GenAI and data management for tax

Generative artificial intelligence (GenAI) can unlock data in powerful new ways, but it can't change the fact that good outputs still depend on good inputs. That has made effective data management crucial for companies that want to get the most out of their GenAI investments.

GenAI has revolutionized not only access and processing of data, but also its definition. Data is no longer just the information sitting in select files on a server. It's now also every document, email, chat, notice, filing, policy, law, rule, video, or visualization—it's everything your business can see. That amplifies the need to understand the principles of effective data management as well as the data challenges faced by tax departments when implementing Gen AI.

But first, it's important to understand how tax has changed.

The evolution of your tax department

Tax departments have been facing a rapidly changing landscape, continually asked to do more—and then more again—with less.

While 15 years ago, compliance was the top day-to-day priority, tax departments subsequently added risk management, cost reduction, and process automation as critical factors. Then, the focus shifted again to the value tax is expected to bring to an organization through analytics, insights, and alignment. Part of this process of unlocking business value created an accelerated drive to automation—which has now brought tax departments to focus on GenAI and predictive analytics.

But all these new responsibilities and changing priorities require high-quality data. In its absence, automation just generates bad answers quicker. And as the definition of data expands, so does the complexity of quality control. *(See sidebar on GenAI data hurdles.)*

KPMG survey of 500 C-suite executives: Is your organization currently using GenAI in tax or finance?

- **59%** Yes
- **29%** No, but have plans to do so in next 12 months
- **12%** No, but interested in doing so
- **0%** No, and have no interest in doing so

100% interest is not a data point we find often in our surveys!

Global trends

The pace of change facing today's tax function is unprecedented, requiring new tools and skills even as investment has lagged, workload has increased, and the talent pool is shrinking due to steadily declining enrollment for accounting students. That brings tax departments to four central questions:

1 Do we have the right skills?

2 Do we have the right tools and technology?

3 Do we have an optimized target operating model?

4 **Do we have the right data?**

The tax data problem

Most companies don't realize that their tax departments are among the largest consumers of data in their organization, often leaving tax out of their data and technology discussions. As a result, tax professionals spend more time obtaining and correcting data than actually doing the tax analysis itself—a highly inefficient process with missed opportunities and a fatigued tax department.

That becomes a root cause of the tax data problem.

The tax data problem: The inability to identify, collect, and use data to efficiently support compliance, planning, and opportunity management within the tax function.
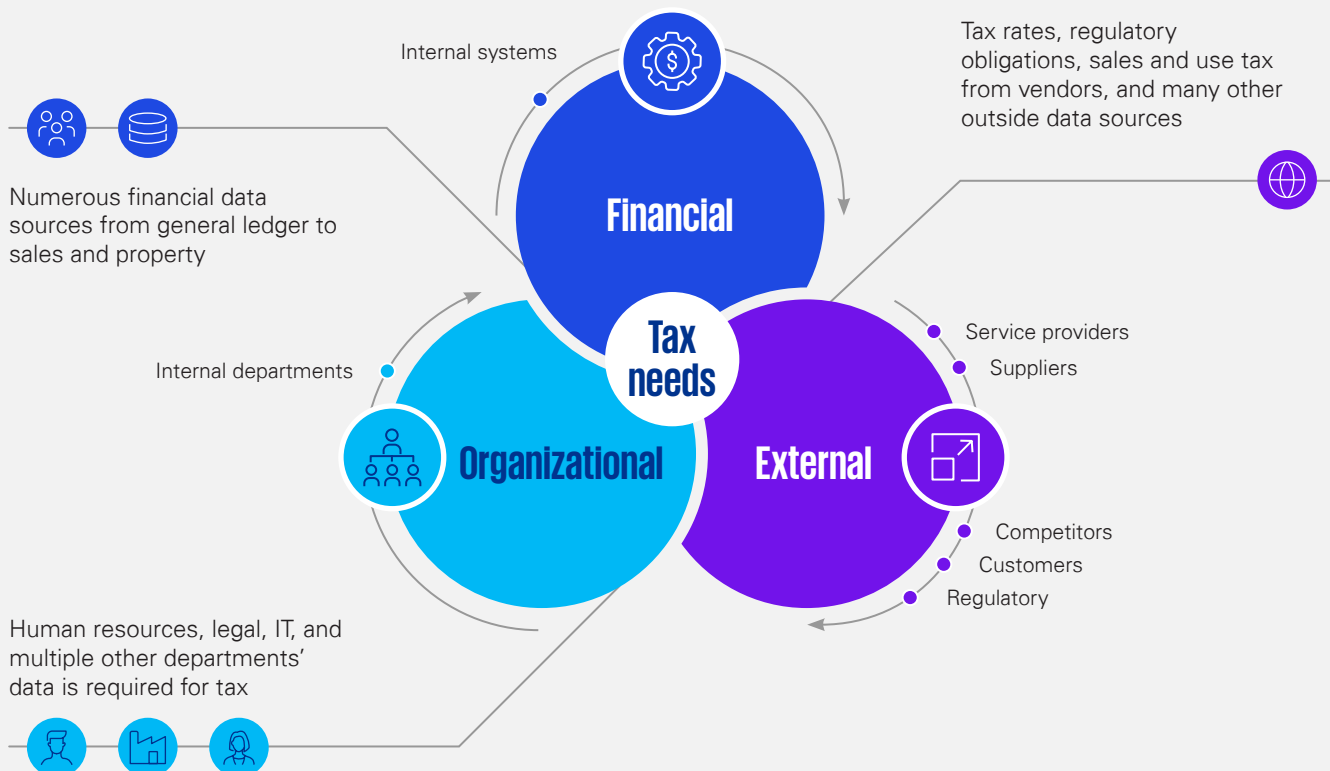
The biggest hurdle is often disparate data. Tax department data needs are vast, fluid, and often scattered across your organization. Some may be controlled by other groups like human resources, information technology (IT), legal, finance, or accounting. And some may reside outside of your organization, where it is maintained by third parties like service providers, business partners, or even regulators. *(See Image 1)*

