

# Das *IfADo*-PPM: Ein Ansatz zur Messung und Bewertung der Leistungen von Forschungseinrichtungen

von

Klaus-Helmut Schmidt und Herbert Heuer

## Problemstellung

Der Ruf nach einer Evaluation von Forschungsleistungen scheint in dem Maße lauter zu werden, in dem knappe Finanzmittel eine möglichst wirkungsvolle Verteilung erforderlich machen. Inzwischen gibt es eine Reihe von Vorschlägen für verschiedene Verfahrensweisen, und vor allem in den Nachbarländern werden solche Verfahren auch schon praktiziert. Im folgenden beschreiben wir ein Verfahren, das seit zwei Jahren am *IfADo* verwendet wird, sowie die ihm zugrunde liegenden Überlegungen. Es ist nicht nur prinzipiell auch für andere Einrichtungen geeignet, sondern es weist auch einige Besonderheiten auf, insbesondere eine explizite Diskussion und Festlegung von Bewertungskriterien. Es handelt sich um eine Variante eines Verfahrens, das sich bei sehr vielen Anwendungen in Industrie, Verwaltung und Handwerk bewährt hat.

Bewertung von Forschungsleistungen ist nicht neu, sondern schon seit langem fast allgegenwärtig. Sie erfolgt in der Regel in Form von Gutachten etwa über ein Manuskript, eine Dissertation, eine Habilitation oder einen Antrag auf Forschungsmittel etc. Solche Begutachtungen gehören quasi zum kulturell gewachsenen Instrumentarium der Selbstorganisation und Qualitätssicherung im Bereich von Wissenschaft allgemein. Die Kriterien, die hierbei angelegt werden, sind allerdings fachspezifischer Art und reflektieren die jeweils als gültig anerkannten Bezugssysteme innerhalb der einzelnen Wis-

senschaftsdisziplinen. D.h., die Forschungsleistung eines Ingenieurs wird nach anderen Kriterien bemessen als die eines Biologen oder eines Philosophen. Diese fachspezifischen Bezugssysteme entwickeln sich im Prozeß der Ausdifferenzierung von Aufgaben, die die einzelnen Fachdisziplinen arbeitsteilig übernommen haben. In Abhängigkeit von ihrem jeweiligen Entwicklungsstand können sich selbst innerhalb einer Fachdisziplin wieder unterschiedliche und neue Aufgaben oder Spezialgebiete herausbilden, die dann wiederum eigene Bezugssysteme zur Beurteilung der hier erbrachten Leistungen entstehen lassen.

Das, was die neue Entwicklung bei der Evaluation von Forschungsleistungen auszeichnet, ist nicht der Sachverhalt der Evaluation an sich, sondern der Gedanke, man könne die Leistungen von Forschungseinrichtungen mit unterschiedlichen Aufgabenprofilen und Bezugssystemen vergleichend auf den Prüfstand stellen. Die gemeinsame Evaluation der verschiedenen Fakultäten und Institute einer Universität oder der sehr heterogenen Institute der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (früher Wissenschaftsgemeinschaft Blaue Liste) ist ein Beispiel hierfür. Bei einer derartigen fachübergreifenden Evaluation werden die innerhalb der einzelnen Fachdisziplinen geltenden Bezugssysteme nicht selten außer Kraft gesetzt. An ihre Stelle treten dann mehr oder weniger explizierte Kriterien der Bewertung, die sich oft

an vermeintlich objektiven und leicht verfügbaren Kenngrößen orientieren, wie z. B. an der Anzahl von Publikationen in Zeitschriften mit hohem Impact-Faktor oder der Summe der eingeworbenen Drittmittel. Es ist offensichtlich, daß derartige Kenngrößen den jeweiligen Besonderheiten der Aufgaben und Leistungen der einzelnen Fachdisziplinen und Forschungseinrichtungen nicht immer gerecht werden. So ist die Summe der eingeworbenen Drittmittel eng mit dem Ausmaß verknüpft, in dem die Bearbeitung bestimmter Forschungsfragen den Einsatz teurer Laborgeräte erforderlich macht. Im Lichte dieses Kriteriums schneiden z. B. die Naturwissenschaften, bedingt allein schon durch die Art ihrer Forschungsfragen, immer besser ab als die Sozial- oder Geisteswissenschaften. Der Impact-Faktor von wissenschaftlichen Zeitschriften verleitet sogar zu einer unterschiedlichen Bewertung von Leistungen selbst innerhalb einzelner Fachdisziplinen. Gemessen an diesem Kriterium wären dann z. B. in der Medizin biochemische oder molekularbiologische Arbeiten weitaus höher einzustufen als Beiträge aus dem Bereich der Anatomie oder Physiologie.

Eine Gefahr der einfachen und im Prinzip mißbräuchlichen Verwendung einzelner weniger Kenngrößen besteht nicht zuletzt darin, daß die Bewertung auf den Wissenschaftsbetrieb zurückwirkt - ob gewollt oder nicht, und zwar in wenigstens zweifacher Hinsicht. Das Fälschen von Daten, um die Publikationschancen in angesehenen Zeitschriften zu erhöhen, ist wohl die krasseste Form einer solchen Rückwirkung; die für die Bewertung herangezogenen Kriterien verlieren durch diese und weniger drastische Maßnahmen zur gezielten Beeinflussung der Kenngrößen letzten Endes ihre Validität.

Weitaus subtiler und weniger spektakulär fallen die Rückwirkungen aus, wenn Forschungseinrichtungen ihre ursprünglichen Aufgabengebiete verlassen bzw. derart verlegen, daß sie gemessen an den Kriterien "gut dastehen". Dies hätte langfristig nicht nur eine Verarmung der Forschungsland-

schaft innerhalb einzelner Fachdisziplinen zur Folge, mit dem Fokus auf dann nur noch wenige, aber hochbewertete Forschungsfelder. Fachübergreifende Leistungsvergleiche mittels solcher Kenngrößen dürften einige Fachgebiete oder Forschungseinrichtungen auch gänzlich ins wissenschaftliche Abseits stellen und in ihrem Bestand bedrohen. Gleichwohl können diese Fachgebiete und Einrichtungen mit der wissenschaftlichen Bearbeitung von Aufgaben befaßt sein, die von einem anerkannt großen gesellschaftlichen Nutzen sind.

