Copilot und Datenqualität: Warum Copilot nur so gut ist wie Ihre Dokumente

Über Prompts wird viel geredet, über Lizenzen noch mehr. Über das, was Copilot tatsächlich liest, fast gar nicht. Dabei entscheidet genau das über die Qualität der Antworten: Copilot formuliert brillant, aber er formuliert aus dem, was er in Ihrem Tenant findet. Liegt dort die Preisliste von 2021 neben der aktuellen, drei Fassungen desselben Angebots und ein Vertragsentwurf, der nie unterschrieben wurde, dann bekommen Sie eine Antwort, die überzeugend klingt und trotzdem nicht stimmt.

Dieser Artikel zeigt, warum Datenqualität der unterschätzte Erfolgsfaktor jeder Copilot-Einführung ist, woran es in der Praxis hakt und wie Sie Dokumente so anreichern, dass Copilot sie richtig einordnen kann, ohne Ihre Belegschaft mit Pflichtfeldern zu quälen.

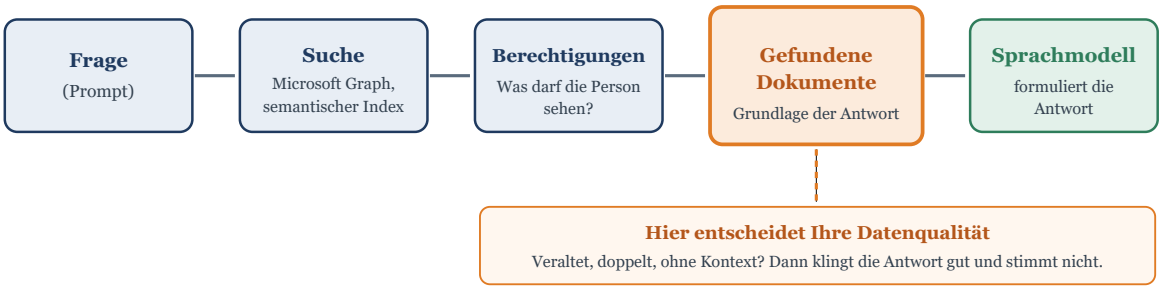
Das Wichtigste in Kürze

- **Copilot erfindet keine Fakten, er sucht sie.** Die Antwort beruht auf den Dokumenten, die die Suche im Microsoft Graph liefert. Schlechte Treffer ergeben schlechte Antworten.
- **Gefährlich ist nicht die offensichtlich falsche Antwort,** sondern die plausibel klingende, die auf einem veralteten Entwurf beruht.
- **Metadaten liefern den fehlenden Kontext:** Status, Gültigkeit, Kunde, Dokumentart. Ohne sie kann Copilot einen Entwurf nicht von der freigegebenen Fassung unterscheiden.
- **Datenqualität und Oversharing sind zwei Seiten derselben Medaille.** Wer das eine prüft, sollte das andere gleich mitprüfen.

Das Missverständnis: Copilot ist kein Hellseher

Viele Unternehmen behandeln Microsoft 365 Copilot wie ein Orakel, das „die Firma kennt“. Technisch passiert etwas viel Nüchterneres: Copilot nimmt Ihre Frage, sucht über den Microsoft Graph und den semantischen Index nach passenden Inhalten, filtert sie nach den Berechtigungen der fragenden Person und übergibt die Treffer dem Sprachmodell. Erst dann wird formuliert.

So entsteht eine Copilot-Antwort



Das Sprachmodell ist dabei der stärkste Teil der Kette. Es fasst zusammen, übersetzt, glättet und klingt immer souverän. Genau das ist das Problem: **Es glättet auch Widersprüche weg, die es gar nicht erkennen kann.** Findet die Suche zwei Preislisten, nimmt Copilot womöglich die falsche und präsentiert sie mit derselben Selbstsicherheit wie die richtige. Das alte Prinzip „Garbage in, garbage out" gilt also weiterhin, nur mit einem gefährlichen Zusatz: Der Müll kommt jetzt hervorragend formuliert wieder heraus.

