M-Files
The Smarter Way to Work.

UNLOCKING THE POWER OF GENERATIVE AI WITH MICROSOFT 365

M-Files

Table of Contents

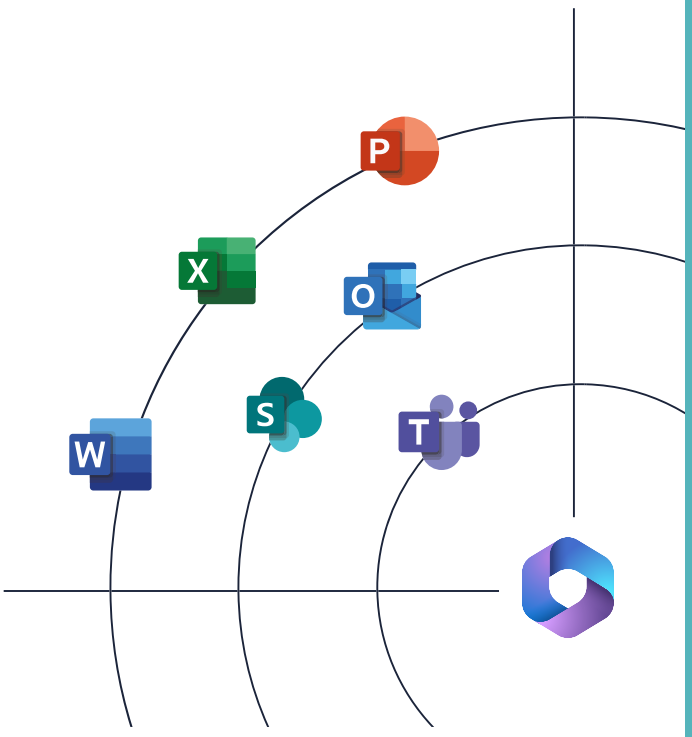
M-Files

INTRODUCTION

The rapid emergence of Generative AI in 2023 has awakened in businesses the need to seriously consider their content strategy and data governance.

Is Generative AI a solution to our content chaos? Can we tap into the value of Generative AI, even though we may have overlooked how documents and data are governed? How far can general purpose office productivity applications —such as Microsoft 365's SharePoint, Teams, OneDrive and now Copilot—take us?

This paper will discuss the challenges organizations face with disorganized content as well as provide insights on solving those challenges as organizations seek to scale new heights of productivity with Generative AI.



M-Files

EXECUTIVE SUMMARY

Most organizations use general-purpose office productivity suites such as Microsoft 365 in their daily operations. These suites provide tools for storing and collaborating on content produced on those platforms.

These tools focus on authoring and collaborating. In addition, governance and control require rigor to achieve and maintain. Most organizations recognize this information challenge, which may cause a loss of employee productivity and increased compliance risks. There's several ways to address these challenges—some more successful than others.

The promise of Generative AI could level up the potential value of organizational knowledge hidden in documents, emails, and conversations. An increasing number of organizations are taking on the challenge of solving this state of content chaos. Successful organizations are addressing the business challenge, the technology challenge, and the people challenge in order to unlock the full value of Generative AI.

