

Evaluation von Artefakten in der Wirtschaftsinformatik

Ulrich Frank

Universität Koblenz-Landau, Institut für Wirtschaftsinformatik
Rheinau 1, D-56075 Koblenz

„There are potentially at least as many ways of dividing up the world into object systems as there are scientists to undertake the task.“

Clyde H. Coombs, Howard Raiffa, Robert M. Thrall

Inhalt

- 1 Evaluation: Ein vielschichtiges Thema
- 2 Artefakte im Gegenstandsbereich der Wirtschaftsinformatik
- 3 Zur Bewertung von Artefakten in der betrieblichen Praxis
- 4 Evaluation als inhärenter Bestandteil wissenschaftlicher Forschung
- 5 Abschließende Bemerkungen

Zusammenfassung

Der Gegenstand der Wirtschaftsinformatik ist wesentlich durch Artefakte geprägt. Das gilt einerseits für den Anwendungsbereich der Disziplin, andererseits ist die Forschung selbst auf die Konstruktion und Beurteilung von Artefakten, wie etwa von Modellen und Prototypen, gerichtet. Der vorliegende Beitrag untersucht die spezifischen Schwierigkeiten, die mit einer wissenschaftlichen Beurteilung informationstechnologischer Artefakte verbunden sind. Ausgehend von einem einführenden Überblick über das Phänomen Evaluation und die Bedeutung von Artefakten in der Wirtschaftsinformatik werden verschiedene Ansätze zur Bewertung von Artefakten näher betrachtet - wie der Entwurf von Bezugsrahmen oder empirische Untersuchungen. Dabei werden die Rahmenbedingungen betriebswirtschaftlich motivierter Evaluationen in der Praxis berücksichtigt. Daneben werden die Bedeutung und die eigentümlichen Schwierigkeiten thematisiert, die eine Bewertung solcher Artefakte mit sich bringt, die Ergebnisse einschlägiger Forschung in der Wirtschaftsinformatik darstellen. Dabei zeigt sich, daß Maßstäbe, wie sie von der Wissenschaftstheorie zur Beurteilung von Forschungsergebnissen vorgeschlagen werden, allenfalls teilweise verwendet werden können.

Abstract

Artefacts are of pivotal importance for the information systems discipline. On the one hand, they constitute a major part of the domain that is subject of information systems research. On the other hand, research is producing artefacts, such as models and prototypes, itself. This paper provides an investigation of the specific problems that are related to a rational evaluation of information technology artefacts. After a general overview of evaluation and the role of artefacts in information systems research, we will discuss selected approaches to evaluating artefacts. This includes both, the economic evaluation of artefacts which are part of business information systems and the evaluation of those artefacts which are produced as results of information systems research. It will be shown that general epistemological criteria to evaluate research results are of limited help.

1 Evaluation - ein vielschichtiges Phänomen

Evaluation wird gemeinhin verstanden als die gezielte Bewertung von materiellen oder immateriellen Gegenständen unter Rückgriff auf Kriterien und Verfahren, deren Angemessenheit erläutert oder begründet werden sollte (vgl. [Hou93], S. 1). In dieser ersten Begriffsabgrenzung deuten sich schon die Schwierigkeiten an, die mit der Durchführung von Evaluationen verbunden sind. Gleichzeitig ist es offensichtlich, daß Evaluationen - auch wenn der Begriff eher selten verwendet wird - in vielfältiger Form nahezu alle Bereiche moderner Industriegesellschaften durchdringen. Sie sind typischerweise auf facettenreiche Sachverhalte gerichtet, also solche, die sich gegen eine einfache Erfassung wesentlicher Merkmale sperren. Sie reduzieren damit, jedenfalls vordergründig, Komplexität. Evaluationen dienen der Vorbereitung und nicht zuletzt der Legitimation von Entscheidungen. Es ist deshalb wenig überraschend, daß sie auch zur Instrumentalisierung von Interessen verwendet werden.

Evaluation ist - jedenfalls dem Anschein nach - durch das Bemühen um *Objektivität* gekennzeichnet. Schon bei oberflächlicher Betrachtung liegt es auf der Hand, daß die Chance, Evaluationen durchzuführen, die gemeinhin als objektiv bezeichnet würden, wesentlich von der Art des zu bewertenden Sachverhalts abhängen: Offensichtlich ist dies für die Kulturpolitik eines Landes schwieriger als für ein Medikament zur Behandlung von Schnupfen. Evaluation setzt Wahrnehmung und Urteilsvermögen voraus. Seit den Anfangszeiten abendländischer Wissenschaften wurden immer wieder die Möglichkeiten und Beschränkungen dieser menschlichen Fähigkeiten thematisiert. Dabei ist einerseits an erkenntnistheoretische Fragen nach den grundlegenden Grenzen unseres Urteilsvermögens zu denken. Beispielfhaft dafür sind die Arbeiten Kants zur "Kritik der reinen Vernunft" sowie zur "Kritik der Urteilskraft". Daneben sind Arbeiten in der kognitiven Psychologie zu nennen, die auf die Erfassung von Verzerrungen und Anomalien der Urteilsfähigkeit gerichtet sind. Hier ist u.a. an umfangreiche Untersuchungen zu denken, die das Entscheidungsverhalten unter Unsicherheit

zum Gegenstand hatten. Dabei hat sich u.a. die verbreitete Disposition gezeigt, in komplexen Entscheidungssituationen zu selektiver Wahrnehmung zu neigen, wobei vor allem die leicht verfügbaren Informationen berücksichtigt werden ("availability", [TvKa82]).

