

*Eckpunkte der Evaluierung von
Forschungs- und Technologiepolitik*

Januar 2004

Endbericht („Modul I“) im Rahmen des Projektes

Zusammenarbeit des Rates für Forschung und Technologieentwicklung
mit der Plattform Forschungs- und Technologieevaluierung

Kontakt:

Plattform Forschungs- und Technologieevaluierung GesbR

Wiedner Hauptstraße 76, A-1040 Wien

office@fteval.at

www.fteval.at

Dezember 2003

"Nun hat der ständige Lebensaufenthalt in einem wohlgeordneten Staat aber durchaus etwas Gespenstisches: man kann weder auf die Straße treten, noch ein Glas Wasser trinken oder die Elektrische besteigen, ohne die ausgewogenen Hebel eines riesigen Apparates von Gesetzen und Beziehungen zu berühren, sie in Bewegung zu setzen oder sich von ihnen in der Ruhe seines Daseins erhalten zu lassen; man kennt die wenigsten von ihnen, die tief ins Innere greifen, während sie auf der anderen Seite sich in ein Netzwerk verlieren, dessen ganze Zusammensetzung überhaupt noch kein Mensch entwirrt hat ..." (Robert Musil, Der Mann ohne Eigenschaften)

Einleitung

Ziel dieses im Zuge der Zusammenarbeit der Plattform Forschungs- und Technologieevaluierung und des Rates für Forschung und Technologieentwicklung entstandenen Papiers „Eckpunkte der Evaluierung in Forschungs- und Technologiepolitik“ ist es, die wichtigsten Konzepte von Evaluierungen in diesen Politikbereichen darzustellen. Nachfolgend werden nun die Ziele, Methoden und Abläufe von Evaluierungen und der Einsatz unterschiedlicher Instrumente für unterschiedliche Ziele und Zwecke dargestellt.

Zweck dieses Dokumentes ist es nicht zuletzt, zu einem gemeinsamen Wissensstand hinsichtlich der Möglichkeiten, Arten, Methoden etc. von Evaluierungen zu kommen. Dies ist Voraussetzung dafür, dass gemeinsame Diskussionen (und Dispute) auf einer vergleichbaren Grundlage beruhen.

Weitere Dokumente – etwa die **Standards** der Evaluierung in Österreichs Forschungs- und Technologiepolitik der Plattform – hängen eng mit diesem Papier zusammen. Es soll eine gemeinsame, d.h. verständliche und gemeinsam getragene Basis für die Bemühungen um die Hebung der Evaluierungskultur in Österreich geschaffen werden.

Das Projekt „Zusammenarbeit des Rates für Forschung und Technologieentwicklung mit der Plattform Forschungs- und Technologieevaluierung“ besteht aus drei Modulen:

- Modul I: „Eckpunkte der Evaluierung in Forschungs- und Technologiepolitik“
- Modul II: „Evaluierung der Sondermittelprogramme. Bausteine einer Evaluierungsstrategie für den Rat für Forschung und Technologieentwicklung“
- Modul III: „Vorschläge für den Rat, wie generell die Evaluierungskultur in Österreich verbessert werden könnte“.

Themen

Einleitung	3
Themen.....	4
Die Evaluierung von Forschungs- und Technologiepolitik.....	5
Funktionen von Evaluierung.....	5
Arten der Evaluierung	6
Zeitliche Dimension der Evaluierung.....	7
Evaluierungsgegenstand.....	7
Evaluierung im Policy Cycle.....	8
Methoden der Evaluierung	9
Evaluierung im Kontext anderer Formen der Bewertung und wissenschaftlicher Unterstützung im Politikprozess	13
Anhang: Wo stehen wir in Österreich? Auf welchen „Standards“ beruht die laufende Evaluierungsarbeit?	15

Die Evaluierung von Forschungs- und Technologiepolitik

Forschungs- und Technologiepolitik umfasst alle öffentlichen Initiativen zur Gestaltung von Forschungs- und Innovationssystemen - d.h. neben öffentlichen Programmen, Politiken, Strategien und Regulierungen auch die "Landschaft" der forschenden und Technologie entwickelnden Institutionen. Die Besonderheit der Forschungs- und Technologiepolitik im Vergleich zu anderen Politikfeldern besteht darin, dass Programme, Politiken und Institutionen sich aus bestimmten Phänomenen des Markt- und Systemversagens heraus legitimieren, die der Forschung, technologischer Entwicklung und innovativer Dynamik inhärent sind. Markt- und Systemversagen entsteht in erster Linie dadurch, dass Wissen und Forschungsergebnisse nur schwer von den ErfinderInnen exklusiv genutzt werden können und dadurch, dass Erfindung, Anwendung und Verbreitung von Innovationen in komplexen Systemen von Politik, Forschung und Markt passieren, die insgesamt hoch störungsanfällig sind.