M-Files

ORGANIZATIONAL KNOWLEDGE IS HIDDEN IN DOCUMENT REPOSITORIES

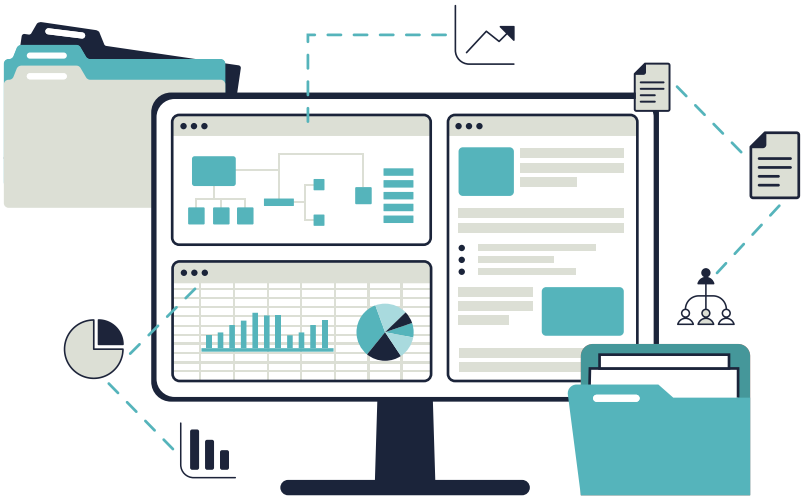
How new knowledge is created with Microsoft 365

Microsoft Office 365 is a leading office productivity suite deploying tools such as Word, PowerPoint, and Excel. Microsoft renamed this package strike Office 365 when migrating these tools to the cloud and started marketing them as a SaaS subscription. Subsequently, Microsoft gradually bundled most of its existing software—SharePoint and Dynamics—as well as new software—Teams and Power BI—into the same cloud subscription and renamed it Microsoft 365.

Today, most organizations globally use Microsoft Word, PowerPoint, and Excel. When a user subscribes to these tools, they automatically gain access to OneDrive, Teams, and SharePoint for content storage and collaboration.

Before Microsoft started bundling office productivity tools, many companies chose SharePoint as their intranet platform or for some other use case. As a result, few organizations globally don't have SharePoint or decide not to use it at all.

What kind of content gets created with Microsoft 365? Perhaps a better question is, what kind of content is NOT created with these office tools? Leaders, managers, and employees at all levels look at numbers in Excel spreadsheets, read reports in Word, prepare presentations with PowerPoint, send emails with Outlook, and collaborate in Teams. Specialized applications are used across many processes, but eventually the output of those applications will be exported into common office formats for others to consume. Even if the origin of data is outside the organization (such as from a customer, partner, or from the Internet) that data is usually stored with the rest of related information—for example in a Teams project folder.



Why the “illusion of easy” leads to problems downstream

When companies subscribe to Microsoft 365, getting started is easy. End users can start authoring documents and saving them to their personal OneDrive. Similarly, when a group of knowledge workers start working together on a project, setting up a new workspace and folder structure in Teams only takes a few minutes or, the IT department has already established a central SharePoint for common documents such as employee handbooks.

Things run smoothly at first—until the company reaches a certain size threshold, hires more new knowledge workers, or simply accumulates a critical mass of information spanning across tens, hundreds, or even thousands of data silos. OneDrive, SharePoint sites, and libraries, Teams channels, folders, or email attachments now store this disparate information. Lack of an information-management plan or the inability to enforce information governance starts to take a heavy toll.

The result?

Daily work becomes chaotic. Finding the right information is difficult, and quickly leads to rework, duplication, version-control issues, and overall frustration. In addition, with varying access-rights protocols across these systems, it’s virtually impossible to know who has access to the correct information or how compliance of regulations or internal policies is maintained, resulting in skyrocketing compliance risks.

What’s the root cause of this problem? Why do organizations lose productivity with a productivity suite?

Simply put, Microsoft 365 is only a collection of tools—applying these tools to solve business problems is the responsibility of the user organization. That is, Microsoft 365 doesn’t implement any specific business processes on behalf of the user organization. It’s a general-purpose toolset. While the Microsoft 365 toolset is easy to use on an individual and a team level, the "illusion of easy" disappears at the organizational level, especially among more regulated industries.

The end result?

Organizations possess a storehouse of valuable insight hidden within documents. Knowledge workers may not know such a storehouse of actionable data exists and they can’t access it. In short, they’re accumulating a vast amount of intellectual capital into their SharePoint libraries, but fail to capitalize and leverage the knowledge hidden inside those documents.

