opentext | InfoArchive

eBOOK

Modernize information archiving and simplify data compliance

How consolidation of legacy data and content opens new opportunities for digital transformation and long-term agility

opentext™

Content

opentext™

Compliance risks for business-complete data

Over decades, organizations accumulate terabytes (or petabytes) of **business-complete application data maintained** on various IT systems. Aging data affects application **scalability**, data **relevance** and **costs** and represents a mounting risk of **non-compliance**. Host applications must be kept running, licensed and updated. Servers must be powered, protected, backed up and monitored. And when mergers, acquisitions and reorganizations occur, all of this must be refactored with new processes and security requirements. The expense is formidable to **retain data** that is rarely accessed.

There is a better way...

Up to **80%** of IT **spending** may go towards operations and maintenance of **legacy systems**

US Government Accountability Office (2019, GAO-19-471)

opentext™
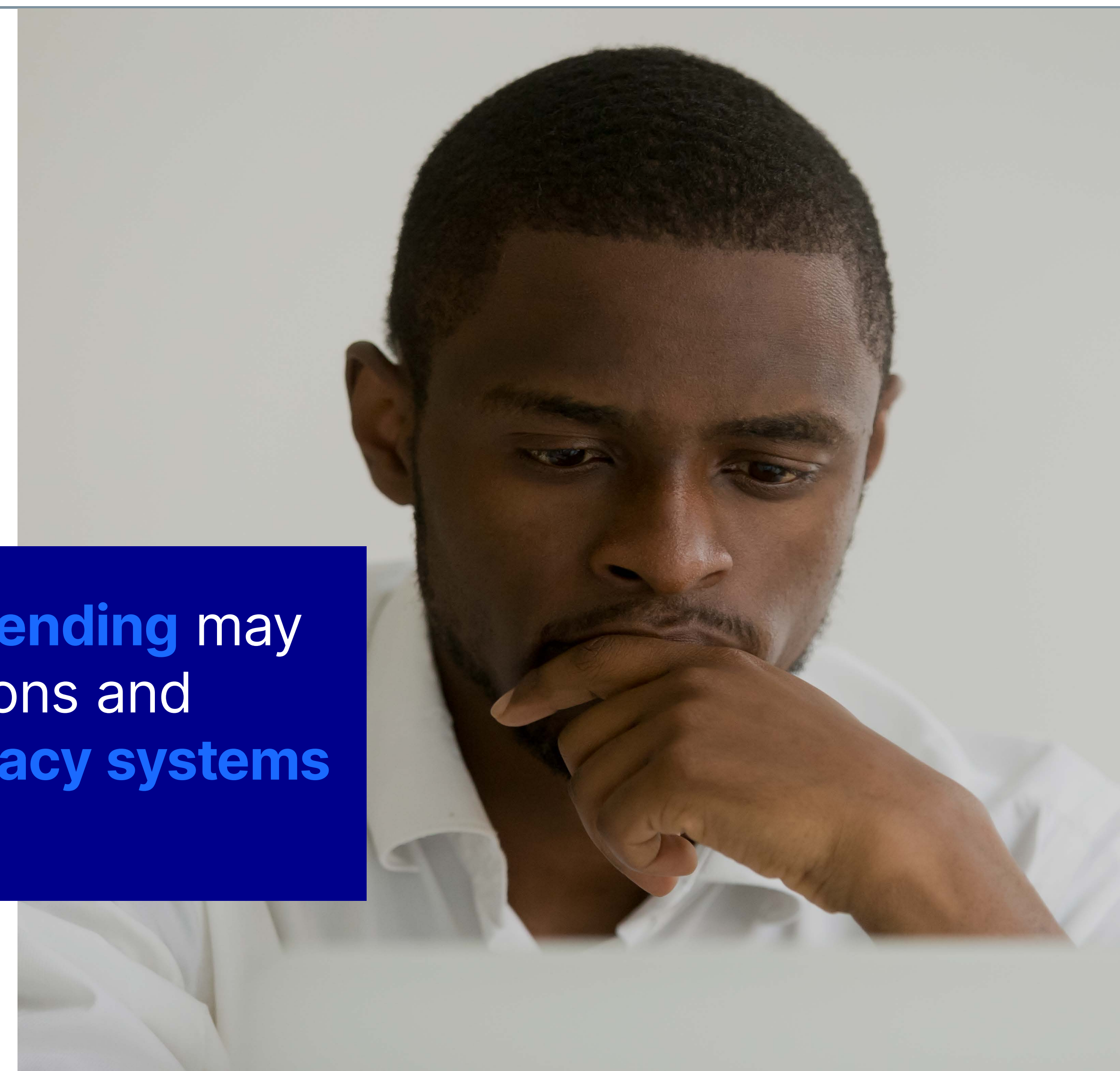
The vast landscape of data types requires a new, long-term approach

Long-term IT and budget planning is daunting when confronted with dozens of individual applications, billions of records and terabytes of historical data.