Forschungs- und Technologiepolitik soll zeigen können, dass sich die Investitionen in sie lohnen. EntscheidungsträgerInnen im Bereich der österreichischen Forschungs- und Technologiepolitik benötigen – nicht zuletzt in einem europäischen Kontext - Instrumente und Techniken, die dem steigenden Informations- und Lernbedarf z. B. über die Funktionsweise des österreichischen Forschungs- und Innovationssystems, die Qualität der österreichischen Forschung und Entwicklung im internationalen Vergleich und die Wirksamkeit der Forschungs- und Technologieförderung Rechnung tragen. So ist die Schaffung und Stärkung eines europäischen Forschungsraumes als Instrument zur Erreichung "der wettbewerbsfähigsten und dynamischsten wissensbasierten Ökonomie" bis 2010 (Ziele von Lissabon) die konsensuale Vorgabe aller Mitgliederstaaten der Europäischen Union. Dies bedeutet, dass die "österreichische Forschungslandschaft" zur Erreichung die Ziele von Lissabon verstärkt auf den Europäischen Forschungsraum auszurichten sein wird, wobei Evaluierung ein zentrales Element sein kann.

Eine entwickelte Evaluierungskultur ist zentraler Bestandteil einer lernenden und strategisch orientierten Forschungs- und Technologiepolitik. Sie ist gleichsam Voraussetzung und Folge einer guten Politikgestaltung. Gut heißt hier: Effizient, fair, transparent.

Funktionen von Evaluierung

Man kann Evaluierung so definieren: Es geht um methodisch basierte Analyse und Bewertung der Eignung forschungs- und technologiepolitischer Annahmen und Zielsetzungen sowie der daraus abgeleiteten Maßnahmen, deren Wirkung und Zielerreichung. Unterschiedliche Arten von Evaluierungen haben unterschiedliche Funktionen und Auswirkungen. Gemeinsam ist ihnen, dass eine ganze Reihe von Aufgaben im Politikzyklus durch bzw. mit Hilfe von Evaluierungen bewältigt werden können. Evaluierung kann hierbei, abhängig von den jeweiligen Rahmenbedingungen, verschiedene Funktionen erfüllen:

- eine **Legitimationsfunktion** – z. B. eine Rechtfertigung der Verwendung von öffentlichen Mitteln.
- eine **Informationsfunktion** - eine Information an die Öffentlichkeit über den Umgang mit und die Verwendung von öffentlichen Mitteln,
- eine informationsorientierte **Lernfunktion** für ProgrammträgerInnen, TrägerInnen technologiepolitischer Entscheidungen, WissenschaftlerInnen etc.,
- eine **Lenkungsfunktion** für die zukünftige Gestaltung von Politikzielen, von Maßnahmenplanung etc.

- und eine **Kontrollfunktion** im Sinne eines Controllings eines privatwirtschaftlichen Unternehmens.

Auch kann Evaluierung eine moderierende Funktion zwischen den wetteifernden Interessen unterschiedlicher Akteure der Forschungs- und Technologiepolitik erfüllen.

Diese Funktionen treffen grosso modo auf alle Arten, Ebenen und Zeitpunkte von Evaluierungen zu, freilich in unterschiedlicher Ausprägung und Auswirkung.

Arten der Evaluierung

Neben den verschiedenen Funktionen sind auch unterschiedliche Arten der Evaluierung zu unterscheiden:

- **Konzeptevaluierungen** überprüfen Mission, Annahmen, grundlegende Thesen und Rahmenbedingungen der Programme (Institutionen, Projekte). Bewertet wird hierbei, wie berechtigt der Einsatz der Politik, des Programms/Projektes (inklusive der darin enthaltenen Lösungsansätze) zur Bekämpfung der vorher identifizierten Probleme ist. Konzeptevaluierungen sind typischerweise Elemente von ex ante Evaluierungen.
- **Designevaluierungen** beschäftigen sich mit der Zweckmäßigkeit der Ausgestaltung eines Programms und seiner organisatorischen Struktur. Dabei wird überprüft, inwieweit die vorgesehenen operationalen Ziele den jeweiligen Problemsituationen entsprechen, ob die eingesetzten Instrumente diesen Zielsetzungen auch angemessen sind und die angestrebten Ziele auch erreicht werden können bzw. konnten. Darüber hinaus soll auch die organisatorische Struktur, also alle geplanten Aktionen sowie deren Abfolge hinsichtlich ihrer zeitlichen Konsistenz sowie der Übereinstimmung mit dem gewählten Lösungsansatz bewertet werden. Designevaluierungen sind typischerweise Elemente der Evaluierung von Programmen zu allen Zeitpunkten (ex ante, interim und ex post).
- Das Abrücken der Evaluierung von einem blinden Gegenüberstellen der Situation vor und nach einer Maßnahme ermöglicht es, im Rahmen der **Prozessevaluierung** die Durchführung v. a. eines Programms im Detail zu untersuchen. Themen sind hier die Qualität der Implementierung, die laufende Kontrolle des Erfolgs sowie Effizienz und Effektivität der Abläufe. Prozessevaluierungen sind typischerweise Elemente von interim Evaluierungen und stehen bei Programmen mit hohen Anforderungen an das Programm- und Projektmanagement im Vordergrund.
- **Wirkungsanalysen** schließlich bewerten die Erreichung der Programmziele (Effektivitätsanalyse) und versuchen alle durch das Programm direkt oder indirekt ausgelösten Effekte zu identifizieren und möglichst auch in ökonomischen Einheiten zu quantifizieren (z. B. induzierte Umsatzsteigerungen etc.). Im Zentrum steht die Erfassung der sozialen, über die privaten hinausgehenden, Erträge des Programms/Projekts. Wirkungsanalysen sind wichtige Elemente von ex-post Evaluierungen.