Angesichts dieser Probleme fachübergreifender Leistungsvergleiche mittels einheitlicher Bewertungskriterien dürfte es geboten sein, alternative Verfahrenswege zu beschreiten und zu erproben. Ein wichtiger Schritt ist, den Gedanken aufzugeben, daß es fachübergreifende Kriterien der wissenschaftlichen Leistungsfähigkeit gebe, und einen Schritt zurückzugehen: Leistung ist nicht "an sich" definiert, sondern stets im Hinblick auf eine Aufgabe oder auf ein Ziel. Es ist daher sinnvoll, zunächst einmal von den Aufgaben und Zielen auszugehen, die einzelne Forschungseinrichtungen bearbeiten und verfolgen. Denn diese Aufgaben und Ziele bilden in der Regel den Legitimationskern, über den sie sich in ihrer Existenz begründen und der ihnen ihr spezifisches Profil verleiht. Erst in einem zweiten Schritt wäre dann zu prüfen, an Hand welcher Kriterien beurteilt werden kann, in welchem Ausmaß eine Forschungseinrichtung ihre jeweiligen Aufgaben tatsächlich auch erfüllt und ihre Ziele erreicht. Ein solches Vorgehen macht die spezifischen Aufgaben- und Zielprofile der einzelnen Einrichtungen transparent und erlaubt zugleich eine quantitative Beurteilung ihrer aufgabenspezifischen Leistungen. Leistungsvergleiche zwischen verschiedenen Einrichtungen mit unterschiedlichen Aufgaben- und Zielprofilen könnten dann ohne die unbegründete Verwendung einheitlicher und aufgabenspezifischer Bewertungskriterien auskommen, in dem sie sich an eben diesem Kriterium des Ausmaßes der jeweiligen Aufgabenerfüllung und Zielerreichung orientieren.

Dieses Kriterium ist einerseits auf verschiedene Forschungseinrichtungen anwendbar, berücksichtigt andererseits aber auch deren spezifische Aufgabenstellungen und Ziele.

Ganz ähnliche Vorstellungen zur Beurteilungen von Arbeitsleistungen allgemein liegen einem Ansatz zugrunde, der im deutschen Sprachraum als "Partizipatives Produktivitätsmanagement" (PPM) bekannt geworden ist (Pritchard et al. 1993; Werthebach & Schmidt 1995). Dieser Ansatz, der im wesentlichen Prinzipien für das Vorgehen bei der Entwicklung von Leistungsmeß- und Beurteilungssystemen beinhaltet, wurde bisher hauptsächlich im Produktions- und Dienstleistungsbereich eingesetzt und erprobt. Die hier gewonnenen positiven Erfahrungen und Ergebnisse (siehe Schmidt & Kleinbeck 1997; van Tuijl et al. 1997) gaben Anlaß, PPM auch bei der Erstellung eines Leistungsbeurteilungssystems für das Institut für Arbeitsphysiologie zu nutzen. Von der Anwendung dieses Systems erwarten wir, daß der Institutskörper als Ganzes sich an den hier festgelegten Beurteilungskriterien orientiert und die internen Aktivitäten auf das Erfüllen dieser Kriterien ausrichtet - ganz im Sinne der oben angedeuteten Rückwirkung.

### **Merkmale von PPM und Entwicklungsschritte**

Der PPM-Ansatz folgt zwei einfachen, aber bedeutsamen Grundsätzen. Der erste Grundsatz lautet: "Nur das, was man mißt, kann man auch verbessern". D.h., Messungen machen es erst möglich, Leistungsdefizite und -verbesserungen als solche zu erkennen. Der zweite Grundsatz besagt: "Nur das, was man mißt, findet auch Beachtung". Aus diesen beiden Grundsätzen leiten sich eine Reihe von Anforderungen ab, die ein wirksames Leistungsmeß- und Beurteilungssystem (auch für Forschungseinrichtungen) erfüllen sollte.

Die erste Anforderung betrifft die *Vollständigkeit der Messung*. Das heißt, die zur Leistungsmessung herangezogenen Größen müssen möglichst umfassend sein und alle Leistungsfacetten vollständig abbilden. Aktivitäten und Leistungsaspekte, die nicht erfaßt werden, verlieren erfahrungsgemäß über kurz oder lang an Bedeutung und werden schließlich gänzlich vernachlässigt. Dies kann u.U. äußerst konsequenzenreich sein. Sollen die verwendeten Leistungsmaße zudem Maßnahmen zu ihrer Verbesserung anregen, dann müssen sie auch vom Verhalten der Personen beeinflussbar sein. Erst die *Beeinflussbarkeit der Leistungsmaße* ermöglicht es, Zusammenhänge zwischen eigenem Verhalten und den hieraus resultierenden Leistungseffekten zu erkennen und bei der Auswahl von Verhaltensalternativen zu nutzen. Nicht beeinflussbare Leistungsmaße lassen dagegen diese Zusammenhänge im Unklaren und führen gewöhnlich zum Erleben von Hilflosigkeit und Frustration (Schmidt & Hollmann 1998).

Wirksame Leistungsmeß- und Beurteilungssysteme sollten ferner in der Lage sein, die verschiedenen *Leistungsmaße in ihrer Bedeutsamkeit zu gewichten*. Denn nicht alle Leistungsaspekte tragen in gleicher Weise zur Gesamtleistung einer Arbeitseinheit bei; vielmehr unterscheiden sie sich im Gewicht und in ihrer Bedeutsamkeit. Darüber hinaus sollten wirksame Meß- und Beurteilungssysteme neben den rein deskriptiven Leistungsmaßen auch Informationen darüber vermitteln, ob eine Leistung in einem bestimmten Aspekt auf ein eher gutes, schlechtes oder normales Arbeitsergebnis verweist. Die rein deskriptiven Leistungsmaße sind also um Informationen zur *Bewertung der einzelnen Leistungsergebnisse* zu ergänzen.

Bei dieser expliziten Bewertung ist für jeden einzelnen Leistungsaspekt zu klären, welchen Beitrag eine Leistungssteigerung in diesem Aspekt zur Gesamtleistung der Arbeitseinheit leistet. Liegen mehrere Aspekte vor, und dies dürfte der Normalfall sein, muß zudem beurteilt werden, ob ein Anstieg um eine Maßeinheit beim Leistungsaspekt A

mehr oder weniger zur Gesamtleistung beiträgt als ein Anstieg um eine Maßeinheit im Leistungsaspekt B. Hierdurch werden die unterschiedlichen Arten von Leistungsmaßen in eine gemeinsame Metrik überführt, die der Bewertung. Wenn man so will, werden hier Äpfel und Birnen dadurch vergleichbar gemacht, daß man zwei Zahlen definiert, die jeweils für Äpfel und Birnen den "Wert für die Sättigung" kennzeichnen. Auf diese Weise werden für alle Komponenten der Leistungsprofile Bewertungsfunktionen erstellt, die Aussagen darüber machen, welche Bewertungen aus dem Erreichen bestimmter Leistungsergebnisse resultieren.

