

# Big Data: Best Practices Guide

01

Introduction | 1

02

Increasing Demands of the Information Driven  
Enterprise | 2

03

Setting Your Information Management Goals | 5

04

Guide to Best Practices for Information  
Management | 8

05

Serrala Data and Document Management Solutions | 11

# 01

## Introduction

**Businesses are experiencing an exponential growth of data in today's digitized world and companies that know how to manage this valuable asset will be better prepared to innovate and compete in the years to come. Managing the continuous influx of data and documents ("information") in today's digitized business, however, is complex. Organizations need to control the storage of and access to information from the moment of its creation throughout the information lifecycle, in accordance with policies and regulations. Companies that choose to ignore the importance of managing information are at risk of data breaches or hacks which can be costly, due to a loss of reputation, loss of business, or possibly fines. Even if a company has an information management strategy in place, it can be difficult to ensure the correct processes are being followed consistently. With limited IT staff and resources, many companies are simply trying to balance the need for an information management strategy against other critical IT projects.**

This white paper will provide data/information managers with a set of best practices that they can follow to ensure they have a future-proof and risk-proof data and document management strategy in place. It will discuss the importance of automating information management best practices to ensure compliance with policies and regulations; the need for cost effective storage platforms to reduce the cost of retaining information for lengthy retention periods; and how to provide users with seamless yet secure access to historical information for reporting, audit and other requirements. With these best practices in place, CIOs and IT teams will be able to support their organization's rapid data growth, ensure secure and compliant information retention, and improve the ability to access information for reporting and audits.

# 02

## Increasing Demands of the Information Driven Enterprise

Businesses have always collected information about their customers, suppliers, and employees, but the rapid digitalization of business is rapidly increasing the amount and type of information collected. According to IBM, 90% of all data has been created in the last 2 years and Hitachi Data Systems estimates that the data it manages is growing at a rate of 40 - 60% each year.

The incredible success of companies such as Google, Amazon, and Facebook illustrate how powerful and valuable data can be for a business. Every interaction an individual has with these companies creates a digital footprint, which can provide valuable insights about current and future behavior. The Internet of Things is also turning the things we use - thermostat, car, or commercial airplane - into data generators. These things collecting real time data in order to adjust conditions (temperature, driving speed, or direction) and schedule maintenance.

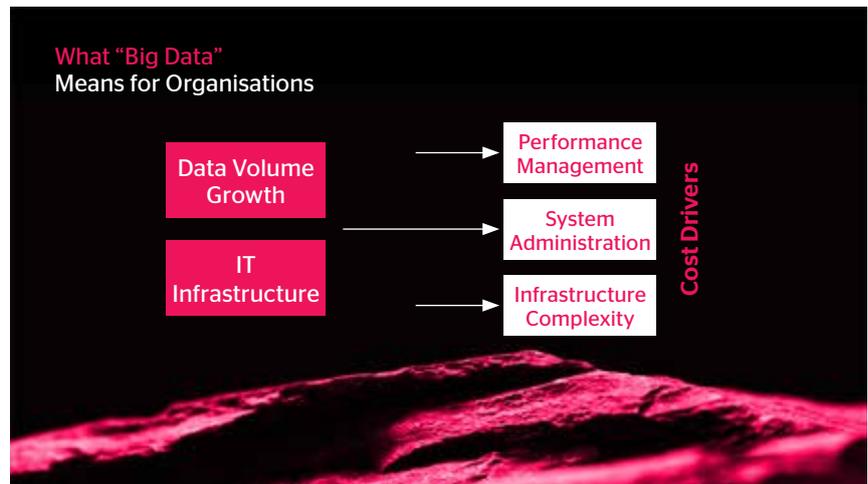
The value of all of this information is turning data into "the oil of the digital era", an asset that will fuel the success of businesses by providing greater insights and enabling new business opportunities. It will also provide the fuel for the machine learning and artificial intelligence capabilities that businesses will be implementing now or in the future. However, to capitalize on this increasingly valuable asset, companies need to understand the risks this asset can pose to their enterprise systems and the responsibilities they have for retaining data. Therefore, there must be a future-proof plan for managing the information the company collects and retains.

