

UNAUSGESCHÖPFTER WERT: Was jede Führungskraft über unstrukturierte Daten wissen sollte



Holly Muscolino
Group Vice President,
Content Strategies and the
Future of Work, IDC



Amy Machado
Research Manager,
Enterprise Content and Knowledge
Management Strategies, IDC



John Rydning
Research Vice President,
Global Datasphere, IDC



Dan Vesset
Group Vice President,
Analytics and Information Management, IDC

Inhaltsverzeichnis



KLICKEN SIE AUF EINE DER ÜBERSCHRIFTEN, UM DIREKT ZU DIESER SEITE ZU NAVIGIEREN.

Einleitung	3
In diesem Whitepaper	4
Das Versprechen von KI und die Bedeutung von unstrukturierten Daten	5
Ein einheitlicher Ansatz zur Verwaltung unstrukturierter Daten	9
Unzureichend genutzt, unterbewertet, unterfinanziert	9
Komplexität	11
Geschäftsrisiken	12
Compliance-Herausforderungen	14
Produktivität	14
Prioritäten	15
Vorteile	16
Empfehlungen	17
Methodik	19
Demographische Daten	20
Anhang: Ergänzende Daten	21
Über die IDC-Analysten	22

Einleitung

Im Jahr 2022 waren 90 % der von Unternehmen generierten Daten unstrukturiert und nur 10 % strukturiert. In diesem Jahr erzeugten Unternehmen weltweit 57.280 Exabyte unstrukturierter Daten – ein Volumen, das bis 2023 voraussichtlich um 28 % auf über 73.000 Exabyte ansteigen wird. Zum Vergleich: Ein Exabyte entspricht 1 Million Terabyte oder 1 Milliarde Gigabyte. 73.000 Exabyte unstrukturierter Daten entsprechen der Datenmenge von über 97 Billionen sequenzierter menschlicher Genome; sie entsprechen auch der Menge an Videostreams, die ein ganzes Jahr lang täglich rund um die Uhr auf 2,7 Milliarden Bildschirme übertragen werden.

Unstrukturierte Daten, oder Inhalte, sind für Unternehmen geschäftskritisch. Sie sind ein wesentlicher Bestandteil von Lieferketten, wo sie in Bestellungen, Produktbeständen und Import-/Exportdaten enthalten sind. Sie finden sich in Vertriebsvereinbarungen, Marketinginhalten, Verträgen, Patenten, Filmdrehbüchern, Behandlungsvermerken von Patienten. Sie sind in Finanzberichten und Leistungsnachweisen von Mitarbeitern enthalten; sie bilden den Kern von Design- und Konstruktionsdokumenten, Produktspezifikationen und Produktplänen. Sie machen den überwältigenden Teil der menschlichen Kommunikation in Form von E-Mails, Sitzungsprotokollen und -notizen, Präsentationen und Sofortnachrichten aus. Am wichtigsten ist vielleicht, dass der größte Teil unseres Wissens in Form von unstrukturierten Daten erfasst, kuratiert und weitergegeben wird. Inhalte sind daher für den Betrieb eines Unternehmens unerlässlich. Sie ermöglichen es Unternehmen, die Komplexität zu bewältigen, Geschäftsrisiken zu managen und die Produktivität im Zeitalter von Daten und künstlicher Intelligenz (KI) zu steigern.

Das Verständnis der Rolle unstrukturierter Daten in einem Unternehmen ist besonders wichtig im Zusammenhang mit generativer KI (GenKI), die auf große Sprachmodelle (LLMs) angewiesen ist, die wiederum anhand riesiger Mengen von Text, Bildern, Audio und Video trainiert wurden mit dem Ziel, Muster zu erkennen und die Produktivität, Kreativität und Entscheidungsfindung in großem Umfang zu fördern. Führende Unternehmen erkennen zunehmend, dass der Einsatz von GenKI von einer Mischung aus privaten und öffentlich verfügbaren unstrukturierten Daten abhängt.

Mit unstrukturierten Daten sitzen Unternehmen auf einem wahren Goldschatz von Informationen. Laut einer IDC-Studie wird jedoch nur die Hälfte der unstrukturierten Daten eines Unternehmens analysiert, um daraus Wert zu schöpfen, und nur 58 % der unstrukturierten Daten werden nach ihrer ersten Verwendung mehr als einmal wiederverwendet. Wenn alle Daten als Vermögenswert betrachtet werden sollen – wie von vielen Führungskräften behauptet – dann müssen sie auch als solche behandelt werden. Stellen Sie sich die Reaktion der Investoren oder Mitarbeiter vor, wenn ein CEO oder CFO nur die Hälfte der finanziellen Mittel eines Unternehmens produktiv zur Erzielung einer Rendite einsetzen will.



der unstrukturierten Daten werden nach der ersten Verwendung mehr als einmal wiederverwendet.

Dies ist jedoch bei unstrukturierten Daten der Fall, die routinemäßig vergeudet oder vernachlässigt werden, wodurch das Unternehmen Sicherheitsrisiken ausgesetzt und der Möglichkeit beraubt wird, die Produktivität zu steigern, Innovationen mit den neuesten KI-Funktionen zu entwickeln und die Komplexität der Daten zu nutzen, um Kunden zu begeistern und einen Beitrag für alle Interessengruppen zu leisten.

Die meisten Unternehmen befinden sich in dieser Situation, weil sie mit Silos von unstrukturierten Daten konfrontiert sind, die in unzähligen Anwendungen, Tools und Systemen erstellt, repliziert, gespeichert und verwaltet werden. Vielen Unternehmen fehlt ein einheitlicher Rahmen für die Sicherheit unstrukturierter Daten, der ein Gleichgewicht zwischen Governance und Benutzerzugriffsanforderungen herstellt. Andere verlassen sich auf eine manuelle Datenklassifizierung, die zeitaufwendig ist, die Sicherheit der Daten beeinträchtigt und das schnelle Auffinden relevanter Daten behindert.

IN DIESEM WHITEPAPER

Um besser zu verstehen, was führende Unternehmen und ihre Herangehensweise an die Aktivierung und Nutzung unstrukturierter Daten unterscheidet, führte IDC eine von Box unterstützte Marktforschungsstudie durch, in der mehr als 400 Business- und IT-Entscheidungsträger aus großen und mittelständischen Unternehmen aus verschiedenen Branchen und geografischen Regionen befragt wurden. Die IDC-Analysten führten außerdem mehrere ausführliche Interviews mit Box-Kunden durch und stützten sich auf bestehende syndizierte IDC-Studien, wie IDC Global Datasphere, über die Erzeugung, Replikation, Verwaltung und Nutzung von Daten. Die Ergebnisse unserer Forschung zeigen die Vorteile eines einheitlichen Ansatzes für die Verwaltung unstrukturierter Daten als grundlegende Fähigkeit für die Wertschöpfung im Zeitalter der KI.

Das Versprechen von KI und die Bedeutung von unstrukturierten Daten

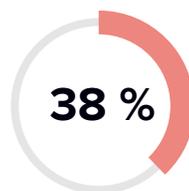
KI ist nicht neu, und viele von uns nutzen sie heute bei der Arbeit, eingebettet in Tools und Anwendungen, die uns bei der Bewältigung unserer täglichen Aufgaben helfen. Klassische KI und ML stoßen auf eine breite Akzeptanz. Allerdings gibt es seit kurzem einen neuen Zweig der KI, die generative KI.

