

# Skalierbare Datenqualität in anwendungsspezifischen Szenarien\*

Matthias Virgin  
DZNE e.V. Bonn  
Standort Rostock / Greifswald  
Gehlsheimer Str. 20  
18147 Rostock  
matthias.virgin@dzne.de

Tobias Umbria  
Universität Rostock  
Institut für Informatik  
Lehrstuhl für Datenbank und  
Informationssysteme  
18051 Rostock  
tu@informatik.uni-  
rostock.de

Andreas Heuer  
Universität Rostock  
Institut für Informatik  
Lehrstuhl für Datenbank und  
Informationssysteme  
18051 Rostock  
heuer@informatik.uni-  
rostock.de

## ABSTRACT

Konzepte der Datenqualität haben sich neben vielen Disziplinen auch im Umfeld der Informationstechnologie etabliert und somit einen festen Platz in Aspekten der Entwicklung neuer Informationssysteme eingenommen. Das Vorhandensein vieler unterschiedlicher objektiver und subjektiver Anforderungen an die Qualität von Daten zwingt den Entwickler mittlerweile dazu, einen immer größeren Teil seiner Entwicklungszeit für die Modellierung von Konzepten zur Qualitätserfüllung zu verwenden. In diesem Beitrag wird eine erste Idee vorgestellt, um diesen Aufwand frühzeitig zu verringern. Dazu wird zunächst eine Betrachtung der sicherheitsrelevanten Qualitätsfaktoren vorgenommen. Daraufhin erfolgt eine anwendungsspezifische Einordnung der Qualitätskriterien in Gruppen ähnlicher Charakteristika mit anschließender Skalierung. Somit besteht die Möglichkeit einzelne Gruppen mit besonders hohen Qualitätsbedürfnissen zu identifizieren und zu behandeln. Ressourcen können dadurch von Anfang an schonender und auf bestimmte Qualitätskriterien hin gezielter eingesetzt werden.

## Categories and Subject Descriptors

H.1.1 [Systems and Information Theory]: Value of information

## Keywords

Datenqualität, Qualitätsbewertung, Qualitätsmodell

## 1. MOTIVATION

Die Qualität von Daten hat schon seit längerer Zeit einen hohen Stellenwert auf dem Gebiet der Informationsverarbeitung erlangt. Die große Problematik von verfälschten, un-

zutreffenden oder fehlenden Daten kann in verschiedenen Situationen und Anwendungsszenarien den täglichen Umgang mit den Daten und deren Informationsgehalt erschweren und führt zu falschen Ergebnissen in der Bearbeitung [2]. Falsche Ergebnisse, die z.B. im medizinischen Bereich Diagnosen darstellen, können durch falsche Behandlungen sogar gesundheitsschädliche Folgen mit sich führen. Auch auf dem Wirtschaftssektor können Entscheidungen aufgrund fehlerhafter Datenbestände zu gravierenden finanziellen und personellen Auswirkungen führen.

Datenqualität ist somit ein unverzichtbarer Begleiter jeglicher Informationssysteme. Ursprünglich und auch heute noch hat das Konzept der Datenqualität in der Integration heterogener Datenbestände einen besonders hohen Stellenwert. Aber natürlich sind auch für jedes andere Softwaresystem die Techniken zur Erhöhung der Datenqualität einsetzbar und empfehlenswert [9]. Deshalb ist schon bei der Planung von Softwaresystemen, unabhängig vom Anwendungsszenario, darauf zu achten, dass von vornherein ein verstärkter Blick auf die Datenqualität gerichtet wird. Qualität bedeutet eine bestimmte Eignung für einen Zweck. Das heißt, dass für jeden Zweck der Qualitätsbegriff bzw. dessen Ausprägung neu bestimmt werden muss [4]. Damit wird ausgesagt, dass jedes Anwendungsszenario einen anderen Blickwinkel auf bestimmte Qualitätsaspekte von Daten besitzt. Das bedeutet, dass jedes Szenario und dessen Zielsetzung im Hinblick auf den Einsatz bestimmter Methoden und Konzepte der Datenqualität immer wieder neu diskutiert werden muss. Diese Maßnahmen resultieren aus den verschiedenen Anforderungen an die Datenqualität, die in einem besonderen Maße betrachtet werden müssen. In der Literatur werden zwar Vollständigkeit, Fehlerfreiheit, Konsistenz und Aktualität als besonders wichtige Anforderungen beschrieben; diese können aber in unterschiedlichen Anwendungsszenarien, neben weiteren Anforderungen, verschieden wichtige Stellen einnehmen. Bei geografischen Anwendungen ist es z.B. die spezielle Anforderung an die Positionsgenauigkeit, die einen sehr hohen Stellenwert besitzt, aber in kaum einem anderen Anwendungsszenario wichtig ist [6], oder auch Forderungen nach referenzieller Integrität als Qualitätsanforderung an relationale Datenbanken [2].