Woran Copilot in der Praxis scheitert

Die Ursachen sind selten exotisch. Es sind die Altlasten, die sich in jedem gewachsenen SharePoint, jedem Teams-Bereich und jedem OneDrive über Jahre angesammelt haben.

Versionschaos

„Angebot_final.docx", „Angebot_final_v2.docx", „Angebot_final_v3_neu.docx": Für Menschen ist halbwegs klar, welche Datei gilt, meistens jedenfalls. Für die Suche sind das drei gleichwertige Treffer. Dass die Versionierung in SharePoint genau dieses Problem längst gelöst hätte, hilft nichts, wenn die Belegschaft weiterhin Kopien statt Versionen anlegt.

Veraltete Inhalte

Preislisten, Organigramme, Verfahrensanweisungen und Konditionen aus den Vorjahren liegen oft unverändert neben den aktuellen Fassungen. Niemand löscht sie, weil „man das vielleicht noch braucht". Copilot kann ohne ein Gültigkeitsmerkmal nicht erkennen, dass die Anweisung von 2019 längst ersetzt wurde.

Fehlender Kontext

Ein Dokument, dessen einziges Merkmal der Dateiname ist, verrät der Suche wenig. Zu welchem Kunden gehört es? Ist es ein Entwurf oder freigegeben? Wer ist zuständig? Was ein Mensch aus dem Ablageort und seinem Erfahrungswissen ableitet, steht für Copilot schlicht nirgends.

Dubletten und Schattenablagen

Dieselbe Datei als Anhang in Mails, als Kopie im OneDrive, als Upload im Teams-Chat und noch einmal in der Projektbibliothek: Jede Kopie lebt ihr eigenes Leben, wird anders bearbeitet und landet als eigener Treffer in der Suche.

Ein typisches Szenario

Eine Vertriebsmitarbeiterin fragt Copilot nach den Wartungskonditionen für einen Bestandskunden. Copilot antwortet prompt und mit Quellenangabe. Die Quelle ist allerdings ein Angebotsentwurf aus dem Vorjahr, der nie rausging, abgelegt im OneDrive eines Kollegen, der das Unternehmen längst verlassen hat. Der freigegebene Vertrag liegt in der Vertragsbibliothek, trägt aber den Dateinamen „Scan_0047.pdf".

Copilot hat hier nichts falsch gemacht. Er hat das Beste aus dem herausgeholt, was er finden konnte.

Dokumente schlau anreichern: Was Metadaten leisten

Metadaten sind die Beschriftung, die einem Dokument seinen Kontext mitgibt, unabhängig von Dateiname und Ablageort. In SharePoint sind das Spalten, Inhaltstypen und Begriffe aus dem verwalteten Terminologiespeicher. Sie fließen in den Suchindex ein, auf den auch Copilot zugreift, und sie machen aus einer namenlosen Datei ein einordbares Dokument.



Entscheidend ist nicht die Menge der Felder, sondern ihre Aussagekraft. Diese fünf bringen in der Praxis den größten Nutzen:

Metadatenfeld	Zweck	Wirkung für Copilot
Dokumentart	Angebot, Vertrag, Anweisung, Protokoll	Treffer passen zur Frage („Welcher Vertrag …?")
Status	Entwurf, in Prüfung, freigegeben, archiviert	Freigegebenes lässt sich von Entwürfen unterscheiden
Gültig bis	Ablauf von Preisen, Konditionen, Anweisungen	Veraltetes ist als solches erkennbar
Kunde / Projekt	Zuordnung über Bibliotheksgrenzen hinweg	Kontext auch bei nichtssagendem Dateinamen
Verantwortlich	Fachliche Zuständigkeit, nicht der Ersteller	Rückfrageweg bei unklaren Antworten

Ein schöner Nebeneffekt: Metadaten helfen nicht nur Copilot. Ihre klassische SharePoint-Suche, Ansichten, Filter und Aufbewahrungsregeln werden mit denselben Feldern schlagartig besser. Mehr zur Grundlage finden Sie im Kompetenzbereich SharePoint.

