

Stefan Kuhlmann und Thomas Heinze

Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf

Teil II: Produktion und Verwendung evaluativer Information
sowie Möglichkeiten ihrer künftigen Organisation

I. Einleitung

Die Evaluation von Forschungsleistungen gewinnt an Bedeutung. Im ersten Teil dieses Beitrages¹ argumentierten wir, dass (1) das Interesse an Information über Leistungen und Qualitäten staatlich geförderter Wissenschaft und Forschung zunimmt, (2) die Einführung von Globalhaushalten in Verbindung mit leistungsorientierten Modi der Verteilung von Forschungsmitteln Bewertungsverfahren erfordert, (3) der wachsende Wettbewerb zwischen Forschungseinrichtungen zur Ausprägung unverwechselbarer, wettbewerbsfähiger Forschungsprofile zwingt und schließlich (4) auch die fortschreitende Europäisierung der Forschungslandschaft nach Leistungsvergleichen verlangt.

Der Text dieses Beitrages beruht auf einem Gutachten für die Geschäftsstelle der Deutschen Forschungsgemeinschaft (DFG).² Der erste Teil des Beitrages präziserte, was begrifflich und konzeptionell unter „evaluative Informationen über Forschungsleistungen“ zu verstehen ist. Wir präsentierten eine Auswahl von anerkannten Indikatoren, mit deren Hilfe evaluative Informationen generiert werden können. Zur Sprache kamen auch Methoden und Zwecke der Evaluation sowie die Grenzen solcher Verfahren.

¹ S. Kuhlmann/Th. Heinze, Evaluation von Forschungsleistungen in Deutschland: Erzeuger und Bedarf; Teil I: Konzeptionelle Grundlagen. In: *Wissenschaftsrecht*, 37 (2004), 53–69.

² 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>).

Der vorliegende zweite Teil beginnt mit einer kurzen Übersicht der Evaluationspraxis in Deutschland (Abschnitt II) und skizziert dann explorativ, wie und von wem evaluative Informationen von Forschungsleistungen bisher erbracht (Abschnitt III) und nachgefragt werden (Abschnitt IV). Außerdem diskutieren wir die Frage (Abschnitt V), ob eine Dienstleistungseinrichtung geschaffen werden sollte, die Informationen zu Forschungsleistungen systematisch sammelt, dokumentiert und analysiert. Vorgeschlagen wird die Errichtung einer „Gemeinschaftsstelle Clearinghouse“ für Forschungsstrukturinformation.

II. Evaluationspraxis in Deutschland

In historischer Perspektive kann man in Deutschland wie auch in anderen industrialisierten Ländern³ wissenschaftsinterne und -externe Ansätze als zwei heterogene Entwicklungslinien der Evaluation von Wissenschaft sowie der Förderung von Forschung und Innovation unterscheiden. Sie lassen sich als ein Drei-Schalenmodell darstellen⁴:

Erste Schale – Individuelle Forschungsleistungen: Den „Kern“ bilden peer review-Verfahren und später zusätzlich Verfahren zur Messung der Forschungsleistung einzelner Forscher und Gruppen (Bibliometrie etc.) als *wissenschaftsinterne* Instrumente für die Entscheidung über die Allo-

³ Die deutsche Evaluationspraxis im Bereich von Forschungs- und Innovationspolitik steht in enger Beziehung zu ähnlichen „Evaluationskulturen“ in anderen Industrieländern; Übersichten bei *B. Bozeman, B./J. Melkers* (Hrsg.), *Evaluating R&D Impacts: Methods and Practice*, Boston et al. 1993 (Kluwer); *L. Georghiou*, *Assessing the Framework Programmes – a meta-evaluation*, *Evaluation* Vol. 1, No.2, October 1995 pp.171–188; *OECD* (ed.), *Policy Evaluation in Innovation and Technology, Towards Best Practices*. Paris 1997; *R. Hirasawa*, *Framework on Public R&D Evaluation from the Viewpoint of New Public Management: Related to the Present Japanese Situation*, *International Conference on Socio-economic Evaluation of Public Research and Development-Perspectives*, Tokyo, 18./19. April 2002; *Ph. Shapira, S., Kublmann* (Hrsg.), *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, Cheltenham 2003 (E. Elgar). In den angelsächsischen Ländern setzte sich eine eher quantitativ orientierte Leistungsmessung durch, während in Deutschland und Frankreich eher qualitative Verfahren dominieren. In Japan verfolgt man seit Mitte der 1990er offiziell eine Kombination beider Linien, wobei die Implementation noch in den Kinderschuhen steckt.

⁴ Dieser Abschnitt beruht auf *S. Kublmann*, *Evaluation in der Forschungs- und Innovationspolitik*. In: *Stockmann, R.* (Hg.): *Evaluationsforschung*, Leverkusen 2003 (Leske+Budrich), 2., überarbeitete und aktualisierte Auflage, 287–307; sowie *S. Kublmann*, *Evaluation of research and innovation policies: a discussion of trends with examples from Germany*. In: *Int. J. Technology Management*, Vol. 26, Nos. 2/3/4, 2003, 131–149.

kation von Fördermitteln in Forschungseinrichtungen⁵. Das Peer Review-Verfahren ist das vorherrschende Evaluationsinstrument der Deutschen Forschungsgemeinschaft (DFG). Förderanträge werden von Peers bewertet, die im Vier-Jahres-Rhythmus von der gesamten wissenschaftlichen Gemeinschaft gewählt werden⁶. Jeder Gutachter ist angehalten, den Antrag allein auf der Grundlage seiner wissenschaftlichen Qualität zu beurteilen. Mittlerweile haben Senat und Hauptausschuss der DFG beschlossen, dieses bewährte System („bottom-up“) um strategierorientierte, interdisziplinäre und internationale Komponenten zu ergänzen.

Zweite Schale – Programme: Um diesen Kern der Peer Review-Verfahren legt sich eine Schale von Evaluationsstudien im Sinne von *impact analyses* forschungs- und innovationspolitischer Programme, die im Unterschied zur institutionellen Förderung gezielte Gestaltungsabsichten verfolgen. In Deutschland zeigte sich insbesondere das Bundesforschungsministerium aktiv; zwischen 1985 und 1993 ließ man etwa 50 größere Evaluationsstudien anfertigen⁷, in den folgenden Jahren nicht weniger. Solche Studien schließen an die amerikanische Wirkungsforschung und *policy analysis* an⁸. Dieses Evaluationskonzept kann als *wissenschafts-extern* bezeichnet werden, denn es wird überwiegend vom politisch-administrativen System in Gang gesetzt, um zu prüfen, inwieweit politisch gesetzte Ziele umgesetzt wurden. Als Evaluatoren treten dabei üblicherweise unabhängige Forschungsinstitute im Auftrag der Administration auf; sie wenden ein relativ breites Spektrum von Evaluationskonzepten und -instrumenten an⁹ und sind seit den 1990er Jahren auch professionell in einer „Deutschen Gesellschaft für Evaluation (DeGEval)“ organisiert.

⁵ D. Guston, The expanding role of peer review processes in the United States. In: Shapira, P./Kuhlmann, S.; S. Hornbostel, Wissenschaftsindikatoren. Bewertungen in der Wissenschaft, Opladen 1997 (Westdeutscher Vlg.); H.-D. Daniel, Guardians of Science. Fairness and Reliability of Peer Review, Weinheim u.a. 1993 (VCH); A. van Raan (ed.), Handbook of Quantitative Studies of Science and Technology, Amsterdam 1988 (Elsevier).

⁶ F. Neidhardt, Selbststeuerung in der Forschungsförderung. Das Gutachterwesen der DFG, Opladen 1988 (Westdeutscher Vlg.).

⁷ S. Kuhlmann, /D. Holland, Evaluation von Technologiepolitik in Deutschland – Konzepte, Anwendung, Perspektiven, Heidelberg 1995 (Physica/Springer).

⁸ „Policy-analytic movement“; D. Schön, /M. Rein, Frame Reflection. Toward the Resolution of Intractable Policy Controversies, New York 1994 (BasicBooks); Übersichten bei Th. Widmer, Meta-Evaluation. Kriterien zur Bewertung von Evaluationen, Bern et al. 1996 (Haupt); M. Q. Patton, Utilization-Focused Evaluation. The New Century Text, Thousand Oaks et al. 1997 (Sage); S. Kuhlmann, Politikmoderation. Evaluationsverfahren in der Forschungs- und Technologiepolitik, Baden-Baden 1998 (Nomos).

⁹ Shapira/Kuhlmann, aaO.