Wenn Gegenstand und Zielsetzung einer Evaluation festgelegt sind, ist damit nicht notwendigerweise gesagt, welche Merkmale bzw. Indikatoren betrachtet und wie ihre Ausprägungen erfaßt werden sollen. Diese Anforderung an die für eine Evaluation ausgewählten Indikatoren ist vergleichbar mit gängigen Anforderungen an die Erhebung von Merkmalen im Rahmen empirischer Forschung - wie Reliabilität und Validität. So läßt sich etwa eine Vielzahl von Merkmalen denken, um die Qualität von Forschung und Lehre an einer Universitätseinrichtung zu beurteilen. Dieser Umstand hat subtile Konsequenzen. Mit der Auswahl von Merkmalen ist der Anspruch verbunden, eine angemessene Bewertung des betrachteten Sachverhalts zu realisieren. Die Merkmale haben danach einen instrumentellen Charakter. Tatsächlich hat aber die Auswahl der Merkmale einen erheblichen, wenn auch latenten, Einfluß auf das Ergebnis der Evaluation. Überzeichnend formuliert: Nicht das Ziel der Evaluation bestimmt die auszuwählenden Merkmale, sondern die ausgewählten Merkmale legen das Ziel der Evaluation fest - und mögen damit letztlich den zu beurteilenden Gegenstand beeinflussen, indem die Gestaltung des Gegenstands an den Indikatoren ausgerichtet wird. Die damit verbundenen, mehr oder weniger deutlichen Verzerrungen des eigentlichen Gegenstands der Evaluation erinnern an eine traditionsreiche Diskussion in der Ökonomie: die Frage nach der Wertneutralität von Mitteln bei gegebenen Zielen. Schon früh hat Myrdal darauf verwiesen, daß eine solche Trennung nicht möglich ist: "Die Wertsetzung bezieht sich jeweils auf einen ganzen Verlauf ...", ([Myr33], S. 310). In diesem Zusammenhang ist auch der "instrumentalistische Trugschluß" ([Alb60], S. 217) zu nennen.

In angelsächsischen Ländern gibt es seit langem eine rege Evaluationsforschung. Sie ist vor allem auf *Evaluationsverfahren* gerichtet, die im weitesten Sinne der Politikberatung dienen. Eine besondere Rolle spielen dabei das Bildungs- und das Gesundheitswesen (ein Überblick findet sich in [Hou93]). Die erhebliche Bedeutung, die derartige Evaluationen in den USA erlangt haben, spiegelt sich u.a. in dem Umstand, daß mit der "American Evaluation Association" eine Art Standesorganisation für diejenigen entstanden ist, die professionell Evaluationen durchführen. Die Ziele, denen sich dieser Verband mit mehreren tausend Mitgliedern ([Hou93], S. vii) verschrieben hat, sind bemerkenswert. So wird u.a. angestrebt "evaluation as a profession" zu etablieren und einen Beitrag zur "generation of theory and knowledge about effective human action" (<http://www.eval.org/>) zu leisten. Ein weiteres Indiz für die Professionalisierung von Evaluationen ist in der Existenz einer dedizierten Fachzeitschrift (American Journal of Evaluation) und einer beachtlichen Zahl einschlägiger Monographien (z.B. [Hou93], [PaTi97], [Scr80]) zu sehen.

Auch im deutschsprachigen Raum spielen aufwendige Evaluationsverfahren eine wichtige Rolle - nicht zuletzt in der Politikberatung. Dabei wird allerdings zumeist der Begriff "Gutachten" verwendet. Der Terminus Evaluationsforschung ist vor allem mit einem Teilgebiet der Psychologie verbunden (allgemein: Gesundheitswesen - ein Überblick findet sich in [Koc90]). Daneben gibt es in den Sozial- und Wirtschaftswissenschaften eine Fülle ähnlich ausgerichteter Forschung. Dazu zählen u.a. Untersuchungen der vielfältigen Folgewirkungen des Technologie-Einsatzes wie auch solche, die auf eine Bewertung wirtschaftspolitischer Maßnahmen zielen. In jüngerer Zeit hat auch im deutschsprachigen Raum die explizite Verwendung des Begriffs Evaluation deutlich zugenommen. Dabei ist vor allem an die Beurteilung der Forschung und Lehre an Hochschulen bzw. an die Beurteilung von Hochschullehrern und Hochschulen zu denken.

Jenseits dedizierter Forschungsprojekte und institutioneller Rahmenbedingungen ist Evaluation gleichsam inhärenter Bestandteil wissenschaftlicher Forschung: Forschung impliziert nach gängigem Verständnis das Bemühen um Erkenntnisfortschritt. Das wiederum erfordert die vergleichende Bewertung alternativer Angebote zur Beschreibung bzw. Erklärung der je interessierenden Phänomene. In der Wirtschaftsinformatik spielt Evaluation neben dieser grundsätzlichen Bedeutung auch im Anwendungsbereich eine wichtige Rolle: Entscheidungen im Informationsmanagement erfolgen in einem Umfeld von Systemen und Verfahrensweisen, deren Beurteilung in technischer wie auch in wirtschaftlicher Hinsicht erhebliche Herausforderungen mit sich bringt. Die Wirtschaftsinformatik erbt dabei gleichsam die spezifischen Evaluationsprobleme der Informatik und der Ökonomie (Aufwand, Nutzen!). Daneben sieht sich die Forschung in der Wirtschaftsinformatik einigen eigentümlichen Evaluationsproblemen gegenüber, die sich einerseits aus ihrem interdisziplinären Standort zwischen den Wirtschaftswissenschaften und der Informatik ergeben, andererseits eine Folge der Besonderheiten des Fortschritts im Bereich der Informationstechnologie darstellen.

Der vorliegende Beitrag ist darauf gerichtet, ein spezifisches Evaluationsproblem der Wirtschaftsinformatik zu analysieren: die Beurteilung informationstechnischer Artefakte - wie etwa Software, Architekturen von Informationssystemen, Datenmodellen etc. Nach einem kurzen Überblick über die eigentümlichen Besonderheiten, die mit informationstechnischen Artefakten einhergehen, betrachten wir zunächst die Anforderungen und Rahmenbedingungen der Bewertung solcher Artefakte in der betrieblichen Praxis. Anschließend wenden wir uns der Evaluation jener Artefakte zu, die Gegenstand und Resultat einschlägiger Forschung in der Wirtschaftsinformatik sind.