GenKI-Modelle, die auf Basis riesiger, meist unstrukturierter, öffentlich zugänglicher Datensätze angeleitet werden, haben die Vorstellungskraft von Führungskräften und Verbrauchern gleichermaßen geweckt. Das Versprechen dieser Technologie, die Art und Weise, wie wir mit Maschinen arbeiten und interagieren, zu verändern, wirkt sich bereits auf die Entwicklung von kreativen Inhalten, auf der Verarbeitung natürlicher Sprache (NLP) basierenden Chatbots, Text- und Videoanalysen und unzählige andere Anwendungsfälle aus.

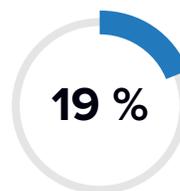
AUS UNSERER STUDIE GEHT HERVOR ...



haben generative KI in ihrem Unternehmen bereits umfassend eingesetzt



haben generative KI in einigen Bereichen eingesetzt



erforschen aktiv den Einsatz generativer KI



erwägen den Einsatz generativer KI, haben aber noch keine Maßnahmen getroffen

Nur 3 % der Befragten erwägen derzeit nicht den Einsatz dieser Technologie.

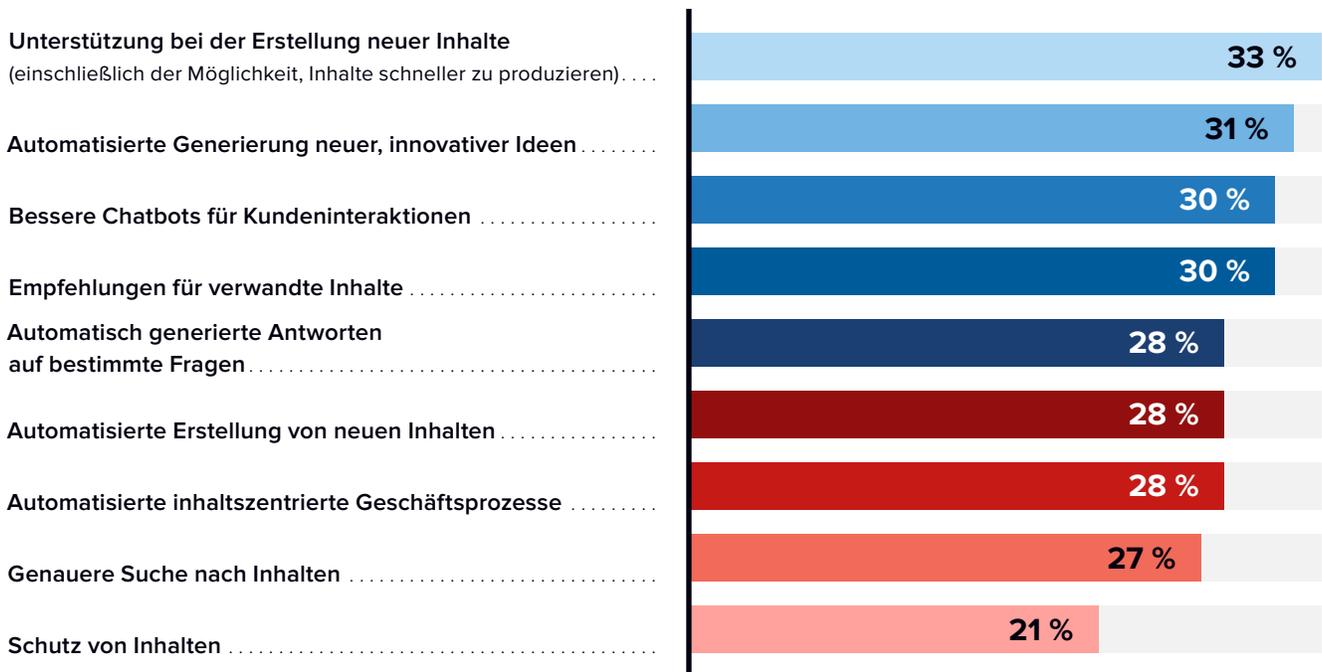
Unternehmen versprechen sich von GenKI mehrere Vorteile. **Abbildung 1** zeigt die Anwendungsfälle mit den größten erwarteten positiven Auswirkungen von GenKI. Ein grundlegendes Thema dieser Vorteile ist die verbesserte Produktivität – bei der Optimierung von Geschäftsprozessen, der Erstellung kreativer Inhalte, der Interaktion zwischen Mensch und Computer, der Schaffung und dem Austausch von Wissen und in anderen Bereichen. Dies ist besonders wichtig in einem Klima, in dem Unternehmen versuchen, mit weniger mehr zu erreichen. In der CEO-Umfrage von IDC gaben im Jahr 2023 61 % der nordamerikanischen CEOs und 78 % der europäischen CEOs an, dass sie erwarten, den „Run“-Teil ihrer IT-Budgets zu reduzieren, um die neuen „Build“-Initiativen im Jahr 2023 zu finanzieren. Unternehmen, die diese neue Chance nicht nutzen, gehen das Risiko ein, gegenüber ihren Wettbewerbern an Boden zu verlieren.

ABBILDUNG 1

Auswirkungen der generativen KI

Wo erwarten Sie die größten positiven Auswirkungen aus dem Einsatz generativer KI?

(% der Befragten)



n = 414, Quelle: IDC, Umfrage zur Ausschöpfung des Potenzials von unstrukturierten Daten, Mai 2023

Einige spezifische Anwendungsfälle von GenKI:

- Identifizierung riskanter Klauseln in einem Vertrag
- Auslösen eines Echtzeit-Geschäftsprozesses über KI-extrahierte Metadaten
- Mehr Sicherheit im großen Maßstab durch KI-Automatisierung anstelle manueller Eingriffe
- Schnelle Suche und Validierung von Finanzdaten von Drittanbietern, um festzustellen, ob ein potenzieller Kunde für ein Angebot in Frage kommt
- Beurteilung der Absicht und der Tonalität einer E-Mail, um den richtigen Arbeitsablauf – automatisiert oder menschlich – für ein besseres Kundenerlebnis zu bestimmen

In dem Maße, wie die Begeisterung über das Potenzial von GenKI wächst, nimmt auch die Erkenntnis zu, dass ihre Effizienz von der strategischen Verwaltung der unstrukturierten Daten eines Unternehmens abhängt. Sowohl GenKI- als auch klassische KI-Anwendungsfälle hängen von verschiedenen KI-Modellen ab, die auf einer Mischung aus internen, proprietären unstrukturierten Daten und externen offenen Daten angelehnt werden. Ganz gleich, ob es darum geht, neue Modelle aus langen Text-, Video- oder Bilddateien anzulernen, wird erwartet, dass Technologieführer eine Strategie entwickeln, um alle verfügbaren unstrukturierten Daten auf sichere Weise zu identifizieren, zu klassifizieren und zu vereinheitlichen und sie KI- und Data-Science-Experten zur Verfügung zu stellen, ohne dass die Gefahr besteht, dass geistiges Eigentum verloren geht.

Tatsächlich ist das von den Teilnehmern unserer Umfrage am häufigsten genannte Hindernis für die Einführung von GenKI die Sorge um die Freigabe unternehmenseigener Inhalte für die großen Sprachmodelle der GenKI-Technologieanbieter. Hinzu kommt die Unklarheit bezüglich der Rechte am geistigen Eigentum an den Inhalten, die im Rahmen der automatischen Generierung neuer Inhalte durch GenKI verwendet werden, sowie der Umgang mit der Wahrnehmung der Mitarbeiter hinsichtlich der Automatisierung bestehender Arbeitsfunktionen.