Dritte Schale – Institutionen: Hier geht es um die Leistungsfähigkeit von Forschungsinstitutionen oder größeren Wissenschaftsgebieten, wie sie u.a. der Wissenschaftsrat (WR) durchführt. Nach der deutschen Vereinigung nahmen die WR-Evaluationen gestaltenden Einfluss auf die neue Struktur der Forschungslandschaft Ostdeutschlands¹⁰. Seit den 1990er Jahren wurden institutionelle Evaluationen immer häufiger durchgeführt¹¹. Hierzu zählen auch die „Systemevaluations“ aller großen Forschungseinrichtungen in Deutschland im Zeitraum 1998 und 2002. Die Leitfrage dieser Systemevaluations war, ob und in welcher Weise die evaluierten Einrichtungen einen wesentlichen Beitrag zur Produktion von Wissen und Innovation leisten. Bei der Evaluation von Forschungsinstitutionen werden zunehmend wissenschaftsinterne und -externe Fragestellungen und Informationsquellen kombiniert. Parallel dazu widmet auch die Privatwirtschaft der Effizienz, Effektivität und strategischen Ausrichtung ihrer Laboratorien erhöhte Aufmerksamkeit¹².

Zusammenfassend: Die deutsche Evaluationspraxis ist einerseits *entwickelt* hinsichtlich des hohen Maßes an Selbstorganisation bei wissenschaftsinternen Verfahren, die Konsens und Verpflichtung zwischen den Forschern fördern, sowie *dynamisch* wegen vielfältiger neuer Evaluationsansätze und Strategieentwicklungen in allen Institutionen. Sie ist jedoch auch *unsystematisch* und *zersplittert*, weil die verschiedenen Evaluationsbemühungen bisher nur wenig aufeinander abgestimmt werden konnten.

Weiterhin ist festzustellen, dass sich viele Evaluationsinitiativen (interne und externe Ansätze) nicht mehr allein auf die Feststellung der Qualität einzelner Forschungsleistungen beschränken. Zusehends soll evaluative Information auch zur Erhellung *strukturbildender* sowie *sozio-ökonomischer Effekte* von Forschung und Forschungsförderung beitragen. Ein gemeinsames Raster von Indikatoren zur Erfassung und Analyse solcher Effekte wurde bisher jedoch noch nicht vereinbart.

¹⁰ H.-J. Block/W. Krull, What are the consequences? Reflections on the impact of evaluations conducted by a science policy advisory body, in: *Scientometrics*, Vol. 19, 1990, Nos. 5–6, 427–437.

¹¹ S. Kuhlmann/D. Holland, Erfolgsfaktoren der wirtschaftsnahen Forschung, Heidelberg 1995 (Physica/Springer).

¹² Z.B. K. Brockhoff, Forschung und Entwicklung. Planung und Kontrolle, München/Wien 1999 (R. Oldenbourg), 5. erg. u. erw. Aufl.; H.D. Bürgel/C. Haller/M. Binder, F & E-Management, München 1996 (Vahlen).

III. Erzeugung evaluativer Information

Im ersten Teil unseres Beitrages wurden Indikatoren zur Beschreibung von Forschungsleistungen vorgestellt (siehe dort die Tabellen 1, 2 und 3). Im Folgenden skizzieren wir explorativ die Produktion entsprechender Information, strukturiert nach (1) Investitionen in Forschung (Input-Maße als Referenzgrößen zur Bewertung von Forschungsleistung) sowie (2) Produktivitäts- bzw. Wirkungsinformationen¹³. Evaluative Informationen werden auf unterschiedlichen Aggregationsebenen erzeugt. Wie im ersten Teil dargestellt (dort Tabelle 4), können vertikal die Ebenen von Organisationen, Instituten oder Arbeitsgruppen und Einzelwissenschaftlern sowie horizontal die institutionelle „Landschaft“, innerhalb welcher Forschung stattfindet, Forschungsprogramme, Fachdisziplinen und Forschungsfelder unterschieden werden.

Input-Information

Auf den Aggregationsebenen *Programm und Institutionen-“Landschaft“* finden wir eine Reihe regelmäßiger Publikationen¹⁴ sowie datenbankgestützter Informationssysteme¹⁵. Insbesondere die der ministeriellen und der amtlichen Statistik zugehörigen Berichte enthalten hoch aggregierte

¹³ Die folgenden Angaben erheben nicht den Anspruch auf Vollständigkeit; die Informationen wurden 2001/02 erhoben.

¹⁴ Z.B. *Bundesministerium Bildung und Forschung (BMBF)*, Bundesbericht Forschung 2000, Bonn; *BMBF* (Hrsg.), Grund- und Strukturdaten 2000/2001, Bonn 2001; *Statistisches Bundesamt (StBA)* (Hrsg.), Wissenschafts- und Technologieindikatoren, Wiesbaden 2001, *StBA* (Hrsg.), Zur finanziellen Lage der Hochschulen, in: *Wirtschaft und Statistik* 12/2001, Wiesbaden, S. 1021–1026; Abschlussberichte zu den Systemevaluationen: *Wissenschaftsrat (WR)* (Hrsg.), Systemevaluation der Blauen Liste – Stellungnahme des Wissenschaftsrates zum Abschluss der Bewertung der Einrichtungen der Blauen Liste, Köln, 2000; *WR* (Hrsg.), Systemevaluation der HGF – Stellungnahme des Wissenschaftsrates zur Hermann von Helmholtz-Gemeinschaft Deutscher Forschungszentren, Köln 2001; *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; *Evaluierungskommission* (Hrsg.), Systemevaluierung der Fraunhofer-Gesellschaft. Bericht der Evaluierungskommission, München 1998; Jahresberichte von Forschungseinrichtungen und übergeordneten Berichtsstellen der großen Forschungseinrichtungen wie z.B. der Max-Planck-Gesellschaft und der Fraunhofer-Gesellschaft.

¹⁵ BMBF und Deutschen Forschungsgemeinschaft (DFG), die Informationen zu Forschungsausgaben nach ihrer Aufgabenstruktur und -entwicklung in Deutschland (DASTAT), hochschulspezifische Datenmengen (ICEbmbf) sowie Angaben zu DFG-geförderten Programmen bereit halten (GEPRIS).

Beschreibungen der finanziellen und personellen Ausstattung der institutionellen Teilbereiche der deutschen Forschungslandschaft (Hochschulen, außeruniversitäre Forschung, Wirtschaftsforschung), aufgeschlüsselt nach regionalen (Bundesländer), disziplinären und sektoralen Kriterien.

Hochschulbezogene Statistiken weisen spezifische Abgrenzungsprobleme bei Forschung und universitärer Lehre auf. Die vorhandene Berichterstattung macht nur in Ansätzen Strukturen sichtbar. Zwar dokumentiert das Statistische Bundesamt relativ ausführlich Strukturkennziffern zur Ausstattung der Hochschulen, beispielsweise Grundmittel je Student, Drittmittel je Professorenstelle oder Ausgaben der Hochschulen nach Fächergruppen¹⁶; doch erfährt man nichts über den Anteil promovierter Wissenschaftler in Hochschulen, an außeruniversitären Einrichtungen oder Forschungseinrichtungen der Wirtschaft, ebenso wenig über deren Alterstruktur insgesamt oder aufgeschlüsselt nach Sektionen bzw. Fachdisziplinen. Die bisherige Aufschlüsselung der Personalstruktur nach Besoldungs- und Statusgruppen trägt die Handschrift einer für ministerielle Bedarfe zugeschnittenen Datenerhebungspraxis.

Für die *Aggregationsebene Fachdisziplinen und Forschungsfelder* (auf organisationaler und departmentaler Ebene) liegen Querschnittsbegutachtungen des Wissenschaftsrates in den Feldern Umwelt-, Material- und Energieforschung vor.¹⁷ Hier wurden durch Selbstbeschreibungen (schriftliche Befragung der Einrichtungen) und Peer Reviews Informationen zur personellen und materiellen Ausstattung der untersuchten Forschungseinrichtungen erhoben. Weiterhin weist die DFG die seit den 1990er Jahren bewilligten Mittelflüsse an Einrichtungen der universitären und außeruniversitären Forschung aus, gegliedert nach 16 Fachgebieten. Deutlich wird dadurch, an welchen Orten DFG-finanzierte Forschung fachlich ihre Schwerpunkte findet – sowohl mit Blick auf einzelne Einrichtungen, als auch mit Blick auf Förderregionen, die in der Dokumentation kartographisch dargestellt sind¹⁸. Weiterhin hat Niedersachsen seit März 1999 mit der systematischen Evaluation der universitären Fachdisziplinen begonnen, die auf Fakultäts- bzw. Fachbereichsebene evaluative Input-Informationen erheben. Diese von der Wissenschaftlichen Kommission Niedersachsen durchgeführten Evaluationen liegen für die Fächer Chemie, Geschichte, Biologie, Bauingenieurwesen/Architektur, Elektrotechnik, Maschinenbau,

¹⁶ *StBA*, Zur finanziellen Lage der Hochschulen, aaO.

¹⁷ *WR* (Hrsg.), Stellungnahme zur Umweltforschung in Deutschland, Bd.1 und 2, Köln 1994; *WR* (Hrsg.), Stellungnahme zur außeruniversitären Materialwissenschaft, Köln 1996; *WR* (Hrsg.), Stellungnahme zur Energieforschung, Köln 1999.

