Copilot-Datenqualität: Dokumente fit machen für Microsoft 365 Copilot

Microsoft 365 Copilot ist nur so gut wie die Inhalte, die er in Ihrem Tenant findet. Das klingt banal, wird aber in fast jeder Copilot-Einführung unterschätzt. Lizenzen werden beschafft, Prompt-Schulungen gebucht, Berechtigungen geprüft, und dann wundern sich alle, warum Copilot souverän aus der Preisliste von 2021 zitiert.

Dieser Kompetenzbereich bündelt alles, was Sie wissen müssen, damit Copilot das Richtige findet: wie Copilot Inhalte auswählt, woran Antworten in der Praxis scheitern, wie Sie Dokumente mit Metadaten anreichern, Altlasten aufräumen und Datenqualität messbar machen. Ohne Beraterlyrik, mit konkreten Werkzeugen und einem Vorgehen, das in Wochen funktioniert, nicht in Jahren.

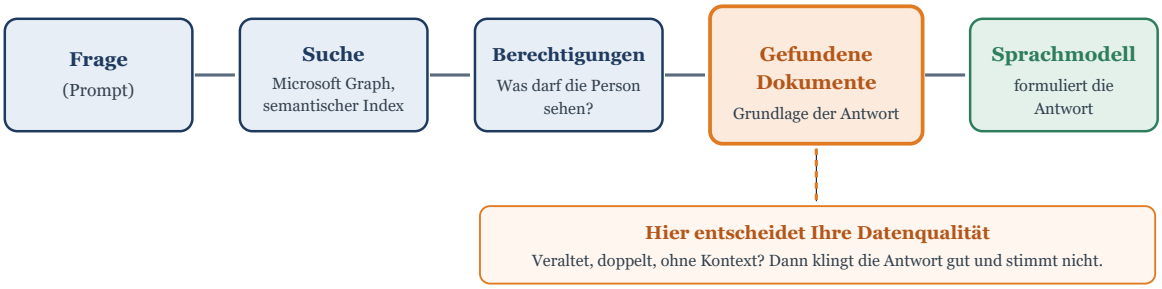
Das Wichtigste in Kürze

- **Copilot sucht, er weiß nichts.** Jede Antwort beruht auf den Dokumenten, die die Suche im Microsoft Graph liefert. Die Qualität dieser Treffer begrenzt die Qualität der Antwort.
- **Copilot behandelt jede Quelle als maßgeblich.** Er unterscheidet nicht von sich aus zwischen der gültigen Richtlinie und dem veralteten Entwurf daneben.
- **Vier Dimensionen entscheiden:** Aktualität, Eindeutigkeit, Kontext und Zugriff. Wer eine davon ignoriert, bekommt unzuverlässige oder gefährliche Antworten.
- **Sie müssen nicht den ganzen Tenant sanieren.** Entscheidend sind die Kronjuwelen, also die Inhalte, auf die sich Copilot-Antworten stützen sollen.

Warum Datenqualität über Copilot entscheidet

Copilot beantwortet Fragen nicht aus einem allwissenden Firmengedächtnis. Er sucht über den Microsoft Graph und den semantischen Index nach passenden Inhalten, filtert sie nach den Berechtigungen der fragenden Person und lässt das Sprachmodell daraus eine Antwort formulieren. Das Modell ist dabei hervorragend im Zusammenfassen, aber es prüft nicht, ob die gefundene Grundlage noch stimmt.

