

Meta-Kennzahlen für die Bewertung von Dienstleistungen

Edzard Weber und André Nimmich

Bei der Bewertung von Systemen bzw. Dienstleistungen ist es unzureichend, wenn die relevanten Systemeigenschaften nur auf einzelne Kennzahlen verdichtet werden. Es sollte nicht nur das zu bewertende Objekt betrachtet werden, sondern auch das bewertende System und das Bewertungssystem selbst. Diese können über Metriken beschrieben werden. Die so gewonnenen Meta-Kennzahlen helfen, die primären Kennzahlen zu interpretieren und auf ihre Stabilität und Verlässlichkeit hin zu überprüfen. Dieser Beitrag stellt entsprechende Meta-Kennzahlen und Entwurfsentscheidungen vor. Diese fließen in ein entsprechendes Kennzahlen-Cockpit ein, welches die Aussagekraft von Bewertungsergebnissen steigern kann.

1 Einleitung

Ein Bewertungsverfahren ist ein systematischer Vorgang der Zuordnung eines bewertenden und eines zu bewertenden Systems zu einem Werturteil (Bechmann 1991, Bechmann 1998). Einerseits gibt es das zu bewertende Objekt, welches in Form eines Sachmodells repräsentiert wird. Andererseits entsteht ein Wertesystem des bewertenden Systems (das wertende Subjekt). Hierbei gibt es per se keine Einschränkungen, ob ausschließlich Experten oder auch Laien am Bewertungsvorgang teilnehmen dürfen.

Die Menge der möglichen konzeptionellen Ansätze zur Bewertung von Dienstleistungen ist vielfältig. So kommen beispielsweise das Benchmarking (Griehle 2003), Petri-Netze (Winkelmann und Luczak 2006) oder die VOFI (Becker et al. 2008) zum Einsatz. Die Bewertung beschreibt nach bestimmten Kriterien ein System und seine Eigenschaften. Eine Meta-Bewertung, mit der sich dieser Beitrag beschäftigt, beschreibt nach bestimmten Kriterien bzw. mit bestimmten Metriken die Ergebnisse einer Systembewertung und die Eigenschaften der Datenbasis.

Diese Arbeit stellt zunächst die Notwendigkeit und die Anforderungen an Meta-Bewertungsansätze dar. Anschließend werden ein Bewertungsmodell und die daraus ableitbaren Meta-Kennzahlen vorgestellt. Abschließend wird auf die toolbasierte Meta-Bewertung und seine aktuellen Anwendungsbereiche eingegangen.

Der Forschungsschwerpunkt und gleichzeitig auch die Herausforderung liegen in der Verteilung von Prioritäten oder Gewichtungen für einzelne Teilaspekte. Wie wirken sich einzelne Veränderungen von Gewichtungen der Teilaspekte auf die gesamte Bewertung aus? Beispielsweise priorisieren Individuen Prozessketten von Dienstleistungen unterschiedlich. Manche legen Wert auf eine schnelle Durchführung, während andere der nachträglichen Betreuung mehr Beachtung schenken. Für den Forscher sollte dies jedoch in dem Sinne unerheblich sein, als dass er die Gewichtungen aller Bewertungen für seine Zwecke neu ordnen kann, um Effekte auf das Gesamtergebnis beurteilen zu können. Weiterhin kann die Kompensation zwischen einzelnen Bewertungsaspekten untersucht werden. Was passiert, wenn versucht wird, zwei unterschiedliche Bewertungsaspekte anzugleichen? Wie verändern sich dabei andere Aspekte und Gewichtungen?

2 Herausforderungen bei der Bewertung von Dienstleistungen

Es soll zunächst dargestellt werden, inwiefern sich eine Dienstleistungsbewertung von einer Sachleistungsbewertung unterscheidet. Danach werden generelle Herausforderungen an kollektive Bewertungen vorgestellt.

2.1 Dienstleistungsspezifische Rahmenbedingungen

In der Norm DIN EN ISO 9000:2005 ist festgelegt: „Eine Dienstleistung ist das Ergebnis mindestens einer Tätigkeit, die notwendigerweise an der Schnittstelle zwischen dem Lieferanten und dem Kunden ausgeführt wird und üblicherweise immateriell ist.“ Bei der Bewertung von Dienstleistungen gibt es gegenüber der Bewertungen von Sachleistungen andere Voraussetzungen, die sich auf den Erstellungsprozess auswirken (Scheer et al. 2002). Sachleistungen können dem Kunden in seiner konkreten Ausprägung bereits vor einer Vertragseinigung zugänglich, anschaulich oder anfassbar sein. Dienstleistungen hingegen können nicht auf Lager produziert werden. Hier findet eine Einigung zwischen Anbieter und Nachfrager statt, bevor eine Leistung erbracht wird. Dadurch wird der Kunde selbst zum Akteur im Leistungserstellungsprozess. Und während Sachleistungen vorab, kundenspezifisch und immer wieder gleichartig erstellt werden können, wird beim Dienstleistungsprozess durch die individuelle Einbindung des Kunden jede Instanz zu einem Unikat.

Für eine Bewertung von Leistungen bedeutet dies: Die Instanz eines materiellen Produktes bzw. einer Sachleistung kann von mehreren Personen und zu beliebigen Zeitpunkten getestet werden. Um Dienstleistungen zu testen/bewerten, benötigt jeder Proband eine eigene Instanz und ein Test ist auch nur dann möglich, wenn Ersteller und Proband zeitgleich zusammenkommen.

Die Bewertung von Dienstleistungen wird dann interessant, wenn der Anbieter dieser immateriellen Leistungen die Frage nach der Sicherung der Qualität dieser

stellt. Qualitätssicherung von Dienstleistungen wird für zukünftige Aktivitäten interessant und entwickelt beim Kunden das Vertrauensverhältnis. Die Messung der Qualität kann aber nicht wie bei Sachleistungen über mögliche Ausfallzeiten oder Reklamationen von entwickelten Gütern gemessen werden. Ferner wird für die Messung von Qualität der erbrachten Leistungen ein Werkzeug benötigt, welches flexibel genug ist, sich neuen Rahmenbedingungen anzupassen, aber eine hohe Aussagekraft hat. Bei der Bewertung der Dienstleistungsqualität ist es erforderlich, eine erwartete Qualität der tatsächlich wahrgenommenen Qualität gegenüberzustellen (Bieger 2002).