¹⁸ *Deutsche Forschungsgemeinschaft: Bewilligungen an Hochschulen 1991 bis 1995*, Bonn 1997, sowie Folgepublikationen.

Wirtschaftswissenschaften, Informatik und Berufswissenschaften bereits vor. Bemerkenswerte Ansätze in dieser Richtung werden mittlerweile auch in Baden-Württemberg und Bayern verfolgt. Eine bundesweite Übersicht der universitären Fachdisziplinen ist derzeit nicht verfügbar.

Einen ersten Schritt in diese Richtung geht die Studie *Forschungslandkarte Hochschulmedizin*, die vom Fraunhofer-Institut für Systemtechnik und Innovationsforschung im Auftrag des BMBF und des Medizinischen Fakultätentages durchgeführt wurde¹⁹. Neben dem Schwerpunkt der Untersuchung – Forschungsschwerpunkte der Medizinischen Fakultäten in Deutschland – wurden auch zahlreiche Daten zum Forschungsinput, z.B. Anzahl der Professuren, wissenschaftliche Vollzeitkräfte auf Planstellen sowie Höhe des Landeszuschusses für Forschung und Lehre über alle medizinischen Fakultäten Deutschlands erhoben.

Daten über *Universitäten* und ihre Untereinheiten werden in der Regel von den Wissenschaftsministerien der Bundesländer erfasst. In Baden-Württemberg stellt der FuE-Monitor des dortigen Statistischen Landesamtes Finanzdaten für jede baden-württembergische Universität zusammen²⁰. So sind die Forschungsausgaben nach Herkunft (z.B. Grund- und Drittmittel) und Fachgebieten ausgewiesen und in Relation zur Zahl der Studierenden gestellt. Allerdings fehlen weitgehend Strukturmaße, die Profilunterschiede zwischen einzelnen Universitäten ausweisen würden. Ähnlich liegt es mit dem Bericht des Wissenschaftsrates zur Drittmittelsituation der Hochschulen²¹. Hier werden relativ ausführlich die eingeworbenen Drittmittel strukturell untersucht, etwa auf Lehrstuhl- oder An-Instituts-Ebene bzw. im regionalen Vergleich. Der Bericht kritisiert zu Recht, dass die für solche Analysen erforderlichen Strukturinformationen bisher nur schwer oder gar nicht erhältlich sind. Einen wichtigen Fortschritt erzielt hier das erwähnte DFG-Ranking, das die bewilligten Forschungsmittel nicht nur nach Regionen und Fachgebieten, sondern auch nach Empfängerorganisationen (Hochschulen; andere) differenziert ausweist.

Auf der *Aggregationsebene einzelner Forschungsträgerorganisationen* und ihrer Einzelinstitute bzw. Forschungszentren ist die Datenbasis vergleichsweise heterogen. Relativ gut beschrieben sind die außeruniversitä-

¹⁹ BMBF und Medizinischer Fakultätentag (Hrsg.), *Forschungslandkarte Hochschulmedizin*. (Bearbeitung: Susanne Bühner, Fraunhofer ISI). Stuttgart 2002 (Fraunhofer IRB-Verlag).

²⁰ *Statistisches Landesamt Baden-Württemberg*, (Hrsg.), *FuE-Monitor Baden-Württemberg*, Stuttgart 2001.

²¹ *WR* (Hrsg.), *Drittmittel und Grundmittel der Hochschulen 1993–1998*, Köln 2000.

ren Einzelinstitute, die einerseits regelmäßige, zumeist jährliche Selbstbeschreibungen in Form von Jahresberichten anfertigen, und von welchen andererseits im Rahmen der Systemevaluationen Fremdbeschreibungen angefertigt worden sind. Im Einzelnen ergibt sich folgendes Bild.

Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL).²² Im Rahmen der Systemevaluation der WGL wurden alle Einzelinstitute hinsichtlich ihrer Personal- und Finanzausstattung beschrieben. Die Angaben sind quantitativ gut, gehen jedoch nicht über eine zaghafte qualitative Beschreibung hinaus. So werden zwar Drittmittelquoten der Gesamtfinanzierung und das disziplinäre Portfolio ausgewiesen. Es fehlen jedoch relevante Strukturmaße wie Alterstruktur des Wissenschaftspersonals oder Infrastrukturmaße. Dies änderte sich allerdings: Der Senat der WGL hat 2002 Grundsätze zu Aufgaben und Verfahren der Evaluation der WGL-Institute verabschiedet, welche eine interne Evaluation durch Beiräte und eine externe Evaluation durch einen Senatsausschuss Evaluierung vorsieht.

Helmholtz-Gemeinschaft deutscher Forschungszentren (HGF).²³ Im Rahmen der Systemevaluation der HGF wurden alle Zentren hinsichtlich ihrer Personal- und Finanzausstattung beschrieben. Die Angaben sind quantitativ und qualitativ deutlich präziser und zudem umfangreicher als bei der WGL. So werden insbesondere Strukturmaße systematisch erhoben; Finanz- und Personaleinsatz werden zugleich zentren- und themenbezogen aufgeschlüsselt; die Personalstruktur wird ausführlich differenziert (Doktoranden, Diplomanden, Gastwissenschaftler, Frauenanteil) sowie deren Entwicklungsdynamik in einem Fünfjahresfenster dargestellt, insbesondere mit Blick auf Berufungs- und Qualifikationsentwicklungen. Zu beachten ist allerdings, dass ein Zentrum der HGF nicht vergleichbar ist mit einzelnen Instituten der WGL, FhG oder MPG. HGF-Zentren sind selbst Institutsverbände und erheblich größer als einzelne Institute der anderen Forschungsorganisationen.

Fraunhofer-Gesellschaft (FhG).²⁴ Die von der FhG-Zentrale publizierten Gesamtjahres- und Finanzberichte enthalten nach Instituten aufgeschlüsselte Finanzdaten. Dieses Bild spiegelt sich auch in den Jahresbe-

²² Grundlage der folgenden Ausführungen sind Dokumentrecherchen der Jahresberichte von fünf WGL-Instituten sowie der Abschlussbericht des Wissenschaftsrates zur Systemevaluation der WGL.

²³ Grundlage der folgenden Ausführungen sind Dokumentrecherchen der Jahresberichte von fünf HGF-Zentren sowie der Abschlussbericht des Wissenschaftsrates zur Systemevaluation der HGF.

²⁴ Grundlage der folgenden Ausführungen sind Dokumentrecherchen der Jahresberichte von fünf FhG-Instituten sowie der Abschlussbericht des Wissenschaftsrates zur Systemevaluation der FhG.

richten der einzelnen Institute wider, die in vergleichsweise homogener Form Finanz Eckdaten bereithalten. Personaldaten werden in den Jahresberichten der gesamten FhG nur überblickshaft und nicht institutsbezogen ausgewiesen. Solche Informationen sind dagegen in den einzelnen Institutsdokumentationen in quantitativer Hinsicht gut dokumentiert. Insgesamt fällt auf, dass weiterführende qualitative und strukturelle Dimensionen fehlen.

*Max-Planck-Gesellschaft (MPG)*²⁵: Die MPG veröffentlicht keine Gesamtberichte, welche Input-Größen institutsbezogenen ausweisen. Die allgemeinen Informationen über Personal- und Finanzgrößen der MPG, die sich aus den institutsübergreifenden Jahresberichten ableiten lassen, werden in den Jahresberichten der Einzelinstitute in quantitativer Hinsicht nur wenig disaggregiert. Allerdings finden sich – im Unterschied zur FhG – qualitative Informationen, z.B. der Anteil promovierter Wissenschaftler oder nach Disziplinen aufgeschlüsselte Beschäftigtenkategorien (Wissenschaftler, Nicht-Wissenschaftler, Studentische Hilfskräfte usw.). Infrastrukturinformationen sind dagegen nur in aggregierter Form erhältlich und werden nur zum Teil in den Jahresberichten der Einzelinstitute ausgewiesen.

Output- und Wirkungsinformation

Wir haben außerdem untersucht, welche Produktivitäts- bzw. Wirkungsinformationen gegenwärtig in Deutschland auf unterschiedlichen Aggregationsebenen verfügbar sind. Dabei ergibt sich folgendes Bild.

Für die *Aggregationsebene Institutionen-“Landschaft“* ist die Situation bei Leistungsindikatoren derjenigen der Input-Informationen vergleichbar. Die meisten der oben erwähnten Publikationen – und darüber hinaus der Bericht zur Leistungsfähigkeit Deutschlands und die Patentstatistiken des Deutschen Patent- und Markenamts²⁶ – dokumentieren neben Input-Parametern auch Output- und Wirkungs-Größen. Diese Indikatoren beschreiben in der Regel die institutionellen Teilbereiche der deutschen Forschungslandschaft (Hochschulen, außeruniversitäre Forschung, Wirtschaftsforschung) als Ganzes und schlüsseln sie nach regionalen, disziplinären und sektoralen Kriterien auf. So werden im Bun-

²⁵ Grundlage der folgenden Ausführungen sind Dokumentrecherchen der Jahresberichte von fünf MPG-Instituten sowie der Abschlussbericht der Internationalen Kommission zur Systemevaluation der DFG und der MPG.

