

# 18 Ensuring Data Accuracy

For some organizations, AI will open Pandora's box, unleashing the evils of unmanaged data onto the screens of the innocent. This chapter addresses the challenges of data hygiene and steps toward a clean, accurate data estate.

Unlike the last chapter, this one will

- Be less specific to Copilot, and more about data governance as a whole
- Be short!

Because in the AI history books, Copilot will be a short chapter. Preparing the organization for a long future with AI will provide a lasting competitive advantage.

---

## Garbage in, You Know What Out

Copilot is just an AI assistant, not a system of record or data repository. If the data in the Microsoft 365 tenant is already well-managed and governed, Copilot will be optimally efficient. For the rest of us living in the real-world, **Copilot will dig through multiple copies of similar documents, unaware of the source of truth.** It could respond with content from the data that isn't the most useful.

Most times, the Pilot won't know what they are missing. They are likely more concerned with getting helpful information from Copilot rather than absolute data perfection. Copilot can provide value even if the underlying data has minor inconsistencies.

The importance of providing accurate data for Copilot can't be understated in disciplines like healthcare, finance, and law.

Data hygiene and governance is a cultural challenge for most organizations, and a long journey for the few who choose to take it. Those few who do **will have a competitive advantage, especially as data becomes the new currency in the AI age.**

But for those without such discipline, Copilot may be to data governance what a sore toe is on the first day of vacation. If you keep using it without proper treatment, it'll become a major nuisance down the road.

---

## The Problem is Often the People

Data in SharePoint and OneDrive can come from various sources, including manual uploads, automated processes, and downloaded reports from other systems. Additionally, multiple users can edit and modify the same files, potentially introducing inconsistencies or conflicting information. When's the last time you spent any time deleting duplicate photos from 2018, or files from your cloud storage from 2017? Without proper deletion, governance and management, the Microsoft 365 data sources can become bloated, leading to Copilot retrieving outdated or inaccurate information.

Few businesspeople are willing to sit with IT leadership and develop a retention or deletion policy. A pervasive approach is to keep everything, just in case.

---

## The Solution is Data Governance

Establishing a data governance framework is essential for ensuring data quality in Microsoft 365.

This includes:

- Defining clear policies and procedures for data ownership, access controls, and lifecycle management.
- Complying with standards like GDPR and HIPAA.
- Setting retention labels and deletion policies to purge old data from Microsoft 365.
- Archiving files to other data stores if necessary.
- Educating users to improve their data literacy.

That won't happen overnight, due to people, existing processes, and organizational politics. Some suggestions to get started follow.

## The Journey Begins with a Small Step

Organizational culture can't change overnight, and IT cannot make data governance decisions on their own. Yet unfortunately, they will get lots of questions when GenAI produces inaccurate results.

To help decide what's a live/current document, data owners should be named for each department (or SharePoint site). Owners are responsible for overseeing data quality and implementing validation processes. Delegating these individuals should be part of the role of the AI Center of Excellence, discussed in Chapter 28. This requires buy-in from individual managers, which requires buy-in from the top. **To gradually shift your organization to a data-centric culture, it's this kind of distributed data ownership, outside of IT, that will be required.**

If yours is one of the many organizations where folks don't delete anything, consider this trending technique. Some CIOs have been able to move stale data to an archive (like blob storage in Azure) just in case someone asks for it in the future. That keeps the data from being discovered by Copilot (or easily exfiltrated in the event of a breach). That's better than the status quo but doesn't change users' hoarding behavior!

---

## Is this Go-Time?

Will GenAI compel line of business leaders to care more about data quality? Could GenAI convert organizations into a data-driven culture? One where employees understand the importance of data quality and are willing to be trained on and follow data management practices.

It could be now or never. With all the news about the promise of AI, CxOs will be more open than ever to change. It's critical to know the risks of doing nothing. They can all relate to the maxim, "Garbage in, garbage out." That fact will persist into the near future.

Organizations that start to improve their data literacy will have a competitive advantage as data and AI become more integral into organizational strategy. For now, naming data owners and educating people on the "why" behind data sensitivity, deletion, and retention is an effective way to begin the journey.