Sachleistungen besitzen eine materielle Hülle, die sich für objektive Messungen besser eignet. Wenn es anerkannte objektive Methoden gibt, reicht auch eine kleine Gruppe von Testern aus. Wenn der Anteil subjektiver Meinungen in der Bewertung hoch ist, sind für repräsentative Bewertungsergebnisse mehr Bewertungsdurchläufe und unterschiedliche Perspektiven notwendig. Die Dienstleistungserstellung kann zudem noch einer zeitlichen Veränderung unterliegen. Während materielle Produkte in höheren Stückzahlen in einem Produktionsdurchgang mit gleichbleibender Qualität erzeugt werden können, gewinnt der Anbieter einer Dienstleistung bei jeder Instanz neue Erfahrungen hinzu und kann eine permanente und mitunter unbewusste Verbesserung betreiben. Folgende Erschwernisse gibt es somit bei der Bewertung von Dienstleistungen:

- Die Erhebung der Bewertungsdaten ist zeitlich verteilt.
- Das zu bewertende System verändert sich über die Zeit.
- Die Datenbasis muss zwecks Repräsentativität größer sein.
- Unterschiedliche Perspektiven mit unterschiedlichen Kompetenzgraden müssen betrachtet werden.

2.2 Allgemeine Herausforderungen bei Bewertungen

Neben speziellen Unterschieden für Bewertungen von Sach- und Dienstleistungen existieren allgemeine Herausforderungen an Bewertungssysteme, die bisher in den Hintergrund getreten sind. Das jeweils verwendete Bewertungssystem wird in der Praxis häufig als gegeben hingenommen. Bedenken zur Stabilität und Interpretierbarkeit der Bewertungsergebnisse werden nur selten angestellt, da hierfür auch ein systematisches Vorgehen fehlt.

Ein einzelner Wert für die Ausprägung einer Eigenschaft eines Systems bzw. einer Dienstleistung oder seine Güte insgesamt ist nicht ausreichend. Hierbei findet zwangsläufig eine zu starke Verdichtung der Charakteristika des zu bewertenden aber auch des bewertenden Systems statt. Ein Beispiel liefert die Benotung von schulischen Leistungen, die sich aus pragmatischen Gründen aber bewährt hat. Zwei Schüler unterschiedlicher Klassen haben jeweils in Mathematik die Note 3. Nun kann sich die eine Note aus einer guten schriftlichen und einer ausreichend mündlichen Leistung zusammensetzen während die andere sich aus zwei befriedi-

genden Teilen zusammensetzt. Ebenso bleibt verborgen, ob beide Schüler gleich viele schriftliche Tests bewältigen mussten, wie fähig der Lehrer in pädagogischer oder fachlicher Hinsicht ist, ob die Anzahl Fehlstunden gleich ist, ob die Klassengröße ähnlich ist, wo der Schüler zum Klassendurchschnitt liegt, ob eine Veränderung zum Vorjahr stattfand usw. All diese Zusatzinformationen lassen die verdichteten Gesamtnoten in einem jeweils anderen Lichte stehen und können erklären, warum sich die beiden Schüler in ihrer mathematischen Kompetenz trotz gleicher Noten stark voneinander unterscheiden. Aus pragmatischen Gründen wird immer gerne auf stark verdichtende Bewertungssysteme, wie z. B. Schulnoten, zurückgegriffen. Wenn es jedoch um weitreichende Entscheidungen auf Basis einer Bewertung geht, kann aber sehr wohl der Mehraufwand gerechtfertigt sein, dass auch die Struktur des Bewertungssystems selbst in die Bewertung mit einfließt.

Weiterhin werden Anwendungen benötigt, welche „die Bewertung und Verbesserung der wirtschaftlichen Anwendbarkeit von Referenzmodellierungstechniken fokussieren sowie die konkrete Anwendung der Referenzmodellierung in Form von Gestaltungsempfehlungen für bestimmte Arbeitsbereiche thematisieren“ (Thomas und Scheer 2003). Die Aufgabe der Bewertung unter diesem Betrachtungswinkel kann durch ein Forschungswerkzeug dieser Art erfüllt werden.

Bewertungsmaßstäbe können sehr subjektiv, indirekt und umfangreich sein. Die Aussagekraft der gewonnenen Ergebnisse verschiebt sich je nach dem, welche Bewertungsaspekte in den Vordergrund treten. Diese Maßstäbe sollten jedoch flexibel für viele verschiedene Rahmenbedingungen sein, da Betrachtungsweisen sehr differieren können.

Neben subjektiven Messansätzen existieren auch objektive Ansätze, die kundenorientierte Anforderungen an die Dienstleistungsqualität messbar machen. Objektive Messansätze, wie Expertenbeobachtungen, Silent-Shopper-Verfahren oder Dienstleistungs- und Warentests, erheben Messdaten aus der Kundensicht, nicht jedoch auf Grundlage subjektiver Einzelwertungen der Kunden. Dies kann durch subjektive Messansätze, die merkmals-, ereignis- oder problemorientiert sein können, erfolgen (Bruhn 2008). Diese unterschiedlichen Ansätze sind in ihrer eigenen Betrachtungsweise hilfreich für die Erforschung und Bewertung spezifischer Anforderungen von Dienstleistungen. Ein flexibles Meta-Analyse-Werkzeug kann den Nutzen dieser Ansätze technisch aufwerten.

Auf Grund subjektiver Unterschiede einzelner Bewertungsindividuen wird die Qualität der Gesamtbeurteilung beeinflusst. Unterschiedliche Wissensstufen über das zu bewertende Objekt behindern eine qualitativ gleichmäßige Verteilung der erhobenen Daten. Für manche Individuen ist ein Aspekt der Bewertung wichtiger, als für andere, die mehr Gewicht auf einen Aspekt legen, der bei anderen Akteuren kaum Beachtung findet. Die Bewertung an sich ist dabei jedoch nicht quantitativ, ferner werden alle Beurteilungen qualitativ betrachtet und auch so behandelt. Es besteht nicht die Absicht, eine Anwendung zu schaffen, welche einfache Umfragen bewältigt. Vielmehr ist es die Aufgabe, ein Werkzeug zu schaffen, welches Forschungsfragen über die Sinnhaftigkeit einzelner Bewertungskriterien und Meta-Kennzahlen beantworten kann.

