

Die Rollen von Data Stewards und Data Scientists



Successful and effective data stewards reside in the business, are visible, respected and influential – they must have the vision to understand the importance of data quality to the overall business objectives, as well as the impact of quality issues on downstream business processes. The leader of the steward team, i.e. the corporate sponsor for data quality, will serve as the ultimate point of decision on issues and conflicts occurring across stewards and teams. Ted Friedman, Gartner Group¹

Informationsmanagement im Unternehmen

Bereits vor dem Zeitalter von Big Data war die Frage, wer für welche Daten und ihre Qualität im Unternehmen verantwortlich ist, mehr als essentiell, denn Information entscheidet alles. Es ging schon immer darum, dem richtigen Kunden rechtzeitig das richtige Produkt mit der richtigen Rechnung an den richtigen Ort zur richtigen Zeit in der richtigen Menge zu liefern. Das war und ist die Zielsetzung von Informationsmanagement. Die Wichtigkeit von Informationsmanagement wurde und wird nicht immer anerkannt. Wenn aber Information nicht korrekt, vollständig und rechtzeitig zur Verfügung steht, dann gibt es Prozessabbrüche oder -verzögerungen, Stornos, Retouren, Stillstand, Verärgerung auf der Kundenseite und einiges an Unbequemem und Ärgerlichem mehr. Nicht die richtige Information zu haben kostet viel Geld, Umsatz und verursacht auch noch Kosten. Prozesse

können nur dann (optimal) ablaufen, wenn die Information stimmt. Informationsmanagement tut also Not. Bleibt die Frage, wer für Informationsmanagement zuständig und folglich verantwortlich ist, wenn die Information nicht stimmt.

Die Aufgaben von Data Stewards

Die Aufgaben des Informationsmanagements und der Qualitätssicherung von Daten im Unternehmen nehmen immer häufiger Data Stewards wahr. Sie arbeiten im Kontext von Information Governance, die die Regeln und Prozesse definiert, die im Informationsmanagement gelten beziehungsweise ablaufen. So soll sichergestellt werden, dass Information korrekt, konsistent und durchgängig im ganzen Unternehmen angelegt, gemanagt und genutzt wird. Data Stewards sorgen dafür, dass die Information Governance von allen im Unternehmen bei allen Tätigkeiten beachtet und umgesetzt wird. Diese

Aufgabe sollten sie nicht als polizeiliche Kontrollinstanz erfüllen, sondern im Gegenteil als Dienstleister für die Fachabteilungen. Dazu gehören unterschiedliche Tätigkeiten. Data Stewards sollten eine führende Rolle bei der Entwicklung von Datendefinitionen einnehmen. Sie sollten Data Profiling unterstützen, um Fehler in Daten aufzudecken und die Auswirkungen solcher Fehler abschätzen. Sie sollten die Nutzung von Daten propagieren und für Datensicherheit sorgen. Sie sollten auch über die Einhaltung aller Regeln und über die Qualität der Daten mittels Monitoring wachen. Sie sind auch beteiligt, wenn es um die Priorisierung von Datenqualitätsmaßnahmen geht.

Die Aufgaben von Data Stewards sollten natürlich auch im Rahmen der Unternehmensstrategie und der Unternehmensziele priorisiert werden. So sollte beispielsweise ein Softwareunternehmen, wie es die in Pforzheim

¹ siehe <http://www.gartner.com/newsroom/id/589207>

“Data scientists turn big data into big value, delivering products that delight users, and insight that informs business decisions. Strong analytical skills are given: above all a data scientist needs to be able to derive robust conclusions from data.”
Daniel Tunkelang, Principal Data Scientist, LinkedIn

ansässige Uniserv praktiziert, den Schwerpunkt der Arbeit von Data Stewards auf die Kundendaten legen, die in den Vertriebs- und Marketing-Prozessen genutzt werden, auf denen das Unternehmen aktuell einen Fokus gelegt hat. Einen Fokus auf Kundendaten ganz allgemein legen, wäre hier eine zu unscharfe Vorgabe, die in der Regel nicht zum Erfolg führen wird.

Die Organisation von Data Stewards

Data Stewards können mit ihrer Arbeit im Unternehmen starten, auch wenn noch keine ausgefeilte Information Governance besteht. Ein kleines gemischtes Team aus Fachabteilungen und der IT kann mit der Formalisierung von Datendefinitionen und Informationsmanagement-Prozessen und Nutzungsregeln beginnen. Die Mitarbeiter/innen sollten gute Kommunikatoren und Vermittler sein sowie geschickt verhandeln können. Sie müssen nicht unbedingt Vollzeit in das Team delegiert werden, aber man sollte darauf achten, dass genügend Zeit für die Tätigkeit als Data Steward eingeplant wird. Das Einstellen neuer Mitarbeiter für diese Aufgaben hat sich in der Regel nicht bewährt, da Data Stewards das Unternehmen inklusive der informellen Informationskanäle gut kennen müssen, um erfolgreich arbeiten zu können.

