

Kontext statt Metadaten

Heiko Robert, Gründer, ecm4u GmbH, Stuttgart, www.ecm4u.de, info@ecm4u.de

In den letzten 20 Jahren haben viele Organisationen Dokumentenmanagementsysteme eingeführt. Die Speicherung von Metadaten an den Dokumenten wird zwar von Anwendern gewünscht, wenn es um das Auffinden und die Strukturierung von Informationen und Dokumenten geht, jedoch nicht selten bei der Speicherung umgangen. Die Verlässlichkeit dieser Metadaten beschränkt sich daher meist auf automatisch gepflegte Informationen. Ganz neue Möglichkeiten ergeben sich durch Paradigmenwechsel, die von Cloud-Technologien ausgelöst werden: Metadaten werden nicht mehr (ausschließlich) im System gespeichert, sondern über Kontext-Beziehungen unter Verwendung vereinfachter Technologien referenziert. Konsequenz zu Ende gedacht kann dies zu einer wesentlich besseren Vernetzung von Unternehmensinformationen führen und die Unternehmens-IT-Infrastruktur revolutionieren, die heute noch von Datensilos geprägt ist.

Das Malheur der (Meta-)Datenpflege

Anwender haben es über viele Jahre hingenommen, dass die Pflege von Metadaten in Dokumentenmanagement- (DMS), aber auch in anderen Systemen ein mühsamer Prozess sein kann. Häufig stehen nur einfache Eingabefelder in mehr oder weniger umfangreichen Formularen zur Verfügung, um die gewünschten Informationen zu pflegen. Hierbei werden jedoch Daten erfasst, die vor allem in anderen Systemen verwaltet werden. Nehmen wir ein einfaches Beispiel zur Illustration: Es soll ein Geschäftsbrief als Korrespondenz abgelegt werden. Im DMS ist vorgesehen, unter anderem den Geschäftspartner, die Kontaktinformation, ein Betreff und ggf. einen Vorgang anzugeben. Der fleißige Mitarbeiter tippt daher gewissenhaft ein:

Geschäftspartner: Mustermann AG

Kontakt: Hans Meier

Betreff: Änderungen unserer Lieferbedingungen

Vorgang: Überprüfung der Lieferbedingungen

Gehen wir davon aus, dass diese Metadaten in einfachen Textfeldern ohne zusätzliche Hilfsmechanismen gespeichert werden, birgt diese Art der Metadaten gleich mehrere Nachteile:

Die Informationen zu Geschäftspartner oder Kontakt könnten falsch oder unvollständig übermittelt worden sein, verschiedene Schreibweisen nutzen oder die Namen könnten sich sogar im Laufe der Zeit ändern. Die Konsequenz ist, dass der Geschäftsbrief später nicht gefunden wird. Genauso ist es aber auch möglich, dass beispielsweise der Name des Kontakts nicht eindeutig ist, weil es mehrere Ansprechpartner mit dem Namen Hans Meier gibt.

Einfache Eingabefelder können den Anwender auch nicht durch Validierung und die Anzeige von Zusatzinformationen unterstützen. Dadurch werden Fehler kaum erkannt und die Eingabe von redundanten Informationen kann nicht vermieden werden. Noch schwieriger wird es, wenn der Anwender zwingend weitere Kontextinformationen wie beim Zuordnen des korrekten Vorgangs benötigt. Der Vorgangstitel kann kaum genutzt werden, um durch den Anwender einen bereits vorhandenen Vorgang, der beispielsweise im CRM geführt wird, eindeutig zu identifizieren. Der Titel ist viel zu allgemein und die Vorgangsnummer dem Anwender nicht geläufig.

Umgekehrt eignen sich einfache Textfelder kaum, um von anderen Systemen verlässlich auf gespeicherte Dokumente zuzugreifen, ohne dass eindeutige Referenzen gespeichert werden.

Einige DMS-Hersteller oder auch Systemintegratoren haben daher in der Vergangenheit zwei Strategien angewendet, um dies zu verbessern:

- Verwaltung bestimmter Stammdaten, also Entitäten wie Organisationen und Kontakte im DMS selbst oder
- die Integration führender Systeme wie CRM- oder ERP-Systeme in das DMS über diverse, meist proprietäre Schnittstellen.

