

Stefan Kuhlmann und Thomas Heinze\*

## Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf

### *Teil I: Konzeptionelle Grundlagen*

#### *I. Forschungsleistungen: Informationsbedarf*

Über ein Drittel aller in Deutschland betriebenen Wissenschaft und Forschung wird von der öffentlichen Hand finanziert, knapp zwei Drittel von Unternehmen der Wirtschaft. Die staatlichen Forschungsmittel fließen überwiegend an Universitäten und außeruniversitäre öffentliche Forschungseinrichtungen. Allein die Deutsche Forschungsgemeinschaft (DFG) finanziert etwa 40 % aller Drittmittel an Hochschulen. Traditionell fördert die DFG Forschungsprojekte unter der Bedingung einer ex ante-Evaluation der wissenschaftlichen Qualität von Projektvorschlägen durch Fachkollegen (*peer review*). Dieses Verfahren ist international anerkannt. Eine ex post-Bewertung der Ergebnisse und Auswirkungen geförderter Forschung, etwa Projekte oder Programme der DFG und anderer Forschungsförderungseinrichtungen erfolgt, abgesehen von einzelnen ad-hoc-Evaluationen, bis heute nicht systematisch und regelmäßig. Ein Grund dafür liegt in dem Umstand, dass das deutsche Forschungssystem – bekanntlich institutionell stark „versäult“<sup>1</sup> – bis heute kein nach einheit-

---

\* Prof. Dr. Stefan Kuhlmann, Fraunhofer-Institut für Systemtechnik und Innovationsforschung (ISI) sowie Copernicus Institute, Universität Utrecht; Dipl.-Soziologe Thomas Heinze, ISI sowie Stanford University.

<sup>1</sup> In den vergangenen zehn Jahren haben mehrere „Systemevalutionen“ der großen deutschen Forschungsorganisationen kritisiert, dass diese institutionell zu sehr gegeneinander abgeschottet sind; vgl. *Evaluierungskommission* (Hrsg.), Systemevaluierung der Fraunhofer-Gesellschaft. Bericht der Evaluierungskommission, München, 1998, <http://www.fraunhofer.de/german/company/fhmodel/evaluierungsbericht.pdf>; *Internationale Kommission*, Forschungsförderung in Deutschland. Bericht der internationalen Kommission, zur Systemevaluation der Deutschen Forschungsgemeinschaft und der Max-Planck-Gesellschaft, Studie im Auftrag der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK), herausgegeben von der Volkswagen-Stiftung, Hannover, 1999; *Wissenschaftsrat* (Hrsg.), Systemevaluation der Blauen

lichen Kriterien organisiertes Verfahren zur Erhebung von Input- und Output-Parametern kennt<sup>2</sup>, weder auf der Arbeitsebene des Forschungsgeschehens, noch auf höheren Aggregatebenen, also Instituten, Fakultäten, Universitäten und Forschungsorganisationen.

Dem stehen drei Entwicklungen gegenüber: (1) Der Ruf nach Rechenschaft über Leistungen und Qualitäten staatlich geförderter Wissenschaft und Forschung wird lauter, (2) Durch Einführung von Globalhaushalten in Verbindung mit geplanten oder bereits etablierten neuen, leistungsorientierten Modi der Verteilung von Forschungsmitteln wachsen die Autonomiespielräume der Hochschulen. (3) Das in Gang gesetzte Prinzip des Wettbewerbs zwischen Forschungseinrichtungen fordert vom Wissenschaftsmanagement der Universitäten und Institute die Ausprägung unverwechselbarer, wettbewerbsfähiger Forschungsprofile. Diese Entwicklungen erfordern ein Mindestmaß an Informiertheit über universitäre und außeruniversitäre Forschungsleistungen. Auch auf europäischer Ebene stellt sich diese Herausforderung: Mit Blick auf den künftigen „Europäischen Forschungsraum“ versucht die Europäische Kommission gegenwärtig Leistungsindikatoren zu identifizieren und in ein Berichtssystem zu implementieren.<sup>3</sup> In der Folge entwickeln die für Forschungsplanung, -management, und -politik Verantwortlichen einen wachsenden Bedarf nach Informationen über erbrachte Forschungsleistungen und die Rolle institutioneller und finanzieller Rahmenbedingungen von Forschungstätigkeit.

Liste – Stellungnahme des Wissenschaftsrates zum Abschluss der Bewertung der Einrichtungen der Blauen Liste, Köln, 2000; *Wissenschaftsrat* (Hrsg.), *Systemevaluation der HGF – Stellungnahme des Wissenschaftsrates zur Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren*, Köln, 2001.

<sup>2</sup> Auch in anderen Ländern finden sich keine völlig vereinheitlichten Evaluationssysteme. Anders als in Deutschland gibt es in einigen Ländern allerdings systematische, ganze institutionelle Kontexte umspannende Evaluationsverfahren, wie etwa die regelmäßige *Research Assessment Exercise* in Großbritannien (vgl. *L. Georghiou et al.*, *Impact of the Research Assessment Exercise and the Future of Quality Assurance in the Light of Changes in the Research Landscape*, A report produced by PREST, University of Manchester for HEFCE, 2000) oder das generelle Erfordernis eines Leistungsnachweises für jede staatliche Investition, also auch im Bereich öffentlicher Forschungsförderung, wie im Falle des *Government Performance and Results Act* in den USA (vgl. *S. Cozzens*, *Frameworks for evaluating S&T policy in the United States*, in: *Ph. Shapira/S. Kuhlmann* (Hrsg.), *Learning from Science and Technology Policy Evaluation, Experiences from the United States and Europe*, Cheltenham/UK 2003 (E. Elgar), 54–64).

<sup>3</sup> Vgl. *R. Ragnarson*, *Zukünftiger europäischer Forschungsraum*, in: *FuE-Info 2/2001*, S. 8–11; *European Commission*, *Towards a European Research Area*, Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions, Brussels, 18 January 2000, COM (2000) 6.

