

DV-Dialog

IBM BUSINESS SYSTEMS • SYSTEM i • AS/400



Interview mit
Milco Gehrman und *Frank Joecks*,
zwei Analytics-Experten der Profi AG

Ein Kontra dem Datenchaos

FOTO: JÖRG LADWIG

Mit der Menge an Daten wächst auch deren Bedeutung. Für das Jahr 2022 geht das Marktforschungsunternehmen IDC von 97 Zettabyte Daten aus, anders ausgedrückt 97 Milliarden Terabytes oder 97 Sextillionen Bytes. Eine Sextillion ist eine Eins mit 36 Nullen.

Kein Wunder, dass der Digitalgipfel der Bundesregierung im Dezember unter folgendem Motto stand: „Daten – Gemeinsam digitale Werte schöpfen“. Dabei ging es auch um Dateninfrastrukturen, -plattformen, Datenräume sowie -verfügbarkeit und -nutzung. Aus Sicht der Deutschsprachigen SAP-Anwendergruppe sind all das wichtige Themen, um die digitale Transformation von Wirtschaft und Gesellschaft aktiv und nachhaltig zu gestalten. Dabei seien generell industriespezifische Plattformen

wie z. B. Gaia-X oder Catena-X erstrebenswert, um die immer fragileren Wertschöpfungsketten zu stabilisieren und es den Unternehmen zu ermöglichen, ihre Effizienz zu steigern. Das Ziel sollten vernetzte Unternehmen sein, doch aktuell enden Systemarchitekturen oft immer noch an den Unternehmensgrenzen.

Fortschritte an dieser Stelle haben eine große Hebelwirkung. Beispielsweise sind Unternehmen, die ein datengesteuertes Change-Management etabliert haben, in ihren Veränderungsprozessen erfolgreicher als Unternehmen, die dies noch nicht umsetzen. Das ergab eine aktuelle Studie „Datengestützte Organisationen: Mit Daten den Veränderungserfolg beflügeln“ von Capgemini. Demnach steigert die hohe Datenreife einer Organisation den Erfolg um 27 Prozent, durch eine datengesteuerte Führung um 23 Prozent und um 26 Prozent durch eine datenbasierte Unternehmenskultur. 61 Prozent der Befrag-

tungen gaben an, sich bei Entscheidungen auf eine Datenanalyse zu stützen. Mit Hilfe von Daten können objektive KPIs definiert werden, um den Fortschritt in Veränderungsprozessen zu messen, der bisher nur durch eine Vision und daraus abgeleitete Ziele und Maßnahmen vorangetrieben wurde.

Mittelständler sind längst noch nicht immer soweit, denn die Herausforderung besteht darin, Daten intelligent zu nutzen. Dabei hilft eine klare Datenstrategie, Vertrauen in die Nutzung von Daten im gesamten Unternehmen, eine moderne Datenlandschaft und die Fähigkeit, datengestützte Entscheidungen zu treffen. Es gilt, datenbasiertes Arbeiten fördern, um das Vertrauen der Belegschaft in Veränderungsprojekte zu stärken. Wie das gelingen kann und welche Rolle das Datenmanagement dabei spielt, erklären Frank Joecks und Milco Gehrman, zwei ausgewiesene Analytics-Experten der Profi AG.



„Die ständige Pflege der Daten ist das A und O!“

Interview mit *Milco Gehrman* (links) und *Frank Joecks*, zwei der Analytics-Experten der Profi AG

TEXT: BERTHOLD WESSELER | FOTOS: JÖRG LADWIG

Die laufende Digitalisierung und innovative Projekte wie der Einstieg in die Industrie 4.0 generieren immer mehr Daten. Jeder Sensor, jede neue Workload und jedes Monitoring-Tool zur Prozesssteuerung und -kontrolle erzeugt zusätzliche, oftmals auch neue Typen von Daten. Mittelständler wollen diese Daten speichern, auswerten, verknüpfen, juristischen Vorgaben entsprechend pflegen und archivieren. Um die resultierende Datenflut aber nach-

haltig in den Griff zu bekommen, reicht die Erweiterung des Storage-Pools oder die Erweiterung um Cloud-Storage nicht aus. Denn das eigentliche Problem der Datenflut ist ein anderes: Wer nicht weiß, welche Daten überhaupt vorhanden bzw. wo sie gespeichert sind, kann sie weder nutzen noch löschen.