Die Datenqualität ist durch die Anforderungen an sie messbar. Also sind die Anforderungen auch messbar. Somit soll-

\*Copyright is held by the authors.  
*GvD Workshop '10*, 25.-28.05.2010, Bad Helmstedt, Germany.

ten eher applikations- oder kundenbezogene Anspruchsniveaus (Soll-Wert) auf Anforderungen definiert werden, an denen die Qualität der Daten im laufenden Prozess beurteilt werden kann [2]. Somit kann es kein allgemeines Qualitätsmodell für Daten bei der Entwicklung von Softwaresystemen geben. Deshalb stellen wir hier einen Lösungsansatz vor, der ein objektives Vorgehen definiert ohne die benötigten dynamischen Aspekte einzelner Anwendungsszenarien zu vernachlässigen.

## 2. STAND DER TECHNIK

Eingangs wurde bereits die Möglichkeit einer Gruppierung von Qualitätsdimensionen erwähnt. Diese Idee, als Qualitätsmodell benannt, ist an sich nichts Neues. Um einen Überblick über vorhandene Modelle und deren Charakteristika zu geben, führen wir anschließend vier von ihnen an.

Mitte der 1990er Jahre wurde durch Wang und Strong eine Befragung bei IT-Managern durchgeführt [10]. Diese sehr oft zitierte Umfrage ergab als Ergebnis einen Umfang von 179 Anforderungen an die Qualität von Daten. Von den Autoren wurden anschließend 15 Anforderungen ausgewählt, die von der Mehrzahl der Befragten genannt wurden.

Ein Großteil der heutigen Autoren geht von annähernd diesen 15 grundlegenden, allgemeinen Anforderungen, den sogenannten Qualitätsdimensionen, aus. Auch wenn diese in diversen Publikationen oft syntaktisch unterschiedlich erscheinen, eröffnet die Kurzerklärung meist einen semantischen Zusammenhang. Die in Deutschland gebräuchlichen Dimensionen<sup>1</sup> wurden durch die Deutsche Gesellschaft für Informations- und Datenqualität auf Grundlage von Robert Wang [10] erstellt. Dabei wurden sie ins Deutsche übersetzt und mit festen Definitionen versehen. Ein Ziel dieser Festlegung war, dass Dimensionen so ausgelegt werden, dass sie zum einen überschneidungsfrei sind und zum anderen entweder einzeln oder in eindeutiger Kombination miteinander verwendet werden können [9].

Die Autoren äußerten sich zusätzlich zu einer Kategorisierung der Dimensionen nach bestimmten Merkmalen. Das Ergebnis von Wang und Strong war eine erste Zerlegungssemantik, die die Dimensionen in 4 Kategorien nach folgendem Muster einordnete:

- Informationszugang (Systemzugang, Zugangssicherheit)
- Darstellung (Interpretierbarkeit, Verständlichkeit, Manipulationsfähigkeit, Integrität und Widerspruchsfreiheit)
- Informationszusammenhang (Relevanz, Zusatznutzen, Aktualität, Vollständigkeit, Informationsumfang)
- Eigenwert (Richtigkeit, Objektivität, Glaubwürdigkeit, Reputation)