So entsteht eine Copilot-Antwort

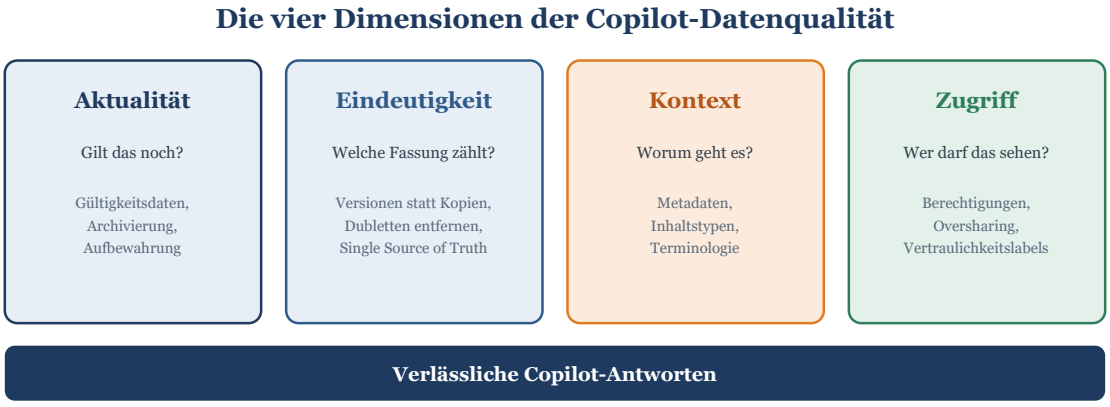


Daraus folgt eine unbequeme Wahrheit: Die meisten „falschen“ Copilot-Antworten sind keine Halluzinationen. Copilot gibt getreu wieder, was er gefunden hat, und das war eben der Entwurf aus dem Vorjahr. Wie das im Alltag aussieht und warum Metadaten den Unterschied machen, zeigt der Einstiegsartikel Copilot und Datenqualität: Warum Copilot nur so gut ist wie Ihre Dokumente.

> **Merksatz**
>
> **Copilot macht aus mittelmäßigen Daten keine guten Antworten.** Er macht aus mittelmäßigen Daten gut formulierte Antworten, und das ist gefährlicher.

Die vier Dimensionen der Copilot-Datenqualität

Datenqualität ist kein diffuses Gefühl, sondern lässt sich an vier Fragen festmachen. Jede Dimension hat ihre eigenen typischen Fehler und ihre eigenen Gegenmittel.



Dimension	Typische Probleme	Gegenmittel
Aktualität	Abgelaufene Preislisten, ersetzte Anweisungen, Projekte von vor fünf Jahren	Gültigkeitsdaten, Archivierung, Aufbewahrungsrichtlinien
Eindeutigkeit	Kopien statt Versionen, Anhänge in Mails und Chats, „final_v3_neu“	Versionierung nutzen, Dubletten entfernen, eine maßgebliche Ablage
Kontext	Nichtssagende Dateinamen, fehlender Status, keine Zuordnung zu Kunde oder Projekt	Metadaten, Inhaltstypen, Terminologiespeicher
Zugriff	Zu breite Freigaben, verwaiste Websites, Links für „Jeder“	Berechtigungen bereinigen, Vertraulichkeitslabels, eingeschränkte Auffindbarkeit

Aktualität: Gilt das noch?

Veraltete Inhalte sind der häufigste Grund für falsche Antworten. Niemand löscht gern, also bleibt alles liegen. Der wirksamste Hebel ist hier oft nicht das Löschen, sondern das

Archivieren: Inhalte, die in Microsoft 365 Archive verschoben werden, nutzt Copilot laut Microsoft nicht mehr für seine Antworten. Sie bleiben aufbewahrt, stören aber nicht mehr.

Eindeutigkeit: Welche Fassung zählt?

Wo es fünf Kopien eines Dokuments gibt, hat Copilot fünf gleichwertige Treffer. SharePoint kann Versionen seit Jahrzehnten sauber verwalten. Das Problem ist die Gewohnheit, Kopien anzulegen und per Mail zu verschicken. Hier hilft nur eine Kombination aus Aufräumen und besseren Ablagewegen.

Kontext: Worum geht es überhaupt?

Ein Dokument ohne Metadaten ist für die Suche nur ein Dateiname mit Text. Status, Gültigkeit, Dokumentart, Kunde oder Projekt machen daraus ein einordbares Dokument. Das ist die Dimension, die am häufigsten übersehen wird, und gleichzeitig die mit dem größten Hebel.

Zugriff: Wer darf das sehen?

Copilot zeigt nur, was die fragende Person ohnehin sehen darf. Was über Jahre zu breit freigegeben wurde, findet er aber zuverlässig. Oversharing ist deshalb die vierte Dimension der Datenqualität, auch wenn sie meist unter Sicherheit läuft. Vertiefung dazu finden Sie im Kompetenzbereich Microsoft Purview.

Die Themenfelder im Überblick

Die Fachartikel dieses Kompetenzbereichs ordnen sich sechs Themenfeldern zu. Sie können der Reihe nach lesen oder direkt dort einsteigen, wo es bei Ihnen gerade hakt.