Anreichern, ohne die Belegschaft zu quälen

Der Klassiker unter den gescheiterten Metadatenprojekten: zwanzig Pflichtfelder, ein Dialog beim Speichern, und nach drei Wochen steht überall „Sonstiges“. Metadaten funktionieren nur, wenn sie überwiegend automatisch entstehen.

Inhaltstypen und Standardwerte

Legen Sie pro Bibliothek fest, welche Dokumentarten dort hingehören, und hinterlegen Sie Standardwerte für Spalten. Was in der Bibliothek „Verträge Kunde X“ landet, hat Kunde und Dokumentart automatisch, ohne dass jemand etwas eintippt.

Beim Ablegen gleich richtig beschriften

Die meisten Metadaten gehen nicht im Altbestand verloren, sondern jeden Tag neu, beim Ablegen. Wer eine Mail samt Anhang erst speichern, dann hochladen und anschließend in der Bibliothek die Eigenschaften nachpflegen muss, lässt den letzten Schritt verständlicherweise weg. Genau hier setzt DragPad an: Mail, Anhang oder Datei auf das angedockte Pad ziehen, Ziel wählen, Metadaten setzen, fertig. Das ist eine Geste statt fünf Klicks, und gerade bei Massenablagen wie Fotos von der Baustelle der Unterschied zwischen „machen wir später“ und „ist erledigt“.

Aus Struktur und Inhalt ableiten

Vieles steckt bereits in der Ordnerstruktur, im Dateinamen oder im Text selbst: Kundennummern, Datumsangaben, Vertragsbezeichnungen. Ein Skript oder Flow kann daraus Metadaten befüllen, auch rückwirkend für den Altbestand.

Besonders lohnend ist das bei einer Migration, denn dann wird ohnehin jedes Dokument einmal angefasst. CosyTrack.Migrate beherrscht deshalb neben der klassischen Migration auch die Anreicherung nach Struktur und Inhalt: Aus Ordnerpfaden und Dokumentinhalten werden beim Umzug saubere Metadaten, statt dass das alte Ordnerchaos eins zu eins in die neue Umgebung wandert.

KI-gestützt nachbefüllen

Für den großen Altbestand lohnt sich ein KI-Durchlauf, der Dokumente liest und Metadaten vorschlägt. Microsoft bietet dafür Funktionen in SharePoint Premium an, die allerdings zusätzliche Lizenzkosten verursachen. Alternativ lassen sich eigene Lösungen mit einem Sprachmodell bauen, die genau auf Ihre Felder und Begriffe zugeschnitten sind. Wichtig in beiden Fällen: Stichproben prüfen, bevor Sie Tausende Dokumente automatisch beschriften.

Aufräumen statt nur beschriften

Nicht jedes Dokument verdient Metadaten. Vieles gehört schlicht weg: Dubletten, abgelaufene Entwürfe, Ablagen ausgeschiedener Mitarbeiterinnen und Mitarbeiter. Aufbewahrungsrichtlinien in Microsoft Purview sorgen dafür, dass das Aufräumen kein einmaliger Kraftakt bleibt.

> **Faustregel**
>
> **Lieber fünf gepflegte Felder als zwanzig leere.** Ein Feld, das die Belegschaft von Hand ausfüllen muss, sollte einen echten Nutzen haben, sonst wird es nach kurzer Zeit nicht mehr gepflegt.

Vorgehen in vier Schritten

Datenqualität klingt nach Mammutprojekt, ist es aber nicht, wenn Sie gezielt vorgehen. Sie müssen nicht den ganzen Tenant sanieren, bevor Copilot startet. Sie müssen nur wissen, wo die wichtigen Inhalte liegen, und diese in Ordnung bringen.

