

Kann der gesellschaftliche Impact von Forschung gemessen werden? Herausforderungen und alternative Evaluationssansätze

Christoph E. Müller & Birge Wolf¹

Measuring societal impacts of research is a challenging task in research evaluation. In this article, we describe several of these challenges with regard to causal inference, time lag, side-effects, operationalization, comparability between disciplines, and availability of required data. We show how different approaches deal with these challenges in evaluation practice and focus on a particular approach named “practice impact” in more detail. This approach includes an improved documentation and is sought to have positive effects on innovation processes and synergies with research and research funding. Moreover, dialogue with different user groups is fostered and serves to make evaluation beyond scientific impact desired, feasible, and efficient.

Forschungsevaluation erhebt, analysiert, interpretiert und bewertet Daten zu Forschungstätigkeiten. Die Evaluation der Forschung von Hochschulen wird zunehmend wichtiger und liegt sowohl im Interesse der staatlichen Kontrollinstanzen, der breiten Öffentlichkeit als auch der Hochschulen selbst (Schiene, 2004). Dies gilt auch für außeruniversitäre Forschungseinrichtungen sowie generell für die Empfänger öffentlicher Forschungsgelder.

Gemäß den Funktionen von Evaluation (Stockmann, 2006) kann die Forschungsevaluation dazu dienen, eingesetzte Ressourcen zu legitimieren, deren angemessene Verwendung zu kontrollieren, sowie Lern- und Entwicklungspotenziale für Forschung und Forschungspolitik zu identifizieren. Darüber hinaus kann ihr auch eine Steuerungsfunktion zugeschrieben werden, indem sie Anreize setzt und Lenkungseffekte hervorruft.

Bislang fokussierte die Forschungsevaluation vor allem auf die Resonanz von Forschung innerhalb des Wissenschaftssystems, häufig als ‚wissenschaftlicher Impact‘ bezeichnet. Der gesellschaftliche Nutzen der Forschung rückt jedoch stärker in den Mittelpunkt. So verweisen Akteure aus Politik, Wissenschaft und Zivilgesellschaft zunehmend auf drängende gesellschaftliche Problemlagen, zu deren Lösung die Forschung Beiträge leisten sollte (Wolf, Häring, & Heß, 2015). Vor dem Hintergrund begrenzter Ressourcen in der Forschungsförderung stellt sich somit die Frage einer Legitimation der verausgabten öffentlichen Mittel (z.B. Bornmann, 2012) und geeigneter Maßnahmen zur Steigerung des gesellschaftlichen Nutzens der Forschung.

¹ Autoren in alphabetischer Reihenfolge.

Eine Forschungsevaluierung mit Bezug zum gesellschaftlichen Impact wird daher zunehmend angestrebt. Zum einen, weil die Erarbeitung umsetzbarer Lösungsbeiträge für gesellschaftliche Problemlagen, wie z.B. in inter- und transdisziplinären Forschungsansätzen, von Indikatoren des wissenschaftlichen Impacts nicht erfasst wird (Wolf et al., 2016; Rafols et al., 2012). Zum anderen, weil ein hoher wissenschaftlicher Impact nicht zwangsläufig einen hohen gesellschaftlichen Impact impliziert und somit auch nicht als stellvertretender Indikator verwendet werden kann (Bornmann, 2013; Donovan, 2011). Zweifelsohne ist jedoch sowohl für die wissenschaftliche als auch die gesellschaftliche Wirkung von Forschung eine hohe wissenschaftliche Qualität erforderlich, sodass es sich nicht um konkurrierende, sondern sich ergänzende Evaluierungsansätze handelt. Darüber hinaus sollte die zweckfreie Grundlagenforschung von diesen Evaluierungsansätzen unberührt bleiben.

Die Messung und Bewertung des gesellschaftlichen Impacts von Forschung ist mit zahlreichen mehr oder weniger schwerwiegenden methodischen und forschungspraktischen Problemen konfrontiert (siehe z.B. Bornmann, 2013), ohne deren Lösung eine effektive Evaluation von Forschungsleistungen kaum möglich ist. Dieser Beitrag widmet sich zunächst einigen dieser grundsätzlichen Herausforderungen. Er stellt zudem Lösungsvorschläge vor und präsentiert mit einem Dokumentationssystem aus dem Projekt „Praxis Impact II“ einen eigenen Ansatz (Wolf et al., 2016, 2014), der unter anderem das Problem der mangelnden Datenverfügbarkeit im Bereich der Forschungsevaluation sowie weitere Probleme bei der Bewertung des gesellschaftlichen Impacts von Forschung in Deutschland zu lösen versucht. Eingangs werden zunächst jedoch grundlegende Merkmale wissenschaftlichen und gesellschaftlichen Impacts skizziert.

1. Wissenschaftlicher und gesellschaftlicher Impact von Forschung

Der wissenschaftliche Impact von Forschung wird im Rahmen der Forschungsevaluation meist anhand klar definierter, numerischer Indikatoren gemessen (z.B. Publikationsoutput in begutachteten Fachzeitschriften oder zitationsbasierte Indizes wie der ‚Journal Impact Factor‘ oder der *h*-index). Diese Indikatoren entstehen aufgrund wissenschaftlicher Kommunikations- und Arbeitsprozesse. Da veröffentlichte Artikel in Fachzeitschriften und zugehörige Zitationen in etablierten Datenbanken wie dem ‚(Social) Science Citation Index‘ oder SCOPUS erfasst werden, können die erforderlichen Kennzahlen zu Zwecken der Forschungsevaluation jederzeit ohne übermäßigen Aufwand abgerufen werden. Streng genommen erfassen diese Indikatoren jedoch nicht den wissenschaftlichen Impact im Sinne einer Wirkung, sondern verschiedene Dimensionen von Produktivität, Qualität und Resonanz innerhalb des Wissenschaftssystems. Denn mit Zitationen wird nicht der tatsächliche Beitrag einer Forschungsaktivität zum Erfolg künftiger Forschung abgebildet, sondern die Kenntnisnahme einer Forschungsarbeit unter Wissenschaftlern.