Beide Ansätze sind teuer, aufwändig und eingeschränkt: Die Verwaltung von Organisationen und Kontakten in einem DMS kann nur eine Notlösung darstellen, da es kaum das führende System für die Verwaltung dieser Stammdaten werden kann und auch nie dem Anspruch an ein System zur Kontaktverwaltung erfüllen wird. Auch das regelmäßige Im- und Exportieren von Stammdaten erzeugt neue Probleme.

Werden die Entitäten hingegen durch eine DMS-spezifische Integration mit dem führenden System des Unternehmens referenziert und angezeigt, wird meist nur der interne Schlüssel der referenzierten Entität und ggf. der Name gespeichert. Die Nutzung solcher Schnittstellen kann mit erheblichen Kosten verbunden und dennoch stark auf die Funktionalität eingeschränkt sein, die ein DMS-Anbieter oder dessen Partner vorgesehen hat, da diese nur ganz bestimmten Kombinationen von DMS und führendem System umgesetzt sind.

Sollte im Grunde nicht *eigentlich* die Anforderung sein, das Dokument mit den Kontakten und Geschäftspartnern (Entitäten) zu verknüpfen, gleich in welchem System diese verwaltet werden, quasi in einer Föderation von Stammdaten? Eine Änderung des Kontaktnamens oder der Firmierung des Geschäftspartners sollte dabei genauso wenig Einfluss auf diese Verknüpfung haben, wie die Ablösung eines führenden Systems durch ein anderes. Ein nicht unerheblicher Teil der Kosten und Aufwände bei der Umsetzung von DMS-Vorhaben werden durch die Vernetzung mit führenden Systemen verursacht.

Für die erste Forderung ist es notwendig, systemintern einen stabilen, unveränderbaren Fremdschlüssel zu speichern. Dem Anwender soll jedoch immer die sprechende Eigenschaft wie Name, Ort und weitere Referenzen aus dem führenden System angezeigt werden. Für die zweite Forderung ist es notwendig, als stabilen Fremdschlüssel niemals die internen IDs eines Fremdsystems zu verwenden, sondern einen fachlichen Fremdschlüssel, der auch nach einem Systemwechsel noch zur Verfügung steht und möglichst einfach auch fernmündlich übermittelt werden kann. Beispiele hierfür kennt jeder aus dem Alltag: Kunden-Nr., Vorgangs-Nr., Auftrags-Nr.

Vernetzte Stammdaten

Für manchen werden die zuvor genannten Forderungen selbstverständlich erscheinen und wenig Neues enthalten. Da diese jedoch nur unzureichend umgesetzt sind, lohnt es sich, die Gründe hierfür näher zu beleuchten:

- Anbieter von Softwaresystemen, in denen wichtige Stammdaten verwaltet werden (ERP, CRM, Groupware etc.) verfolgen meist eine Monopolstrategie, welche auch One- oder Single-Vendor-Strategie genannt wird. Die Föderation von Stammdaten steht kaum im Fokus.
- Bisher bekannte und genutzte Konzepte und Technologien wie EAI (Enterprise Application Integration), SOA (Service Oriented Architecture) und Webservices sind für die pragmatische Integration von Stammdaten zu komplex, für die direkte Einbindung in Browser-Technologien ungeeignet und selten für den synchronen Abruf durch viele, nahezu beliebige Anwendungen ausgelegt.

Durchdenkt man tiefgreifender die Abbildung von Geschäftsprozessen durch die IT, wird man sich früher oder später damit auseinandersetzen, dass die wichtigen Entitäten, auf die man referenzieren möchte, in mehreren Systemen verteilt sind und dass diese nicht selten redundant gepflegt werden.

Statt also mit Geschäftspartnern, Kontakten, Fällen zu arbeiten, schlägt sich die IT damit herum, diese Entitäten mehr oder weniger gut zwischen Systemen wie ERP, CRM, Groupware zu synchronisieren oder noch schlechter: von den Anwendern händisch synchronisieren zu lassen.