<sup>1</sup>15 Qualitätsdimensionen: Zugänglichkeit, Angemessener Umfang, Glaubwürdigkeit, Vollständigkeit, Übersichtlichkeit, Einheitliche Darstellung, Bearbeitbarkeit, Fehlerfreiheit, Eindeutige Auslegbarkeit, Objektivität, Relevanz, Hohes Ansehen, Aktualität, Verständlichkeit, Wertschöpfung

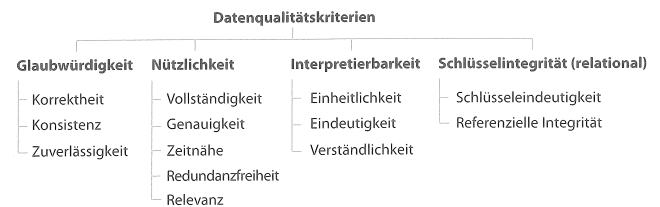


Figure 1: Qualitätsmodell nach Hinrichs [3]

80% der befragten IT-Manager durch Wang hätten ebenfalls diese Kategorisierung gewählt [9, 10]. Diese erste Systematisierung machte es möglich Datenqualität auf einer strukturierten Ebene zu behandeln, zu steuern, zu messen und im Anschluß daran auch zu bewerten. Auf Grundlage dieses Modelles sind viele weitere Modelle entstanden, die sich in ihrer Zerlegungssemantik unterscheiden und somit auch Unstimmigkeiten untereinander erzeugen. Es seien an dieser Stelle noch 3 weitere Modelle genannt, die in der Literatur erhöhtes Interesse aufweisen.

Ein Modell, das direkt von Wang abgeleitet ist, ist das von Felix Naumann (2002). Er veränderte die Kategorien (leicht, nach eigenen Aussagen) und die Einordnung der Dimensionen (teilweise anders benannt) und bemerkte zudem, dass viele Kriterien immer noch sehr subjektiv geprägt seien und eine feste Definition deshalb auch nicht sinnvoll erscheint [4].

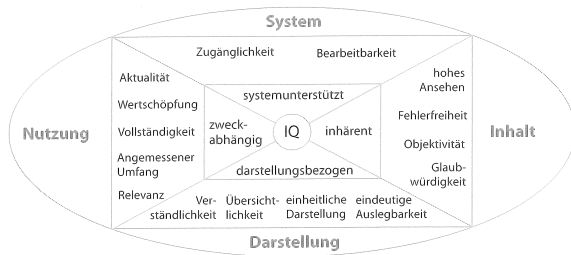
- Inhaltsbezogene Kriterien (Interpretierbarkeit, Dokumentation, ...)
- Technische K. (Verfügbarkeit, Preis, ...)
- Intellektuelle K. (Glaubwürdigkeit, Reputation, ...)
- Instanzbezogene K. (Datenmenge, Verständlichkeit, ...)

Allerdings, so Naumann, ist eine solche Liste nützlich, um für bestimmte Anwendungen die relevante Teilmenge aus den Kategorien festlegen zu können [4]. Es wird allerdings nicht auf die Art und Weise bzw. auf die Vorgehensgrundlage sowie auf die weiterführende Zerlegungssemantik eingegangen.

Um noch einmal, wie in der Motivation schon einleitend erwähnt, zu verdeutlichen, dass bestimmte Szenarien neben den grundsätzlich wichtigen Dimensionen, wie z.B. Vollständigkeit und Fehlerfreiheit, auch anwendungsspezifische Dimensionen nutzen, um die Relevanz anwendungsspezifischer Gegebenheiten auszudrücken, zeigt das Qualitätsmodell nach Hinrichs (Figure 1).

Hier finden sich in der rechten Kategorie die Dimensionen Schlüsseleindeutigkeit und Referenzielle Integrität, die speziell auf relationale Datenbanken ausgerichtet sind. Bei der Verbreitung relationaler Datenbanken wäre der Einsatz dieses Modells allerdings nach [2] als legitim anzusehen.