M-Files

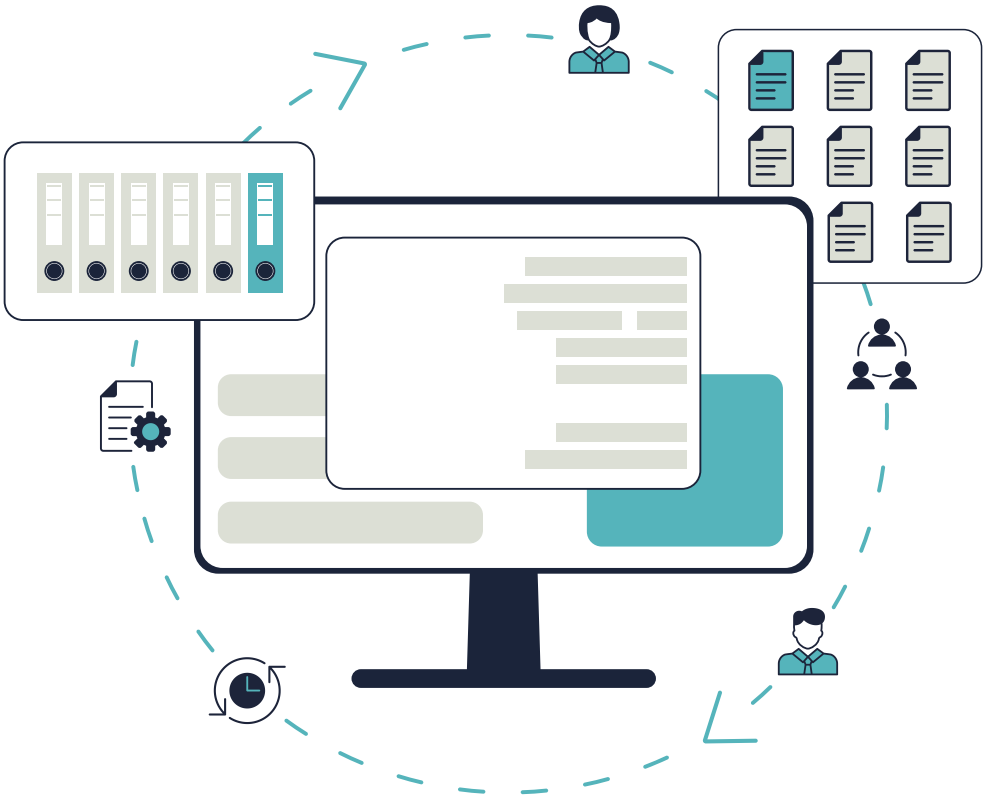
What role do content processes play in creating and leveraging knowledge?

The challenges outlined above relate to the way content is created and consumed. Modern work is rarely performed in a vacuum but, rather, in a team of multiple contributors. Any given document has a specific lifecycle, with various knowledge workers editing, reviewing, and consuming the content of the document across its lifecycle.

Usually, the content processes also change form during the document's lifecycle, starting from highly collaborative processes—where people might even co-author the document—towards increasing control where every change is carefully reviewed and vetted.

At the end of the lifecycle, a document may even be frozen as an immutable record that acts as evidence of a business transaction.

Optimally, workflows travel seamlessly, with every knowledge worker knowing what to do and when to do it, easily finding the information they need and "passing the ball" to the next person in the process. When considering the challenge of disorganized data and applying the additional layer of processes, we begin to see why organizations struggle.