Diese Bewertungsfunktionen sind im einfachsten Fall linear, d.h., mit einem Anstieg im Leistungsmaß X geht ein linearer Anstieg in der Bewertung einher. Oft dürfte dies jedoch nicht ausreichen, um die tatsächlichen Bewertungen unterschiedlicher Leistungsaspekte abbilden zu können. Denn in vielen Fällen scheint es so etwas wie einen abnehmenden (oder auch zunehmenden) Beitragsnutzen zu geben; man denke nur an die Unterschiede in der Bewertung eines 1. oder eines 101. Apfels, die bei vielen Menschen vorhanden sein dürften. Ähnliche Verhältnisse dürften auch für eine Reihe von Produkten einer Forschungseinrichtung gelten, wie z. B. die Anzahl gehaltener Kongreßreferate oder auch die Anzahl von Gutachtertätigkeiten der verschiedensten Art. Eine gewisse Anzahl derartiger Produkte ist wichtig für die Einbettung in den Wissenschaftsbetrieb, aber irgendwann nimmt der Nutzen, der aus weiteren Referaten und Gutachten resultiert, ab. Um diese Sachverhalte angemessen abzubilden, können Bewertungsfunktionen auch nichtlinear sein.

Meß- und Bewertungssysteme, die diesen Anforderungen entsprechen, lassen sich nicht nur auf einzelne Forschungseinrichtungen isoliert anwenden; sie können auch für die vergleichende Bewertung der Gesamtleistungen unterschiedlicher Forschungseinrichtungen herangezogen werden.

Grundlage für eine solche vergleichende Bewertung sind natürlich nicht die berücksichtigten Maßzahlen selbst, sondern ihre Bewertung. Der Vergleich erfolgt dann danach, wie gut die einzelnen Einrichtungen ihre jeweils spezifischen Aufgaben erfüllen, die ja - wie gesagt - sehr unterschiedlich sein können. Die Bewertungen geben genau hierüber Auskunft. Ohne diesen zusätzlichen Schritt der Bewertung des Ausmaßes der Aufgabenerfüllung, der die Beurteilungsmaßstäbe innerhalb der einzelnen Fachkulturen explizit macht und Vergleiche unter Einbeziehung der unterschiedlichen Maßstäbe und Aufgabenstellungen erlaubt, kann keine fachübergreifende Evaluation wissenschaftlicher Diversität gerecht werden. Denn es dürfte keine unmittelbare Maßzahl geben, mit deren Hilfe etwa Ingenieur- und Geisteswissenschaften in ihren Leistungen verglichen werden können. Aber auf der Basis fachspezifischer zielorientierter Bewertungen ergibt sich eine gemeinsame Metrik, die den Vergleich verschiedener Forschungseinrichtungen gestattet.

Der PPM-Ansatz zielt darauf ab, Leistungsmeß- und Bewertungssysteme zu entwickeln und einzuführen, die genau diese Merkmale und Möglichkeiten beinhalten. Bei der Entwicklung und Einführung sind die in Abb. 1 aufgeführten Schritte zu durchlaufen und die zugeordneten Fragen zu beantworten.

Bevor diese Schritte in Angriff genommen werden, ist noch die Frage nach den "Einheiten" zu klären, für die die einzelnen Schritte realisiert werden sollen. Wir haben uns entschieden, das gesamte Institut als "Einheit" zu betrachten und nicht mit unterschiedlichen Aufgabenkatalogen für verschiedene Arbeitsgruppen bzw. im Institut vertretene Fächer zu beginnen. Die Gründe für eine solche Entscheidung dürften auf der Hand liegen, da das Festlegen von "Bezugseinheiten" für das Meß- und Bewertungssystem auf die kooperativen Strukturen innerhalb und zwischen diesen Einheiten zurückwirken kann.

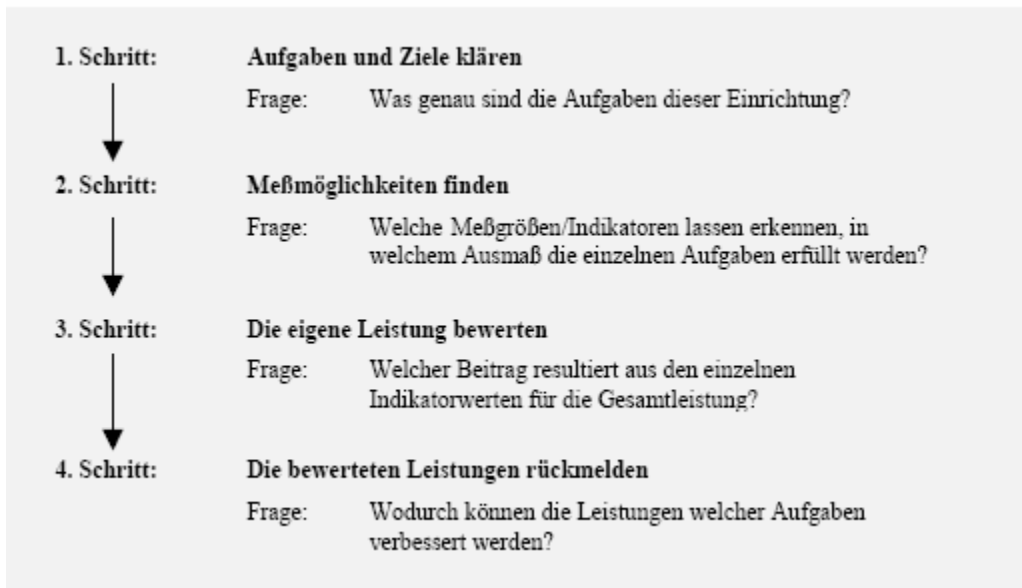


Abb. 1: Schritte bei der Entwicklung und Einführung eines Leistungsmeß- und Bewertungssystems

### **Voraussetzungen für eine erfolgreiche Systementwicklung und -einführung**

Zahlreiche praktische Erfahrungen aus Produktions- und Dienstleistungsunternehmen wie auch unsere eigenen Erfahrungen am Institut belegen, daß für eine erfolgreiche Systementwicklung und -einführung eine Reihe von Randbedingungen erfüllt sein müssen. Ein zentraler Faktor ist die möglichst weitgehende Einbeziehung der von der Bewertung betroffenen Personen, hier der am Institut arbeitenden Wissenschaftler. Die Bedeutsamkeit ihrer Beteiligung an der Systementwicklung und -einführung ergibt sich zum einen aus ihren fundierten Kenntnissen der fachspezifischen Erwartungen an die Leistungen einer Forschungseinrichtung. Zum anderen muß das entwickelte System von ihnen akzeptiert und im Alltagshandeln gelebt und umgesetzt werden.