Das letzte hier genannte Qualitätsmodell stellt das Referenzmodell im deutschsprachigen Raum dar (Figure 2). Im Jahr



**Figure 2: Qualitätsmodell von der DGIQ [8] nach R. Wang [10]**

2007 überarbeitete die Deutsche Gesellschaft für Informations- und Datenqualität (DGIQ) die von Robert Wang ursprünglich 15 ermittelten Qualitätsdimensionen und übersetzte sie „final“ ins Deutsche und passte die Definitionen an (eindeutig, nach eigener Aussage) [8]. Die 4 Kategorien, die hier genutzt wurden, sind von der Zerlegungssemantik her dem jeweiligen Untersuchungsgegenstand zugeordnet. Also „systemunterstützt zu System“, „inhärent zu Inhalt“, „darstellungsbezogen zu Darstellung“ und „zweckabhängig zu Nutzung“ [9].

Die Anforderungen an die Qualität von Daten stellen eine Grundlage für die Messbarkeit und Bewertung der Qualität, also der Eignung von Daten für einen bestimmten Zweck, dar. Die Messung und Bewertung von Qualität ist immer noch sehr stark subjektiv geprägt. Das wird belegt durch generelle Aussagen wie: „Die Qualität von Daten ist dann gut, wenn Sie den Anforderungen an den Gebrauch der Daten erfüllen.“ [1] oder: „Die Qualität der erhobenen Daten ist abhängig von der Qualität der Erhebung.“ [7]. Die Qualität der Daten wird also immer an anderen Maßstäben gemessen.

Durch diese, wie schon erwähnt, meist subjektiven Bewertungen ist es erforderlich, zumindest eine objektive Vorgehensweise zu definieren, die die gesetzten Qualitätsdimensionen nutzt und diese dynamisch auf verschiedene Anwendungsszenarien anwendet. Die hier gezeigten Qualitätsmodelle versuchen bzw. haben die einzelnen Qualitätsdimensionen in Kategorien zusammengefasst, um den Umgang mit ihnen zu erleichtern. Allerdings fallen 2 Aspekte auf:

1. **Spezifische Dimensionen** Die Nutzung von anwendungsspezifischen Dimensionen passt das Qualitätsmodell für diese bestimmte Anwendung an, macht es aber gegenüber etablierten Modellen nicht vergleichbar. Erfahrungswerte von vorhergehenden Modellen können so nicht übernommen werden.
2. **Statische Modelle** Die hier vorgestellten Modelle sind alle sehr statisch aufgebaut. Anwendungsspezifische Szenarien können damit nicht flexibel genug beschrieben werden.

Unser Ziel ist die flexible Gruppierung durch vorhandene und anerkannte Qualitätsdimensionen für anwendungsspezifische Szenarien. Und dies so, dass gesammelte Erfahrungswerte auf später folgende Modelle mit übernommen werden können.