2 Die Bedeutung von Artefakten in der Wirtschaftsinformatik

Der Gegenstand der Wirtschaftsinformatik ist wesentlich durch Artefakte geprägt, um nicht zu sagen konstituiert: Informationsbestände in unterschiedlichen Ausprägungen, Anwendungssysteme, Architekturen, Software-Werkzeuge, Standards, Informationsmodelle, wie auch Modellierungssprachen und diverse Analyse- und Entwurfsmethoden. Auch andere Disziplinen, wie etwa die Betriebswirtschaftslehre, die Rechtswissenschaften und vor allem die Ingenieurwissenschaften sind zu einem erheblichen Teil durch die Auseinandersetzung mit Artefakten geprägt. Die Artefakte, denen sich die Wirtschaftsinformatik gegenüberstellt, sind allerdings durch einige Besonderheiten gekennzeichnet. So wird die Erfassung existierender Artefakte durch die häufig erhebliche Komplexität informationstechnologischer Systeme bzw. Konzepte erschwert. Ein Problem, das durch den raschen technologischen Wandel noch verschärft wird. Dabei ist auch zu berücksichtigen, daß marktgängige Technologien häufig eine spezielle Begrifflichkeit prägen, die einerseits den Nachvollzug der wesentlichen Zusammenhänge behindert und andererseits zu der verwirrenden Begriffsvielfalt in der Informationstechnologie beiträgt. Zudem ist daran zu denken, daß häufig verwendete Artefakte, wie etwa Texteditoren, Datenbanken etc., Wahrnehmung und Urteilsvermögen auch der beteiligten Wissenschaftler nachhaltig beeinflussen: Sie konstruieren eine bestimmte Realität, die nach einer gewissen Zeit als selbstverständlich angesehen wird - obgleich grundsätzlich auch andere Ausprägungen denkbar wären.

Aus der Sicht der Unternehmen führt die skizzierte Komplexität zu einem hohen Maß an Unsicherheit und damit zu einem erheblichen Beratungsbedarf. Die seriöse Beschreibung einzelner Produkte mag deshalb durchaus eine sinnvolle Aufgabe von Wissenschaft sein. Gleichzeitig ist allerdings zu berücksichtigen, daß die Auseinandersetzung mit solch singulären, allerdings eben ausgesprochen komplexen Facetten des Gegenstandsbereichs geeignet ist, den Blick für wesentlichere Zusammenhänge zu verstellen. Beispiele für diese Gefahr finden sich zuhauf: objektorientierte Technologien, betriebswirtschaftliche "Standardanwendungen" oder Internet-Technologien.

Die Wirtschaftsinformatik als anwendungsorientierte Disziplin befindet sich hier in einer schwierigen Situation. Einerseits kann sie sich wohl nicht auf ein Abstraktionsniveau zurückziehen, wie es in weiten Teilen der Informatik üblich ist, da sie sich sonst zu weit von ihrer Praxis entfernt. Andererseits opfert sie den Anspruch auf Wissenschaftlichkeit, der ja wesentlich durch das Bemühen um Abstraktion und eine sorgsam entwickelte Terminologie gekennzeichnet ist, wenn sie sich allein darauf beschränkt, faktische Erscheinungsformen von Informationstechnologie zu untersuchen. Das skizzierte Problem erhält dadurch zusätzliches Gewicht, daß seit einiger Zeit die Durchführung von Projekten mit Beratungscharakter durch politische Vorgaben gefördert wird, die Drittmittel-erwerb zu einem Wert eo ipso erheben. In solchen Projekten wird häufig, mit

gutem Grund, die Praxis die Randbedingungen der Kooperation wesentlich prägen - also etwa die zu verwendenden Werkzeuge und damit auch: die Begrifflichkeit, die diesen Werkzeugen zugrunde liegt.

3 Zur Bewertung von Artefakten in der betrieblichen Praxis

Aus betriebswirtschaftlicher Sicht stellen Anwendungssysteme, Werkzeuge, Entwurfsmethoden etc. keinen Selbstzweck dar. Sie dienen vielmehr der Erreichung jener - ggfs. durchaus konfligierender Ziele - auf die die Handlungskomplexe ausgerichtet sind, in die sie eingebettet sind. Die Evaluation von Artefakten spielt einerseits bei Auswahlentscheidungen eine Rolle, andererseits empfiehlt die betriebliche Kosten- und Leistungsrechnung eine regelmäßige Bewertung nach Maßgabe relevanter Kriterien. Insofern kommt der Erfassung der für einen Bewertungszweck als wichtig angesehenen Alternativen sowie deren angemessener Beurteilung eine erhebliche Bedeutung zu. Gleichzeitig ist der mit einem solchen Bewertungsverfahren verbundene Aufwand häufig so hoch, daß er von einem Unternehmen allein nicht vollständig erbracht werden kann. Daraus ergibt sich ein Bedarf an externer Unterstützung. Während am Markt eine Vielzahl von Unternehmen - wie etwa Marktinformationsdienste oder einschlägige Beratungsfirmen - entsprechende Hilfen anbieten, wollen wir im folgenden die Möglichkeiten einer wissenschaftlichen Unterstützung des dargestellten Bewertungsproblems betrachten.

