

Überlegungen zur Qualitätssicherung geisteswissenschaftlicher Forschungsleistungen

Ausgearbeitet von der ad-hoc-Kommission der phil.-hist. Klasse zur Bewertung geisteswissenschaftlicher Forschungsleistungen (Fassung 11. 6. 2008)

ALLGEMEINE GESICHTSPUNKTE¹

1. Das explosionsartige Wachstum von Mechanismen zur Qualitätskontrolle in der Forschung führt vor allem in den Geisteswissenschaften zu verbreiteter Abneigung, teils fundamentaler Kritik. Gleichzeitig fordern Entscheidungsträger außerhalb der ÖAW eine verstärkte „Messung“ und Kontrolle von Forschungsleistungen. Das entspricht einer gesellschaftlichen Entwicklung, die man mehr oder weniger kritisch sehen kann, der sich die ÖAW als Forschungsträgerorganisation aber nicht entziehen kann.
2. Die bisher bekannten Instrumente der Qualitätskontrolle erreichen noch lange nicht die Qualität jener Forschung, die sie kontrollieren sollen. Vieles ist noch in Erprobung. Selbst mit einem wesentlich verbesserten Instrumentarium wird es kaum möglich sein, Forschungsleistungen adäquat zu messen. Ein repräsentatives Bündel an Kennzahlen kann aber für viele Fragen recht aussagekräftig sein. Indikatoren erlauben vor allem Schwächen zu erkennen; gutes Mittelmaß mit gezielter Erfüllung der Kriterien lässt sich weniger leicht von Exzellenz unterscheiden.
3. Die ÖAW hat mit beträchtlichem Aufwand ein eigenes System zur Bemessung von Forschungsleistungen implementiert. Die Daten werden in AkademIS erfasst und dann zu Kennzahlen für die Wissensbilanz aggregiert. AkademIS hat den Nachteil, dass die Eingabe sehr aufwändig und wenig benutzerfreundlich ist. Das ist zum Teil ein technisches Problem und im Prinzip lösbar. Ein andere Frage ist, welche Aussagekraft die Wissensbilanz bzw. einzelne ihrer Kennzahlen haben.
4. Dabei ergibt sich für die geisteswissenschaftliche Forschung ein Problem. In vielen Naturwissenschaften ist eine Bewertung auf Grund weniger Kennzahlen üblich. Das sind, einzeln oder in Kombination: Publikationen in *peer-reviewed journals*; Impact-Faktoren (Zitationshäufigkeit) nach immer weiter entwickelten Formeln (z.B. H-Index); und die Drittmittelquote. Viele Geistes- und Kulturwissenschaften sind aus

¹ Dieses Papier bezieht sich auf die Situation in den Geistes- und Kulturwissenschaften und versucht, bei allen disziplinären Unterschieden, gemeinsame Probleme und Grundsätze herauszuarbeiten. In den meisten Sozialwissenschaften ist die direkte Aussagekraft quantitativer Indikatoren (z.B. Impact-Faktoren) wesentlich breiter akzeptiert und daher weniger problematisch als hier dargestellt; ISI, Google Scholar und Melvyl (s.u.) bieten vielfach entscheidende Kriterien. Die Besonderheiten in den Rechtswissenschaften wiederum sind so ausgeprägt, dass ihnen in Folge ebenfalls nicht hinreichend Rechnung getragen werden kann; hingewiesen sei in diesem Zusammenhang auf die im März 2008 von der rechtswissenschaftlichen Fakultät der Universität Wien abgegebene Stellungnahme zur Frage der Evaluierung von Lehr- und Forschungsleistungen an dieser Fakultät und auf das Positionspapier des internationalen Wissenschaftlichen Beirats dieser Fakultät vom Mai 2008.

verschiedenen Gründen schwieriger zu evaluieren. Damit grobe Verzerrungen vermieden werden, ist eine Vielzahl von Indikatoren nötig. Das macht aber wiederum die Erfassung schwierig und aufwändig und ruft dadurch Kritik hervor.