Symptome erkennen

Copilot zitiert alte Richtlinien, nimmt die falsche Version oder liefert vage Antworten, obwohl die Informationen im Tenant liegen. Hier geht es darum, solche Symptome richtig zu deuten und von echten Halluzinationen oder Berechtigungslücken zu unterscheiden.

Aufräumen und Archivieren

Redundante, veraltete und triviale Inhalte, kurz ROT, verstopfen die Suche. Wie Sie sie finden, was weg kann, was ins Archiv gehört und wie Aufbewahrungsrichtlinien dafür sorgen, dass das Aufräumen keine einmalige Aktion bleibt.

Metadaten und Informationsarchitektur

Spalten, Inhaltstypen, Standardwerte und der Terminologiespeicher: das Handwerkszeug, mit dem Dokumente ihren Kontext bekommen. Dazu die Frage, welche Felder sich wirklich lohnen und welche nach drei Wochen niemand mehr pflegt. Grundlagen dazu im Kompetenzbereich SharePoint.

Anreichern im Alltag und im Altbestand

Metadaten entstehen entweder beim Ablegen oder gar nicht. Für die tägliche Ablage zählt jede gesparte Handbewegung, hier setzt DragPad an: Mail oder Datei auf das Pad ziehen und

die Metadaten in einem Schritt setzen. Für den Altbestand lohnt sich die automatische Ableitung aus Struktur und Inhalt, idealerweise im Zuge einer Migration mit CosyTrack.Migrate.

Copilot Studio und Agenten

Agenten, die auf ausgewählte SharePoint-Bibliotheken zugreifen, verschärfen das Thema. Sie werten pro Antwort nur wenige Suchtreffer aus, und wenn darunter ein veraltetes Dokument ist, zitiert der Agent es mit voller Überzeugung. Kuratierte, gepflegte Wissensquellen sind hier keine Kür, sondern Voraussetzung.

Messen und verantworten

Datenqualität verbessert sich nur, was gemessen wird und wofür jemand zuständig ist. Dazu gehören Kennzahlen wie Metadaten-Füllgrad und Altersstruktur, Testfragen zur Antwortqualität und eine klare fachliche Verantwortung pro Bibliothek.

Wo stehen Sie? Ein einfaches Reifegradmodell

Die folgende Einordnung hilft, den eigenen Ausgangspunkt realistisch einzuschätzen. Die meisten Unternehmen, die Copilot einführen, stehen auf Stufe 1 oder 2, und das ist kein Grund zur Panik, sondern der normale Startpunkt.

Stufe	Merkmale	Folge für Copilot
1 Gewachsen	Ablage nach Gewohnheit, Kopien überall, kaum Metadaten, niemand fühlt sich zuständig	Antworten wirken zufällig, Vertrauen sinkt schnell
2 Strukturiert	Bibliotheken nach Themen, erste Inhaltstypen, Altbestand weitgehend unberührt	Gute Treffer in neuen Bereichen, Ausreißer aus dem Altbestand
3 Angereichert	Kronjuwelen mit Metadaten, Status und Gültigkeit, Altbestand archiviert oder bereinigt	Verlässliche Antworten für die wichtigen Fragen
4 Gepflegt	Anreicherung läuft automatisch, Kennzahlen werden verfolgt, Zuständigkeiten sind gelebt	Copilot wird zur tatsächlichen Wissensquelle

Vorgehen in vier Schritten

Der Weg von Stufe 1 nach Stufe 3 folgt fast immer demselben Muster. Entscheidend ist, mit Fakten zu beginnen statt mit einem Bauchgefühl, und sich zuerst auf die Inhalte zu konzentrieren, die wirklich zählen.



- **Bestandsaufnahme:** Ablagen, Alter, Dubletten, Metadaten-Füllgrad und Freigaben erfassen. Der Copilot Diagnostiker liefert dieses Bild in einem Durchlauf, inklusive Oversharing.
- **Kronjuwelen festlegen:** Welche Inhalte müssen für Copilot-Antworten stimmen? Verträge, Preise, Anweisungen, Produktwissen.
- **Anreichern und aufräumen:** Inhaltstypen und Standardwerte einführen, Altbestand automatisch anreichern, ROT-Inhalte archivieren oder löschen.
- **Freischalten und messen:** Copilot schrittweise freigeben und die Antwortqualität mit festen Testfragen prüfen.