²⁶ BMBF (Hrsg.), Zur technologischen Leistungsfähigkeit Deutschland. Zusammenfassender Endbericht 2000, Bonn 2001; Deutsches Patent- und Markenamt, DPMA, Jahresbericht 2000, München 2001.

desbericht Forschung, dem Grund- und Strukturdaten-Bericht und der Patentstatistik des DPMA durchaus detaillierte Hinweise zur Produktivität der Wissensgenese und der Produktion von Hochschulabsolventen und Promotionen geben. Weiterführende qualitative oder strukturelle Informationen – so wie sie etwa im US-amerikanischen Report on Science and Engineering Indicators zu finden sind²⁷ – stehen jedoch kaum zur Verfügung. Zu solchen Informationstypen gehören etwa das Verhältnis nationaler/internationaler Publikationen und Patente in unterschiedlichen Technologiesektoren und differenziert nach institutionellen Bereichen (Universitäten, Forschungseinrichtungen etc.).

Weiterhin fällt auf, dass das Gewicht der Dokumentation eher auf Output-Größen liegt und weniger auf Wirkungsindikatoren. Es gibt unterschiedliche Gründe, warum dies der Fall ist. Zu vermuten ist, dass Outputs leichter zu erheben und auszuwerten sind als Wirkungs-Größen, weil die sachlichen und zeitlichen Zurechenbarkeiten überschaubarer sind. Damit ist zum einen gemeint, dass die „Wirkungen“ von Forschung schwieriger zu operationalisieren und zu messen sind als „Outputs“ und daher einen erheblichen Datenerhebungsaufwand nach sich ziehen. Zum anderen gestaltet sich die Beantwortung der Frage, auf wen oder was eingetretene „Wirkungen“ zuzurechnen sind, in aller Regel schwieriger als die Dokumentation von Outputs²⁸.

Relativ gut dokumentiert ist die Patentproduktivität der beiden großen Forschungsorganisationen FhG und MPG. Die Patentstelle der Fraunhofer Gesellschaft dokumentiert in ihrem Jahresbericht relativ detailliert Erfindungsmeldungen, Patenterteilungen oder Markenmeldungen nach Technologiegebieten und institutioneller Herkunft (Universitäten, Privatpersonen etc.), und schlüsselt darüber hinaus auch die Entwicklung der Verwertungseinnahmen auf²⁹. Ebenso dokumentiert die Verwertungsgesellschaft der Max-Planck-Gesellschaft, Garching Innovation GmbH, seit den 1980er Jahren – wenngleich weniger detailliert – die Zahl der Erfindungen, Patente und Lizenzverträge der gesamten MPG.

Für die *Aggregationsebene einzelner Forschungsträgerorganisationen* und ihrer Einzelinstitute bzw. Forschungszentren, welche einzelne For-

²⁷ National Science Board (Hrsg), Science and Engineering Indicators, Washington 2000, Kap. 4, 6.

²⁸ Vgl. A. Airaghi, N. E. Georghiou, L. Kuhlmann, S. Ledoux, M. J. van Raan, A. F. J. Viana Baptista, J., Options and Limits for Assessing the Socio-Economic Impact of European RTD Programmes. Report to the European Commission, DG XII, Evaluation Unit, Brussels/Luxembourg 1999 (Office for Official Publications of the European Communities).

²⁹ Fraunhofer Patentstelle, Jahresbericht der Fraunhofer Patentstelle für die Deutsche Forschung, 2000/2001, München 2001.

schungsorganisationen im Sinne von Universitäten oder Einzelinstituten der außeruniversitären Forschungsverbände ausweist, stellt sich das Informationsangebot sehr heterogen dar. Wiederum – wie bereits bei den Input-Größen festgestellt – sind die außeruniversitären Einzelinstitute umfangreicher beschrieben als Universitäten und zwar einerseits in Form von Jahresberichten, und andererseits im Zuge der durch die Systemevaluations angefertigten Fremdbeschreibungen.

Die von der *FhG*-Zentrale publizierten Gesamtjahres- und Finanzberichte enthalten nach Instituten aufgeschlüsselte Output- und Wirkungsdaten. Relativ ausführlich sind Informationen zur Produktivität des Forschungshandelns und der Wissensprodukte vorhanden, und zwar in quantitativer und qualitativer Hinsicht. So werden Projekte auf Institutsebene sehr gut aufgeschlüsselt nach Laufzeit und Finanzierungsgeber, ebenso wie Patentanmeldungen, erteilte Patente und Lizenzeinnahmen nach Technologiesektoren dargestellt werden. Weniger umfangreich sind Daten zu Reputationsaspekten und Aspekte der Sichtbarkeit des generierten Wissens dargestellt. Insgesamt wird der Darstellung von Forschungswirkungen weniger Raum eingeräumt als der Darstellung von Forschungsoutputs.

Die *MPG* veröffentlicht keine Gesamtberichte, welche Output- und Wirkungsdaten institutsbezogenen ausweisen. Wiederum – wie bereits bei den Input-Größen festgestellt – werden entsprechende Informationen jedoch durch die Jahresberichte der Einzelinstitute vergleichsweise gut abgebildet. Zwar sind Strukturindikatoren nicht so häufig nachweisbar wie quantitative Maße. Allerdings ergibt sich der Eindruck, dass die Institute der *MPG* durchaus über ihre Produktivität aussagefähige Selbstbeschreibungen anfertigen. Dies gilt eingeschränkt auch für die Wirkungsseite.

Im Rahmen der Systemevaluation der *WGL* wurden weder die Gesamtebene, noch die Einzelinstitute hinsichtlich ihrer Forschungsoutputs und -wirkungen beschrieben, zumindest wurden solche evaluativen Informationen keiner breiteren Öffentlichkeit zugänglich gemacht. Zwar werden bei einigen Instituten Publikationen und Projekte relativ umfangreich ausgewiesen. Allerdings fällt die Beschreibung heterogen aus, so dass von einem einheitlichen Bild noch nicht die Rede sein kann.

Im Rahmen der Systemevaluation der *HGF* wurden die Zentren – im Vergleich zu den Instituten der *WGL* – relativ ausführlich beschrieben. Der Abschlussbericht hält ein Tableau von „Leistungsdaten“ bereit, das vor allem nicht-wissenschaftliche Leistungsbezüge thematisiert. So werden etwa Patent- und Lizenzdaten, Wirtschaftskooperationen und Ausgründungen nicht nur in quantitativer, sondern auch in struktureller Hinsicht für ein Dreijahreszeitfenster und zentrenbezogen dargestellt.

Auffällig ist, dass wissenschaftliche Leistungen nicht zentrenbezogen ausgewiesen werden. Zwar dokumentiert die HGF als Ganzes die Kooperationsanteile ihrer Publikationen nach Internationalität und institutioneller Herkunft, jedoch nicht auf der Ebene von Zentren oder Instituten.

Universitäten dokumentieren ihre Forschungsleistungen in sehr heterogener Weise. Einen ersten Ansatzpunkt bietet das bereits erwähnte DFG-Ranking, welches die Bewilligungsempfänger der DFG nach Hochschulen aufschlüsselt. Neben dem absoluten Gesamtvolumen wird das über die Zahl der an einer Hochschule tätigen Professoren und Hochschullehrer relativierte Volumen ausgewiesen. Interpretiert man eine DFG-Bewilligung nicht als bloßen Mittelzufluss und damit als „Input“, sondern als durch den Begutachtungsprozess zertifizierte Qualität geplanter Forschungsarbeit, dann stellt die DFG – zumindest in Näherung – eine Leistungskennziffer für allgemeine Forschungsleistung zur Verfügung³⁰. Generell gilt, dass eine für Universitäten übergreifende Dokumentation von Output- und Wirkungsindikatoren weitgehend fehlt. Zwar ist im Hochschulrahmengesetz (HRG) seit 1998 die Qualitätskontrolle von Forschung (und Lehre) rechtlich festgeschrieben worden. Allerdings erscheint dessen Umsetzung, sowohl was die Ländergesetzgebung als auch die gegenwärtig Praxis angeht, relativ uneinheitlich.

Anbieter evaluativer Information

Eine beachtliche Zahl von Einrichtungen erzeugt evaluative Informationen, jedoch in unterschiedlichen Formaten und nur begrenzt öffentlich zugänglich. Zum einen kann man die Anbieter danach unterscheiden, ob sie Informationen über sich selbst (Selbstbeschreibung) oder über andere erstellen (Fremdbeschreibung). Die Jahresberichte der Institute der MPG und FhG können als regelmäßig verfasste Selbstbeschreibungen gelten. Hier werden evaluative Informationen z.T. für Zwecke der strategischen Forschungsplanung, z.T. aufgrund bestehender Rechenschaftspflichten als Zuwendungsempfänger von Bund und Ländern erhoben, vor allem im Bereich der Input-Daten. Die Systemevaluationen der großen Forschungseinrichtungen und die genannten Einzelstudien sind dagegen Fremdbeschreibungen, bei welchen externe Akteure evaluative Informationen in Kooperation mit den und über die zu evaluierenden Einheiten erheben.