5. Einige Gründe, die eine Beurteilung geisteswissenschaftlicher Forschungsleistungen erschweren: komplexere Forschungsfragen und Darstellungsmethoden; unterschiedliche Karriereverläufe und Forschertypen; Dezentralität der Publikationsorgane; methodische Unterschiede selbst innerhalb der Disziplinen. Die Schwierigkeit liegt nicht darin, dass, wie oft gesagt, Geisteswissenschaften ‚soft‘ und weniger kompetitiv wären. In den großen Fächern mit tausenden Studenten etwa erreicht nur ein ganz geringer Anteil der Absolvent/innen eine wissenschaftliche Karriere, die Konkurrenz ist also stärker.

ÜBERLEGUNGEN ZU EINZELNEN BEWERTUNGSKRITERIEN

1. **Publikationen:** Entscheidend in den meisten Fächern sind Monographien; sie begründen den wissenschaftlichen Ruf eines Forschenden. Ebenso können aufwändige Quelleditionen Spitzenprodukte der Forschung darstellen. Artikel sind demgegenüber oft sekundär, insgesamt jedoch durchaus aussagekräftig. Dabei ist auch die Bedeutung der Zeitschriften unterschiedlich. In vielen Fächern gibt es kaum ‚flagship journals‘ (wie Science, Nature). Eine Einladung zur Publikation durch den Herausgeber kann, muss aber nicht wichtiger sein als formelles Peer review. Viele der besten Artikel erscheinen in (thematisch fokussierten) Sammelwerken; zu unterscheiden von Gelegenheitspublikationen (z.B. Festschriften). Von den wissenschaftlichen Publikationen im engeren Sinn zu unterscheiden sind (durchaus verdienstvolle) Sach- oder Studienbücher, Einführungen, Überblickswerke, Lexika (mit Ausnahme wissenschaftlicher Enzyklopädien), Ausstellungskataloge, etc.

Bewertungsmöglichkeiten:

- a. **Quantitativ:** Die Zahl der (wissenschaftlichen) Publikationen ist ein zentrales, aber manchmal irreführendes Kriterium (z.B. viele Beiträge zum selben Thema, schlechte oder wenig innovative Bücher, unzugängliche Veröffentlichungsformen)
- b. **Verbreitung** der Publikationen: Die Verbreitung von Zeitschriftenartikeln stellt der European Reference Index for the Humanities (ERIH) dar, der Zeitschriften mit einer gewissen Relevanz aufnimmt („listed journals“) und dann je nach ihrer internationalen oder nur regionalen Verbreitung unterscheidet (Kategorien A, B, C). Die jeweilige Auswahl wurde viel kritisiert, wird aber auch ständig verbessert. Die Verbreitung von

Monographien erfassen internationale Bibliothekskataloge (der besten Indikator ist MELVYL). In raschem Wachstum ist die Erfassung von Publikationen über Google Scholar und ISI-Web of Knowledge, die in vielen Disziplinen (aber nicht in allen) schon recht aussagekräftige Ergebnisse liefern. Ein prinzipielles Problem stellt bei all diesen Instrumenten die Sprache dar; englische Publikation verbessert meist die Zugänglichkeit, das sollte aber nicht zu einer Diskriminierung anderer Sprachen führen. Das gilt besonders für Disziplinen wie Sprach- oder Rechtswissenschaften. Eine gewisse Lösungsmöglichkeit bieten englische Abstracts.

