Cleaning for Copilot Readiness

FROM **INTENT** TO **CONSUMPTION**

Infotechtion

About Infotechtion

Smart Data Security and Governance Solutions for Microsoft 365 and Beyond

Our vision is to be a global advisory-of-choice for companies and organisations who desire holistic integrated information Governance maximising the value of information wherever it lives.

Information Protection and Governance Specialization to help customers safeguard their entire organization with integrated security, compliance, retention, and identity solutions.

Data Security includes implementing managed services to manage data security threats leveraging endpoint solutions, information protection, data loss prevention solutions and establish policies to protect insider threats.

Infotechtion is Fasttrack ready and Jumpstart ready partner with existing hands-on experience of preparing early adopter customers to be enterprise-ready for ‘Copilot for M365’.

Microsoft
Solutions Partner
Modern Work
Specialist
Adoption and Change Management

Microsoft
Solutions Partner
Security
Specialist
Information Protection and Governance

Gen-AI represents a new branch of the knowledge management

18%

jump in productivity over the next 12-18 months from GenAI

6.5%

Of their 2025 budget, leaders plan to allocate to GenAI in 2024.

68%

of leaders engaged in a pilot with Copilot confirm the productivity gains possibilities but unsure on the costs to integrate in business flow of work.

Knowledge Creation

Knowledge Extraction

60%

Organisations took 3+ months longer than planned to remediate risks related to copilot.

73%

Organizations with low Information maturity paused their Copilot implementations after a small pilot.

26%

Decided to upgrade to add-on governance solutions to accelerate their information maturity for copilot.

Key Emerging Risk Options to Evaluate.

01

Source of Information to CoPilot.

02

Confident Inaccuracy

03

Content Sprawl

Dark data exposure risks are a significant blocker for customers to move on from intent to consumption.

Why

- Long-term planning and decision-making that affect the overall direction of an organization.
  - **Market Analysis and Forecasting**: Analyze large datasets from various sources, including market reports, news, and social media, to predict trends and provide insights into future market conditions.
  - **Competitive Intelligence**: By processing and summarizing information about competitors, LLMs help organizations understand their competitive landscape, identify threats, and spot opportunities.
  - **Innovation and R&D**: LLMs can assist in generating ideas for new products, services, or business models by analyzing trends, customer feedback, and scientific literature.
  - **Policy and Strategy Formulation**: Organizations can use LLMs to model different strategic scenarios, evaluate potential outcomes, and make informed decisions on policy and strategy.