3.1 Bezugsrahmen zur Unterstützung von Evaluationen

Eine systematische Evaluation empfiehlt zunächst die Festlegung der Anforderungen, die das zu bewertende Artefakt erfüllen soll. Anschließend sind Kriterien zu ermitteln, die im Hinblick auf die Erfüllung der Anforderungen bedeutsam sind. Entsprechende Kriterienlisten, zusammen mit mehr oder weniger gehaltvollen Erläuterungen zu ihrer Anwendung, werden i.d.R. Bezugsrahmen genannt. Solche Bezugsrahmen zur Unterstützung der Evaluation informationstechnischer Artefakte existieren in großer Zahl. Sie sind u.a. auf die vergleichende Bewertung von Anwendungssoftware (z.B. [RaHa95]), Werkzeugen der Software-Erstellung (z.B. [Her94], [Sch91]), Benutzungsschnittstellen (z.B. [OpRe92]) oder auf die Evaluation von Modellierungssprachen und -methoden (z.B. [HoGo93], [MoPu92], [MoSh94]) gerichtet.

Die in einem Bezugsrahmen angegebenen Kriterien sind auf verschiedenen Abstraktions- und Detaillierungsstufen denkbar. Während einige Kriterien mit gut nachvollziehbaren Meßvorschriften einhergehen (z.B. Benchmarks), sperren sich andere Kriterien gegen eine objektivierte Erfassung (z.B. "Verständlichkeit der Dokumentation"). Gleichzeitig ist zu berücksichtigen, daß auch solche Kriterien, deren Ausprägung in eindeutiger Weise erfaßt werden kann, im Hinblick auf ihre Beziehung zu den jeweiligen Evaluationsanforderungen er-

hebliche Interpretationsspielräume offenlassen können (z.B. Metriken zur Bewertung von Software-Qualität). Die direkte Unterstützung einer Auswahlentscheidung durch einen Bezugsrahmen wird zudem häufig dadurch erschwert, daß ein Artefakt u.U. verschiedenen Zwecken dienen soll, die unterschiedliche Bewertungskriterien nahelegen. Dabei sind auch die Besonderheiten der Handlungskomplexe, innerhalb derer die Artefakte genutzt werden, zu beachten. Hier ist beispielsweise an die Kompatibilität mit existierenden Artefakten oder an Konflikte bei der Nutzung knapper Ressourcen zu denken. Nicht zuletzt sind es die Fähigkeiten, Präferenzen und Dispositionen, oder allgemeiner: die *Perspektiven* der verschiedenen Benutzer, die für eine differenzierte Bewertung zu berücksichtigen sind.

Auswahlentscheidungen erfordern eine Aggregation der zunächst an mehreren Kriterien orientierten Bewertung. Nicht selten werden solche Aggregationen durch das Berechnen von Mittelwerten durchgeführt. Mitunter werden zusätzlich Gewichte für die verschiedenen Kriterien eingeführt. Oft genug sind die Skalenniveaus der verfügbaren Daten für die verwendeten Operationen aber nicht hinreichend - wenn etwa Mittelwerte ordinalskalierten Daten gebildet werden. Auch ist die jeweilige Gewichtung häufig nicht hinreichend zu begründen. Dennoch dürfte ein gewisses Maß an Verdichtung vielen, die eine Entscheidungsunterstützung wünschen, entgegenkommen - auch um den Preis damit einhergehender Verzerrungen. Dieser Umstand weist auf ein gewisses Dilemma der Evaluation hin: Einerseits ist das Bemühen um Objektivität eine wesentliche Orientierung für Evaluationsvorhaben, andererseits gibt es die Erwartung, daß Evaluationsergebnisse möglichst eindeutig sein sollten. Anders gewendet: Je seriöser, differenzierter und damit: zurückhaltender eine Evaluation ausfällt, desto schlechter mag sie geeignet sein, die von manchem Entscheidungsträger gewünschte Unterstützung zu bieten.

Der Entwurf solcher Bezugsrahmen stellt sicher eine wichtige Aufgabe für die Wirtschaftsinformatik dar: Nicht nur, daß Artefakte einen wesentlichen Teil des Gegenstands der Disziplin konstituieren, darüber hinaus gibt es in der Praxis eine deutliche Nachfrage nach einer entsprechenden Unterstützung. Gleichzeitig kann nicht übersehen werden, daß der Entwurf von Bezugsrahmen mit erheblichen Herausforderungen verbunden ist. Informationstechnische Artefakte durchdringen die Praxis der wissenschaftlichen Wirtschaftsinformatik in subtiler Weise: Wir sind an den Umgang mit gewissen Geräten und Software so sehr gewöhnt, daß es mitunter schwerfällt, von den täglich erfahrenen Ausprägungsformen dieser Artefakte zu abstrahieren. Zudem wird es häufig kaum möglich sein, objektiv nachzuweisen, daß eine vorgeschlagene Kriterienliste vollständig ist. Daneben mag es als wichtig erachtete Kriterien geben, deren nachvollziehbare Erfassung am Evaluationsobjekt kaum zu beschreiben ist. Hier ist etwa an Kriterien wie "Benutzerfreundlichkeit" oder "Anschaulichkeit" zu denken. Während in solchen Fällen in weniger anspruchsvollen Bezugsrahmen ordinalskalierte Schätzungen als Maß dienen, werden in wissenschaftlich ausgerichteten Bezugsrahmen einzelne Kriterien mitunter durch eine Operationalisierung in meßbare Größen überführt. Ein Beispiel dafür liefern die

rung in meßbare Größen überführt. Ein Beispiel dafür liefern die bereits erwähnten Metriken zur "Messung" von Software-Qualität. Dabei ist allerdings nicht zu übersehen, daß u.U. die gemessenen Größen nicht mehr das repräsentieren, was eigentlich erfaßt werden sollte.