Daher wird es notwendig, Bewertungen zu gewichten und somit bestimmten Indikatoren bzw. Kriterien mehr Bedeutung beizumessen als anderen. Weiterhin müssen Indikatoren entwickelt werden, die eine Bewertung überhaupt erst möglich machen. Zur Bewertung von Qualitätsmerkmalen können verschiedene Metriken herangezogen werden. Metriken sind Funktionen, die bestimmten Eigenschaften des zu bewertenden Objekts numerische Werte zuweisen (Globke 2005). Meta-Metriken beschreiben die Eigenschaften von Metriken. Das eigentlich zu bewertende System wird durch Meta-Metriken nicht direkt beschrieben. Diese dienen dazu, die von den Metriken gelieferten Werte interpretieren zu können.

Für die im Folgenden dargestellte Art und Weise zur Bewertung von Systemen, wurden folgende Anforderungen angesetzt:

- Die Bewertung muss durch eine Menge von Akteuren durchführbar sein, ohne dass die benutzerindividuellen Angaben verloren gehen. Sie müssen einzeln, als Menge und in aggregierte Form gehalten werden.
- Das Bewertungssystem muss dynamisch erweiterbar sein. Akteure müssen weitere Bewertungskriterien hinzufügen können und Systemelemente in Subelemente aufgliedern können.
- Die aus der Datenbasis ableitbaren Kennzahlen müssen gebildet werden und nachvollziehbar (visuell) den betroffenen Bereichen des Bewertungssystems zugeordnet werden.

Als wesentliche nicht-funktionale Anforderung gilt, dass die Bedienung webbasiert und anwenderfreundlich sein muss. Diese Anforderung bildet die Grundlagen für das konzeptionelle Bewertungsmodell und seine technische Realisierung in Form eines Kennzahlen-Cockpits.

3 Metriken

In diesem Abschnitt wird das konzeptionelle Meta-Modell zur Bewertung kurz vorgestellt. Die Beschreibung dieses Konzepts soll die Verortung der Daten verdeutlichen. Weiterhin wird definiert, welche Daten zur Gewinnung von Kennzahlen herangezogen werden. Das zu Grunde liegende Modell basiert auf einem Bewertungsbaum, welcher die herangezogenen Daten zur Bewertung und zur Gewinnung von Kennzahlen beherbergt.

3.1 Bewertungsbaum

Die Bewertung erfolgt anhand einer Baumstruktur *tree*. Die Verzweigung des Bewertungsbaumes entspricht der logischen Zergliederung des zu bewertenden Objektes und der Bewertungskriterien. Teilaspekte können so schnell extrahiert und einzeln untersucht werden. Der Vergleich von einzelnen Zweigen untereinander kann durch eine dynamische Struktur ermöglicht werden.

$$tree = \{t\} \text{ mit } t = t, t | \{t\} | \{\}$$

Bei Tiefe und Breite sind lediglich Gründe der Pragmatik und Effizienz zu berücksichtigen. Ein Bewertungsbaum T kann einerseits als verschachtelter Klammersausdruck dargestellt werden oder als verzweigter Klammersausdruck.

$$T = \{ \{ \{ \{ \} \dots \} \{ \{ \{ \} \dots \} \{ \{ \{ \} \dots \} \dots \} \}$$

$$\underbrace{\hspace{10em}}_{\{ \}_{T_1}}$$

$$\underbrace{\{ \}_{T_{11}} \quad \{ \}_{T_{12}} \quad \{ \}_{T_{13}}}_{\{ \}_{T_1}}$$

$$\underbrace{\{ \}_{T_{111}} \quad \{ \}_{T_{112}} \quad \dots \quad \{ \}_{T_{121}} \quad \{ \}_{T_{122}} \quad \dots \quad \{ \}_{T_{131}} \quad \{ \}_{T_{132}} \quad \dots}_{\{ \}_{T_{11}} \quad \{ \}_{T_{12}} \quad \dots \quad \{ \}_{T_{13}} \quad \dots}$$

Jeder bewertende Akteur erstellt zum gegebenen Baum eine Bewertung q und eine Gewichtung p . Es sind nur die Blätter des Baumes, die eine explizite Bewertung erhalten. Die Werte der Verzweigungen werden aus diesen errechnet.

$$q_{x_1 \dots x_n}^T = \begin{cases} a \text{ mit } a \in Z^+ & \text{wenn } T_{x_1 \dots x_n} = \{\} \\ null & \text{sonst} \end{cases}$$

Bei der Gewichtung wird (optional) analog zum AHP (Analytic Network Process) ausgewertet (Saaty 2005, Meixner 2008). Es findet ein paarweise Vergleich aller angrenzenden Verzweigungen statt. Durch die Berechnung des Eigenvektors beträgt die Summe aller Gewichtungswerte den Wert 1. Weiterhin ist es möglich, Vorschläge für weitere Zweige einzubringen. Diese sind den anderen Akteuren zunächst unbekannt und erhalten automatisch eine minimale Gewichtung.

$$p_{x_1 \dots x_n}^T = \begin{cases} a \text{ mit } a \in [0 \dots 1] \text{ und } \sum_{i=1}^{\#T_{x_1 \dots x_{n-1}}} p_{x_1 \dots x_{n-1} x_i} & \text{wenn } T_{x_1 \dots x_n} \neq null \\ null & \text{sonst} \end{cases}$$

Die Gesamtbeurteilung V eines Akteurs lässt sich als rekursive Formel darstellen. Der Wert V errechnet sich aus der Summe der jeweils gewichteten Werte der angrenzenden Zweige.

$$V^T(S) = \begin{cases} q_{x_1 \dots x_{\#S}}^T & \text{wenn } T_{x_1 \dots x_{\#S}} = \{\} \\ \sum_{i=1}^{\#T_{x_1 \dots x_{\#S}}} [p_{x_1 \dots x_{\#S} x_i}^T \cdot V^T(S + \{i\})] & \text{sonst} \end{cases}$$

Die Variable S ist ein Vektor, in dem der Pfad zur jeweiligen Verzweigung gespeichert wird. Als Aufrufparameter dient der Pfad zu derjenigen Verzweigung, welche berechnet werden soll, z. B. $V^T(\{1\})$ für den Ausgangszweig (Urknoten).