Deshalb haben wir zwei Experten der Profi AG befragt, wie sich die Datenflut so kanalisieren lässt, dass sie effektiv nutzbar ist; Milco Gehrman ist Senior Sales & Solution Manager des Darmstädter IT-Dienstleisters und Frank Joecks Geschäftsfeldleiter Soft-

ware-Lösungen. Und Stefan Held, Teamleiter Software-Lösungen West bei der Profi AG, zeigt als dritter im Bunde auf, wie sich das angesichts der steigenden Flut vielerorts drohende Datenchaos elegant vermeiden lässt.

Herr Joecks, Daten gelten gemeinhin als wertvoll, wenn nicht gar als „Schatz“.

Würden Sie dieser These zustimmen

– oder gibt es nicht auch wertlose

Daten und regelrechten Datenmüll?

Frank Joecks: Grundsätzlich sind Daten wertvoll – das würde ich so unterschreiben. Natürlich ist nicht jeder einzelne Datensatz per se wertvoll, sondern vielmehr die Gesamtheit aller Daten, abhängig von deren Zustand. Der eigentliche Wert der Daten entsteht ja erst durch ihre Verknüpfung und Verarbeitung. Und zwar nur dann, wenn alle Daten aktuell, korrekt und vollständig sind. Also ist der Zustand der Daten das Wertvolle – nicht der Datensatz an sich.

„Sobald die Daten gespeichert sind, darf man sie nicht sich selbst überlassen, denn Daten altern.“

Milco Gehrman

Profi spricht in einem Whitepaper aber auch vom Datenchaos. Wann droht das?

Milco Gehrman: Datenchaos finden wir immer dann vor, wenn Daten zwar erhoben und gesammelt, aber nicht gepflegt werden. Und wenn es weder Regelungen für die Sammlung der Daten noch Strukturen für die plausible, konsistente Aufbewahrung und den Zugriff auf die Daten gibt. Dann passen die Daten nicht zueinander, sind unvollständig oder einfach auch irrelevant. Datenchaos entsteht immer dann, wenn alle Daten einfach in einen Topf geschmissen und dann sich selbst überlassen werden. Das ist der größte Fehler, den man machen kann.

Man muss die Daten also nicht nur sammeln, sondern auch permanent pflegen, eventuell auch noch aufbereiten und zusammenfügen zu einem Datensatz?

Joecks: Exakt – die ständige Pflege der Daten ist das A und O! Man muss natürlich auch von vornherein die Strukturen dafür schaffen, dass alle Daten zuverlässig, sicher und sinnvoll verknüpfbar abgelegt werden können.

Gehrman: Sobald die Daten gespeichert sind, darf man sie nicht sich selbst überlassen, denn Daten altern. Man muss also bestimmte Dinge mit den Daten tun, damit sie ihren Wert behalten. Das nennen wir Datenpflege.

Verstehe ich das richtig: Gefragt ist so etwas wie „Datenmanagement“?

Gehrmann: Genau das – ein funktionierendes Datenmanagement. Dann ist klar geregelt: Wie kommen neue Daten in den Bestand, wie werden sie verknüpft, wie werden sie bereitgestellt und wie werden sie gepflegt.

Joecks: Das Datenmanagement regelt natürlich auch, wie vorhandene, auch sehr alte Datenbestände gepflegt werden. Dann ist auch festgelegt, wie alte und neue Datenbestände sinnvoll kombiniert werden, welche Randbedingungen dabei zu beachten sind und welche Methoden dafür in Frage kommen. Wichtig ist es die richtige Balance zu finden, damit die bürokratischen Hürden beim Datensammeln und bei der Datenpflege nicht zu hoch werden. Andernfalls wird man keine Datenschätze anhäufen können.