Gegenwärtig wird dieser Bedarf in Deutschland von den vorhandenen Einrichtungen – den Verwaltungen der großen Forschungsträgerorganisationen, den Hochschulen sowie ihren Verbänden, den Forschungsförderungseinrichtungen oder auch dem Wissenschaftsrat – nicht hinreichend gedeckt. Es ist nicht nur schwierig, einen deskriptiven Überblick über die Forschungslandschaft zu erhalten. Es ist auch kaum möglich, relevante Leistungsparameter – sofern sie gemessen werden – in Relation zu aussagefähigen, gleich aggregierten Input-Größen zu setzen, um Analysen der Dynamik sowie der Stärken und Schwächen des Forschungssystems zu betreiben. Strukturbildende Effekte von Forschungsförderung – etwa die Herausbildung international anerkannter Forschungsschwerpunkte (*Centres of Excellence*) oder Auswirkungen auf das Bildungsangebot der Hochschulen sind ebenfalls bisher nicht systematisch erfassbar. Es fehlen vor allem kompatible Formate und Rhythmen der Informationserfassung und -aufbereitung, einerseits quer zu den Forschungseinrichtungen und andererseits anschlussfähig an internationale Vergleiche.

Vor diesem Hintergrund drängt sich die Frage auf, ob eine Dienstleistungseinrichtung eingerichtet werden sollte, die Informationen zu Forschungsleistungen systematisch sammelt, dokumentiert und analysiert. Vorgeschlagen wird die Schaffung einer „Gemeinschaftsstelle Clearinghouse“ für Forschungsstrukturinformation, die von einer freiwilligen Kooperation der deutschen Forschungseinrichtungen getragen wird und als Knotenpunkt der systematischen Informationsbeschaffung, -analyse und -dokumentation dient. Die „Gemeinschaftsstelle Clearinghouse“ sollte von einem Konsortium gesteuert werden, dem alle relevanten deutschen Einrichtungen der Forschungspolitik, der Forschungsförderung, der Forschungsträgerinstitutionen sowie ausgewählten Einrichtungen der Wissenschafts- und Innovationsforschung zugehören.

Der vorliegende Text ist die überarbeitete Fassung eines Gutachtens für die Geschäftsstelle der Deutschen Forschungsgemeinschaft (DFG).<sup>4</sup> Der Text präzisiert, was begrifflich und konzeptionell unter „evaluative Informationen über Forschungsleistungen“ zu verstehen ist. Hierzu wird nach einer begrifflichen Bestimmung (Abschnitt II) eine Auswahl von anerkannten Indikatoren präsentiert, mit deren Hilfe evaluative Informationen über Forschungsleistungen generiert werden können (Abschnitt III). Zur Sprache kommen auch methodische Evaluationsansätze (Abschnitt IV), Zwecke der Evaluation (Abschnitt V) und die Grenzen sol-

<sup>4</sup> S. *Kuhlmann/Th. Heinze*, Informationen zur Forschungsevaluation in Deutschland – Erzeuger und Bedarf, Gutachten für die Geschäftsstelle der Deutschen Forschungsgemeinschaft, Karlsruhe (Fraunhofer ISI Discussion Papers Innovation System and Policy Analysis, 3/2003; <http://www.isi.fhg.de/ti/Downloads>).

cher Verfahren (Abschnitt VI). In einer zweiten, späteren Veröffentlichung soll und explorativ skizziert werden, wie und von wem evaluative Informationen von Forschungsleistungen in Deutschland bisher erbracht und dokumentiert werden. Auch werden Empfehlungen zur Gestaltung einer „Gemeinschaftsstelle Clearinghouse“ für Forschungsstrukturinformation gegeben.

## II. Was sind Forschungsleistungen?

Unter Forschung versteht man wissenschaftliche Arbeit, welche die Produktion neuen Wissens und auch neuer Technologien zum Ziel hat. Dazu nutzt Forschung bestimmte Ressourcen (z.B. Personal, Infrastruktur) und erbringt Leistungen. Ein solches Input-Throughput-Output-Modell des Forschungsprozesses findet sich in unterschiedlichen Spielarten in der Literatur.<sup>5</sup> Was Forschungsleistung materiell bedeutet, lässt sich jedoch nicht umstandslos definieren. Im Unterschied zur Wirtschaftsleistung, die über das Periodenergebnis von Unternehmen ermittelt wird, verfügt die Wissenschaft über keinen vergleichbar eindeutigen Ansatz, Leistung zu messen. Forschungsleistung wird aus diesem Grund anhand unterschiedlicher Kriterien definiert, etwa als Leistungsmenge, hinsichtlich ihrer Kreativität und Originalität, ihrer Sichtbarkeit und Anschlussfähigkeit, der durch sie erzeugten Wirkung oder der Effizienz ihrer Erstellung.<sup>6</sup>

Diese an sich nützlichen, jedoch in Form einer Liste unübersichtlichen Kriterien lassen sich ordnen. Naheliegend erscheinen die folgenden Dimensionspaare: (1) Quantität und Qualität, (2) Effizienz und Effektivität, (3) Produktivität und Wirkung. Das erste und zweite Dimensionspaar bezeichnen sachliche Aspekte. Zum einen geht es um Leistungsmenge und Leistungseigenschaften. Forschungsleistung lässt sich in diesem Sinne „abzählen“, aber auch „wiegen“. Zum zweiten wird Leistungserstellung mit Blick auf eingesetzte Ressourceneinheiten (Effizienz) und zum anderen hinsichtlich definierter Zielgrößen (Effektivität) evaluiert. Das dritte

---

<sup>5</sup> Vgl. *I. Liefner*, Leistungsorientierte Ressourcensteuerung in Hochschulsystemen. Ein internationaler Vergleich, Berlin, 2001, 39 f.

<sup>6</sup> *H. Bolsenkötter*, Ansätze zur Erfassung und Beurteilung von Forschungsleistungen, in: *Fisch/Daniel* (Hrsg.), Messung und Förderung von Forschungsleistung, Konstanz, 1986, 41; *B. Martin/J. Skea*, Academic research performance indicators: An assessment of the possibilities. Science Policy Research Unit, University of Sussex 1992; *B. Bozeman, B./J. Melkers* (Hrsg.), 1993, Evaluating R&D Impacts: Methods and Practice, Boston et al. 1993 (Kluwer); *G.B. Jordan/D. Streit*, Recognizing the competing values in science and technology organizations: implications for evaluation, in: *Shapira/Kuhlmann*, aaO., 2003, 316–338.