Zeitliche Dimension der Evaluierung

Eine gebräuchliche Einteilung von Evaluierungen setzt am Zeitpunkt der Durchführung in Bezug zum Programmverlauf an:

- **Ex ante Evaluierungen** setzen vor Beginn (des Programms / Projekts / der Politik, der Institution) an und haben einen zukunftsgerichteten Charakter. Sie verbessern die innere Architektur vor allem von Programmen und erweitern die Steuerungsmöglichkeiten innerhalb dieser Vorhaben.
- **Interim Evaluierungen** werden während der Laufzeit eines Programms/ Projektes/ einer Institution/ Politik durchgeführt. Sie sind entweder begleitend über die gesamte Laufzeit angelegt (häufig auch: formative Evaluierung) oder finden zu festgelegten Zeitpunkten statt.
- **Ex post Evaluierungen** starten nach Programm- bzw. Projektende (Ablauf einer Politik, Auslaufen einer Institution) und betrachten die Entwicklung, Wirkungen und Kosten bzw. Nutzen (sofern möglich) retrospektiv. Dabei kann auch zwischen ex-post Evaluierungen direkt nach Abschluss eines Programms oder Projekts und **Backward-look Evaluierungen** einige Jahre später unterschieden werden.

Evaluierungsgegenstand

Evaluierung ist ein sehr weitreichendes Konzept. Evaluierungen werden hauptsächlich dort eingesetzt, wo Märkte fehlen, also im innerbetrieblichen Bereich oder in der Politik. In beiden Fällen erlebten Evaluierungen in den letzten Jahrzehnten einen enormen Aufschwung, da sich hoheitliche bzw. direkte Zugriffe als inadäquat oder als ineffizient erwiesen haben. Im Kontext der Forschungs- und Technologiepolitik können die folgenden Einheiten Untersuchungsgegenstand von Evaluierungen sein:

- **Politik** ist ein Bündel von Aktivitäten (Programme, Verfahren, Regelungen etc.), welche in ihrer Art zwar unterschiedlich und verschiedenartig ausgestaltet sein mögen, die aber ein gemeinsames Motiv oder Ziel haben. Im Gegensatz zu Projekten und Programmen ist eine Politik in der Regel nicht zeitlich oder haushaltsmäßig begrenzt.
- **Institution** - Dabei wird auf den (meist dauernd angelegten) Charakter physischer Strukturen abgestellt. Jede Art von Institutionen kann Gegenstand von Evaluierungen sein. In unserem Zusammenhang sind vor allem drei große Blöcke zu erwähnen: Universitäten und Fachhochschulen¹ mit ihrer Verknüpfung von Forschung und Lehre, Forschungseinrichtungen und Förderungseinrichtungen bzw. Agenturen.
- Ein **Programm** ist eine Kombination von intentional aufeinander bezogener Interventionen, Projekten, Maßnahmen oder Teilprogrammen, die auf das Erreichen

¹ Universitäten und Fachhochschulen haben eigene Evaluierungsfunktionen und -strukturen. Die Plattform grenzt sich hier insofern ab, dass sie sich ausschließlich mit Aspekten der Forschungsevaluierung auseinandersetzt, nicht mit jenen der Lehre.

eines spezifischen, vordefinierten Zieles abzielen. Ein Programm ist mit einer festen Laufzeit ausgestattet, verfügt über ein zugeordnetes, häufig zentral verwaltetes Budget und eine klare Struktur.