3.2 Meta-Kennzahlen

Das Konzept der Meta-Kennzahlen wurde für die Erforschung von Veränderungen des (Teil-)Baums unter bestimmten Betrachtungsweisen eingeführt. Da Bewertungen kollektiv für ein Objekt erfolgen sollen, differieren auch die jeweiligen Gewichtungen der Bewertungsaspekte. Die Datenbasis für eine kollektive Bewertung besteht somit aus folgenden Elementen,

- eine Menge von Akteuren $A = \{a_1 \dots a_n\}$, die unterschiedlichen Kompetenzteams bzw. Gruppen zugeordnet werden können
- einem Entscheidungsbaum T
- die Bewertung Q^T vom Akteur a bezüglich des Entscheidungsbaumes T
- die Gewichtung P^T vom Akteur a bezüglich des Entscheidungsbaumes T
- die Zeitpunkte, wann Akteur a die jeweilige Bewertung oder Gewichtung vorgenommen hat.

Obwohl die Verschiedenheit der Datenbasis sehr gering ist, lassen sich eine Vielzahl von Kennzahlen herausbilden, um den eigentlichen Zielwert und seine Verlässlichkeit interpretieren zu können.

Aus dieser Datenbasis lassen sich weitere Daten erzeugen. Es sind die Baumstruktur, Gewichtung, Bewertung und Erstellungsdatum basierend auf den Mengen jener individuellen Daten (Daten werden als Menge zusammengefasst) sowie basierend auf den aggregierten individuellen Daten (Daten werden auf einen Wert verdichtet).

Diese Daten lassen sich durch Angabe von Kosten erweitern. Dabei entspricht die Verteilung der Kosten der Verteilung der Gewichtung. Aspekte höherer Gewichtung erhalten einen höheren Kostenwert als andere.

Meta-Kennzahlen zur Bewertung Q^T

Hier werden die Verhältnisse der konkreten Bewertungen einzelner Elemente bzw. Teilbäume durch die bewertenden Akteure zueinander betrachtet. Eine Gesamtbewertung von 3, die sich aus den Werten 2 und 4 zusammensetzt, besitzt eine höhere Streuung als eine Bewertung aus den Teilwerten 3 und 3. Die Meta-Streuung

betrachtet hingegen diese Streuungswerte. Eine Bewertung mit hoher Streuung der Einzelbewertungen kann also eine geringe Meta-Streuung haben. Derartige Streuungsbetrachtungen sind notwendig, um die Existenz von über- oder unterdurchschnittlich gut bewerteten Teilzweigen in die Interpretation der Gesamtbewertung einfließen zu lassen.

- **Streuung auf Ebene bei Einzelakteur**
Wie im obigen Beispiel dargestellt, können zusammengesetzte Bewertungen aus verschiedenen Teilwerten bestehen. Die Streuung gibt die Verteilung dieser Unterschiede für einen Einzelakteur an.
- **Meta-Streuung bei Einzelakteur**
Die Meta-Streuung gibt die Streuung der Streuung einzelner Bewertungen für einen Einzelakteur an.
- **Streuung auf Ebene bei Akteursgruppen**
Auf der Ebene von Akteursgruppen treten Streuungen vermehrt auf und können mittels dieser Kennzahl abgebildet werden.
- **Meta-Streuung bei Akteursgruppen**
Die Verteilung der Streuung in einer Akteursgruppe wird durch diese Kennzahl ermittelt.
- **Streuung in Verzweigung**
Verzweigungen können nicht direkt bewertet werden. Die Berechnung erfolgt auf der Grundlage der eingegebenen Daten und der Gewichtung. Die Streuung der Berechnungen basiert auf den Streuungen der Ebene von Akteuren.
- **Meta-Streuung in Verzweigungen**
Eine Streuung der Streuung in den berechneten Wertungen für Teilsysteme kann durch diese Kennzahl ausgedrückt werden.
- **Homogenitätstest**
- **Quantilvergleich**

Der Homogenitätstest bzw. U-Test oder auch Mann-Whitney-Test (Mann und Whitney 1947) genannt dient dazu, festzustellen, ob sich zwei Gruppen von Akteuren bei der Bewertung signifikant unterschiedlich verhalten. Bewerten bspw. die Praktiker und die Wissenschaftler ein Prozessmodell bzw. einzelne Aspekte von diesem ähnlich oder ist eine der beiden Gruppen optimistischer eingestellt? Der Quantilvergleich kann dazu herangezogen werden, um der Bewertung einzelner Akteure eine Position innerhalb der Menge aller Akteure zu geben. Dies liefert individuelle Akteursprofile mit den Angaben, wie häufig dieser Akteur beispielsweise im 10%-Quantil der Positiv- oder Negativ-Bewerter liegt.

Meta-Kennzahlen zur Gewichtung P^T

Die Verteilung der Gewichtungen ist je nach Untersuchungsaspekt verschieden. Meta-Kennzahlen beschreiben in diesem Fall die Varianz der Gewichtungen. Diese sind analog zu den Kennzahlen zur Bewertung Q^T .

Meta-Kennzahlen zur gewichteten Bewertung $Q^T P^T$

Die Streuung von Bewertungen und Gewichtungen können ebenfalls über Meta-Kennzahlen gemessen werden. Diese sind analog zu den Kennzahlen zur Bewertung Q^T .

Gewichtete Bewertung orthogonaler Knotenmengen

Meta-Kennzahlen müssen sich nicht zwangsweise an der durch den Bewertungsbaum vorgegebenen Struktur orientieren. Die einzelnen Elemente und ihre jeweiligen Bewertungen können beliebig rekombiniert werden. Dabei können andere Klassifizierungsmerkmale herangezogen werden, die nicht in der vorgegebenen Baumstruktur erfasst werden können:

- Streuung in Elementgruppe bei Einzelakteur
- Streuung in Elementgruppe bei Akteursgruppen
- Homogenitätstest
- Quantilvergleich

Ebenfalls können die ungewichteten Bewertungen und Gewichtungen separat betrachtet werden.

Meta-Kennzahlen zur Sensitivität

Über eine Sensitivitätsanalyse wird untersucht, wie stabil eine Gesamtbewertung auf Basis seiner gewichteten Einzelbewertungen ist. Wie groß dürfen Werteschwankungen sein, damit sich diese auf die Gesamtbewertung auswirken.

- Sensitivität der individuellen Bewertungen
 - Sensitivität der aggregierten Bewertungen
 - Sensitivität der individuellen Gewichtungen
 - Sensitivität der aggregierten Gewichtungen
 - Sensitivität der individuellen gewichteten Bewertungen
 - Sensitivität der aggregierten gewichteten Bewertungen
- Auch hier können zusätzlich Streuung und Meta-Streuung betrachtet werden.