Dimensionspaar bezeichnet ein Zeitverhältnis und unterscheidet unmittelbar Produziertes von mittelbarer Wirkung. Forschungsleistungen werden differenziert nach bloßer Existenz und erzeugter Resonanz. In der Summe geben die drei Dimensionen somit Auskunft über das „wie“ von Forschungsleistung: Wie viel?, Wie gut?, Wie produktiv?, Wie effizient? usw.

### III. Typen evaluativer Information

Forschungsleistungen lassen sich durch eine Vielzahl von Indikatoren beschreiben. Im Folgenden soll eine Auswahl solcher Wissenschafts- und Forschungsindikatoren entlang der erwähnten Dimensionspaare vorgestellt werden. In den vergangenen Jahrzehnten wurde eine relativ elaborierte Wissenschafts- und Innovationsindikatorik entwickelt, die als Grundlage für eine Matrix von Informationstypen dienen kann.<sup>7</sup> Die Tabellen 1, 2 und 3 stellen ein Satz von Indikatoren vor, die sich zum einen auf Produktivitäts- bzw. Wirkungsindikatoren beziehen, und zum anderen auf Input-Maße, welche Referenzgrößen zur Bewertung von Forschungsleistung darstellen. Die Tabellen sind nicht als vollständig zu verstehen, sondern zeigen vor allem standardisierte und mit der deskriptiven sowie schließenden Statistik auswertbare Kennziffern.

Zunächst umfasst Forschungsleistung Produktivitätsaspekte. Hierzu zählen Forschungshandeln, das Wissensprodukte erzeugt, und die Ausbildung wissenschaftlichen Nachwuchses (vgl. Tabelle 1). Im quantitativen Sinne kann Forschungshandeln über die Zahl laufender Projekte und Forschungsaufenthalte bestimmt werden. Diese Indikatoren können in qualitativer Hinsicht näher beschrieben werden durch die Länge der Laufzeit einer Forschungsaktivität oder ihre Finanzierungsquelle. So deuten längerfristig angelegte DFG-Projekte eher auf Grundlagenforschung hin, während im Rahmen von Regierungsprogrammen finanzierte Forschungsprojekte mit geringer Laufzeit eher zur angewandten oder strategischen Forschung gerechnet werden. Ein solcher Profilunterschied kann

---

<sup>7</sup> Vgl. R. Fisch/H.-D. Daniel (Hrsg.), Messung und Förderung von Forschungsleistung, Konstanz, 1986; H.-D. Daniel/R. Fisch (Hrsg.), Evaluation von Forschung, Konstanz, 1988; P. Weingart/R. Sebringer/M. Winterbager, (Hrsg.), Indikatoren der Wissenschaft und Technik, 1991; W. G. Stock, Wissenschaftsevaluation. Die Bewertung wissenschaftlicher Forschung und Lehre, ifo Diskussionsbeiträge, München, 1994; S. Hornbostel, Wissenschaftsindikatoren. Bewertungen in der Wissenschaft, Opladen, 1997; S. Hornbostel, Kurzgutachten für die DFG zum Stand der Forschung im Bereich Forschungsevaluation, Jena, 2000.

sich auch in der Wissensproduktion niederschlagen, so dass im erstgenannten Fall mehr und vor allem in referierten Fachzeitschriften publiziert wird, während im zweiten Fall Forschungsberichte erstellt werden, die eher für die Forschungsfinanciers als für die wissenschaftliche Gemeinschaft von Interesse sind.

Tabelle 1:  
Indikatoren zur Messung wissenschaftlicher Produktivität (Auswahl)

	Quantitative Information	Qualitative Information
Forschungshandeln	Projekte	Laufzeit Finanzierung (DFG, Ministerien, Stiftungen) Beitrag zur Profilbildung
	Forschungsaufenthalte	Regelmäßigkeit Finanzierung (Stipendien, Kooperationsverträge)
Wissensprodukte	Publikationen (alle Wissenschaften) Patentanmeldungen (Technikwissenschaften)	Monographie/Zeitschriftenbeiträge referiert/nicht-referiert, national/international Technologiegebiet, nationales/europäisches Patentamt
Ausbildung wissenschaftlichen Nachwuchses	Promotionen	Benotung Promotion Dauer Promotion

Forschungshandeln und Wissensprodukte bzw. die Ausbildung wissenschaftlichen Nachwuchses sind nicht notwendigerweise korreliert. Denkbar ist, dass Universitätsforscher, die keine drittmittelfinanzierten Projekte durchführen, zahlreiche theoretische Aufsätze in international angesehenen Fachzeitschriften publizieren und eine geringe Zahl von Doktoranden beschäftigen, die in kurzer Zeit promovieren. In welcher Form Wissensprodukte zur Verfügung gestellt werden, variiert nach Disziplinen. Für Geisteswissenschaftler hat die Monographie eine höhere Bedeutung als für Physiker und Chemiker, bei denen die Forschungskommunikation weitgehend in internationalen Fachzeitschriften stattfindet.<sup>8</sup>

<sup>8</sup> H.-D. Daniel/R. Fisch (Hrsg.), aaO., 1988, Kapitel II.

Tabelle 2:  
Indikatoren zur Messung wissenschaftlicher Wirkung (Auswahl)

	Quantitative Information	Qualitative Information
Sichtbarkeit (in der Wissenschaft)	Zitation Rezensionen, Repliken	Nachhaltigkeit/Halbwertszeit der Zitationen in referierten/intern. Fachzeitschriften
Reputation (in der Wissenschaft)	Berufungen Forschungspreise, Stipendien	Reputation der rufenden Universität Art und Status des Preises bzw. der Stipendien
Ausstrahlung (in die Gesellschaft)	Lizenzen, Patenzitate Gutachten, Beratungs- tätigkeiten Vorträge	Art der Erlösverwendung Technologiegebiet Art der auftraggebenden Einrich- tung (Stiftungen, Ministerien) Art des Veranstalters (z.B. Industrie- kammer, Investorenkonferenzen)

Forschungsproduktivität ist somit kein eindimensionales, quantitativ fassbares Konstrukt, sondern weist entscheidende qualitative Dimensionen auf.

