White paper

Iron Mountain InSight DXP:

AImodeltraining,security,andcompliance

IRON MOUNTAIN®

Contents

03/ Executivesummary

04/ AI-powered functionality

07/ Data privacy and security

08/ Ethical AI and bias mitigation

09/ Reliability and accuracy

10/ Integration and implementation

10/ Transparency and explainability

10/ Regulatory compliance

11/ AI governance and shared compliance responsibility framework

12/ Customer responsibilities and Iron Mountain commitments

13/ Summary

Executive summary

Iron Mountain InSight® DXP empowers businesses to turn their information into actionable data through a secure, scalable platform. As a composable intelligent content platform, InSight DXP ingests and processes unstructured content and integrates with key business processes and systems via pre-built connectors or application programming interfaces (APIs). Built from the ground up with an AI-native core rather than a bolted-on framework, InSight DXP pairs an advanced multi-modal AI extraction layer with human-in-the-loop (HITL) exception handling to mitigate excessive manual review overhead while maintaining high accuracy at volume scale. The platform's advanced AI capabilities enable high levels of accuracy for automated data extraction and document classification across high volumes of enterprise data. The platform processes DOCX, TIF, and PDF files across industries and functions, including HR, invoice processing, auto lending, healthcare, finance, insurance, and government. InSight DXP is a portfolio, and functionality may vary by product. Iron Mountain can work with you to support your organization's needs.

This whitepaper delves into the critical aspects of the AI-based functionality available in InSight DXP, related data privacy and security, as well as customer responsibilities and Iron Mountain commitments. It provides a comprehensive overview of how Iron Mountain enables responsible and secure AI deployment while supporting adherence to evolving regulations (such as the EU AI Act) and best market practices.

AI-powered functionality

Native AI orchestration

While many platforms add AI as an afterthought, InSight DXP is built with AI as its native core. By pairing an advanced multi-modal AI extraction layer with HITL exception handling, the platform mitigates excessive manual review, delivering high data accuracy at enterprise scale. Recognized by independent research firms, including placement as a Strong Performer in **The Forrester Wave™: Document Mining and Analytics Platforms, Q2 2026**, InSight DXP elevates AI from experimental to enterprise-ready through:

- **Integrated governance tools:** Embedding real-time guardrails directly into the system to align AI interactions with your internal policies and business logic from the start.
- **Model neutrality:** A model-agnostic approach that enables you to use the best AI tools for your specific needs, delivering the freedom to leverage the latest innovations without provider lock-in.
- **Your data, your context:** Grounding AI in your specific business rules and routing queries through advanced permission frameworks to strictly limit data access to authorized users. By continuously learning your operations, the system refines its accuracy with every interaction, delivering precise results that reflect how your business actually works.

AI models

InSight DXP is designed to be model-agnostic, leveraging the power of premier foundational Large Language Models (LLMs) provided and hosted by tier-1 enterprise hyperscalers. The platform actively operates a multi-model tiering system that supports a wide array of frontier models, which are hot-swappable without altering the underlying prompt templates or configurations.

The platform uses these capabilities to drive advanced document processing (including automated classification, extraction, and splitting), AI chat, and operational agents. These underlying engines are explicitly selected to meet the rigorous compliance demands of global risk frameworks and to align cleanly with the transparency,

risk evaluation, and copyright management practices established by international benchmarks, including the regional guidelines for General-Purpose AI (GPAI) models.

Deployed via enterprise-grade cloud infrastructure, these foundational models utilize robust, localized regional data-residency configurations tailored directly to the customer's jurisdiction. For example, in European deployments, the architecture enforces strict "EU Data Boundary controls". This deployment enforces regional data controls so that prompt data, context windows, and generated outputs physically remain entirely within the European Union and are excluded from global model training datasets by the cloud provider.

While the underlying LLMs are broad in scope and hosted via these certified hyperscale environments, Iron Mountain deeply customizes and controls their execution. The platform applies a "model attention" mechanism to specific document contexts so the LLM can effectively interpret highly dense files. Iron Mountain then refines these foundational capabilities through a unified suite of application-level techniques, including prompt engineering, chunking, embedding, and Retrieval-Augmented Generation (RAG).

