

Datenqualitäts-Checkliste für neue Datenhaushalte

Dr. Marcus Gebauer

WestLB AG, Datenqualitätsbeauftragter, Düsseldorf, Germany
Marcus_Gebauer@WestLB.de

Abstract: Ursachen schlechter Datenqualität finden sich in vielen Bereichen von Organisationen. Neben den direkten Einflussfaktoren wie IT und IT-Systemlandschaft, Prozesse und nicht zuletzt die Mitarbeiter, sind es vor allem die Projekte, in denen die Grundlage der zukünftigen Datenqualität gelegt wird. Wir zeigen worauf schon in der Projektphase neuer Datenhaushalte geachtet werden muss, um in der Produktivphase qualitativ hochwertig Informationen bereitstellen zu können.

1 Einleitung

Unternehmen erkennen heute die Notwendigkeit, ihre Organisation nach abstrakten Stellgrößen zu steuern, da die Einflussfaktoren auf die Unternehmenssteuerung zunehmend vielfältiger werden. So kennen die meisten Banken die Positionen des Chief Financial Officers (CFO) und Chief Risk Officers (CRO), die die wesentlichen Steuerungszahlen verantworten. Daneben haben viele die Position des Chief Information Officers (CIO) implementiert, der für die Bereitstellung der für den Betrieb notwendigen IT und IT-Infrastruktur verantwortlich ist. Während in den meisten Unternehmen das Hauptaugenmerk des CIO auf die Bereitstellung von Hardware, Netzwerktechnik und Software in Projekten gelegt wird, sind die in den Systemen lebenden Daten und deren Qualität häufig nicht im Fokus. Dabei sind es gerade diese Daten und Informationen, die den wesentlichen Produktionsfaktor einer Bank mit ihren virtuellen Produkten ausmachen.

Doch welche Risiken für die Informationsqualität¹ werden in den Projekten, seien es nun Projekte der IT oder der Fachbereiche, gelegt? Diese Frage zu beantworten, nimmt sich die Generische Checklist als Aufgabe.

2 Informationsqualität

Daten spielen in der Informationsgewinnung eines Unternehmens eine zentrale Rolle. Dabei ist ihre Qualität entscheidend abhängig von Prozessen, Personal und IT und wirkt sich am Ende auf das Einkommen und die Kosten eines Unternehmens aus [JG98].

¹ Die Begriffe Daten- und Informationsqualität werden synonym verwendet.

2.1 Definition und Zusammenhang

Daten werden klar strukturiert und haben immer einen offiziellen, formellen, zweckorientierten Charakter. Neben den Daten hat auch der Kontext, in dem Daten leben einen offiziellen, formellen Charakter. Daten und Kontext charakterisieren zusammen „Formelle Informationen“, die zusammen mit „Informellen Informationen“ die Informationslandschaft einer Organisation bilden. Formelle Informationen dienen zur Planung, Steuerung und Darstellung der Organisationsstruktur, entstehen in Prozessen und werden für Prozesse benötigt entstehen (Abb. 1). Abbildung 1 stellt die Merkmale des Kontext und der Daten dar. Der Kontext gibt eine Aussage über den Zweck und den Prozess, für den Informationen benötigt werden. Daneben klassifizieren Format, Inhalt, Objekt (auf das sich das Datum bezieht), Stabilität, Struktur und Verarbeitung die Daten, die im Kontext ‚leben‘. Datenhaushalte müssen daher so gestaltet sein, dass sie den Anforderungen der mit ihnen abgebildeten Informationen gerecht werden.