What

Remove unvaluable stuff

- ROT
- Expired records
- Outdated info
- Not content

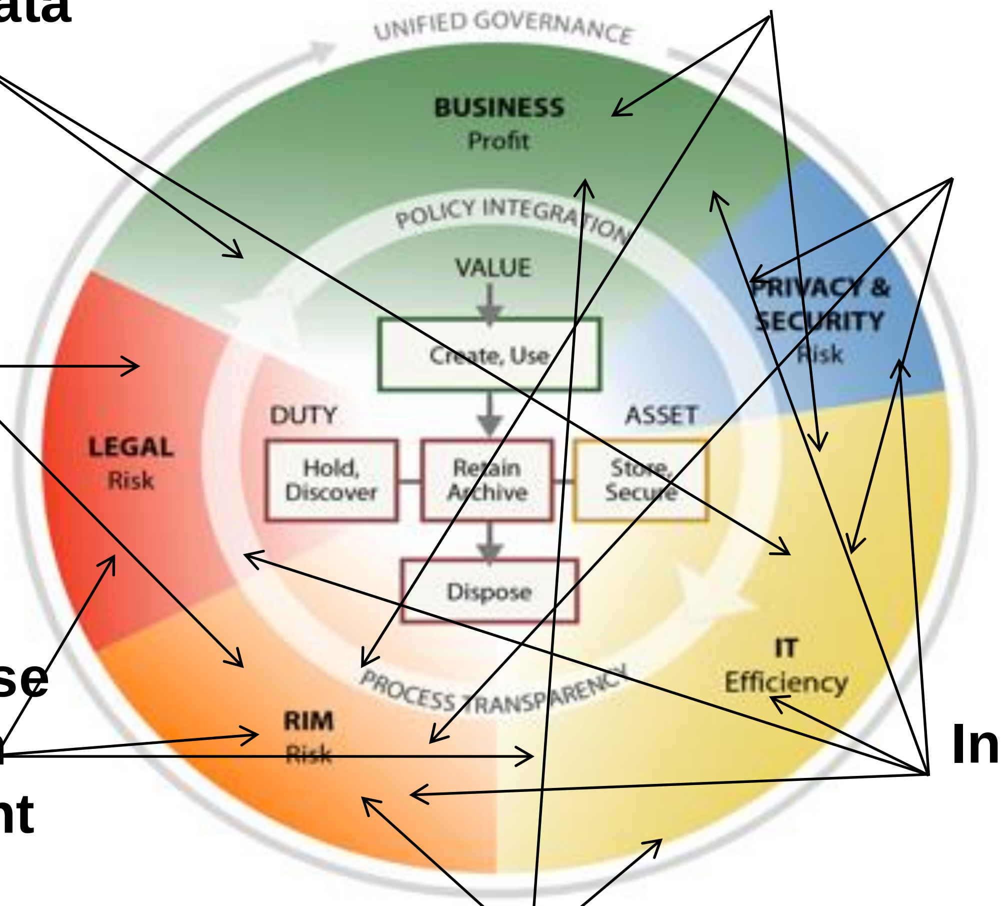
Identify and protect sensitive stuff

- Internally and externally Confidential
- Privacy data

Promote good stuff

- Valuable knowledge
- Accurate transactions

Working Together - IGRM



Copilot Planning and Governance

Big Data

Information Protection

eDiscovery

Normal course information management

Investigation/Audit (internal)

Records Management

Source: IGRM www.edrm.net

Traditional Drivers for Content Cleanup

- Information Management Compliance Risk Mitigation
- M365 Cloud migration
- Reduce cost and level of effort when litigation strikes
- Improve employee productivity
- Enhance access to institutional memory
- Reduce storage costs
- Dispose of eTrash
- Eliminate expired records
- Enhance disaster recovery

- Expand third party crawler software offerings to include proactive e-file clean up
- Find and protect abandoned content, home directories of people no longer in the organization
- Shut down updates of PST files, email archives for retention and risk purposes
- Remove and protect databases on shared drives, isolate dedicated applications or data sets
- Isolate non-business content
- Remove obsolete systems and content