Werkzeuge im Überblick

Für jede Phase gibt es passende Werkzeuge, teils als Bordmittel von Microsoft, teils als spezialisierte Lösungen. Die folgende Übersicht ordnet sie den Aufgaben zu.

Werkzeug	Aufgabe	Wann sinnvoll
SharePoint Advanced Management	Berichte zu Freigaben und Websites, eingeschränkte Auffindbarkeit	Oversharing eindämmen, verwaiste Websites finden
Microsoft Purview	Aufbewahrung, Vertraulichkeitslabels, Löschrichtlinien	Lebenszyklus dauerhaft regeln
Microsoft 365 Archive	Inaktive Websites kostengünstig archivieren	Altbestand aus Copilot heraushalten, ohne zu löschen
SharePoint Premium	KI-gestütztes Ausfüllen von Spalten	Anreicherung, wenn die Lizenzkosten passen
Copilot Diagnostiker	Analyse von Oversharing, Labels und Metadaten-Füllgrad	Bestandsaufnahme vor und nach der Einführung
DragPad	Ablage von Mails und Dateien mit Metadaten in einer Geste	Tägliche Ablage, Massenablagen
CosyTrack.Migrate	Migration mit Anreicherung nach Struktur und Inhalt	Altbestand, Umzug von Fileservern oder alten SharePoint-Versionen

Häufige Fragen

Was bedeutet Datenqualität im Zusammenhang mit Copilot?

Gemeint ist, ob die Inhalte, auf die Copilot zugreift, aktuell, eindeutig, mit Kontext versehen und richtig berechtigt sind. Ist eine dieser vier Dimensionen schwach, leidet die Antwortqualität, auch wenn Lizenzen, Schulung und Technik stimmen.

Halluziniert Copilot, oder liegt es an unseren Daten?

In Unternehmensszenarien liegt es meist an den Daten. Prüfen Sie bei einer falschen Antwort zuerst die angegebenen Quellen. Ist die Quelle veraltet oder ein Entwurf, hat Copilot korrekt wiedergegeben, was er gefunden hat.

Müssen wir vor der Einführung alles aufräumen?

Nein. Konzentrieren Sie sich auf die Kronjuwelen und halten Sie den Altbestand per Archiv oder eingeschränkter Auffindbarkeit heraus. Alles andere kann schrittweise folgen.

Welche Metadaten sind für Copilot am wichtigsten?

In der Praxis bringen Dokumentart, Status, Gültigkeit, Zuordnung zu Kunde oder Projekt und fachliche Verantwortung den größten Nutzen. Wichtiger als die Anzahl der Felder ist, dass sie überwiegend automatisch befüllt werden.

Welche Werkzeuge helfen beim Analysieren und Anreichern?

Für die Bestandsaufnahme der Copilot Diagnostiker, für die tägliche Ablage mit Metadaten DragPad und für den Altbestand, gerade im Zuge einer Migration, CosyTrack.Migrate. Dazu kommen die Bordmittel von SharePoint, Purview und Microsoft 365 Archive.

Wie lange dauert es, bis Copilot spürbar bessere Antworten liefert?

Für einen klar abgegrenzten Bereich, etwa die Vertragsablage oder das Qualitätshandbuch, sind wenige Wochen realistisch. Der Effekt ist sofort messbar, wenn Sie vorher und nachher dieselben Testfragen stellen.

Fazit

Copilot-Datenqualität ist kein Nebenschauplatz der Einführung, sondern ihr Fundament. Wer Aktualität, Eindeutigkeit, Kontext und Zugriff im Griff hat, bekommt einen Copilot, dem die Belegschaft vertraut. Wer sie ignoriert, bekommt einen eloquenten Assistenten mit veralteten Unterlagen und verliert das Vertrauen der Anwender schneller, als es sich zurückgewinnen lässt.

> **Weiterlesen auf boddenberg.de**
>
> - Copilot und Datenqualität: Warum Copilot nur so gut ist wie Ihre Dokumente
> - Copilot Diagnostiker: Oversharing und Datenqualität messen
> - DragPad: Mails und Dateien in einer Geste ablegen und beschriften

- CosyTrack.Migrate: Migration mit Metadatenanreicherung nach Struktur und Inhalt
- Kompetenzbereich Microsoft 365 Copilot
- Kompetenzbereich SharePoint
- Kompetenzbereich Microsoft Purview