



— WEBINAR SERIES —

Sandy Serkes
Co-Founder & CEO

Jennifer Nelson
VP Strategic Solutions

Unhostaging Your Data: Reclaiming Control from Troublesome Applications & Systems



Webinar Series



www.ValoraTech.com



ROT Remediation:
File Share Clean-up
January 21



From Observational to
Operational Governance
July 23



Letting the Robots
Classify: Automating IG
February 18



Unhostaging Your Data:
Reclaiming Control
August 19



AutoClassifying the 3 Rs:
ROT, Records & Retention
March 18



Managing Data Privacy &
While Managing Records
September 16



Structured Data
Management & Disposition
April 15



AI: What Happens when
the Model Collapses
October 21



10 Ways AutoClassification
Can Impact AI
May 13



AI Policy & Governance
with Lewis Eisen
November 18



How AC Transforms
Enterprise Search
June 17



AutoClassification & AI
Trends in Enterprise
December 16

About Valora



- AutoClassification platform for Information Governance
- Automates **discovery**, **identification**, **classification** & **defensible disposition**
- Trusted by IG teams, Records & IM teams, Legal, Compliance, IT teams
- Brings **multiple enterprise repositories** into a single view, and **multiple IG use cases** under one umbrella





Poll Results

Who's here today
& what are your interests?



