

Checkliste für eine bessere Datenqualität

Was können Sie mit der Checkliste erreichen?

- Häufige Datenqualitätsprobleme vorausschauend erkennen
- Anregung für eine Datenqualitätsstrategie erhalten
- Strategisches Vorgehen bei Datenqualitätsproblemen als Teil einer Data Governance implementieren

15 Dimensionen zur Messung von Daten- bzw. Informationsqualität

Diese 15 Dimensionen können Sie dabei unterstützen, Datenqualität systematisch zu beurteilen.

<input type="checkbox"/>	Zugänglichkeit (accessibility) Kann ich die Daten auf einfache Weise abrufen?	System
<input type="checkbox"/>	Bearbeitbarkeit (ease of manipulation) Ist es möglich, dass ich die Daten ändern/transformieren kann?	
<input type="checkbox"/>	Glaubwürdigkeit (believability) Sind den Daten weitere Informationen (z.B. Zertifikate oder ausführliche Berichte über Datengewinnung und -aufbereitung) beigefügt, die Vertrauen in die Datenqualität erwecken?	Inhalt
<input type="checkbox"/>	Hohes Ansehen (reputation) Steht die Datenquelle für kontinuierlich qualitativ hochwertige Daten? (ähnlich Glaubwürdigkeit)	
<input type="checkbox"/>	Objektivität (objectivity) Enthalten die Informationen keine Wertungen und sind sachlich?	
<input type="checkbox"/>	Fehlerfreiheit (free of error) Entsprechen die Informationen dem, was sie abbilden sollen?	
<input type="checkbox"/>	Aktualität (timeliness) Sind die Informationen so aktuell, wie es der Zweck erfordert?	Nutzung
<input type="checkbox"/>	Wertschöpfung (value-added) Kann ich aus den Daten einen gesellschaftlichen oder wirtschaftlichen Wert schöpfen?	
<input type="checkbox"/>	Vollständigkeit (completeness) Sind alle notwendigen Datenfelder vollständig gefüllt?	
<input type="checkbox"/>	Angemessener Umfang (appropriate amount of data) Reicht mir der Umfang der Daten aus, um meine Frage zu beantworten?	
<input type="checkbox"/>	Relevanz (relevancy) Sind die Informationen für meinen Zweck notwendig?	
<input type="checkbox"/>	Übersichtlichkeit (concise representation) Kann ich auf die benötigten Informationen innerhalb eines Datensatzes einfach zugreifen? (Zugänglichkeit innerhalb eines Datensatzes)	Darstellung
<input type="checkbox"/>	Einheitliche Darstellung (consistent representation) Entsprechen die Informationen fortlaufend einem festgelegten Format?	
<input type="checkbox"/>	Verständlichkeit (understandability) Sind die Informationen inhaltlich einfach zugänglich?	
<input type="checkbox"/>	Eindeutige Auslegbarkeit (interpretability) Wenn verschiedene Personen mit den Daten unabhängig voneinander arbeiten, werden sie die Daten in der gleichen (korrekten) Weise interpretieren?	

Anmerkungen zu den 15 Dimensionen

- Die vorgestellten 15 Dimensionen basieren auf einer Studie der MIT-Wissenschaftler*innen Wang und Strong von 1996¹. Sie sind eine gängige Form, Datenqualität zu bestimmen. Natürlich existieren auch noch andere Herangehensweisen.
- In den seltensten Fällen sind alle Dimensionen relevant. Vielmehr sollten Sie eine Auswahl aus den 15 Dimensionen entsprechend der Sinnhaftigkeit und dem Zweck treffen.

Von 15 Dimensionen zu einer Strategie

- Sie können die 15 Dimensionen zum losen Reflektieren von vorliegender Datenqualität oder möglichen Datenqualitätsproblemen bei noch unbekanntem Datenquellen nutzen. Sie können die Dimensionen auch noch näher definieren und daraufhin Datenqualität messen und eine Bestandsaufnahme durchführen.
- Eine Messung ist empfehlenswert, weil Sie damit Veränderungen nachverfolgen können und entsprechende Maßnahmen treffen können.
- Je mehr Sie von ad-hoc-Maßnahmen zu einer ganzheitlichen Strategie wechseln, desto mehr müssen Sie Prozesse zur Datenqualitätssicherung verstetigen. Damit sind Sie auf dem besten Wege, eine Data Governance zu implementieren.
- Eine Data Governance bezeichnet ein klar definiertes Konzept, das den Umgang mit Daten innerhalb einer Organisation regelt. Das können zum einen Prozesse zur Datenqualität, aber eben auch Datenschutz und Datensicherheit sein.
- Folgende Schritte empfehlen wir zu einer Data Governance für eine höhere Datenqualität:
 - o Ein innerhalb der Organisation einheitliches Verständnis des Kernprozesses bei Datenqualitätsproblemen festlegen und verstetigen (standardisieren)
 - o Vorhandene Datenstandards verwenden und Metadaten pflegen
 - o Alle betroffenen Stellen in der Organisation einbeziehen und Verantwortlichkeiten klar benennen
 - o Prozesse transparent und nachvollziehbar machen (z.B. durch eine Versionskontrolle bei Datensatzänderungen)
 - o Auswahl der Bewertungsdimensionen stetig hinterfragen und ggf. anpassen
 - o Regelmäßig Datenqualität messen
- Metadaten: Metadaten geben mehr Informationen zum Datensatz, werden jedoch leider häufig vernachlässigt. Metadaten für eigene Datensätze zu pflegen, sollten Sie hoch priorisieren. Damit sichern Sie ab, dass man auch nach einiger Zeit einen qualitativ hochwertigen Datensatz ohne große Mehrarbeit verwenden kann. Bei externen Datensätzen müssen Sie davon ausgehen, dass diese nicht ihren eigenen Qualitätsstandards entsprechen. Das gilt auch für die entsprechenden Metadaten.

mFUND-Begleitforschung zu Data Governance

Das iRights.Lab führt ein Begleitforschungsprojekt zu „Data-Governance im Innovationsprozess“ durch. Ziel ist es, dass sich Organisationen, die an datengetriebenen Innovationen arbeiten, selbstständig und frühzeitig eine Data-Governance auferlegen, um Aspekte wie Datenschutz, Datensicherheit und Datenqualität zu gewährleisten und als Wettbewerbsvorteil nutzen zu können.
www.iriights-lab.de

Lizenz: CC BY-ND (Namensnennung-Keine Bearbeitungen 4.0 International, <https://creativecommons.org/licenses/by-nd/4.0/legalcode.de>)

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages

¹ Richard Y. Wang, Diane M. Strong: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12:4, 5–34, 1996.