Abb. 1: Stammdaten-Entitäten



Abb. 2: IT-Systeme

Ginge es jedoch nach dem Anwender, würde dieser eigentlich in jeder Situation Dokumente und Informationen möglichst nur mit den relevanten Entitäten eines Unternehmens wie Geschäftspartner, Kontakt, Fall usw. assoziieren, also verknüpfen und danach darauf zugreifen und auswerten wollen. Man könnte auch sagen: Der Anwender möchte eigentlich den gewünschten Kontext herstellen, anstatt stupide Metadaten einzugeben. Es sollte auch möglich sein, über mehrere Verknüpfungen hinweg zu navigieren – vergleichbar zu einem Assoziationsnetz (z.B. „Die Mail kam im Kontext des Projekts X. Absender und Betreff sind allerdings vergessen“). Diese Art des

Zugriffs entspräche viel eher der menschlichen Denkweise und ermöglichte außerdem auch mehrere, direkte und indirekte Zugriffspfade auf dieselben Informationen und Dokumente.

Die Konzepte und Technologien hierzu kennt die IT unter Master Data Management (MDM). Man könnte auch sagen, es handelt sich bei MDM um eine Übersetzung von Konzepten relationaler Datenbanken für eine systemübergreifende Anwendung. Die Definition im englischen Wikipedia-Artikel trifft den Kern deutlich besser als die deutsche Fassung über Stammdatenmanagement. Frei übersetzt: „Stammdatenmanagement umfasst die Prozesse, das Einführen und Überwachen von Richtlinien, Standards und Werkzeugen, um im Geschäftsalltag die kritischen Daten einer Organisation konsistent zu definieren und über eindeutige Referenzen zu verwalten“

Die bisher verfügbaren Produkte und Lösungen sind meist sehr hochpreisig und eingeschränkt auf bestimmte Plattformen und Technologien. Daher sind diese Konzepte noch nicht sehr verbreitet und nur wenige Softwareanbieter verwenden diese in den eigenen Produkten und Oberflächen. Ungeachtet dessen: sollte es nicht ein strategisches Ziel von Unternehmen sein, Referenzen auf Stammdaten statt Metadaten in Freitextfeldern, Suchmaschinen, Fach- und Workflowsystemen zu verwenden?

Voraussetzungen für die konsequente Nutzung von MDM-Konzepten sind die Verfügbarkeit (technisch wie kaufmännisch) von möglichst technologieübergreifenden und einfach einzubindenden

- Oberflächen-Komponenten, mit denen man dem Anwender die relevanten Daten statt der gespeicherten Referenzen anzeigt und
- Service-Schnittstellen, über die die Unternehmenssysteme auch in der Geschäftslogik Zugriff auf die referenzierten Informationen haben.

Die Tragweite dieser Voraussetzungen legt eigentlich nahe, dass diese nur erfüllt werden können, wenn entweder offene Standards vereinbart werden, die von allen Anbietern umzusetzen sind (der erfahrene ITler wird nun fortsetzen: „... und dessen Umsetzung sehr unwahrscheinlich ist.“). Alternativ könnten sich Anbieter mit Monopolstellung im Markt mit eigenen Lösungen durchsetzen (Beispiel DATEV).

Glücklicherweise haben sich durch andere Entwicklungen im Markt ganz neue, technologische Konzepte etabliert, die hervorragend für die Umsetzung und Adaption von MDM-Konzepten in Unternehmen geeignet sind, sodass nicht auf die mühsame Verabschiedung von Standards gewartet werden muss. Es geht manchmal sehr viel pragmatischer. Dazu später mehr.

Führende und konsumierende Systeme

Ein Umdenken ist vor allem in der Betrachtung und Bewertung von sogenannten führenden Systemen notwendig. Noch vor gar nicht allzu langer Zeit hat sich ein IT-Leiter mit der Aussage wie „SAP ist unser führendes System“ durchsetzen können - fälschlicherweise. Die Praxis hat gezeigt, dass es *das* führende System gar nicht gibt. Besser: In jedem Unternehmen sollte es möglichst nur *ein* führendes System *je* Stammdaten-Entität, beispielsweise für die Pflege von Unternehmensmetadaten, geben. Ist dies nicht möglich, müssen Konzepte und Konventionen umgesetzt werden, die faktisch nur *ein* System zum verlässlichen, führenden System für *bestimmte* Metadaten werden lässt. Dieses muss die Daten wiederum allen anderen Anwendungen zur Verfügung stellen. Selbst bei den großen Softwareanbietern mit ganzheitlichem Ansatz gibt es schon lange nicht mehr ein zentrales, integriertes System, was deutlich macht: Jedes IT-System in einem Unternehmen muss in der Lage sein, (direkt oder indirekt) Stammdaten sowohl anderen Systemen synchron zugänglich zu machen als auch die Daten anderer Systeme zu konsumieren.