Das von den Teilnehmern unserer Umfrage am häufigsten genannte Hindernis für die Einführung von GenKI ist die Sorge um die Freigabe unternehmenseigener Inhalte für die großen Sprachmodelle der GenKI-Technologieanbieter.

Wie bei jeder neuen Technologie gibt es auch hier Herausforderungen und Bedenken. Auf die Frage nach den größten Hindernissen für die Einführung von GenKI gab die Hälfte (49 %) der Befragten in unserer Studie an, dass sie Bedenken haben, die firmeneigenen Inhalte ihres Unternehmens für die großen Sprachmodelle der GenKI-Technologieanbieter freizugeben, und fast die Hälfte (47 %) der Befragten führten an, dass es keine Klarheit über die geistigen Eigentumsrechte an den Inhalten gäbe, die zum Anlernen der großen Sprachmodelle verwendet werden. Das am dritthäufigsten genannte Hindernis für die Einführung von GenKI ist der Umgang mit der Meinung der Mitarbeiter über die Automatisierung bestehender Arbeitsfunktionen (41 %).

Um diese Bedenken auszuräumen, bieten Technologieanbieter mit Lösungen, die GenKI beinhalten, folgende Ansätze:

- ✓ Verpflichtung zu Transparenz über KI-Praktiken, Technologie, Anbieter und Datennutzung
- ✓ Vollständige Kundenkontrolle über KI-Nutzung, Daten und Prozesse. Kunden können die Nutzung von KI aktivieren oder deaktivieren und entscheiden, ob KI auf ihre unstrukturierten Daten angewendet werden soll
- ✓ Gewährleistung, dass keine LLMs ohne ausdrückliche Genehmigung mit Kundendaten angelernt werden
- ✓ Vermittlung eines klaren Verständnisses der Funktionsweise ihres KI-Systems und der Logik hinter den KI-Ergebnissen für Nutzer, um den Kontext zu verdeutlichen
- ✓ Einsatz von vertrauenswürdigen KI-Modellen, die die Genauigkeit, Zuverlässigkeit und Sicherheit von KI-Lösungen unterstützen
- ✓ Sicherstellung, dass KI-Systeme die gleichen Kontrollen und Datenzugriffsrichtlinien einhalten, die den Zugriff auf unstrukturierte Daten innerhalb ihrer Plattform und ihres Systems insgesamt bestimmen: Dazu gehört auch der Schutz von Kundendaten durch die Implementierung robuster Sicherheitsprotokolle, einschließlich Verschlüsselung

Eine grundlegende Fähigkeit, um GenKI oder andere KI-Modelle zur Gewinnung von Mehrwert aus unstrukturierten Daten zu ermöglichen, ist eine einheitliche, kontrollierte, sichere und zugängliche Plattform für alle unstrukturierten Daten.

Ein einheitlicher Ansatz zur Verwaltung unstrukturierter Daten

Wir haben unsere Studienteilnehmer gefragt, welche Auswirkungen es auf ihr Unternehmen hätte, wenn sie eine Software (oder einen Cloud-Service) zur Verwaltung aller unstrukturierten Daten in einer einheitlichen, kontrollierten, sicheren und zugänglichen Plattform implementieren könnten.

92 Prozent der Befragten sagten uns, dass dies sich mäßig bis äußerst positiv auf Innovation und Kosten auswirken würde. 55 Prozent der Befragten erwarten eine mäßige bis äußerst positive Auswirkung auf die Sicherheit, während weitere 25 Prozent eine leicht positive Auswirkung auf die Sicherheit erwarten. Die meisten Unternehmen stehen jedoch vor der Herausforderung, alle ihre unstrukturierten Daten zu vereinheitlichen.

Unzureichend genutzt, unterbewertet, unterfinanziert

Unsere laufenden Untersuchungen zeigen, dass die Ausgaben für Technologien zur Nutzung unstrukturierter Daten geringer sind als die Ausgaben für strukturierte Daten, obwohl das Volumen und die Vielfalt unstrukturierter Daten weitaus größer ist als die strukturierter Daten.

- 55 Prozent sagen, dass weniger als die Hälfte aller unstrukturierter Daten zwischen Mitarbeitern oder Systemen ausgetauscht werden.
- 41 Prozent geben an, dass weniger als die Hälfte aller unstrukturierter Daten wiederverwendet werden (d. h. mehr als einmal nach der ersten Verwendung abgerufen werden).
- 22 Prozent der unstrukturierter Daten werden unnötigerweise repliziert, weil Unternehmen einfach nicht wissen, welche Daten sie haben oder wie sie diese finden können.

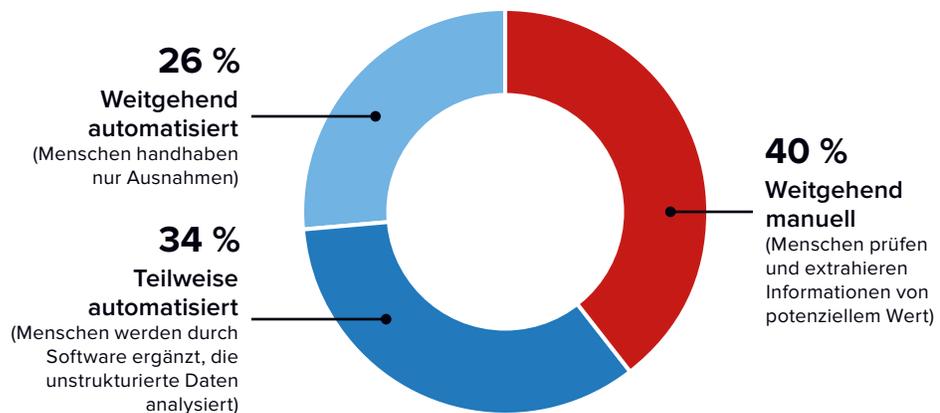
- 46 Prozent geben an, dass weniger als die Hälfte aller unstrukturierten Daten analysiert wird, um daraus einen Nutzen zu ziehen. Selbst dann, wenn unstrukturierte Daten analysiert werden, geschieht dies zu oft manuell, wie in **Abbildung 2** zu sehen ist.

ABBILDUNG 2

Methoden zur Analyse von unstrukturierten Daten

Wie hoch schätzen Sie den Anteil der unstrukturierten Daten, die mit den folgenden Methoden analysiert werden?

(% der Befragten)



n = 414, Quelle: IDC, Umfrage zur Ausschöpfung des Potenzials von unstrukturierten Daten, Mai 2023

Es besteht eindeutig die Möglichkeit einer zusätzlichen Automatisierung, die nicht nur die Produktivität verbessert, sondern es Unternehmen auch ermöglicht, den Wert unstrukturierter Daten besser zu nutzen.

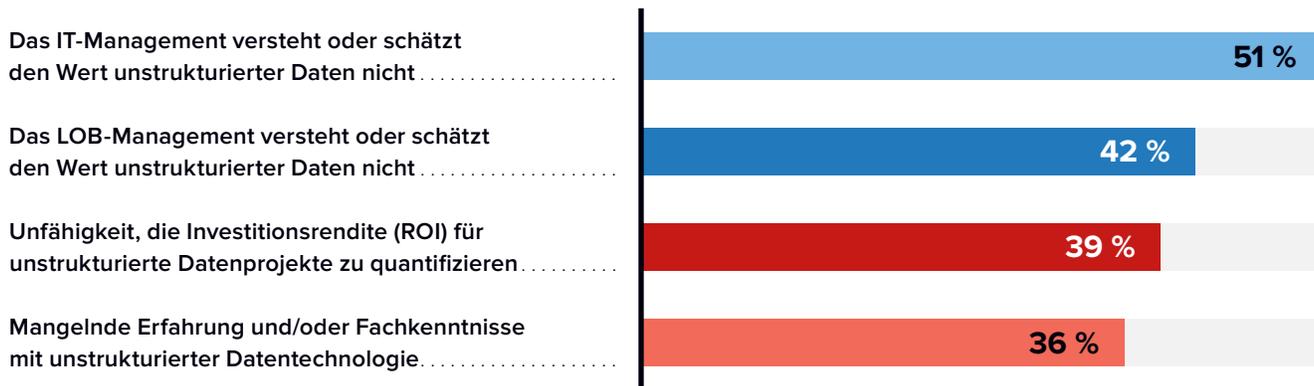
Diese Kennzahlen deuten darauf hin, dass zu viele der potenziell wertvollen Ressourcen in Form von unstrukturierten Daten verschwendet oder vernachlässigt werden. Sie generieren weder einen Return on Asset (ROA) noch einen nachgelagerten ROI.