Forschungsleistung kann außerdem über Wirkungsindikatoren ermittelt werden, die Aufschluss geben, in welchem Umfang Forschungsleistungen in der Wissenschaft sichtbar und anschlussfähig sind, welche wissenschaftliche Reputation sie langfristig erzeugen und wie sie in die gesellschaftliche Umwelt ausstrahlen (vgl. Tabelle 2). Werden publizierte Schriften zitiert, in Rezensionen besprochen und in Repliken bzw. durch Zitationen zum Gegenstand einer Diskussion gemacht, deutet dies darauf hin, dass geleistete Forschung wissenschaftlich relevant ist. Die Frage, welche Qualität eine solche Sichtbarkeit hat, kann über ihre Dauerhaftigkeit (Halbwertszeit) und den Ort der Sichtbarkeitserzeugung (referierte Fachzeitschriften, Sammelbände etc.) bestimmt werden. Werden Schriften häufig in viel gelesenen Fachzeitschriften aufgegriffen und diskutiert, erzeugt dies eine bessere Sichtbarkeit als beispielsweise in Sammelbänden mit geringer Auflage. Eine weitere Wirkung besteht in aufgebauter Reputation, dokumentiert durch Forschungspreise, Stipendien oder Berufungen. Je nachdem, welche Universität ruft, welcher Preis verliehen und welches Stipendium gewährt wird, erwirbt der Forscher mehr oder weniger Reputation. Weiterhin kann auch die Ausstrahlung von Forschung in Politik und Wirtschaft mit Indikatoren beschrieben werden. Fließen Patent-

anmelden Lizenzannahmen zu, indiziert dies eine hohe Verwertbarkeit des produzierten Wissens. Wird von Unternehmen oder öffentlichen Einrichtungen Expertise in Form von Gutachten oder Beratungen nachgefragt, so weist dies auf außerwissenschaftliche Sichtbarkeit und Verwertbarkeit wissenschaftlichen Wissens hin.

In der Evaluation können Produktivitäts- und Wirkungsindikatoren systematisch auf Referenzkategorien, zumeist Inputgrößen zurückbezogen werden. Produktivität und Wirkung können mit verschiedenen Input-Indikatoren korrelieren, etwa unterschiedlichen Altersstufen und Gruppen von Wissenschaftlern (Mittelbau versus Professoren), oder einen Zusammenhang zwischen der Qualität der Infrastruktur und der erbrachten Forschungsleistungen anzeigen. Auf diese Weise entstehen Informationen mit Strukturaussage, die Vorbedingung für Wissenschaftsmanagement und -politik sind. Tabelle 3 fasst eine Auswahl von Input-Indikatoren unter den Aspekten Personal, Finanzierung und Infrastruktur zusammen.

Tabelle 3:  
*Input-Indikatoren (Auswahl)*

	Quantitative Information	Qualitative Information
Personal	Anzahl Wissenschaftler Anzahl Nicht-Wissenschaftler	Anteil der promovierten Wissenschaftler Relation Wiss./Nicht-Wiss. Relation Mittelbau/C3–C4 Stellen Altersstruktur Disziplinenportfolio
Finanzierung	Volumen Veränderung des Volumens	Herkunft (Grundfinanzierung, Drittmittel)
Infrastruktur	Quadratmeter Fläche Anzahl Personal-computer und Rechensysteme	Alter, Baulicher Zustand technische Leistungsmerkmale

Forschungsleistungen werden in der Evaluationspraxis auf unterschiedlichen Aggregationsebenen gemessen. Wie Tabelle 4 im Überblick zeigt, bezieht sich die horizontale Ebene auf Fachdisziplinen und Forschungsfelder, auf Forschungsprogramme und schließlich auf die institutionelle „Landschaft“, innerhalb welcher Forschung stattfindet. Sie er-

streckt sich quer zu konkreten Personen und Organisationen. Die vertikale Ebene reicht von Organisationen über Institute und Arbeitsgruppen bis hin zum Einzelwissenschaftler. Sie gibt die Tiefenschärfe an, mit welcher Leistungsindikatoren in Forschungseinrichtungen erhoben werden können. Beide Ebenen sind zumeist verschränkt: So wurde das Feld Energieforschung durch den deutschen Wissenschaftsrat auf disziplinärer und organisationaler Ebene untersucht;<sup>9</sup> die Evaluation der Fächer Wirtschaftswissenschaften und Geschichte durch die Wissenschaftliche Kommission Niedersachsen<sup>10</sup> wurde auf Fakultäts- und Forschergruppen-Ebene durchgeführt.

Tabelle 4:  
*Aggregations-Ebenen der Wissenschaftsindikatorik*

<b>Horizontale Ebene</b>
Fachdisziplinen (z.B. Chemie, Biologie), Forschungsfelder (z.B. Energieforschung)
Programme (z.B. BMBF, EU-Kommission, DFG)
Institutionelle Forschungslandschaft (z.B. Systemevalationen)
<b>Vertikale Ebene</b>
Organisation (z.B. Universität, Max-Planck-Institut, Fraunhofer-Institut)
Department (z.B. Fachbereiche, Fakultäten, Abteilungen)
Forschergruppen (innerhalb von Departmentstrukturen)
Einzelwissenschaftler

#### *IV. Methodische Ansätze der Evaluationsforschung*

Wir kennen heute vielfältige Methoden zur Analyse erzielter oder potentieller Produktivität und Wirkungen von Forschung; diese Methoden stützen sich in unterschiedlicher Weise auf die skizzierten Indikatoren.<sup>11</sup> Neben der „klassischen“ Bewertung von Forschungsleistungen durch

<sup>9</sup> *Wissenschaftsrat* (Hrsg.), Stellungnahme zur Energieforschung, Köln, 1999.

<sup>10</sup> *Wissenschaftliche Kommission Niedersachsen* (Hrsg.), Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen, Chemie, Bericht und Empfehlungen, Hannover, 1999; *Wissenschaftliche Kommission Niedersachsen* (Hrsg.), Forschungsevaluation an niedersächsischen Hochschulen und Forschungseinrichtungen. Wirtschaftswissenschaften. Bericht und Empfehlungen, Hannover, 2001.