Der gesellschaftliche Impact geht über die innerwissenschaftliche Resonanz hinaus und beschreibt den gesellschaftlichen Nutzen, der durch Forschungsergebnisse oder daraus entwickelte Outputs hervorgerufen wird (Bornmann, 2013). Dies können beispielsweise ökologische, ökonomische, soziale oder kulturelle Wirkungen sein (Donovan, 2008). Evaluierungskonzepte, die sich auf gesellschaftlichen Nutzen beziehen, umfassen ebenfalls vielfältige Kriterien in den Dimensionen von Produktivität, Qualität und Impact. Die Erhebung von Daten ist aufgrund der Vielfalt der Kriterien meist mit einem deutlich größeren Aufwand verbunden (siehe Abschnitt 2.6). Evaluierungsverfahren, die ergänzend zum wissenschaftlichen Impact auch den gesellschaftlichen Impact berücksichtigen, sind bisher nur in Einzelfällen etabliert (wie z.B. im ‚Research Excellence Framework‘ (REF) in Großbritannien oder im ‚Standard Evaluation Protocol‘ (SEP) in den Niederlanden). Verfahren zur Evaluierung von Forschung sind grundsätzlich nicht unumstritten und weisen sowohl für die Erfassung wissenschaftlichen Impacts als auch gesellschaftlicher Leistungen spezifische Vor- und Nachteile auf.

2. Herausforderungen bei der Messung und Bewertung gesellschaftlicher Impacts

Die dargestellten Herausforderungen ergeben sich aus einem rigorosen Verständnis der Wirkungsevaluation, die das Ziel verfolgt, möglichst fehlerfrei die Nettowirkungen einer sozialen Aktivität zu schätzen. Die Einhaltung der aufgeführten Anforderungen der Wirkungsevaluation steht dabei teilweise in Konflikt mit der Natur gesellschaftlicher Wirkungen von Forschung. Dadurch wird die Wirkungsevaluation jedoch nicht obsolet, sondern es wird der Grundstein gelegt für die Suche nach alternativen Strategien zur Bewertung des gesellschaftlichen Impacts.

2.1 Kausalitäts-/Attribuierungsproblem

Eine zentrale Aufgabe der Wirkungsevaluation von Forschung ist die Identifikation eines kausalen Zusammenhangs zwischen einer Forschungstätigkeit und einer beobachtbaren Veränderung in einem gesellschaftlichen Bereich. Diesen Vorgang bezeichnet man als kausale Attribuierung (z.B. Martin, 2007), die in komplexen Innovationssystemen mit besonderen Herausforderungen verbunden ist. Es ist nämlich davon auszugehen, dass beobachtbare Veränderungen in einem gesellschaftlichen Bereich nicht ausschließlich auf die evaluierte Forschungsaktivität zurückzuführen sind. Eine Vielzahl externer Faktoren sowie auch andere Forschungsaktivitäten können beteiligt sein. Somit besteht die Herausforderung darin, festzustellen, ob ein eindeutiger Zusammenhang zwischen Forschungstätigkeit und Veränderung besteht (siehe Abbildung 1). Weitere Schwierigkeiten liegen in der Quantifizierung des Anteils der Forschungsaktivität an der beobachteten Wirkung. Darüber hinaus beeinflussen einzelne Faktoren im Innovationssystem den Umfang und die spezifische Ausprägung der beobachteten Wirkung. Beispielsweise können Forschungsergebnisse auf vielfältige Art angewendet werden und im jeweiligen Anwendungskontext entsprechend unterschiedliche gesellschaftliche Auswirkungen zur Folge haben (vgl. UNEP, 2012; Miedzinski et al., 2013).

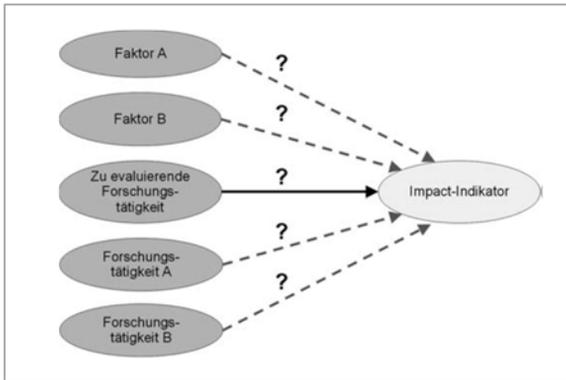


Abbildung 1: Das Attribuierungsproblem

Zur Lösung dieser Herausforderungen müssen aufwändige Untersuchungsanordnungen – beispielsweise Vergleichsgruppendesigns oder Vorher-Nachher-Studien – konzipiert und angewendet werden. Da oftmals keine geeigneten Vergleichsgruppeneinheiten verfügbar sind und Forschungsevaluationen in der Regel erst zum Ende von Forschungsaktivitäten umgesetzt werden, sind solche Designs in der Praxis meist nicht umsetzbar. Der hohe Ressourcenbedarf erschwert die Umsetzung zusätzlich.

2.2 Die Zeitdimension

Der gesellschaftliche Impact von Forschungsaktivitäten kann sich zu unterschiedlichen Zeitpunkten manifestieren, sodass die Festlegung eines geeigneten Zeitpunktes für die Erfassung von Wirkungen eine Herausforderung darstellt. Wirkungen können während des Projekts sowie unmittelbar oder (sehr) lange nach Projektabschluss auftreten (Martin, 2007). Um auch spät eintretende gesellschaftliche Impacts erfassen zu können, empfiehlt Pregernig (2007) daher einen Zeitraum von zehn Jahren zur Nachbeobachtung, das britische REF arbeitet mit einem Zeithorizont von 20 Jahren (HEFCE, 2014). Zudem können Wirkungen über die Zeit hinweg auch nachlassen oder sogar ganz verschwinden. Wird die Forschungsevaluation zu spät durchgeführt, können diese Wirkungen nicht mehr identifiziert werden. Unabhängig davon, welcher Zeitraum von der Forschungsevaluation schlussendlich gewählt wird: Eine absolute Sicherheit für die Wahl des adäquaten Zeitpunktes für die Wirkungserfassung gibt es nicht (Abbildung 2).