Meta-Kennzahlen zur Konsistenz

Auf Konsistenz werden die paarweisen Vergleiche der Elemente untersucht, aus denen die Gewichtungswerte ermittelt werden. Vollständige Konsistenz liegt dann vor, wenn zwischen einzelnen Vergleichen keine Widersprüche zu erkennen sind. Ein Widerspruch liegt z.B. dann vor, wenn $a > b$ und $b > c$ aber auch $c > a$ bewertet worden sind.

- Konsistenz der individuellen Gewichtung zusammenlaufender Zweige
- Konsistenz der aggregierten Gewichtung zusammenlaufender Zweige
- Streuung der individuellen Konsistenzwerte
- Meta-Streuung der individuellen Konsistenzwerte
- Streuung der aggregierten Konsistenzwerte
- Meta-Streuung der aggregierten Konsistenzwerte
- Streuung in der Menge der individuellen Konsistenzwerte
- Meta-Streuung in der Menge der individuellen Konsistenzwerte

Meta-Kennzahlen zur Datenqualität

Die Datenqualität kann unter verschiedenen Aspekten beurteilt werden.

- Das Datenalter ist bekannt, bezüglich dessen die Annahme getroffen werden könnte, so dass neuere Daten verlässlicher sind als alte.
- Die Detaillierungstiefe des Bewertungsbaumes ist ein weiterer Aspekt. Dabei ist nicht nur interessant, wie groß er insgesamt ist, sondern auch wie ausbalanciert der Baum in der vertikalen ist.
- Analog hierzu betrachtet die Detaillierungsbreite die horizontale Balance.
- Die Anzahl der beteiligten Akteure, deren Beurteilungen aggregiert werden, kann durchaus in den einzelnen Zweigen des Baumes variieren. Auch hier können absolute Zahlen und Balance- also Streuungswerte ermittelt werden.
- Auch die Kompetenz der Akteure kann in den einzelnen Bewertungszeigen unterschiedlich angemessen sein. Urteile von ausgewiesenen Fachexperten sind dann sicherlich anders zu betrachten als die von offensichtlichen Laien.

All diese Aspekte können jeweils durch Minimal-, Maximal- und Durchschnittswerte sowie nach Streuung und Meta-Streuung beschrieben werden.

Meta-Kennzahlen für Aggregationspfade

Wenn es einen Bewertungsbaum gibt, zu dem mehrere individuelle Bewertungen vorliegen, muss beachtet werden, auf welcher Ebene im Baum eine Aggregation der individuellen Bewertungen vorgenommen wird. Es können bereits die Bewer-

tungen der Blätter aggregiert werden. Der Gesamtwert ermittelt sich dann aus den gewichteten und aggregierten Blattwerten. Es können aber auch erst die individuellen Bäume ausgewertet werden. Der Gesamtwert wird dann aus den individuellen Gesamtbewertungen ermittelt. Bei einem n -schichtigen Bewertungsbaum gibt es n verschiedene Ansatzpunkte, an denen die Aggregation der individuellen Bewertungen vorgenommen werden kann (Aggregationspfade).

Unterschiedliche Aggregationspfade können trotz identischer Datenbasis unterschiedliche Gesamtwerte liefern. Dies kann einerseits durch Rundungsfehler geschehen. Andererseits können individuelle Bewertungen nach der Aggregation anders gewichtet werden als bei der Auswertung im individuellen Baum.

Kennzahlen auf Aggregationspfade lassen sich auf Bewertungen und Gewichtungen anwenden.

Meta-Kennzahlen zur Wirtschaftlichkeit

Betrachtet werden können jeweils die Kosten für den Erhalt eines Ist-Zustandes, oder das Erreichen eines geplanten oder ungeplanten Soll-Zustandes. Der Standard-AHP berücksichtigt bereits eine Wirtschaftlichkeitsberechnung. Durch die kollektive Beurteilung und die explizite Berücksichtigung von Teilsystemen werden aber bestimmte Kennzahlen zusätzlich in diesem Kontext anwendbar:

- Individuelle Kostenschätzung
- Kollektive Kostenschätzung
- Streuung der individuellen Kostenschätzungen
- Streuung der aggregierten Kostenschätzungen
- Streuung in der Menge der individuellen Kostenschätzungen
- Meta-Streuung der individuellen Kostenschätzungen
- Meta-Streuung der aggregierten Kostenschätzungen
- Meta-Streuung in der Menge der individuellen Kostenschätzungen
- Individuelle kostengewichtete Elastizität
- Kollektive kostengewichtete Elastizität
- Homogenitätstest
- Quantilvergleich

Meta-Kennzahlen zur zeitlichen Veränderung

Zu bewertende Systeme können sich mit der Zeit verändern. Das Bewertungssystem kann sich verändern. Und auch die Menge der Bewertenden und deren Meinungen können sich mit der Zeit verändern. Wie stark diesbezüglich jeweils die Konstanz ist, kann über weitere Kennzahlen abgebildet werden. Es gibt die Verän-

derungshäufigkeit und -regelmäßigkeit, die Existenz von Tendenzen, der Schwankungsbereich (Min/Max) oder die Streuung. Diese sechs Kennzahlentypen lassen sich im Prinzip auf all jene bisher genannten Kennzahlen anwenden. Und für jeden dieser sechs Kennzahlentypen können wiederum Streuung und Meta-Streuung untersucht werden, so dass am Ende jede Kennzahl um 18 weitere Kennzahlen ergänzt werden kann.

4 Das Kennzahlen-Cockpit

Im vorherigen Abschnitt wurden 101 verschiedene (Meta-)Kennzahlen vorgestellt. Jede Kennzahl ließe sich durch weitere 18 Kennzahlen ergänzen, wenn der zeitliche Verlauf berücksichtigt werden soll. Der Umfang ist enorm und in seiner Gesamtheit nicht mehr pragmatisch, intuitiv handhabbar oder nachvollziehbar: Was ist beispielsweise die Meta-Streuung der Streuung der zeitlichen Streuung der Meta-Streuung der Streuung einer individuellen Gewichtung? Dieser skurrile Ausdruck beschreibt, ob zeitliche Bewertungsschwankungen zwischen Teilzweigen des Bewertungsbaumes auftreten und wie stark sich diese innerhalb des Bewertungssystems unterscheiden. Dies kann in bestimmten Situationen durchaus eine hilfreiche Information darstellen. Eine Kennzahl sollte daher solange nicht als sinnlos abgegolten werden, solange ihre Sinnlosigkeit nicht allgemein nachgewiesen werden konnte. Möglich wird dies aber nie sein. Daher muss der Ansatz gewählt werden, für zukünftige Kennzahlen-Cockpits diejenigen Kennzahlen zu ermitteln, deren Wertveränderungen einen erkennbaren und interpretierbaren Effekt auf die Beurteilung eines Objektes haben. Um diese Effekte und auch die Interdependenzen zwischen den einzelnen Kennzahlen erkennen zu können, ist eine toolbasierte Herangehensweise in realistischen Anwendungsfällen unumgänglich.