Complexity is the enemy.

The structured data of a running financial, ERP or HR application, along with document attachments, define critical business context that must be preserved to maintain compliance. The data must also remain accessible to users who may need it in their daily work, reporting or data mining, but is costly in the form of a live application.

Simplifying such a large variety of data and applications cost-effectively requires more than just careful planning. It requires a whole new approach.

Retaining essential business context that includes both relational data and content is more economical and defensible in the long term. It also better supports a move to the cloud, now or in the future

Most business transactions are a combination of structured data and content

Structured, relational data, databases

LOB APP

COBOL

Legacy application

Mainframe application

+

Unstructured content, files and metadata

DOC

CAD

PDF

Documents

Drawings

Images

OR

High-volume customer communications

Print streams

opentext™

Compliance is a moving target in the long run

Keeping data compliant with various regulations and business standard requirements is an ongoing task and can change frequently. Such changes strain IT departments to keep environments performant, secure, compliant and up to date, especially when aging, business-complete information is in play.

"By the end of **2023**, modern privacy laws will cover the personal information of **75%** of the world's population."

Gartner (2021) "The Top 8 Cybersecurity Predictions for 2021-2022"

California Consumer Privacy Act (CCPA) 2.0

Effective January 1, 2023

New **"Sensitive Personal Information"** category protects and allows consumer restriction of:

- SSN, driver's license number, financial info
- location, race, ethnicity, religion
- union membership
- personal communications
- genetic, biometric or health information
- sex life and sexual orientation

Compliance requirements that must be reviewed and adapted regularly include:

- Records retention requirements, which are often based on key business events, such as contract termination or patient data, that affect both the period of retention and the immutability of the business record.

- Privacy regulations, such as GDPR or CCPA, which can affect access control, data visibility, retention and deletion policies.

- Security best practices, which must be responsive to emerging threats, a task that is all but impossible with an aging application at the end of its lifecycle.

- Legal hold requirements, which are unpredictable, making collections and discovery difficult without sufficient control to pause retention periods and extract the right data in the proper format.

Responding to these and other external requirements quickly—and without continually upgrading data schemas on billions of records—is a critical element of maintaining compliance in a stable environment.

opentext™

Archiving imperatives

An organization must reduce complexity and remove as many components as possible while still carrying out its core mission and maintaining compliance.

At a minimum, it will need an archive that:

- Combines active archiving and legacy data into a single infrastructure for the best cost savings.

- Preserves the original context and fidelity of the archived data to assure defensible preservation.

- Incorporates archived data securely into daily processes so that it is available within daily work and doesn't otherwise hinder productivity.

- Creates automated systems of information governance that can adapt to future compliance requirements.

Information archiving must seamlessly combine both **structured and unstructured data** into discrete records

Source application → Analysis, extraction and ingestion →

Logical records

Master Record

Data Records

Document attachments

DOC CAD PDF

Enhanced metadata with key values

The result must provide **secure data access** and faithfully support **productivity and compliance**

The Compliant Archive

Data accessibility through search and view interface

APP Data portals, dashboards and integrated applications

Searchable index

Compliance with records, privacy and holds

Encrypted, immutable storage

opentext™

The relevance and value of archived data

The value of any data in an organization will change over its lifecycle. The tendency is to believe that once the data is no longer part of an active business cycle or transaction, it can be archived and forgotten; this is a fallacy. Users may need access to data long after a contract terminates, a project is completed or the bill is paid.

Examples of scenarios where users may require frequent access include customer service applications, field service reference and data mining.

IT leaders may fear decommissioning applications out of a concern that the data will become obscure and inoperable. They may fear that they won't maintain the same level of usability for customer service users or others who may need quick and straightforward access to the data.