Um die Voraussetzungen für digital gestützte Unternehmensprozesse zu schaffen, müssen die Unternehmenslenker und IT-Leiter umdenken: Die IT muss als eine einheitliche Schicht für den Zugriff von Unternehmensstammdaten über Anwendungsgrenzen hinweg verstanden werden.

Bekannt und größtenteils realisiert ist die Anforderung, dass Unternehmenssoftware wie ECM nicht als Anwendung, sondern als Infrastruktur zu verstehen ist. Dies sagt aber noch nichts über die Speicherung von Kontext aus. Es ist kaum vorstellbar, dass ein Eingangsschreiben nur automatisch verarbeitet werden kann, wenn *zuvor* durch einen Experten entschieden wird, in welcher Rolle der Absender handelte (Kunde, Interessent, Lieferant, Vertragspartner, ...), damit erst danach das zu referenzierende System für Kontextdaten ausgewählt werden kann oder wenn zuvor der Absender im ERP als Geschäftspartner angelegt wurde. Genau das ist aber heute noch in den meisten Unternehmen der Fall, wenn es beispielsweise um die Zuordnung eines externen Geschäftspartners geht. Die Dokumentenmanagementsysteme sind die Anwendungen, die es gewohnt sind, sich den führenden

Systemen unterzuordnen und sich in diese zu integrieren. Aber: auch eine Umkehrung dieses Prinzips muss möglich sein. Ein DMS kann für bestimmte Entitäten des Unternehmens das führende System sein (Beispiel: Korrespondenz in Form von Geschäftsbriefen per Mail, Fax oder gescanntem Papier). Erfolgt die Recherche und Anzeige im DMS, sollte es möglich sein, Kontextinformationen aus einem ERP-, CRM- oder Vorgangssystem anzuzeigen, ohne dass diese im DMS gespeichert sein müssen. Erfolgt die Recherche im CRM-, ERP- oder Vorgangssystem, müssen die relevanten Dokumente ermittelt werden können, ohne dass diese erst im entsprechenden System zusätzlich verlinkt/referenziert werden. Mit anderen Worten: zukünftig sollte jede Anwendung darauf vorbereitet sein, sowohl führendes System für bestimmte Daten zu sein, als auch führende Daten anderer Systeme zu konsumieren.

Was verändern die Cloud-Technologien?

Die beschriebenen Konzepte und Forderungen klingen nach unerfüllbaren Wunschträumen. Ganz unerwartet bekommen diesen jedoch starken Rückenwind durch die Einführung von Cloud- und die aktuellen Entwicklungen in den Web-Technologien: die flächendeckende Verwendung von REST (Representational state transfer) als Transport-Protokoll und JSON (JavaScript Object Notation) als Datenformat.

Der IT-Laie wird nun denken, es handle sich wieder einmal um eine neue Marketing-Welle und den damit verbundenen Buzz-Words der Branche. Es lohnt sich aber, den Unterschied zu bisherigen Integrations-Technologien zu verstehen (die von Webservices über SOAP und Daten in Form von XML geprägt waren):

- Bei JSON handelt es sich um ein auch vom Menschen einfach zu lesendes Datenformat, welches speziell für die Verwendung von JavaScript entwickelt wurde. Im Gegensatz zu XML können die Daten direkt von jedem aktuellen Internet-Browser oder anderen Laufzeitumgebungen für JavaScript verarbeitet werden, ohne dass es hierfür einer zusätzlichen Serversoftware bedarf. Das Format ist ganz bewusst sehr einfach gehalten und deren Komplexität nicht mit der von XML vergleichbar.
- REST ist das ideale Protokoll für die Verwendung von JSON. Es handelt sich um eine Erweiterung von HTTP und ermöglicht zusätzliche Methoden für den einfachen und pragmatischen lesenden wie schreibenden Zugriff auf Daten. REST ist im Gegensatz zu SOAP (umgangssprachlich auch Web-Services genannt) grundsätzlich synchron (es wird sofort eine Antwort geliefert). REST stellt eine einfache Alternative zu ähnlichen Verfahren wie SOAP und WSDL.
- JSON über REST und JavaScript sind heute die Grundlage für fast alle modernen Web-Technologien und Anwendungen. Kaum ein Software-Entwickler kommt heute noch ohne diese Technologien und die hierfür notwendigen Erfahrungen aus.
- Nicht nur in der ECM- bzw. DMS-Branche hat sich JavaScript als Standard-Skriptsprache für die Umsetzung von einfacher Business-Logik durchgesetzt. JavaScript wurde u. a. als ECMAScript (ECMA-262, ISO/IEC 16262) standardisiert.