### **Increasing Demands on Enterprise Systems**

The rapid increase in data volume experienced by companies today, can have a significant impact on the cost and the performance of business systems. Many of these systems were implemented several years ago and do not have the capacity to handle the 40-60% year over year increase in data volume. For these systems, larger than expected data volumes mean that reports take longer to run and transactions take longer to complete. The larger data volume also requires more memory and storage and even in some cases, an investment in specialized systems that are capable of processing extremely large volumes of data. As the volume of data increases, therefore, the enterprise systems increase in cost and complexity and require more time to administer, which increases the overall cost of information for the business.

To control IT costs and reduce IT complexity, organizations need a data volume management strategy, that will enable the business to align the cost of the data with its value to the business. There is no need to invest in specialized systems, such as large servers or in-memory systems, if the volume of data is not of great value to the business.

# 02



However, there might be plenty of information which is considered “business complete”, and thus will not be changed anymore. This is important for reporting and analysis purposes, as well as for information requests such as (tax) audits, and needs to be kept for a reasonable time period due to legal requirements such as tax audits. In fact, in a survey of its enterprise software customers, SAP found that 50% and more of all data in SAP systems was two years old or older. As this mature data typically is fix and thus not changed anymore, it can be removed from the initial SAP database system and be archived to more cost efficient storage location where it will still be available for display access and analysis/reporting needs.

This data archiving process is a well proven data management technique and should be an initial and central component of any information management plan practicing good “data hygiene” Companies that put an archiving strategy in place can archive data and documents based on business needs and policies and store it in more cost effective archive repositories. Data Archiving is thus a method for retaining data for long periods of time outside the productive database environment. It reduces the volume of data in enterprise systems so companies can maximize the use of their existing IT resources, avoid the need to invest in new systems, and reduce the cost and complexity associated with managing large volumes of data.

## **The Risks and Responsibility of Information Stewardship**

The risks associated with retaining information is another issue that arises when companies collect large volumes of data about their customers, suppliers or employees. Companies have an ethical and legal responsibility to ensure that any information is retained securely while it is under a company’s stewardship. Recent news stories about hackers and data leaks at major corporations and government agencies illustrate how risky it can be for companies to retain information such as credit card data or birthdates and other personally identifiable data. While some of these leaks have been caused by external factors, many are caused by employees who unknowingly or intentionally leak information that they should not.

# 02

The risk of leaked data to a company is greater than simply a decrease in the corporate reputation or a temporary loss in business. Many data privacy and security regulations exist which make companies legally responsible for the data under their stewardship. Failure to meet some of these regulations (listed below), can put companies at risk of legal action and hefty fines:

- GDPR: General Data Protection Regulation (EU)
- HIPAA: Health Information Protection Act (USA HIPAA)
- PII: Personally Identifiable Information
- PIPEDA: Personal Information Protection & Electronic Documents Act
- PHIA: Personal Health Information Act
- PCI DSS: Payment Card Industry Data Security

Companies, therefore, need to ensure that they have a plan in place to mitigate this risk and protect the data that is under their care. In general, a strong data retention plan should include:

1. A corporate data retention policy that addresses the various business requirements and data security and privacy regulations that apply to the business, reviewed by the Legal, Risk, IT and Business teams.
2. A secure IT infrastructure that ensures that information is safe from both external threats (e.g., firewalls) and internal threats (e.g., access controls).
3. A method or process for securing the data and documents under the company's control according to the policies no matter where it resides: production systems, archive repositories, content management systems.