Ein weiterer bedeutsamer Faktor ist die Konsensbereitschaft unter den an der Systementwicklung beteiligten Personen. Diese kann u.U. sehr gering sein, etwa dann, wenn die Systementwicklung von der Vorstellung geleitet wird, das jeweils erwartete eigene Abschneiden bei den verschiedenen Leistungskriterien in den Vordergrund zu stellen. Um dieser Gefahr vorzubeugen, hat

es sich als hilfreich erwiesen, das System mit Unterstützung eines externen Moderators zu entwickeln. Dieser sollte sowohl über vertiefte Kenntnisse der PPM-Methodik verfügen als auch in der Lage sein, die Diskussion so zu leiten und anzuregen, daß ein breiter Konsens möglich wird. Hierbei haben wir die bemerkenswerte Erfahrung gemacht, daß zunächst prinzipiell gegensätzliche Vorstellungen von Aufgaben und deren Prioritäten an scheinbarer Widersprüchlichkeit verlieren können, wenn es um die konkreten Abbildungen auf Kennzahlen des Bewertungssystems geht und wenn gefragt wird, welche Bedeutung diese Aufgaben denn im Gesamtspektrum aller Anforderungen haben.

Wichtig ist schließlich ebenfalls die Überzeugung, daß eine Leistungsmessung und -bewertung notwendige Voraussetzungen für Leistungsverbesserungen darstellen. Diese Leistungsverbesserungen muß man aber letztlich auch wollen. Systeme der hier vorgestellten Art helfen, dieses Wollen in die richtigen Bahnen zu lenken.

### **Ausgangsbedingungen des IfADo-PPM**

Das Institut für Arbeitsphysiologie an der Universität Dortmund (*IfADo*) betreibt an-

wendungsorientierte Grundlagenforschung zu Sicherung der Arbeitsfähigkeit des Menschen unter Vermeidung tätigkeits- und arbeitsumweltbedingter Beeinträchtigungen. Aus dieser Kernaufgabe ergeben sich einige Besonderheiten des Instituts, die für die Entwicklung eines Leistungsmeß- und Beurteilungssystems von Bedeutung sind. So setzt die Bearbeitung dieser Kernaufgabe interdisziplinäre Forschungsansätze voraus. Im Institut sind daher verschiedene Fachdisziplinen vertreten (wie z. B. Arbeitswissenschaft, Arbeitsmedizin, Toxikologie, Physiologie, Psychologie) sowie weitere Disziplinen (z. B. Mathematik, Chemie), denen eine vorwiegend unterstützende Funktion zukommt. Alle diese Wissenschaftsdisziplinen sind wiederum in fachspezifische Bezugssysteme zur Bewertung ihrer jeweiligen Forschungsleistungen eingebettet.



Abb. 2: Schnittstellen zwischen Institut und Außenwelt

Ein weiteres für die Entwicklung eines Bewertungssystems bedeutsames Merkmal des *IfADo* - und natürlich auch jeder anderen Forschungseinrichtung - ist sein Umfeld, also die "Schnittstelle zwischen Institut und Außenwelt". Wie Abb. 2 verdeutlicht, gehören zu diesem Umfeld nicht nur Institutionen des Wissenschaftsbetriebs und des Arbeitsschutzes, sondern auch die im Vorstand der Trägergesellschaft vertretenen Sozialpartner sowie die betriebliche Praxis. Die Erwartungen und Anforderungen, die aus diesem Umfeld an die Arbeit des Instituts herangetragen werden, sind unterschiedlicher Art und bestimmen nachhaltig sein heterogenes und breites Aufgabenspektrum. Ein solches Umfeld definiert quasi die un-

mittelbaren Adressaten der Arbeit des Instituts, an denen sich die Entwicklung eines Bewertungssystems eng zu orientieren hat. Forschungseinrichtungen dürften sich nicht zuletzt in ihrem jeweiligen Umfeld und den hieraus erwachsenden Aufgaben und erwarteten Leistungen unterscheiden.

Der Interdisziplinarität des Instituts und den Besonderheiten seines Umfeldes wurde bei der Entwicklung des *IfADo*-PPM dadurch Rechnung getragen, daß Wissenschaftler aus verschiedenen Fachdisziplinen und Experten für die jeweiligen Umfeldsegmente des Instituts beteiligt waren. Der Entwicklungsgruppe gehörte ferner der externe Moderator an. Außerdem war ein Mitglied des Betriebsrats stets als Gast vertreten.

Für die Entwicklung des *IfADo*-PPM gab es schließlich auch einen äußeren Anlaß, die Begehung des Instituts durch eine Gutachtergruppe des Wissenschaftsrats. In der Diskussion mit dieser Gruppe, in der die im Institut angesiedelten Fachdisziplinen nur zum Teil vertreten waren, wurde deutlich, daß die angelegten Bewertungsmaßstäbe sehr heterogen und teilweise auch fachfremd waren. Gleichzeitig schienen die Prinzipien der Selbstbewertung des Instituts den Gutachtern nur wenig klar und durchsichtig zu sein. Mit der Entwicklung des *IfADo*-PPM sollten diese Prinzipien auch nach außen hin transparenter und besser vermittelbar werden.

## Entwicklung des *IfADo*-PPM

Die Entwicklungsgruppe nahm ihre Arbeit Ende Mai 1996 auf; sie endete nach insgesamt 15 zwei-stündigen Sitzungen im Mai 1997. Bei der Entwicklung wurde dem in Abb. 1 dargestellten schrittweisen Vorgehen gefolgt. Dementsprechend war zunächst zu klären:

*"Welche Aufgaben hat das Institut zu bearbeiten?"*

Bei der Diskussion dieser Frage wurden fünf Aufgabenbereiche als wesentlich für die Institutsarbeit ausdifferenziert.