M-Files

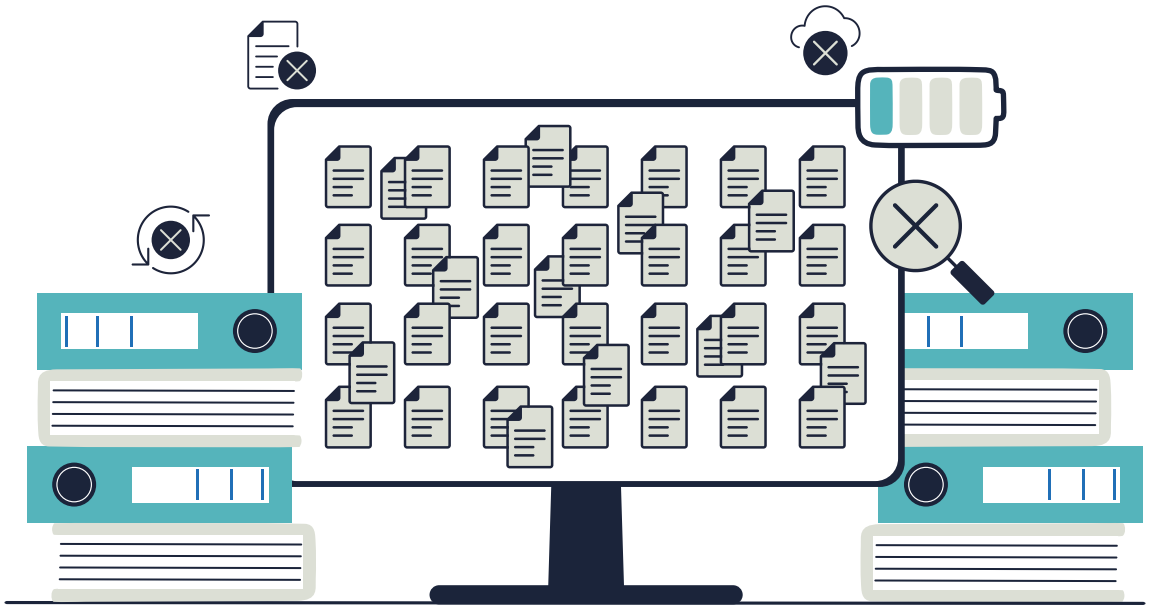
WE RECOGNIZE THE CHALLENGES, BUT FAIL TO FIND A SOLUTION

End-user friction and the path of least resistance

Organizations deploy a diversity of knowledge workers with a variety of backgrounds, competencies, and interests. Not everyone is a technology enthusiast. Not everyone enjoys learning new tools or eagerly changes their way of working as new technologies arise.

If the organization starts with minimal or no information governance, adding new rules, policies, or mandatory steps in daily knowledge worker workflows can be easily perceived as detrimental to productivity. As a result, knowledge workers in charge of producing business results may quickly shut down the process. In general, any new system, process, or change in daily tasks may fail if the knowledge workers don't see a personal benefit and resist adopting it.

Selling change is hard. Many organizations give up and follow the path of least resistance. The attitude? "Let the end users do what they do and try to figure out ways to sort out the issues without our involvement." That's one of the prime reasons organization seek to leverage the capabilities of Microsoft 365—not only for creating and collaborating, but also for governance and control. As we'll discuss in the next chapter, the other reason is IT cost.



M-Files

The inevitability: The build-versus-buy analysis

For any given business challenge, companies must balance between the need and available solutions. There's always the option to build a solution to an exact need or buy existing software that's close enough. Most enterprise software is configurable and can be adapted to the given need—within limits.

Let's look at a company that understands the limitations of the out-of-the-box approach but wants to maximize their existing Microsoft 365 license. What hurdles do they need to overcome to solve their information-management, compliance, and automation challenges?

The first step is to understand the big picture.

- What's the information architecture of the business?
- Who is creating each piece of information?
- How does it flow?
- What are the governance requirements for different types of information?

The next step is to create an information-management plan and the processes to govern it. Finally, the organization must be ready to evaluate implementation and enforcement of that plan.