The Real Issue

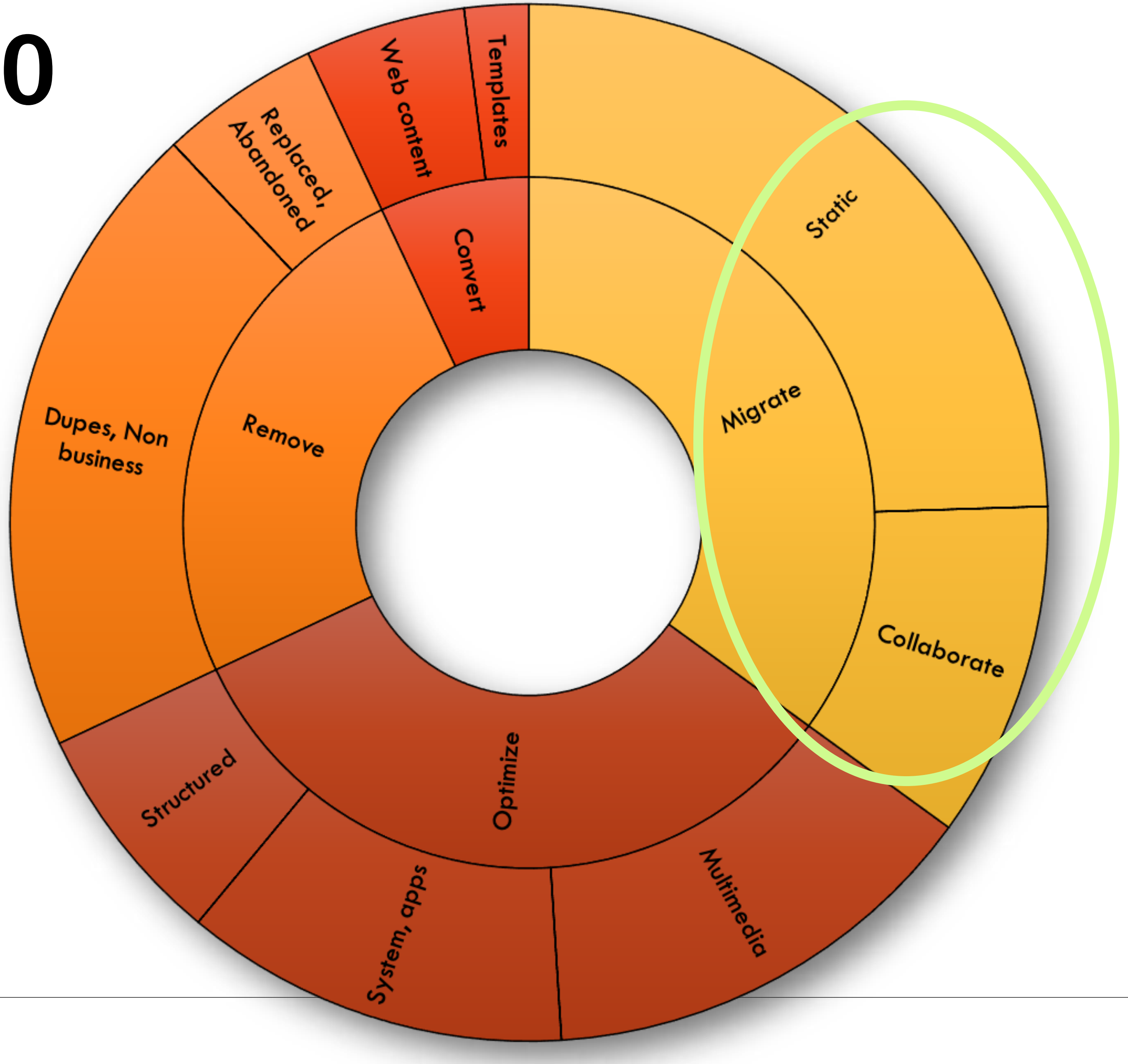
- In large corporate spaces:
  - Out of sight, out of mind – but not copilot
  - No one knows what a record is
  - Information is not uniformly protected
  - Value, importance, accuracy and risk change over time
- Bring the content to the solution and the solution to the content
  - Work is done in M365
  - The same tools you use to evaluate legacy content are what your use to classify future content

Getting to the facts

To govern unstructured data to reduce cost and risk, and add value using a repeatable six stage process:

1. **Identify** and inventory repositories and content to make sense of murky pools of dark unstructured data
2. **Extract** additional information from files to facilitate understanding and classification
3. **Understand** the age, ownership, format, and content of each item; and what that tells you about practices and issues.
4. **Classify** content based on the facts in context to determine whether information is an asset or a liability
5. **Review** rules and results with business users, content owners to get approval for action
6. **Act** and execute on governance decisions: optimize storage systems; classify, migrate and protect data; or make it more readily available to the business