4.1 Grundgedanke

Ein entsprechendes Kennzahlen-Cockpit wird momentan realisiert. Dieses webbasierte Tool ermöglicht verteilten Gruppen von Experten bzw. Praktikern, Bewertungen von Objekten vorzunehmen. Dabei wird in einzelnen Sitzungen jeweils nur ein System(-zustand) bewertet. Die Ergebnisse der individuellen Sitzungen werden zusammengefasst. Die Datenerhebung ist durch eine intuitive Benutzeroberfläche leicht durchzuführen, so dass sich Werte und Gewichtungen einfach zuordnen lassen und Metriken abgebildet werden können.

Das Bewertungs-Tool hat somit die Aufgaben, eine kollektive Bewertung zu ermöglichen sowie die Auswirkungen der Eingabedaten auf die neuen Meta-Kennzahlen nachvollziehbar zu machen. Meta-Kennzahlen können evtl. nur kontextspezifisch interpretiert werden. In welchen Fällen eine hohe Meta-Streuung positiv, negativ oder als irrelevant zu interpretieren ist, soll zunächst nur von Domänenexperten eingeschätzt werden. Eine evtl. vorhandene Systematik in Abhängigkeit von der Bewertungsintention soll dann daraus abgeleitet werden.

Das Bewertungssystem basiert auf einer dynamischen Baumstruktur. Eine wichtige Grundlage ist daher die Entwicklung eines dynamischen Datenmodells, welches diese Baumstruktur abbilden kann und neue Zweige in Form von Indikatoren und Systemelementen jederzeit durch die Akteure oder Forscher hinzufügen lässt. Diese Dynamik wird dann interessant, wenn auf der Grundlage der bereits gewonnenen Werte Gewichtungen neu verteilt werden. Ergebnisse lassen sich so unterschiedlich interpretieren und eröffnen ein weitreichendes Spektrum an Berechnungs- und Deutungsmöglichkeiten.

Der Aufbau des Bewertungskatalogs erfolgt dabei sehr dynamisch und lässt sich für bestimmte Betrachtungen entsprechend leicht anpassen. Weiterhin ist für sich wiederholende Bewertungsdurchgänge ein Versionsmanagement integriert, so dass Bewertungen über die Zeit betrachtet werden können. Andererseits können einmal erstellte Bewertungsmuster wiederverwendet und neu rekombiniert werden.

Dieses Tool ist zum einen für Experten und Praktiker gedacht, welche die Aufgabe haben, Systeme wie Sachleistungen oder immaterielle Leistungen zu bewerten. Wissenschaftler besitzen zum anderen die Möglichkeit, neue Indikatoren zu erproben oder anhand von Kennzahlen die Aussagekraft eines bestimmten Indikators und die Kennzahlen selbst erforschen zu können. Auswertungsfunktionen geben ferner Aufschluss über die Qualität des bewerteten Objekts aus verschiedenen Betrachtungswinkeln.

Im Zusammenhang mit gestalterischen Mitteln der Softwarekartographie (Lankes et al. 2005) kann die Anwendung die Bewertung von Dienstleistungen oder entsprechenden Modellen erleichtern. Interpretationen gewichteter Ergebnisse lassen sich somit schneller erschließen und können objektiv veranschaulicht werden.

Erweiterungen des bestehenden Systems sind denkbar. Expertenbewertungen können auf Grundlage von bisherigen Gewohnheiten oder Empfehlungen durch eine sog. Recommendation Engine (Diestelhorst und Matthes 2001) verarbeitet werden. Der Einsatz von Visualisierungskonzepten aus der Softwarekartographie rundet die Benutzbarkeit und Anpassbarkeit ab.

4.2 Sichten-Konzept

Eine Trennung des zu bewertenden Systems in Teilsysteme ermöglicht eine modulare Bewertung und eine einfache Erschließung markanter Kennzahlen. Diese Kennzahlen treten meist als Teilergebnisse dieser Unterstrukturen auf. Die Berechnung der Teilergebnisse erfolgt durch die Aggregation der Blätter eines Teilbaums. Bekannte Bewertungsverfahren besitzen eine statische Struktur, d.h. sie betrachten nur eine bestimmte Sichtweise möglicher Aggregationen. Diese Sichtweise berücksichtigt jeweils nur die Aggregation von Teilbäumen eines Benutzers zu einer bestimmten Kennzahl, welche die Bewertung des Gesamtsystems darstellen soll. Die Berechnung des Endwertes erfolgt dann ausschließlich über die Kombination der einzelnen Bewertungen des Gesamtsystems. Das arithmetische

Mittel über diese Bewertungsergebnisse soll dann eine Aussage über das bewertete System in einem Gesichtspunkt geben.

Nach kurzer Überlegung entstehen Zweifel über die Aussagekraft des berechneten Endwertes. Neben Rundungsfehlern, die beim Aggregieren von Teilbäumen auftreten, werden relevante Informationen über die Herkunft der Daten verschwiegen. Die Kennzahlen des Gesamtsystems geben kaum Aufschluss über die einzelnen Teilsysteme und würden durch eine andere Sichtweise oder Aggregation einen anderen Endwert ergeben. Dabei ist es möglich, für jede Ebene des Bewertungsbaumes eine eigene Sicht zu erstellen und gesondert zu betrachten.

Weiterhin sind Sichtweisen denkbar, die jeweils nur ein Teilsystem unterschiedlicher Akteure betrachten. Somit kann eine Bewertung auf Teilsystemebene erfolgen. Danach können die aggregierten Evaluationen der Teilsysteme wieder zu einem Gesamtsystem zusammengesetzt werden. Dabei wird ersichtlich, dass nur im Einzelfall das gleiche Ergebnis berechnet wird, wie etwa bei der Aggregation der einzelnen Bäume der Akteure. Ein weiterer Vorteil ist die Betrachtung der Teilsysteme an sich. Bewertete Gesamtsysteme geben nur bedingt Aufschluss über die zugehörigen Teilsysteme.