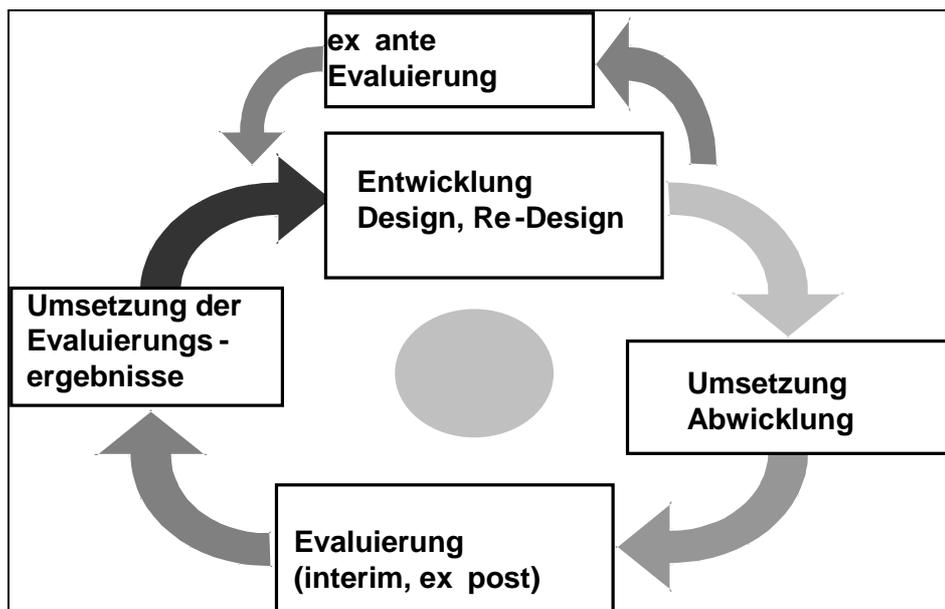
- Ein **Projekt** ist eine einzelne, nicht teilbare Maßnahme, ausgestattet mit einem fixen Zeitplan und einem eigenen Budget.

Evaluierung im Policy Cycle

Ein Programm wird auf Grundlage einer Problemstellung entworfen und entwickelt. Die ex-ante Evaluierung überprüft Angemessenheit und Design des Programms. Nach allfälligen Modifikationen wird das Programm umgesetzt - d.h. in der Regel an eine mit Abwicklung und Programm-Management befasste Einrichtung abgegeben. Eine interim Evaluierung zieht Zwischenbilanz: Sie hat sowohl den Stand der Projekte als auch das Programm-Management im Blick und gibt Empfehlungen für die Weiterführung. Die Umsetzung der Evaluierungsergebnisse - und zwar sowohl der Zwischenevaluierung als auch der Endevaluierung hat Konsequenzen: Das betreffende Programm wird weitergeführt oder auch nicht, es wird modifiziert oder auch nicht, es wird größer oder kleiner gestaltet etc.

Die folgende Abbildung zeigt nun einen idealtypischen "Policy Cycle" für FTE-Programme:

Abbildung 1: Idealtypischer „Policy Cycle“



Die Abbildung zeigt auch, dass in Evaluierungen stets mehrere Ebenen eingebunden sind, nämlich eine strategische Ebene, die Entwicklung und Design des Programms verantwortet, eine abwickelnde Management-Ebene und die Zielgruppe des Programms, die Kunden und Empfänger. Zwischen den unterschiedlichen Ebenen gibt es die Notwendigkeiten klarer Arbeitstrennung, aber auch ebenso klarer Kommunikationswege. Nur auf einen soll hier hingewiesen werden: Die Ergebnisse von Evaluierungen müssen „hinauf“, nämlich Teil der Bewertungs-, Lern- und Entscheidungsaktivitäten derer werden, die FTE-Politik machen und das hört eben nicht bei den Programmverantwortlichen auf. Dinge, die bei Nichtschließen dieses „Policy Cycle“ drohen, sind Mikroaktionen, unerwünschte Kontinuitäten ebenso wie unerwünschte Brüche, Handeln wider besseres Wissen (bzw. Wissen-Sollen).

Timing der Evaluierungen: Damit Evaluierungsergebnisse genutzt werden können, ist eine zeitgerechte Planung und Durchführung der Evaluierung, sowie die zeitgerechte Abgabe der Evaluierungsberichte notwendig. Die Auftraggeberin einer Evaluierung und die EvaluatorInnen müssen sicherstellen, dass die Ergebnisse dann vorliegen, wenn die Informationen auch den größten potenziellen Nutzen haben. Das bedeutet etwa im Falle von Programmevaluierungen, dass die Ergebnisse der Evaluierung bereits vor der Entscheidung über die Verlängerung oder das Auslaufenlassen eines Programms verfügbar sein müssen, damit sie genutzt werden können. Dieser Anspruch setzt ein Planungsverfahren bei Evaluierungen voraus, der von zukünftigen Deadlines zeitlich zurückrechnet. Realistische Planungsvorgaben erleichtern diesen Prozess.

Kommunikationsstrategien frühzeitig festlegen: Mit der Erstellung eines Berichtes über die Ergebnisse der Evaluierung ist es nicht getan. Damit die Ergebnisse von Evaluierungen optimal genutzt werden können, braucht es eine qualifizierte Öffentlichkeit, um die Ergebnisse zu diskutieren und Schlüsse für notwendige weitere Schritte zu ziehen.