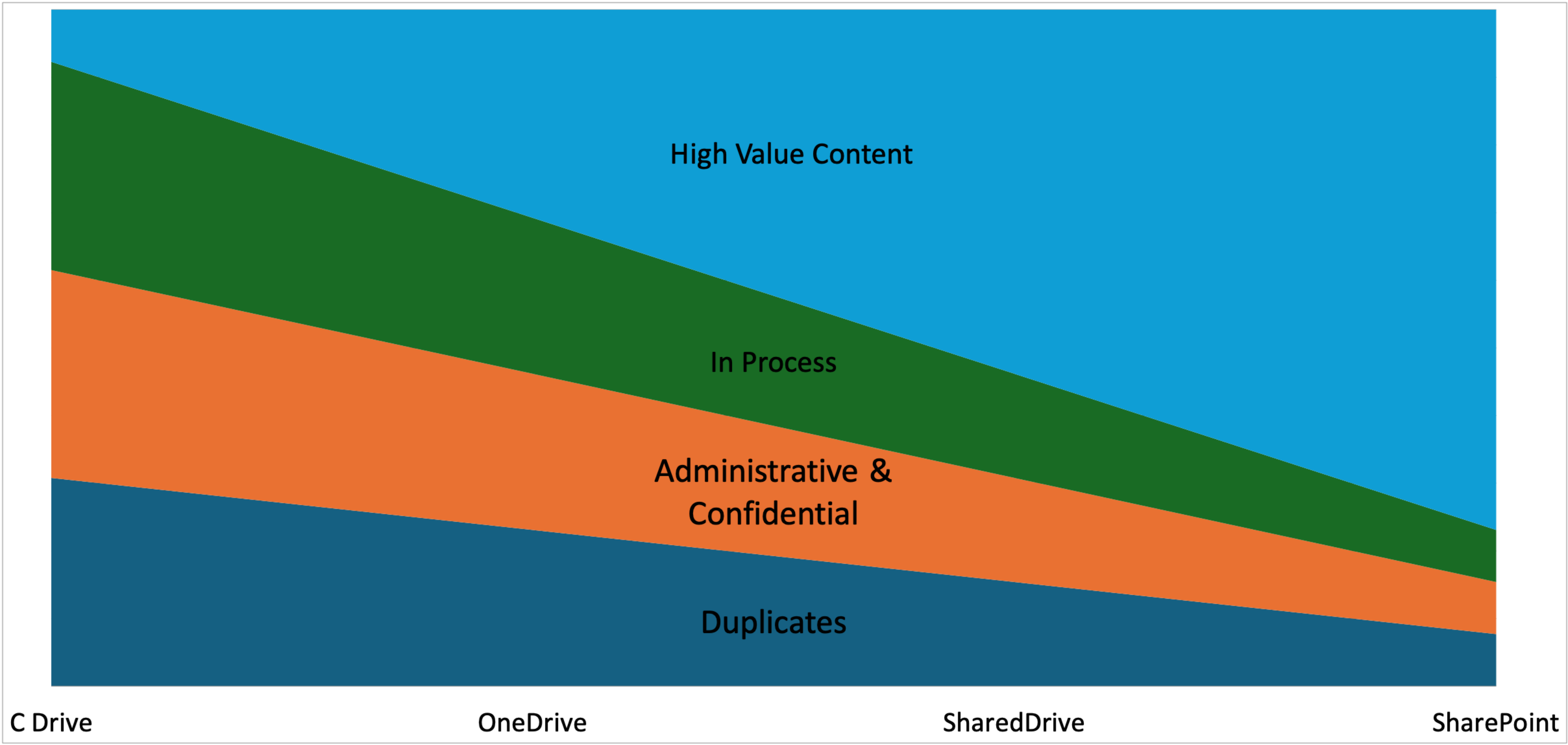
Fortune 20


Templates
Web content
Replaced, Abandoned
Convert
Remove
Dupes, Non business
Static
Migrate
Collaborate
Optimize
Structured
System, apps
Multimedia

Content by Source

	Shared Drive	Exchange	Lift and Shift	SPO
Expired Records	X	X	X	X
Sensitive Info	X	X	X	X
Duplicates	X	XXXX	X	X
Temporary	X		X	
Photos/Media	X	X	X	X
Applications	X		X	
Databases	X		X	
Multiple Formats	X		X	X

Location Personalities



Targeted Remediation

- Understand your starting point
- Determine rules and roles – Expand RRS
- Preserve and Protect legal holds
- Put away apps, databases, compound docs, web content, exceptions
- Isolate protected content
  - Isolate
  - Label
- Cleanup
  - Unilateral – Enterprise wide consistent policy based decisions
  - Uniform – RM Categories, workgroup
  - Unified – Review by the custodian or owner
- **<u>Address larger IG issues</u>**

Defensibility

- Provide systematized deletion process rather than individual or arbitrary process
- Implement consistent approach across the company
- Approve and document standard (reusable) queries
- Approve and document custom queries and results
- Customize company-specific forms to document approvals from Legal, RM, IT and Business Unit
- Isolate and preserve documents subject to litigation holds
- Human validation of query results to document “reasonableness”
- Provide audit trails of work performed and documents deleted

Duplicate factoids

- Duplicates around 20-30% of storage
  - Much of that is also eTrash
  - Average of 3 documents per duplicate set
  - 15% of the duplicates cannot be deleted
  - 85% of the duplicates were created by humans, 15% by systems
  - Storage level deduplication does not benefit productivity, backups, eDiscovery, retention management

The Classification Imperative

1. In RRS, **Identify content that has real content value** – research, strategies, specifications, standards, legal documents, etc.