Eine wichtige Rolle bei der Aggregation von Teilbäumen spielen Gewichtungen. Diese Gewichtungen einzelner Kennzahlen stellen dabei die Wichtigkeit des zu betrachtenden Wertes dar. Einerseits stellen sie das Stimmgewicht einzelner Akteure dar. Experten könnten mehr Gewicht in ihrer Bewertung erhalten als Laien. Andererseits geben sie an, welche Teilsysteme mehr Aussagekraft besitzen als andere.

Diese Faktoren lassen die Komplexität von Bewertungen und deren Auswertungen in die Höhe steigen. Um dabei nicht den Überblick zu verlieren, wird es notwendig, entsprechende technische Hilfsmittel zu verwenden. Jedoch basieren aktuelle Bewertungssysteme auf statischen Strukturen und lassen nur eine bestimmte Sichtweise zu. Auf Grund wachsender Anforderungen und dem oben beschriebenen Bedarf nach dynamischen Strukturen wurde das Konzept des Kennzahlen-Cockpits entwickelt. Dieses webbasierte Werkzeug verinnerlicht die oben betrachteten Sichtweisen und ermöglicht die bedarfsgerechte Anpassung an unterschiedliche Gegebenheiten.

4.3 Systemarchitektur

Entwickelt wird das Kennzahlen-Cockpit anhand einer Architektur, die sich in drei Bereiche gliedert (vgl. Abb. 1). Sämtliche anfallende Daten werden in der Datenbasis persistent gehalten. Dabei werden die Daten klassifiziert nach deren Herkunft. Die *VersionBase* enthält alle Datenbestände, die zur Versionierung von Evaluationen und Bäumen notwendig ist. Werte, die bei Evaluationen entstehen, werden in der *ValueBase* gehalten, um einen schnellen Zugriff auf die Werte für Berechnungen zu ermöglichen. In der *WeightBase* sind die Gewichtungen gespeichert, die je nach Bewertungsdurchgang oder Auswertungssitzung verschieden sein kann.

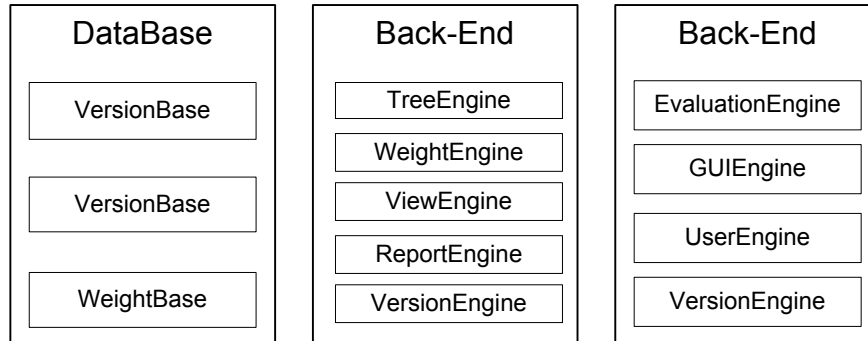


Abb. 1. Architektur des Kennzahlen-Cockpits

Die logische Steuerung übernehmen die sog. Engines. Diese werden nach ihrem Bedarf unterschieden. Für das Back-End beispielsweise, in welchem die Bewertungsbäume entstehen, wird vorrangig die *TreeEngine* verwendet. Die *TreeEngine* erzeugt die dynamischen Bäume und stellt Möglichkeiten bereit, diese an verschiedene Gegebenheiten anzupassen. Zusammen mit der *WeightEngine* und der *ViewEngine* können komplexe Systeme in einem Bewertungsbaum abgebildet werden. Entsprechende Sichten werden mit Hilfe der *ViewEngine* dargestellt und steuern das oben angesprochene Sichten-Konzept. Hauptaufgabe der *ReportEngine* ist die Berechnung und Auswertung abgegebener Bewertungen. Sie bildet das Kernstück des Systems. Alle erzeugten Berechnungen und Ausgaben werden im System hinterlegt und können nach Bedarf wiederhergestellt werden. Diese Herausforderung wird mit Hilfe der *VersionEngine* gemeistert und setzt ein leistungsstarkes Versions- und Changemanagement um.

Im Front-End spielen vorrangig Anzeigesteuern eine Rolle. Die *EvaluationEngine* stellt einzelne Bewertungssitzungen für Akteure bereit und verarbeitet sie. Verwaltet werden sämtliche beteiligten Akteure durch die *UserEngine*. Um entsprechende Evaluationssitzungen durchzuführen, erzeugt die *GUIEngine* web-basierte Benutzerschnittstellen, welche an die bestimmten Anforderungen angepasst sind. Auch im Front-End wird die *VersionEngine* für die Versionierung von Sitzungen verwendet.

5 Anwendung und Ausblick

Die beschriebene Anwendung kann als dezentrales Bewertungstool für Dienstleistungen verwendet werden. Beispielsweise ist es denkbar, die Qualitätssicherung, welche aus verschiedenen Betrachtungsebenen erfolgen kann, über ein solches Werkzeug abzubilden. Der Aufbau der zu sichernden bzw. zu überprüfenden Teilprozesse einer Dienstleistung kann mit Hilfe des dynamisch aufspannenden Baumes direkt abgebildet werden. Akteure, die in den Prozess der Qualitätssicherung involviert sind, bekommen so eine direkte Möglichkeit, die Dienstleistung aus ih-

rer Perspektive zu evaluieren. Der Leistungserbinger erhält jedoch ein Werkzeug, welches nach seinen Bedürfnissen bzw. Forschungsschwerpunkten individuell anpassbar ist.