Who you are & what your challenges are



Who You Are

Records / Information Management

Knowledge Management

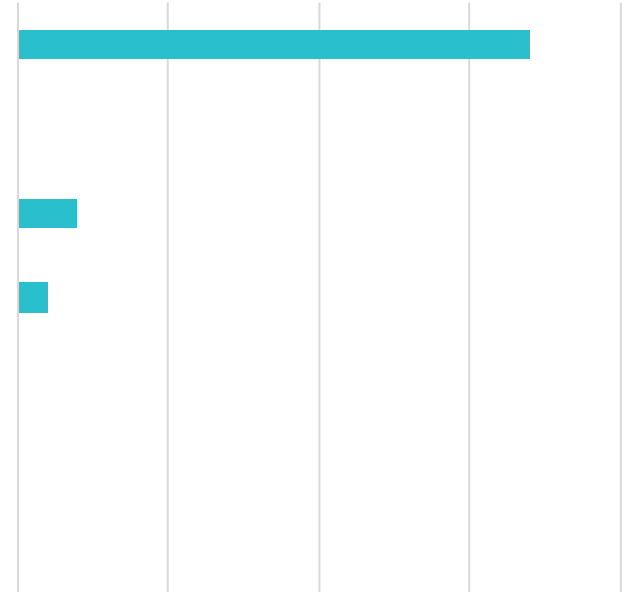
Legal

Compliance

IT

Data Privacy / Security

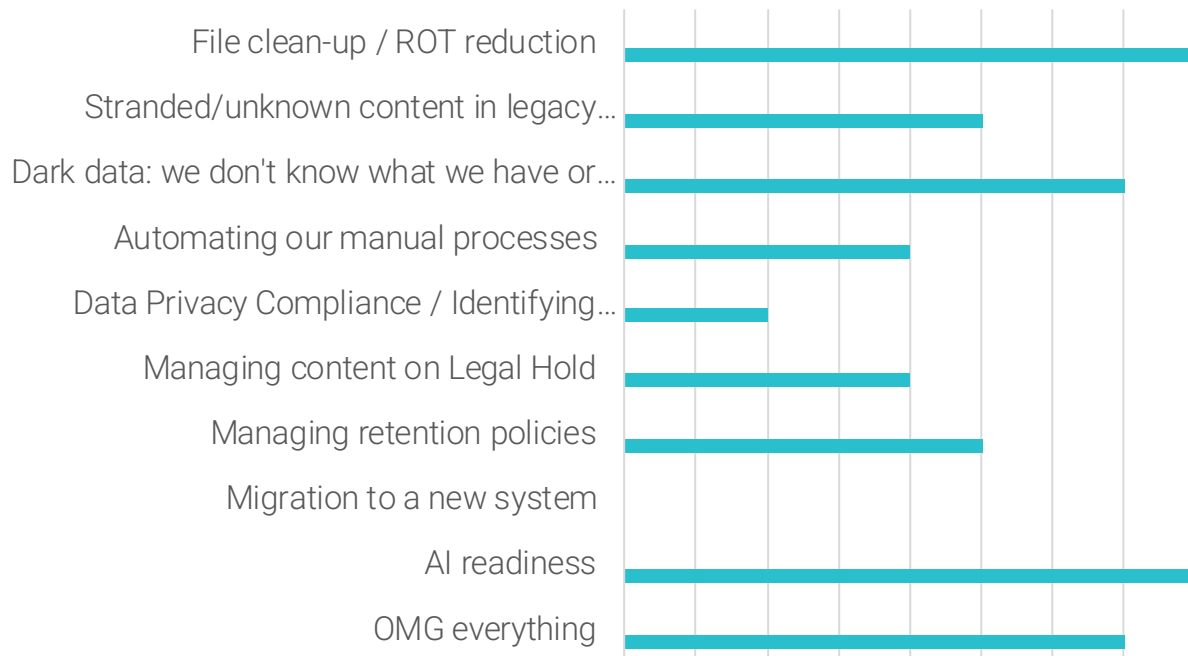
Other



Who you are & what your challenges are



Your RIM & IG Challenges



Housekeeping



Q&A



Recording



Slides



Feedback

Agenda



- Understanding data hostaging
- How to unhostage data
- How unhostaging plays into the bigger InfoGov narrative (including AI)
- How to prevent future hostaging attempts (future-proofing)

When is data “hostaged”?



- Don't know what it is/contains and no easy/good way to find out
- Exorbitant costs to examine data, use data, retrieve/remove data, retention manage it
- Unintentionally unusable systems/media (ex: don't know password, don't have access, license expired, legal hold never lifted)

Data is **hostaged** when it is legally yours, but operationally inaccessible.

Examples of hostaged data

- Data trapped in SaaS & 3rd party apps
- Throttled API access & transactions
- Proprietary formats
- No ability to access/edit/remove metadata
- No lifecycle or retention management
- Blocked (or forgotten) access & permissions, expired licenses
- Abandoned data (& abandoned data under legal hold)
- M&A & divestiture data
- Egress, access, editing & management fees
- Decommissioned & EOL systems
- Ransomware

Hostaged Data = information the organization remains responsible for, but cannot readily understand, access, govern, use, retain, migrate, or dispose of.



How does data become “hostaged”?

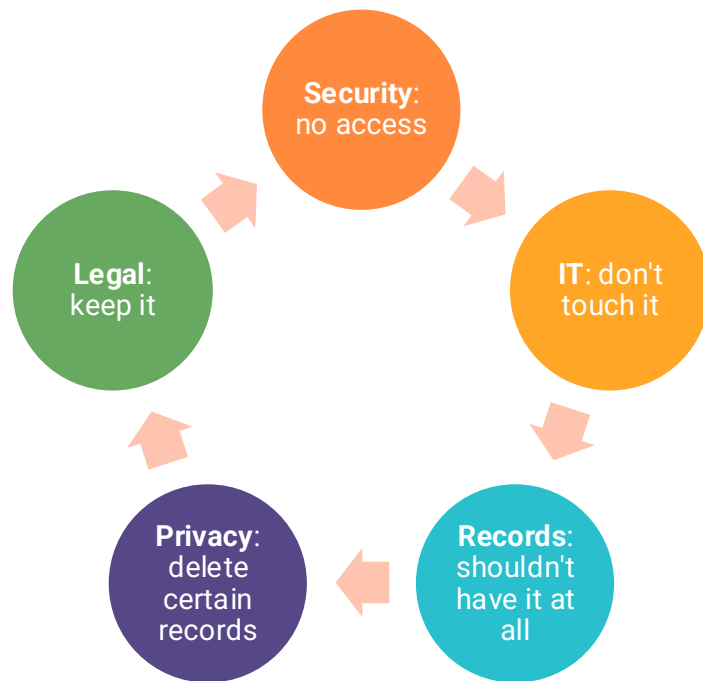
- Systems (and people/knowledge) age out.
No one remembers/has access/understands.
- Intentional barriers (usually put in place by 3rd parties w/ incentive to keep you hostaged). More on this one later.
- Increased security or privacy concerns lock out non-essential parties (and attempts at lifecycle mgmt. or remediation)
- Bad/threat actors (ex: breach & ransomware)