1. Aufgabenbereich "**Forschungsorganisation**": Hierunter fallen alle Aktivitäten, die der Vorbereitung und Durchführung von Forschungsarbeiten dienen.
2. Aufgabenbereich "**Verbreitung**": Dieser Aufgabenbereich beinhaltet alle Arbeiten, durch welche das Institut für Arbeitsphysiologie wissenschaftliches Gedankengut und Forschungsergebnisse der wissenschaftlichen Öffentlichkeit zugänglich macht. Die Verbreitung kann in Form von Schriften und in Form von Vorträgen erfolgen. Entscheidend für die Aufnahme eines geleisteten Beitrages in diesem Bereich ist es, daß er sich an einen wissenschaftlichen Adressatenkreis wendet (andernfalls wird er dem Aufgabenbereich "Umsetzung und Anwendung", ggf. auch dem Bereich "Lehre und wissenschaftlicher Nachwuchs" zugeordnet).
3. Aufgabenbereich "**Umsetzung und Anwendung**": In diesen Aufgabenbereich fallen alle Aktivitäten, die dem Transfer von Forschungsergebnissen in die Praxis der Arbeitswelt dienen.
4. Aufgabenbereich "**Lehre und wissenschaftlicher Nachwuchs**": Dieser Aufgabenbereich umfaßt alle Bildungsangebote, die vom Institut geleistet werden, sowie alle Aktivitäten zur Förderung des wissenschaftlichen Nachwuchses.
5. Aufgabenbereich "**Wissenschaftliche Selbstverwaltung**": In diesen Aufgabenbereich fallen alle Aktivitäten, die innerhalb der Strukturen des Wissenschaftsbetriebes von Institutsmitgliedern übernommen werden (wie z. B. Begutachtungen, Organisation von Tagungen usw.).

Das durch diese Bereiche gekennzeichnete Aufgabenprofil verdeutlicht die doppelte Verankerung des Instituts im Wissenschaftsbetrieb einerseits und im Anwendungsfeld der betrieblichen Praxis andererseits. Aufbauend auf diesem Aufgabenprofil war im nächsten Entwicklungsschritt die folgende Frage zu beantworten:

*"In welchen Meßgrößen/Indikatoren spiegeln sich die in den Aufgabenbereichen erbrachten Leistungen des Instituts angemessen wider?"*

Bei der Festlegung bzw. Entwicklung solcher Meßgrößen wurde weitgehend dem Prinzip gefolgt, nur solche Leistungen zu berücksichtigen, die das Institut "nach außen" abgibt. Denn nur solche Leistungen dürften für eine Bewertung von Bedeutung sein. Alle übrigen Leistungen, die quasi im Innern einer Forschungseinrichtung verbleiben, sind nur insofern von Belang, als sie beeinflussen, was nach außen abgegeben wird. Zur Erfassung aller in den Aufgabenbereichen erbrachten Leistungen wurden insgesamt 15 Indikatoren berücksichtigt, die in Tabelle 1 im Überblick aufgeführt sind.

Bei diesen Indikatoren handelt es sich im wesentlichen um zusammengesetzte Größen, z. B. um gewichtete Häufigkeiten unterschiedlicher Arten von Publikationen. Die festgelegten Gewichtungen waren dabei weniger am Arbeitsaufwand für die entsprechenden Leistungen orientiert, sondern vielmehr an der Bedeutung der Leistung oder an den Wirkungsmöglichkeiten des Instituts (z. B. Gremien des Arbeitsschutzes). Zur Verdeutlichung dieses Gewichtungsprinzips sei auf einen Indikator beispielhaft genauer eingegangen, der wohl die umfassendsten Diskussionen ausgelöst hat - der Indikator 3 für den Aufgabenbereich "Verbreitung".

**Tabelle 1: IfADo - PPM: Aufgabenbereiche und Indikatoren**

**I. Aufgabenbereich ‘Forschungsorganisation’**

Indikator 1: Gewichtete Anzahl von Projekten mit abgeschlossener Vorbereitung

Indikator 2: Gewichtete Anzahl von hausinternen Berichten/Vorträgen über zuvor erfolgte Untersuchungen

**II. Aufgabenbereich ‘Verbreitung’**

Indikator 3: Anzahl von veröffentlichten Schriften für wissenschaftlichen Adressatenkreis, gewichtet nach ihrer wissenschaftlichen Bedeutung

Indikator 4: Anzahl von gehaltenen Vorträgen für wissenschaftlichen Adressatenkreis, gewichtet nach ihrer wissenschaftlichen Bedeutung

**III. Aufgabenbereich ‘Umsetzung und Anwendung’**

Indikator 5: Gewichtete Anzahl von veröffentlichten Schriften für anwendungsorientierten Adressatenkreis

Indikator 6: Gewichtete Anzahl von gehaltenen Vorträgen für anwendungsorientierten Adressatenkreis

Indikator 7: Gewichtete Anzahl von Mitgliedschaften in Umsetzungsgremien

Indikator 8: Anzahl von sachbezogenen Stellungnahmen zu Anfragen von Betrieben, Verbänden, öffentlichen Stellen etc.

Indikator 9: Anzahl prototypischer Problemlösungen und Patente

**IV. Aufgabenbereich ‘Lehre und wissenschaftlicher Nachwuchs’**

Indikator 10: Gewichtete Anzahl von Aktivitäten zur akademischen Ausbildung

Indikator 11: Gewichtete Anzahl von nicht-universitären Aktivitäten zur Fort- und Weiterbildung

Indikator 12: Gewichtete Anzahl abgeschlossener Promotionen und Habilitationen

**V. Aufgabenbereich ‘Wissenschaftliche Selbstverwaltung’**

Indikator 13: Gewichtete Anzahl erstellter Gutachten bzw. Schriftleitungen

Indikator 14: Gewichtete Anzahl von Mitgliedschaften in wissenschaftlichen/akademischen Gremien

Indikator 15: Gewichtete Anzahl von Tätigkeiten zur Organisation von Kongressen/Workshops