Data input

Iron Mountain utilizes a diverse range of data sources for AI model training, including the customer's document repository, open-source data, internal data, and simulated data, based on the customer's scope of work. To minimize manual exception handling and rework in high-volume operations, the platform pairs native multimodal extraction layers with HITL exception handling workflows.

To maintain data quality and mitigate bias, the platform incorporates the following measures:

- **High-volume ingestion and discrepancy flagging:** The platform processes high-volume external data streams by running parallel ingestion nodes that automatically structure unorganized files and flag visual data discrepancies at intake.
- **Layout-aware multimodal grounding:** The platform parses complex document layout geometries, signatures, and nested tables to structure context before it interfaces with a reasoning model, mitigating layout formatting inconsistencies across diverse document types.
- **Neuro-symbolic validation chains:** To mitigate generative hallucination risks in regulated environments, InSight DXP sequences probabilistic deep learning models with deterministic, symbolic validation chains. The platform applies strict linguistic rules, Regex patterns, and cross-field mathematical checks across more than 2,400 complex field extraction validation paths. This aligns contextual adaptability with rigid logic engines, mapping extracted business data directly to auditable page coordinates.
- **Human labeler evaluation:** Multi-labeler consensus workflows, including double-blind review queues and double-keying quality controls, to help mitigate indexing bias and maintain high data accuracy.
- **Validation processes:** Techniques like mock data generators and subject matter expert (SME) validation contribute to robust data validation.

Model training

InSight DXP offers a spectrum of AI model training solutions. Pre-trained AI models (both general and specialized) are available, along with model fine-tuning capabilities. Iron Mountain uses a blend of training methodologies depending on the specific model type and data availability.

Here's a breakdown of the training methodologies our implementation team employs, structured into client-centric deployment tiers:

Tier 1: Out-of-the-box generative automation

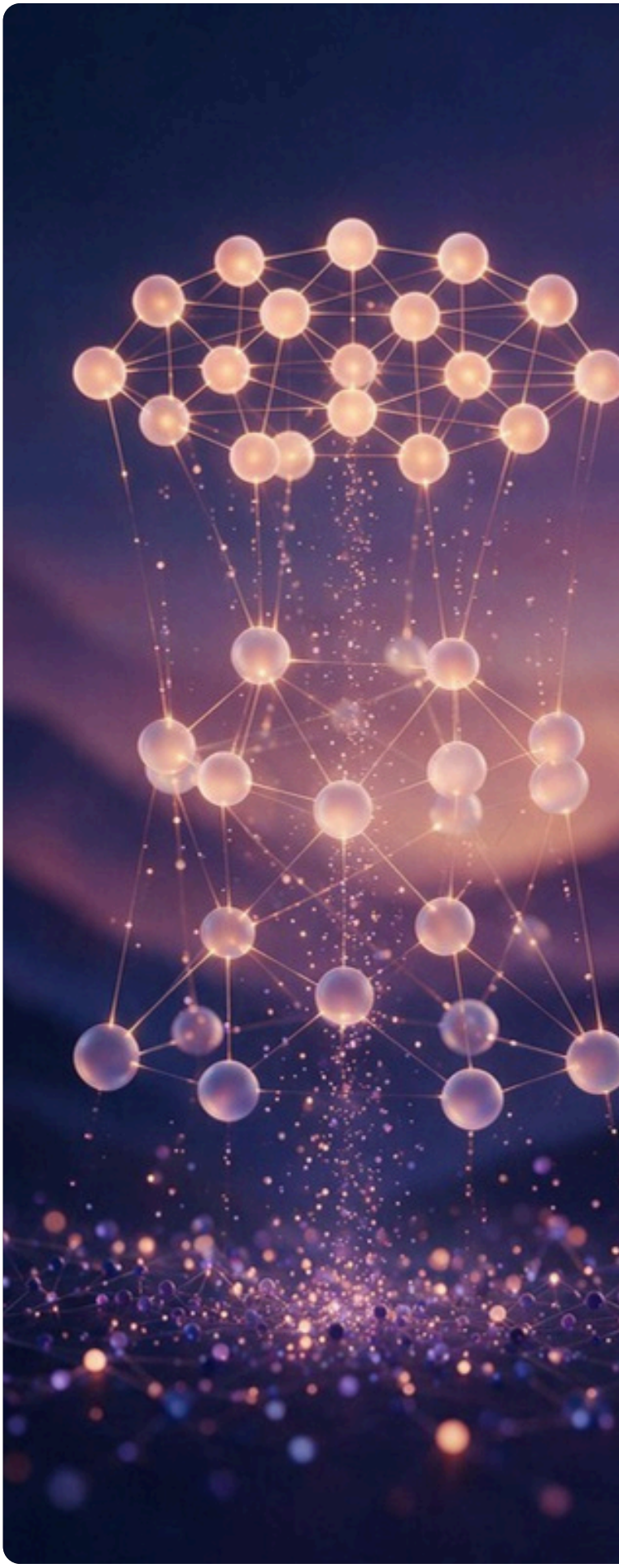
- **Zero-shot/few-shot generative AI:** Leverages prompts and example context learning to guide the model towards generating desired results.
- **One/few-shot computer vision (CV) solutions:** Ideal for scenarios involving a limited number of documents or templates, this approach utilizes document alignment and document fingerprint detection techniques.