Gehrmann: Viele Unternehmen verzetteln sich in Details zu den Regelungen, welche Daten überhaupt erfasst und gesammelt werden sollen. Oder sie werden schon verworfen, bevor man ihren Wert im Gesamtbestand überhaupt erkennen konnte. Das darf auf keinen Fall passieren! Zunächst einmal sollte ich alle Daten willkommen heißen und in den Datenbestand aufnehmen. Falls ich dann erkenne, dass sie keinen Wert haben, kann ich sie immer noch verwerfen.

„Die wichtigsten Daten sind die, die mir bei den dringendsten Problemen helfen.“

Frank Joecks

Joecks: Zu große Vorsicht ist wirklich nicht empfehlenswert! Vorausgesetzt, man darf die Daten überhaupt erheben, empfehle ich das auch zu tun. Lieber zuerst etwas mehr Daten sammeln und dann eventuell einige davon verwerfen als sich später ärgern, zu sparsam gewesen zu sein. Denn man darf nicht vergessen: Wer datengetriebene Entscheidungen treffen möchte, wer Analyse-Modelle aufbauen will und wer mit dem Gedanken an datenbasierende Prognosen spielt, der muss zwingend über einen gewissen Datenbestand verfügen. Dieser Bestand muss zuvor erst einmal aufgebaut werden – und das geht nicht über Nacht. Deshalb ist klar im Vorteil, wer im Laufe der Jahre einen gewissen Datenvorrat angelegt hat, quasi auf Verdacht.

Wie baut man solch ein Datenmanagement sinnvollerweise auf – von vornherein unternehmensweit oder sukzessive in den wichtigsten Unternehmensbereichen? Gibt es „Best Practices“?

Gehrmann: Empfehlenswert ist der Aufbau eines Data Lake, der klaren, aber möglichst wenigen Regeln folgt. Im Data Lake sollten Daten abgelegt werden können, ohne dass ein langwieriger, bürokratischer Prozess durchlaufen wird. Dann stehen die Daten für unterschiedlichste Anwendungsfälle parat – ein gutes Datenmanagement einmal vorausgesetzt, das die Daten hegt, pflegt und überarbeitet. Hier trennt sich recht schnell die Spreu vom Weizen. Das Unternehmen sollten den Weg einschlagen, auf dem es am einfachsten ist, die Daten und sammeln und zu pflegen.

Wie setzt man dabei Prioritäten und identifiziert die Bereiche, in denen die wichtigsten Daten stecken?

Joecks: Die wichtigsten Daten sind die, die mir bei den dringendsten Problemen helfen. Die Daten werden ja nicht aus Jux und Dollei gesammelt, sondern um damit Aufgaben zu lösen oder Entscheidungen zu treffen.

Sind diese Daten nicht schon im ERP-System vorhanden?

Gehrmann: Manche sicherlich, aber bei weitem nicht alle. Man muss sich eines vor Augen halten: Die ERP-Daten sind ja immer für einen ganz bestimmten Zweck erhoben worden. ERP-Daten wurden zum Beispiel erfasst, um Produktions- oder Einkaufsprozesse effizient zu steuern oder Rechnungen zu schreiben.

Anderes Beispiel: In der Materialwirtschaft kann man mit den richtigen Daten die Mengen und das Timing der Bestellungen optimieren. Da fehlen dann aber noch wichtige Daten, zum Beispiel: Laufen die Maschinen so „rund“, dass tatsächlich auch pünktlich wie geplant in der gewohnten Qualität geliefert werden kann.

Sämtliche Daten zu speichern, scheidet ja oft allein aus Kosten- und Performance-Gründen aus...

Joecks: Aus Kostengründen wohl nur selten! Speicher ist heutzutage günstig, speziell wenn ich an Cloud-Speicher denke. Performance kann schon eher ein Thema werden, falls das Datenvolumen sehr groß wird. Aber um Performance geht es im Bereich Analytics nur selten, sondern es geht um die Resultate. Es geht also vor allem darum: Welche Granularität der Daten brauche ich? Und über welche Zeiträume brauche ich diese Daten?