Was haben diese Technologien nun aber mit den Cloud-Technologien zu tun?

Cloud-Anbieter von Unternehmenssoftware wie Dropbox, Google, Microsoft und Salesforce müssen eine große Hürde nehmen: Die Integration in die unternehmensinternen Anwendungen.

Unternehmen sind es gewohnt, Software an die eigenen Anforderungen anzupassen und in die oft gewachsene Infrastruktur zu integrieren. Cloud-Anbieter können jedoch für all' Ihre Kunden nur ein und dieselbe Software in unveränderter Form anbieten. Eine klassische Integration über Datenschnittstellen in die führenden Systeme des Unternehmens kommt für die Cloud-Anbieter nicht in Frage: Die Unternehmensnetze sind durch mehrere Firewalls voneinander getrennt. Der Cloud-Anbieter kann unmöglich all' die proprietären Schnittstellen und Datenformate der beim Kunden laufenden Anwendungen unterstützen. Dies entspricht ganz und gar nicht den Erwartungen und Erfahrungen der Unternehmen. Auf den ersten Blick ein Dilemma.

Anbieter wie Salesforce haben damit vor Jahren begonnen, die Integration der CRM-Daten auf den (Web-)Client zu verlagern. Per REST-Protokoll werden die Daten vom Cloud-Dienstleister direkt vom Browser durch JavaScript in die unternehmenseigene Anwendungsoberfläche eingebunden – und zwar ohne auf die APIs der eigenen, unternehmensinternen Anwendungen zuzugreifen zu müssen. JavaScript fügt die Daten nachträglich im Browser auf der Seite ein, nachdem diese vom Anwendungsserver des Unternehmens empfangen wurde. Es ist sogar möglich, ganz neue, angepasste Web-Anwendungen zu erstellen, die mehrere, unterschiedliche REST-Dienste

nutzen, ohne dass hierfür überhaupt eine aktive Serverkomponente notwendig wäre. Alles geschieht direkt aus dem Browser heraus.

Da jedoch Stammdaten nicht nur in der Anwenderoberfläche eine Rolle spielen, werden immer mehr Serversysteme nachgerüstet, ebenfalls über REST und JSON auf die Daten anderer Systeme zuzugreifen, sodass sich dieser Mechanismus klammheimlich als Quasi-Standard etabliert. Und zwar ohne dass sich große Unternehmen um dessen Standardisierung bemüht hätten. Dies ist vor allem deshalb möglich, weil die genutzten Technologien so einfach gehalten sind, die vorhandenen Transportwege des Internets genutzt werden und weil diese schnell umzusetzen sind.

Was heißt das für die Unternehmen heute?

Die ursprünglich für den Internet-Browser entwickelten Technologien haben bereits auch in vielen Serversystemen und Anwendungen Einzug erhalten. Es lohnt sich also, die bisher gewohnte Verwendung und Integration von Metadaten zu überdenken. Mit diesen lassen sich Daten aus anderen Anwendungen viel leichter in den Oberflächen und in der Anwendungslogik der Unternehmenssoftware-Landschaft integrieren, auch wenn dies in vielen Unternehmen und bei einigen etablierten Anbietern noch nicht angekommen zu sein scheint. Mancher Anbieter von Unternehmenssoftware hängt diese Veränderungen in den eigenen Produkten auch nicht an die große Glocke. Sie sind ja schließlich für den Betrieb der eigenen Cloud-Angebote gedacht. Der aktuelle Exchange-Server beispielsweise verfügt über eine beachtliche REST-Schnittstelle, die auch unternehmensintern ohne das Cloud-Angebot genutzt werden kann, um MDM-Konzepte zu realisieren.