Schaffung eines Evaluierungssystems mit Feedback-Schleifen: Evaluierungen sind nicht Selbstzweck. Sie sollen einen Beitrag zur Verbesserung der Performance des forschungs- und technologiepolitischen Systems leisten und dazu organisatorische Lernprozesse bei den beteiligten Akteuren fördern. Voraussetzung dafür sind systematisch geplante und durchgeführte Evaluierungen, die ausreichend Feedbackschleifen vorsehen. Erst durch die Rückkopplung der Ergebnisse können strategisch-orientierte Verbesserungsmaßnahmen Gestalt annehmen, neue Instrumente geschaffen, und die Umsetzung der Ziele der Forschungs- und Technologiepolitik neuen Herausforderungen angepasst werden.

Feedback-Schleifen können institutionalisiert werden, wie Beispiele aus anderen Ländern zeigen. In Großbritannien etwa stellen besondere Anforderungen an das Programm- und Maßnahmendesign (ROAME-Prozess) die Einbindung von Erfahrungen in den Politikprozess sicher. Allerdings müssen natürlich die geschaffenen Rückkopplungsstrukturen flexibel genug sein, damit die involvierten Akteure rasch auf verändernde Rahmenbedingungen reagieren können.

Die durch Evaluierungen gewonnenen Informationen können während des gesamten Politikzyklus einen wichtigen Beitrag zur Förderung der Zielorientierung sowie der Effektivität und Effizienz von Forschungs- und Technologiepolitik leisten.

Methoden der Evaluierung

Die Wahl der Methode ist im Kontext von Umfang und Zielsetzung des Programms/der Institution/des Problems zu treffen. Neben einem adäquaten Kosten/Nutzen-Verhältnis sind auch pragmatische Gründe, wie z. B. die Verfügbarkeit von Daten ausschlaggebend für die Wahl der Methode.

Ein generelles Spannungsfeld ergibt sich hier daraus, dass Evaluierungen einerseits analytisch und unabhängig andererseits aber problem- und kundenorientiert sind. Summative Ansätze betonen eher den analytischen Aspekt, bei formativen Ansätzen verstehen sich die EvaluatorInnen stärker als "Coach".

Quantitative Methoden

Bei quantitativen Methoden sind zwei große Gruppen unterscheidbar:

Beschreibende statistische Analysen (Surveys) generiert aus Befragungen, Daten des Monitorings und/oder sekundärstatistischem Datenmaterial stellen die grundlegendste quantitative Methodengruppe dar. Diese Analysen sind Teil jeder Evaluierung und umfassen neben vergleichenden Tabellen und

Grafiken auch fortgeschrittenere Methoden wie z. B. Geographische Informationssysteme oder Faktor- und Clusteranalysen. Wichtig ist hier, dass die gesammelten und strukturierten Informationen zumindest teilweise zu quantitativen Indikatoren verdichtet werden.

Methoden zur Abschätzung von Nettoeffekten sind notwendige Bestandteile einer Evaluierung, da nur auf diese Weise Thesen hinsichtlich der Wirkung von Programmen, Projekten oder Politiken verifiziert bzw. falsifiziert werden können (Stichwort **Additionalität**). Hier sind wiederum drei Methodengruppen unterscheidbar, die alle relativ hohe Anforderungen an das vorhandene Datenmaterial stellen:

- **Vergleichende Analysen** bilden keine gesamtwirtschaftlichen Effekte ab, sondern versuchen einen Vergleich der Welt mit der betrachteten Intervention mit der Welt ohne die betrachtete Intervention (im Gegensatz zu beschreibenden statistischen Methoden, die in der Regel keinen mit/ohne Vergleich sondern einen davor/danach Vergleich durchführen). Hierunter fallen Kontrollgruppen- bzw. Vergleichsgruppen-Ansätze.
- **Ökonometrische Modelle** sind die vermutlich am weitesten verbreitete Methode zur Messung der ökonomischen Effekte von F&E-Programmen. Sie werden sowohl auf der Unternehmensebene wie auch der sektoralen und der gesamtwirtschaftlichen Ebene angewandt, um Zusammenhänge zwischen Eingriffen der öffentlichen Hand und den davon ausgehenden Wirkungen zu testen bzw. quantitativ abzuschätzen. Auf gesamtwirtschaftlicher Ebene können nur größere Interventionen untersucht werden (z. B. im Bereich der Bewertung der indirekten (steuerlichen) F&E-Förderung). Mikroökonometrische Arbeiten, die beispielsweise die Entwicklung von Performance-Indikatoren von Firmen mit der Teilnahme an einem Technologieprogramm in Verbindung setzen gibt es in Österreich noch kaum.
- **Kosten-Nutzen-Analysen** oder auch **Fiskalanalysen** stellen einen sehr strikten Rahmen zur strukturierten Untersuchung forschungs- und technologiepolitischer Interventionen dar. Bei großen öffentlichen Infrastrukturinvestitionen (Autobahnen, Tunnels) wie auch in anderen Politikfeldern (vor allem in der Arbeitsmarktpolitik) zählen Kosten-Nutzen-Analysen zur Standard-Methodologie bei Evaluierungen, in der Forschungs- und Technologiepolitik hat dieses Analyseinstrument, welches bei vieldimensionalen und strategischen Zielsetzungen schwer anwendbar ist, Seltenheitswert. Der Vorteil von Kosten-Nutzen-Analysen ist, dass nicht nur explizite Aussagen in bezug auf die Art der Kosten und Nutzen getroffen werden, sondern diese Aussagen auch quantifiziert werden müssen. Kosten-nutzen-analytische Überlegungen spielen immer dann eine wichtige Rolle, wo es um soziale Erträge oder volkswirtschaftliche Effekte technologiepolitischer Programme geht. Dies heißt im wesentlichen, dass
 - a. auch jene – positiven und negativen – Effekte abgeschätzt werden, die nicht bei den direkten Programmteilnehmern auftreten,
 - b. versucht wird, sogenannte ‚intangible‘, keiner Marktbewertung zugängliche, Effekte zu bewerten,
 - c. zu unterschiedlichen Zeitpunkten anfallende Kosten- und Nutzenströme in ihrem Gegenwartswert dargestellt und damit verglichen werden können.