Information archiving plans should consider these key capabilities:

by 2025

more than **80%** of organizations will use **accessibility** as a key criterion for digital workplace technology investments—up from 20% in 2021

massive generational shifts will force **75%** of organizations to adapt their **hybrid work** strategies to include demands for radical **flexibility**

Gartner (2021) "Predicts 2022: Digital Workplace is Foundational for Employee Experience"

Requirement	Benefits
Hierarchical data record display that helps visualize the complete record	Familiarity of the record display to the original application for usability
Secure, in-situ access to relevant historical data from customer service, financial or other applications	Makes it easy to reach archive data from current application, increasing productivity
Securing or masking sensitive personal information	Protects sensitive, high-risk data that could represent a data leak
Display of older or obscure file formats that may no longer be available, such as CAD drawings or spreadsheets	Offers broad access to historical data that would otherwise require expensive or rare desktop apps
Access to structured data through APIs, SQL or Python	Provides integration with customer-facing mobile apps or web portals

opentext™

Adaptability is essential to long-term information governance and compliance

Plans rarely survive implementation. When making plans for consolidating and preserving data, what made perfect sense years ago no longer meets the stringent requirements in place today. In-process data in live applications must be maintained according to the original application's technical requirements and limitations. But once that data reaches its natural completion—business-complete data—those requirements may no longer make sense for long-term preservation or access.

Your organization may wish to migrate to a centralized data center and infrastructure to reduce hardware overhead, monitoring and disaster-recovery requirements. Your organization may also be considering a move to cloud-based management and storage.

Security may also be a key concern as the original application may be designed around decades-old security practices that fall short of modern best practices. As a result, organizations must often deploy expensive and cumbersome remote access controls to protect an antiquated application. Moving to a modern architecture can free you of these external solutions.

Some other issues that may affect future planning include:

- Permanent, read-only preservation of regulated data.

- Improved auditing of data access and lifecycle.

- Accessibility of the data, particularly customer communications.

Accommodating all these future capabilities can be expensive and complex, especially if it is an afterthought.

By **2024**, the majority of **legacy applications** will receive some **modernization** investment

IDC. "IDC Futurescape: Worldwide Cloud 2022 Predictions" (2021) Doc# US47241821, October 2021

opentext™

Rapid adaptability and scalability in the cloud

When developing an information archiving plan for your organization, options with rapid scalability, low overhead and compliant storage cannot be ignored. Operating a well-contained, compliant infrastructure that meets all of the previously discussed requirements for information archiving is more feasible in the cloud.

When considering how data will be archived, organizations need the flexibility to deploy in the manner that best suits the long-term needs of the business, with the ability to change plans if conditions change.

Cloud deployment benefits

Data sovereignty
Comply with jurisdictional storage requirements

Hybrid flexibility
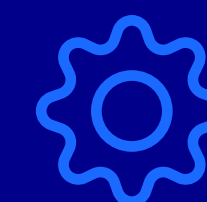Methodically move to the cloud one dataset at a time
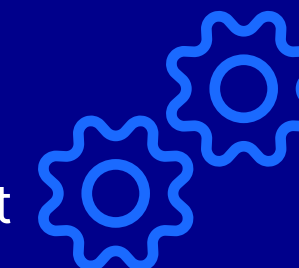
On-premises storage
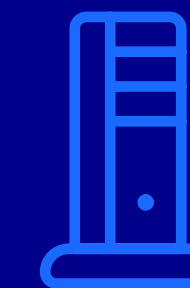
Cloud storage

