

Link: <https://www.computerwoche.de/a/der-ungehobene-schatz,1908337>

Unstrukturierte Daten

Der ungehobene Schatz

Datum: 02.11.2009

Autor(en): Holger Eriksson

Eine Flut an unstrukturierten Daten überrollt seit Jahren die Unternehmen. Der steigende Bedarf an Storage-Kapazitäten ist dabei angesichts sinkender Hardware-Preise nur ein Randaspekt. Vielmehr entgeht den Firmen wertvolles Wissen, das unerschlossen vor sich hinschlummert. Auch birgt die Masse der unstrukturierten Daten erhebliche rechtliche Risiken.

Dass die Datenmenge in den Unternehmen zunimmt, ist keine Überraschung: Experten gehen davon aus, dass sie sich etwa alle zwei bis drei Jahre verdoppelt. Eigentlich kein großes Problem - die Speicherpreise pro Gigabyte sinken seit Jahren. Es sind vor allem unstrukturierte Daten, die zum Anschwellen der Datenflut beitragen. Schätzungen gehen davon aus, dass heute nicht einmal ein Viertel aller Daten in strukturierter Form in den Datenbanken der Unternehmens-Applikationen abgelegt ist. Der große Rest ist "**unstrukturiert**"¹ und verteilt sich auf **E-Mail**²- und Messaging-Systeme, PDFs, Office-Daten, und immer häufiger auch Audio- und Video-Dateien. Der Versuch, dem wachsenden Datenvolumen allein mit dem Ausbau der Storage-Kapazität zu begegnen, ist allerdings zu kurz gedacht.



"Wahre Produktivitätssprünge können erst erreicht werden, wenn strukturierte und unstrukturierte Daten kombiniert werden", so IDC-Analyst Rüdiger Spies.

Foto: IDC

Das Problem: Obwohl diese Daten oft relevante Informationen enthalten, sind sie dem Zugriff von analytischen Applikationen weitgehend entzogen. Denn BI-Applikationen greifen in der Regel auf ein Data-Warehouse zu, das wiederum mit den (strukturierten) Daten der Unternehmens-Applikationen befüllt wird. "Die **Data-Warehouse-Technologie**³ kann für sich auch in Anspruch nehmen, dass aus den Unternehmensdaten wertvolle Informationen entstehen, die zu besseren Entscheidungen führen. Das gilt aber nur für strukturierte transaktionsorientierte Daten", schreibt IDC-Analyst Rüdiger Spies in seiner CIO-Kolumne.

Während BI-Lösungen heute als Basis von Geschäftsentscheidungen und für die Steuerung und Planung kaum mehr aus dem Unternehmensalltag wegzudenken sind, bleibt doch das Manko, dass unstrukturierte Daten - und damit wichtige Informationen - regelmäßig außen vor bleiben. "Wahre Produktivitätssprünge können erst erreicht werden, wenn strukturierte und unstrukturierte Daten kombiniert werden", resümiert Analyst Spies.

Rechtliche Fallstricke

Zudem sind gerade unstrukturierte Daten oft von besonderer rechtlicher Relevanz: Vereinbarungen mit Kunden und Partnern werden **per E-Mail**⁴ getroffen, Geschäftsberichte und Verträge sind in Textverarbeitungs- oder PDF-Dateien gespeichert, Listen und Aufstellungen sind in Tabellenblättern erfasst und Absprachen über technische Details liegen in Präsentationen oder digitalen Zeichnungen vor. "Wer seine **unstrukturierten Dat**⁵en nicht gesetzeskonform aufbereitet, riskiert die digitale Amnesie. Diese liegt vor, wenn Daten oder Informationen nicht oder nicht in angemessener Zeit aufgefunden werden können", sagt der Rechtsanwalt Wilfried Reiners aus München. Seine Kanzlei PRW ist auf IT-Themen spezialisiert und hat das Whitepaper "IT-Compliance - Ausgewählte rechtliche Aspekte zur Relevanz unstrukturierter Daten in Unternehmen" verfasst.