Gehrmann: Nehmen wir das Beispiel „Markttrends“ und die Frage: Wie wird sich der Absatz bestimmter Produkte in den nächsten Monaten entwickeln? Dann spielen Daten, die ich vor zwei, drei Jahren für eine ähnliche Prognose erhoben habe, vermutlich gar keine Rolle mehr. Vielleicht lohnt es sich, für den Vergleich ein paar Aggregate aufzubewahren: Summen über Regionen oder Zeiträume und Kennzahlen zu Ausreißern. In der Regel

Beseitigung des Datenchaos

▣ **Bevor ein systematisches Datenmanagement überhaupt möglich ist, muss das Datenchaos beseitigt werden, das üblicherweise zuvor durch unzureichende Sammlung und Pflege der Daten historische gewachsen ist. Anders ist eine homogene, strukturierte und weitgehend automatisierte Datenverwaltung nicht möglich.**

Bei der Beseitigung des Datenchaos muss zunächst festgestellt werden, welche Daten überhaupt verfügbar –

und welche dieser Daten relevant sind. Dazu müssen die relevanten Daten, die bisher in den unterschiedlichsten Formaten an unterschiedlichsten Stellen gesammelt wurden, digitalisiert und sauber zusammengeführt werden.



Stefan Held,
Teamleiter
Software-
Lösungen West bei
der Profi AG.

Aktuelle Daten lassen sich dagegen mit Sensoren und Überwachungssystemen in regelmäßigen Abständen automatisch erfassen. Diese aktuellen Daten können dann mit historischen Werten, Ist- und Soll-Werten sowie unstrukturierten Daten, z. B. Fotos oder Grafiken, kombiniert werden. Fehlerhafte Daten müssen dabei aussortiert bzw. gelöscht, die entstehenden Datensätze bereinigt und lückenhafte Daten aufgefüllt werden. Die gründliche Aufbereitung aller Daten ist zwar zeit- und kostenintensiv, aber essenziell, denn erst nach einer gründlichen Aufarbeitung und nach der Beseitigung des Datenchaos kann die Datenanalyse und Datenmodellierung brauchbare Ergebnisse liefern.

Ist das Datenchaos beseitigt, kann das eigentliche Datenmanagement weitestgehend automatisiert erfolgen. Der Kunde sollte dann selbst entscheiden können, ob ein Dienstleister das Projekt begleiten soll oder ob die Datenpflege ohne externe Hilfe fortgesetzt werden soll. ▣

Stefan Held

sind für solche Prognosen aber keine uralten Detaildaten mehr erforderlich. Anders ist das zum Beispiel bei der baulichen Infrastruktur: Hier kann ein einzelnes Ereignis, das zwanzig Jahre zurückliegt, auch heute noch eine Rolle spielen, beispielsweise für die Wartung. **Joecks:** Der größte Fehler, den wir dabei machen können, ist es, einen Datenfriedhof zu schaffen! Damit meine ich Daten, die niemand mehr braucht und die demzufolge auch niemand pflegt. Ich vergleiche das gerne mit den Dias einer Rucksack-Tour durch Asien vor 30 Jahren. Mangels Projektors schaut ▣



sich die heute niemand mehr an. Doch im Keller steht eine Kiste, die Platz verbraucht. Eine Alternative wäre es, Abzüge zu erstellen, eine Collage daraus zu machen und diese an die Wand zu hängen. Dann hätte man etwas davon.

„Wichtig ist es die richtige Balance zu finden, damit die bürokratischen Hürden beim Datensammeln und bei der Datenpflege nicht zu hoch werden.“

Frank Joecks

Gehrmann: Genauso ist das auch mit Daten. Sobald sie eine gewisse Aktualität verloren haben, macht man besser etwas damit, anstatt sie vergammeln zu lassen. Beispielsweise kann es wichtig sein, Metadaten zu ergänzen, damit die Daten aussagekräftig bleiben. Vielleicht auch deshalb, weil man sich beim Erfassen der Daten um die Metadaten noch gar keine Gedanken gemacht hatte.

Worauf kommt es an, damit das implementierte Datenmanagement dann auch effizient und zuverlässig funktioniert?