³⁰ DFG, Förder- und Forschungsmonitoring durch die DFG: Aktueller Stand und Perspektiven. In: Perspektiven der Forschung, 2002.

Eine weitere Unterscheidung betrifft den institutionellen Ort, an dem evaluative Informationen erzeugt werden. Wie Daniel und Hornbostel zeigen³¹, sind in Westeuropa und den USA Körperschaften mit Berichtspflichten an parlamentarische Gremien oder Parlamente entstanden, die außerhalb des Wissenschafts- und Forschungssystems stehen. Beispiele dafür sind die Research Assessment Exercises des Higher Education Funding Council of England (HEFCE) im Vereinigten Königreich oder das Centre National de Evaluation (CNE) in Frankreich. Darüber hinaus existieren Einrichtungen, die enger an das Wissenschaftssystem gebunden sind. Zu nennen sind hier die Untereinheit der amerikanischen National Science Foundation, die den Report on Science and Engineering Indicators herausgibt (Science Resources Statistics, SRS), das Observatoire des Sciences et des Techniques (OST) in Frankreich oder die Assessment-Panels der Vereniging van Universiteiten (VSNU) in den Niederlanden.³²

In Deutschland hat die Praxis der Forschungsevaluation keinen gemeinsamen institutionellen Ort und kennt bisher keine gemeinsamen Formate; hier koexistiert eine Vielzahl evaluativer Selbst- und Fremdbeschreibungen, die mit unterschiedlichen Absichten erstellt werden. Im Folgenden werden die relevanten Anbieter evaluativer Informationen kurz skizziert.

In den Zentralverwaltungen der *außeruniversitären* „Forschungssäulen“ (MPG, HFG, WGL, FhG) Deutschlands laufen traditionell vielfältige Informationen, auch solche zu den Forschungsleistungen der Einzelinstitute zusammen. Dies geschieht jedoch in unterschiedlicher Weise. Die Autoren sind der Auffassung, dass der Umfang der erzeugten Informationen derzeit bei der MPG und der FhG noch höher ist und eine bessere Qualität aufweist als bei den übrigen Einrichtungen. Dies liegt nicht zuletzt daran, dass MPG und FhG relativ klare „Missionen“ verfolgen, die in regelmäßiger Weise durch Selbstbeschreibungen und interne Steuerungskommissionen (Beiräte) validiert werden. Ein solcher Zustand wird nach den Systemevaluationen auch bei der HGF und der WGL angestrebt; Berichtssysteme sind derzeit im Aufbau. Künftig werden alle vier Zentralverwaltungen wichtige Informationslieferanten sein.

³¹ H.-D. Daniel, Wissenschaftsevaluation. Neuere Entwicklungen und heutiger Stand der Forschungs- und Hochschulevaluation in ausgewählten Ländern, CEST, Bern 2001; S. Hornbostel, Kurzgutachten für die DFG zum Stand der Forschung im Bereich Forschungsevaluation, Jena 2000.

³² Weitere Information bei den entsprechenden websites: HEFCE: <http://www.hefce.ac.uk>; CNE: <http://www.plan.gouv.fr/organismes/cne>; SRS: <http://www.nsf.gov/sbe/srs>; OST: <http://www.obs-ost.fr>; VSNU: <http://www.vsnunl.nl>.

Die DFG als zentrale Selbstverwaltungseinrichtung der deutschen Wissenschaft zur Förderung der Forschung an Hochschulen und öffentlich finanzierten Forschungsinstituten stellt mit ihren Informationssystemen GEPRIS und DFG-Ranking – wie erwähnt – zwei Informationstools zur Verfügung, die in den letzten Jahren systematisch ausgebaut wurden und Strukturinformationen zur Forschungslandschaft bereithalten.

Der *Stifterverband* für die deutsche Wissenschaft unterstützt die Forschungsberichterstattung der Bundesregierung und stellt Daten für EUROSTAT, OECD und UNESCO bereit.

Das *Hochschulinformationssystem HIS* führt als von Bund und Ländern finanzierte Einrichtung Informationserhebungen durch, die vorwiegend die Entwicklung von Verfahren zur Rationalisierung der Hochschulverwaltung, Mitwirkung bei deren Einführung und Anwendung sowie Gutachten zur Schaffung von Entscheidungsgrundlagen im Hochschulwesen zum Zweck haben. Einen dezidiert forschungspolitischen Auftrag hat das HIS nicht.

Der *Wissenschaftsrat (WR)* besitzt als „Fremdbeschreibungsagentur“ wissenschaftlicher Einrichtungen eine hohe Reputation. Er hat neben den Querschnittsevaluationen, die zeitlich und sachlich begrenzte evaluative Informationen dokumentieren³³, bisher allerdings kein systematisches Berichtssystem entwickelt, das Forschung in Deutschland beschreibt. Auch die Systemevaluationen der HFG und WGL sind – obgleich umfangreich – ihrem Zustandekommen und ihrer Anlage nach ad-hoc-Untersuchungen.

Das *BMBF* erhebt mit seinen Dokumentationen zur strukturellen Situation des deutschen Forschungssystems und der technologischen Leistungsfähigkeit Deutschlands systematisch wichtige evaluative Informationen. Als Bundeseinrichtung verarbeitet das BMBF diese Fremdbeschreibungen auf hoch aggregierter Ebene. Seit vielen Jahren führt das Forschungsministerium Programmevaluationen durch³⁴; bis heute wird aber nur ein Teil dieser von externen Evaluationsexperten durchgeführten Studien (häufig ex post-Wirkungsanalysen) der Öffentlichkeit zugänglich gemacht oder regulär publiziert. Wenig evaluative Information liegt bisher bezüglich der vom Bund betriebenen Einrichtungen der problemorientierten Ressortforschung vor.

Einige Wissenschaftsministerien der *Länder* (insbesondere Baden-Württemberg, Bayern, Nordrhein-Westfalen, Niedersachsen) haben eine Reihe von Initiativen zur Evaluation der Hochschulforschung angestoßen.

³³ Vgl. Fußnote 17.

³⁴ Übersicht bei Kuhlmann/Holland, Evaluation von Technologiepolitik, aaO.

Das Generaldirektorat Forschung der *Europäischen Kommission* hat im Rahmen des Auf- und Ausbaus der europäischen Forschungsrahmenprogramme nach und nach eine relativ systematische Evaluationspraxis entwickelt³⁵. Dazu gehören das Five Year Assessment des Rahmenprogramms durch ein Expertenpanel, jährliche Monitoring-Berichte, Evaluationen der Spezifischen Programme, horizontal Evaluationen übergreifender Fragestellungen und Studien der Wirkungen europäischer Programme im nationalen Forschungssystem³⁶. Im Rahmen der Kommissionsinitiative zur Schaffung eines „europäischen Forschungsraums“³⁷ gibt es Bemühungen, Indikatorenbündel zur Identifikation europäischer wissenschaftlicher „Centres of Excellence“ zu identifizieren, exemplarisch anzuwenden und in ein öffentlich zugängliches Berichtssystem zu implementieren; hierzu gehören Bibliometrie und Patentanalysen sowie Auswertungen vielfältiger Quellen evaluativer Informationen aus nationalen Kontexten³⁸. Für die strategische evaluative Analyse ganzer Institutionen-„Landschaften“ können solche Datenaggregationen künftig von großer Bedeutung sein. Außerdem erstellt das Generaldirektorat Forschung – ähnlich wie die OECD³⁹ – verschiedene Indikatorenreports auf der Grundlage evaluativer Information, etwa die „Indicators for benchmarking of national research policies“ mit aggregierten Daten zu „Human resources in RTD“; „Public and private investment in R&D“; „Scientific and technological productivity“; „Impact of RTD on economic competitiveness and employment“⁴⁰.

³⁵ L. Georghiu, aaO.

³⁶ Z.B. G. Reger/S. Kuhlmann, *European Technology Policy in Germany. The Impact of European Community Policies upon Science and Technology in Germany*, Heidelberg 1995 (Physica/Springer).

³⁷ Z.B. *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; S. Kuhlmann/J. Edler, *Changing Governance in European Research and Technology Policy*. In: Edler, J./Kuhlmann, S./Behrens, M. (eds.) (2003): *Changing Governance of Research and Technology Policy – the European Research Area*, Cheltenham (E. Elgar), 3–32.