Dennoch sagen nur 44 % der Unternehmen, dass es einfach ist, die Finanzierung von Projekten zu rechtfertigen, die eine neue oder bessere Nutzung unstrukturierter Daten beinhalten. Die größten Hindernisse für die Rechtfertigung und Finanzierung von Projekten zur Nutzung unstrukturierter Daten sind in **Abbildung 3** (nächste Seite) aufgeführt.

ABBILDUNG 3

Hemmnisse für die Finanzierung von Projekten zur Nutzung von unstrukturierten Daten

Was sind die größten Hindernisse bei der Rechtfertigung und Finanzierung von Projekten zur Nutzung unstrukturierter Daten? (Bitte alle zutreffenden Angaben auswählen)
(% der Befragten)



n = 414, Quelle: IDC, Umfrage zur Ausschöpfung des Potenzials von unstrukturierten Daten, Mai 2023

Auf der Grundlage der laufenden Untersuchungen von IDC zu Technologieausgaben schätzen wir, dass nur 40 % der Gesamtausgaben für „Daten“-Technologien für Projekte und Initiativen mit Schwerpunkt auf unstrukturierten Daten verwendet werden. Es sind mehr finanzielle Mittel erforderlich, da Unternehmen sowohl vor Herausforderungen stehen als auch nach Möglichkeiten suchen, aus unstrukturierten Daten, die 90 % aller Daten ausmachen, Nutzen zu ziehen. Um die Chancen und den Wert unstrukturierter Daten zu nutzen, müssen Unternehmen vier Faktoren angehen, die sich auf unstrukturierte Daten auswirken und von ihnen beeinflusst werden: Komplexität, Geschäftsrisiken, Compliance-Herausforderungen und Produktivität.

Komplexität

Die Hälfte (50 %) unserer Umfrageteilnehmer sagte uns, dass die unstrukturierten Daten ihres Unternehmens größtenteils oder vollständig in Silos gespeichert sind. Ein Viertel (25 %) der Unternehmen schätzt sich selbst als unzureichend darin ein, alle Quellen von unstrukturierten Daten im Unternehmen zu kennen und/oder zu katalogisieren. Das ist nicht verwunderlich, wenn man bedenkt, dass der typische Angestellte 37 Software-Tools für seine tägliche Arbeit nutzt. 70 % davon werden verwendet, um unstrukturierte Daten zu erstellen, zu konsumieren oder zu bearbeiten. Das führt zu einer erheblichen Komplexität in der Arbeitsumgebung.

Das am häufigsten genannte Problem für Unternehmen war die Tatsache, dass die Anzahl der Verbindungen oder Beziehungen zwischen verschiedenen unstrukturierten Datenquellen und/oder Repositorien schneller wächst als die Fähigkeit der Mitarbeiter, diese zu nutzen, zu verarbeiten und zu verwalten. 28 Prozent der Umfrageteilnehmer gaben an, dass die schnelle

Zunahme von Verbindungen oder Beziehungen zwischen verschiedenen unstrukturierten Datenquellen und/oder -speichern eine der größten Herausforderungen für sie darstellt, während 40 Prozent der Befragten Schwierigkeiten bei der Integration von Technologien für unstrukturierte Daten mit anderen Technologien haben.

Auch die Geschwindigkeit und das Volumen der Daten sind von Bedeutung. 25 Prozent der Befragten sagten uns, dass das Datenvolumen schneller wächst als ihre Fähigkeit, es zu nutzen, zu verarbeiten und/oder zu verwalten. 27 Prozent der Befragten gaben an, dass die größte Herausforderung darin besteht, dass die Geschwindigkeit unstrukturierter Daten schneller zunimmt, als sie diese verarbeiten können, und weitere 20 Prozent führten auf, dass die größte Herausforderung das schnelle Wachstum der Vielfalt unstrukturierter Daten sei. Ein leitender Verwaltungsangestellter einer großen Versicherungsgesellschaft sagte uns zum Beispiel: „Wenn [unsere Kunden] in einen Unfall verwickelt werden, können sie hochauflösende 4K-Videos aufnehmen ... [die Daten] werden immer größer und komplexer.“ Aber wir dürfen nicht vergessen, dass eine einheitliche Plattform die Verwaltung unstrukturierter Daten vereinfachen kann, auch wenn alle wachsenden Systeme durch ihre Expansion immer komplexer werden. Kurz gesagt: Komplexität muss nicht kompliziert sein.

Geschäftsrisiken

Die ausufernden Anwendungen und die häufig daraus resultierende Fragmentierung unstrukturierter Daten schaffen zusätzliche Sicherheits- und Compliance-Risiken für Unternehmen. Sowohl die Ausbreitung als auch die Fragmentierung führen zu mehr potenziellen Angriffsflächen, die gesichert werden müssen – oft mit verschiedenen Identitäts- und Authentifizierungsmodellen und unterschiedlichen Verwaltungsfunktionen, die weitere potenzielle Fehlerquellen darstellen.

Unsere Studienteilnehmer gaben an, dass ihre durchschnittlichen jährlichen Kosten für Verletzungen der Datensicherheit bei 3,8 Millionen Dollar liegen. IDC schätzt, dass diese Kosten im Jahr 2023 auf fast 4,5 Millionen Dollar steigen werden. Unternehmen mit stärker fragmentierten, unstrukturierten Daten zahlen jedoch einen höheren Preis für Sicherheitsverletzungen. Wir haben diese Fragmentierung anhand der häufigsten Methode des Austauschs von unstrukturierten Daten gemessen – zum einen die Nutzung von Punkt-zu-Punkt-Verbindungen über E-Mail, Chats, USB-Sticks, FTP usw. und zum anderen die Nutzung von zentralisierten Plattformen für den Austausch, die Zusammenarbeit und die Verwaltung von Inhalten. Eine stärkere Fragmentierung führte zu einer Verdoppelung der jährlichen Kosten für Sicherheitsverletzungen (4,5 Millionen Dollar gegenüber 2,2 Millionen Dollar).

Unternehmen mit eher fragmentierten, unstrukturierten Datenansätzen zahlen einen höheren Preis für Sicherheitsverstöße. Eine stärkere Fragmentierung führte zu einer Verdoppelung der jährlichen Kosten für Sicherheitsverletzungen (4,5 Millionen Dollar gegenüber 2,2 Millionen Dollar).

