

ENTERPRISE DATA HEALTH ASSESSMENT

How Big is Your Unstructured Data Problem?

An assessment to benchmark your visibility, data footprint, infrastructure efficiency, security posture, governance, and AI readiness.

Visualize

Minimize

Optimize

Secure

Govern

A practical self-assessment for enterprise IT, infrastructure, security, data, privacy, and governance leaders.

Use the results of this assessment to identify where assumptions may be masking unnecessary cost, risk, and operational friction.

You can't manage what you can't see.

Unstructured data is everywhere: file shares, NAS, cloud repositories, collaboration platforms, archives, and business applications. As the estate grows, organizations often lose the ability to answer basic questions about what they have and what should happen to it.

Before investing in another tool or launching another transformation project, determine how confidently your organization can answer the questions below.

- What unstructured data exists across the enterprise — and where does it live?
- Which data is sensitive, regulated, stale, redundant, or no longer needed?
- Who can access sensitive information — and do they still need that access?
- How much expensive storage, backup, and recovery capacity is being consumed by low-value data?
- Are retention and governance policies actually being enforced across repositories?
- Which data is relevant, appropriate, and trustworthy enough to support AI initiatives?

THE OBJECTIVE

Identify where assumptions may be making unnecessary cost, risk, and operational friction.

How to Score Your Data Health

For each statement, choose the answer that best reflects your current state. Be conservative: score based on what you can consistently prove across your environment, not what policy says should be happening.

Score	Maturity	What it means
3	Strong	Continuous, enterprise-wide visibility and control
2	Partial	Capability exists, but only for selected repositories, teams, or processes
1	Limited	Mostly manual, point-in-time, or dependent on investigation
0	Unknown	You cannot reliably answer or prove the statement today

Maximum score: 45 points. Record your score for each question, then total the five sections.

SECTION 1

VISUALIZE — Do You Understand Your Data Estate?

#	Assessment question	Score (0–3)
1	Can you identify the major sources of unstructured data across on-premises, cloud, and hybrid environments?	
2	Can you understand data by age, owner, access, type, sensitivity, and business value?	
3	Can you continuously identify meaningful changes in your data estate without relying on one-time scans or manual inventories?	

Section subtotal: **0** / 9

SECTION 2

MINIMIZE — Are You Keeping More Data Than You Need?

#	Assessment question	Score (0–3)
4	Can you quantify inactive, obsolete, redundant, or trivial data across your environment?	
5	Can you identify data that has exceeded business, legal, or retention requirements?	
6	Do you have a repeatable, defensible process for deleting unnecessary data?	

Section subtotal: **0** / 9

SECTION 3

OPTIMIZE — Is Your Data Consuming Unnecessary Resources?

#	Assessment question	Score (0–3)
7	Do you know how much high-cost production storage is occupied by inactive or low-value data?	
8	Can you determine which data could be moved, archived, tiered, or deleted based on policy and business value?	
9	Can you identify which enterprise data is relevant and appropriate for AI, analytics, or other strategic initiatives?	

Section subtotal: **0** / 9

SECTION 4

SECURE — Is Unnecessary Data Creating Unnecessary Risk?

#	Assessment question	Score (0–3)
10	Can you identify sensitive and regulated information regardless of where it resides?	
11	Do you know when sensitive data is broadly accessible, exposed, or available to users who no longer require access?	
12	Can you systematically reduce unnecessary sensitive-data exposure and attack surface?	

Section subtotal: **0** / 9

SECTION 5

GOVERN — Can You Enforce the Policies You Create?

#	Assessment question	Score (0–3)
13	Are retention, privacy, and governance policies consistently applied across repositories?	
14	Can you produce evidence showing how sensitive, regulated, and aging data is being managed?	
15	Can you automatically take or orchestrate action when data violates policy?	

Section subtotal: **0** / 9

TOTAL YOUR FIVE SECTIONS

Your total score: **0** / 45

What Does Your Score Tell You?

Add your five section subtotals. Use the ranges below as a directional indicator of your current ability to understand and control unstructured data.

36-45 **CONTROLLED**

You have strong visibility and governance over much of your unstructured data estate. Opportunities may still exist to reduce cost, automate action, improve AI readiness, or close gaps between repositories.

24-35 **REACTIVE**

You have meaningful tools and processes in place, but gaps between systems, teams, and repositories may be creating unnecessary cost, risk, and manual work.

12-23 **EXPOSED**

Significant portions of your data estate may be unmanaged, unnecessarily expensive, overexposed, or difficult to govern. Quantifying the highest-risk and highest-cost areas should be a priority.

0-11 **UNKNOWN**

The greatest risk may be what you cannot currently prove. Without visibility, it is difficult to accurately quantify cost, exposure, compliance posture, or AI readiness.