³⁸ Z.B. E. C. M. Noyons, R. K. Buter, A. F. J. van Raan, U. Schmoch, T. Heinze, S. Hinze, R. Rangnow, 2003: *Mapping Excellence in Science and Technology across Europe*, Life Sciences, Brussels (EC-PPLS CT-2002-0001); E.C.M. Noyons, R. K. Buter, A. F. J. van Raan, U. Schmoch, T. Heinze, S. Hinze, R. Rangnow, 2003: *Mapping Excellence in Science and Technology across Europe*, Nanoscience and Nanotechnology, Brussels (EC-PPN CT-2002-0001).

³⁹ Z.B. *OECD*, *Science, Technology and Industry Scoreboard 2001*, Towards a knowledge-based economy, Paris 2001.

⁴⁰ *European Commission*, *Indicators for benchmarking of national research policies*. Luxembourg 2001 (Office for Official Publications of the European Communities).

Ein wichtigen Beitrag zur Generierung und zum Angebot evaluativer Informationen liefern Forschungseinrichtungen mit einem Schwerpunkt im Bereich *Wissenschafts- und Innovationsforschung*. Hier sind vor allem die folgenden zu nennen (wobei darauf hinzuweisen ist, dass eine ganze Reihe weiterer Forschergruppen gelegentlich Beiträge liefert): Das Institut für Wissenschafts- und Technikforschung (IWT) an der Universität Bielefeld; das Centrum für Hochschulentwicklung (CHE) in Gütersloh; das Wissenschaftszentrum Berlin (WZB); das Fraunhofer-Institut für Systemtechnik und Innovationsforschung (ISI) in Karlsruhe; das Wissenschaftliche Zentrum für Berufs- und Hochschulforschung an der Universität Kassel. Zunehmend tragen auch ausländische Forschungseinrichtungen evaluative Studien zum deutschen Forschungssystem bei. Zu nennen sind hier Einrichtungen wie das Centre for Science and Technology Studies (CWTS) an der Universität Leiden, das Department of Philosophy of Science and Technology der Universität Twente, Niederlande, oder auch spezialisierte Beratungsunternehmen mit dem Schwerpunkt Wissenschafts-, Forschungs-, Technologie- und Innovationspolitik, beispielsweise Technopolis (Hauptsitz Brighton, UK).

IV. Nachfrage und Verwendung evaluativer Information

Im folgenden werden Desiderata der Bereitstellung evaluativer Information mit Blick auf mögliche Nutzer behandelt. Die Angaben zur Nachfrage beruhen teils auf explorativen Interviews, teils auf Plausibilitätserwägungen der Autoren.

Nutzerinteressen

Die *Deutsche Forschungsgemeinschaft* hat im Jahr 2000 einen Senatsausschuss „Perspektiven der Forschung“ ins Leben gerufen. Er soll als „Frühwarnsystem“ im Rahmen wissenschaftlicher Bedarfsanalysen innovative Forschungsfelder identifizieren und die forschungsstrategischen Initiativen des Senats vorbereiten. Damit ergänzt die DFG ihre bisherige Praxis, dass zu jedem Thema zu jeder Zeit ein Antrag gestellt werden kann. Die erforderlichen Bedarfsanalysen können nur auf der Basis evaluativer Information erstellt werden. Hierzu gehören Input-Informationen (Wer fördert bereits was, national und international?), Output-Informationen auf mittlerem und höherem Niveau (Welche Einrichtungen leisten bereits exzellente Forschung in bestimmten Gebieten? Welche Einrichtungen sollten gestärkt werden?) sowie Information über erzielte

oder potentielle Wirkungen von Forschung und Förderung (z.B. in der akademischen Ausbildung; im Bereich Technologie und Innovation; zur Lösung sozialer oder ökologischer Probleme, etc.).

Strategische Planung großer Forschungseinrichtungen: Die Forschungsträgerorganisationen wurden im Rahmen der Systemevaluationen aufgefordert, deutlicher als bisher explizite strategische Entwicklungspläne zu betreiben. Sie müssen daher künftig – auf der Grundlage einer regelmäßig aktualisierten, kritischen Analyse ihres eigenen Leistungsspektrums – ihr Forschungsportfolio im Kontext der nationalen und internationalen Forschungslandschaft positionieren. Ähnlich wie die DFG benötigen sie dabei Input-Informationen (Wer investiert bereits in welchem Bereich, national und international?), Output-Informationen auf mittlerem und höherem Niveau (Welche anderen Einrichtungen leisten bereits exzellente Forschung in bestimmten Gebieten?) sowie Information über erzielte oder potentielle Wirkungen von Forschung und Förderung (z.B. im Bereich Technologie und Innovation; zur Lösung sozialer oder ökologischer Probleme, etc.).

Bundes- und Landeseinrichtungen der Forschungspolitik (BMBF, BMWi; Bundestag; Wissenschaftsministerien der Länder) sind verantwortlich für die Pflege der regionalen, nationalen und institutionellen Forschungslandschaft im internationalen Kontext. Strategieentwicklung und Planungsarbeiten können in dieser Hinsicht bisher nur auf fragmentarische oder ad hoc erstellte Aggregate evaluativer Information zurückgreifen. Weder in der Interaktion mit Forschungsträgerinstitutionen und Universitäten noch gegenüber den Integrationsversuchen der Europäischen Union (Europäischer Forschungsraum; „Mapping of Centres of Excellence“) können Bund und Länder bisher ihre Interessen, Entscheidungen und Vorgehensweisen auf systematische Analysen stützen (abgesehen von – zum Teil guten, aber aufwändigen – ad hoc-Studien durch Einrichtungen der Wissenschafts- und Innovationsforschung). Benötigt werden Input-Informationen (Welche Forschungskapazitäten gibt es bereits, und wer fördert was, national und international?), Output-Informationen auf mittlerem und höherem Aggregationsniveau (Welche Einrichtungen leisten bereits exzellente Forschung in bestimmten Gebieten? Welche Einrichtungen sollten gestärkt werden?) sowie Information über erzielte oder potentielle Wirkungen von Forschung und Förderung (z.B. in der akademischen und sonstigen Bildung; im Bereich Technologie, Innovation und Wirtschaftsentwicklung; zur Lösung sozialer oder ökologischer Probleme, etc.).

Das *Generaldirektorat Forschung der Europäischen Kommission* wird im Rahmen der Initiative „Europäischer Forschungsraum“ die Arbeiten

an der Identifikation wissenschaftlicher „Centres of Excellence“ fortsetzen und entsprechende Ergebnisse einem öffentlich zugänglichen Berichtssystem zuführen. Hier entsteht künftig eine dauerhafte Nachfrage nach verlässlichen, regelmäßig aktualisierten, mittel bis hoch aggregierten evaluativen Informationen aus nationalen Kontexten. Je hochwertiger diese Nachfrage gedeckt wird, desto adäquater wird die Forschung in Deutschland in der europäischen Landschaft abgebildet.

Auch der *Wissenschaftsrat* wird bei künftigen Evaluationen von Forschungsinstitutionen oder Themenfeldern ebenso wie bei möglichen weiteren Systemevaluationen Aggregatinformationen heranziehen müssen. Ein besonders dringender Bedarf besteht hinsichtlich Strukturinformationen zur Forschung in Universitäten.

Einrichtungen der *Wissenschafts- und Innovationsforschung* konnten ihre – teils aus wissenschaftlichem Interesse, teils im Auftrag forschungspolitischer Akteure durchgeführten – Studien zu Stärken, Schwächen und Dynamik des deutschen Forschungssystems bisher nicht auf ein verlässliches, allgemein akzeptiertes System evaluativer Information stützen. Vielfach müssen Fundamentaldaten aufwändig im Rahmen von Einzelstudien erhoben werden.

Desiderata

Zusammenfassend lassen sich die möglichen Verwendungszwecke evaluativer Information, die nach kompatiblen Formaten generiert, systematisch und regelmäßig gesammelt, nach festgelegten Fragestellungen strukturiert und unterschiedlichen Nutzungen aggregiert wurde, darstellen wie in Tabelle 1. Sie ordnet die Relevanz von evaluativen Input-Informationen, Output-Informationen sowie Wirkungsinformationen nach Evaluationszwecken und Aggregationsebenen. Tabelle 1 beruht auf Plausibilitätserwägungen der Autoren.