Rapid provisioning
Scale up in a fraction of the time

Increase ingestion throughput

Scale storage or access new storage classes

opentext™

Complete information archiving efforts require the right tools and expertise

A large-scale information archival project may appear to be a daunting task. Beyond the sheer volume of data, the complexity and preservation of the original context, meaning and usability of the data is critical and affects the overall workload. That's why it's so important to have the right tools at your disposal and experienced professionals available for facilitation.
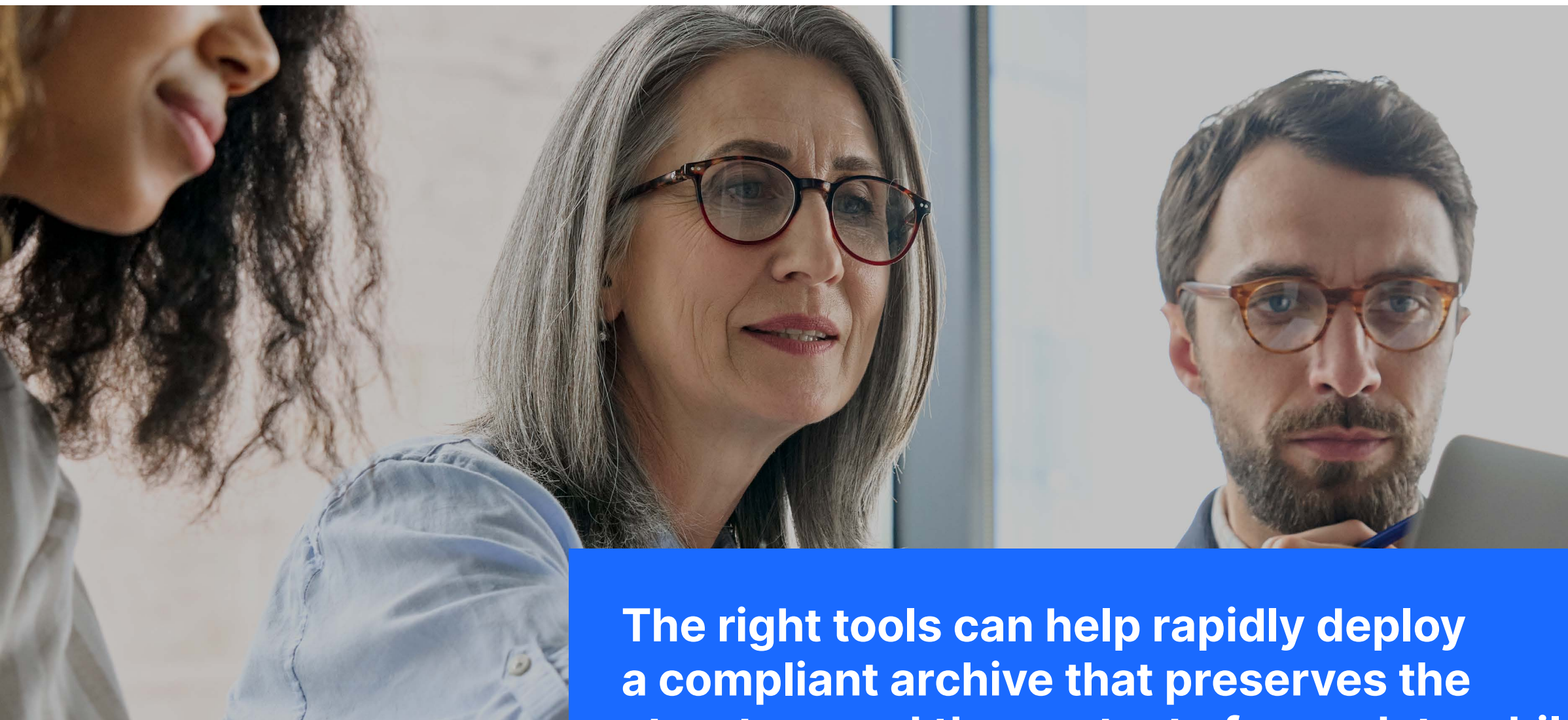
The right tools can help rapidly deploy a compliant archive that preserves the structure and the context of your data while maintaining familiarity for your users

Some aspects of migration that you may want to consider include:

- Faithfully representing the original data schema so that it can be searched or interpreted later.

- Masking or partitioning privacy data that is protected by privacy laws to avoid non-compliance and data leaks.

- Finding the optimal archival format to balance storage and cost limitations against compliance and archival fidelity requirements.

- Extraction of metadata, including key values, display and searchable identifying information and soft metadata from text analytics, such as summary text or keywords and entities in the content.

- Classifying archival records so that they can be effectively secured, searched and assigned records policy or retention during ingest.

- Retention requirements, including retention event metadata and record identification within complex formats such as AFP print streams.

- Encryption of content to comply with security requirements, such as HIPAA or data protection requirements.

opentext™

Revolutionizing archiving with OpenText™ InfoArchive Platform

Long-term planning for information archiving one application or dataset at a time may seem beneficial as each situation has unique requirements. However, when combined with the legacy data and the sum of the requirements for a robust information archiving infrastructure, an overly simplistic approach may not even meet basic compliance requirements.

OpenText InfoArchive is a complete platform for ongoing active archiving and legacy decommissioning at enterprise scale. InfoArchive combines enterprise-class information management and unstructured content services with sophisticated structured data archiving capabilities. The result is that users never lose access to required information to complete their work and the necessary controls are in place for a compliant and safe preservation store.

InfoArchive also provides a modern, cloud-first architecture designed to accommodate multiple deployment scenarios and adapt over time.