<sup>11</sup> Vgl. *R. Fisch/H.-D. Daniel* (Hrsg.), aaO., 1986; *H.-D. Daniel/R. Fisch* (Hrsg.), aaO., 1988; *A. F. J. van Raan* (Hrsg.), Handbook of Quantitative Studies of Science and

Fachkollegen (Peer Review, Plausibilitätsüberprüfungen, Schätzurteile) sind die wichtigsten der Vorher/Nachher-Vergleich, wissenschaftsinterne/externe Evaluation, Kosten/Nutzen-Analysen sowie der Kontroll- oder Vergleichsgruppenansatz. Sie greifen auf die genannten Indikatoren zurück, die mit Hilfe von Datensammlungsverfahren (z.B. primäre: mündliche oder schriftliche Befragungen, bibliometrische Datensammlung; oder sekundäre, prozessproduzierte Daten) zusammengetragen und Datenanalyseverfahren (deskriptive und schließende Statistik) einzeln oder kombiniert ausgewertet werden. Zu den anerkannten Methoden zählen die folgenden:

*Vorher/Nachher-Vergleich:* Peer Review wird meist für die ex ante-Prüfung wissenschaftlicher Leistung verwendet. So werden im Normalverfahren der Deutschen Forschungsgemeinschaft (DFG) eingereichte Forschungsprojekte ex ante auf ihre wissenschaftliche Eignung hin überprüft.<sup>12</sup> Eine ex post-Bewertung fragt demgegenüber: Was wurde erreicht? Mit welchem Aufwand? Auf welche Weise? Mit welchen Wirkungen und Nebenfolgen? Das sind typische Fragestellungen, die Vorher/Nachher-Vergleiche kennzeichnen. Gestellt werden solche Fragen vor allem von Akteuren der Forschungs- und Technologiepolitik, die neben der institutionellen Förderung auch Forschungsprojekte und -programme finanzieren und auf Informationen zum – im engeren und weiteren Sinne – Erfolg bzw. Misserfolg solcher Maßnahmen angewiesen sind, nicht zuletzt um politisch-staatliches Handeln zu legitimieren.

Bei genauem Hinsehen fällt auf, dass der Vorher/Nachher-Vergleich keine eigenständige Methode zur Evaluation von Forschungsleistungen ist – dies wird er erst in Kombination mit beispielsweise Kosten/Nutzen-Analysen oder Kontroll- bzw. Vergleichsgruppenansätzen (siehe unten) –, sondern gegenüber dem Peer Review vor allem ein Zeitverhältnis problematisiert. Es geht um die Frage, an welchem Zeitpunkt des Forschungs-

---

Technology, Amsterdam: Elsevier, 1988; *H.-D. Daniel*, Guardians of Science – Fairness and Reliability of Peer Review, Weinheim, 1993; *B. Bozeman/J. Melkers*, aaO. 1993; *H. Grupp/U. Kuntze/U. Schmoch*, New Technology Indicators for the Evaluation of Research and Development Programmes, in: Becher/Kuhlmann (Hrsg.): Evaluation of Technology Policy Programmes in Germany, Boston/Dordrecht/London: Kluwer Academic Publishers, 1995, 243–284; *S. Hornbostel*, aaO., 1997; *S. Hornbostel*, aaO., 2000; *J. Youtie/B. Bozeman/P. Shapira*, Using an evaluability assessment to select methods for evaluating technology development programs: the case of Georgia Research Alliance, in: Evaluation and Program Planning 22 (1999), 55–64; *S. Kublmann*, Evaluation of Research an Innovation Policies: A Discussion of Trends with Examples from Germany, in: International Journal of Technology Management, 26 (2003), 131–149.

<sup>12</sup> Vgl. *F. Neidhardt*, Selbststeuerung in der Forschungsförderung. Das Gutachterwesen der DFG, Opladen: Westdeutscher Verlag, 1988.

prozesses Steuerungsinstanzen eingebaut werden und welche Wirkungen dies auf den Verlauf und den Typ der geleisteten Forschungsarbeit hat. Das „klassische“ Modell war die Antizipation von Forschungsqualität, nicht ihre retrospektive Bewertung. Schließlich impliziert Wissenschaftsfreiheit im voll ausgeschöpften Wortsinne Handlungsentlastung der Wissensproduzenten von außerwissenschaftlichen Interessen und Ansprüchen. Auf dieses Prinzip wird immer wieder hingewiesen, wenn die Diskussion ex post-Evaluationen thematisiert. An dieser Stelle soll nicht abschließend bewertet werden, ob ein resultativer Zugang zum Phänomen der Forschungsleistung dem antizipativen vorzuziehen ist oder nicht. Die Entwicklung der letzten Jahre, insbesondere im europäischen Ausland (z.B. Finnland, Großbritannien, Niederlande), lässt es aber angeraten erscheinen, ex post-Evaluationen des Wissenschaftsbetriebes zu etablieren, um komplementär zum weiter laufenden antizipativen Peer Review Steuerungswissen zu erzeugen.

*Kosten/Nutzen-Analysen* fragen danach, in welchem Verhältnis das Forschungsprodukt oder die Forschungswirkung auf der einen Seite und die eingesetzten Ressourcen auf der anderen Seite stehen. Kosten-Nutzen-Analysen versuchen, dieses Verhältnis in eine numerische Relation zu bringen. Angaben wie die Publikationsanzahl je eingesetzter Geldsumme können diesbezüglich nur ein grobes erstes Bild erzeugen, das analog zu den Beispielen in Tabelle 1 qualitativ erweitert werden muss. Dabei ist im Auge zu behalten, dass konkrete Kosten-Nutzen-Analysen immer nur so gut sind wie die jeweils verwendeten Rechenmodelle, die Angemessenheit und Realitätsnähe der modellierten Annahmen sowie die Datenqualität. Die Veränderung von Modellannahmen, die Nichtverfügbarkeit bestimmter Informationen, allgemeine Messprobleme oder nicht-antizipierte Nebeneffekte können die Anwendbarkeit und Zuverlässigkeit dieses methodischen Ansatzes erschweren. Eine wesentliche Bedingung seines Anwendungserfolges besteht in einer vertieften Kenntnis der zu evaluierenden Untersuchungseinheit.