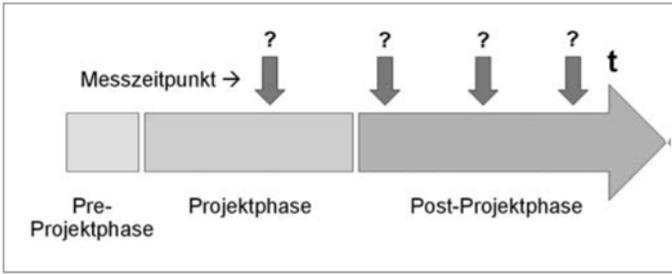


Abbildung 2: Das Problem der Zeitdimension

2.3 Nicht-intendierte Wirkungen

Neben den intendierten Auswirkungen für die Gesellschaft können Forschungsaktivitäten auch nicht-intendierte Wirkungen nach sich ziehen (z.B. Rymer, 2011), die sowohl positiv als auch negativ sein können (Stockmann, 2006; Abbildung 3). Um eine umfangreiche Bewertung des Nutzens einer Forschungsaktivität für die Gesellschaft vornehmen zu können, muss die Forschungsevaluation positive und negative (letztere sind i.d.R. die nicht-intendierten) Wirkungen berücksichtigen.

Die damit einhergehende Problematik ist offensichtlich: Um nicht-intendierte Wirkungen prüfen zu können, muss man wissen, in welchen gesellschaftlichen Bereichen positive oder negative ungeplante Wirkungen überhaupt auftreten können. Aufgrund der Vielfalt möglicher Wirkungsfelder besteht das Risiko, dass potenziell relevante Bereiche nicht berücksichtigt werden.

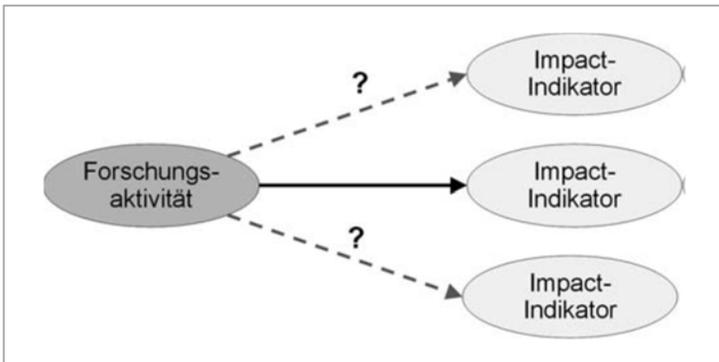


Abbildung 3: Nicht-intendierte Wirkungen („side-effects“)

2.4 Operationalisierung

Um den gesellschaftlichen Impact von Forschung präzise erfassen zu können, muss festgelegt werden, anhand welcher Indikatoren dies erfolgen soll. Dabei sollte ein Indikator SMART sein. Dies bedeutet, „er trifft eindeutig den angestrebten theoretischen Sachverhalt, ist messbar und für den gewünschten Arbeitskontext anwendbar, zur Lösung der anstehenden Probleme relevant und für die Beteiligten nachvollziehbar (...). Ein guter Indikator entspricht dementsprechend sowohl theoretischen, methodischen, praktischen und politischen Anforderungen.“ (Meyer, 2010, S. 24).

Aufgrund der Originalität von Forschungsaktivitäten sowie der großen Vielfalt gesellschaftlicher Wirkungsfelder ergeben sich Probleme bei der Operationalisierung. Einerseits muss bei der Auswahl bereits vorhandener Indikatoren sichergestellt werden können, dass das individuelle Wirkungsfeld der jeweiligen Forschungsaktivität passgenau erfasst wird. Andererseits existieren erforderliche Indikatoren oftmals nicht (z.B. Godin & Doré, 2005), weshalb sie erst aufwändig entwickelt und validiert werden müssen.

2.5 Interdisziplinäre Vergleichbarkeit

Eng mit der Operationalisierung verbunden ist das Problem der interdisziplinären Vergleichbarkeit. Es dürfte einleuchten, dass der gesellschaftliche Impact von Forschungsaktivitäten unterschiedlicher Disziplinen nicht ohne Weiteres anhand derselben Indikatoren gemessen und bewertet werden kann. Bornmann (2013, S. 219) konstatiert daher: „Since it can be expected that the scientific work of an engineer has a different impact than the work of a sociologist or historian and because research results affect many different aspects of society (...), it will hardly be possible to have a single assessment mechanism.“ Die Forschungsevaluation steht damit vor der Herausforderung, die Spezifika unterschiedlicher Disziplinen bei der Bewertung des gesellschaftlichen Impacts angemessen zu berücksichtigen.

2.6 Datenverfügbarkeit

Die Messung und Bewertung des gesellschaftlichen Impacts sind untrennbar mit der Erhebung und Analyse empirischer Daten verbunden. Aufgrund des individuellen Charakters von Forschungsaktivitäten kann die Menge und Qualität verfügbarer Daten stark variieren, je nachdem, welche Forschungstätigkeit evaluiert wird. In der Praxis stellt sich zudem häufig das Problem, dass robuste Daten nicht verfügbar sind (z.B. Spaapen et al., 2011b) und erst erhoben werden müssen. Sofern Primärdaten erhoben werden können, ist deren Qualität nicht zwangsläufig hinreichend. Die weit verbreitete Nutzung von Selbsteinschätzungen aus Interviews oder Fragebögen ist zum Beispiel mit Problemen wie sozialer Erwünschtheit oder Erinnerungseffekten verbunden. In einigen Fällen ist die Erhebung von Daten aufgrund situativer Restriktionen – z.B. kein Datenzugang oder unzureichende Ressourcen – dagegen überhaupt nicht möglich, in vielen anderen mit einem großen Ressourcen- und Zeitaufwand verbunden. Zudem ist die Datenerhebung ineffizient, wenn bestimmte Daten für jede Evaluation neu erhoben werden

müssen, weil die kontinuierliche Erfassung von Informationen nicht etabliert ist. Eine kontinuierliche Erfassung wäre grundsätzlich wünschenswert, erfordert jedoch Ressourcen für den Aufbau von Dokumentationssystemen und ein Wissensmanagement, welches deren zuverlässige Nutzung gewährleistet.