Die Erstellung wissenschaftlich akzeptabler Bezugsrahmen sollte eine Reihe von Kriterien berücksichtigen:

- *Zielorientierung*: Die Evaluationsziele sollten explizit gemacht und ggfs. erläutert werden.
- *Begründung*: Die vorgeschlagenen Kriterien sollten grundsätzlich in nachvollziehbarer Weise begründet werden.
- *Seriösität*: Auch wenn das Bemühen um Praxisorientierung nahelegt, den Erwartungen an möglichst eindeutige Evaluationsverfahren nachzugeben, sollte dabei auf keinen Fall ein wissenschaftlicher Anspruch geopfert werden. Er empfiehlt einen zurückhaltenden Umgang mit Operationalisierungen und verbietet arithmetische Operationen, die das jeweilige Skalenniveau nicht zuläßt. Dazu gehört auch der explizite Hinweis auf Grenzen bzw. Schwächen eines Bezugsrahmens.

Die Anwendung von Bezugsrahmen auf konkrete Artefakte, z.B. zum Zweck eines Produktvergleichs, kann durchaus auch noch Gegenstand einer wissenschaftlichen Untersuchung sein. Dabei ist allerdings Zurückhaltung geboten. Die erhebliche Dynamik der einschlägigen Märkte führt dazu, daß auch seriös durchgeführte Evaluationen rasch obsolet werden.

3.2 Empirische Untersuchungen

Da Artefakte in betriebswirtschaftlichen Kontexten i.d.R. keinen Selbstzweck darstellen und ihre wissenschaftliche Bewertung überdies mit erheblichen Problemen behaftet ist, scheint eine indirekte, ex post durchgeführte Evaluation durchaus sinnvoll. Dabei kann von spezifischen Eigenschaften, etwa softwaretechnischen Konzepten, weitgehend abstrahiert werden. Statt dessen wird untersucht, welchen Erfolgsbeitrag Artefakte in bestimmten Nutzungskontexten aufweisen. Entsprechende empirische Untersuchungen basieren auf der Hypothese, daß sich der praxisrelevante Wert eines informationstechnischen Artefakts erst bei einer Nutzung unter realen Bedingungen offenbart.

Vor allem im "Information Systems Research" angelsächsischer Prägung gibt es eine Fülle empirischer Untersuchungen, die auf eine Evaluation in diesem Sinne gerichtet sind - häufig verbunden mit dem Schlagwort "IS effectiveness" ([SeSt98]). Die Bandbreite der einschlägigen Forschung ist dabei beachtlich. Sie zielt u.a. auf den Einfluß bestimmter Software-Arten auf individuelle Arbeitsstile und Kreativität [Mas96], die Auswirkungen der konzeptionellen Mo-

dellierung auf die Problemlösungsfähigkeit verschiedener Interessengruppen [WeSi98] und - immer wieder - die Wirkung des Einsatzes von Artefakten auf die Produktivität [Wei92]. Auch wenn derartige Untersuchungen wichtige Aufschlüsse darüber liefern mögen, wie bestimmte Artefakte genutzt werden, haftet ihnen doch ein kaum gewichtiges Problem an: jenseits statistisch meßbarer Korrelationen festzustellen, ob eine erfaßte Variable tatsächlich kausal vom Einsatz eines Artefakts abhängt [DeMc92].

Aus der Sicht eines Unternehmens, das vor einer Investitionsentscheidung steht, ist es interessant, einen Hinweis auf die beste Option zu erhalten. Da in diesem Zusammenhang die Vermeidung größerer Risiken häufig eine erhebliche Rolle spielt, kommt der Bewährung in der Praxis eine große Bedeutung zu. Es geht also nicht vorrangig um die - wie auch immer zu ermittelnden - besten Konzepte oder Technologien, sondern um den Rückgriff auf solche Lösungen, die sich bereits als besonders erfolgreich erwiesen haben. Seit einiger Zeit wird in der Forschungsförderung häufig auf "best practice" als Bewertungsgrundlage hingewiesen. Auch in der Wirtschaftsinformatik wird dieses Kriterium in Erwägung gezogen ([Buh97], [ReWi97], [Mer97]). Darin drückt sich auch die Betonung einer anwenderorientierten bzw. marktorientierten Sicht aus. In diesem Sinn unterstreicht die EU-Kommission in einer Ausschreibung für die Entwicklung von Entwurfswerkzeugen die Bedeutung von "best practice" [Eur97]: "Best practice in the use of appropriate design tools and methods is a key factor for the efficient, reliable and rapid realisation of fault-free systems which reach the market ahead of the competition and at the right price."

Es mag durchaus interessant sein, zu untersuchen, welche Technologien unter den Randbedingungen der Praxis zu herausragenden Resultaten führen. Dabei ist allerdings - wie grundsätzlich bei empirischen ex post Untersuchungen - zu berücksichtigen, daß auch bei statistisch signifikanten Zusammenhängen eine Kausalität zwischen dem Einsatz eines Artefakts und einem betriebswirtschaftlich interessanten Erfolgskriterium kaum unterstellt werden kann: Zu komplex ist i.d.R. das Gefüge weiterer Einflußfaktoren. Unabhängig von solchen methodischen Schwierigkeiten bleibt eine grundsätzliche Kritik am "best practice" Kriterium: Da es sich um eine Bewertung nach einem längeren Praxiseinsatz handelt, eignet es sich nicht für neue Artefakte, denen angesichts der hohen Dynamik in der Informations- und Kommunikationstechnologie eine besondere Bedeutung zukommt.