Archive requirements	InfoArchive benefits
Preservation of structured, relational data	• Preserves relational data in open formats • Full table archiving • Cloud and on-site storage integrations
Content services for unstructured data	• Complete, searchable metadata index • Full retention lifecycle management • Legal holds and export capabilities
Search and analytics	• Metadata and text search • Text analytics for enhanced metadata • API access for data mining support
Data and content viewing	• Presentation of complex business schemas • Available integrated web-based file viewer • Integrate into production LOB systems
Security and privacy controls	• Encrypted storage with customer keys • Privacy masks to protect personal information • Access control lists

opentext™

Context and usability keep archive data relevant

Typically, archives try to normalize all data into a uniform format for search and display. Unfortunately, this reformatting often means losing fidelity and making the data far less approachable. It also affects user productivity with tradeoffs in expense, preservation fidelity and adaptability. Given that organizations typically have dozens of legacy applications to resolve, it makes sense to have a rapid application development environment that searches and displays the data. This approach enables frequent process change without affecting platform architecture, security or scalability.

InfoArchive provides a built-in, configurable web interface to search and retrieve any metadata, archived data and view content in hundreds of formats. Within days, organizations can deploy a simple, familiar user interface to their archive, which meets most key user-retrieval requirements. Search and data display screens can be incorporated into existing applications. The built-in access control ensures that only authorized users view only the data relevant to their role in the organization.

InfoArchive also provides full REST APIs to include archived data and content into mobile applications, data portals or third-party applications.
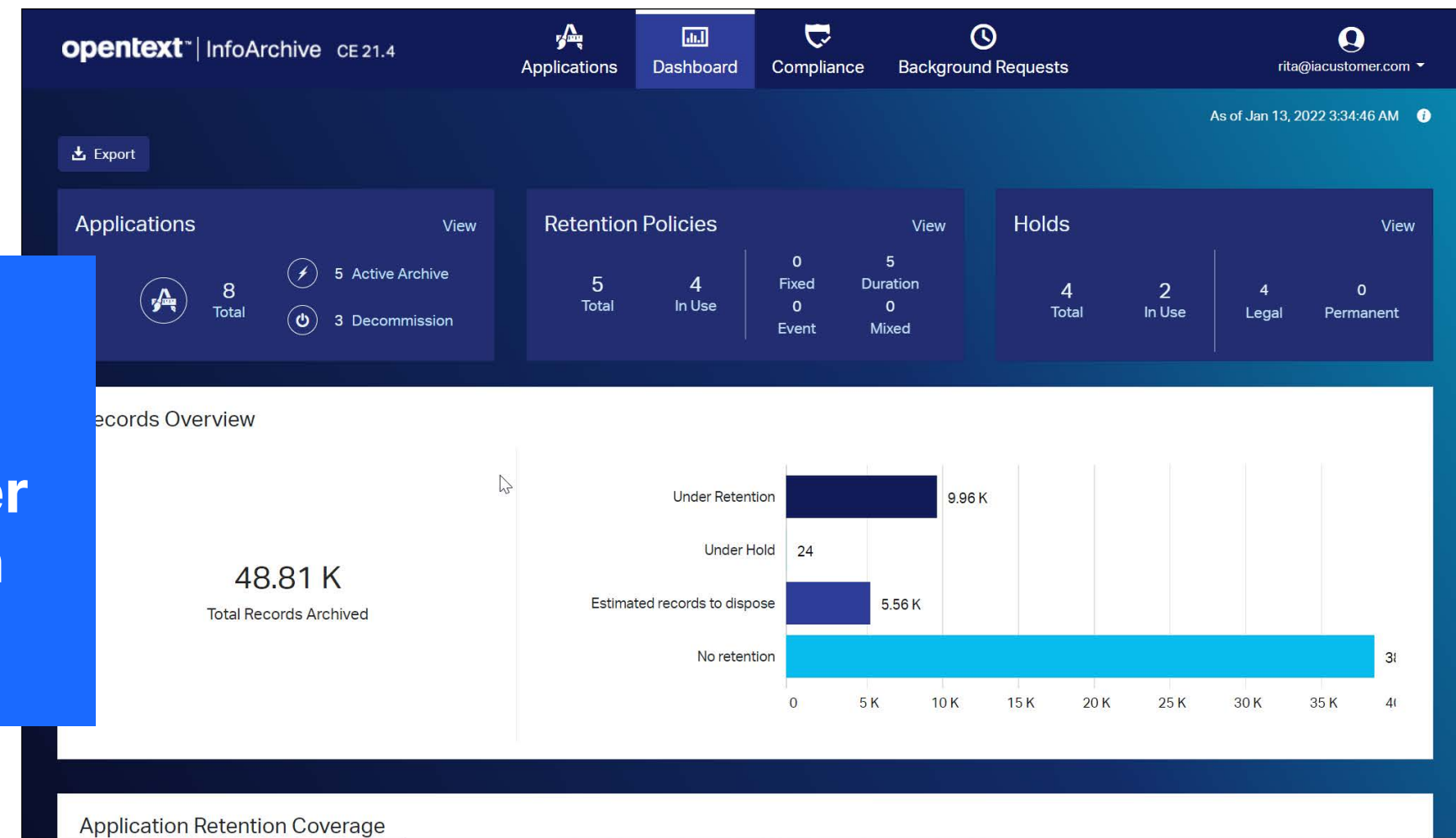