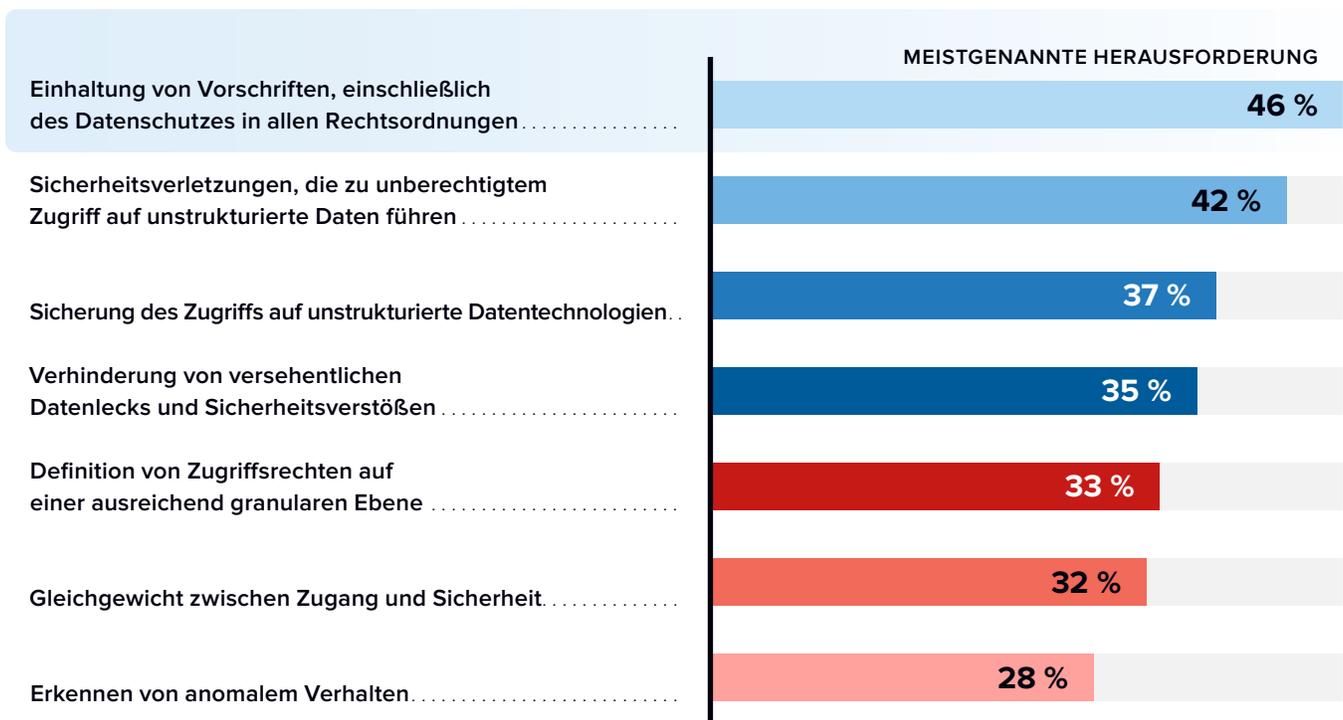
Wenn es um unstrukturierte Daten geht, stehen Unternehmen vor einer Reihe von Sicherheitsherausforderungen. Dazu gehören der unbefugte Zugriff auf unstrukturierte Daten, versehentliche Datenlecks, das Gleichgewicht zwischen Zugriff und Sicherheit und die Erkennung von anomalem Verhalten (siehe **Abbildung 4**).

ABBILDUNG 4

Sicherheits- und Compliance-Herausforderungen

Welche Aussagen beschreiben am besten die Sicherheitsherausforderungen Ihres Unternehmens in Bezug auf unstrukturierte Daten?

(% der Befragten)



n = 414, Quelle: IDC, Umfrage zur Ausschöpfung des Potenzials von unstrukturierten Daten, Mai 2023



51 %

der Befragten gaben an, in den letzten 12 Monaten mit Verstößen gegen Datengesetze konfrontiert gewesen zu sein, wobei die durchschnittlichen Gesamtkosten 1,03 Millionen Dollar betragen.

Compliance-Herausforderungen

Die am häufigsten genannte Herausforderung, wie in **Abbildung 4** (vorherige Seite) dargestellt, ist die Einhaltung von Vorschriften. Diese Herausforderung ist keine Überraschung für Unternehmen, die mit der Fragmentierung von Inhalten über viele Apps in verschiedenen Regionen zu kämpfen haben. Und diese, in verschiedenen Ländern weltweit erlassenen Daten- und KI-Vorschriften, werden auch in Zukunft weiter zunehmen.

Wir haben unsere Studienteilnehmer gefragt, wie zuversichtlich sie sind, dass ihr Unternehmen unstrukturierte Daten aufbewahrt und/oder entsorgt und sie, falls erforderlich, auf der Grundlage aller aktuellen internen Richtlinien oder externen gesetzlichen Anforderungen zur Verfügung stellt. Fast die Hälfte (46 %) der Befragten gab an, dass die Einhaltung von Vorschriften, einschließlich des Datenschutzes in allen Rechtsordnungen, eine der größten Herausforderungen für ihr Unternehmen darstellt. Mehr als die Hälfte (51 %) der Befragten gab an, in den letzten 12 Monaten mit Verstößen gegen Datengesetze konfrontiert gewesen zu sein, wobei die durchschnittlichen Gesamtkosten 1,03 Millionen Dollar betragen. Zusätzlich zu den Kosten, die mit Bußgeldern verbunden sind, öffnet die Nichteinhaltung von Vorschriften die Tür zu Marken- und Reputationsrisiken sowie zu Wettbewerbsrisiken durch den Verlust von geistigem Eigentum in Form von unstrukturierten Daten.

Unsere Studie zeigt auch, dass das Vertrauen in die interne und externe Compliance in Bezug auf die Aufbewahrung, Entsorgung und/oder Weitergabe von unstrukturierten Daten sehr unterschiedlich ist. Während 73 % der Unternehmen „etwas“ oder „sehr zuversichtlich“ sind, was die Einhaltung ihrer eigenen internen Vorschriften angeht, schwindet diese Zuversicht, wenn es um externe Vorschriften geht. Nur 59 % haben das gleiche Vertrauen in die Einhaltung externer gesetzlicher Vorschriften. Natürlich erhöht die Fragmentierung unstrukturierter Daten das Risiko der Nichteinhaltung von Branchen- und Regierungsvorschriften wie ITAR, GxP, DSGVO oder HIPAA.

Produktivität

Viele der in diesem Whitepaper bereits erläuterten Mängel und Herausforderungen sind doppelt problematisch, da sie die Produktivität der Mitarbeiter beeinträchtigen. Die IT-Mitarbeiter sind gezwungen, mehr Zeit für die Sicherung unstrukturierter Datenbestände oder die Integration von Silo-Systemen aufzuwenden, Ineffizienzen bei der Zusammenarbeit mit Mitarbeitern des Unternehmens oder Datenwissenschaftlern zu beseitigen.

Für Mitarbeiter im Unternehmen ist es schwierig, relevante Informationen zu finden oder sie in anderen Anwendungen im Arbeitsablauf zur Verfügung gestellt zu bekommen. Die Unfähigkeit, die richtigen Informationen zur richtigen Zeit zu finden, führt wiederum zu unnötiger Doppelarbeit bei der Erstellung neuer unstrukturierter Datenbestände.