3. Lösungsansätze

Im Folgenden wird beispielhaft dargestellt, wie bestehende Evaluierungsansätze mit Bezug zum gesellschaftlichen Impact mit den Herausforderungen der Wirkungsevaluation umgehen.

Im REF, dem nationalen Evaluierungsverfahren in Großbritannien, stellen Universitätsinstitute ihren positiven gesellschaftlichen Impact in ‚Impact Case Studies‘ dar. Gefordert werden darin der Beleg und die Quantifizierung von Impacts (z.B. externe Belege, Nutzerbefragungen) sowie eine Attribuierung über ‚translational activities‘². Die Impact Case Studies werden von Expertengremien bewertet. Das Bewertungsergebnis hat Einfluss auf 20 % der künftigen Mittelzuweisung, wodurch Anreize gesetzt werden, das Management gesellschaftlicher Impacts und von translational activities sowie deren kontinuierliche Dokumentation zu optimieren (HEFCE, 2014, Vertigo-Ventures et al., 2016). Dafür wurden beispielsweise universitäre Forschungsinformationssysteme um Indikatoren erweitert, die die gesellschaftlichen Leistungen in Anlehnung an das REF abbilden. Auch in der Forschungsförderung wurden Systeme zur digitalen Dokumentation gesellschaftlicher Leistungen eingeführt und es wurde mit Harmonisierungsprozessen zwischen den Datenerfassungssystemen begonnen (Jörg et al., 2014). In Deutschland erfolgt derzeit mit der Spezifikation für einen „Kerndatensatz Forschung“ (IfQ et al., 2015) eine Standardisierung von Forschungsinformationen, die jedoch vorwiegend Indikatoren zur Messung des wissenschaftlichen Impacts enthält.

In Evaluierungskonzepten in der entwicklungsbezogenen Forschung sowie in der Innovations-/Technikfolgenabschätzung wird viel Wert daraufgelegt, intendierte und nicht-intendierte Wirkungen umfassend abzuschätzen. U.a. werden dafür Wirkungspfade und -prozesse mit Stakeholdern reflektiert (z.B. Douthwaite, 2007) und durch vordefinierte mögliche Wirkungsbereiche wird eine umfassende Perspektive in der Bewertung hergestellt (z.B. UNEP, 2012). Die Darstellung eines Zusammenhangs zwischen Forschungsaktivitäten und Wirkungen wird häufig durch die Verwendung von ‚logic models‘³ der Impactentstehung unterstützt, wie z.B. im Payback-Framework (Donovan et al., 2011). Aus dem Feld der Nachhaltigkeitsbewertung wird jedoch auf die Notwendigkeit wesentlich komplexerer Modelle verwiesen, in denen Wechselwirkungen sowie Rebound-Effekte mitberücksichtigt werden (Miedzinski et al., 2013).

² ‚Translational activities‘ umfassen alle Aktivitäten, die Forschungsergebnisse für die Gesellschaft nutzbar machen (z.B. Wissensaustausch, zielgruppenspezifische Veröffentlichungen u.a. Outputs, Verwertungsrechte).

³ Logic Models stellen einen linearen Bezug zwischen Maßnahmen und Resultaten her. Klassisch gliedern sie sich in Input, Aktivitäten, Outputs, Outcomes, Impact.

Eine weitere Möglichkeit für den Umgang mit den Herausforderungen des Impact Assessments besteht darin, nicht direkt den gesellschaftlichen Impact zu evaluieren, sondern mit stellvertretenden (proxy) Indikatoren wie Prozessen und Outputs zu arbeiten, die geeignet erscheinen, gesellschaftlichen Impact hervorzuheben. Diese finden sich beispielsweise in Konzepten zur Evaluierung transdisziplinärer Forschung (z.B. Bergmann et al., 2005). Im Konzept der ‚produktiven Interaktionen‘ werden direkte (z.B. Workshops), indirekte (über Outputs) oder finanzielle (Co-finanzierung) Interaktionen zwischen Forschung und außerwissenschaftlichen Akteuren und deren ‚Produktivität‘ erfasst. “The interaction is *productive* when it leads to efforts by stakeholders to somehow use or apply research results or practical information or experiences” (Spaapen et al., 2011, S. 212). Eng an dieses Verfahren angelehnt ist die Evaluierung gesellschaftlicher Relevanz im SEP in den Niederlanden. Den Universitätsinstituten obliegt dort die Auswahl von Indikatoren, die im Kontext ihrer Zielsetzungen geeignet erscheinen, Outputs der Forschung für gesellschaftliche Gruppen sowie ihre Nutzung und Anerkennung darzulegen (VSNU et al., 2014).

Nachfolgend wird beispielhaft ein Lösungsansatz für die Agrarforschung beschrieben, der die vorgestellten Ansätze mit einbezieht.

4. Der Lösungsansatz von ‚Praxis Impact‘

Die Projekte Praxis-Impact (PI) I und II⁴ waren von der Zielsetzung getragen, ein anwendbares Konzept für die Evaluierung praxisorientierter, gesellschaftlicher Leistungen der Forschung am Beispiel der deutschen Agrarforschung zu entwickeln. Ziel einer solchen Evaluierung ist es, ein großes Engagement der Forschung für einen gesellschaftlichen Nutzen zu fördern und dieses im Wissenschaftssystem stärker anzuerkennen.