Tier 2: Deep learning custom optimization

- **Transformer/convolutional neural network (CNN)-based solutions:** Suitable for scenarios with a sufficient volume of labeled documents, this method involves fine-tuning deep neural networks.
- **Generative AI-based solutions (small data):** Employs fine-tuning of large foundation models (large language and vision models) through techniques like prefix/ prompt tuning and low-rank adaptation of large language models (LoRA).
- **Composite solutions:** Combines Tier 1 and Tier 2 methodologies for optimal results across complex enterprise workflows.

Customization and user control

The platform's AI functionality can be turned on or off depending on the customer's specific requirements. If the customer chooses to disable this functionality, they should contact their Iron Mountain account representative.



Data privacy and security

InSight DXP prioritizes data privacy and security through robust measures. The processing of data is subject to the relevant data privacy and security standards, in accordance with applicable regulations and industry-standard practices, designed to maintain confidentiality, integrity, and controlled access.

Encryption: The platform employs industry-standard TLS 1.2 or higher for data in transit and AES 256 encryption for data at rest. This helps protect customer data from unauthorized access.

Exclusive deployment: AI models are deployed within Iron Mountain's cloud partner infrastructure to provide exclusive access and help prevent data exposure to third-party AI or cloud vendors.

Model behavior and controls: Our sovereign gateway and K8s isolation restrict agents from raw data, moving us from "honor-system" governance to "infrastructure-enforced" security.

Localized training: InSight DXP utilizes only the customer's own data for model training and validation, providing data isolation so that one customer's data is not used to train models for another customer.

Access control: The platform implements multi-factor authentication and allows users to define data access boundaries, maintaining strict control over who can access sensitive information.

Data loss prevention: The platform leverages data loss prevention (DLP) mechanisms to identify and mask personally identifiable information (PII), while preserving both original and de-identified data copies, helping maintain compliance with data privacy regulations.

Partner selection: Iron Mountain partners with vendors that adhere to strict data security and privacy standards and controls (technical, administrative, and contractual).

Logging and monitoring: The platform logs and monitors AI prompts and responses to detect and help prevent exposure of sensitive data to the LLM, supporting data protection, auditability, and compliance requirements.

Ethical AI and bias mitigation

Iron Mountain employs a multi-pronged approach to promote ethical AI development and mitigate bias:

Diverse data sources: The platform utilizes a range of data sources, including customer data, open-source data, and simulated data, so that model training is comprehensive and representative.

Human evaluation: Multi-labeler consistency methods and SME validation are employed to maintain data quality and reduce bias in training datasets.