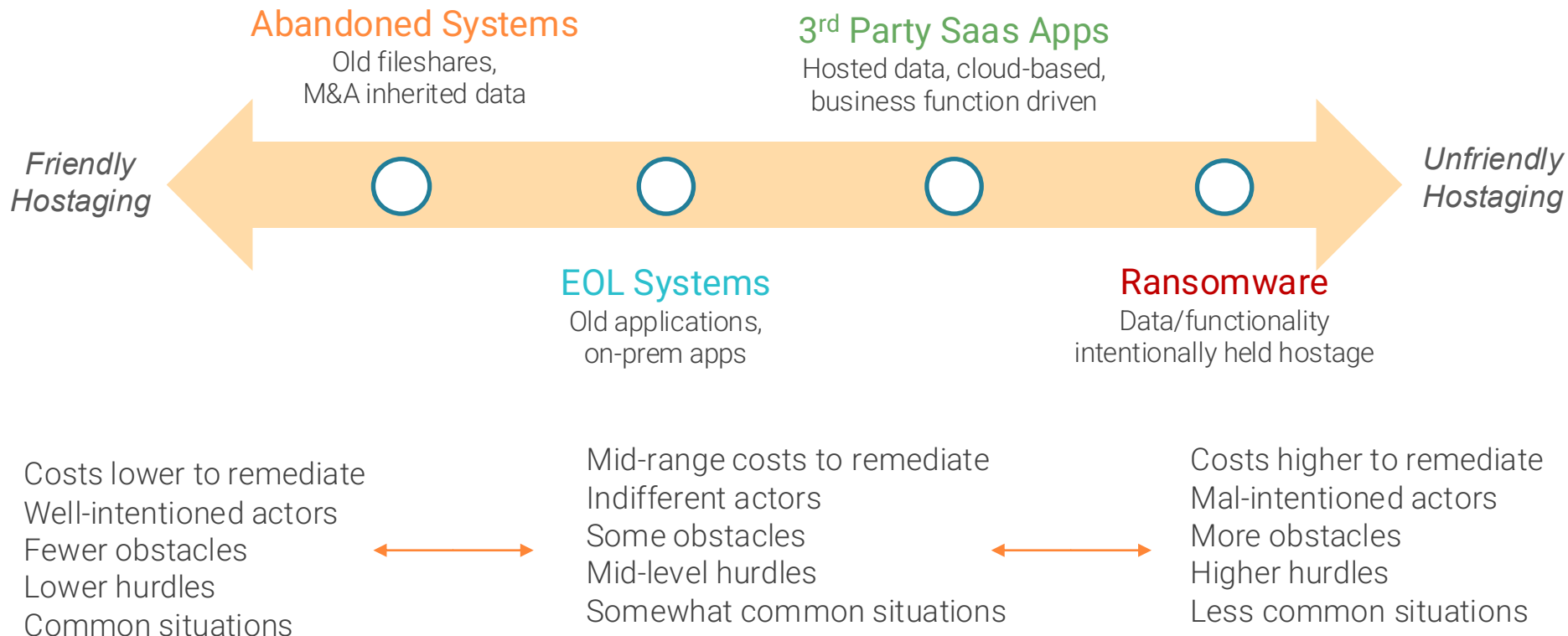
How does data become “hostaged”?

A lot of it is self-inflicted.

- Legal says preserve it.
- Security says nobody gets access.
- IT says the application can't be touched.
- Records says it should have been destroyed seven years ago.
- Privacy says we need to find and delete certain records.
- Nobody has enough knowledge or authority to reconcile those positions.



The Hostaging Continuum



Understanding data hostaging

These things are not the same...



Ownership

- Who's data is it?
- Who created the data?
- Who can edit it?
- Who has the right to use the data?
- Who is responsible for the data's accuracy, usage, safety, disposition?



Access

- Who can see the data?
- What are they allowed to see/do/know?
- Who can download or use the data? How?
- What systems can interact with the data?



Control

- How can we lock this data down?
- How can we remediate this data?
- How can we bulk manipulate or access this data?
- How can we classify this data?



Governance

- How do we ensure prompt & accurate deletion?
- How do we dictate what can be used, when, and by whom (like AI)?
- How do we account for change over time?
- How do we maintain our rights over our data?
- How do patients/ customers/ employees retain their rights?

Data Hostaging is not new. It's just becoming more sophisticated.