Entwickelt wurde ein Konzept für die Evaluierung von Agrarforschungsprojekten, welches sowohl einen Kriterienkatalog als auch einen Vorschlag für die Durchführung der Evaluation unter Einbindung von Stakeholdern umfasst. Um die Umsetzung der Evaluierung zu gewährleisten, wurde ein Dokumentationskonzept entwickelt, welches die Grundlage für die Erfassung von Informationen zu Aktivitäten, Outputs, Beteiligten, Anwendungsmöglichkeiten und (potenziellem) Impact durch Forschende darstellt. Das Dokumentationskonzept ist ein Entwurf, um Forschungsinformationssysteme mit den aufgeführten Informationen zu erweitern und damit Teile von Anträgen und Berichten für den Fördermittelgeber zu ersetzen. Die Entwicklungen des Dokumentations- und Evaluierungskonzeptes basieren auf explorativen Interviews in PI I (Wolf et al., 2013) sowie einer umfangreichen Literaturliteraturanalyse in PI I und II zu bestehenden Konzepten, deren Inhalte kombiniert, anhand ihrer Eignung für die Zielsetzung ausgewählt und mit neuen

⁴ FKZ 2806OE307 und 2812NA102 / 103. Gefördert durch das Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz (BMELV) aufgrund eines Beschlusses des Deutschen Bundestages im Rahmen des ‚Bundesprogramms Ökologischer Landbau und andere Formen nachhaltiger Landwirtschaft‘.

Elementen angereichert wurden. Die Entwicklungen wurden in PI II mit Forschungsförderern diskutiert und mit Forschenden aus dem Agrarbereich erprobt (Wolf et al., 2016).

Nachfolgend wird das Dokumentations- und Evaluierungskonzept entlang seiner Entwicklungsprämissen dargestellt, namentlich die Umsetzbarkeit für die deutsche Agrarforschung und der positive Einfluss auf Innovationsprozesse.

4.1 Umsetzbarkeit der Evaluierung

Um erweiterte Verfahren der Forschungsevaluierung umsetzen zu können, lag der Fokus von PI II darauf, hierfür erforderliche Informationen in angemessener Qualität mit vertretbarem Aufwand zu erfassen bzw. bereits vorhandene Dokumentationsprozesse nutzbar zu machen. Ein gutes Beispiel für die Nutzung vorhandener Dokumentationsprozesse ist die Erfassung der Indikatoren für den wissenschaftlichen Impact. Diese werden nicht explizit für eine Evaluierung erhoben, sondern entstehen über die elektronische Erfassung von Publikationen und Zitationen als „Nebenprodukt“ wissenschaftlicher Arbeit und Kommunikation.

Mit dem in PI II entwickelten Konzept sollen Förderprozesse der Bundesforschung für die Dokumentation nutzbar gemacht werden, da dort ein Teil der für die Evaluation notwendigen Informationen bereits erfasst wird. Die Förderprogramme der Bundesministerien fokussieren häufig auf praktische/gesellschaftliche Problemlagen und beinhalten Kooperations- und Transferprozesse mit außerwissenschaftlichen Akteuren, sodass damit verbundene Informationen von den Förderern in Abschlussberichten und der obligatorischen Verwertungsplanung⁵ gewünscht und gefordert werden. Diese Textdokumente sind bislang nicht nach Evaluierungskriterien strukturiert und zu späteren Zeitpunkten auftretende Wirkungen werden nur bedingt erfasst.

In PI II wurden bestehende Dokumentationsansätze für gesellschaftliche Leistungen (s.o.) aufgegriffen und um Evaluierungskonzepte erweitert, die für die Agrarforschung als geeignet eingestuft wurden. Die Dokumentation wurde darauf ausgerichtet, Anforderungen der Forschungsförderung zu erfüllen, um damit Teile von Anträgen und Berichten ersetzen zu können. Dies dient dazu, den Dokumentationsaufwand für Forschende zu begrenzen und die Nutzbarkeit der Informationen für die Förderprozesse zu gewährleisten.

Konkret wird mit dem Dokumentationskonzept mehreren methodischen Herausforderungen bei der Bewertung gesellschaftlicher Wirkungen von Forschung begegnet. Die Operationalisierung wird unterstützt, indem für ein breites Kriterienspektrum definiert wurde, welche Attribute zur Beschreibung von Aktivitäten, Outputs oder Wirkungen notwendig sind. Hierdurch können sie auf einer einheitlichen Basis bewertet und somit besser verglichen werden. Nichtsdestotrotz erlaubt das Dokumentationskonzept die Abbildung der individuellen Wirkungspfade von Projekten, die sich in unterschiedlichen Leistungen zeigen können.

⁵ Die Verwertungsplanung umfasst sowohl die Nutzung(smöglichkeiten) der Ergebnisse durch die Konsortialpartner als auch durch Dritte.

Eine digitale Erfassung von Forschungsleistungen ist kontinuierlich und über das Projektende hinaus möglich. Sie wird damit den unterschiedlichen und oftmals langen Zeiträumen der Entwicklung gesellschaftlicher Wirkungen gerecht. Da die Dokumentation unmittelbar zum Zeitpunkt von Aktivitäten, Outputs oder beobachteten Wirkungen erfolgen kann und einer definierten Struktur folgt, wird eine höhere Qualität der Informationen erreicht als z.B. durch Interviews im Rahmen einer ex-post Evaluation. Bei Informationen, die in der Projektberichterstattung genutzt werden (z.B. Aktivitäten oder Verwertungspläne), findet zusätzlich eine Qualitätssicherung/Verifizierung durch den Förderer statt.

Um den Wissenstransfer zu unterstützen und die Dokumentationsanforderungen der Förderer zu erfüllen, umfasst die Dokumentation Outputs für und Aktivitäten mit außerwissenschaftliche(n) Akteure(n). Diese Informationen unterstützen die Ableitung eines plausiblen Zusammenhanges zwischen Projekt und Wirkungen – wenngleich sie für eine eindeutige Attribuierung von Wirkungen noch nicht hinreichend sind.

4.2 Positiver Einfluss des Dokumentations- und Evaluierungskonzeptes auf Innovationsprozesse

Mit dem Ziel von PI II, dass die Evaluierung einen positiven Einfluss auf Innovationsprozesse hat, gehen weitgehende Implikationen für die Frage der Impacteffizienz und die damit verbundenen Herausforderungen einher.