Deploying such a plan takes some time and effort. **Translating it to Microsoft 365 and SharePoint environments require deep expertise.**

- What's the role of each system?
- Are they capable of storing content?
- How does the company enforce requirements that end-users save the information to the right place across every step?
- What kind of SharePoint site hierarchy should the company have?
- Does the company allow anyone to create new Teams groups?
- Can OneDrive be used to share content?
- How can access rights, term libraries, and retention policies be applied across systems?
- Should Power Automate be used to reduce manual steps?
- How about Syntex and Purview?

Solving this puzzle requires the correct configuration of several interconnected solutions within the Microsoft toolset—not only doing this once, but also maintaining this configuration as the business evolves, as the Microsoft toolset evolves, and as knowledge workers come and go. **That again raises the larger question about investments.**

- Are the required features included in the current licensing tier?
- How long does it take and how much does it cost to build it all?
- Does the company consider these factors to fall under in-house competence or should they outsource the work?

Microsoft 365 is a powerful toolset with significant potential. However, it's also vast, complex, and evolves at a rapid pace, requiring dedicated professionals to build and run it and making the cost-benefit math a necessity before long-term commitments can be executed. Can a company build the desired solution with Microsoft 365 or is there another solution that gets closer to the vision more rapidly? What do the long tail of maintenance and the evolution of business needs look like?

Considering the whole IT landscape

So far, we've mostly examined the issue of organizing documents within the Microsoft 365 environment. However, the average number of applications a knowledge worker uses is 11 and more than 40% of digital workers use more than that. No document lives in isolation. It's always a part of a variety of business processes. How do we build efficient, end-to-end processes that include documents and other IT systems?

That's easier said than done. However, there are tools within Microsoft 365 for that option. SharePoint can be used to manage information in the form of lists. Each column in the list can have custom data types, allowing users to store documents and tag them with additional information potentially related to in-place business processes. It's also possible to retrieve the values of those columns from non-Microsoft systems through integrations. For example, there could be a SharePoint site for archiving contracts—that site could have a list with columns specifying type of contract, contracting parties, and other information used to expose data outside the document. **It's clear that achieving end-to-end automation is possible, but the challenge reverts to the big picture.**

- To what extent are we trying to automate our processes and how many different SharePoint sites and lists do we need to achieve that?
- What's the cost of that implementation project?
- What happens if a new employee is unaware of the company policy regarding the storage of agreements in the contracts archive?
- How can people discover all the various
- processes and practices related to storing information according to the master plan?
- How can AI help?
- Can generative AI and Microsoft Copilot finally solve the problem?

of professional services organizations rely on Microsoft office productivity tools.

Source: The Total Economic Impact™ Of Microsoft 365 E3, Forrester, October 2022

Generative AI alone won't solve the problem

Generative AI can help in solving the problem of information chaos. Large language models (LLM) can harvest insights from vast amounts of data and save a tremendous amount of time that knowledge workers spend searching for the correct information. However, as demonstrated by general purpose chatbots and early attempts at enterprise GenAI pilots, the quality of data has an impact on the quality of the responses by LLMs. In the worst-case scenario, the answer to a question is not based on any real fact but simply generated by the AI. For a GenAI application to provide accurate answers, input given to the AI must be accurate. Summarizing a piece of text is simple because all the needed information is there.

Answering a question based on enterprise data is a whole different ball game. How do we know which information to give to the AI? Can't we just give all the information? That would essentially be the same as asking a search engine to start indexing all data from scratch every time someone performs a search. So real-world GenAI applications that are tasked with harvesting organizational knowledge from enterprise data need to first index that data. Such an index is not your typical keyword counter but a semantic index across all data inside all documents, encapsulating the meaning of each text snippet.