Das Team vernetzt sich dann im Unternehmen und identifiziert in den verschiedenen Fachabteilungen und in der IT dezidierte Ansprechpartner. Es hat auf diese Weise auch die Aufgabe, eine funktionierende Partnerschaft zwischen den Fachabteilungen und der IT zu etablieren. Organisatorisch fasst man am besten das Team mit seinen Ansprechpartnern in den Fachabteilungen und der IT in einem Kompetenzzentrum zusammen. Ein solches hat einen Sponsor auf der Führungsebene des Unternehmens und wird von einer Doppelspitze geleitet, einem Mitarbeiter aus der Fachabteilung, der als Primus inter Pares agiert, und einem Mitarbeiter

aus der IT. Es hat ein Budget, verantwortet das Programm und entwickelt seine Methodologie, die Information Governance.

Data Steward-Programme stoßen mitunter im Unternehmen auf kulturelle Probleme. Beispielsweise wollen manchmal von einer Fachabteilung Probleme mit schlechter Datenqualität nicht gesehen werden, weil in dieser Fachabteilung die Daten zwar entstehen, aber hauptsächlich an anderer Stelle genutzt werden. Daher hat eine solche Fachabteilung möglicherweise wenig Interesse, in Datenqualität zu investieren. Hier können Data Stewards für Transparenz sorgen und aufzeigen, welche Probleme dem Gesamtunternehmen entstehen und wie man diese Fachabteilung in Sachen Datenqualität unterstützen kann. Kommt man allerdings auch dann nicht weiter, so wird man im Unternehmen an dieser Stelle ein formales Change Management einsetzen müssen.

Big Data-Analytik

Data Stewards haben die Verantwortung für die Unternehmensdaten und/oder – im jeweiligen Umfang – für Daten aus den Fachabteilungen. Jetzt kommt Big Data. Wie beeinflusst das Management von Big Data die bisherigen Aufgaben und Zielsetzungen von Data Stewards? Schauen wir zuerst auf die Anforderungen von Big Data-Analytik. Sie erfordert gegenüber der traditionellen Analytik neue Skills und Rollen, die sich organisatorisch gesehen am besten in einem erweiterten BI-Kompetenzzentrum ansiedeln lassen. In einigen Unternehmen wie Amazon, eBay, Facebook, Google, Twitter u.a., die sich schon einige Zeit mit Big Data beschäftigen, haben sich solche neuen Rollen gebildet, die zusammen ein Big Data-Team ausmachen. Die fünf wichtigsten Rollen in einem solchen Team findet man in einem Vorschlag von McKinsey & Company². Dieser Beitrag beschäftigt sich zwar mehr mit Big Data-Marketing-

projekten, aber er ist allgemein genug geschrieben, um auf andere Big Data-Projekttypen übertragen werden zu können. Insbesondere identifiziert der McKinsey-Ansatz Rollen statt Job-Bezeichnungen. Daraus lassen sich dann direkt die notwendige Expertise und die Job-Beschreibung ableiten. Es geht um die fünf folgenden Rollen (Ich habe die amerikanischen Bezeichnungen stehen gelassen):

- Data Hygienists stellen sicher, dass die Daten bereinigt und richtig sind und auch über den Lebenszyklus der Daten so bleiben. Dieses Data Profiling und Cleansing beginnt ganz am Anfang des Projektes, wenn die ersten Daten erfasst werden. Daran sind alle Team-Mitglieder beteiligt, die diese Daten nutzen wollen.
- Data Explorers durchsuchen das Big Data-Universum, um die Daten aufzufinden, die man im Projekt braucht. Dazu gehört auch die Aufbereitung der Daten für das Projekt, denn die meisten Daten draußen waren niemals erzeugt, um analytisch untersucht zu werden, sind also weder für eine Analyse geeignet noch gespeichert noch organisiert.
- Business Solution Architects haben die Aufgabe, die identifizierten Daten zusammenzustellen und für die Analyse vorzubereiten. Dazu werden die Daten auch für die erwarteten Abfragen strukturiert. Daten, die im Minuten- oder Stundentakt benötigt werden, müssen dann auch entsprechend aufgefrischt werden.
- Data Scientists übernehmen das Organisieren der Daten und das Bauen von analytischen Modellen im Rahmen des Projektes. Dazu gehört auch das Überprüfen, Ändern und Ersetzen von Modellen, wenn notwendig, sowie die Kommunikation von analytischen Resultaten gegenüber dem Vorstand und dem gesamten Unternehmen.
- Campaign Experts haben die Aufgaben des Interpretierens der Ergebnisse und dem entsprechenden Umsetzen in Aktionen. Dazu gehören

² siehe im Blog „Five Roles You Need on Your Big Data Team“ des Harvard Business Review (Juli 2013) http://blogs.hbr.org/cs/2013/07/five_roles_you_need_on_your_bi.html?utm_source=Socialflow&utm_medium=Tweet&utm_campaign=Socialflow

auch das Priorisieren von Kanälen und das Festlegen der Kampagnen-Sequenzen.