Many companies assume that setting up a secure IT infrastructure and access controls (step 2 above) are sufficient to ensure the security of the data they hold. However, as proven by the countless news stories of data breaches, these measures are often not enough. This is because information can reside in many different systems, not all of which have the same level of security measures. The constant evolution of data management policies also makes securing data according to regulations difficult.

For example, companies might have put a process in place to protect customer credit card data in production systems; however this data might also exist in other systems that are not as tightly controlled (i.e., system copies that exist (for example for QA), in BW systems, or archive repositories).

Practicing good information management is the best way for companies to reduce the risk of data leaks. Regularly archiving information according to polices and regulations reduces the amount of information available in production and adjacent systems. It moves all "business complete" data and documents from online systems into a single archive repository, where it is more secure.

# 02

An additional level of fraud resistance is therefore added. The increased security is the result of the following conditions specific to archived data:

- It is frozen and not editable, making it harder to tamper with
- It is extracted from all online systems and moved to a central repository, reducing the number of locations where is stored
- If it is no longer needed, it can be purged

Users often resist archiving data because they are concerned the information will be lost or deleted, or that they will no longer have access to it; however, archiving is not the same as purging data: Archiving simply freezes the data and moves it to another location, increasing its security while retaining accessibility to the archived data.

Archiving also provides a defined point in time at which companies can purge data if it is at the end of its useful life and has met its legal compliance requirements – an important step which many companies fail to do.

# 03

## Setting Your Information Management Goals

Knowing the risks associated with retaining large volumes of information, companies must have a clear understanding of what information they have in their online systems and why they need it.

- Do they want to be able to search through a set of transactions to troubleshoot specific issues?
- Perform a historical analysis of a business unit or predict future trends or outcomes?
- Analyze very large data sets to support better business decisions?

Questions like these will help define how to manage the information life cycle to meet specific business needs.

### Understanding the Information Lifecycle

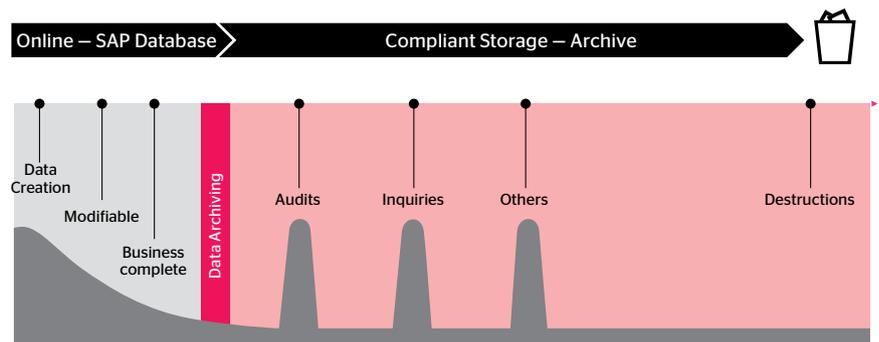
In general, organizations generate or collect data as part of a business transaction and should only retain data if it serves a legitimate business purpose. A typical information lifecycle pattern is shown in the figure below.

**Data Residence** is the period of time that business data objects need to remain accessible in the online database environment. Data passes through phases and only becomes eligible for archiving once it is **business complete**. At this point, data is not able to be modified anymore. There can be a varying duration between this unmodifiable business complete state and the time when objects can be archived. The overall “residence time” is based on business requirements including how frequently users access the data.

# 03

After data is archived, it is still accessible from archive storage, however, users access that data less frequently. Typically archived data is access for audits or other inquiries. When an object meets its **Retention Time** it finally becomes eligible for purging. At this point, data has met all legal compliance and other corporate retention requirements.

## Typical Lifecycle of Data and Documents



## 5 Things You Need to Know Before Putting an Information Management Strategy in Place

Once you understand the information lifecycle and the reasons why your business want to retain certain information, you will need to ensure you have the proper infrastructure in place to meet your data retention goals.