Configurable search forms help users navigate unfamiliar data

Search results are multi-faceted, allowing the user to drill into more detail or even view attachments

Critical governance, compliance and privacy controls

Organizations are compelled to maintain perhaps dozens of costly legacy applications beyond their useful lifecycle due to non-discretionary compliance requirements. And because of cumbersome methods to comply, IT is retaining data beyond the records policy requirements; they are flatly unable to identify and defensibly remove the data.

InfoArchive is a flagship product within the OpenText suite of information governance applications. As such, InfoArchive incorporates several essential compliance functions:

Privacy masking: Sensitive personal information remains secure during everyday use

Defensible governance: Content is under active retention management

Capability	Benefits
Records retention lifecycle	Event-based, multi-phase retention lifecycles
Legal hold	Hold policy and export content for eDiscovery
Privacy masks	Mask high-risk privacy data from leaks
Audit trails	Full auditability to document actions
Access control	Protect data by application and role

opentext™

## High-volume print streams or customer communications

When archiving print streams in AFP or PDF formats from large-scale applications and even mainframes, IT has typically had to weigh some unwelcome choices. Print stream content generated by these applications has historically been handled by specialty content archives that are now aging and expensive to operate. The content is challenging to extract and store without the right tools, and compliance can also be difficult.

OpenText provides a complete solution for handling high-volume customer communications and other print-stream applications:

- OpenText InfoArchive creates an efficient store and governance platform for compressed print streams in AFP or PDF formats.

- OpenText™ Exstream™ is a market-leading solution for omnichannel customer communication management (CCM).

- OpenText™ Output Transformation Services handles conversion to PDF or run-time presentation in PDF for AFP-based content.