Allerdings sollten die Schwierigkeiten nicht davon abhalten, diesen Ansatz intelligent einzusetzen. Aus Sicht der Governance von Forschungseinrichtungen (universitär und außeruniversitär) erscheint es beispielsweise wichtig, die Relation der administrativen gegenüber den forschungsproduktiven Kosten zu ermitteln, um Ansatzpunkte für die Steuerung von Forschungsorganisationen zu erhalten. Das Ausmaß der Verwaltungskosten variiert institutionenspezifisch und zeitabhängig; Informationen darüber sind nützlich für Entscheidungsträger im Wissenschaftsmanagement und Forschungspolitik, aber auch für die Evaluierten selbst.

*Kontroll- und Vergleichsgruppenansätze:* Ein weiterer methodischer Ansatz ist der Rückbezug von Forschungsleistungen und –wirkungen auf geeignete Referenzeinheiten, die mit der zu evaluierenden Untersuchungseinheit bestimmte Eigenschaften teilen. Dieser Ansatz stammt ursprünglich aus der experimentellen Psychologie und hat auch in der Evaluationsforschung Fuß gefasst. Das Erkenntnisinteresse ist hier darauf gerichtet, die in den Tabellen 1 und 2 überblicks- und ausschnitthaft dargestellten Informationstypen interpretationsfähig zu machen, indem man für eine Mehrheit von Untersuchungsobjekten ein gültiges Eichmaß (Benchmark) zur Bewertung bestimmt. Forschungsprogramme lassen sich auf diese Weise mit ähnlich ausgerichteten und budgetierten Forschungsprogrammen vergleichen; Universitäten, Fakultäten oder außer-universitäre Forschungsinstitute werden im Hinblick auf thematische Schwerpunkte oder Größe verglichen.

Die Methode verlangt die Verwendung von Untersuchungseinheiten, die in möglichst vielen Hinsichten vergleichbar sind – bis auf die zu evaluierende Inputgröße, z.B. die Teilnahme an einem Forschungsförderungsprogramm. Die Leistungsfähigkeit (Produktivität, Wirkung) solcher Vergleichseinrichtungen dient als „quasi-natürlicher Nullpunkt“ für jene Untersuchungseinheiten, für die man evaluieren möchte, welchen Einfluss bestimmte Inputgrößen oder Forschungsprogramme haben oder hatten. In diesem Sinne sind Vergleichs- oder Kontrolleinheiten als methodischer Versuch zu werten, das Fehlen echter experimenteller Designs bei der Suche nach realen Kausal- und Wirkungsmechanismen zumindest ansatzweise auszugleichen. Ob das „Nachstellen“ einer experimentellen Anordnung gelingt, hängt im Wesentlichen von der Verfügbarkeit geeigneter Einheiten ab.

*Wissenschaftsinterne und -externe Evaluation:* Evaluationsverfahren können auch nach Verfahrensaspekten betrachtet werden. Fertigen die zu evaluierenden Untersuchungseinheiten (Organisationen, Forschungsprogramme, vgl. Tabelle 3) Beschreibungen über ihre Forschungsarbeit an, dann ist dies ein interner Vorgang, der zwar nach außen kommuniziert werden kann, aber in der Regie der zu evaluierenden Einrichtung bleibt. Evaluationssubjekt und Evaluationsobjekt sind identisch. Wird dagegen eine Leistungsbeschreibung durch Akteure angefertigt, die zur Umwelt der zu evaluierenden Einheit gehören, hat man es mit Fremdbeschreibungen zu tun. Werden diese Evaluationen von anderen Wissenschaftseinrichtungen in Gang gesetzt, sind solche Fremdbeschreibungen wissenschaftsintern, z.B. im Sinne von empirischer Wissenschaftsforschung. Diese Forschung ist selbst keine Evaluation im engeren Sinne, sondern erstellt evaluative Informationen als ersten Schritt eines Prozesses, bei dem die

Identifikation von Strukturen und die Bedingungen der Produktivität sowie Wirkungsbreite wissenschaftlicher Arbeit im Vordergrund steht. Im Gegensatz zur empirischen Wissenschaftsforschung sind von Akteuren des politisch-administrativen Systems in Gang gesetzte Fremdbeschreibungen als wissenschaftsextern zu bezeichnen. Die seit mehreren Jahren laufende Evaluation der Hochschulforschung in Niedersachsen ist in diesem Sinne wissenschaftsextern, weil sie trotz der Beteiligung von Fachkollegen aus anderen Bundesländern und Institutionen zur politischen Entscheidungsfindung verwendet wird.<sup>13</sup>

#### V. Zwecke der Evaluation

In der Literatur besteht Einigkeit darüber, dass die Evaluation von Forschungsleistungen mit erheblichem, derzeit eher wachsendem Zeit- und Geldaufwand verbunden und schon deshalb kein bloßer Selbstzweck sein kann, sondern einer legitimen Begründung bedarf. Über diese Begründung müssen sich Evaluatoren im Klaren sein und sie auch gegenüber den zu Evaluierenden transparent machen.<sup>14</sup> Welche Zwecke können dies sein?

*Qualitätssicherung:* Es liegt zunächst im Interesse der Wissenschaft selbst, die Produktion wissenschaftlicher Arbeiten von hoher Qualität sicherzustellen und Bewertungsverfahren zur qualitativen Normung von Forschungsleistungen durchzuführen. Dieses Interesse lässt sich aus forschungslogischer Sicht mit der Verfügbarkeit kreativer, origineller und bestehende Wissensbestände übersteigender Einsichtsgewinne begründen, die allen am Forschungsprozess Beteiligten zugute kommen. Qualitativ nicht nur gute, sondern exzellente Forschung kann als Ideal der Wissenschaftsgemeinschaft schlechthin gelten. Außerdem erzeugt Forschung, deren Qualität einer beständigen Überprüfung unterliegt, auch außerhalb der Wissenschaft hohe Glaubwürdigkeit und Legitimität. In Zeiten knapp bemessener öffentlicher Haushalte wird es aus dieser Perspektive für den staatlichen Fördergeber schwieriger Kürzungen für exzellente Forschungseinheiten zu rechtfertigen.