4 Evaluation als inhärenter Bestandteil wissenschaftlicher Forschung

Es sind nicht nur Produkte oder am Markt vorherrschende Konzepte, die den Umgang der Wirtschaftsinformatik mit Artefakten prägen. Daneben ist die Forschung in der Wirtschaftsinformatik durch eine Fülle selbst geschaffener Artefakte, wie Modelle und Prototypen gekennzeichnet. Die Konstruktion solcher

Artefakte ist m.E. ein zentraler Bestandteil der Forschung in unserem Fach: Die Erkenntnisse, die dabei gewonnen werden können, fördern ein tieferes Verständnis von Informationssystemen und liefern gleichzeitig einen Beitrag zu einer elaborierten Entwurfslehre. Tatsächlich ist der Entwurf von Artefakten in der Wirtschaftsinformatik (ähnliches gilt für die angewandte Informatik) auch von weniger erfreulichen Erscheinungen begleitet. So sind viele wissenschaftliche Arbeiten und auch Lehrbücher auf die - zumeist nur rudimentäre - Darstellung solcher Entwürfe gerichtet. Eine kritische Auseinandersetzung oder gar ein detaillierter Vergleich einzelner Arbeiten bleibt allerdings i.d.R. aus. Besonders augenfällig wird dieser Umstand auf einschlägigen Konferenzen: Nacheinander werden grobe Zusammenfassungen einzelner Modelle oder Prototypen präsentiert. Auf diese Weise entsteht mitunter der Eindruck – und dies ist die (selbst-)kritische Feststellung eines Betroffenen, daß die Forschung hier einer Spielwiese für die Schaffung von Artefakten gleicht, die vor allem einem dient: der Befriedigung der Konstrukteure.

Vordergründig könnte man die Auffassung vertreten, daß die in der Wirtschaftsinformatik als Ergebnisse einschlägiger Forschung geschaffenen Artefakte in gleicher Weise zu evaluieren seien wie jene, die in der betrieblichen Praxis eingesetzt werden. Ein solcher Ansatz sieht sich allerdings zwei schwerwiegenden Einwänden ausgesetzt. So werden die meisten Forschungsartefakte den Weg in die Anwendung entweder gar nicht oder allenfalls mit erheblicher zeitlicher Verzögerung und zudem in modifizierter Form finden. Daneben - und dies ist m.E. der wichtigere Einwand - sollte sich auch eine anwendungsorientierte Wissenschaft um Kriterien zur Evaluation ihrer Forschungsergebnisse bemühen, die nicht allein die praktische Verwertung umfassen. Die Bewertung der Ergebnisse wissenschaftlicher Forschung ist zentraler Bestandteil der Wissenschaftstheorie. Bevor wir auf die spezifischen Herausforderungen eingehen, die sich durch die Evaluation von Artefakten ergeben, werden wir deshalb zunächst Orientierungen betrachten, die von der Wissenschaftstheorie zur Beurteilung wissenschaftlicher Aussagen angeboten werden.

4.1 Zur Beurteilung wissenschaftlicher Aussagen

Die Anwendung angemessener Kriterien zur Beurteilung von Forschungsergebnissen ist ein wesentliches, wenn auch mitunter implizites Kennzeichen von Wissenschaft. Das gilt in zweifacher Hinsicht. So sind einerseits Abgrenzungskriterien erforderlich: Was sind wesentliche Merkmale wissenschaftlicher Erkenntnis, was unterscheidet sie von den "Träumereien eines Geistersehers" (Kant). Daneben - und durchaus damit zusammenhängend - impliziert das Bemühen um Erkenntnisfortschritt die vergleichende Beurteilung konkurrierender Erkenntnisangebote. Bei der Untersuchung dieser Frage sind einige miteinander verwobene Aspekte zu berücksichtigen [Ham82]. So ist zu klären, welche Sachverhalte überhaupt als Gegenstand wissenschaftlicher Untersuchungen zu akzeptieren sind. Anders gewendet: Welche Sachverhalte sind prinzipiell einer

Beurteilung durch wissenschaftliche Verfahren zugänglich, welche nicht? Wodurch wird subjektives Wissen zu wissenschaftlicher Erkenntnis? Welches ist die angemessene sprachliche Form der Darstellung und Vermittlung wissenschaftlicher Erkenntnisse?

Jenseits der in der Vergangenheit mitunter heftig geführten Grabenkämpfe zwischen den verschiedenen wissenschaftstheoretischen Schulen¹, lassen sich m.E. durchaus deutliche Gemeinsamkeiten im Hinblick auf grundsätzliche Anforderungen an realwissenschaftliche Aussagen rekonstruieren² - die Unterschiede zeichnen sich eher in der Umsetzung ab. Danach sollte eine sprachliche Darstellung von Erkenntnissen in den Realwissenschaften folgende Merkmale aufweisen:

- Es sollten generelle Aussagen, also Aussagen über *Klassen* von realen Sachverhalten, enthalten sein.
- Die beschriebenen Merkmale bzw. Eigenschaften dieser Sachverhalte sollten in ihrem Verhältnis zu Zeit und räumlicher Verteilung möglichst stabil, also *invariant* sein.
- Eine - wie auch immer geartete - *intersubjektive Überprüfung* der Aussagen sollte möglich sein.
- Es sollte neues Wissen und möglichst *informatives Wissen* enthalten sein - wobei die Vorstellungen darüber, was informativ bedeutet, divergieren. Nach der vom Kritischen Rationalismus vertretenen Ansicht nimmt der Informationsgehalt einer Aussage mit der Anzahl der durch sie ausgeschlossenen denkmöglichen Konstellationen zu.
- Die Aussagen sollten *bewährt* sein, d.h. eine Reihe von Überprüfungsversuchen erfolgreich überstanden haben.

Im Hinblick auf die Überprüfung wissenschaftlicher Aussagen gibt es eine Reihe unterschiedlicher Auffassungen. Häufig wird auf den Ansatz des Kritischen Rationalismus verwiesen, wonach eine Überprüfung durch die Konfrontation mit der Realität als objektiver Instanz zu erfolgen hat ("Fallibilismus"). Das setzt geeignete Abbildungs- bzw. Meßvorschriften voraus, um die Korrespondenz der Hypothesen mit der Wirklichkeit wiederum in einem intersubjektiv nachvollziehbaren Verfahren feststellen zu können. Daneben wird in den Wirtschafts- und Sozialwissenschaften mitunter auf diskursive Überprüfungs- bzw. Begründungsverfahren, auf die wir im nächsten Abschnitt kurz eingehen, verwiesen.