Viele Ineffizienzen sind auf die bereits erwähnten Faktoren zurückzuführen: Anwendungswildwuchs und Fragmentierung unstrukturierter Daten und der Anwendungen, die diese Daten verwalten. Wir haben herausgefunden, dass weniger als die Hälfte (47 %) des Austausches der unstrukturierten Daten über zentralisierte Freigabe-, Content-Collaboration- oder Content-Management-Tools erfolgt, während 53 % der unstrukturierten Daten von Punkt-zu-Punkt oder von Person-zu-Person über Kanäle wie E-Mail, Chats, USB-Sticks, FTP oder gemeinsam genutzte Laufwerke ausgetauscht werden. Der Austausch unstrukturierter Daten über Punkt-zu-Punkt- oder Person-zu-Person-Methoden hemmt die Produktivität, da es schwierig ist, diese Daten zu finden und/oder die richtige Version zu identifizieren. Außerdem besteht das Risiko eines Datenverlusts, wenn der Speicher nicht verwaltet wird und Unbefugte darauf zugreifen oder die Daten verloren gehen können.

Auch zu beachten ist, dass die Fragmentierung den Wissensaustausch zwischen Mitarbeitern und externen Interessengruppen wie Partnern und Lieferanten behindert. Ein kritisches Problem für Unternehmen entsteht dann, wenn entweder die Abwanderung von Wissensarbeitern hoch ist oder ein erheblicher Teil der Belegschaft in absehbarer Zeit in den Ruhestand gehen wird. Die Klassifizierung und Weitergabe von Wissen ist jedoch auch eine grundlegende Fähigkeit für produktivere Innovationen, die Weiterbildung von Mitarbeitern und die Verbesserung des allgemeinen Wissensniveaus eines Unternehmens.

Trotz der verschiedenen Herausforderungen hilft das Potenzial, aus unstrukturierten Daten Wert zu schöpfen – ob mit GenKI, klassischer KI oder Unternehmensanwendungen –, dabei, unternehmerische Prioritäten zu setzen. Das bedeutet, dass IT-Führungskräfte eine Content-Strategie benötigen, die das „Content-Problem“ und die Vorteile erkennt, die sich aus Investitionen zur Bewältigung dieser Herausforderungen ergeben. Eine erfolgreiche Content-Strategie nutzt, schätzt und finanziert unstrukturierte Datenprojekte.

Prioritäten

Der Finanzierungsprozess für unstrukturierte Datenprojekte verbessert sich: 40 % der Unternehmen erwarten, dass die Finanzierung von Projekten mit unstrukturierten Daten in den nächsten 1–3 Jahren einfacher wird, während 40 % davon ausgehen, dass der Finanzierungsprozess gleichbleibt. Das scheint sicherlich einer der Vorteile der Popularisierung von GenKI zu sein.

DIE 3 WICHTIGSTEN PRIORITÄTEN VON UNTERNEHMEN IM JAHR 2023, WENN ES UM **GESCHÄFTSDATEN** GEHT:



1
Compliance und
Datenschutz



2
Entscheidungsfindung
unter Einbeziehung
von Erkenntnissen aus
unstrukturierten Daten



3
Daten-
sicherheit

DIE 3 WICHTIGSTEN PRIORITÄTEN VON UNTERNEHMEN IM JAHR 2023, WENN ES UM GESCHÄFTSDATEN-WORKFLOWS GEHT:



1 Wissensmanagement



2 Kontextualisierung von Daten für die Verwendung im gesamten Unternehmen



3 Sicherstellung einer schnelleren Datenverfügbarkeit

Vorteile

Wir haben die Teilnehmer gefragt, wie sie die Fähigkeit ihres Unternehmens einschätzen würden, alle Quellen unstrukturierter Daten, die in verschiedenen Geschäftsbereichen oder Abteilungen anfallen, zu kennen und/oder zu katalogisieren. Mit anderen Worten: Wie gut wissen sie, welche Arten von unstrukturierten Daten sie haben, wofür sie verwendet werden und wo sie sich befinden? 43 Prozent bewerteten ihr Unternehmen in dieser Hinsicht als „sehr gut“ oder „ausgezeichnet“, während 32 % sich selbst als „gut“ einstufen und 25 % sagten, sie seien „mittelmäßig“ oder „schlecht“.

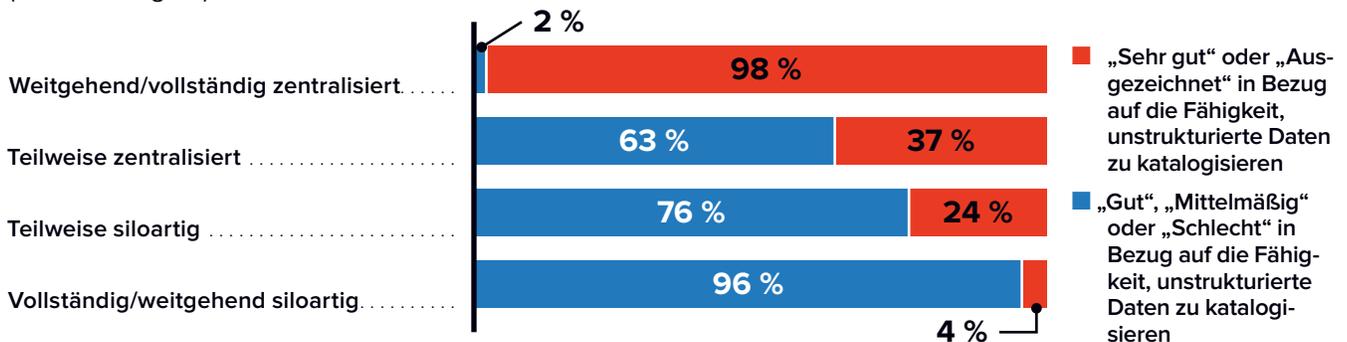
Von den 43 %, die sagen, dass sie über „sehr gute“ oder „ausgezeichnete“ Kenntnisse über ihre Daten verfügen, haben fast alle ihre Daten überwiegend oder vollständig zentralisiert. Andererseits haben die meisten Unternehmen, deren unstrukturierte Daten vollständig oder größtenteils in Silos gespeichert sind, Schwierigkeiten, die benötigten Ressourcen zu finden (siehe **Abbildung 5**).

ABBILDUNG 5

Die Zentralisierung unstrukturierter Daten verbessert die Fähigkeit zur Wertschöpfung beträchtlich

Inwieweit werden unstrukturierte Daten in den von Ihnen ausgewählten Tools, Anwendungen und Systemen zentral (zentralisiert) oder in jedem einzelnen Tool bzw. jeder einzelnen Anwendung (silosartig) gespeichert und verwaltet? Wie würden Sie die Fähigkeit Ihres Unternehmens einschätzen, alle Quellen unstrukturierter Daten, die in verschiedenen Geschäftsbereichen oder Abteilungen anfallen, zu kennen und/oder zu katalogisieren?

(% der Befragten)



n = 414, Quelle: IDC, Umfrage zur Ausschöpfung des Potenzials von unstrukturierten Daten, Mai 2023

Für einen Zugriff auf die Daten in dieser Abbildung siehe [Abbildung 5 Ergänzende Daten](#) im Anhang.

Mit anderen Worten: Daten, die größtenteils oder vollständig zentralisiert sind, helfen Unternehmen sehr dabei, ihre Daten besser zu kennen und diese bei Bedarf auch zu finden. Datensilos beeinträchtigen die Fähigkeit, Daten für Geschäftseinblicke, Wissen und verbesserte Mitarbeiter- und/oder Kundenerfahrungen zu nutzen. Sie können die Nichteinhaltung von Vorschriften und Sicherheitsverstöße wahrscheinlicher machen.

Die Zentralisierung bietet auch den Vorteil des leichteren Zugriffs auf eine große Menge an Daten an einem Ort, um große Sprachmodelle für generative KI anzulernen. Kurz gesagt, Unternehmen ohne eine einheitliche Content-Strategie werden wahrscheinlich den Anschluss verlieren.