Qualitative Methoden

Qualitative Methoden sind im wesentlichen aus der Kritik an der quantitativen Sozialforschung entstanden und verstehen sich als interpretative Verfahren. Im Rahmen von Evaluierungen wird der Begriff meist in den folgenden Zusammenhängen angewendet:

- **Qualitative Verfahren zur Gewinnung von Daten:** Hierzu zählen offene, halboffene und narrative Interviews, teilnehmende und direkte Beobachtung, und Fokus Gruppen. Fallstudien und Textanalysen haben auch primär qualitativen Charakter. Die so gewonnenen Daten können in der Folge unterschiedlich ausgewertet werden.
- **Qualitative Forschung zur Generierung von Hypothesen:** Hier werden auch auf der Basis von qualitativ gewonnenen Daten Hypothesen zur Wirkungsweise bestimmter Interventionen formuliert. In diesem Kontext, der meist bei sehr komplexen Fragestellungen Sinn macht, ist qualitative Forschung keine eigenständige Methode sondern Vorbereitung oder Begleitung anderer Forschungsprozesse (so kann beispielsweise bei der Evaluierung komplexer Programme mittels Interview das Verständnis für das Programm verbessert werden).
- **Sinnverstandendes Vorgehen im Rahmen interpretativer Sozialforschung:** Hier werden individuelle Bedeutungszuschreibungen, situative Kontexte und gesellschaftliche Rahmenbedingungen in den Forschungsprozess mit einbezogen und durch kommunikatives Handeln zugänglich gemacht. Diese Methoden der qualitativen Interpretation sind dann zentral, wenn komplexe soziale Prozesse im Vordergrund stehen, aber auch wenn Evaluierungen stark formativen oder moderierenden Charakter haben. Wesentliche Kennzeichen sind: Verzahnung von Durchführung und Auswertung, zahlreiche Rückkopplungen und Feed-back Schleifen und Lernen in sozialen Kontexten.

Eine weitere qualitative Methode, die bei der Bewertung von Projekten vor allem im Bereich der Forschung eingesetzt wird, ist das **Peer Review Verfahren**. Peer Review Verfahren werden traditionellerweise bei der Bewertung von Artikeln für wissenschaftliche Zeitschriften und für die ex ante Projektauswahl eingesetzt. Peer Review Verfahren sind unverzichtbar, da es für sehr viele Aspekte der Antragsqualität keine „objektiven“, beispielsweise durch Indikatoren oder Messverfahren ermittelbaren Kriterien gibt. In diesem Kontext gibt es auch kritische Literatur, die teilweise auch empirisch gestützt ist (mangelnde Reliabilität, mangelnde Fairness, mangelnde Validität). Bei der ex-ante Bewertung von Projekten kommt die Unmöglichkeit hinzu, künftiges vollständig bewerten zu können. Das heißt, selbst wenn die Peers fair sind und keine echte Diskriminierung vorliegt, bleibt die Bewertung immer bis zu einem gewissen Grad mangelhaft, da niemals vollständige Information über die künftigen Projekte vorliegen kann. Peer Review Verfahren sind andererseits unverzichtbar, da es für sehr viele Aspekte der Antragsqualität keine „objektiven“, beispielsweise durch Indikatoren oder Messverfahren ermittelbaren Kriterien gibt. Um der Komplexität von Projekten gerade im technologiepolitischen Bereich gerecht zu werden, diskutiert man in jüngster Zeit vielfach das Konzept der „Extended Peer Review“. Hier werden FachexpertInnen, also klassische Peers mit Know-how TrägerInnen aus anderen Bereichen (wie z. B. ManagementexpertInnen, IndustrievertreterInnen, ManagementexpertInnen, EvaluatorInnen etc.) zusammengeführt.