Dieser Indikator erfaßt die Anzahl veröffentlichter Schriften, die die Autorenschaft zumindest eines Mitarbeiters aus dem Institut für Arbeitsphysiologie tragen. Die Schriften wurden dabei nach ihrer wissenschaftlichen Bedeutung gewichtet. Die Gewichte beziehen sich auf Artikel in Zeitschriften, Monografien, herausgebene Bücher, Buchbeiträge sowie Sonstiges. Die leitende Frage bei der Gewichtung war die nach der Bedeutung und Wirkung einer Schrift (Wie viele Leser erreicht sie? Wie wird ihre Aufnahme bei den Lesern sein?). Der Ausgangspunkt waren Artikel in Zeitschriften, die in Abhängigkeit vom Impact-Faktor (SCI, SSCI) der Zeitschrift ihrer Veröffentlichung wie folgt gewichtet wurden:

*begutachtete Zeitschriften*

IF > 10	x 10
IF > 2	x 8
IF > 0.4	x 6
IF > 0	x 4
kein IF	x 2
nicht begutachtete Zeitschriften	x 0.2

Die Gewichte der anderen Kategorien von Schriften wurden folgendermaßen hierauf abgestimmt:

*Monografien*

Lehrbücher	x 10
Übersichten	x 9
Habilitationen	x 4
Dissertationen und Forschungsberichte	x 2

*Herausgegebene Bücher*

Lehrbücher und Handbücher	x 5
Proceedings und sonstige Bücher	x 2

*Buchbeiträge*

Lehrbücher und Handbücher	x 5
Proceedings und sonstige Bücher	x 2

*Sonstiges*

Abstracts	x 0.1
(in Relation zur veröffentlichten Quelle)	
kurze Beiträge	
2 Seiten oder weniger	x 0.2
(in Relation zur veröffentlichten Quelle)	

Es ist klar, daß es in jedem Fach eine Rangordnung von Zeitschriften gibt. Diese spiegelt sich in prinzipiell brauchbarer Weise auch in den Impact-Faktoren des SCI und des SSCI wider, die die meisten vom Institut genutzten Zeitschriften enthalten. Allerdings sind diese Impact-Faktoren - wie gesagt - in hohem Maße fachspezifisch, und das gilt auch für die im Institut für Arbeitsphysiologie repräsentierten Fachrichtungen. Beispielsweise sind in der Toxikologie wesentlich höhere Impact-Faktoren erreichbar als etwa im Bereich der Arbeitswissenschaft. Zum teilweisen Ausgleich dieser fachspezifischen Unterschiede in den erreichbaren Impact-Faktoren wurde die aufgeführte Klassenbildung von Gewichtungen gewählt, ohne diese Unterschiede jedoch völlig zu nivellieren. Denn diese Unterschiede dürften auch so etwas wie Unterschiede in der momentanen gesellschaftlichen Bedeutung verschiedener Fächer widerspiegeln.

Die hier vorgenommene Klassenbildung ist in gewissem Maße spezifisch für die im *I-fAdo* vertretenen Fachdisziplinen. Ihre Grundlage stellt eine Übersicht über die Impact-Faktoren derjenigen Zeitschriften dar, in denen Mitarbeiter des Instituts in den letzten Jahren veröffentlicht haben. Sie dürfte folglich auch für verschiedene Forschungseinrichtungen unterschiedlich ausfallen. Die Gewichtungen der Meßwerte der meisten anderen Indikatoren folgen einem ganz ähnlichen Bedeutsamkeitskalkül.

Nach Festlegung aller Indikatoren sah der nächste Entwicklungsschritt die Klärung der folgenden Frage vor:

*"Wie sind die Leistungen in den einzelnen Indikatoren zu bewerten?"*

Zur Beantwortung dieser Frage sieht der PPM-Ansatz ein stark formalisiertes Verfahren vor, das in der Festlegung von Bewertungsfunktionen besteht. Dabei ist zunächst der potentielle Wertebereich eines jeden Indikators zu bestimmen, der sich von einem unter ungünstigsten Bedingungen zur erwartenden Indikatorminimum bis zu einem

unter günstigsten Bedingungen zu erwartenden Indikatormaximum erstreckt. Darüber hinaus muß für jeden einzelnen Indikator festgelegt werden, welche Indikatorausprägungen ein neutrales Ergebnis anzeigen, das unter normalen Bedingungen zu erreichen ist und gewöhnlich erwartet werden kann. Um einen Eindruck von der potentiellen Streubreite der Indikatorwerte zu erhalten, können vorliegende Daten aus vorherigen Zeitperioden genutzt werden. Sofern in der Entwicklungsgruppe die Meinung vorherrschte, daß die Leistung des Instituts in dem einen oder anderen Indikator bisher eher unbefriedigend war, wurde der neutrale Wert höher angesetzt. Wegen des Rückgriffs auf Daten zurückliegender Jahre beziehen sich die kalkulierten potentiellen Wertebereiche der Indikatoren auf eine Institutsgröße von 41 Wissenschaftlern. Dies entspricht dem Stellenplan des Jahres 1997. Auf der Grundlage der festgelegten mögli-

chen Wertebereiche eines jeden Indikators war anschließend zu ermitteln, welche Beiträge die einzelnen Indikatorwerte zur Gesamtleistung oder der Produktivität des Instituts leisten. Das genaue Verfahren zur Erstellung solcher Bewertungsfunktionen ist bei Pritchard et al. (1993) beschrieben. Abb. 3 zeigt zwei Beispiele für derartige Funktionen. Im oberen Teil der Abbildung ist die festgelegte Bewertungsfunktion für den Indikator "Gewichtete Anzahl von veröffentlichten Schriften für wissenschaftlichen Adressatenkreis" dargestellt, im unteren Teil die Bewertungsfunktion für den Indikator "Gewichtete Anzahl von veröffentlichten Schriften für anwendungsorientierten Adressatenkreis".