Wir haben auch bewertet, inwieweit die Nutzung unstrukturierter Daten in den letzten 12 Monaten die folgenden Fähigkeiten des Unternehmens der Befragten verbessert hat.

Die sechs am häufigsten genannten Ergebnisse mit dem größten Nutzen:

1. Kundenzufriedenheit, -engagement und -bindung
2. Datengovernance
3. Einhaltung der Vorschriften
4. Innovation und/oder Kreativität
5. Produktivität der Mitarbeiter
6. Zusammenarbeit



der Technologieausgaben werden für **unstrukturierte Dateninitiativen** verwendet.



der Technologieausgaben werden für **strukturierte Dateninitiativen** verwendet.

Diese Geschäftsvorteile ergeben sich aus der Wertschätzung – und anschließenden Nutzung – unstrukturierter Daten mit einer Investition in eine zentralisierte, sichere und zugängliche Plattform.

Empfehlungen

Um im Zeitalter von KI nicht nur wettbewerbsfähig zu bleiben, sondern auch langfristig Erfolg zu haben, müssen Unternehmen ihre Daten wie einen Vermögenswert behandeln. Bei strukturierten Daten, die in Datenbanken und Data Warehouses gespeichert sind, ist dies bereits weitgehend der Fall. Dennoch ist die Behandlung unstrukturierter Daten nach wie vor nicht so gut wie die von strukturierten Daten.

Unsere Untersuchung hat gezeigt, dass Unternehmen mit siloartigen und sehr unterschiedlichen unstrukturierten Daten vor vielen Herausforderungen stehen. Wir haben auch festgestellt, dass Unternehmen daran interessiert sind, mehr in Initiativen für unstrukturierte Daten zu investieren, da sie den Bedarf an solchen Daten für das Anlernen von GenKI-Modellen erkennen. Sie erkennen die Risiken und Kosten, die entstehen, wenn sie nicht mehr in Technologien, Fähigkeiten und Prozesse investieren, um aus unstrukturierten Daten Wert zu schöpfen.

Jüngste Fortschritte bei Plattformen zur Vereinheitlichung unstrukturierter Daten, um kreative und geschäftliche Prozesse, die Einhaltung von Vorschriften, die Produktivität der Mitarbeiter und Produkt- oder Serviceinnovationen zu erleichtern, werden es Unternehmen leichter machen, unstrukturierte Daten für einen größeren Geschäftswert zu nutzen.

UNTERNEHMENSLEITUNG UND IT SOLLTEN FOLGENDE ASPEKTE BERÜCKSICHTIGEN:

- 1** Machen Sie eine Bestandsaufnahme der unstrukturierten Daten Ihres Unternehmens sowie der Prozesse, auf die sie angewiesen sind.
- 2** Beurteilen Sie die neuesten Technologieplattformen zur Vereinheitlichung unstrukturierter Daten.
- 3** Wenn Sie solche Plattformen in Betracht ziehen, sollten Sie die Skalierbarkeit, Leistung, Verwaltbarkeit, Interoperabilität und Sicherheit der Plattform bewerten.
- 4** Implementieren Sie ein Datenklassifizierungsschema zur Unterstützung des Zugriffs auf unstrukturierte Daten und ihrer Nutzung.
- 5** Beurteilen Sie die aktuellen KI-Angebote der Anbieter von Unified-Data-Plattform-Technologien und ihre Roadmaps in der sich schnell verändernden KI-Technologie.
- 6** Experimentieren Sie fortlaufend mit GenKI, um den produktiven Einsatz dieser neuen Technologie zu ermitteln.
- 7** Initiieren oder erweitern Sie ein Programm zur Vermittlung von Datenkompetenz, um den Umgang der Mitarbeiter mit KI-gestützten Datenlösungen zu verbessern.
- 8** Verdeutlichen Sie den Wert des Human-in-the-Loop-Ansatzes bei der Nutzung von GenKI als Werkzeug zur Unterstützung von Mitarbeitern, indem Sie wiederholbare Schritte in ihrem Arbeitsablauf automatisieren, anstatt menschliche Expertise, Urteilsvermögen und zwischenmenschliche Kommunikationsfähigkeiten zu ersetzen.
- 9** Nutzen Sie GenKI, um die Arbeit des Compliance-Teams zu unterstützen und die Auswirkungen unstrukturierter Daten auf die Einhaltung von Vorschriften zu verstehen.
- 10** Identifizieren Sie Möglichkeiten zur Integration von unstrukturierten und strukturierten Daten, um Prozesse zu optimieren und die datengestützte Entscheidungsfindung zu verbessern.

Diese Schritte ermöglichen es Unternehmen, den ungenutzten Wert ihrer unstrukturierten Daten auszuschöpfen.

Methodik

Die in diesem Whitepaper dargestellten Ansichten, Analysen und Empfehlungen basieren auf der Marktforschung von IDC zur Nutzung und Verwaltung unstrukturierter Daten. Teil der Studie war eine Umfrage unter 414 Unternehmens- und IT-Entscheidungsträgern aus großen und mittelgroßen Unternehmen aus verschiedenen Branchen und geografischen Regionen, die in ihren Unternehmen für den Zugriff und/oder die Nutzung und Verwaltung unstrukturierter Daten zuständig sind. Die IDC-Analysten führten ergänzend zur Umfrage mehrere ausführliche Interviews mit Box-Kunden durch und stützten sich auf bestehende syndizierte IDC-Studien, wie IDC Global Datasphere, über die Erzeugung, Replikation, Verwaltung und Nutzung von Daten.

IDC-Forscher definierten unstrukturierte Daten als Daten, die nicht in einer vordefinierten Weise organisiert sind und nicht einem semantischen Datenbankmodell entsprechen, das die Adressierung und/oder Analyse der Daten erleichtert.

Unstrukturierte Daten umfassen:

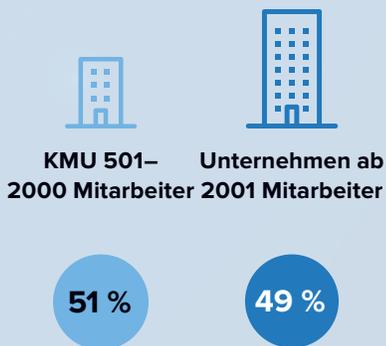
- **Lange Dokumente** (z. B. Manuskripte, Vorschläge, Präsentationen, Tabellen, Memos und andere Textdokumente wie Auszüge/Rechnungen und Formulare)
- **Kurze Texte** (z. B. E-Mails, Sofortnachrichten oder Texte, Posts in sozialen Netzwerken)
- **Erfasste Rich-Media-Inhalte** (z. B. Überwachungsfotos/Videos/Audioaufnahmen, Besprechungsvideos oder Videos, die im Rahmen von Betriebsabläufen aufgenommen wurden)
- **Erstellte Rich-Media-Inhalte** (z. B. Videos, Audioaufnahmen, Bilder für kreative Zwecke, zur Unterhaltung, für Marketing- und Verkaufszwecke)

Strukturierte Daten hingegen sind hochgradig organisierte Daten, die in der Regel in einer Datenbank gespeichert sind, welche einem Datenmodell entspricht und in der die Daten üblicherweise tabellarisch mit Spalten und Zeilen aufgebaut sind. Strukturierte Daten umfassen Sensor- und IoT-Daten. Zu den strukturierten Daten können Transaktions- oder Betriebsdaten wie Verkaufstransaktionen, Finanzkonten, Kundenlisten oder Mitarbeiterdaten gehören.