Wenn durch die Evaluierung Impulse für vermehrte gesellschaftliche Leistungen der Forschung gesetzt werden sollen, muss der Evaluierungszeitraum so gewählt werden, dass Evaluierungsergebnisse für die Forschenden relevant sind (z.B. für ihren Karriereweg nutzbar). Dies spricht für einen Evaluierungszeitraum von bis zu drei Jahren nach Projektende und die Honorierung produktiver Interaktionen (z.B. genutzte Outputs). Produktive Interaktionen können sich in einem solchen Zeitraum bereits entfalten, wohingegen Impacts in diesem Zeitraum oftmals noch nicht erfassbar sind.

Projekte bieten sich als Evaluierungsgegenstand für gesellschaftliche Leistungen besonders an, da der Bezug zu einer spezifischen Forschungstätigkeit hergestellt wird und produktive Interaktionen vielfach durch Kooperation verschiedener Partner in einem Projekt entstehen. Positiv evaluierte Projekte sollten somit allen am Projekt beteiligten Forschungseinrichtungen als ‚Teamerfolg‘ zugerechnet werden, weil eine weitere Attribuierung auf die einzelnen beteiligten Institutionen nicht nur schwierig ist, sondern auch Kooperationsprozesse konterkarieren kann. Das Wissenschaftsmanagement von Forschungseinrichtungen profitiert jedoch nicht nur von den Evaluierungsergebnissen, sondern könnte über Schnittstellen oder Exportfunktionen auch die Inhalte des Dokumentationssystems nutzen (unter Berücksichtigung der informationellen Selbstbestimmung der Forschenden).

Wenn in der Evaluierung über produktive Interaktion hinaus auch gesellschaftliche Impacts erfasst werden sollen – beispielsweise um erweiterte Erkennt-

nisse über Förderprogramme oder Innovationssysteme zu erlangen – ist es zielführend, die komplexen Wechselwirkungen in Innovationssystemen zu berücksichtigen⁶, anstatt von einer linearen Ursache-Wirkungsbeziehung auszugehen.

Die in diesem Abschnitt bislang dargestellten Sachverhalte werden im PI II-Konzept berücksichtigt, indem

- a) in der Dokumentation Prozesse, Outputs sowie potenzielle und de facto Anwendungen und Wirkungen erfasst werden können,
- b) die Projektevaluierung auf produktive Interaktionen, eine Bewertung der Anwendbarkeit von Forschungsergebnissen und Outputs sowie potenzielle Wirkungen fokussiert.

Zudem werden Rahmenbedingungen erfasst, die die tatsächliche Anwendung von Forschungsergebnissen fördern oder hemmen können (und somit oft entscheidend dafür sind, ob vom Grundsatz her anwendbare Lösungsbeiträge der Forschung tatsächlich angewendet werden). Eine Berücksichtigung von Rahmenbedingungen unterstützt es somit, Projekte fair zu bewerten und zielführende Lenkungseffekte zu erzeugen: Denn Forschung sollte sich auch gesellschaftlichen Fragestellungen widmen, bei denen die Implementierung von Lösungsbeiträgen von ungünstigen Rahmenbedingungen geprägt ist. Dies kann zum Beispiel bei Nachhaltigkeitsfragen der Fall sein.

Die Produktivität einer Interaktion kann im PI II Dokumentations- und Evaluierungskonzept u.a. über Nutzungsdaten erfasst werden. Dies ist möglich, weil das Dokumentationskonzept vorsieht, dass Outputs der Forschung in einem erweiterten Forschungsinformationssystem zugänglich gemacht werden. Dies kann in Form von Uploads oder der Verlinkung von Informationen erfolgen und unterstützt die Umsetzung von Open Access-Strategien. Die Produktivität einer Interaktion im Sinne von „es gibt Akteure, die an der Nutzung ein Interesse haben“ wird durch eine inhaltliche Bewertung ergänzt. Diese umfasst die Beurteilung der Anwendbarkeit bzw. der Vor- und Nachteile einer tatsächlichen Anwendung sowie das zugehörige gesellschaftliche Wirkungspotenzial in den vier Dimensionen der Nachhaltigkeit (ökologisch, ökonomisch, sozial, kulturell). In diese Einschätzung werden auch außerwissenschaftliche Akteure eingebunden. Sie ist über qualitative Feedbacks bereits im Dokumentationskonzept implementiert und der Bewertungsprozess soll gleichermaßen durch wissenschaftliche und außerwissenschaftliche Akteure durchgeführt werden. Damit kann das Evaluierungskonzept einen positiven Einfluss auf Forschungs- und Innovationsprozesse ausüben, weil der (Impact fördernde) Dialog zwischen Wissenschaft und Praxis/Gesellschaft verstärkt wird. Darüber hinaus bietet die fundierte Einschätzung der Anwendbarkeit von Forschungsergebnissen für Praxisakteure eine Orientierung für die Erprobung von innovativen Lösungen.

Somit begegnet das beschriebene Konzept mehreren Herausforderungen der Impact-Bewertung: Die präzise Erfassung von Wirkungen und ihre Attribuierung

⁶ Vgl. hierzu das Special Issue der Zeitschrift “Historical Social Research” (2015, Jg. 40, Nr. 3) mit dem Titel “Methods of Innovation Research: Qualitative, Quantitative and Mixed Methods Approaches”.

auf die zu evaluierende Forschungsaktivität sind für den Ansatz nicht erforderlich. Stattdessen erfolgt die Betrachtung eines breiten Spektrums von Wirkungspotenzialen, deren Plausibilität durch die verschiedenen Perspektiven aus Wissenschaft und Praxis/Gesellschaft gesteigert wird. So kann auch die Wahrscheinlichkeit erhöht werden, negative bzw. nicht-intendierte Wirkungen bzw. Wirkungspotenziale von Innovationen zu erfassen. Darüber hinaus wird mit diesem Vorgehen angemessen berücksichtigt, dass gesellschaftliche Wirkungen nicht nur durch die Forschungsergebnisse als solche determiniert werden, sondern in Abhängigkeit von der spezifischen Anwendung unterschiedlich ausgeprägt sein können. Zusammen mit der Berücksichtigung von Rahmenbedingungen können die Evaluierungsergebnisse entsprechend dazu beitragen, Innovationsprozesse gezielt im Hinblick auf gesellschaftlichen Nutzen zu optimieren und strategische Entwicklungen im Wissenschaftsmanagement und in der Forschungsförderung zu unterstützen.