¹ Hier ist im deutschsprachigen Raum vor allem an den Kritischen Rationalismus, die Kritische Theorie und den Konstruktivismus zu denken.

² Diese Annahme beruht darauf, daß meines Wissens die im folgenden dargestellten Anforderungen in den wissenschaftstheoretischen Auseinandersetzungen im wesentlichen von keiner Partei in Frage gestellt wurde.

Die skizzierten Anforderungen sind hilfreich für die Abgrenzung wissenschaftlicher Aussagen. Zudem liefern sie eine Orientierung für deren Bewertung, denn die einzelnen Kriterien sind durchaus in unterschiedlichen Intensitäten denkbar. Gleichzeitig ist ihre Anwendbarkeit eingeschränkt. So ist die vergleichende Beurteilung des Informationsgehalts oder der intersubjektiven Überprüfbarkeit von Aussagen in den Wirtschafts- und Sozialwissenschaften ausgesprochen problematisch. Darüber hinaus begrenzt sie den Zuständigkeitsbereich von Wissenschaft auf solche Urteile, die auf eine Beschreibung faktischer Realität zielen. Es wird sich zeigen, daß gerade in der Wirtschaftsinformatik eine solche Einschränkung häufig nicht praktikabel ist.

4.2 Diskursive Urteilsfindung

Im Unterschied zur oben skizzierten Wissenschaftsauffassung des Kritischen Rationalismus gehen die Vertreter hermeneutischer Ansätze davon aus, daß sich soziale Realität allenfalls oberflächlich durch Messungen im Sinne der Naturwissenschaften beschreiben läßt. Statt dessen sei es erforderlich, sich um angemessene Interpretationen der je betrachteten Sachverhalte zu bemühen. Angemessenheit wird dabei in einem diskursiven Prozeß sachkundiger Wissenschaftler beurteilt. Um solche Prozesse dem Ideal *rationaler* Urteilsfindung anzunähern, nennen sowohl die Kritische Theorie ("herrschaftsfreie Kommunikation") wie auch der Konstruktivismus ("vernünftige Beratung") eine Reihe von Bedingungen ([Hab81], ([Lor74]):

- Die Teilnehmer sollten bemüht sein, in gemeinsamer Beratung einen Konsens zu erzielen.
- Es sollten keine Aussagen wider besseren Wissens gemacht werden (sie sollten nicht persuasiv sein).
- Die Teilnehmer sollten bemüht sein, eigene Interessen zu transzendieren und die Sichtweisen anderer Teilnehmer nachzuvollziehen.
- Die Teilnehmer sollten sachkundig sein.
- Die Diskurse sollten herrschaftsfrei sein. Lediglich der "eigentümlich zwanglosen Zwang des besseren Arguments" ([Hab81], S. 52 f.) ist zugelassen.

Im Unterschied zu den vom Kritischen Rationalismus vorgeschlagenen Prüfungsverfahren ist der Gegenstand wissenschaftlicher Urteilsfindung nicht auf falsifizierbare Aussagen über die Wirklichkeit eingeschränkt. Prinzipiell können danach auch Werturteile vernünftig begründet werden. Damit sind solche Ansätze diskursiver Begründung grundsätzlich gut geeignet, eine - wenn auch abstrakte - Methode für Evaluationen zu liefern. Gleichzeitig sind ihre Unzulänglichkeiten kaum zu übersehen. Im wesentlichen münden sie in die Frage, wie entschieden werden soll, daß die Teilnehmer die geforderten

wie entschieden werden soll, daß die Teilnehmer die geforderten Voraussetzungen erfüllen. Es wird sich zeigen, daß diese Frage für unsere weitere Betrachtung von erheblicher Bedeutung ist.

5 Abschließende Bemerkungen

Die geradezu dominierende Rolle, die Artefakten im Gegenstandsbereich der Wirtschaftsinformatik zukommt, macht eine Evaluation konkurrierender Angebote unvermeidlich. Das gilt für Entscheidungssituationen in der Praxis wie auch für die Forschung in einer Disziplin, zu deren wesentlichen Gegenständen die Untersuchung von Gestaltungs- und Einsatzbedingungen solcher Artefakte zählt. Es wäre jedoch Ausdruck eines naiven Positivismus, würde man eine vollständige und dabei objektive Evaluation für möglich halten. Die Möglichkeiten der Wirtschaftsinformatik sind hier ausgesprochen bescheiden. Dennoch - oder besser: gerade deshalb - sollte die Evaluation von Artefakten mit großem Nachdruck und der nötigen kritischen Distanz betrieben werden. Nur so kann einer Forschungspraxis entgegengewirkt werden, in der der Entwurf von Artefakten zum Selbstzweck wird ohne daß nach Erkenntnisfortschritt gefragt wird. Während der Rekurs auf Bewährung in der Praxis ("best practice") oder die Befragung von Orakeln (Ergebnisse sog. Expertenbefragungen) wichtige Randbedingungen für den erfolgreichen Einsatz von Artefakten in Unternehmen aufzeigen mag, sind solche Ansätze für die Bewertung von Forschungsergebnissen nicht zu akzeptieren, da sie zur Preisgabe eines wissenschaftlichen Anspruchs führen. Die in der hermeneutischen Wissenschaftstheorie skizzierten Verfahren diskursiver Urteilsfindung liefern eine Orientierung auch für die Evaluation von Artefakten, die im Rahmen der Forschung entstanden sind. Ihre Anwendung ist allerdings mit einem erheblichen Aufwand verbunden und zudem nicht frei von Schwächen.