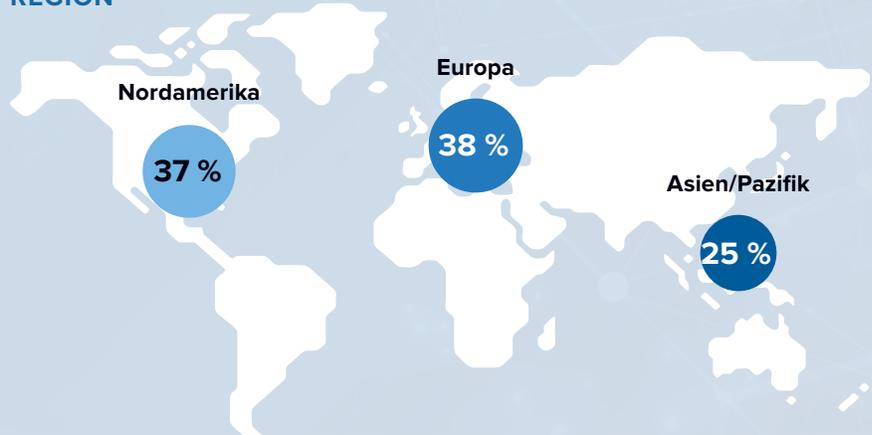
Auf der nächsten Seite finden Sie die wichtigsten demografischen Daten der Umfrage.

DEMOGRAPHISCHE DATEN

FIRMENGRÖSSE



REGION



BRANCHEN



POSITION



Anhang: Ergänzende Daten

Dieser Anhang bietet Zugang zu den Daten für alle komplexen Abbildungen in diesem Dokument. Klicken Sie auf „Zurück zur Originalabbildung“, um zur ursprünglichen Datenabbildung zurückzukehren.

ABBILDUNG 5, ERGÄNZENDE DATEN

Die Zentralisierung unstrukturierter Daten verbessert die Fähigkeit zur Wertschöpfung beträchtlich

	„Sehr gut“ oder „Ausgezeichnet“ in Bezug auf die Fähigkeit, unstrukturierte Daten zu katalogisieren	„Gut“, „Mittelmäßig“ oder „Schlecht“ in Bezug auf die Fähigkeit, unstrukturierte Daten zu katalogisieren
Weitgehend/vollständig zentralisiert	98 %	2 %
Teilweise zentralisiert	37 %	63 %
Teilweise siloartig	24 %	76 %
Vollständig/weitgehend siloartig	4 %	96 %

n = 414, Quelle: IDC, Umfrage zur Ausschöpfung des Potenzials von unstrukturierten Daten, Mai 2023

[Zurück zur Abbildung](#)

Über die IDC-Analysten



Holly Muscolino

Group Vice President, Content Strategies and the Future of Work, IDC

Holly Muscolino ist Group Vice President, Content Strategies and the Future of Work. Sie ist verantwortlich für die Forschung im Bereich Innovation und Transformation bei Content-Lösungen, einschließlich intelligenter Dokumentenverarbeitung, elektronischer Signatur, Imaging und Druck sowie anderer Content-Workflow-Services. Zu Muscolinos Kernkompetenzen gehören auch die Umgestaltung der Arbeitswelt, die Erforschung von Technologien und digitalen Fähigkeiten sowie die Rolle der Technologie bei der Gestaltung der Zukunft unserer Arbeit.

[Mehr über Holly Muscolino](#)



Amy Machado

Research Manager, Enterprise Content and Knowledge Management Strategies

Amy Machado ist Research Manager, Enterprise Content and Knowledge Management Strategies bei IDC. Sie arbeitet auch an der Forschung zur intelligenten Dokumentenverarbeitung mit. Ihr Beratungsprogramm befasst sich mit der grundlegenden Neugestaltung von Content-Diensten und damit verbundenen Anwendungsfällen, die durch die Digitalisierung und die Anwendung innovativer Technologien verändert werden. Angetrieben durch die zunehmende Verlagerung in die Cloud und die Nachfrage nach der Nutzung von Daten für Geschäftsentscheidungen, schaffen diese neuen Technologien viele Möglichkeiten für die Automatisierung und den Einblick in Daten, die traditionell undefiniert und unstrukturiert waren. Machado arbeitet an Software für das Management der Kundenkommunikation, am Enterprise Content Management sowie an der gemeinsamen Nutzung von Inhalten und Zusammenarbeit. Sie nutzt Primärforschung, um die Größe, Prognosen und die Segmentierung dieser Märkte für Tools und Software für das Content- und Wissensmanagement in Unternehmen zu ermitteln.

[Mehr über Amy Machado](#)



John Rydning

Research Vice President, Global Datasphere, IDC

John Rydning ist verantwortlich für die Global DataSphere-Prognosen, in denen die jährlich erzeugten Datenmengen gemessen werden, sowie für die Global StorageSphere-Prognosen zur Messung der weltweit installierten Speicherkapazitäten und der in einem bestimmten Jahr gespeicherten Datenmenge. Darüber hinaus leitet Rydning aufschlussreiche Forschungsarbeiten, die sich mit den wichtigsten Trends, Anwendungsfällen, Technologien und anderen Faktoren befassen, die sowohl die Global DataSphere als auch die StorageSphere beeinflussen.

[Mehr über John Rydning](#)



Dan Vesset

Group Vice President, Analytics and Information Management, IDC

Dan Vesset ist Group Vice President der IDC-Marktforschungs- und Beratungsabteilung für Analytik und Informationsmanagement. In dieser Funktion leitet er eine Gruppe von Analysten, die sich mit allen Aspekten der Verarbeitung strukturierter Daten und unstrukturierter Inhalte, ihrer Integration, Verwaltung, Steuerung, Analyse, Visualisierung und Monetarisierung befasst. Vesset leitet außerdem den globalen Forschungsbereich Big Data und Analytics von IDC.

[Mehr über Dan Vesset](#)

IDC Custom Solutions

Diese Veröffentlichung wurde von IDC Custom Solutions erstellt. Die hierin vorgestellten Meinungen, Analysen und Forschungsergebnisse stammen aus detaillierteren Untersuchungen und Analysen, die von IDC unabhängig durchgeführt und veröffentlicht wurden, sofern nicht ein spezifisches Anbietersponsoring angegeben wird. IDC Custom Solutions stellt IDC-Inhalte in einer Vielzahl von Formaten für den Vertrieb durch verschiedene Unternehmen zur Verfügung. Dieses IDC-Material ist für den externen Gebrauch lizenziert und die Verwendung oder Veröffentlichung von IDC-Forschungsergebnissen bedeutet in keiner Weise, dass IDC die Produkte oder Strategien des Sponsors oder Lizenznehmers befürwortet.



IDC Research, Inc.
140 Kendrick Street, Building B, Needham, MA 02494, USA
T +1 508 872 8200

 @idc

 @idc

[idc.com](https://www.idc.com)

International Data Corporation (IDC) ist der weltweit führende Anbieter von Marktinformationen, Beratungsdiensten und Veranstaltungen für die Märkte der Informationstechnologie, Telekommunikation und Verbrauchertechnologie. Mit mehr als 1300 Analysten bietet IDC globales, regionales und lokales Fachwissen über Technologie- und Branchenmöglichkeiten und -trends in mehr als 110 Ländern weltweit. Die Analysen und Erkenntnisse von IDC helfen IT-Fachleuten, Geschäftsführern und Investoren, faktenbasierte Technologieentscheidungen zu treffen, um ihre wichtigsten Geschäftsziele zu erreichen.

©2023 IDC. Die Vervielfältigung ist verboten, sofern sie nicht genehmigt wurde. Alle Rechte vorbehalten. [CCPA](#)