Mit dem Dokumentations- und Evaluierungskonzept wird zudem die Operationalisierung unterstützt, da zum einen die vielfältigen, überwiegend qualitativen Indikatoren standardisiert dokumentiert werden. Zum anderen wird mit dieser Vielfalt adäquat und in vergleichbarer Weise umgegangen, indem mit fachlichem Sachverstand aus mehreren Perspektiven eine Bewertung in übergeordneten Kategorien erfolgt. Die übergeordneten Bewertungskategorien sowie Indikatorenbeispiele sind in Abbildung 4 dargestellt. Vermutlich lassen sich die übergeordneten Bewertungskategorien in verschiedenen Disziplinen oder Forschungsfeldern einsetzen (interdisziplinäre Vergleichbarkeit), während die in darunterliegenden Ebenen angesiedelten Kriterien und Indikatoren spezifisch auf eine Disziplin ausgerichtet sein können.

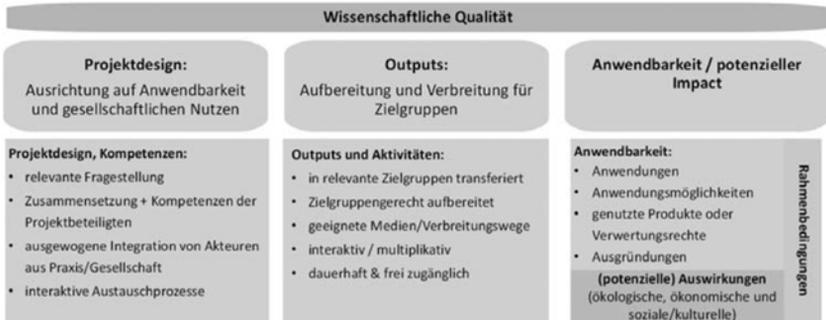


Abbildung 4: Bewertungskriterien und Indikatorenbeispiele

Schließlich schützt die Kombination aus dem Einsatz vielfältiger Indikatoren und einem Begutachtungsverfahren davor, einzelne Indikatoren zu manipulieren. Hierdurch wird die dauerhafte Nutzbarkeit des Konzeptes gesichert.

5. Schlussfolgerungen

Der vorliegende Beitrag hat unterschiedliche, mehr oder weniger gewichtige Herausforderungen aufgezeigt, die mit der Messung und Bewertung gesellschaftlicher Impacts von Forschung verbunden sind. Es gibt jedoch zahlreiche Ansätze, die sich auf unterschiedliche Art und Weise mit der Lösung dieser Herausforderungen auseinandersetzen.

Das PI II-Konzept fokussiert darauf, produktive Interaktionen, Anwendbarkeit und potenzielle Impacts in einer Weise zu evaluieren, die praktisch umsetzbar ist und Innovationsprozesse fördert. Die Orientierung an diesen beiden Zielsetzungen ist gleichbedeutend mit einem Abschied vom ‚messbaren Impact‘ gemäß einem rigorosen Verständnis der Wirkungsevaluation, weil dessen Anforderungen nur bedingt umsetzbar und den Zielsetzungen teilweise nicht zuträglich sind.

Die tatsächliche Umsetzung des PI II-Konzeptes ist maßgeblich von einem Interesse der Forschungsförderer abhängig. Dabei lassen sich neue Konzepte umso leichter etablieren, je mehr Akteure einen Nutzen davon haben. Investitionen, die das Engagement der Forschung für einen gesellschaftlichen Nutzen fördern und honorieren, sollten sich somit nicht nur für die Forschungsförderer auszahlen, sondern für möglichst viele Akteure einen Nutzen entfalten. Entsprechend wurden (und werden weiterhin) die Perspektiven von Förderern, Forschenden und Forschungseinrichtungen sowie Akteuren aus Praxis und Gesellschaft bei der Weiterentwicklung einbezogen.

Angaben zu den Autoren

Christoph E. Müller ist Forschungsreferent am Institut für Gesetzesfolgenabschätzung und Evaluation am Deutschen Forschungsinstitut für öffentliche Verwaltung in Speyer. Seine Forschungsinteressen liegen in den Bereichen Evaluationsforschung, Wirkungsmessung und Methoden der empirischen Sozialforschung.

Birge Wolf ist wissenschaftliche Mitarbeiterin an Fachbereich Ökologische Agrarwissenschaften der Universität Kassel. Ihre Forschungsinteressen liegen in der Evaluation praxisorientierter Agrarforschung.

Kontakt:

Dr. Christoph E. Müller
Institut für Gesetzesfolgenabschätzung und Evaluation
Deutsches Forschungsinstitut für öffentliche Verwaltung Speyer
Freiherr-vom-Stein-Str. 2
67346 Speyer
Telefon: +49 6232 654-291
Mail: mueller@foev-speyer.de
URL: <http://www.foev-speyer.de/de/beratung/ingfa.php>

Birge Wolf
Universität Kassel
Fachbereich Ökologische Agrarwissenschaften
Nordbahnhofstr. 1 a
37213 Witzenhausen
Telefon: +49 5542 981536
Mail: birge.wolf@uni-kassel.de