### 3. LÖSUNGSANSATZ

#### 3.1 Rechtliche Anforderungen als Qualitätsdimension

Einige Autoren ([1, 10] etc.) führten innerhalb ihrer Qualitätsdimensionen auch die der Sicherheit an. Grundlegend kann man sagen, dass die Dimension der „Sicherheit“ inhaltlich zu einem großen Teil aus Anforderungen an den Datenschutz bestehen und somit auch rechtliche Vorgaben widerspiegeln. Diese rechtlichen Vorgaben müssen erfüllt werden und können, wie später in diesem Beitrag zu sehen, zunächst nicht in ihrer Wichtigkeit skaliert werden. Deshalb gehört unserer Meinung nach der rechtliche Teil der Dimension Sicherheit auch nicht zur Datenqualität. Allerdings kann die Sicherheit auf Kundenwunsch hin über den Anforderungen von rechtlichen Vorgaben oder leitlinienerstellenden Einrichtungen liegen. Wir sprechen hier von einer Übererfüllung von sicherheitsrelevanten Anforderungen, die wiederum einen qualitativen Charakter besitzen. Dieser Mehrwert kann in Zahlen, und somit skaliert, ausgedrückt werden und durch den erhöhten Schutz der Daten auch indirekt die Qualität anderer Dimensionen erhöhen. Insofern werden wir in weiterer Betrachtung Aspekte der Sicherheit unabhängig der rechtlichen Anforderungen mit betrachten. Ungeachtet dessen haben die DGIQ und Autoren aktueller Publikationen die Sicherheit aus den Qualitätsdimensionen und deren Umfeld komplett entfernt. Wie wir aber im Stand der Technik festgestellt haben, sind diese Modelle recht starr und unflexibel gegenüber anwendungsspezifischen Szenarien. Deswegen sind wir der Meinung, dass man in gewissen Szenarien die Sicherheit als zusätzliche Dimension mitbetrachten kann. Auch wenn die DGIQ die Sicherheit als Dimension mit der Begründung entfernt hat, dass sie zwar zu einer Qualitätssteigerung der restlichen Dimensionen führt, aber als erschöpfend bewertet angesehen werden sollte [9], müsste man sie immer noch nebenher zu den bestehenden Dimensionen mit betrachten und nicht gänzlich aus dem Blickfeld dieser entfernen. Denn auch die Umsetzung rechtlich relevanter Sicherheitsmaßnahmen muss auf seine Qualität hin überprüft und bewertet werden.

#### 3.2 Anwendungsspezifische Gruppen

In mehreren Arbeiten, wie auch in [4], wird davon gesprochen, dass für jede Anwendung die Qualität neu bewertet werden muss. Denn jedes Unternehmen und somit auch jedes Anwendungsszenario ist anders, so dass die Datenqualität nach [9] sinnvoll umgesetzt werden muss. Allerdings bleibt es meist bei diesen Aussagen. Ein konkreter weiterführender Vorschlag ist unseres Erachtens nicht eindeutig zu finden. Für unseren Ansatz gehen wir nicht von einer strikten anwendungsunabhängigen Einordnung der Dimensionen in Kategorien aus, wie in den Qualitätsmodellen im Stand der Technik gezeigt. Eher schlagen wir vor, je nach Anwendungsszenario eigene Kategorien zu erzeugen und die Dimensionen angepasst an das Projektvorhaben einzusortieren. Die Zerlegungssemantik ist also hier sehr flexibel angesetzt. Wir nutzen dazu nur die 15 von der DGIQ vorgeschlagenen Dimensionen, um eine gewisse Möglichkeit für Vergleiche beizubehalten. Eigens definierte Dimensionen, wie in Hinrichs Modell [3], wollen wir vermeiden, damit begrifflich auf einer Ebene gearbeitet werden kann. So kann man aber z.B. eine eigene Kategorie „Referenzielle Integrität“ erzeugen, die mit bestimmten Qualitätsdimensionen durch Nutzer oder Ent-

wicklerwünschen befüllt wird. Bei dieser sehr flexiblen Erstellung einer eigenen Zerlegungssemantik ist natürlich eine sehr bewusste Auswahl der Qualitätsmerkmale und dessen Einordnung in die Kategorien zu treffen. Durchaus kann es dabei vorkommen, dass einige Dimensionen gar nicht betrachtet werden, da sie für das Vorhaben nicht relevant sind oder durch andere Dimensionen mit abgedeckt werden. Auch ist es durchaus denkbar, eine Qualitätsdimension mehreren Gruppen zuzuordnen, je nachdem, ob mehrere Kategorien ähnliche Qualitätsanforderungen besitzen.

Kategorie	DQ Dimensionen
A	Zugänglichkeit, Bearbeitbarkeit, Aktualität
B	Übersichtlichkeit, Einheitliche Darstellung, Verständlichkeit, Eindeutige Auslegbarkeit
C	Wertschöpfung
D	Relevanz
E	Vollständigkeit, Fehlerfreiheit
F	Objektivität, Glaubwürdigkeit, Hohes Ansehen
G	Sicherheit, Vertraulichkeit

**Table 1: Beispielhafte Gruppierung von Qualitätsdimensionen (in Anlehnung an [8] und [5]).**

In der Tabelle 1 haben wir exemplarisch die Dimensionen in 7 Gruppen nach einer einfachen Ähnlichkeit der Begriffe der Dimensionen erstellt. Hier sind die Kategorien noch mit Buchstaben benannt, eine sinnvolle Namensgebung auf Grundlage bestimmter anwendungsszenarischer Anforderungen oder Gebiete ist natürlich möglich und erwünscht.