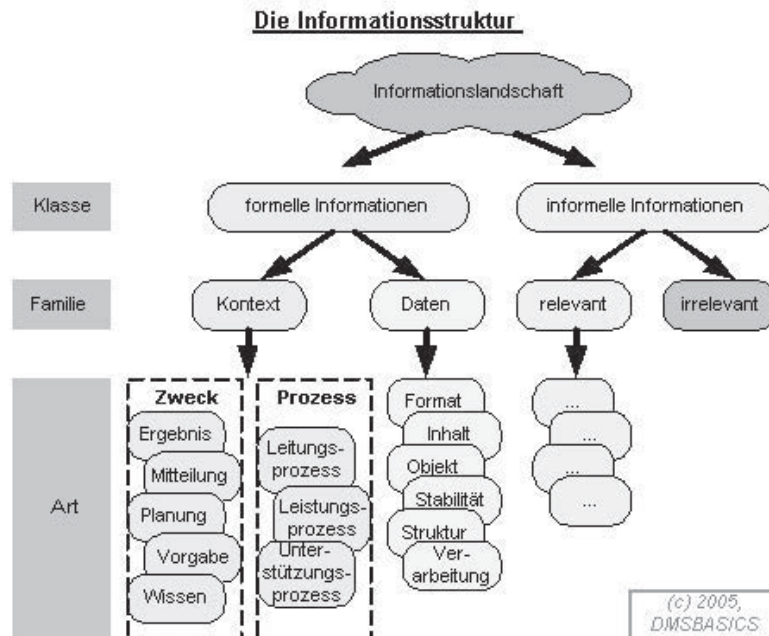


Abb. 1: Differenzierung der Informationslandschaft nach DMSBASICS [RT05]

2.2 Einflussfaktoren der Informationsqualität

Das Management der Informationsqualität beeinflusst nahezu alle Bereiche der Bank und wird ebenso durch diese Bereiche beeinflusst. Hierbei ist die Betrachtung der technischen und der funktionalen Bereiche notwendig. Die technischen Subfaktoren bestehen aus denen der Informationstechnologie und die funktionalen unterteilen sich nochmals nach Prozessen und Personal (siehe Abb. 2).

technisch	Funktional	
IT & IT Landschaft	Prozesse	Personal
- IT-Strategie - Hardware und Software - Systemdokumentation - Schnittstellenmanagement	- Dokumentation - Prozessmanagement - Kontrollen - Verantwortungen	- Kommunikation - DQ-Bewusstsein - Unternehmenskultur - Personalausstattung
objektiv messbar		subjektiv messbar

Abb. 2: Die drei Einflussfaktoren IT, Personal und Organisation mit ihren Komponenten.

3 Umsetzung

Aus den im Kapitel 2 beschriebenen Definitionen und Einflussfaktoren leiten sich die Untersuchungsbereiche der DQ-Checkliste für (neue) Datenhaushalte ab. Im Folgenden werden die Anforderungen in den Bereichen

- Organisatorische Spezifikationen,
- Prozessdefinitionen,
- Datenerfassung, -bearbeitung und -nutzung und
- Management, Messen, Maßnahmen und Monitoring von Datenqualität

näher beschrieben und die wesentlichen Fragen dargestellt. In der Vollversion der Checklist werden diese vier Bereiche durch ‚Management-Anforderungen‘ und ‚Technische Anforderungen‘ ergänzt. Da die Anforderungen dieser im Allgemeinen recht gut erfüllt sind, wird an dieser Stelle nicht weiter darauf eingegangen.

3.1 Organisatorische Spezifikationen

Jede Information und jedes Datum erfüllt einen definierten Zweck und ist eindeutig einem Leistungsprozess (Abb. 1) zuzuordnen. Damit ist die Qualität des Leistungsprozesses direkt mit der der Information verknüpft. Um die Qualität der Daten zu garantieren sind die Rollen des Dateneigentümers, -erfassers, -bearbeiters und -kunden für alle zu verarbeitenden Datenfelder (inkl. Aufgaben) zu spezifizieren. Dies kann in so genannten Informationsgruppen geschehen, um den Aufwand pragmatisch zu halten.

- Dateneigentümer, -erfasser, -bearbeiter und -kunde sind bestimmt
- Aufgaben und Verantwortungen sind klar beschrieben und abgestimmt
- Die Prozessbeteiligten haben an einem DQ-Training teilgenommen
- Dateneigentümer, -erfasser, -bearbeiter und -kunde stimmen sich ab

3.2 Prozessdefinitionen

Schlechte Informationsqualität hat seine Ursachen hauptsächlich in den Prozessen einer Organisation. Sind die unter Abschnitt 3.1 beschriebenen Rollen definiert, so sind die dazugehörigen Prozesse zu spezifizieren.