Vier Schritte zu Copilot-tauglichen Daten



Schritt	Was passiert	Ergebnis
1 Bestandsaufnahme	Analyse von Ablagen, Alter, Dubletten, Metadaten-Füllgrad und Freigaben	Faktenbasiertes Bild statt Bauchgefühl
2 Kronjuwelen	Festlegen, welche Inhalte für Copilot-Antworten stimmen müssen (Verträge, Preise, Anweisungen)	Klare Priorität für die Sanierung
3 Anreichern und aufräumen	Inhaltstypen, Standardwerte, automatische und KI-gestützte Befüllung, Löschen von Altlasten	Copilot-taugliche Kernbestände
4 Freischalten und messen	Copilot schrittweise freigeben, Antwortqualität mit Testfragen prüfen	Nachweisbar bessere Antworten

Für Schritt 1 müssen Sie nicht in Excel-Exporten versinken: Der Copilot Diagnostiker erfasst in einem Durchlauf Oversharing, Freigaben, Sensitivity Labels und den Stand der Metadatenanreicherung und liefert damit genau die Grundlage, auf der Sie die Kronjuwelen festlegen.

Datenqualität und Oversharing gehören zusammen

Copilot hält sich strikt an die vorhandenen Berechtigungen. Das ist beruhigend und gefährlich zugleich: Was eine Person sehen darf, kann Copilot ihr auch zusammenfassen, inklusive der Gehaltsliste, die seit Jahren versehentlich für „Jeder außer externen Benutzern“ freigegeben ist. Datenqualität beantwortet die Frage, ob Copilot das Richtige findet. Oversharing beantwortet die Frage, ob er Dinge findet, die er nie hätte finden dürfen.

Beide Fragen klären Sie am besten in einem Durchgang, denn sie betreffen dieselben Bibliotheken, dieselben Altlasten und dieselben Zuständigen. Hintergründe zu Aufbewahrung, Labels und Richtlinien finden Sie im Kompetenzbereich Microsoft Purview.

Häufige Fragen

Muss ich vor der Copilot-Einführung den gesamten Tenant aufräumen?

Nein. Konzentrieren Sie sich auf die Inhalte, auf die sich Copilot-Antworten stützen sollen: Verträge, Preise, Anweisungen, Produktwissen. Der Rest kann schrittweise folgen oder über Aufbewahrungsrichtlinien auslaufen. Wo Sie anfangen sollten, zeigt eine Analyse mit dem Copilot Diagnostiker in wenigen Tagen.

Liest Copilot Metadaten überhaupt?

Metadaten aus SharePoint fließen in den Suchindex ein, auf dem Copilot aufsetzt. Sie verbessern damit die Auswahl der Treffer und geben Copilot Kontext, etwa ob ein Dokument freigegeben ist. Zusätzlich lassen sich Agenten gezielt auf bestimmte, gepflegte Bibliotheken beschränken.

Reicht nicht eine gute Ordnerstruktur?

Ordner helfen Menschen beim Navigieren, der Suche aber nur bedingt. Ein Dokument kann nur in einem Ordner liegen, aber mehrere Merkmale tragen. Ordner und Metadaten schließen sich nicht aus: Aus der Ordnerstruktur lassen sich Metadaten oft automatisch ableiten.

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 und Purview, die mit sauberen Metadaten erst richtig wirksam werden.

Wer ist für die Datenqualität zuständig, die IT oder die Fachbereiche?

Beide. Die IT stellt Strukturen, Automatismen und Prüfwerkzeuge bereit. Die Fachbereiche wissen, welche Fassung gilt und was weg kann. Ohne fachliche Verantwortung pro Bibliothek bleibt jedes Aufräumprojekt Stückwerk.

Fazit

Copilot ist ein hervorragender Formulierer, aber kein Archivar. Er findet, was Sie ihm hinlegen, und macht daraus eine überzeugende Antwort, egal ob die Grundlage stimmt. Wer Copilot einführt, ohne auf die Datenqualität zu schauen, kauft sich einen eloquenten Assistenten, der mit veralteten Unterlagen arbeitet.

Die gute Nachricht: Mit einer gezielten Bestandsaufnahme, klar definierten Kronjuwelen und überwiegend automatischer Anreicherung bekommen Sie das in Wochen hin, nicht in Jahren. Und Sie verbessern nebenbei Suche, Ablage und Compliance gleich mit.

Weiterlesen auf boddenberg.de

- Microsoft 365 Copilot: Security, Compliance und Governance
- 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 SharePoint
- Kompetenzbereich Microsoft Purview