*Gesellschaftlicher Nutzen:* Forschung betreibende Einrichtungen stehen einer vielgestaltigen gesellschaftlichen Umwelt gegenüber, die „For-

---

<sup>13</sup> *Wissenschaftliche Kommission Niedersachsen* (Hrsg.), aaO., 1999; *Wissenschaftliche Kommission Niedersachsen* (Hrsg.), aaO., 2001.

<sup>14</sup> H.-D. Daniel, *Wissenschaftsevaluation. Neuere Entwicklungen und heutiger Stand der Forschungs- und Hochschulevaluation in ausgewählten Ländern*, Bern: CEST, 2001, 49; auch R. Fisch/H.-D. Daniel (Hrsg.), aaO., 1986, 11.

schungsleistung“ in unterschiedlicher Weise beurteilt. Unternehmen etwa werden sich vorzugsweise für diejenigen Forschungsleistungen interessieren, die im wirtschaftlichen Sinne verwertbar sind. Eine unter dieser Prämisse durchgeführte Evaluation wird vorzugsweise darauf abstellen, wie hoch der Anteil wirtschaftlich anschlussfähigen Wissens an der gesamten Wissensproduktion einer Forschungseinrichtung ist und wie gut es sich tatsächlich für wirtschaftliche Verwertungsprozesse verwenden lässt. Insbesondere für den Fall wissenschaftsbasierter Technologien (z.B. Biotechnologie, Nanotechnologie) hat diese Perspektive in den letzten Jahren an Bedeutung gewonnen. Das Wissenschaftssystem produziert jedoch nicht nur Wissen für die Wirtschaft: Die Entwicklung des „genetischen Fingerabdrucks“ oder moderne „Breitbandkommunikation“ beruhen auf Forschungsleistungen des Wissenschaftssystems für andere gesellschaftliche Bereiche, namentlich Rechtssystem und Massenmedien.

*Effizienz- und Effektivitätskontrolle:* Akteure der Forschungs- und Technologiepolitik haben ein vitales Interesse an der Klärung der Frage, wie effizient und effektiv aus Steuergeldern finanzierte Forschungstätigkeiten durchgeführt werden. Damit werden Forschungserträge zum einen in Relation zu den eingesetzten Ressourcen wie z.B. Personal, Infrastruktur, Fördermittel (Effizienz) und zum anderen in Relation zu vereinbarten Zielvorstellungen (Effektivität) gesetzt. Politische Akteure wollen Forschungsförderung auf ihren Zielerreichungsgrad hin überprüfen, um den Erfolg der Ressourcenallokation zu dokumentieren und politisches Handeln zu legitimieren.

*Institutioneller und organisationaler Wandel:* In der Perspektive der empirischen Wissenschaftsforschung interessiert Forschungsleistung als kreatives, neues Wissen, mit dem innerwissenschaftlich definierte Probleme besser als bisher behandelt bzw. gelöst werden können. Der Zweck von Evaluation wird folglich darin bestehen, nach den Bedingungen produktiver Forschung zu fragen, d.h. Forschungshypothesen aufzustellen und mit evaluativer Informationen zu überprüfen, warum bestimmte Forscher oder wissenschaftliche Einrichtungen produktiver sind als andere. So gewonnenes Wissen kann auch genutzt werden, um bestehende institutionelle Strukturen zu verändern. In diesem Zusammenhang ist die Beschreibung und Bewertung von Forschungsleistungen für einzelne Forschungseinrichtungen von Interesse. Je nach der spezifischen „Mission“ einer Einrichtung (z.B. Grundlagenforschung; angewandte Forschung und Innovation; problemorientierte Forschung) helfen evaluative Informationen zur Produktivität, Arbeitsweise und wissenschaftlichen Reputation dabei, organisationsinterne Stärken und Schwächen zu identifizieren und gegebenenfalls steuernd zu reagieren. In diesem Sinne wird Evaluation von

„innen“ her begründet und kann als „Wissenschaftsmanagement“ professionalisiert werden.

## *VI. Grenzen der Evaluation*

Die Evaluation von Forschungsleistungen ist kein einfaches Unterfangen und sondern stößt auf vielfältige Schwierigkeiten. Ein grundsätzliches Problem besteht in der Akzeptanz der Evaluation als Verfahren schlechthin. Insbesondere über den klassischen Peer Review hinaus gehende Evaluationen werden von vielen Akteuren des Wissenschafts- und Forschungssystems als Eingriff in das grundgesetzlich verbrieft Selbstbestimmungsrecht wissenschaftlicher Tätigkeit wahrgenommen. Diese Skepsis gegenüber Evaluation kann in offene Ablehnung umschlagen, wenn evaluative Informationen zur Entscheidungsfindung bei der Allokation von Forschungsmitteln herangezogen werden sollen. Es bedarf wenig Phantasie, um sich vorzustellen, dass Widerstände oder Blockaden gegen Evaluation auch deren Ergebnisse verzerren oder sogar unmöglich machen können. Sofern Evaluation ihren Zweck erfüllen soll, bedarf es der partnerschaftlichen und transparenten Zusammenarbeit von Evaluatoren und zu Evaluierenden.

Ein zweites Problem, das eng mit dem ersten verwandt ist, liegt bei den verwendeten Indikatoren. Die vorgestellten Indikatoren lassen sich nicht umstandslos in verschiedenen Fachdisziplinen und Forschungsfeldern anwenden. So bestehen zwischen Fachdisziplinen „kulturelle“ Unterschiede im Publikations- und Zitierverhalten, deren Nichtbeachtung zu Verzerrungen der Interpretationen führen kann. Die bereits erwähnten unterschiedlichen Publikationskulturen bei Geisteswissenschaftlern und Naturwissenschaftlern oder auch nationale Wissenschaftsstile, die mehr oder weniger stark normieren, was gute Wissenschaft ist<sup>15</sup>, fallen in diesen Problembereich. Bloße zahlenmäßige Vergleiche der Publikationstätigkeit entlang der vorgestellten Aggregatebenen (Einzelwissenschaftler, Institut, Organisation etc.) würde ohne den systematischen Einbezug qualitativer Aspekte zu kurz greifen. Bei allen notwendigen Bemühungen um objektivierende Verfahren und geeignete Indikatoren muss davor gewarnt werden, quantitative Indikatoren allein als hinreichend für die Evaluation von

---

<sup>15</sup> Vgl. *J. Galtung*, Structure, Culture and Intellectual Style. An Essay Comparing Saxonic, Teutonic, Gallic and Nipponic Approaches, in: *Social Science Information*, 20 (1981) 6.