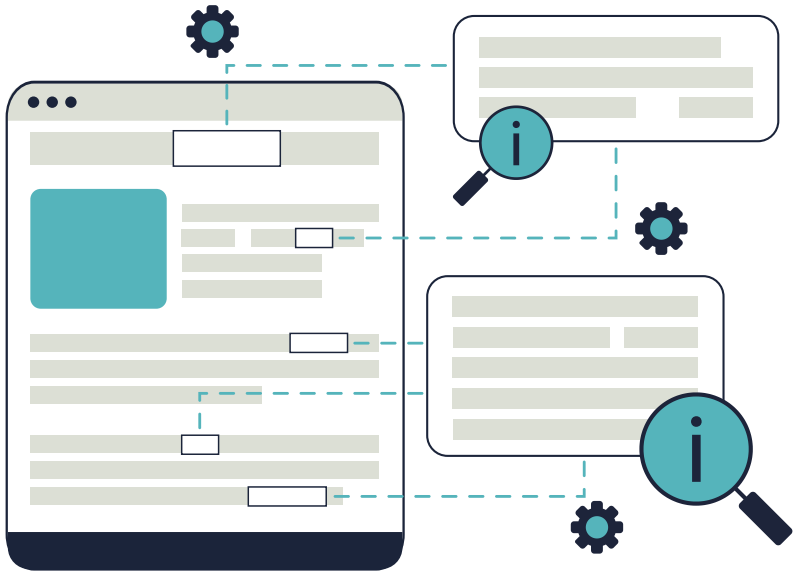
This system then needs to be able to understand the meaning of any question to retrieve the right information from the index. Both requirements are counterproductive in an environment of information chaos.

Analysts estimate that
90% of documents that enterprises store are obsolete, duplicates, or irrelevant to the business.

Indexing all that data will incur an unreasonable amount of new costs and will include content that the shiny new Generative AI should never read. Additionally, if none of the content is classified and governed, the AI won't be able to know which document was a poorly written draft, a test document created by a summer intern, or an authoritative piece of knowledge crafted by a seasoned subject-matter expert. When an employee is looking for good examples of project plans, the results might be influenced by poor-quality content.

The question might also lack proper context—the GenAI may base the answer on a variety of projects plans such as IT projects, marketing projects, and office-relocation projects. If the employee were a business consultant completing projects for clients, they wouldn't appreciate the answer. Even if the prompt would be sufficiently rich in giving context to the AI, there's no guarantee that semantic index could match that context.

To summarize, a successful GenAI application requires well curated source content to build a robust index and it needs the user prompt to be bolstered by additional context. Optimally, the prompt offers an explicit context, such as limiting the answer to specific type of project-plan documents.



M-Files

SUCCESS MEANS ADDRESSING THREE ESSENTIAL DIMENSIONS

Focusing on the essentials for business impact

There are always more ideas than time, attention, and budget allow to execute them. Any attempt to fix information chaos must start with areas that produce the best business impact in the least amount of time. Is there a process, a department, or a business unit that suffers most from these challenges? What is the smallest amount of change the company can institute that leverages the most measurable results? For example, is there a specific type of information that could easily be isolated?

Sometimes, organizations start the change with commercial documents like proposals and contracts. Other times, they relate best to the execution of projects and operations. For some, it's management of assets. Industries and organizations are different but usually the prime candidates can be clearly identified.

Showing tangible results and proving that the approach works will get the attention of company leadership, the buy-in of the business stakeholders, and promote confidence in pursuing the long-term journey.

That doesn't mean organizations don't have to think about the big picture. Quite the opposite—only by cataloguing the businesses, processes, and systems is it possible to evaluate the starting point. Only by having a basic grasp of the overall changes required will executives endorse the first experiments.

There's no shortcut to cleaning up data, but there are best practices in achieving that goal. Organizations need to adopt an approach that prioritizes business impact, makes technology choices analytically, and ensures that end users are on board across the entire journey.

Performing due diligence in choosing the right technology

After the optimal scope has been identified, it's time to address the technology challenge. It may feel overwhelming, especially if there have been prior failed attempts, but in the end it's just analysis work. Any analysis starts with understanding the current state, defining the desired future state, and figuring out execution.