Die Rollen der Data Explorers und Campaign Experts benötigen Expertisen wie Cognitive Scientists und Behavioral Economists. Solche Expertise ist notwendig, um zu identifizieren, welche Daten für das Projekt wichtig sind und welche nicht. Sie ist auch von großer Hilfe in der Interpretation von Ergebnissen und entsprechenden Umsetzungen. Daher ist McKinseys Rollenmodell so wichtig. Wenn man Aufgaben und Rollen verstanden hat, weiß man, welche Experten man im Projekt braucht.

Die Rollen von Data Scientists

Auf die Data Scientists wollen wir noch genauer eingehen. Das sind Mitarbeiter, die eher in der Business Intelligence angesiedelt sind, mit folgendem weiter detaillierten Profil:

- Technische Expertise: Tiefe Kenntnisse in einer Natur- oder Ingenieurs-Wissenschaft sind notwendig. Sie bilden die Grundlage, um als Data Scientist erfolgreich arbeiten zu können. Insofern sollte man zukünftige Data Scientists in dieser Gruppe suchen und dann auch die weiteren geforderten Eigenschaften testen.
- Problembewusstsein: die Fähigkeit, ein Problem in testbare Hypothesen aufzubrechen.
- Kommunikation: die Fähigkeit, komplexe Dinge per Anekdoten durch einfach verständliche und gut kommunizierbare Sachverhalte darzustellen.
- Kreativität: die Fähigkeit, Probleme mit anderen Augen zu sehen und anzugehen („thinking out of the box“).

Data Scientists und Data Stewards: ein „Dream-Team“?

Wie arbeiten jetzt das Business Intelligence-Kompetenzzentrum, in dem die Data Scientists sitzen, und das Informationsmanagement-Kompetenzzentrum, in dem die Data Stewards sitzen, zusammen?

Die Schnittstelle bilden die Data Hygienists, die die Aufgaben von Data Stewards in Big Data-Initiativen übernehmen. Hier sehen wir bereits den Unterschied in der Arbeitsweise von Data Stewards. Im Unternehmen sind Data Stewards das ausführende Organ der Information Governance, die unternehmensweit in gleichem Maße für alle Unternehmensdaten Gültigkeit hat. In Big Data-Initiativen wird die Richtigkeit und Vollständigkeit der Daten pro Projekt jedes Mal neu festgelegt. Es gibt keine durchgängige Data Governance, sondern eine Data Governance, die pro Projekt vor allem von den Data Scientists bestimmt wird.

Das kann auch bedeuten, dass manche Big Data-Projekte komplett ohne Governance ablaufen. Data Scientists argumentieren hier, dass eine Bereinigung der Daten aus den verschiedenen Big Data-Quellen die analytischen Ergebnisse beeinflussen könnte, weil dann die Daten durch Annahmen der Data Stewards zur Datenqualität verfälscht werden könnten. Auch könnten gerade in Ausreißern wichtige Informationen stecken, die man im Rahmen von Data Discovery unbedingt bräuchte. Dahinter steht auch der gute Gedanke, dass in Big Data-Projekten die Daten dem Zweck des Projektes zu dienen haben, während im Unternehmen die Daten der langfristig angelegten Unternehmensstrategie zu dienen haben.

Der Autor



Dr. Wolfgang Martin ist ein europäischer Experte auf den Gebieten

- Business Intelligence, Performance Management, Analytics, Big Data
- Business Process Management, Information Management, Information Governance
- Customer Relationship Management (CRM)
- Cloud Computing (SaaS, PaaS)

Sein Spezialgebiet sind die Wechselwirkungen technologischer Innovation auf das Business und damit auf die Organisation, die Unternehmenskultur, die Businessarchitekturen und die Geschäftsprozesse.

Weitere Information auf www.wolfgang-martin-team.net oder @wmartinteam.

Fazit: Die Aufgabe von Data Stewards ist das Informationsmanagement der Unternehmensdaten und/oder Daten aus Fachabteilungen unter den Rahmenbedingungen der Information Governance – mit der Zielsetzung, den Fachabteilungen qualitativ hochstehende Daten, die leicht zugänglich sind, in der notwendigen Konsistenz zur Verfügung zu stellen. Sie sind auch dafür verantwortlich, die Information Governance kontinuierlich an der Unternehmensstrategie auszurichten.

Data Scientists haben die Aufgabe, Big Data in „Big Value“ zu wandeln. Sie sind verantwortlich für die Methodologie von Big Data-Analytik sowie die Kommunikation von analytischen Resultaten gegenüber dem Vorstand und dem gesamten Unternehmen.

Data Stewards haben auch in Big Data-Initiativen ihre Rolle, nämlich die der Data Hygienists, die sich aber jeweils an den Projektzielen und nicht an der Unternehmensstrategie ausrichtet. Sie spielen in der Big Data-Analytik die Rolle eines SWAT-Teams, also eines taktisch agierenden Spezialteams, und nicht die strategische Rolle wie im Unternehmen. Insofern sind sie wieder ein Dienstleister, der bei Big Data-Initiativen von den Data Scientists gerufen wird, aber nicht selbständig agiert.