Example: Paper *Iron Mountain*

- Cost to store boxes in safe environment
- Cost to retrieve boxes from environment
- Cost to replace boxes back in environment
- Cost to upgrade contents to scanned images
- Cost to remove/destroy boxes
- Bulk activity very expensive & difficult to obtain



Example: Digital *iManage*

- Cost to store files in safe environment
- Cost to access files inside environment
- Easy to add more data,
Difficult to change or remove data/metadata
- Easy to search/see, hard to action/govern
- Hard/impossible to migrate data
- Hard/impossible to retention manage data
- Bulk activity throttled

These “switching costs” are built in to intentionally deter you from making changes.
They promote “stickiness”, inaction, and “out of sight/out of mind” behavior.

The very opposite of vigilant governance!

The rental house analogy

Imagine you are renting a house.

- You pay a monthly rent to use the premises, the yard, the garage, the utilities, etc.
- Now imagine you had to pay an extra fee every time you:
 - Drove up or down the driveway
 - Used the shower
 - Opened or closed the windows
- Now imagine you were only allowed to do these things 1x/day, between 3-4am only!
- Now imagine **you own the house** and you **still** have to pay fees and restrict your usage!



This is called “Tollgating”

- Deloitte: SaaS Tollgating = restrictions, gating mechanisms, and new charges for data access imposed for accessing data that enterprises previously considered their own.
- “The data organizations thought they owned is increasingly being fenced off by vendors, metered, and sold back to the business that created it.”

● The pricing paradox of agentic SaaS: What to do about tollgating?


Download


Share

Enterprise AI is colliding with a hard truth: The data organizations thought they owned is increasingly being fenced off by vendors, metered, and sold back to the business that created it.

For a long time, many organizations have assumed they could pull data out of vendor software and use it freely throughout their business. They built **systems** like data lakes and integrated platforms based on that assumption.

Now, vendors are actively introducing restrictions, gating mechanisms, and new charges for data access—commonly referred to as “tollgating” fees, which are charges imposed for accessing data that enterprises previously considered their own. Alongside this, practices like “token metering,” where vendors track and bill based on the volume or frequency of data usage, are becoming standard. For instance, a company using a cloud-based CRM system may now incur additional fees when exporting customer data for analytics, directly impacting their AI strategies, budgeting, and architectural decisions. These developments raise a broader strategic question: How should organizations rethink cost structures, sourcing models, and the economics of AI in a tollgated environment?

Why this moment is different

TABLE OF CONTENTS ^

Why this moment matters

- Emerging agentic SaaS architectures
- Preserving data ownership rights
- Evolving vendor ecosystem management
- Adapting cost models to new AI economics
- Balancing cost and capability

Data hostaging conditions are proliferating

- Regulators now explicitly recognize data lock-in as enough of a problem to regulate it.
- AI is creating an economic incentive to put tollbooths around enterprise data.
- The information environment itself is becoming harder to govern. AI as creating unprecedented complexity.

4 statistics on the state of enterprise data security

Organizations face significant data loss and breaches as security and governance lag behind AI use.



Hannah Warfel, Associate Editor
July 13, 2026

2 Min Read

Editor's Spotlight

- **85% of organizations had a major data loss incident within the past 12 months**
- **66% of organizations found an AI tool accessing sensitive data it wasn't supposed to**
- **47% of security professionals said AI threats are the top security risk facing unstructured data**
- **45% of IT decision-makers say they have active concerns about data leaks**

Who is harmed by data hostaging?

Legal / eDiscovery

- Harder and more expensive to preserve, search, collect and produce
- Greater uncertainty about completeness and defensibility
- Legacy repositories persist because nobody can confidently remediate them

Records & Information Governance

- Retention policy exists—but cannot be executed. ROT accumulates.
- Classification and disposition become episodic projects instead of normal operations

Privacy / Compliance

- Cannot reliably find or delete regulated information
- DSARs and remediation become slower and costlier
- Data minimization exists on paper but not in practice

Security

- Unknown, obsolete and inaccessible content expands the attack surface
- Security controls themselves can inadvertently make remediation harder

IT

- Forced to maintain applications and infrastructure nobody wants
- Migration/decommissioning gets deferred
- Technical debt becomes governance debt

The Business / AI

- Valuable information cannot readily be reused
- M&A, migration, analytics and AI initiatives become more expensive
- AI can only exploit data the enterprise can actually reach and govern

...and who benefits?