Finally, there's the day-to-day activity that generates data, including emails, chat messages, workpapers, memos,

deliverables, filings, and more. When you add in the need to collect data multiple times, as well as reconciliation and version issues that creates, the complexity can quickly spiral out of control.

With all that information sitting in various locations, formats, and controls, tax professionals end up spending far too much time scrambling after data and wrangling it to create something that's usable for the tax group. All this becomes amplified once GenAI is brought into the process.

Image 1: Disparate data is the biggest hurdle

Internal systems

Tax rates, regulatory obligations, sales and use tax from vendors, and many other outside data sources

Numerous financial data sources from general ledger to sales and property

Financial

Service providers

Suppliers

Tax needs

Internal departments

Organizational

External

Competitors

Customers

Regulatory

Human resources, legal, IT, and multiple other departments' data is required for tax

Where companies are struggling

Today, most tax departments are struggling to meet the growing requirements they face for both tax reporting and tax planning. New rules are being pushed out and regulators are demanding more granular data, which is driving many companies to data management solutions. Here are the four main pain points we're hearing in the market:

- Our company is getting more complex.
- We are being asked to do more with less.
- We are struggling to comply with new global regulations.
- Our data continues to grow and is more disconnected than ever.

Data management is the key to overcoming these challenges.

How are companies organizing data?

Because it's an ongoing effort, most companies never fully solve their tax data problem. But they are trying to centralize access to data, documents, and other information in ways that range from simply saving all documents on a shared drive to complex architectures where systems and data are fully integrated within a central platform. Most often, companies are using a combination of multiple technologies cobbled together to create a solution.

Whatever the mix of technologies and approaches, here are the things companies need to keep in mind:

Tax data strategy. Your tax data strategy should include GenAI requirements to avoid creating future pain points that don't fit under an overarching strategy.

Alignment with broader organization. Most companies are going to have ongoing data and GenAI projects that are led by finance, IT, even the chief data officer. Try to get a seat at the table to maintain alignment and avoid having to re-create the wheel.

Technology selection and architecture. Select technologies that you're able to integrate into a cohesive solution for your data centralization, while also ensuring that it can integrate with GenAI and has the ability to scale as you grow, add new technology, or bring in new users.

Access controls and security. As you centralize data, make sure only the appropriate people have access to sensitive information. The last thing you want is somebody asking GenAI a question and seeing payroll information as a result.

GenAI data hurdles

While GenAI can significantly improve data retrieval and efficiency, it also comes with its own data hurdles:

1 **Hallucinations.** AI-generated fabrications can lead to the production of convincing but entirely false information, requiring robust verification processes to protect user trust.

2 **Data quality.** The old "garbage-in, garbage-out" principle applies acutely to GenAI, as inconsistent, incomplete, inaccurate, or obsolete data can amplify errors throughout the AI's processing and output generation.

3 **Data security.** GenAI's reliance on large-scale data processing, storage, and sharing increases the risks of data leaks, IP theft, and privacy violations.