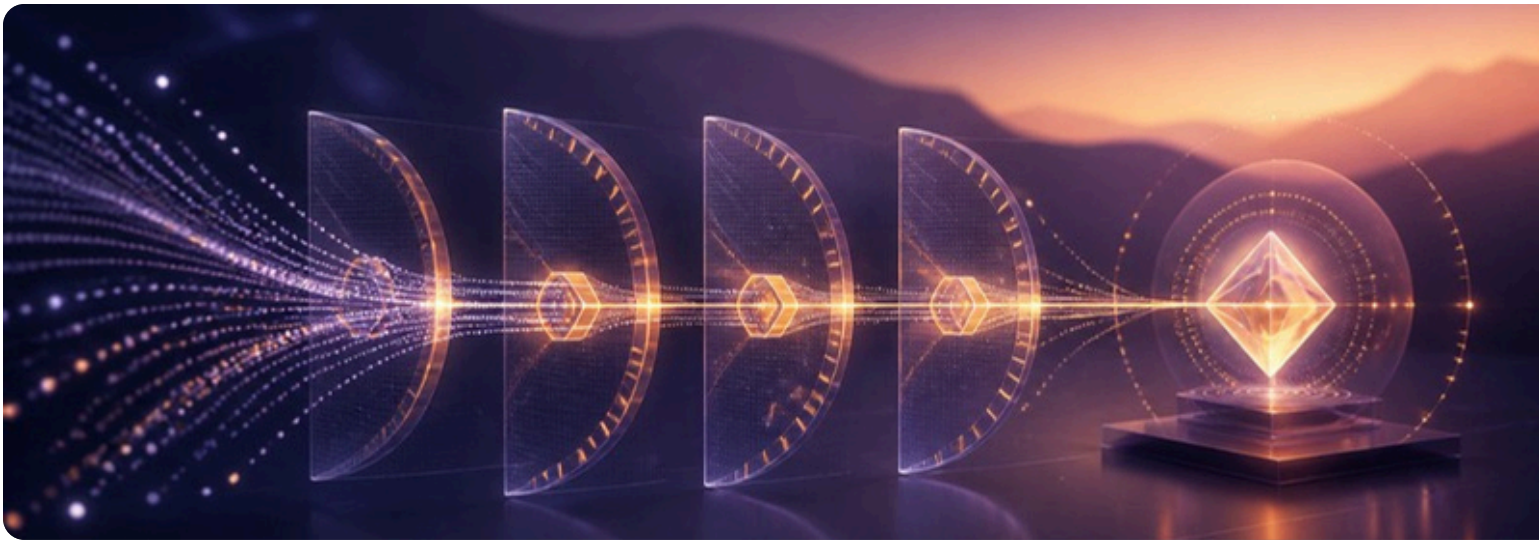
Continuous monitoring: Automated systems and manual processes, including a dedicated security information and event management (SIEM) system and cyber incident response team (CIRT), continuously monitor anomalies and potential biases.

Knowledge base and domain-specific rule engine: Calibrates prediction confidence scores and guides interactions within the context of relevant documents.

Response filters: Constrain interactions within the context of documents, helping to prevent unintended outputs.

Output plausibility: Maintained by our search index (RAG) architecture, which restricts the model's focus solely to the verified customer repository, preventing out-of-context hallucinations.

Bias and input management: Our mandatory PII scrubbing microservice is designed to act as a hard gate, physically masking identifiers before they reach an LLM to help mitigate biased processing.



Reliability and accuracy

Before production deployment, Iron Mountain rigorously tests model reliability and accuracy through:

- **Data collection and cleaning:** Gathering new and diverse datasets, including recent customer data, to maintain continued relevance and reduce noise or errors.
- **Rigorous testing:** New models undergo both quantitative and qualitative evaluation, including accuracy metrics, response latency testing, and human-in-the-loop (HITL) review.
  - **Quantitative evaluation:** Calculating accuracy metrics, response latency, and batch-data stress testing.
  - **Operational quality (non-HITL):** We have implemented automated schema validation to reject malformed or low-confidence outputs at the code level, maintaining accuracy without over-relying on manual review.
  - **Evaluation:** Human labelers use human-in-the-loop (HITL) to check prediction results and submit review reports.
  - **A/B testing:** Comparing the performance of new models with existing models.
- **Model retraining:** Utilizing fine-tuning techniques to optimize model performance.
- **Thorough evaluation:** Rigorous assessment and comparison drive the deployment of the optimal solution.
- **Feedback loop:** User feedback is actively captured and incorporated to identify areas for improvement and enhance model performance. Users can submit tickets to report bugs or performance issues.
- **Industry standard accuracy metrics:** Utilizing established metrics to track model performance.