### 3.3 Skalierung

Nachdem die Dimensionen in die selbst erstellten Kategorien eingeordnet wurden, macht es Sinn, diese nach ihrer Wichtigkeit zu skalieren. Dazu kann man ihnen mehr oder weniger Punkte auf einer Skala zuordnen, um damit ihre Bedeutung in Umfang, Wichtigkeit oder präferierter Behandlung zu kennzeichnen. Eine exemplarische Skalierung haben wir in Abbildung 3) gegeben. Hier haben wir verschiedene globale Anwendungsbereiche definiert, die nach der Aufteilung aus Tabelle 1 skaliert wurden. Alle genannten Anwendungen haben die gleiche Aufteilung der Dimensionen in die Kategorien erhalten. Denkbar ist natürlich auch die individuelle Gruppenbildung bei den einzelnen Anwendungen. Für einen übersichtlichen Vergleich und die allgemeine Erläuterung unserer Idee ziehen wir diese Variante zunächst vor.

Die verschiedenen „System Layouts“ resultieren aus den verschiedenen Anforderungen in dessen Anwendungsgebiete. Gesetzgeber und Behörden (i) legen mehr Wert auf *Vollständigkeit*, *Objektivität* und erhöhte *Sicherheit* (Gruppen E, F & G). Die strengen Anforderungen für die *Zugänglichkeit* und *Aktualität* sowie die *Fehlerfreiheit* (Gruppen A & E) im Bereich der Forschung (ii) oder der Einsatz in der alltäglichen (Berufs-)Praxis (iii), bei der das Wichtigste eine einfache *Zugänglichkeit*, *Vollständigkeit* und *Objektivität* (Gruppen A, E & F) sind, können ebenso abgebildet werden, wie die kla-

re Forderungen nach einer möglichst hohen *Wertschöpfung* (Gruppe C) aus Sicht des Management (iv) unter verringerter Berücksichtigung aller anderen Gruppen.

Die Größe der umschlossenen Fläche, die sich durch das Verbinden der Punkte auf den Kategorieachsen gebildet hat, kann ein erstes Indiz für den mutmaßlichen Implementationsaufwand darstellen, den die Umsetzung des Systems verursachen wird, um allen Anforderungen an die Datenqualität aus der Betrachtung gerecht zu werden. Zudem wird aus der grafischen Betrachtung ersichtlich, auf welche Bereiche sich der Entwickler zuerst konzentrieren sollte, da dies ja einen erhöhten Arbeitsaufwand vermuten lässt.

### 3.4 Entwickler- und Nutzersichten

Nach [9] und [10] ist es zur Beschreibung und Bewertung von Informations- bzw. Datenqualität wichtig, die Qualität der Daten aus der Sicht der Nutzer zu ermitteln (fit for use Konzept). Eine weitere Quelle [4] legt der Qualitätsbewertung ein Qualitätsmodell zugrunde, dessen Subjekte (Nutzer), Prädikate (Anfrageverarbeitung) und Objekte (Datenquelle) nur ganz bestimmte Qualitätskriterien bewerten, und zwar diejenigen, die laut Autor in der Anwendung, Anfrageverarbeitung oder Datenquelle vorliegen. Es wird also überwiegend vom Nutzer ausgegangen, der die Qualitätsansprüche an die Daten vorgibt. Allerdings müsste unseres Erachtens nach auch der Entwickler, der das System erstellt und eventuell auch unterhält, bei der Qualitätsbewertung nach unserem oben beschriebenen Vorgehen maßgeblich unterstützend mitwirken. Das würde in der Praxis folgendermaßen aussehen. Anwender und Nutzer führen die ersten zwei Schritte gemeinsam aus, dabei sollte der Entwickler durch seine Fachkompetenz die Richtung vorgeben.