2. Identify records categories that represent **transactional data that has no re-usable value**, or even potentially harmful data such as privacy information.

3. It may be necessary to keep records around failed projects or activities, but it is not necessary to add them to your AI content. Use metadata, such as product names or numbers, to **separate good quality or successful content from bad quality content** within a specific category.

4. Leverage **sensitivity labels** to improve your AI. A sensitivity label identifying valuable intellectual property should be included, a label identifying privacy information should be excluded

5. Look for **groups or individuals that you know are the "brains"** of the organization and intentionally create content that is high quality (regardless of its record status.)

Takeaways

- Every project has different goals
- Consistent principals
  - Delete, put away, organize, migrate
  - Shared drives are organized – inconsistently
- Content is as valuable as context
- Large volumes take a while to deal with
- Perfection is the enemy of good

Top Information Governance and Security Factors

01

Source of Information to CoPilot. Lack of trust in sources.

02

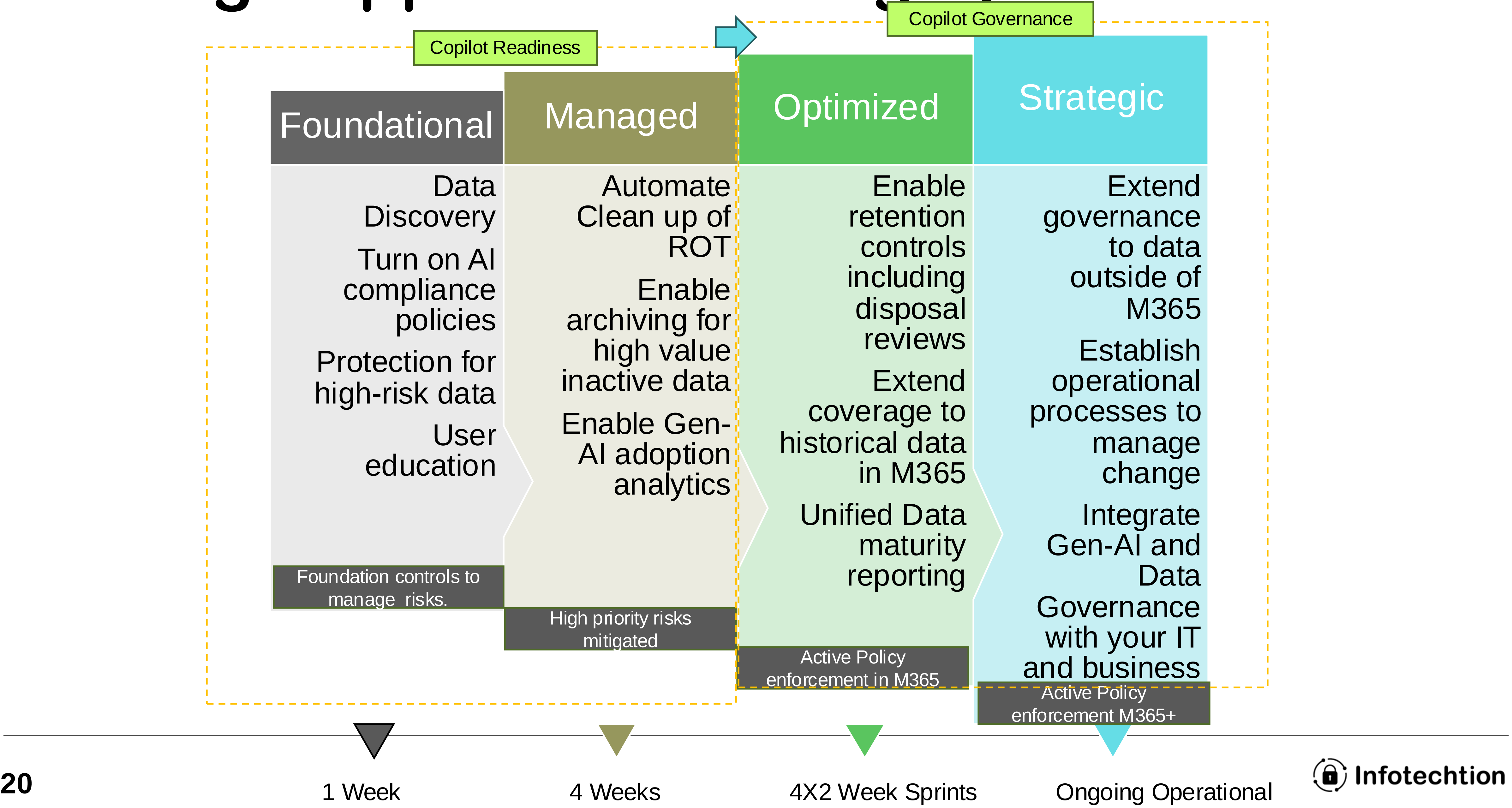
Redundant Obsolete Trivial data led inaccurate results

