relationships. It supports classification models and the development of collection policies.

IBM Enterprise Records

IBM Enterprise Records enables customers to manage all records regardless of type, media or storage location throughout lifespan. It is DoD 5015.02-STD certified and comes with advanced information design, file plan, classification and meta data management including advanced context classification, integrated Records Federation Services for non-IBM repositories. It delivers complex event and records process management including intelligent retention, dynamic legal holds and audited disposition management.

Now

The imperative of immediate use continues to challenge existing practices. This is

particularly the case with electronic archives, where shortening the ingestion time has been an ongoing challenge. Parallelization seems an easy answer and serves when the processes are limited in scope. To design a process that is *smart* as well as *rapid* is not so easy.

Another, often ignored side of *NOW* is the retrieval and presentation side. Here, IBM has a sophisticated asset to bring to bear on the archive use challenge

IBM Mashup Center

Finally, *IBM Mashup Center* is an existing, well-evolved completer product that enables broader use of archived data – allowing it to be used like any other kind of data feed into dashboards and other web application presentation modes, independent of both the archive and the web apps that are its consumer. This extends and completes the circle of information reuse.

Smart

Smart means business-smart. A variety of deployment options, and different levels of data capture and protection (depending on the nature of the information and its use), are one kind of smart because all options leverage a common architecture and common tools, such as those described above. This makes the federated whole, however you deploy it, into a much smarter federated whole.

Smart Range of Deployment Options

IBM Smart Archive Strategy offerings can be deployed as a completely IBM-built solution leveraging the new *IBM Information Archive* purpose-built storage system (see Exhibit 1, on the next page) or as a traditional “on-premise” software implementation owned by the customer.

As an alternative, IBM’s Global Services-hosted SaaS (Software as a Service) offering offers utility pricing and an alternative to on-premises deployment. All these alternatives are built on the same IBM 2.0 Reference Architecture. All deployments can leverage tape, which can reduce the TCO.

IBM’s strategy offers *Policy Harmony* (software to coordinate and reconcile archive policies) and *Data Governance* (software to manage information over the lifecycle of its retention based on business value).

Taken together with the new and existing information management and analytics applications, IBM Smart Archive offers customers a full selection of optional elements. Customers know that all elements of the solution will work well

Exhibit 1 —**IBM Information Archive Provides Next-Generation Retention System**

The new IBM Information Archive (the key enabler for smart archiving) is a flexible and secure storage system. It is a single platform for archiving compliant, non-compliant, structured, and unstructured data – thus reducing the need for multiple systems. It is quite flexible and customizable, due to its internal “collections” architecture. Each collection is a logical container and can be leveraged by different data value or usage). The system can scale from 16 to 304 TB raw (disk only) or higher, if tape is deployed. (This capacity will double in the near future with use of denser drives.)

The IBM Information Archive is offered with three levels of data protection, each of which is appropriate for certain kinds of content.

- The most basic level of protection enables the greatest flexibility for managing an organization’s information retention needs (i.e., documents can be deleted before they expire and retention periods can be increased and decreased).
- The intermediate level of data protection allows IT administrators to increase and decrease retention periods as needed, but information deletion is only allowed after the retention period has expired.
- The third option offers a maximum level of protection where archived files cannot be deleted or removed prior to its expiration. This is also the “compliance” mode that makes the system WORM (Write Once Read Many times). In addition, the enhanced *Tamper Protection* feature locks the administrator root login, providing a time-limited password for root access – for further protection and a clear audit trail.

Source: IBM

together due to the common architecture and the common code base.

Tailoring Deployment Mode to Need

The support for a common architecture across of full spectrum of deployment options strengthens the allure of the IBM Smart Archive strategy. How do you organize a mass of disparate data sources depends on the organization and what it is trying to do. For some kinds of data, centralized management is best. For others, their bulk and usage may dictate that organizing

them where they lie is best. Other data may be best routed to SaaS or cloud. **How you do smart archiving is all a matter of latency, security, and economics.**

- For the first time at enterprise-class archiving, well-targeted and predictably priced professional services (something IBM has developed over the past few years) can assure timely success of projects and reduce the risk of sophisticated initiatives.
- For companies with sensitive information and many low-latency requirements, on-premises, secure purpose-built systems make the most sense.
- For organizations that are more distributed, software as a service may be the way to go (like *salesforce.com*). This approach saves CAPEX and reduces day-to-day administrative tasks
- Cloud offers still better economics and less administration.

A breadth of deployment options has become a new prerequisite. Having adjacent competencies like content management paper capture etc is also useful in most cases. If compliance is an issue, it should be built into the solution because built-in scales better than tacked-on. If you buy for openness and extensibility, you can postpone licenses for functionality until you need them.

Conclusion

A new, expanding data landscape widens the spectrum of choices that apply to optimized data use. It also broadens the concept of an institutional digital archive. *Best of breed* is now a situational concept better articulated as *best fit*. This is determined by the demands of data consumers (be they applications or people) constrained by economic realities and shaped by the tools and enhancements that make the whole a sociable system – one that can serve a variety of users as their needs change. IBM offers a range of capabilities in deployments and in infrastructure and software tools and enhancements that allow organizations to build the archives they need at a price they can afford.



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 a Senior Contributing Analyst 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 in 2001 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.

Disclosures

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.

After publication of a bulletin on *clipper.com*, The Clipper Group offers all vendors and users the opportunity to license its publications for a fee, since linking to Clipper's web pages, posting of Clipper documents on other's websites, and printing of hard-copy reprints is not allowed without payment of related fee(s). Less than half of our publications are licensed in this way. In addition, analysts regularly receive briefings from many vendors. Occasionally, Clipper analysts' travel and/or lodging expenses and/or conference fees have been subsidized by a vendor, in order to participate in briefings. The Clipper Group does not charge any professional fees to participate in these information-gathering events. In addition, some vendors sometime provide binders, USB drives containing presentations, and other conference-related paraphernalia to Clipper's analysts.

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.