Grundsätzlich ist die Anwendung jedoch universell einsetzbar. Der aktuelle Anwendungskontext ist die Wandlungsfähigkeit von Organisationsarchitekturen und Schutzkonzepten (Gronau et al. 2009). Zivile Sicherheit ist hier die zu erbringende Leistung. Wandlungsfähigkeit beschreibt die Fähigkeit von Systemen, sich schnell, effizient und selbständig auf Veränderungen in der Umwelt einstellen zu können (Gronau et al. 2006; Eggert und Gronau 2008). Wandlungsfähigkeit ist nicht direkt messbar und wird daher über indikatorbasierte Erfolgsmuster nachgewiesen. Ob erfolgreiche Muster bzw. die Bewertungsmuster selbst aus einer bestimmten Domäne aber auch auf andere Domänen übertragen werden können, muss erfahrungsgemäß von einer Expertengruppe überprüft werden. Nur selten sind Anwendungskontexte so abgesteckt, dass ein Experte allein den Überblick und das vollständige Wissen hat. Zudem gibt es nur in Ausnahmefällen eine objektive ideale Lösung, so dass eine transparente Entscheidungsfindung eine akzeptanzsteigernde Maßnahme darstellt.

Es ist insbesondere der Aspekt der kollektiven Bewertung, der als Komplexitätstreiber im konzeptionellen wie auch technischen Bewertungssystem fungiert. Aber eben dieser Aspekt wird zukünftig immer mehr an Bedeutung gewinnen. Web-basierte Communities haben sich vielfach etabliert, (verteilte) Gruppenarbeit ist eine Selbstverständlichkeit geworden, die ansteigende Komplexität von Entscheidungen wird für einzelne Akteure unüberschaubar, die Anzahl konsensbasierter Entscheidungen, für die es keine optimale Lösung geben kann, steigt an und Formen der elektronischen Beteiligung werden verstärkt erprobt. Hiermit wachsen auch die Aufgaben und die Verantwortung, die gemeinschaftlich getragen werden müssen. Die Aufgaben können auch Dienstleistungen sein. Kollektive Bewertungen fallen dabei zwangsweise an. Um Vertrauen in die Bewertungsdaten zu verschaffen, helfen eben jene Meta-Kennzahlen. Der einzelne Akteur, der unter Umständen seine Mitstreiter und die Gesamtcharakteristik aller Beteiligten gar nicht kennt, kann sich auf diese Weise ein Bild von der Ergebnisverlässlichkeit machen. Besonders für Dienstleistungen die für größere Kundenkreise übers Internet angeboten werden, hat es sich etabliert kollektive Bewertungsverfahren einzusetzen. In der Regel handelt es sich dabei um Verkaufsplattformen. Die zeitliche Verteilung der erhaltenen Bewertungen und die Anzahl unterschiedlicher Kunden werden dabei bereits häufig angegeben (vgl. www.ebay.de).

Doch auch für kleine Gruppen im unternehmerischen Kontext sind die Kennzahlen ein wichtiges Instrument. Sie dienen zur Dokumentation einer kollektiven Expertenentscheidung und können die Stärken oder die Schwächen in der Bewertung selbst offenlegen und quantifizieren. Letztendlich lassen sich auch Verbesserungspotenziale für zukünftige Bewertungsaktivitäten identifizieren.

6 Literaturverzeichnis

- Becker J, Beverungen D, Knackstedt R, Müller O (2008) Konzeption einer Modellierungssprache zur softwarewerkzeugunterstützten Modellierung, Konfiguration und Bewertung hybrider Leistungsbündel. Physica, Berlin
- Bechmann A (1991) Bewertungsverfahren – der handlungsbezogene Kern von Umweltverträglichkeitsprüfungen. In Hübler K-H, Otto-Zimmermann K (Hrsg) Bewertung der Umweltverträglichkeit – Bewertungsmaßstäbe und Bewertungsverfahren für die Umweltverträglichkeitsprüfung. 2. Aufl, Eberhard Blottner Verlag, Taunusstein
- Bechmann A (1998) Anforderungen an Bewertungsverfahren im Umweltmanagement – dargestellt am Beispiel der Bewertung für die UVP. Bericht 20, Institut für Synergetik und Ökologie (SYNÖK), Barsinghausen
- Bieger T (2002) Dienstleistungsmanagement: Einführung in Strategien und Prozesse bei persönlichen Dienstleistungen. 3. Aufl, Haupt, Bern
- Bruhn M (2008) Qualitätsmanagement für Dienstleistungen: Grundlagen, Konzepte, Methoden. Springer, Berlin
- Diestelhorst L, Florian M (2001) Recommendation Engines. Projektarbeit am Arbeitsbereich Softwaresysteme, Technische Universität Hamburg-Harburg
- Eggert S, Norbert G (2008) Erhöhung der Wandlungsfähigkeit von ECM-Lösungen unter Verwendung kartographischer Gestaltungsmittel. Tagungsband der Multikonferenz Wirtschaftsinformatik (MKWI). München
- Globke W (2005) Software-Metriken. Moderne Softwareentwicklung, Universität Karlsruhe
- Griehle O (2003) Prozessorientiertes Vorgehensmodell für das Benchmarking von Dienstleistungen. Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi), Heft 172, Universität des Saarlandes, Saarbrücken
- Gronau N, Lämmer A, Andresen K (2006) Entwicklung wandlungsfähiger Auftragsabwicklungssysteme. In Gronau N, Lämmer A (Hrsg) Wandlungsfähige ERP-Systeme-Entwicklung, Auswahl und Methoden. GITO, Berlin
- Gronau N, Sielaff S, Weber E, Röchert-Voigt T, Stein M (2009) Change capability of protection systems. In Duncan K, Brebbia C A (Hrsg) Disaster Management und Human Health Risk – Reducing Risk, Improving Outcomes. WIT Transactions, No. 110
- Lankes J, Matthes F, Wittenburg A (2005) Softwarekartographie als Beitrag zum Architekturmanagement. In Aier S, Schönherr M (Hrsg) Unternehmensarchitekturen und Systemintegration. GITO, Berlin
- Mann HB, Whitney DR (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics* 18(1):50–60
- Meixner O, Haas R (2008) Wissensmanagement und Entscheidungsunterstützung. Eigenverlag Institut für Marketing und Innovation, Universität für Bodenkultur, Wien
- Saaty TL (2005) Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks. RWS Publications, Pittsburgh
- Scheer A-W, Griehle O, Klein R (2002) Modellbasiertes Dienstleistungsmanagement. Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi), Heft 171, Universität des Saarlandes, Saarbrücken
- Thomas O, Scheer A-W (2003) Referenzmodell-basiertes (Reverse-)Customizing von Dienstleistungsinformationssystemen. Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi), Heft 173, Universität des Saarlandes, Saarbrücken

Winkelmann K, Luczak H (2006) Modelling, simulation and prospective analysis of cooperative provision of industrial services using coloured petri nets. *International Journal of Simulation Systems, Science and Technology* 7(7):10–26