Neben den beiden großen Methodengruppen "quantitativ" und "qualitativ" werden vor allem im angelsächsischen Raum oft auch **Management-orientierte Systeme** zu den Methoden der Evaluierung gezählt. Beispiele hierfür sind PERT (**Program Evaluierung and Review Technique**), oder CPM (**Critical Path Method**). Diese hoch standardisierten Methoden, welche von privaten Firmen gleichermaßen wie von öffentlichen Einrichtungen angewendet werden, binden Evaluierungen in einen breiteren Kontext von Controlling, strategischer Entscheidung und Organisationsentwicklung ein.

Während noch vor wenigen Jahren Grabenkämpfe zwischen den Vertretern qualitativer und denen quantitativer Denktraditionen gefochten wurden, nehmen heute die **"Mixed-Method"** Ansätze zu: So werden qualitative Methoden häufig im Vorfeld einer quantitativen Analyse verwendet, um den Kontext des Projektes umfassend zu erschließen und damit das Grundverständnis für eine sinnvolle Interpretation der dann folgenden quantitativen Analyse zu schaffen. Auch hängt die Wahl des Methodenmix von der Art und Zielsetzung des zu evaluierenden Programms / Projekts / Politikfeldes ab:

Zielt dieses auf gleiche Ergebnisse bei allen TeilnehmerInnen ab, so ergibt sich eine Präferenz für quantitative Methoden - je individueller die erwarteten Ergebnisse, desto mehr empfiehlt sich der Einsatz qualitativer Methoden.

Im forschungs- und technologiepolitischen Kontext sind quantitative Methoden unentbehrlich, wenn die Evaluierung Aussagen über die Wirkung der Intervention machen soll. Andererseits nehmen qualitative Elemente in ihrer Bedeutung zu: Erstens weil zunehmend mehr Programme nicht nur einzelne Firmen oder Institutionen ansprechen, sondern überhaupt auf die Funktion eines ganzen Innovationssystems zielen und auch stark Managementelemente enthalten. Zum zweiten gehen Evaluierungen von *Institutionen* immer mehr auf ihre Beziehungen zur bzw. wechselweisen Beeinflussung mit anderen Akteuren im Innovationssystem ein (**Systemevaluierung**).

Informationsunterlagen und Datenquellen der Evaluierung

"Collect only the information you are going to use and use all the information you collect" (W.K Kellog, Foundation Evaluation Handbook 1998)

Für die Qualität der Analyse, aber auch für die Wahl der Analysemethode ist die Qualität der vorhandenen bzw. der erhebbaren Daten von zentraler Bedeutung. Neben den projektbezogenen Daten sollen auch die Ergebnisse allgemeiner Primärerhebungen (z. B. Community Innovation Survey, F&E Erhebungen) für Evaluierungen nutzbar gemacht werden.

Auch ist es notwendig, die Frage der Datenweitergabe an EvaluatorInnen bereits im Programmdesign anzusprechen und eventuell in den Förderverträgen mit Dritten zu fixieren.

Datenquellen können unterschieden werden nach:

- **Eigene Primärerhebungen:** Von den EvaluatorInnen durchgeführte Primärerhebungen haben in der Regel die Form von TeilnehmerInnenbefragungen. Hier ist zwischen qualitativen Formen der Datenerhebung und quantitativen Formen der Datenerhebung (strukturierte Fragebögen, standardisierte Interviews) zu unterscheiden.
- **Nutzung von Sekundärdatenbasen:** Sekundärdatenbasen liegen vor allem in aggregierter Form vor, und sind daher eher komplementär als substitutiv zu Primärerhebungen zu sehen. Für Mikrodaten sind teilweise Unternehmensdatenbanken nutzbar, allerdings unterliegen diese Datenbasen in der Regel statistischen Verzerrungen und geben daher die Grundgesamtheit der Unternehmen nicht repräsentativ wieder.
- **Nutzung von Daten aus dem Programm-Monitoring:** dies ist eine ganz zentrale Datenquelle der Evaluierung – ein gutes Monitoring System erleichtert den EvaluatorInnen die Arbeit enorm und vermeidet auch lästige Doppelerhebungen. Das heißt, ex ante Evaluierungen und Arbeiten zum Design von Programmen sollen entsprechende Schienen und Indikatoren für ein Monitoring vorsehen.
- **Primärerhebung im Kontext einer anderen, breiteren Primärerhebung:** Dies stellt im Prinzip einen Idealfall dar. Die für die Evaluierung notwendigen Daten können in diesem Fall entweder direkt aus den Rohdaten der Primärerhebung genommen oder mit evaluierungsspezifischen Zusatzerhebungen kombiniert werden. Auch lassen sich auf diese Weise relativ einfach und genau Kontrollgruppen definieren.