IT teams should evaluate the organization's level of readiness in the following five areas:

### 1. Archive and Document Management Infrastructure:

- Do you have an archive and document management infrastructure in place?
- How old is that infrastructure and does it need to be updated to meet current and future data retention requirements, particularly regarding privacy and security?
- Is the current infrastructure cost effective?
- Is it possible to replace components of this infrastructure with more modern, cloud-based components that would reduce costs?
- Can the current infrastructure scale to meet future business requirements?
- How easy is it for users to access archived data? Do users need better access to archived data?

# 03

## **2. Data Volume Management:**

- Are there any issues with the existing data volume and its growth rate?
- Are your users experiencing performance issues working with this data?
- What will happen if you change your underlying systems or business processes - will data volume be affected?

## **3. Data Privacy:**

- What controls do you have in place to meet privacy regulations such as HIPPA, PIPEDA, or GDPR?
- Are you able to identify, extract and purge data according to these regulations? How long does that take and how many people are involved in the process?

## **4. Legacy Decommissioning:**

- Do you have any old or obsolete systems (non-productive) that are only used for reporting and no longer used for day-to-day business transactions? Can these systems be retired?
- Are such systems obsolete or no longer needed due to a platform migration such as the move to SAP S/4HANA?
- What information needs to be retained from these systems for legal or other corporate requirements?

## **5. Audit Readiness:**

- What are your corporate audit requirements?
- Are you able to ensure exception case management of data and documents for legal case management easily?
- Are you able to perform country specific tax reporting?
- Do you need to implement solutions that make the extraction, validation, and reporting of audit data easier for your staff?

Once you know the answers to these questions, you can decide how you want to proceed to meet your information management goals.

## **3 Information Management Goals and How to Achieve Them**

Ultimately, your goal in preparing your information management strategy should be to lower costs, reduce risk, and increase transparency. However, each company has its own set of information management priorities and concerns and will need to tailor their plan accordingly. Some ways to achieve these goals are listed below.

# 03

## **Lower Costs:**

- Reduce the cost of retaining data by archiving information and moving it to lower cost archive repositories
- Automate data archiving and retention processes or leverage managed services to do so. This will ensure data is consistently archived according to policies and regulations without requiring extra resources
- Retire legacy systems to eliminate unnecessary support and maintenance costs
- Provide data management tools to speed up the processes of extracting data for audit reporting and other requirements

## **Reduce Risk**

- Move information out of production systems to more cost effective and immutable archive storage.
- Decommission systems that are no longer active or that are beyond their recommended life span and that may be susceptible to security breaches.
- Implement tools to support new privacy regulations and that can protect data necessary for legal holds, if necessary

## **Increase Transparency**

- Move all archived information to a central repository to improve access and security
- Provide data access tools that make it faster and easier for users to access archived data

# 04

## Best Practices Guide for Information Management

Whatever your organization's information management goals are, CIOs and data/IT departments should consider using the following best practices to ensure your organization is able to retain information safely and securely over its lifecycle.

### **Archive Information as soon as Possible after it is Business Complete**

Archiving information is the best way to ensure efficient, cost effective long-term retention of information. To achieve the best results, automate the archiving process so that archiving is done consistently and without error.

Adjust the data residency period (the period data is active in your production system) according to the data in question. Some technical data can be archived almost immediately, whereas, other business transactions such as information on a product purchase will need to be retained longer.

# 04

Due to the complex relationships that can exist within and between business objects (for example a complex customer order involves quotes, contracts, delivery information, billing and more), it is a good idea to consult data management experts to determine the best archiving approach for your company. You'll achieve better results in less time. Using software to automate the process will also significantly increase the speed and accuracy of the archiving process.

## **Use Cost Effective Archive Storage for Archived Information**

To ensure maximum performance, your production systems use the highest performing, expensive storage in your IT landscape. Archived information is not accessed as frequently as production data, therefore, it is possible to use more cost effective storage options for your archive repository.