- c. Problematischer als in den meisten Naturwissenschaften (die dabei aber auch keine einheitlichen Standards haben) ist die Konzentration auf **Artikel in „peer-reviewed journals“**. Das würde in vielen Geisteswissenschaften eine Änderung der Publikationskultur erfordern (z.B. weniger Bedeutung von Monographien, Sammelbänden oder Proceedings), die nicht wünschenswert wäre. Vor allem ist die Frage ungelöst, wie Monographien bewertet werden können. Derzeit werden in vielen Fächern vor allem Monographien von Autoren einem Peer review unterzogen, die noch eher unbekannt sind, während höhere Reputation Publikationsmöglichkeiten in großen z.T. kommerziellen Verlagen erschließt. Überlegt wird international eine dem ERIH entsprechende Klassifizierung der Verlage, dazu gibt es aber bisher keine anerkannten Kriterien.
- d. In vielen Geisteswissenschaften werden die Zitationen noch nicht ausreichend erfasst, um eine adäquate Bewertung von **Impact-Faktoren** zu erlauben, wie sie in vielen Natur-, aber auch Sozialwissenschaften üblich (wenn auch nicht ganz unumstritten) sind. Dafür stehen im Prinzip sehr differenzierte numerische Instrumente zur Verfügung (z.B. ISI Web of Science, Harzing's Publish or Perish), die jedoch in den Geisteswissenschaften oft zu irreführenden Ergebnissen führen. Das liegt u.a. an der viel heterogeneren Publikationsstruktur und an der größeren Sprachenvielfalt (das Potential der Geisteswissenschaften würde entscheidend reduziert, wenn alles auf Englisch publiziert werden müsste). Eine weitere Schwierigkeit ist, dass die ‚Halbwertszeit‘ von Publikationen in den Geisteswissenschaften viel länger ist als anderswo. Manche Werke werden erst nach Jahren, ja Jahrzehnten rezipiert. Außerdem kann noch mehr als in den Naturwissenschaften mit populären Themen oder kontroversiellen Thesen ein ‚impact‘ erreicht werden, der völlig im Gegensatz zur wissenschaftlichen Bedeutung steht. In der Wissensbilanz der ÖAW wurde aus den genannten Gründen auf die fächerübergreifende Verwendung von Impact-Faktoren verzichtet. Allerdings gibt es (z.B. von Google

Scholar) große Bemühungen, die Aufnahme geisteswissenschaftlicher Publikationen zu intensivieren.

- e. **Qualitativ/inhaltlich:** Die Universität Wien hat auf Grund des „Grimm-Papieres“ eine formale Einteilung eingeführt, 3 **Qualitätsstufen** bei Publikationen (A: Fortschritt der wissenschaftlichen Erkenntnis, B Überblick über den Forschungsstand, C populärwissenschaftliche Publikationen). Diese Unterscheidung ist wichtig, setzt aber eine Selbsteinstufung voraus (die möglicherweise zu Verzerrungen führt) und muss durch Peer review und Evaluationen überprüft werden. Die Publikationstätigkeit der Forschungseinheiten der ÖAW wird durch den wissenschaftlichen Beirat und in den zyklischen Evaluierungen beurteilt, wobei in unterschiedlichem Maß auf die oben genannten Kriterien zurückgegriffen wird. Wünschenswert wäre auch eine inhaltliche Beurteilung von Schlüsselpublikationen, wie sie bisher im englischen Research Assessment Exercise (RAE) üblich waren; das bedeutet aber einen sehr hohen Aufwand.
2. **Drittmittel-Einwerbung:** Selbständig betreute Drittmittel-Projekte gehören zunehmend zur Karriere und sind vor allem wichtig für die Institution. Sie erlauben wesentlich flexibler als eine Anstellung auf Dienstposten neue Projekte in Angriff zu nehmen und Nachwuchspflege zu betreiben; die Projekte werden dabei in der Regel eingehend evaluiert. Doch sollte die Drittmittel-Quote auch nicht überbewertet werden. Z.B. kann ein Einzelwissenschaftler, der sich auf seine Forschungsarbeit konzentriert, hervorragende Ergebnisse liefern
3. Wichtig für die Geisteswissenschaften ist eine Mehrzahl von **weiteren Kriterien**. Gemeinsam mit den vorgenannten ergibt sich daraus ein differenziertes Profil:
- Vorträge (national oder international, eingeladen oder nicht, *Keynote speakers*)
 - Preise, Ehrungen, Mitgliedschaften und *personal grants*
 - Internationale Vernetzung (Forschungsaufenthalte im Ausland, wissenschaftliche Gäste, Teilnahme an Konferenzen, Kooperationen, Reputation im Ausland)
 - Anbahnung und erfolgreicher Abschluss von Projekten
 - Organisation wissenschaftlicher Veranstaltungen
 - Nachwuchsförderung; Lehre; Habilitation
 - Populärwissenschaftliche Veröffentlichungen und Aktivitäten
 - Wissenschaftliche Organisationstätigkeit: Administrative und beratende Funktionen, Mitgliedschaft in Scientific Boards, Gutachten etc.
 - Bewahrung kulturellen Erbes, Erfassung von Dokumentationsobjekten etc.
- Für die Wissensbilanz der ÖAW werden diese Kriterien bereits verwendet.