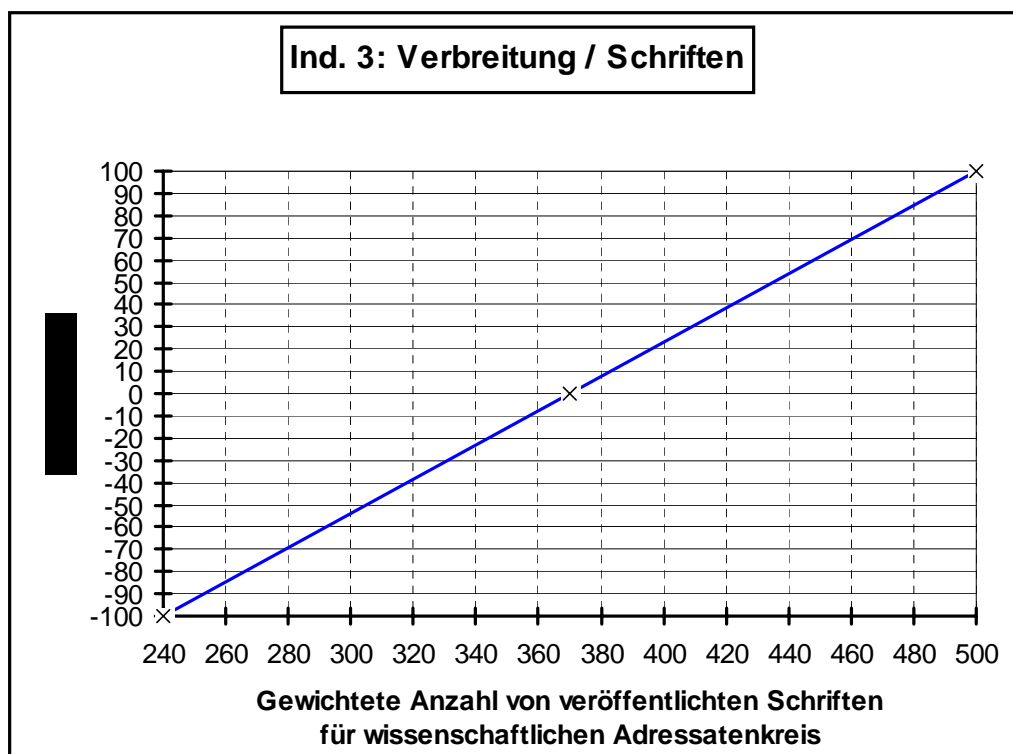


Abb. 3a: Beispiele für Bewertungsfunktionen

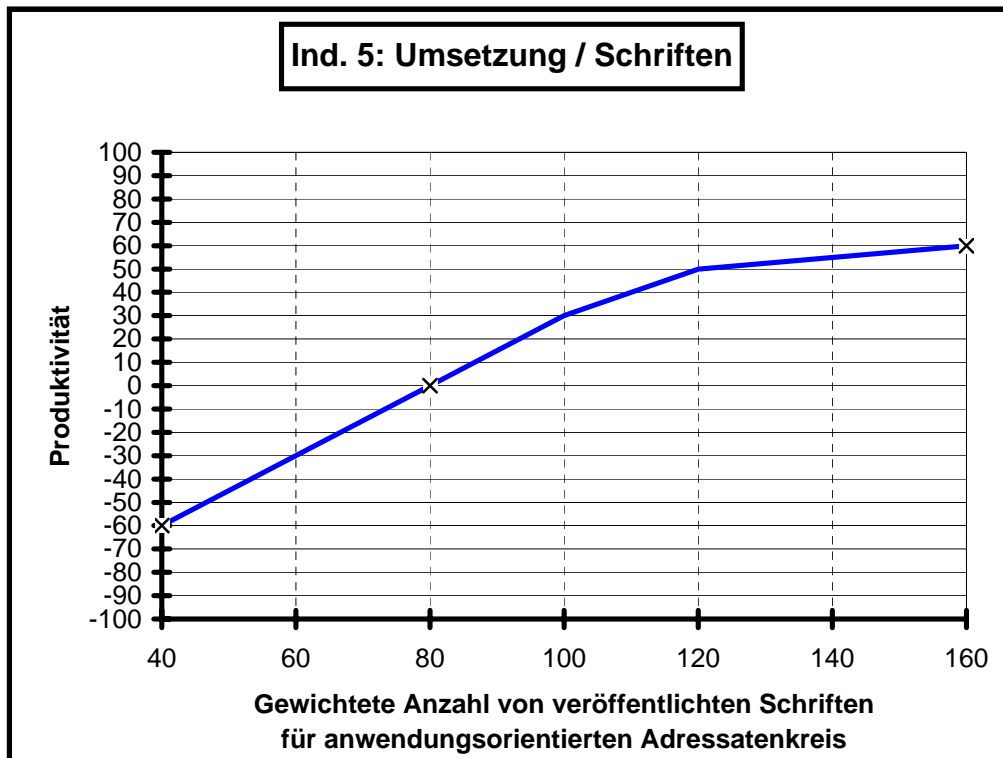


Abb. 3b: Beispiele für Bewertungsfunktionen

Aus dem neutralen Wert für "Schriften an wissenschaftlichen Adressatenkreis" ergibt sich eine Erwartung von 9 Punkten pro Wissenschaftler (die abgetragenen Indikatorwerte beziehen sich auf 41 Wissenschaftler lt. Stellenplan). Dieser neutrale Wert, dem eine Produktivität von 0 entspricht, wäre mit einer guten Publikation in einer Zeitschrift mit einem Impact-Faktor  $>0.4$ , einem Buchbeitrag und einigen Abstracts und Kleinbeiträgen zu erreichen. Das Maximum dieses Indikators liegt bei ca. 12 Punkten pro Wissenschaftler, das Minimum bei ca. 6 Punkten. (Bei diesem Indikator sind wir von dem Prinzip abgewichen, Bewertungen eine obere und untere Asymptote erreichen zu lassen; die Bewertungsfunktion ist hier unbegrenzt linear.)

Die Bewertungen beider Indikatoren lassen zwei bedeutsame Merkmale solcher Bewertungsfunktionen erkennen. Zunächst einmal bringt die Gesamtsteigung der Funktion die relative Bedeutung des Indikators für die Produktivität des Instituts zum Ausdruck. Ein steiler Kurvenverlauf wie für den Indikator "Schriften für wissenschaftlichen Adressatenkreis" weist darauf hin, daß dieser Indikator mit starkem Gewicht zur Gesamt-

leistung des Instituts beiträgt. Das zweite Merkmal, das der Indikator "Schriften für anwendungsorientierten Adressatenkreis" verdeutlicht, betrifft die mögliche Nichtlinearität von Bewertungsfunktionen. D.h., gleich große Veränderungen in verschiedenen Wertebereichen eines Indikators können unterschiedlich starke Produktivitätsveränderungen zur Folge haben. Bewertungsfunktionen mit diesen Merkmalen bieten eine wichtige Orientierungshilfe für die Frage, wie die verfügbaren Ressourcen auf die verschiedenen Aufgabenbereiche und Indikatoren zu verteilen sind, um größtmögliche Leistungssteigerungen zu erzielen. Sie liefern die Grundlage für ein wirksames Ressourcenmanagement. Alle übrigen Bewertungsfunktionen der *IfADo*-Indikatoren sind in Heuer et al. (1998) dokumentiert.