Error detection and correction

Before production deployment, Iron Mountain uses automated systems and manual processes to help detect and correct errors or biases in AI models, such as those previously listed. These systems include:

- **Benchmark datasets:** Regular evaluation using benchmark datasets helps identify errors and biases.
- **Human-in-the-loop (HITL) correction:** Exception-based HITL flagging and correction mechanisms are used to address anomalies.

Integration and implementation

Iron Mountain offers flexible integration options:

- **Custom model development:** The platform supports custom model development through automated machine learning (AutoML) capabilities, allowing users to train models using their own datasets. The platform also offers a low-code labeling experience to create datasets, enabling subject matter experts to train models without extensive technical expertise.
- **Deployment options:** InSight DXP is a cloud-based offering.

Transparency and explainability

Iron Mountain promotes transparency in its AI processes:

- **Model prediction tracing:** The platform connects model predictions to document resources.
- **Transparency and oversight:** Our trace V2 Schema, stored in write-once, read-many (WORM) (tamper-proof) vaults, provides the "flight recorder" transparency that high-compliance customers are now requiring.
- **Information sharing:** Users are provided with information about the model training and updating process, including details on training data methodologies used.

Regulatory compliance

InSight DXP is designed to meet various regulatory requirements:

- **Ongoing compliance:** The platform is designed to support your compliance with evolving data privacy regulations like the General Data Protection Regulation (GDPR) and security standards such as ISO-27001, ISO 27701, ISO 9001, and SOC 2 Type 2.
- **Versatile support:** InSight DXP successfully navigates diverse compliance landscapes, including GDPR in the EU, HIPAA and FedRAMP in the US.

AI governance and shared compliance responsibility framework

To support alignment with international AI frameworks, InSight DXP's compliance obligations operate under a shared responsibility matrix split by technological layers. This division corresponds directly to the industry-standard boundaries established between the foundational model provider and Iron Mountain:

- **Foundational model provider:** Our enterprise hosting partners operate as the General-Purpose AI (GPAI) model providers. Under this framework, their role encompasses the foundational model's core integrity, training dataset compliance, and adherence to established transparency and safety codes of practice. They are expected to compile and maintain exhaustive model technical documentation and provide downstream integrators with the necessary disclosure artifacts regarding capabilities, operational limitations, and systemic risk evaluations. Furthermore, the model provider provides physical security for the AI infrastructure under ISO/IEC 42001 certification (the international standard for AI Management Systems) and is designed to maintain geographic data sovereignty through localized regional data boundaries.

- **Iron Mountain:** We operate as the Downstream AI System Provider, managing the application-level data governance, output plausibility, and operational fairness of the workflows layered on top of the base GPAI engine. To support core data quality and bias evaluation principles, we enforce a mandatory PII-scrubbing microservice that masks demographic or sensitive variables before they interface with the LLM. To maintain strict output plausibility, our Retrieval-Augmented Generation (RAG) architecture confines the model's focus solely to the customer's verified document repository, preventing hallucinations. Finally, to support rigorous logging and record-keeping mandates, our platform captures an immutable operational audit trail using the Trace V2 schema, securely anchored within write-once, read-many (WORM) tamper-proof vaults.

Customer responsibilities and Iron Mountain commitments

Customer responsibilities

Data provision and quality: Customers are responsible for providing accurate, complete, and high-quality data for processing. The effectiveness of AI models, including AutoML and generative AI, heavily depends on the quality of input data. Customers must ensure that their data is free from errors and inconsistencies to achieve optimal results.