Together, these provide a scalable, cloud-based upgrade for customers replacing systems like IBM® Content Manager OnDemand or Mobius from ASG Technologies®.
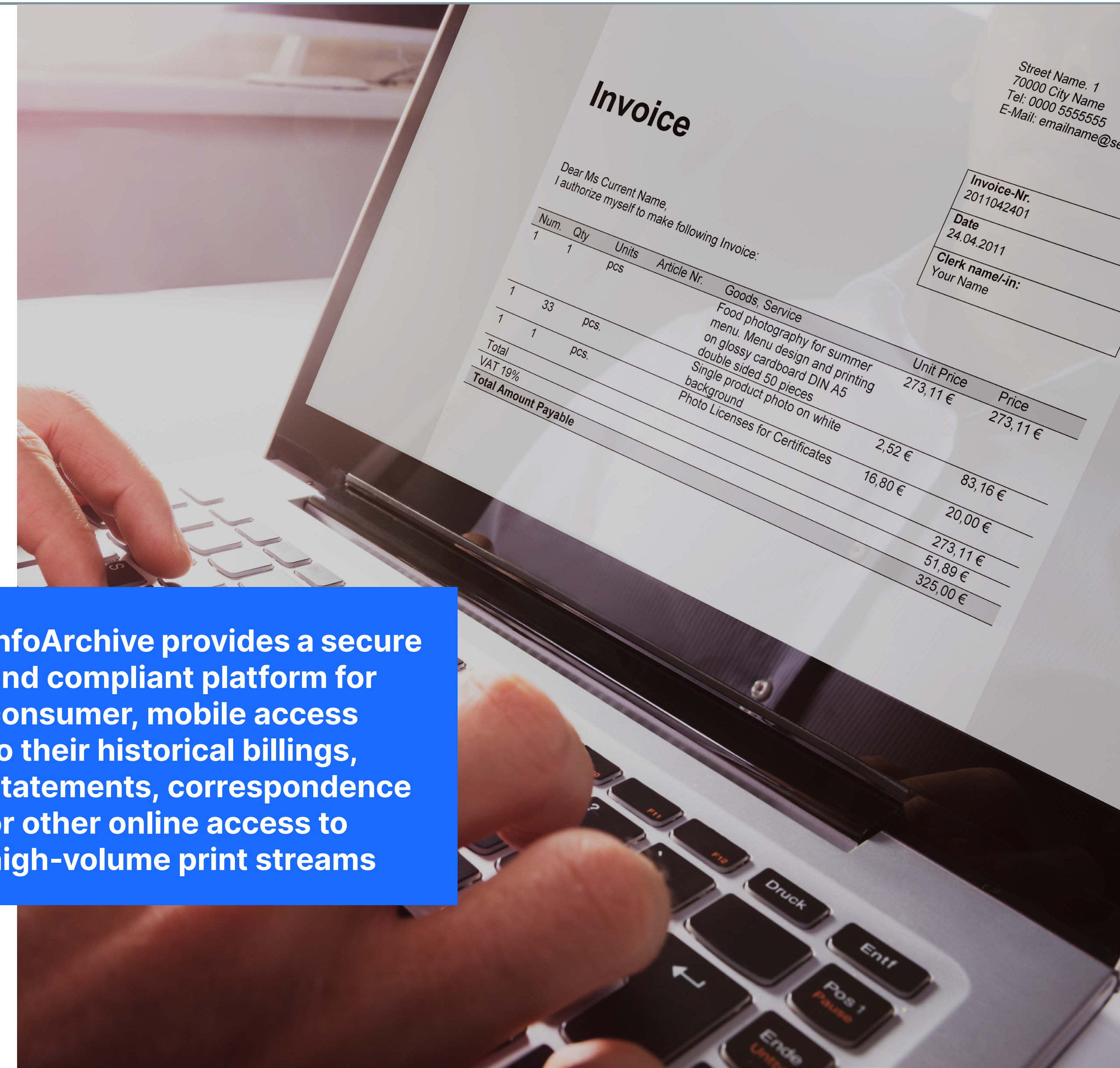
InfoArchive provides a secure and compliant platform for consumer, mobile access to their historical billings, statements, correspondence or other online access to high-volume print streams

opentext™

A flexible and scalable, cloud-first architecture

Effective long-term planning means selecting an archive that provides the flexibility you need for years to come. Many organizations have operational, compliance or data sovereignty requirements that call for the hosting of their data in country or on company-managed infrastructure. During a transition to the cloud, many organizations may require an approach that simultaneously supports cloud and on-premises storage options.

| Operational environment	Benefits
OpenText™ Content Cloud™	Fully managed archive on a subscription basis
Amazon Web Services	Hyperscaler deployment; S3 and Glacier support
Google Cloud	Hyperscaler deployment; Archive or Coldline
Microsoft® Azure®	Hyperscaler deployment; Azure Blob support
Red Hat® OpenShift®	Third-party cloud hosting running OpenShift
On-premises	Self-managed environment in your data center
Hybrid	On-premises with cloud storage for some data

Managed services from
opentext™
Content Cloud

...or deploy on the top hyperscalers

Google Cloud Platform   Azure   amazon web services

Containers facilitate simple deployments and updates
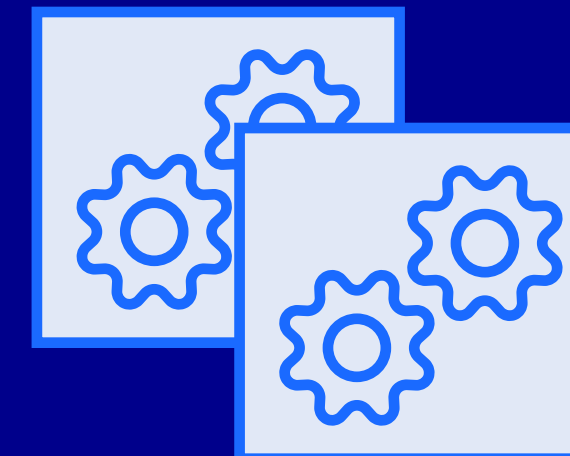
Scale your storage vertically or horizontally as needed

Native cloud   Cold storage   On-premises

opentext™

Unified information governance

As the volume of information and the complexity of regulatory and privacy requirements grows, so does the difficulty of managing a compliance program. Managing your governance program across multiple repositories introduces the risk of inconsistencies and non-compliance. Assuring that your content is governed according to policy and finding content out of compliance is critical to good governance in information management.