4 **Data readiness.** The process of converting diverse, often unstructured data sources for GenAI consumption can require extensive cleansing and integration that understands both the AI model's requirements and the business context.

5 **Governance.** Organizations need to navigate a rapidly shifting landscape of regulations, ethical considerations, and internal policies to ensure data usage and GenAI outputs comply with applicable laws and stakeholder expectations.

6 **Traceability.** The difficulty in mapping the path from input data to output may hinder efforts to validate results, comply with regulations, and explain AI-driven decisions to stakeholders.

7 **Other hurdles.** These include difficulty in incorporating company-specific data, dated inputs, association of unrelated data sets, interpretability, ethical and bias concerns, scalability and performance, vendor and tool selection, cost and return on investment, and the need for continuous learning and adaptation.

Addressing these hurdles requires a strategic approach that includes investing in data management, enhancing security measures, and establishing robust governance practices.

The power of GenAI

GenAI has strongly amplified the existing push within companies to better organize and maintain data. It now goes beyond simply reexamining how they manage data to how they're collecting, analyzing, and streamlining it. And that's where the power of GenAI steps in.

It can alleviate large queries, such as efficiently summarizing new tax legislation, and it can extract and enhance information directly from tax documents that you upload. But perhaps even more powerful, from a data management perspective, is how it revolutionizes access through retrieval augmented generation.

Retrieval augmented generation

With retrieval augmented generation, tax professionals can point GenAI at a specific data set regardless of where it's warehoused. This puts a lot of power into the tax department's hands. Historically, if you wanted access to a large data set you would have to query that information from your enterprise resource planning or other database, which required coding skills that were often lacking within tax. That put tax at the mercy of the IT department's schedule.

But with GenAI, the ability to ask a question in English rather than programming code, while pointing it at a specific data set to gain insights almost instantly, represents a fundamental shift in the tax and finance functions.

Use case #1: Fixed asset depreciation expert

For repetitive tasks like fixed asset depreciation, you can build and save custom instructions and a reusable knowledge base. You can give the AI a set of records, identify assets with keywords such as *repair, replace, remodel, rebuild, renovate,* or *update,* assign it a rule with a value of R&M, specify a useful life, and then add in cost limits—e.g., a useful life of 39 or 40 years, an initial acquisition cost of over $5 million, and a *de minimis* flag for anything below $5,000. This process creates a GenAI persona for a fixed asset depreciation expert.

Now, you can upload and query your fixed asset data as an Excel spreadsheet or other document.

Use case #2: Property tax expert

Property taxes come in a lot of different formats, which can present a challenge. You can set up a property tax expert persona able to extract and structure specific tax-related information from various documents with different formats and labeling. Then you can upload multiple property tax documents. Your GenAI property tax expert persona will then capture the relevant information and put it into a standardized format that includes a description of the property, the address and information for the tax collector, the assessed value, the tax amount, the effective tax rate, the first installment amount, the first installment due date, and other types of information.

Now you can quickly scan this information and export it into an Excel format or some other type of template to pay the bills, enter them into the system, or whatever else you need.

Use case #3: Pointing GenAI at a specific data set

You have 1,000 records reflecting fixed asset data and want to pull up the data on newly acquired fixed assets. You can upload those 1,000 records into GenAI and export them into Excel, organized by acquisition date, asset number, or whatever information you find most relevant. For example, you can ask it to summarize the depreciation by asset class in descending order.

The importance of data security

While all this may seem familiar to people already using publicly available models like ChatGPT, there are several very important differences between internal and external models. By training and querying with your own personas and data sets, you get closer to the information you need from both a legal and internal perspective, without compromising data security.

At KPMG, everything we do is behind firewalls. Knowing your data is secure, you can use GenAI to increase the personal productivity of your tax team. You can optimize your end-to-end business process across both quantitative and qualitative inputs. And you can create significant strategic value by breaking down the boundaries of the tax function, connecting to other areas of the organization, and driving better near-time analysis, more strategic business decisions, and unexpected new opportunities. That's the power of GenAI when it's driven by good data.

Chris Groves
Managing Director,
Ignition Tax—Consulting
E: clgroves@kpmg.com

John Modzelewski
Principal,
Ignition Tax
E: jmodzelewski@kpmg.com

Stephen Spiers
Managing Director,
Ignition Tax—Consulting
E: sspiers@kpmg.com

Patricia Wilson
Managing Director,
Ignition Tax—Consulting
E: pwilson@kpmg.com