Compliance with regulations: While the underlying model infrastructure and platform layers maintain rigorous independent compliance certifications, customers are responsible for ensuring that their specific operational deployment aligns with local legal frameworks (such as GDPR, CCPA, or EU AI Act risk classifications). Because risk tiers are determined entirely by the customer's intended application (e.g., general invoice processing vs. high-stakes use cases), the customer is responsible for evaluating their deployment context, obtaining necessary data usage consents, and ensuring end-to-end regulatory compliance for their specific use case.

Security measures: While Iron Mountain provides a secure environment for data processing, customers are responsible for implementing their own security measures to protect data during transmission to and from the platform. This includes using encryption and secure access protocols.

Usage policies: Customers should adhere to the platform's terms of service and usage policies. This involves using the AI tools provided by Iron Mountain ethically and responsibly, without engaging in activities that could harm other users or violate others' rights.

Iron Mountain commitments

Data security: Iron Mountain is committed to maintaining a secure environment for storing and processing customer data. This includes implementing robust security measures designed to protect against unauthorized access, data breaches, and other cyber threats.

Model accuracy and performance: Iron Mountain is committed to the diligent deployment, maintenance, and continuous tuning of the application-level safeguards described in this document (including context-pinning RAG architectures, input-filtering microservices, and immutable operational logging) pursuant to your specific Scope of Work (SOW) requirements. Because foundational large language models are probabilistic by nature, automated outputs and regional regulatory alignment are subject to the specific operational context, and Iron Mountain continuously optimizes these system safety frameworks to minimize prediction errors.

System availability: Iron Mountain is dedicated to maintaining high availability and reliability of its platform services. In the event of system downtimes or disruptions, Iron Mountain is responsible for promptly addressing issues to restore service continuity.

Support and maintenance: Iron Mountain provides ongoing support and maintenance services to assist customers in effectively using the platform's features. This includes regular updates, troubleshooting assistance, and user training resources.

Summary

By leveraging robust security measures, transparent practices, and continuous improvement, InSight DXP provides a secure and reliable platform for businesses to harness the power of AI while supporting adherence to evolving regulatory requirements and industry-standard practices.

IRON MOUNTAIN®

800.899.IRON | ironmountain.com

About Iron Mountain

ron Mountain Incorporated (NYSE: IRM) partners with more than 240,000 customers across 61 countries, including approximately 95% of the Fortune 1000. We serve as the trusted guardians of what matters most, helping unlock what's possible. By connecting physical, digital, and intelligent workflows, our lifecycle solutions help organizations realize enterprise value at scale. From complex challenges across information management and digital transformation to data security, data centers, and asset lifecycle management, we empower organizations to navigate an ever-changing landscape to prepare for tomorrow's possibilities. Our longstanding commitment to safety, security, sustainability, and continuous innovation defines everything we do. Visit www.ironmountain.com for more information.

⌈ 2026 Iron Mountain, Incorporated and/or its affiliates ("Iron Mountain"). All rights reserved. Information herein is proprietary and confidential to Iron Mountain and/or its licensors, does not represent or imply an invitation or offer, and may not be used for competitive analysis or building a competitive product or otherwise reproduced without Iron Mountain's written permission. Iron Mountain makes no guarantees to develop, deliver, or market any specific product, feature, functionality, or service. All future products, functionality, features, technical specifications, or service offerings, including their availability or release dates, are subject to change or withdrawal at any time without notice. Iron Mountain does not provide a commitment to any regional or future availability and does not represent an affiliation with or endorsement by any other party. Iron Mountain shall not be liable for any direct, indirect, consequential, punitive, special, or incidental damages arising out of the use or inability to use the information, which is subject to change, provided AS-IS with no representations or warranties with respect to the accuracy or completeness of the information provided or fitness for a particular purpose. "Iron Mountain" is a registered trademark of Iron Mountain in the United States and other countries, and Iron Mountain, the Iron Mountain logo, and combinations thereof, and other marks marked by Ð or TM are trademarks of Iron Mountain. All other trademarks may be trademarks of their respective owners.