OpenText™ Core for Federated Compliance delivers cross-repository visibility and policy control for transparency and defensibility of organizational information governance processes and reporting. By centralizing policies, reporting and executing records events and holds, Core for Federated Compliance allows your organization to govern across InfoArchive, OpenText™ Documentum, OpenText™ Extended ECM and even Microsoft 365® repositories from a cloud-based portal.
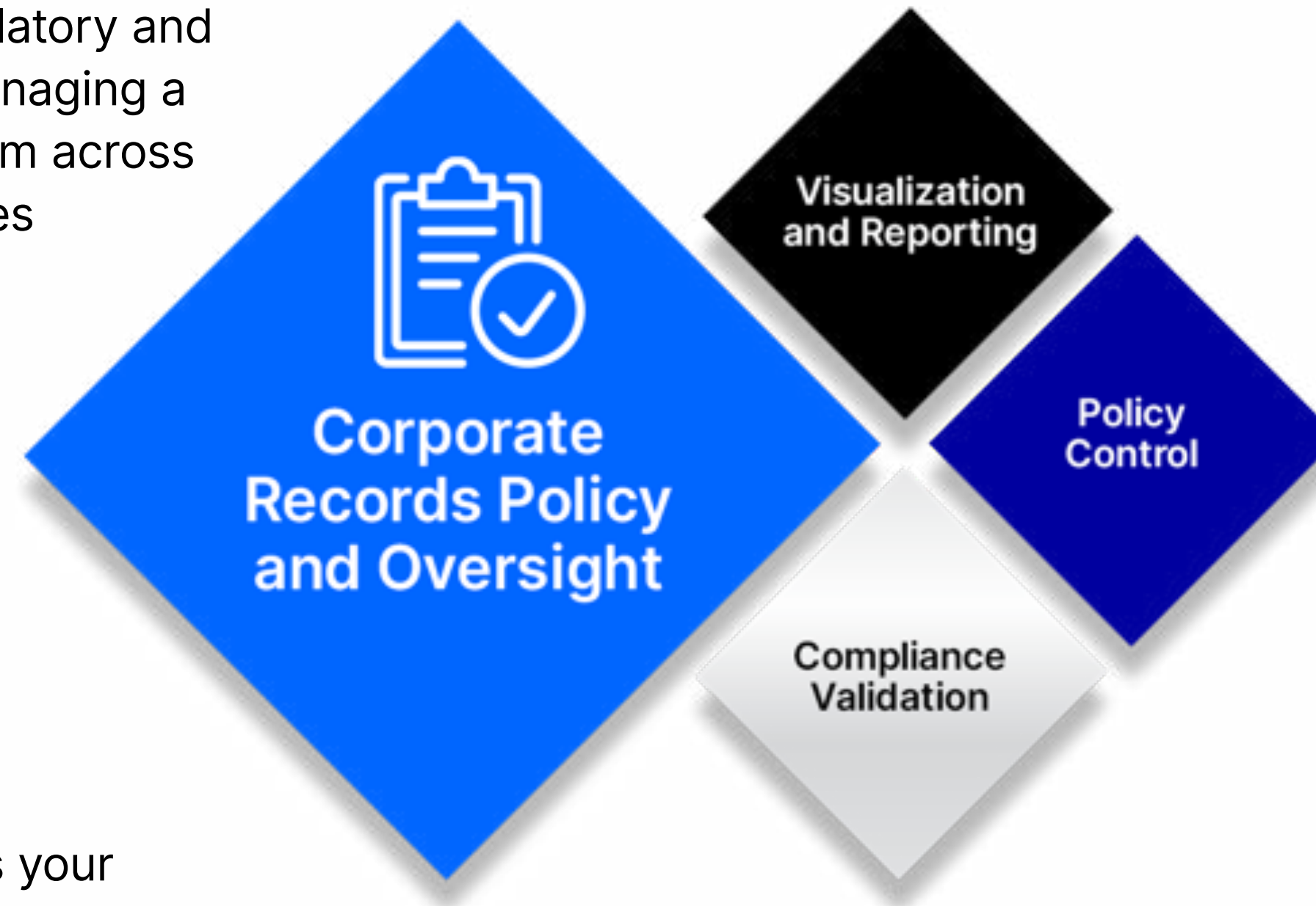
About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit opentext.com.

opentext.com

Twitter | LinkedIn | CEO Blog