Gehrmann: Auf klar definierte Prozesse! Dabei können auch manuelle Tätigkeiten eine Rolle spielen, aber das Gros des Datenmanagements sollte automatisiert erfolgen. Dafür sind aber vernünftige Metadaten erforder-



lich. Dann kann man zum Beispiel automatisch alle Daten zu einem definierten Thema vor einem Zeitpunkt x mit einem gewissen Detaillierungsgrad wahlweise löschen oder auf Monatswerte aggregieren.

Joecks: Das funktioniert nur dann, wenn die Verantwortlichkeiten eindeutig definiert sind. Es muss absolut klar sein, wer bei der Pflege welcher Daten den Hut aufhat und wer vor allem überhaupt weiß, welche Schätze in der Datenbank lagern. Oft besteht ja das Problem auch darin, die richtigen Daten im Unternehmen zu finden.

Gehrmann: Es reicht ja nicht zu wissen, dass sie irgendwo gespeichert sein sollten. Hier hilft es ganz enorm, über klar definierte Prozesse im Datenmanagement festzulegen, wer für bestimmte Daten verantwortlich ist. Die Person kann man fragen – und die Person entscheidet auch, ob Daten irgendwann gelöscht oder kumuliert werden. Ob das dann täglich, wöchentlich, monatlich oder jährlich erfolgt, ist zweitrangig. Hauptsache, es erfolgt automatisch!

Welche Rolle spielen Datenschutz und Datensicherheit in diesem Zusammenhang, beispielsweise bei Kunden- oder Mitarbeiterdaten?

Joecks: Der Schutz sensibler Daten ist teilweise gesetzlich vorgegeben. Oft ist er aber allein schon absolute Pflicht, um Schaden

vom Unternehmen abzuwenden. Andererseits lohnt sich die Mühe einmal zu definieren, welche Daten unter dieser Prämisse überhaupt schützenswert sind. Werden beispielsweise personenbezogene Daten lediglich für technischen Zwecke benötigt, können sie ja vor dem Abspeichern anonymisiert werden.

Gehrmann: Personenbezogene Daten müssen im Rahmen des Datenmanagements anders behandelt werden als anonyme Daten. Das ist ein heikles Thema, denn da gibt uns die DSGVO ganz klar den Rahmen vor. Solche Daten dürfen z. B. nur zweckbezogen und über definierte Verwahrungsfristen gespeichert werden. An dieser Stelle hilft ein automatisiertes Datenmanagement ganz enorm dabei, auf der sicheren Seite zu bleiben.

Wenn Daten so wertvoll sind, müsste man sie doch monetarisieren können! Worauf ist dabei zu achten? Sind heutige Datenbörsen oder IIoT-Ökosysteme überhaupt schon tauglich für den Mittelstand?

Joecks: Prinzipiell funktioniert das schon ganz gut, beispielsweise für Sensor-Daten von Maschinen. Es gibt ja viele Daten, die überhaupt nicht vertraulich sind. Werden die veröffentlicht, dann kann aber die gesamte Branche davon profitieren. Das ist wie bei Open-Source-Software: Viele tragen ihren Teil dazu bei – und alle profitieren davon. Allerdings gibt es durchaus noch viele Teilnehmer im Daten-Markt, die sehr zurückhaltend sind.

Warum?

Joecks: An den Datenbörsen agieren auch Teilnehmer, die irgendetwas anbieten – auch schlecht gepflegte Daten. Irgendwelche „strubbeligen“ Verkehrs- oder Sensor-Daten werden dann für viel Geld angeboten – und niemand kann damit etwas anfangen. Manche Teilnehmer bieten auch Daten an, die ihnen gar nicht gehören. Für Mittelständler lohnen sich Datenbörsen daher nur selten.

Gehrmann: Vielversprechender finde ich hier den Open-Data-Ansatz, bei dem Daten vor einem rein technisch-wissenschaftlichen Hintergrund angeboten werden – auch von Regierungsstellen. Diese Daten sind für jedermann frei zugänglich und können auf Grund von offenen und diskriminierungsfreien Lizenzen unbeschränkt weiterverwendet werden. ☑

„Datenchaos finden wir immer dann vor, wenn Daten zwar erhoben und gesammelt, aber nicht gepflegt werden.“

Milco Gehrmann