Your total score: 0 / 45

Your Score Isn't the Whole Story

Policies and assumptions tell you what should be happening. Your data tells you what is *actually* happening.

Most organizations cannot accurately answer every assessment question until they analyze the data itself. Consider the difference:

Area	What should happen	What you still need to know
Retention policy	A policy may say data should be deleted after seven years.	That does not tell you how much 10-, 15-, or 20-year-old data is still sitting in production.
Access policy	A policy may define who should have access to sensitive data.	That does not tell you who actually can access it today.
Storage strategy	A strategy may move older data to lower-cost tiers.	That does not tell you whether the data needs to exist at all.
AI governance	A policy may define which data AI systems should use.	That does not tell you whether irrelevant, sensitive, stale, or low-quality content is entering the pipeline.

The assessment becomes more valuable when self-reported confidence is compared with evidence from the data itself.

Sample Findings: What a Congruity360 Data Assessment Can Uncover

Illustrative example only. The figures below demonstrate the types of findings an enterprise data assessment can surface; they are not presented as Congruity360 customer averages or industry benchmarks.

Illustrative Customer Sample Data – 10TB Assessment

Finding	Illustrative result	Why it matters
Data untouched for 2+ years	74%	Inactive data that can be tiered to a lower-cost, long-term archive with restricted access.
Redundant, Obsolete, Trivial Data	49%	Cull this data through a combination of tiering and deletion to reduce cyber-attack footprint, reduce storage costs, and improve infrastructure efficiency.
Sensitive data identified	5%	Sensitive information may exist in more locations than teams expect.
Sensitive files beyond retention	140k files	Potential unnecessary exposure and expanded attack surface.
Sensitive files with broad access	212k files	Potential unnecessary exposure from over access.
Data beyond defined retention (7+ years last modified)	39%	Potential governance, discovery, and compliance burden.
Data identified as potentially AI-relevant	28%	The full enterprise estate should not automatically be treated as useful or appropriate AI input.

Translate Data Findings into Business Impact

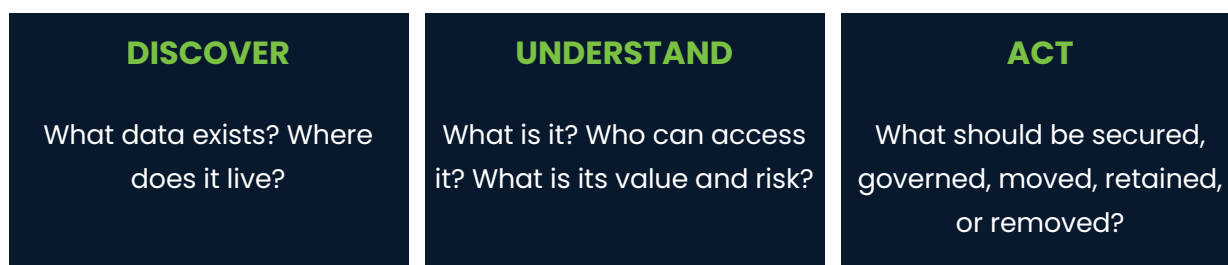
Cost	Reduce unnecessary spend associated with storing, protecting, replicating, backing up, and recovering data that provides little or no business value.
Cyber Risk	Reduce the volume of unnecessary sensitive information and excessive access that can increase the impact of a security incident.
Governance & Compliance	Move from written policies to evidence that retention, privacy, and governance requirements are actually being applied to data.
AI Readiness	Identify data that is relevant, appropriate, trustworthy, and governed before it is introduced into AI and analytics workflows.
Operations	Reduce the data burden associated with migrations, backups, recovery, eDiscovery, audits, and infrastructure management.

One underlying data problem can create multiple business problems.

From Assumptions to Evidence

A self-assessment identifies potential gaps. A data assessment quantifies them.

Congruity360 helps organizations build a complete understanding of their unstructured data estate so they can determine what exists, what matters, what creates risk, and what action to take.



Questions an assessment can help answer

- What exists, and where does it live?
- What is sensitive or regulated?
- Who has access – and where is access broader than necessary?
- What is inactive, stale, redundant, or beyond retention?
- What can be safely moved, archived, or deleted?
- Where can infrastructure spend potentially be reduced?
- Which data is appropriate to support AI and analytics?
- What actions should be prioritized first?

Turn Your Score Into Real Findings

Your self-assessment highlights where gaps may exist. The next step is to quantify what is actually happening across your unstructured data estate and identify the highest-value actions to take first.

>> NEXT STEP: TALK TO US ABOUT A DATA ASSESEMENT <<

Use the findings to prioritize cost reduction, risk reduction, governance, compliance, AI readiness, and data lifecycle initiatives.