Sometimes: the incumbent vendor/provider

- Increased switching costs
- Greater customer “stickiness”
- Continued storage/subscription revenue
- Additional revenue from APIs, extraction, professional services or migration
- Reduced incentive for customers to remediate/delete large volumes

But often: nobody

- Not all hostaging has a hostage-taker.
- Data can be hostaged without anyone benefiting at all.

How to unhostage?

There is a right way & a wrong way (several)

Right way:

- Intentional efforts to understand data, mitigate issues, reduce costs, provide reasonable access to caretakers/governors

Wrong way:

- Bulk deletion/migration/tagging without understanding what data actually is/says/is used for.
- Allowing systematic access without knowledge (Purview? AI?)
- Trading one set of costs for another (storage & productivity -> tokens & inference complexity)
- Ignoring lifecycle management & curation, appropriateness, versioning, etc.



We're going to walk you through the right way



- Via several real-world examples. These will range from easiest (and most common, lowest risk) to hardest (and most complex/dangerous)
- Example 1: Abandoned file share
- Example 2: EOL application
- Example 3: 3rd Party “Friendly” Hostaging
- Not covering today: Example 4: 3rd Party “Unfriendly” Hostaging (i.e., ransomware!)

Commonalities

- 2 main Unhostaging approaches:
 - Lift, Shift, Sift OR
 - Sift, Lift, Shift
- Seek ways to access the data cost-effectively and in bulk (at scale)
 - Ex: API or SDK access, Direct database query, bulk export
- Anticipate pushback
 - Allay replacement fears, negotiate new terms (more later), have benefits & upside arguments ready
 - Ask about throttling and access “chokepoints” in advance

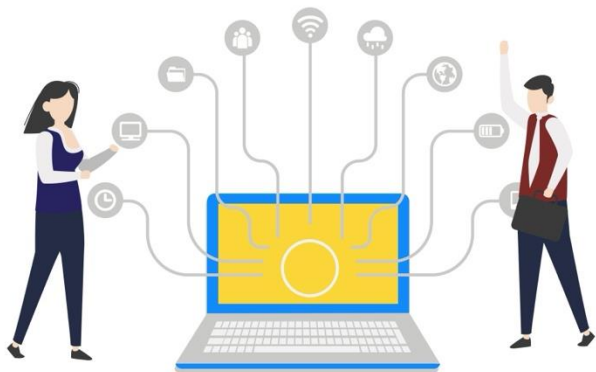


Example 1: unhostaging the abandoned file share



- Ensure access and permissions to the share
 - Ideally, a service account
- Have a manual look at the data
 - How much data, filetypes, last accessed/edited
- Create/install processing connector
- Pull a statistical sample & see what it tells you
 - Run processing benchmarks
- Configure processing to *meet the status of the data*
- Run at high scale
- Determine RFL – Recommended File Location (and status) for all files that will continue their journey
- Watch out for: ancient, unlifted legal holds, unreasonable territoriality, black market data stashes

Sidenote on connectors



What is a connector?

- Connectors are software agents (not AI agents!) that use APIs to connect to various repositories, applications & data stores to scan, analyze, and action content where it lives.

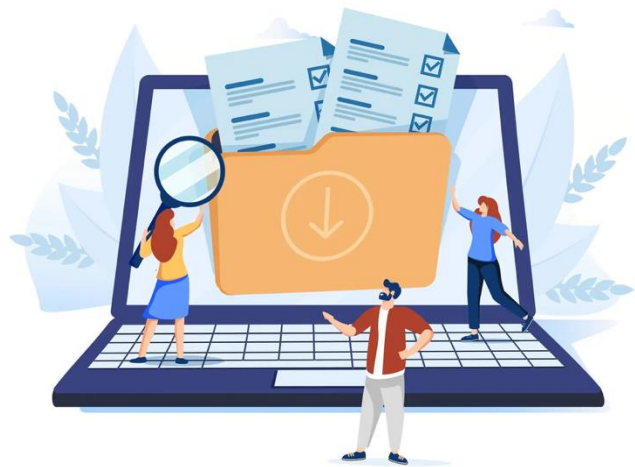
What do they do?

- Provide the ability to access, scan, analyze, and action content programmatically at scale, inside the application (process in place).