- Kategorien bilden** Für das anwendungsspezifische Szenario werden Kategorien gebildet, in denen die Qualitätskriterien später eingeordnet werden. Diese Kategorien können bestimmte Qualitätsaspekte der Anwendung namentlich darstellen.
- Qualitätsdimensionen einfügen** Die 15 festdefinierten Dimensionen werden nach den Spezifika der Anwendung in die Kategorien eingeordnet.
- Skalierung der Kategorien** Nutzer und Anwender bewerten die Wichtigkeit der Kategorien nach ihrem persönlichen Verständnis und ihrer Auffassung gegenüber der Wichtigkeit und Präferenz der Dimensionen in den Kategorien.
- Probleme identifizieren** Beide „System Layouts“ können nun gegeneinander gehalten werden. Große Differenzen zwischen Nutzer- und Entwicklerbewertung einzelner Kategorien identifizieren nochmalige Klärung zum Bedarf und Einsatz bestimmter Konzepte zur Qualitätsbehandlung der Daten.

Durch ein derartiges Vorgehen kann bei der Planung bzw. Anforderungsanalyse einer Anwendung in Hinblick auf den Einsatz des Qualitätsmanagements mit seinen Qualitätsdimensionen von vornherein eine anwendungsspezifische Abstimmung zwischen Nutzer und Entwickler erfolgen. Allen Beteiligten kann dabei ein aussagekräftiges Qualitätskonzept in die Hand gegeben werden.

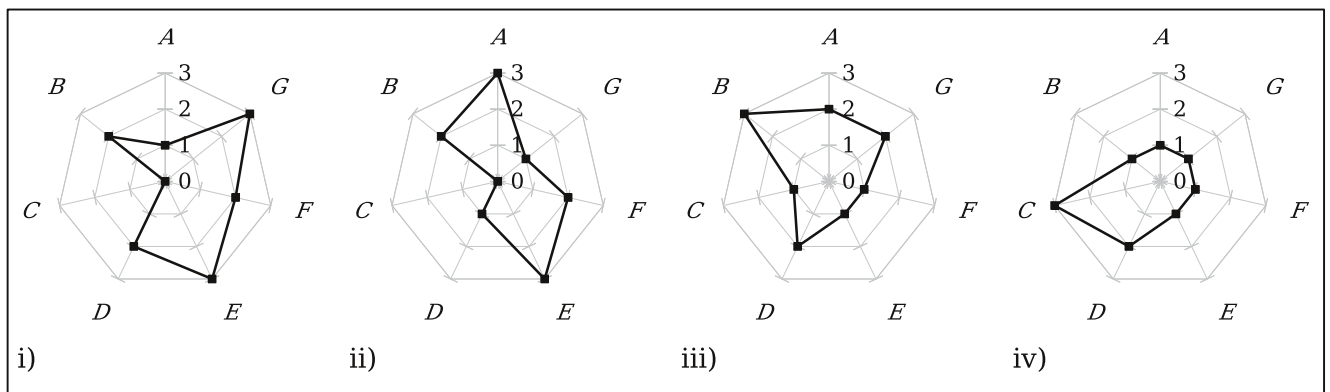


Figure 3: Verschiedene „System Layouts“ bezüglich ihrer Anforderungen der Datenqualität im Bereich der i) Behörden, ii) Forschung, iii) (Berufs-)Praxis and iv) Management; mit Skalierung für 1 - wenig, 2 - viel, 3 - hohe Aufmerksamkeit bzgl. der enthaltenen Qualitätsdimensionen.

#### 4. ZUSAMMENFASSUNG UND SCHLUSSFOLGERUNG

In diesem Beitrag haben wir das Feld der Datenqualität aufgegriffen und ein erstes theoretisches Konzept für den Umgang mit Qualitätsdimensionen vorgestellt. Das generelle Ziel war es, Voraussetzungen für eine hohe Datenqualität zu schaffen, die für ein anwendungsspezifisches Szenario geeignet ist. Denn erst wenn die Funktionalitäten eines Softwaresystems eine, vom Nutzer und Entwickler durch Dimensionen spezifizierte, ausreichende Qualität besitzen, ist die Funktionsfähigkeit eines IT-Systems gewährleistet [9].