### Nutzung des Bewertungssystems

Ein so erstelltes Leistungsmeß- und Bewertungssystem bildet das Leistungsprofil für das Gesamtinstitut ab. Es erlaubt eine Beurteilung dieser Leistung nach wohldefinierten Maßstäben im Zeitverlauf. Das Bewertungssystem liefert darüber hinaus die Grundlage für eine rationale Diskussion über die Stär-

ken und Schwächen im Leistungsprofil eines Instituts. Hieraus lassen sich Maßnahmen zur Verbesserung des Leistungsprofils ableiten, die dann wiederum anhand des Bewertungssystems hinsichtlich ihrer Wirksamkeit überprüft werden können.

Schließlich macht das entwickelte PPM-System auch Außenstehenden transparent, worin ein Institut seinen Auftrag sieht und in welchem Ausmaß es ihn erfüllt. Dies dürfte nicht zuletzt den Prozeß der externen Evaluation der Institutsarbeit auf eine sachbezogene Grundlage stellen. Eine Rückmeldung über die Qualität der Arbeit kann dann - ganz im Sinne des PPM-Gedankens - zu einer kontinuierlichen Weiterentwicklung des Leistungsprofils einer Forschungseinrichtung beitragen.

Abschließend sei bemerkt, daß allein schon die Tatsache der Erstellung und der Existenz eines PPM-Systems eine leistungsförderliche Wirkung entfalten kann, auch ohne das besondere Maßnahmen ergriffen werden. Das belegen nicht nur die eher informellen Erfahrungen mit dem *IfADo*-PPM, sondern auch systematischere Untersuchungen in verschiedenen Anwendungsfeldern. Allein die Existenz definierter und bewerteter Aufgabenbereiche für eine Einrichtung liefert eine Richtschnur für den Einzelnen, an der er die Nutzung seiner Freiräume bei der Arbeit orientieren kann. Forschung ist in der Regel eine Tätigkeit mit erheblichen Freiräumen, oder eine in einem hohen Maße "offene Aufgabenstellung". PPM-Systeme geben bei derartigen Aufgabenstellungen eine einfache Orientierungshilfe für die Verteilung von Zeit und Anstrengung und lenken sie in eine für die Gruppe oder Einrichtung insgesamt leistungssteigernde Richtung.

## Literatur

Heuer H, Fuhrmann H, Schmidt K-H (1998). Die Beurteilung von Forschungsleistungen. Das Beispiel des Instituts für Arbeitsphysiologie an der Universität Dortmund. Frankfurt/M.: P. Lang

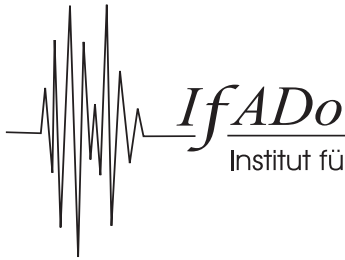
Pritchard RD, Kleinbeck U, Schmidt K-H (1993). Das Managementsystem PPM. Durch Mitarbeiterbeteiligung zu höherer Produktivität. München: C.H. Beck.

Schmidt K-H, Kleinbeck U (1997). Relationships between group-based performance measures, feedback, and organizational context factors. *Europ J Work Organiz Psychol* 6: 303-319.

Schmidt K-H, Hollmann S (1998). Eine deutschsprachige Skala zur Messung verschiedener Ambiguitätsfacetten bei der Arbeit. *Diagnostica* 44: 21-29.

van Tuijl HFJM, Kleingeld A, Schmidt K-H, Kleinbeck U, Pritchard RD, Algera JA (1997). Measuring and enhancing organizational productivity by means of ProMES: Three practical implications. *Europ J Work Organiz Psychol* 6: 279-301.

Werthebach M, Schmidt K-H (1995). Partizipatives Produktivitätsmanagement (PPM): Ein neues Instrument zur zielbezogenen Unterstützung von Gruppenarbeit. In: Knauth P, Wollert A (Hrsg.): *Human Resource Management - Neue Formen betrieblicher Arbeitsorganisation und Mitarbeiterführung* (S 1-33). Köln: Deutscher Wirtschaftsdienst.



*IfADo*

Institut für Arbeitsphysiologie an der Universität Dortmund



*WHO Collaborating Centre  
for Occupational Health*

**H.M. Bolt  
B. Griefahn  
H. Heuer  
W. Laurig  
(Hrsg.)**

# **Arbeitsphysiologie heute**

---

**Bd. 1 (1999)**

---

**Dortmund**

---

ISBN 3-00-004302-0

Als Manuskript gedruckt.

Alle Rechte vorbehalten.

© *IfADo*, Institut für Arbeitsphysiologie an der Universität Dortmund  
Ardeystr. 67, D-44139 Dortmund  
Tel.: 0231/1084-0  
Fax: 0231/1084-308

Druck: Koffler-Druck, Dortmund

Printed in Germany

---

## Vorwort

Mit der Umstrukturierung des Instituts für Arbeitsphysiologie an der Universität Dortmund (*IfADo*) zum 1.5.1997 erhielten die jährlichen Forschungsberichte und Forschungspläne eine neue Form. Die inhaltliche Berichterstattung über die *IfADo*-Projekte wird ab 1999/2000 in den Forschungsplan mit integriert. Hierdurch ergibt sich die Möglichkeit, die Berichterstattung des *IfADo* um Elemente zu erweitern, die sich besonders an Abnehmern der *IfADo*-Forschungsergebnisse orientieren.

Diskussionen über geeignete Formen einer solchen Präsentation wurden im Verlauf des Jahres 1998 im *IfADo* geführt und ergaben den Vorschlag der Publikation einer neuen jährlichen Institutsreihe. Diese möchte anhand wechselnder Beiträge aus dem Institut unterschiedliche Arten des Brückenschlages zwischen arbeitsphysiologischer Grundlagenforschung und der Praxis aufzeigen.

Der erste Band 1999 enthält aktuelle exemplarische Beiträge aus allen fachlichen Bereichen des *IfADo*. Die Institutsleitung des *IfADo* stellt den neuen Ansatz nun zur Diskussion und möchte diesen in den Folgejahren weiterentwickeln.

gez. Univ.-Prof. Dr. Dr. Hermann M. Bolt  
(Institutsdirektor des *IfADo* )