Cloud storage is ideally suited for archived data because it is low cost and infinitely scalable. However, there can be latency issues when accessing data in the cloud, which can make it better suited for storing data that is not accessed frequently and that needs to be retained for long periods of time.

## **Modernize Content Management Solutions**

Many organizations maintain a separate content management repository for documents and other unstructured information. When reviewing your information management requirements, ensure that the content management solution you are using is capable of meeting your current and future information management needs. Many content management systems were implemented years ago and do not have the same features and flexibility of newer cloud-based solutions that provide a better solution at a more reasonable price.

## **Decommission Systems that Are No Longer Viable**

One of the biggest risks many organizations have is the older systems they keep in operation only for reporting purposes. These systems are often vulnerable to hacking because they do not meet modern standards for data privacy and security. They are also costly to operate as companies continue to pay maintenance fees, pay for the staff to support the systems and for the space to store them.

Retiring these legacy systems provides double the value for companies because it can both save money and reduce risk. With an information management structure in place, the critical business information that resides in these systems can be archived and moved to a central archive repository, thus ensuring that the information is stored securely and accessible to meet long term audit requirements.

# 04

## **Automate Information Retention to Ensure Compliance**

Many companies continue to keep data well beyond its useful life and in some cases beyond the time they are legally allowed to retain it. To ensure compliance with policies and regulations, automate the application of retention rules used to enforce retention policies. An automated process will ensure data is automatically reviewed when it reaches end of life and finally purged according to policy. Automated policy enforcement ensures that current staff will manage the retention task which may have been defined by a previous and completely different set of staff members.

## **Mask or Encrypt Data When Necessary**

In some cases, particularly when there is a merger or separation of company divisions, ensuring the privacy and security of data can be very complex and cannot be managed using access controls alone. In this case, consider using data masking or encryption to protect data while it is resident in online systems. Data should be protected at the database level to ensure there are no “back door” ways to access the information, for example, through database administrator tools.

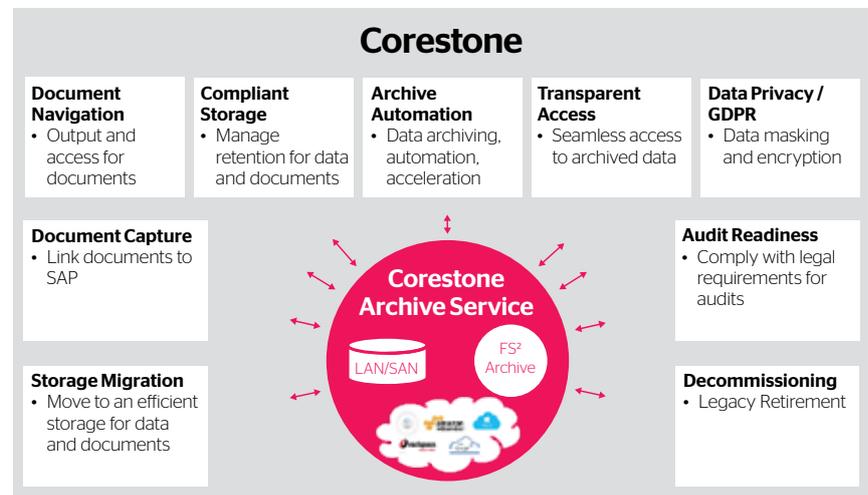
## **Improve User Access to Archived Information for Reporting, Audits and Other Queries**

Two of the biggest challenges of managing large amounts of information are 1) accessing archived information quickly for reporting and other business purposes and 2) extracting the information you need to respond to audit inquiries, customer disputes, or other requests. Business users often have to struggle to access archived data using separate tools and processes. If they need to extract that data, it can take hours, days or even weeks of time working with the IT department to identify, extract and prepare the data required.