When the scope has been decided, the "as-is" analysis has been halfway done. The problems are known—at least on an anecdotal level—and help guide the deeper, system-level analysis of outstanding issues, thus informing the desired state. Such an analysis generates a list of issues that need to be eliminated and a set of requirements and capabilities designed to address each issue.

- How do we go about classifying documents?
- How do we tie those documents to the relevant business processes?
- What kind of workflows would automate and guide workflow?
- What are the governance requirements related to the lifecycle of each type of document?
- What GenAI use cases can we envision and what requirements arise?

The last piece of the puzzle is about "fit-gap" analysis.

- If these are the requirements, how well does our current solution fit the need?
- What are the gaps and what is the estimated cost of filling those gaps?
- What other solutions in the markets could solve these problems and what would the fit-gap analysis look like for them?
- Finally, what do return-on-investment calculations reveal when we compare project costs, license costs, maintenance costs, and expected productivity gains?

The average number of applications a knowledge worker uses is **11**

and more than **40%** of digital workers use more than that.

Source: Gartner 2023

Generative AI: Getting your people on board

The last dimension to success is people—the end-users. They are the ones who experience both the current situation and the envisioned future, including all the positives and negatives. If they are satisfied with the current state of affairs, it will be harder to convince them to change. If they suffer from information chaos and demand something better, the change will be an easier sell.

In both cases, however, the promise of GenAI is a powerful incentive. If a company can demonstrate the simplicity and power of interacting with natural language and can show a significantly better future situation with GenAI, knowledge workers will accept some level of discomfort with a more structured way of working.

As with any system, process, or workflow evolution, a company will need to involve end users early—they are the ones invested in the early analysis to diagnose issues and choose the right scope. They raise practical issues and question with regards to their current way of working. And they can be engaged to envision and promote a better future.

Once the scope and technology have been finalized and built into the first prototype of the future state, the company must engage end-users in piloting and experiencing the new normal.

By respecting all the feedback from the initial pilot groups and by carefully addressing immediate worries, **the company can start to build confidence in the new normal,** while also growing a group of internal advocates who further sell the solution to the next wave of users.

M-Files

CONCLUSION

Most organizations use office productivity suites such as Microsoft 365 to create and collaborate on content. They also end up managing the content on these platforms which can lead to information chaos. As a result, knowledge workers suffer from poor productivity, security and compliance are at risk, and organizations miss the opportunity to accumulate and reuse organizational knowledge hidden inside their documents.

Organizations look to GenAI to solve information chaos but realize that GenAI also needs well-curated content to produce valuable results. There's no shortcut to cleaning up data, but there are best practices in achieving that goal. Organizations need to adopt an approach that prioritizes business impact, makes technology choices analytically, and ensures that end users are onboard across the entire journey.

Organizations who choose to travel this path will gain a head start in the new information economy where knowledge workers are empowered with the entirety of organizational knowledge and where information technology is, not a barrier, but a booster.

ABOUT THE AUTHOR

Ville Somppi has more than 20 years of experience defining, developing, and deploying technology solutions for knowledge intensive industries. Ville has worked with companies across the globe ranging from large corporations to small start-ups and family-owned businesses. He has hands-on experience deploying, optimizing, and using Microsoft solutions.

At M-Files, Ville is in charge of understanding the needs of target industries and improving product-market fit. Ville holds a masters degree (MSc) in software engineering from Tampere University of Technology.

M-Files®

The Smarter Way to Work.

M-Files is the leading platform for knowledge work automation. With the M-Files platform, knowledge workers can find information faster, work smarter, and achieve more. M-Files features an innovative metadata-driven architecture, embedded workflow engine, and advanced artificial intelligence. This enables customers to eliminate information chaos, improve process efficiency, and automate security and compliance.

For more information, visit **m-files.com**

M-Files has offices in eight countries.
To contact one of our regional offices, visit:
m-files.com/en/contact-us