Literatur

- [Alb60] Albert, H.: Wissenschaft und Politik. Zum Problem der Anwendbarkeit einer wertfreien Sozialwissenschaft. In: Topitsch, E.v. (Hg.): Probleme der Wissenschaftstheorie. Wien 1960, S. 201-232
- [Buh97] Buhl, H.U.: Best practices vs. common practices bei der Softwareentwicklung. In: Wirtschaftsinformatik, Heft 6, 1997, S. 639 f.
- [DeMc92] DeLone, W. H.; McLean, E. R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research, Vol. 3, No. 1, 1992, S. 60-95
- [Eur97] European Commission: ESPRIT ESD Best Practice. Technologies for Components and Subsystems (TCS). "Broadening the Use of Design Technologies". Brüssel, Sept. 1997 (<http://www.cordis.lu/esprit/src/esd-best.htm>)

- [Hab81] Habermas, J.: Theorie des kommunikativen Handelns. Bd. 1: Handlungsrationalität und gesellschaftliche Rationalisierung. Frankfurt/M. 1981
- [Ham82] Hamminga, B.: Neoclassical Theory Structure and Theory Development: The Ohlin Samuelson Programme in the Theory of International Trade, Stegmüller, W.; Balzer, W.; Spohn, W. (Hg.), Philosophy of Economics, Berlin, Heidelberg et al. 1982, S. 1-15
- [Her94] Herzwurm, G. (Hg.): CASE-Technologie in Deutschland: Orientierungshilfe und Marktüberblick für Anbieter und Anwender. Universität Köln 1994
- [HoGo93] Hong, S.; Goor, G.: A Formal Approach to the Comparison of Object-Oriented Analysis and Design Methodologies. In: Nunamaker, J.F.; Sprague, R.H. (Hg.): Information Systems: Collaboration Technology, Organizational Systems, and Technology. Proceedings of the 26th International Hawaii International Conference on System Sciences. Los Alamitos 1993, S. 689-698
- [Hou93] House, E.R.: Professional Evaluation: Social Impact and Political Consequences. Newbury Park, Ca. 1993
- [Koc90] Koch, U. (Hg.): Evaluationsforschung. Bewertungsgrundlage von Sozial- und Gesundheitsprogrammen. Berlin u.a. 1990
- [Lor74] Lorenzen, P.: Konstruktive Wissenschaftstheorie. Frankfurt/M. 1974
- [Mai96] Maier, R.: Qualität von Datenmodellen. Wiesbaden 1996
- [Mas96] Massetti, B.: An Empirical Examination of the Value of Creativity Support Systems on Idea Generation. In: MIS Quarterly, Vol. 20, No. 1, 1996, S. 83-98
- [Mer97] Mertens, P.: Best practice. In: Wirtschaftsinformatik, Heft 6, 1997, S. 640
- [MoPu92] Monarchi, D.E.; Puhr, G.: A Research Typology for Object-Oriented Analysis and Design. In: Communications of the ACM, Vol. 35, No. 9, 1992, S. 35-47
- [MoSh94] Moody, D.L.; Shanks, S.: What Makes a Good Data Model? Evaluating the Quality of Entity Relationship Models. In: Loucopoulos, P. (Hg.): Entity-Relationship Approach - ER'94. Business Modelling and Re-Engineering. 13th International Conference on the Entity-Relationship Approach. Berlin, Heidelberg etc. 1994, S. 94-111
- [Myr33] Myrdal, G.: Das Zweck-Mittel-Denken in der Nationalökonomie. In: Zeitschrift für Nationalökonomie. 1933, S. 305-329
- [OpRe92] Oppermann, R.; Reiterer, H.: Evaluation von Benutzerschnittstellen. In: Wirtschaftsinformatik Nr. 3, 1992, S. 283-293
- [PaTi97] Pawson, R.; Tilley, N.: Realistic Evaluation. London 1997
- [RaHa95] Rae, A.K.; Hausen, H.-L.; Robert, P.: Software Evaluation for Certification : Principles, Practice and Legal Liability. London 1995
- [ReWi97] Reitwieser, N.; Will, A.: Best practice oder common practice: Eine Frage der Wirtschaftlichkeit. In: Wirtschaftsinformatik, Heft 6, 1997, S. 640 f.

- [Sch91] Schmidt, H.W.: CASE Products 1990 : A Survey of CASE Products from US Vendors. Arbeitspapiere der GMD, Nr. 518, Sankt Augustin 1991
- [Scr80] Scriven, M.S.: The Logic of Evaluation. Inverness, Ca. 1980
- [SeSt98] Seddon, P.B.; Staples, D.S.; Patnayakuni, R.; Bowtell, M.J.: The IS Effectiveness Matrix: The Importance of Stakeholder and System in Measuring IS Success. In: Hirschheim, R.; Newman, M.; DeGross, J.I. (Hg.): Proceedings of the Nineteenth International Conference on Information Systems. Helsinki 1998, S. 165-176
- [Tim95] Timmreck, T.C.: Planning, Program Development, and Evaluation. A Handbook for Health Promotion, Aging and Health Services. Boston u.a. 1995
- [TvKa82] Tversky, A.; Kahnemann, D.: Judgement under Uncertainty: Heuristics and Biases. In: Kahnemann, D.; Slovic, P.; Tversky, A. (Hg.): Judgement under Uncertainty: Heuristics and Biases. Cambridge 1982
- [Wei92] Weill, P.: The Relationship between Investment in Information Technology and Firm Performance: A Study of the Valve Manufacturing Sector. Information Systems Research, Vol. 3, No. 4, 1992, S. 307-333
- [WeSi98] Weber, R.; Sim, M.; Patel, A.: Stakeholder Experiences with Conceptual Modeling: An Empirical Investigation. In: Hirschheim, R.; Newman, M.; DeGross, J.I. (Hg.): Proceedings of the Nineteenth International Conference on Information Systems. Helsinki 1998, S. 370-375
- [WeWi97] Wernick, P.; Winder, R.L.: Software Engineering as a Kuhnian Discipline. In: Winder, R.L.; Probert, S.K.; Beeson, I.A. (Eds.): Philosophical Aspects of Information Systems. London 1997, S. 117-129