SCHLUSSFOLGERUNGEN

1. Gerade in den Geisteswissenschaften sind die Verhältnisse zwischen den, ja selbst innerhalb der Forschungsrichtungen so unterschiedlich, dass sie schwer miteinander zu vergleichen sind. Viele Kriterien gelten in vielen, aber nicht in allen Disziplinen (z.B. Internationalität ist in vielen rechtswissenschaftlichen Bereichen nicht entscheidend).
2. Peer review durch fachnahe, ausländische Gutachter/innen, das auch qualitative Kriterien anwendet, wird bei der Beurteilung von Forschungsleistungen unersetzlich bleiben und sich wohl auf immer weitere Bereiche ausbreiten. Die Gutachtertätigkeit wird vor allem in einem fortgeschrittenen Karrierestadium zu einer immer größeren Belastung werden, ist aber auch eine wesentliche Aufgabe der ÖAW als Gelehrten-gesellschaft. Schon deshalb muss im Evaluationsprozess der gezielte Einsatz numerischer und formaler Kriterien, wie sie oben erwähnt wurden, zunehmen. Sie erlauben eine rasche erste Orientierung über mögliche Stärken und Schwächen einer Forschungseinrichtung. Über die jeweilige Bedeutung einzelner Kriterien und die Möglichkeit, sie quantitativ zu fassen, sind in den Disziplinen genauere Maßstäbe noch zu entwickeln. Im Zuge der nächsten turnusmäßigen Evaluation von Akademie-Forschungseinheiten sollte daher auch jeweils die prinzipielle Frage der Evaluationskriterien für jede einzelne Forschungseinheit aufgeworfen werden.
3. Wesentlich bleibt die fachspezifische inhaltliche und strategische Beurteilung. Numerische Indikatoren sind letztlich nur dann brauchbar, wenn sie in ein diskursives Umfeld eingebettet sind. Das heißt: Verbindung mit Peer review bzw. Evaluationen; Rückbindung an Forschungsziele und –programme der beurteilten Forschungseinheit; Recht zur Stellungnahme für die Beurteilten; allenfalls Diskussion von Ergebnissen innerhalb der Akademie-Gremien.
4. In dieser Form können die Ergebnisse einer Beurteilung auch (z.B. finanzielle) Konsequenzen haben. Vor allem wäre es wichtig, dass sehr gute Forschungsleistungen positive Folgen haben, sonst nimmt die Motivation leicht ab. Der Spielraum für Sanktionen für die Nichterfüllung geforderter Standards ist im derzeitigen System gering, vor allem, wenn es fix Angestellte betrifft. Das Risiko trägt de facto immer der Nachwuchs in prekären Beschäftigungsverhältnissen.
5. Wie weit sich die Qualität einer Forschungsleistung in Leistungsverträgen eindeutig darstellen lässt, ist daher problematisch. Sehr fragwürdig ist vor allem ein

„formelgebundenes Budget“, bei dem Steigerungen oder Senkungen des Budgets sich automatisch aus der Erreichung oder Nichterreichung bestimmter Zahlen (*benchmarks*) ergeben. Besonders problematisch ist dabei die Bindung an Steigerung bzw. Senkung der bisherigen Kennzahlen. Diese ist zwar leicht und fächerspezifisch feststellbar, aber ein solches Verfahren bevorzugt bisher wenig erfolgreiche Forschungseinrichtungen gegenüber Spitzeneinrichtungen, die ihr Potential bereits voll ausschöpfen und leicht (durch Karrieresprünge erfolgreicher MitarbeiterInnen, Auslaufen von Großprojekten etc.) Schwankungen bei den Indikatoren ausgesetzt sind.

6. Sehr wichtig (wie u.a. die britische Erfahrung zeigt) ist es auch, dass Maßnahmen zur „Qualitätssicherung“ nicht an private Firmen, die von Leuten ohne nennenswerte eigene Forschungserfahrung betrieben werden, oder mit Nicht-Wissenschaftlern besetzte Gremien ausgelagert werden, Die Scientific Community muss selbst die Kontrolle ausüben. Dafür ist die ÖAW in ihrer Verbindung von Gelehrten-gesellschaft und Forschungsträgerorganisation besonders gut geeignet. Sie sollte auch ein Forum bieten, Erfolg und unerwünschte Effekte der Bewertung wissenschaftlicher Leistungen zu diskutieren und ständig zu überprüfen.