"Wer seine unstrukturierten Daten nicht gesetzeskonform aufbereitet, riskiert die digitale Amnesie", warnt Rechtsanwalt Wilfried Reiners.

Bisherige Ansätze in den Bereichen ECM (**Enterprise Content Management**⁶), DMS (Document Management Systems) und ILM (Information Lifecycle Management) im Hinblick auf Storage bieten nur punktuelle Lösungen für einzelne Geschäftsbereiche oder Datensegmente. Grundlage für das Daten-Management ist dabei nicht die inhaltliche Erschließung, sondern die Klassifizierung von Daten nach ihrem (Informations-)Wert. Auf Storage-Ebene basiert die Metadatenklassifizierung auf Kriterien wie Alter, Größe, Dateityp, Besitzer oder Speicherort - in Insellösungen wie E-Mail-, ECM- oder DMS-Systemen erfolgt die Klassifizierung in der Regel manuell durch den Sachbearbeiter.

Bisher fehlt es an herstellernerneutralen unternehmensweiten Lösungen (und erst recht an einer Datenarchitektur), die es prozessübergreifend und über einzelne Geschäfts- oder Anwendungsbereiche hinweg ermöglichen, unstrukturierte Daten zentral zu erfassen und zu verwalten. "Eine stabile, allgemein anerkannte technologische Basis für eine herstellernerneutrale Integration von unterschiedlichen Datentypen fehlt bisher", so IDC-Experte Spies.

Weil aber ein großer und ständig wachsender Teil der relevanten Daten nicht mehr in den Datenbanken von ERP-, **CRM**-⁷ oder BI-Systemen gehalten werden, sondern in Form unstrukturierter Daten im Unternehmensnetz verteilt sind, bleiben sie für den Zugriff von BI- und Management-Informationssystemen verschlossen - ganz gemäß dem geflügelten Wort des Knowledge-Managements: "Wenn das Unternehmen nur wüsste, was das Unternehmen weiß."

"Zwar arbeiten insbesondere die großen Hersteller an neuen Technologien für dieses Problem, aber wirkliche Lösungen - außer Einzellösungen - sind derzeit leider nicht in Sicht", berichtet IDC-Analyst Spies. "Um das Potenzial der Daten in den Unternehmen zu heben, werden eine echte Zusammenführung von strukturierten und unstrukturierten Daten, ECM, BPM, Workflow, **Knowledge-Management**⁸, weitere auf diese Bereiche Bezug nehmenden Technologien und eine neue Datenarchitektur gebraucht."

Links im Artikel:

¹ <https://www.computerwoche.de/software/bi-ecm/1905082/>

² <https://www.computerwoche.de/software/office-collaboration/1885578/index.html>

³ <https://www.computerwoche.de/software/bi-ecm/1905827/index3.html>

⁴ https://www.computerwoche.de/knowledge_center/mittelstands_it/1885624/

⁵ <https://www.computerwoche.de/software/bi-ecm/1868692/>

⁶ <https://www.computerwoche.de/software/bi-ecm/1908169/>

⁷ <https://www.computerwoche.de/software/bi-ecm/1867495/>

⁸ <https://www.computerwoche.de/software/bi-ecm/1869111/>

IDG Tech Media GmbH

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium in Teilen oder als Ganzes bedarf der schriftlichen Zustimmung der IDG Tech Media GmbH. dpa-Texte und Bilder sind urheberrechtlich geschützt und dürfen weder reproduziert noch wiederverwendet oder für gewerbliche Zwecke verwendet werden. Für den Fall, dass auf dieser Webseite unzutreffende Informationen veröffentlicht oder in Programmen oder Datenbanken Fehler enthalten sein sollten, kommt eine Haftung nur bei grober Fahrlässigkeit des Verlages oder seiner Mitarbeiter in Betracht. Die Redaktion übernimmt keine Haftung für unverlangt eingesandte Manuskripte, Fotos und Illustrationen. Für Inhalte externer Seiten, auf die von dieser Webseite aus gelinkt wird, übernimmt die IDG Tech Media GmbH keine Verantwortung.