Wie man sich denken kann, versuchen einige Produkthanbieter die Konzepte von MDM über REST und JSON für Marketing und schnellere Marktdurchdringung zu nutzen, indem diese Techniken primär für die Nutzung durch *andere* Anwendungen umgesetzt und angeboten werden ohne dass die Anwendung selbst diese Technologien verwendet um auf Daten anderer Systeme zuzugreifen. Gelinde gesagt: Wenn Unternehmen die konsequente Nutzung *und* Verwendung von Stammdaten über einfache und transparente Technologien wie JSON über REST von Ihren Software-Anbietern nicht fordern, verringert dies die Wahrscheinlichkeit ihrer Umsetzung. Der Druck ist jedoch bereits vorhanden, da neue Lösungen (vornehmlich aus dem Open-Source-Spektrum und von Cloud-Anbietern) aggressiv diese Konzepte verfolgen. Dies führt bereits zu einer Veränderung der Machtverhältnisse bei den ERP- und CRM-Anbietern, ohne dass dies von den Etablierten bemerkt wird.

Was heißt das für die Nutzung von DMS-, ECM- und Unternehmenssystemen?

Dokumente sind fast immer in einem Kontext eingebunden, der sich auch nach dem Speichern des Dokuments ändern kann. Metadaten sollten möglichst nur noch in Systemen gespeichert werden, wenn dies aus fachlichen und/oder technischen Gründen unbedingt notwendig ist oder es sich um das führende System für diese Daten handelt. Es führt zwangsläufig zu Inkonsistenzen, wenn Stammdaten an verschiedenen Stellen redundant gespeichert werden, ohne dass Automatismen deren Konsistenz sicherstellen. Besser ist es daher, nur eine verlässliche Referenz auf externe Informationen zu speichern.

Einige führende DMS-Anbieter haben aus Sicht des Anwenders in der Vergangenheit einen guten Job geleistet, indem sie sich in die gängigsten Anwendungssysteme und Szenarien wie Mail-Client, Office-Anwendung oder als virtueller Drucker integrieren. Diese Ansätze werden jedoch nie das Potential umsetzen können, wie es durch die Entkoppelung über eine Masterdata-Schicht möglich sein wird. Auch in der klassischen EAI (Enterprise Application Integration) hat man erkannt, dass eine direkte Integration aller Systeme mit allen Systemen über proprietäre Schnittstellen zu einer kaum zu beherrschenden Komplexität führt. Zudem sind so integrierte Systeme sehr unflexibel und anfällig auf Veränderungen.

Das spannende am Masterdatamanagement und der dadurch ermöglichten Referenzierung von Kontext unabhängig einzelner Anwendungen ist, dass es eine sehr viel effizientere Nutzung von IT in Prozessen ermöglicht. Vor allem durch den Mittelstand, der nicht wie die Banken- und Versicherungsbranche fast seine gesamte Unternehmenssoftware selbst entwickelt hat.

Ich wage sogar eine noch weitgehendere These: Heute noch marktführende Systeme werden in absehbarer Zeit an Bedeutung verlieren, wenn diese sich nicht den MDM-Konzepten und Technologien öffnen. Damit einher geht auch die Öffnung und teilweise notwendige Umstrukturierung der heutigen Datensilos, die sich mehr an den MDM-Entitäten als an den Anbietern und Serversystemen orientieren müssen.

Was heute noch für eine pragmatische Umsetzung fehlt, ist das, was heute der Browser für die Cloud-Anbieter leistet: eine unternehmensweite oder gar -übergreifende Schicht, die es beherrscht, per REST auf die wichtigsten

Stammdaten eines typischen Unternehmens zuzugreifen. Brückentechnologien können helfen, proprietäre Schnittstellen über den gleichen Weg erreichbar zu machen. Eine solche Schicht muss auch kaufmännisch und lizenztechnisch für den Einsatz geeignet (sprich: bezahlbar) sein. Die Umsetzung und Verbreitung einer solchen Schicht könnte möglicherweise genauso still und heimlich, rein durch die normative Kraft des faktischen erfolgen. Genauso wie sich auch die Verbreitung von REST und JSON als pragmatischer Ansatz für die Integration von Cloud-Systemen etabliert haben.