03

Overexposed Sensitive / Sensitive personal data

Dark data exposure risks are a significant blocker for customers to move on from intent to consumption.

4 Stage Approach – Manage your Risks



Copilot Readiness

Copilot Governance

Foundational

Data Discovery

Turn on AI compliance policies

Protection for high-risk data

User education

Foundation controls to manage risks.

Managed

Automate Clean up of ROT

Enable archiving for high value inactive data

Enable Gen-AI adoption analytics

High priority risks mitigated

Optimized

Enable retention controls including disposal reviews

Extend coverage to historical data in M365

Unified Data maturity reporting

Active Policy enforcement in M365

Strategic

Extend governance to data outside of M365

Establish operational processes to manage change

Integrate Gen-AI and Data Governance with your IT and business

Active Policy enforcement M365+

1 Week

4 Weeks

4X2 Week Sprints

Ongoing Operational

Case Study- LSEG

Demo: i-ARM for Data Governance

SAMPLE OUTCOME

Workspaces Summary

16,356k Sites

5,785k Sites

⚠ *800 new Teams created per quarter is significantly above industry average*

Enable workspace provisioning approvals

Empty Workspaces.

# Sites ⚠

485

Unmonitored sites are top targets for external threats.

Delete these sites.

Policy Violations

# Exfiltration ⚠

221

Associated users work with critical sensitive data.

Add to priority monitoring.

Data Overexposure

# Domains ⚠

2

disclosure, or theft of sensitive data.

Block these domains.

Infotechtion Observation:

32% of SharePoint sites are currently:

- Wasting storage space and resources (this has cost implications)
- Creating confusion and clutter for users who are looking for relevant and useful sites, and information. This was a common concern upon end-users.

Action Plan for Data Governance Leaders

Crawl:

- ***Assess*** the organizational appetite for moving forward with Data Governance.
- ***Investigate*** ownership costs and business case for a Data risks assessment.