Datensammlung, Datenmanipulation und –korrektur, Datenqualität, Veränderung, Eskalation, Prozesskontrolle

- Prozess sind angemessen beschrieben und veröffentlicht
- Beteiligte haben den Prozessen explizit zugestimmt
- Es gibt einen bekannten Prozesseigner
- Prozess-Dokumentation wird gepflegt und dies wird überprüft

Es ist klar zu regeln, was, wer, wann konkret tun soll und vor allem, was passiert, wenn Regeln nicht eingehalten werden. Erfolgt dies nicht, so sind Leitungs-, Leistungs- und Unterstützungsprozesse (Abb. 1) nicht verlässlich implementierbar.

3.3 Datenerfassung, -bearbeitung und -nutzung

Um Datenredundanzen zu vermeiden, sind neue Datenhaushalte stets gegen existierende Referenzdatenbestände (in- und externe) zu prüfen. Die Vermeidung von Datenredundanzen vereinfacht die Wartung der Daten und mindert die Kosten der Datenhaltung. Datenerfassung und Datenbearbeitung sind klar zu regeln und in den hierfür geeigneten Organisationseinheiten durchzuführen.

- Wo möglich werden Referenzsysteme als einzige Quelle für Daten genutzt
- Qualitätskontrollen für den Datenhaushalt sind implementiert
- Datensammlung erfolgt an dem hierfür geeignetsten Ort
- Prozesse zur Speicherung, Umwandlung und Anreicherung von Daten sind eingeführt

3.4 Management, Messen, Maßnahmen und Monitoring von Datenqualität

Auch wenn Datenmodell und –elemente in ihrer frühen Lebensphase noch gut beschrieben sind, so sind das Management der Datenqualität, deren Messung und Vereinbarungen zu diesen Messungen für die zukünftige Entwicklung der Datenhaushalte von essentieller Bedeutung.

Das Verständnis für die Daten und ihrer logischen Abhängigkeiten verwischt und verändert sich mit der Zeit. So müssen Anforderungen an DQ-Messungen so früh wie möglich schon in der Projektphase spezifiziert werden. Die Anforderungen an DQ und deren Messungen sollten verständlich beschrieben und nachvollziehbar durchgeführt werden. DQ-Messungen sind eng mit den Bedürfnissen der Datenkonsumenten korreliert und kann von Nutzer zu Nutzer selbst des gleichen Datenbestandes variieren. Daher sind die Anforderungen stets von Standpunkt der jeweiligen Geschäftsfelder in Form von Geschäftsregeln zu beschreiben.

- Qualität von Datenhaushalten wird regelmäßig überprüft
- Abstimmungen mit Referenzsystemen werden durchgeführt
- Abstimmungen sind beschrieben bzgl. Wer, Was, Wann und Wie konkret durchzuführen hat
- Geschäftsregeln und ein Standard-Satz von DQ-Metriken

3.5 Messwerte

Die gesamte Checkliste umfasst 87 Fragen, die auf einer sechsstufigen Skala durch einen Assessor auf ihre Einhaltung hin beurteilt werden. Über alle Bewertungen eines Fragenbereiches wird eine gemittelte Maßzahl berechnet. Die gewichteten Ergebnisse der einzelnen Bereiche werden zu einer Qualitätszahl zusammengefasst.

4 Zusammenfassung

Die beschriebene Checkliste bietet einen Satz von Anforderungen, die es dem Assessor erlaubt, den Stand neuer Datenhaushalte bezüglich des Risikos für Datenqualität zu bewerten. Für jeden der besprochenen Bereiche erhält der Assessor eine Maßzahl, aus deren Bewertung Aktionen abgeleitet werden können. Nicht alle Anforderungen werden auf alle Datenhaushalte anwendbar sein. Es liegt am Assessor die richtigen Fragen anzuwenden. In einigen Punkten geht die Checkliste über die Anwendung auf neue Datenhaushalte hinaus. Dies ist jedoch notwendig, um schon in der Projektphase Strukturen für das Management der Datenqualität in der Produktivphase zu legen.

Literaturverzeichnis

- [JG98] Juran, J.M.; Godfrey, A.B.: Juran's Quality Handbook. McGraw-Hill, New York, 1998.
[RT05] Tremba, R.: <http://www.dmsbasics.de>, München, 2005.