Types of Connectors:

- *Read-only*: does not alter existing metadata (last accessed date)
- *Write-only*: moves (or writes) data to a repository
- *Read/Write*: reads (analyzes) & writes (makes changes)

Example 2: unhostaging the EOL Application



- Understand who owns the actual application (SaaS vs. software purchase?) and where it resides (cloud? On prem? Laptops?)
- Ensure access and permissions to the application & data
 - Ideally, a service account
 - If not possible, attempt bulk data download (may need vendor's assistance)
- Create/install processing connector (if possible)
- Pull a statistical sample & see what it tells you
 - Run processing benchmarks
- Configure processing to *meet the status of the data*
- Run at high scale
- Determine RFL – Recommended File Location (and status) for all files that will continue their journey
- Determine bulk/scale migration plan
- Watch out for: old/proprietary file formats, data loss/degradation, inability to access metadata

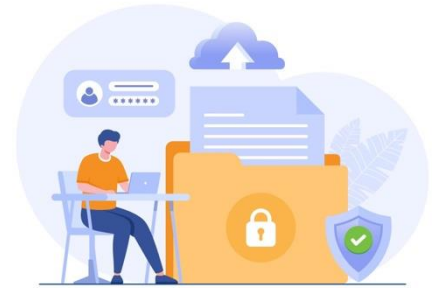
Example 3: unhostaging data from the “friendly” 3rd Party



- Understand who owns the actual data & metadata. Find the contract terms.
- Ensure access and permissions to the application, data & metadata
 - Ideally, a service account and API access
- Create/install processing connector
- Pull a statistical sample & see what it tells you
 - Run processing benchmarks
 - Watch for throttling
- Configure processing to meet the status of the data and *any access limitations*
- Run at high scale (as permitted)
- Determine RFL – Recommended File Location (and status) for all files that will continue their journey
- Determine bulk/scale lifecycle (or migration) plan
- Watch out for: custom file formats, inability to access metadata, egress fees, throttling

How to prevent future hostaging attempts (future-proofing)

- Hostaging is often governance debt.
 - Every time we defer classification, disposition, migration, ownership decisions, or contractual protections, we accrue governance debt.
 - Eventually that debt becomes a constraint on what the enterprise is able to do with its own information.
- Make **exit readiness** a core IG requirement for all new applications/systems/data uses
 - What will be the intended lifecycle and disposition of this data?
 - How will we access it (in bulk) to perform analysis, remediation, management & disposition?
 - What will AI want or need from this data? How can we ensure safe & efficient access?
- Require these in contract negotiations:
 - contract concerns: bulk export rights, metadata completeness, API/bulk-access rights, deletion capabilities, post-termination access, transition assistance, and fees associated with retrieval/export/remediation.
 - Don't negotiate merely for ownership of the data. Negotiate for continuing ability to govern it.
 - Ensure the organization can inventory, classify, preserve, export, remediate, and dispose of its information independently of the vendor.



The hostaging test

For any repository or platform, can we:

- **Know** — determine what information we have?
- **Reach** — access it at enterprise scale?
- **Control** — preserve, remediate, classify and delete it?
- **Action** — edit, move, quarantine, produce, remediate, dispose?
- **Migrate** — extract it with metadata and context intact?
- **Exit** — do all of the above without unreasonable cost or dependency on a third party?

A quick checklist

- ✓ Can we inventory all 20 million objects?
- ✓ Can we identify ROT?
- ✓ Can we apply retention?
- ✓ Can we run classification at scale?
- ✓ Can we export the metadata?
- ✓ Can we delete by governance rule?
- ✓ Can our AI interrogate the corpus?

A “no” or “only at extraordinary cost” answer moves the repository further along the unfriendly hostaging continuum.

The takeaway



- Governance isn't merely knowing what should happen to information. *It is maintaining the enterprise's ability to actually make it happen.*
- The ultimate test of Information Governance is not whether you have policies for your data. *It's whether you (still) have the practical ability to act on them.*



www.valoratech.com



Book a Demo



Webinar Series



[/valora-technologies](https://www.linkedin.com/company/valora-technologies)

Q&A

Thank you



Sandy Serkes
President & CEO
sserkes@valoratech.com



Jennifer Nelson
VP Strategic Solutions
jnelson@valoratech.com