Hauptverursacher unzureichender Datenqualität ist und bleibt der Mensch. Deshalb war es wichtig, die von uns als ersichtlichen Unzulänglichkeiten anderer Qualitätsmodelle aufzugreifen und sie zu behandeln. Dazu haben wir die Dimension der Sicherheit als Konzept einer Übererfüllung von Sicherheitsbedürfnissen über den rechtlichen Vorgaben hinaus in unsere Betrachtungen mit einbezogen. Die Einordnung von 15 fest definierten Dimensionen in selbst erstellte Kategorien, aus einem bestimmten Anwendungsszenario heraus, erlaubt eine Wichtung dieser durch den Nutzer und den Entwickler, um auf ein gemeinsames Qualitätslevel zu gelangen. Damit können Methoden und Werkzeuge in der Anforderungsanalyse bestimmt werden, die die Umsetzung der ermittelten Qualitäten gewährleisten können. Für einen späteren Soll-Ist-Vergleich können die Wichtungen ebenfalls als Bewertungsgrundlage herangezogen werden.

Die nächsten Schritte für die Weiterentwicklung dieser Idee sind zunächst z.B. die Überprüfung, ob eine Mehrfacheinordnung von Qualitätsdimensionen in spezielle Kategorien sinnvoll ist. Auch für die Benennung der Kategorien nach den in jedem Anwendungsszenario unterschiedlichen Anforderungen müsste der Einsatz bestimmter Reglements überprüft werden. Um einfache anwendungsspezifische Kategorien zu nutzen, kann der Einsatz von globalen „System Layouts“ (Szenario-Mustern) Abhilfe schaffen. Auch für die Bewertung der Qualität im Nachhinein, also im laufenden Prozess, muss es sich zeigen ob der hier vorgestellte Ansatz dafür

ausreichend Informationen bietet. Und zum Schluss steht ein genereller praktischer Einsatz noch aus.

#### 5. REFERENCES

- [1] M. Abate, K. Diegert, and H. Allen. A hierarchical approach to improving data quality. *Data Quality*, 4(1):365–369, 1998.
- [2] D. Apel, W. Behme, R. Eberlein, and C. Merighi. *Datenqualität erfolgreich steuern - Praxislösungen für Business-Intelligence-Projekte*. Carl Hanser Verlag München, 2009.
- [3] H. Hinrichs. *Datenqualitätsmanagement in Data Warehouse-Systemen*. PhD thesis, Carl von Ossietzky-Universität Oldenburg, 2002.
- [4] U. Leser and F. Naumann. *Informationsintegration*. dpunkt Verlag Heidelberg, 2007.
- [5] F. Naumann and C. Rolker. Assessment methods for information quality criteria. In *Proceedings of the 2000 International Conference on Information Quality (ICIQ)*, pages 148–162, Cambridge, MA, 2000.
- [6] A. Ostman. The specification and evaluation of spatial data quality. *ICA/ACI International Conference*, 18, 1997.
- [7] E. Poy. Objectives of QC systems and QA function in clinical research. *Quality Assurance (San Diego, Calif.)*, 2(4):326–331, Dec. 1993.
- [8] J. Rohweder, G. Kasten, D. Malzahn, A. Piro, and J. Schmid. Information quality - definitions, dimensions and terms. [http://dgiq.de/\\_data/pdf/IQ-Definition/IQ-Definitionen.pdf](http://dgiq.de/_data/pdf/IQ-Definition/IQ-Definitionen.pdf), 2007.
- [9] J. Rohweder, G. Kasten, D. Malzahn, A. Piro, and J. Schmid. *Daten- und Informationsqualität*. Vieweg Teubner, 2008.
- [10] R. Wang and D. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12:5–33, 1996.