Forschungsleistungen zu betrachten.<sup>16</sup> Der verständliche Wunsch nach einem standardisiert anwendbaren „Indikatoren-Werkzeugkasten“ ist bei der Tendenz zur Verfolgung komplexer politischer Zielsetzungen nicht erfüllbar<sup>17</sup>, auch wenn immer wieder Versuche in dieser Hinsicht unternommen werden.<sup>18</sup> Evaluation von Forschungsleistungen sollte somit „intelligent“ und nicht „mechanisch“ mit dem skizzierten Set an Indikatoren umgehen, um zu sachgerechten Urteilen zu gelangen.

Ein dritter Problemkomplex der Evaluation von Forschungsleistungen bezieht sich vor allem auf Programme der Forschungs- und Technologiepolitik:<sup>19</sup>

- Oftmals werden Evaluationskriterien verwendet, welche die Tatsache vernachlässigen, dass Forschungsprogramme multiple, konfligierende Ziele und im Prozess sich entwickelnde Dynamiken aufweisen.
- Oftmals wird die Perspektive der politisch-administrativen Förderer (oder staatlicher Aufsichtsorgane) eingenommen, jedoch die Sicht der „Betroffenen“ vernachlässigt. Evaluationsergebnisse werden häufig frühzeitig für politische Entscheidungen benötigt, manchmal sogar bevor die Forschungsarbeiten abgeschlossen sind und normalerweise bevor sich die gesamte Breite möglicher Wirkungen entfaltet haben kann.
- Fördermaßnahmen werden oftmals evaluiert, ohne dass ihr Entstehungskontext hinreichend verstanden worden ist („black box problem“). Sie hat dann den Ruf eines „Killers“, wenn sie den (teilweisen) Misserfolg einer Fördermaßnahme dokumentiert, ohne die möglichen Ursachen zu thematisieren.
- Die zeitliche und sachliche Zurechenbarkeit von Input- und Outputs gestaltet sich als schwierig.

Die offensichtlichen Grenzen der Leistungsfähigkeit von Evaluationsverfahren lassen sich nur kompensieren, wenn ihr Verwendungskontext be-

<sup>16</sup> R. Barré, Sense and Nonsense of S&T Productivity Indicators, Science and Public Policy, 2001, vol. 28 (4), 1–8.

<sup>17</sup> A. Airaghi/N. E. Busch/L. Georgiou/S. Kuhlmann/M. J. Ledoux/A. F. J. van Raan/J. Viana Baptista, Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes, Report to the European Commission, DG XII, Evaluation Unit, Brussels/Luxembourg: Office for Official Publications of the European Communities, 1999.

<sup>18</sup> Z.B. European Commission (Hrsg.), RTD Evaluation Toolbox, a report prepared by the Thematic Network „Socio-Economic Evaluation of Public RTD Policies (EPUB)“, 2002.

<sup>19</sup> Vgl. D. Schön/M. Rein, Frame Reflection. Towards the Resolution of Intractable Policy Controversies, New York: BasicBooks, 1994, 12 f.; auch S. Kuhlmann, aaO., 2003.

wusst gehalten und ihre Ergebnisse mit Vorsicht bewertet werden:<sup>20</sup> (1) Evaluatoren und Evaluationsnutzer müssen sich klar darüber werden, was sie eigentlich wissen wollen, welche Fragen in welcher Breite, Tiefe und Radikalität zu stellen sind. Dabei bleibt die analytische Reichweite von Evaluationsstudien immer beschränkt: Wenn man einen prinzipiell unbegrenzten Wirkungsraum annimmt, in den zugleich immer auch andere Wirkkräfte strahlen, dann bereitet die Rekonstruktion von Wirkungszusammenhängen umso mehr Probleme, desto umfassender und allgemeiner die Input- und Outputgrößen definiert werden. (2) Die Evaluationsforschung im Bereich der Forschungs- und Innovationspolitik geht heute davon aus, dass „realistische“ Informationen über ein Evaluationsobjekt nur durch die kombinierte Verwendung verschiedener sozialwissenschaftlicher Methoden und Indikatoren erarbeitet werden können.

Ein viertes Problem liegt darin, geeignete Evaluatoren zu finden. Insbesondere wenn es um externe Evaluation geht, ist es unerlässlich, intermediäre Organisationen zu finden, die von den Akteuren des Wissenschafts- und Forschungssystems akzeptiert sind und sich nicht politisch instrumentalisieren zu lassen.<sup>21</sup> Auf Evaluation gründende Analysen schließlich verlangen ausgewiesenes Expertentum im Bereich der Wissenschaftsindikation und der Evaluationsverfahren.

In der zweiten, zu einem späteren Zeitpunkt folgenden Veröffentlichung wird argumentiert, dass gegenwärtig eine Vielzahl von Einrichtungen evaluative Informationen in sehr unterschiedlichen Formaten, Rhythmen und Qualitäten produziert. Dies steht einem wachsenden Bedarf – z.B. von den Verwaltungen der großen Forschungsträgerorganisationen, den Hochschulen und ihren Verbänden oder dem Wissenschaftsrat – nach verlässlichen, regelmäßigen und vergleichbaren Informationen entgegen. Es wird die Frage diskutiert, ob es sinnvoll ist eine entsprechende Dienstleistungseinrichtung ins Leben zu rufen, und wie eine solche Einrichtung gestaltet werden sollte.

---

<sup>20</sup> Vgl. S. Kuhlmann, Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik, Baden-Baden (Nomos) 1998, 97–106.

<sup>21</sup> H.-D. Daniel, aaO., 2001, 13 f.