Der Bedarf an evaluativer, forschungsstrukturbezogener Information ist um so größer, je *höher aggregiert die Betrachtungs- und Entscheidungsebene der Informationsnutzer* ist. Ein großer Teil des erkennbaren Bedarfs wird vom heute verfügbaren Angebot im deutschen Forschungssystem *nicht gedeckt*. Die Aggregation setzt voraus, dass auch auf niedriger aggregierten Ebenen evaluative Informationen systematisch erhoben und zueinander in Beziehung gesetzt werden. Desiderata bestehen vor allem in folgenden Bereichen:

Tabelle 1:
Zwecke und Aggregationsebenen des Bedarfs an evaluativer Information

Aggregationsebenen evaluativer Information	Evaluationszwecke				
	Qualitäts- sicherung	Gesellschaft- licher Nutzen	Selbststeuerung und organi- sationaler Wandel	Effizienz- und Effektivitäts- kontrolle	Struktur- verbesserung
<i>Vertikal</i>					
Einzelwissen- schaftler	▶				
Forschungs- gruppe; Lehrstuhl	▶ ◆ ▷	▷	▶ ◆ ▷		
Fakultät; Institut; Abteilung	▶ ◆ ▷	▷	▶ ◆ ▷		
Universität; Forschungs- träger- Organisation	▶ ◆ ▷	▷	▶ ◆ ▷	▶ ◆ ▷	▷
<i>Horizontal</i>					
Disziplin; Themenfeld	▶ ◆ ▷	▷	▷	▷	▶ ◆ ▷
Programm		◆ ▷		▶ ◆ ▷	
Institutionen- „Landschaft“	▶ ◆ ▷	▷	▷	▶ ◆ ▷	▶ ◆ ▷

Erläuterung: ▶ = Input-Informationen; ◆ = Output-Informationen; ▷ = Wirkungsinformationen.

1. Es fehlen bisher *kompatible Formate und regelmäßige Rhythmen der Erfassung evaluativer Information* quer zu den Forschungseinrichtungen, auch mit Blick auf die internationale Anschlussfähigkeit wichtiger Daten- und Informationsbestände, sowohl bei Input-Informationen, Output-Informationen wie Wirkungsinformationen. Die Formatierung und Verfahrensabstimmung erfordert ein koordiniertes Vorgehen aller Verantwortlichen im deutschen Wissenschafts- und Forschungssystem.
2. Kompatible Formate bilden die Voraussetzung für *Vergleiche* (institutionelle, internationale, ...) auf *aggregierten Niveaus* – hier ist der Be-

darf besonders ausgeprägt. Dabei müssen die unterschiedlichen Aufgabenstellungen und fachlichen Orientierungen heterogener Forschungseinrichtungen unbedingt berücksichtigt werden – dies erfordert *intelligente Aggregationsstrategien* für evaluative Informationen.

3. Regelmäßige Rhythmen bilden die Voraussetzung für die *zeitliche Verknüpfung von Forschungsinput-, Forschungsoutput- und Forschungsstrukturdaten*. Hierdurch werden die unterschiedlichen Evaluationszwecke adressiert und mögliche Wirkungsmechanismen identifiziert.

V. Mögliche Organisation evaluativer Information

Die folgenden Vorschläge für eine organisatorische *Verbesserung der Bereitstellung evaluativer Information im deutschen Forschungssystem* bleiben aufgrund des explorativen Charakters unserer Untersuchung *bewusst vorläufig* und sollen vor allem zur Diskussion anregen. Dabei wollen wir betonen, dass die Aufmerksamkeit vor allem der Bereitstellung evaluativer Information gilt, hier also nicht Vorschläge zum konkreten institutionellen Ort künftiger Durchführung von Evaluationen oder Forschungsstrategieplanung unterbreitet werden. Diese Aufgaben sind und bleiben in der Verantwortung der Forschungsförderungsinstitutionen, der Forschungsträgereinrichtungen, der Forschungspolitik sowie – als weithin anerkannter neutraler Instanz – des Wissenschaftsrates.

Anregend kann dabei ein Blick auf mögliche Modelle im Ausland wirken. Exemplarisch seien drei Einrichtungen genannt, die verwandte Aufgaben erfüllen, jedoch mit unterschiedlicher Ausprägung:

Überinstitutionelle Datensammlung. Das Observatoire des Sciences et des Techniques (OST) wurde 1990 als öffentliche Einrichtung etabliert. Etwa 15 Personen arbeiten an Projekten für die Trägereinrichtungen des OST, vor allem nationale Ministerien und öffentliche und private Forschungsträgereinrichtungen. OST publiziert jährlich einen FuT-Indikatorenbericht (Science & Technologie – Indicateurs, éditions Economica, Paris) und erstellt Indikatoren, die Entscheidungsträgern der Forschungspolitik und des Forschungsmanagements bei der Strategieentwicklung unterstützen sollen. Dabei werden evaluative Daten bisher überwiegend nur sehr hoch aggregiert präsentiert, kaum nach Institutionen differenziert und wenig analytisch aufbereitet.

Überinstitutionelle Datensammlung, -erhebung und Analyse. Die Abteilung „Science Resources Statistics (SRS)“ der National Science Foundation (NSF) der USA erfüllt die gesetzliche Aufgabe „to provide a central clearinghouse for the collection, interpretation, and analysis of data on scientific and engineering resources, and to provide a source of informa-

tion for policy formulation by other agencies of the Federal Government“ (SRS 2002). Zu diesem Zweck erstellt SRS etwa 14 periodische Erhebungen sowie eine Reihe weiterer Studien und Projekte. Alle US-Bundesbehörden, die FuE betreiben, liefern SRS relevante Daten. Außerdem arbeitet SRS eng mit Universitäten, Industrieunternehmen, Berufsorganisationen und internationalen Organisationen zusammen und hält evaluative Information anderer Erzeuger bereit. Der Schwerpunkt der Aktivität liegt bei SRS allerdings auf der Aggregation und Analyse deskriptiver Daten, weniger hingegen bei der strategischen Auswertung evaluativer Information.

Überinstitutionelle Datensammlung, -erhebung und Analyse kombiniert mit Wissenschafts- und Innovationsforschung, Evaluation, Vorausschau und Beratung. Das schweizerische Zentrum für Wissenschafts- und Technologiestudien (CEST) umfasst ein Kompetenzzentrum Scientometrie, Einheiten für Evaluation, Technologie und Innovation, für Prospektion sowie eine Dokumentationsstelle für Wissenschaftspolitik. CEST beschafft und überprüft Grundlagen für die schweizerische Forschungs-, Hochschul- und Technologiepolitik. Es untersucht den aktuellen Stand und zeigt mögliche Entwicklungen auf. Zu diesem Zweck erstellt es sowohl Fallstudien als auch vergleichende Analysen. Es führt Evaluationen von Disziplinen, von Institutionen sowie bundesstaatlich finanzierten oder mitfinanzierten Wissenschafts- und Technologieprogrammen durch.

Für das deutsche Forschungssystem lassen sich vier alternative Richtungen für eine organisatorische Verbesserung der Bereitstellung evaluativer Information skizzieren:

1. *Fortsetzung des Status quo:* Die aktuelle Situation der Nachfrage nach, der Erzeugung von und des Angebots an evaluativer, forschungsstrukturbezogener Information wird als akzeptabel bewertet; nach dieser Sichtweise besteht also kein Bedarf nach einer neuen Dienstleistung.
2. *Vereinbarung* aller relevanten Akteure im deutschen Forschungssystem (einschließlich Repräsentanten der Wissenschafts- und Innovationsforschung) über ein gewisses *Minimum an kompatiblen Formaten und regelmäßigen Rhythmen* zur Erfassung evaluativer Information, quer zu den Forschungseinrichtungen, bei gleichzeitiger Sicherstellung internationaler Kompatibilität wichtiger Daten- und Informationsbestände, sowohl bei Input-Informationen, Output-Informationen wie Wirkungsinformationen. Die kompatibel formatierten Informationsbestände können als Grundlage für Aggregation, Vergleiche etc. dienen und von allen interessierten Stellen durchgeführt werden.

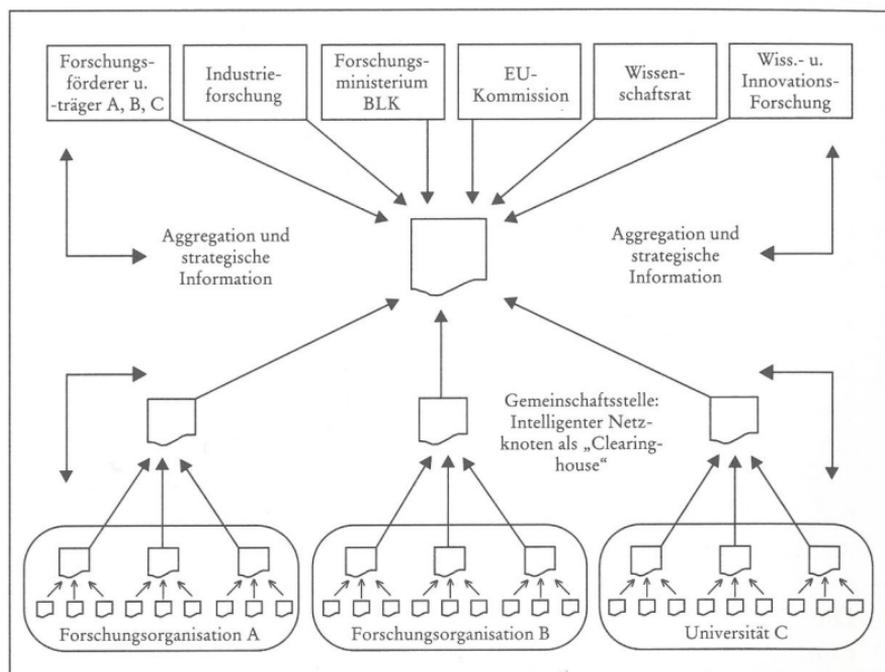
3. Einrichtung einer *Gemeinschaftsstelle* der deutschen Forschung für die Aufbereitung und Vermittlung kompatibel formatierter Forschungsstrukturinformationen. Die Stelle ist als intelligenter „Vermittlungsknoten“ mit der Funktion eines *Clearinghouse* konzipiert, eingebettet in ein Netz von Anbietern und Abnehmern und damit kein monopolistischer Informationsverwalter. Diese Gemeinschaftsstelle sollte auch die Moderation zwischen interessierten Akteuren des Forschungssystem übernehmen.
4. Einrichtung einer *zentralen, übergreifenden Evaluationsagentur*, in Verbindung mit einer zentralen Strategieentwicklungseinheit des deutschen Forschungssystems. Um einer solchen Einrichtung Wirksamkeit zu verleihen, müsste sie mit nennenswerten Machtressourcen gegenüber den relativ autonomen großen Forschungsorganisationen und den Bundesländern ausgestattet werden.