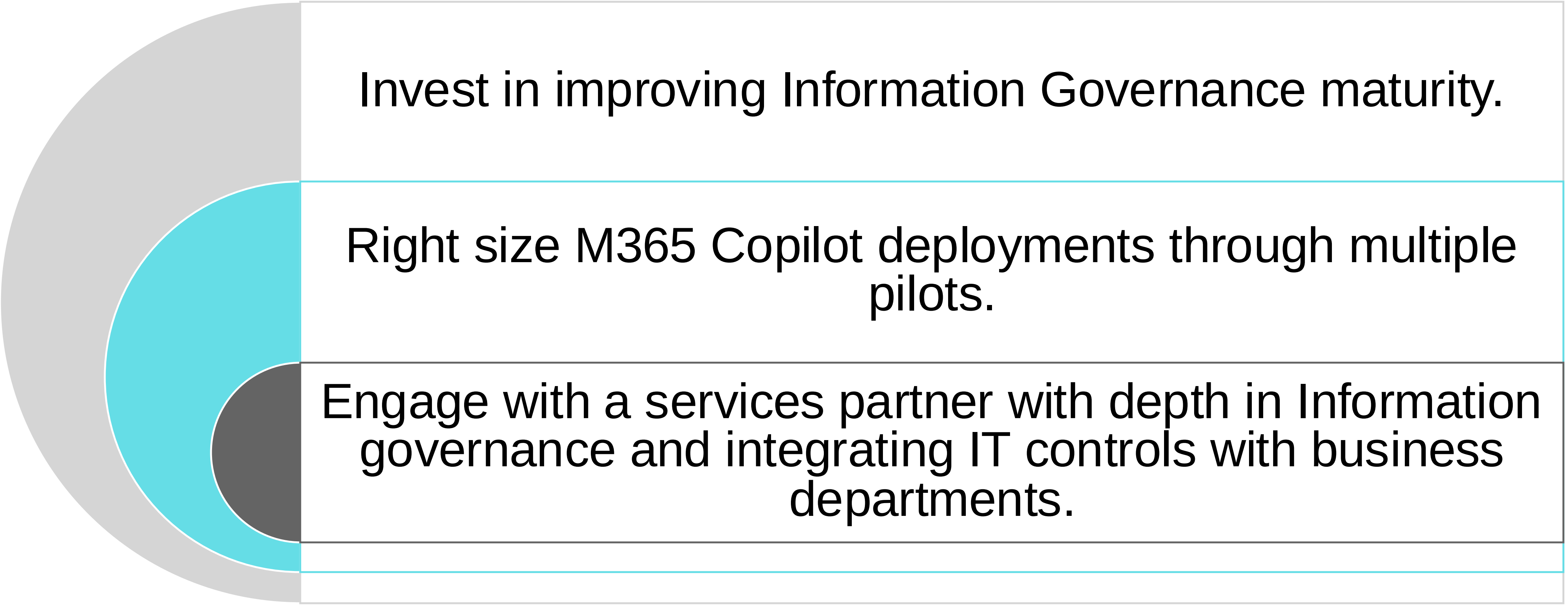
Next 30 days:

- ***Start*** a small pilot to assess data related risks → Risk Score.
- ***Report*** results and *create a remediation plan*.

Next 3 months:

- ***Embed*** a data led process to track, report and remediate your data risks.
- ***Sustain*** the data governance controls to measure the risk improvements over time with adoption of policy-based controls across usage of data by users, Applications and Gen-AI.

Key Recommendations



Invest in improving Information Governance maturity.

Right size M365 Copilot deployments through multiple pilots.

Engage with a services partner with depth in Information governance and integrating IT controls with business departments.

💎 i-ARM value add with advanced readiness path for Microsoft Purview customers.

Call to Action: 1 Week Foundational Governance with i-ARM

M365 Governance Assessment

Identifies the foundational pre-requisites to be established as part of technical readiness.

💎 Workspace Governance

Identify the Sites and Teams pilot users have access explicitly / implicitly to review if any should be included or excluded from the scope of Copilot access.

💎 Data Oversharing

Identifies data in OneDrive for business and SharePoint Sites overshared and implicitly allowing access to Copilot.

💎 Implement Data access and protection controls to filter out data access to Copilot.

Tenant Security Configuration 💎

Identifies changes to your tenant permissions, sharing settings against recommended practices.

Data Loss Prevention

Based on your Current State, recommends foundational DLP configurations to trial with Pilot users.

Legal and Compliance

Recommends Policies to configure for Pilot users to monitor Pilot user interactions with Copilot, report on it for review.

Also includes recommended retention management controls to automatically delete interactions with Copilot.

Advanced security controls to control potential misinformation, bias, copyright violations and other illegitimate or unwanted information generated by Copilot leading to unintended or harmful outcomes 💎

Thank you

✉ VIVEK.B@INFOTECHTION.COM

HTTPS://INFOTECHTION.COM/