Many modern solutions exist to enable users to quickly and accurately report on archived information or to pull archived data from enterprise systems according to specific business requirements such as tax reporting, privacy audits, and more. Consider the cost of these solutions against the time and effort your staff use to perform these tasks and implement software solutions to make this process faster and easier.

## Serrala Data and Document Management Solutions

### Features: Corestone Solution Components for Data and Document Management



Serrala provides a comprehensive set of software solutions and managed services to control information across the entire information lifecycle. Our solutions include:

- **Archive Service:** foundational component to ensure the consistent archiving of information to reduce the volume of data and documents resident in production systems. It enables organizations to leverage a more flexible and broader range of archive storage options, including public Cloud-based storage.
- **Retention Management:** manage and automate the retention of data and documents according to retention rules and legal holds, and ensure that data is purged when it reaches end of life
- **Legacy Decommissioning:** archive data and documents from legacy systems to a central archive repository to enable system decommissioning
- **Data Privacy:** ensure the privacy of data is maintained with data masking and encryption solutions
- **Archive Automation:** create archiving variants, schedules and monitors archiving tasks, ensuring archiving is done according to information management best practices, consistently and with the highest degree of accuracy. It furthermore offers to benchmark organizations against their peers, suggesting improvements to the archiving approach, to improve system performance or reduce risk, if necessary
- **Transparent Access:** provides users seamless and complete access from their SAP user interface. This allows them to efficiently find and use the relevant archived data and documents they need for their daily work without the need to move between application and thus achieve faster reporting.

# 05

- Audit Readiness: extract large volumes of data and documents from online systems and archive repositories in audit-ready formats to meet global standard (SOX, SAF-T) and ad hoc audit requirements
- Document Capture: automatically capture information from unstructured documents (paper, emails, PDFs, etc.), archive it to enable faster, more transparent access to digitized supporting documents, and link it to respective SAP table entries
- Document Navigation: enable users to find and extract documents quickly to respond to audit and other requirements.
- Storage Migration: move data from existing archive repositories or content management systems to a centralized repository for more efficient access to archived information

For more information on how Serrala can help your organization develop an information management strategy, visit us at: [www.serrala.com](http://www.serrala.com).

# Serrala Brings Clarity to Complexity

We are proud to announce our evolution from Hanse Orga Group and welcome you to Serrala. Our new name is inspired by the Sierra de la Serrella, a powerful mountain range in Spain symbolizing confidence and security. With a suite of future-proof end-to-end payment solutions we provide efficient and secure optimization of your financial processes partnered with experienced first class cash visibility and treasury consulting services. Driven by our enterprising nature, we constantly challenge ourselves to bring relevant innovation to market. Serrala supports over 2,500 companies worldwide with advanced technology and personalized consulting to optimize all processes that manage incoming and outgoing payments: from order-to-cash, procure-to-pay and treasury to data and document management. Serrala's software family of best-in-class solutions creates a complete payments ecosystem that provides security and compliance at every step of your financial value chain. Our solutions offer different deployment options to meet your specific requirements – on premise, in the cloud or as a hybrid solution with managed service components. Over 30 years of industry experience provide us with the necessary knowledge and understanding to be your trustworthy partners.



## **Head Office**

Serrala Group GmbH  
Oldesloer Straße 63  
22457 Hamburg  
Germany  
Tel: +49 40 514808-0

## **Contact APAC**

Serrala Hong Kong Ltd.  
Level 9, Central Building  
1-3 Pedder Street, Central  
Hong Kong  
Tel: +852 3958 2826

## **Contact EMEA**

Serrala EMEA B.V.  
Nieuwe Prinsenkade 9  
4811 VC Breda  
Netherlands  
Tel: +31 (0)76 578 05 00

## **Contact North America**

Serrala US Corp.  
205 N. Michigan Avenue / Suite 4110  
Chicago, Illinois 60601  
USA  
Tel: +1 (888) 305 9033

**[serrala.com](http://serrala.com)**