Die Alternative 1 (kein Handlungsbedarf) erscheint angesichts des wachsenden Problemdrucks und Analysebedarfs nicht als wünschenswert. Die Alternative 4 (zentrale Evaluationsagentur) würde kaum Akzeptanz bei der Mehrheit der Forschungsorganisationen finden, nicht zuletzt weil stets die Gefahr einseitiger politischer Vereinnahmung besteht. Außerdem drohen stark institutionalisierte Maßnahmen schneller zu „vergreisen“ als flexible und reversible Einrichtungen.

Wir empfehlen eine *Kombination der Alternativen 2* (mehr Koordination von Formaten und Verfahren der Erfassung evaluativer Information) *und 3* (Informations-Vermittlungsknoten als Clearinghouse). Die Aufgabe einer solchen Gemeinschaftsstelle „Clearinghouse“ besteht in der *Unterstützung* ihrer Trägerorganisationen bei der Bereitstellung von – je nach Institution verschieden ausgeprägten Bedarfen an – evaluativen Informationen: Qualitätssicherung, gesellschaftlicher Nutzen, Effizienz- und Effektivitätskontrolle, Selbststeuerung und organisationaler Wandel, Strukturverbesserung (siehe Abbildung 1).

Die Gemeinschaftsstelle „Clearinghouse“ wird gesteuert und überwacht von einem *Konsortium* (d.h. einer freiwilligen Kooperation) bestehend aus allen relevanten deutschen Einrichtungen der Forschungspolitik, der Forschungsförderung, der Forschungsträgerinstitutionen sowie ausgewählten Einrichtungen der Wissenschafts- und Innovationsforschung. Dieses Konsortium vereinbart den Umfang und die Formate quantitativer und qualitativer evaluativer Information. Das Konsortium empfiehlt, welche Informationen Forschungsorganisationen, Forschungsförderer und andere Konsortialmitglieder beitragen, und setzt den Rahmen für Zielsetzungen, Formate und Verwendungszwecke der evaluativen Informa-

Abbildung 1:
Funktionen einer deutschen Gemeinschaftsstelle „Clearinghouse“
für Forschungsstrukturinformationen: „intelligenter Netzknoten“



tionen. Die Gemeinschaftsstelle wiederum moderiert die erforderlichen Abstimmungsprozesse. Informative Produkte der Gemeinschaftsstelle stehen grundsätzlich allen Konsortialmitgliedern zur Verfügung. Ausnahmen hiervon sind möglich, jedoch einvernehmlich zu regeln. Möglichst viele Informationsprodukte sollen der allgemeinen, interessierten Öffentlichkeit zugänglich gemacht werden (z.B. interaktiv gestaltete Internet-Websites, Berichte und Analysen). Dabei sind mehrere Gestaltungsvarianten für eine Gemeinschaftsstelle „Clearinghouse“ denkbar:

- Arbeit auf der Grundlage *freiwilliger Selbstverpflichtung* aller relevanten Institutionen des deutschen Wissenschafts- und Forschungssystems oder im Rahmen *regulierter Berichtspflichten* (z.B. auf gesetzlicher Grundlage).
- Ansiedlung der Stelle an einer *wissenschaftlichen Einrichtung* (Universitätslehrstuhl; WGL-Einrichtung nach dem Vorbild der GESIS; Bindung an eine andere Forschungsträgerorganisation, z.B. eine – noch zu gründende – nationale Akademie der Wissenschaften) oder bei einer

forschungspolitischen Institution (etwa beim *Wissenschaftsrat* als intermediärer Beratungseinrichtung).

Zur *Arbeitsweise* einer Gemeinschaftsstelle „Clearinghouse“ schlagen die Autoren vor, dass diese vor allem aggregierte und strategische Informationen auf der Basis evaluativer Information erarbeitet. Die realisierten Aggregationsebenen müssen in Relation zu den Verwendungszwecken der generierten Information stehen. Zu den typischen Aufgaben und Verfahrensweisen der Gemeinschaftsstelle zählen:

- Die Erarbeitung und Pflege *kompatibler Formate und regelmäßiger Rhythmen der Erfassung evaluativer Information* quer zu den Aggregationsebenen und Evaluationszwecken.
- Die *Moderation* der Perspektiven und Interessen der verschiedenen Institutionen im deutschen Forschungssystem, die durch die Formatierung und Verfahrensabstimmung erforderlich wird.
- Die Erarbeitung von thematischen, international kompatiblen *Forschungslandkarten*, die einen deskriptiven Überblick über den jeweiligen status quo ermöglichen, „Lücken“ in der Forschungslandschaft identifizieren sowie die Vernetzung von Forschungsaktivitäten erleichtern.
- Die Durchführung von *strategisch orientierten Vergleichen* (institutionelle, internationale, ...), vor allem auf aggregierten Niveaus. Hierzu sind Standards und Modelle zu erarbeiten, wie heterogene evaluative Informationen zueinander ins Verhältnis zu setzen sind.
- Die Durchführung von *Stärken-Schwächen-Analysen* als Basis für Entscheidungen über Mittelallokationen durch Forschungsförderungseinrichtungen (z.B. DFG, BMBF, Landesministerien)

Nach Auffassung der Autoren dürfen evaluative Informationen nicht „blindlings“ gesammelt, aggregiert, analysiert und verbreitet werden, sondern nur im Rahmen vereinbarter konkreter Evaluationszwecke, Rhythmen und Formate. Die Zusammenarbeit im „Clearinghouse“ schließt ein, dass die Trägereinrichtungen „vertrauliche“ Informationen für vereinbarte strategische Analysen bereit stellen.

Die Gemeinschaftsstelle kann nur in begrenztem Umfang eigene Forschungsaufgaben wahrnehmen. Anspruchsvolle Analysen sollten in der Regel von Einrichtungen der *Wissenschafts- und Innovationsforschung* durchgeführt werden. Ein rollierendes System von *Personalabstellungen* aus Nutzereinrichtungen sollte deren Vertrautheit mit den durch die Gemeinschaftsstelle verfügbar zu machenden Informations- und Datenbestände stärken und ihre externe Nutzung unterstützen.

Bei ihren Aggregations- und Analysearbeiten muss die Gemeinschaftsstelle „Clearinghouse“ die Grenzen von Evaluationsverfahren (siehe Ausführungen in unserem ersten Beitrag) bewusst halten. „Blinde Flecken“ des Nichtwissens sollen nicht allein als Risiko sondern auch als Chance für Kreativität akzeptiert werden.

Summary

The article is based on an expert opinion paper on current practice of research evaluation in Germany and future options for organizational change, prepared for the German Research Foundation (DFG) in 2002. The paper starts with the observation that both, the more general demand for accountability in science and research, and more specific developments, such as global budgeting and performance-based allocation of research funds in German universities, call for systematic research evaluations, which in turn depend on the careful collection and analysis of evaluative information, extending traditional peer review in scope. In the first part of the paper, published earlier in WR 37 (1), the authors elaborate on the concept of evaluation of research performance. They discuss different types of evaluative information and indicators, outline established methodical approaches, and critically reflect on the purpose of research evaluations and their limits. The second part provides an account of current research evaluation practice in Germany including data availability, level of data aggregation, and main providers and users. The paper concludes that despite considerable activity and dynamics, current practice remains methodically and organizationally unsystematic and dispersed. Both, compatible data formats and regular series of data collection and analysis across scientific disciplines, research institutions and research programmes are clear desiderata at present. In order to improve the current situation, the authors recommend to establish a *Clearinghouse* for research evaluation whose main mission would be the support of collecting, managing and distributing evaluative information. The Clearinghouse would be borne by all major German research and funding organisations. The discussion concludes with a tentative outline of the functions and institutional design of the proposed new agency.