Portraitfotos



Referenzen

- Bergmann, M., Brohmann, B., Hoffmann, E., Loibl, M. C., Rehaag, R., Schramm, E., Voß, J.-P. (2005):* Qualitätskriterien transdisziplinärer Forschung. Ein Leitfaden für die formative Evaluation von Forschungsprojekten. Frankfurt am Main: Institut für sozial-ökologische Forschung.
- Borrmann, L. (2013):* What is social impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, 64(2), 217–233.
- Borrmann, L. (2012):* Measuring the societal impact of research. *EMBO reports*, 13(8), 673–676.
- Donovan, C. (2011):* State of the art in assessing research impact: Introduction to a special issue. *Research Evaluation*, 20(3), 175–179.
- Donovan, C. (2008):* The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. *New Directions for Evaluation*, 2008(118), 47–60.
- Donovan, C., & Hanney, S. (2011):* The 'Payback Framework' explained. *Research Evaluation*, 20(3), 181–183.
- Douthwaite, B. (2007):* Participatory impact pathways analysis: A practical application of program theory in research-for-development. *Canadian Journal of Program Evaluation*, 22(2), 127–159.
- Godin, B., & Doré, C. (2005):* Measuring the impacts of science: Beyond the economic dimension. Helsinki: Helsinki Institute for Science and Technology Studies. http://www.csiic.ca/PDF/Godin_Dore_Impacts.pdf
- HEFCE (Higher Education and Funding Council for England) (2014):* REF 2014 Key facts. <http://www.ref.ac.uk/media/ref/content/pub/REF%20Brief%20Guide%202014.pdf>
- Institut für Forschungsinformation und Qualitätssicherung (iFQ), Fraunhofer-Institut für Angewandte Informationstechnik (FIT), & Geschäftsstelle des Wissenschaftsrates (2015):* Spezifikation des Kerndatensatz Forschung - Version 1.0. http://www.kerndatensatz-forschung.de/version1/Spezifikation_KDSF_v1.pdf
- Jörg, B., Waddington, S., Jones, R., & Trowell, S. (2014):* Harmonising research reporting in the UK – Experiences and outputs from UKRISS. *Procedia Computer Science* 33, 207–214.
- Martin, B. R. (2007):* Assessing the impact of basic research on society and the economy. Paper presented at the Rethinking the impact of basic research on society and the economy (WF-EST International Conference, 11 May 2007), Vienna, Austria.
- Meyer, W. (2010):* Indikatorenentwicklung: Eine praxisorientierte Einführung. CEval-Arbeitspapier Nr. 10. Saarbrücken: Centrum für Evaluation. http://www.ceval.de/modx/fileadmin/user_upload/PDFs/workpaper10.pdf
- Miedzinski, M., Allinson, R., Arnold, E., Harper, J. C., Doranova, A., Giljum, S., Griniece, E., Kubezko, K., Mahieu, B., Markandya, A., Peter, V., Ploeg, M.,*

- Stasiakowska, A., & van der Veen, G. (2013):* Assessing environmental impacts of research and innovation policy. Study for the European Commission. Brussels: Directorate-General for Research and Innovation.
- Pregernig, M. (2007):* Wirkungsmessung transdisziplinärer Forschung: Es fehlt der Blick aus der Distanz. *Gaia*, 16(1), 46–51.
- Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012):* How journal rankings can suppress interdisciplinary research: A comparison between *Innovation Studies* and *Business & Management*. *Research Policy*, 41(7), 1262–1282
- Rymer, L. (2011):* Measuring the impact of research: The context for metric development. Turner, Australia: The Group of Eight.
- Schiene, C. (2004):* Forschungsevaluation als Element der Qualitätssicherung an Hochschulen. *Zeitschrift für Evaluation*, 3(1), 81–94.
- Spaapen, J., van Droge, L., Propp, T., van der Meulen, B., van den Besselaar, P., & Barker, K. et al. (2011a):* SIAMPI final report. www.siampi.eu/Content/SI-AMPI_Final%20report.pdf
- Spaapen, J., & van Drooge, L. (2011b):* Introducing 'productive interactions' in social impact assessment. *Research Evaluation*, 20(3), 211–218.
- Stockmann, R. (2006):* Evaluation und Qualitätsentwicklung: Eine Grundlage für wirkungsorientiertes Qualitätsmanagement. Münster: Waxmann.
- UNEP (United Nations Environment Programme) (2012):* Application of the sustainability assessment of technologies methodology. Guidance manual. <http://www.unep.org/ietc/InformationResources/Publications/SustainabilityAssessmentofTechnologyManual/tabid/106701/Default.aspx>
- Vertigo-Ventures, Digital Science (2016):* Collecting research impact evidence. Best practice guidance for the research community. <http://www.vertigoven-ures.com/single-post/2016/06/15/Collecting-Research-Impact-Evidence>
- VSNU (Association of Universities in the Netherlands), NOW (Netherlands Organisation for Scientific Research), & KNAW (Royal Netherlands Academy of Arts and Sciences) (2014):* Standard Evaluation Protocol 2015 – 2021. Protocol for research assessments in the Netherlands. <https://www.knaw.nl/shared/resources/actueel/publicaties/pdf/standard-evaluation-protocol-2015-2013-2021>
- Wissenschaftsrat (2013):* Empfehlungen zu einem Kerndatensatz Forschung. Wissenschaftsrat. Berlin. <http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf>
- Wolf, B., Häring, A.-M., & Heß, J. (2015):* Strategies towards evaluation beyond scientific impact. Pathways not only for agricultural research. *Organic Farming*, 1(1), 3–18.
- Wolf, B., Lindenthal, T., Szerencsits, M., Holbrook, J. B., & Heß, J. (2013):* Evaluating research beyond scientific impact - How to include criteria for productive interactions and impact on practice and society. *Gaia*, 22(2), 104–114.

-
- Wolf, B., Szerencsits, M., Gaus, H., Müller, C.E., & Hess, J. (2014):* Developing a documentation system for evaluating the societal impact of science. *Procedia Computer Science*, 33, 289–296.
- Wolf, B., Szerencsits, M., Gaus, H., Müller, C., Heß, J., & Stockmann, R. (2016):* Weiterentwicklung und Erprobung eines Konzeptes zur Dokumentation und Evaluierung von Leistungen der Agrarforschung für Praxis und Gesellschaft. Projektabschlussbericht: <http://orgprints.org/30699/>