Monitoring

Es ist Aufgabe eines guten Monitoringsystems, alle relevanten Daten – und zwar nur diese – in einfacher, systematischer und geschlechtssensibler Weise zu erfassen. Die Daten werden zum einen für das inhaltliche und finanzielle Projektcontrolling verwendet, zum anderen sollen sie EvaluatorenInnen einen guten Einblick in die Projekte geben. Dadurch kann die Datenqualität der Evaluierungen deutlich gesteigert und eine doppelte Erhebung vermieden werden. Es wird von den Fördereinrichtungen zu prüfen sein, in wie weit adäquate Programm-Management-Informationssysteme eingesetzt werden können.

Indikatoren

Indikatoren werden sehr häufig sowohl im Kontext von Evaluierungen als auch im Kontext von Monitoring und Controlling ermittelt. Indikatoren sind jedoch keine für sich selbst sprechenden Kennziffern. Hinter jedem Indikator steht eine modellhafte Annahme über die Realität und ein Operationalisierungskonzept. Indikatoren sollen komplexe, nicht direkt messbare Realität in wenigen Zahlenwerten repräsentieren. Zentrale Indikatoren werden zu Beginn eines Programms festgelegt und - soweit möglich - im Rahmen des Monitorings erhoben. In der Regel werden die folgenden Basisindikatoren unterschieden:

- **Outputindikatoren** (physisch und auf die Implementierung bezogen, z. B. Beratungsinfrastruktur aufgebaut, 2 Berater eingestellt)
- **Ergebnisindikatoren bzw. Outcome Indikatoren** (auf das Ergebnis der Intervention bezogen, z. B. 50 Beratungen im Jahr mit Zielgruppe KMU durchgeführt)
- **Wirkungs- bzw. Impact Indikatoren** (auf das Ziel der Intervention bezogen, z. B. Innovationsfähigkeit von 20 KMUs erhöht gemessen beispielsweise an der Anzahl neuer Produkte)

Aus diesen Basisindikatoren können Effektivitätsmaße (Basisindikator bezogen auf die Plangröße) und Effizienzmaße (Basisindikator bezogen auf die Kosten) abgeleitet werden.

Evaluierung im Kontext anderer Formen der Bewertung und wissenschaftlicher Unterstützung im Politikprozess

Evaluierung ist das zentrale Instrument um die Wirkungen von geplanten, laufenden oder abgeschlossenen Programmen in der Forschungs- und Technologiepolitik beurteilen zu können. Evaluierung steht allerdings als Methode der Bewertung nicht alleine. Gerade in den letzten Jahren hat beispielsweise auf europäischer Ebene Benchmarking zunehmend an Bedeutung gewonnen. Diese versucht, die relative Position einzelner Ländern bzw. Regionen zu bestimmten Aspekten der Forschungs- und Technologiepolitik zu bestimmen.

Evaluierung und Benchmarking blicken in der Regel in die Vergangenheit bzw. analysieren die gegenwärtige Situation. Forschungs- und Technologiepolitik ist jedoch notwendigerweise immer an den (erwarteten) Herausforderungen in der Zukunft orientiert. Die politischen EntscheidungsträgerInnen stehen vor der Aufgabe, auf identifizierte Defizite, vorhandene Potenziale und zukünftig erwartete Chancen und Risiken mit möglichst effektiven und effizienten Maßnahmen zur Verbesserung der Performance von Forschung und technologischer Entwicklung zu reagieren. Lernerfahrungen aus der Vergangenheit aus Evaluierungen und Benchmarking-Studien können dazu natürlich einen wesentlichen Beitrag leisten. Darüber hinaus werden für die Politikgestaltung jedoch auch andere,

stärker prospektiv orientierte Instrumente notwendig sein, um möglichst rationale, nachvollziehbare und sozial akzeptierte Entscheidungen über die forschungs- und technologiepolitischen Schwerpunktsetzungen und Maßnahmen treffen zu können.

Welche Instrumente der Bewertung und der wissenschaftlichen Beratung von Entscheidungsprozessen in welcher Form genutzt und institutionalisiert werden sollen hat auf europäischer Ebene in den letzten Jahren zu zahlreichen Diskussionen und Vorschlägen geführt. Neben der Evaluierung sind es vor allem Methoden der Technologie-Vorausschau (z. B. Foresight-Studien) und die moderneren Instrumente der Technikfolgenabschätzung (z. B. CTA-Constructive Technology Assessment), die in zahlreichen Ländern Europas heute etabliert sind, um Forschungs- und Technologiepolitik zu beraten und weiterzuentwickeln. In Österreich wurde mit dem Technologie-Delphi (ÖAW ITA 1998) ein erster Schritt zur Etablierung von Foresight gesetzt. Konzepte wie "Evaluierung and Learning System Approach" (Guy 1998) oder "Distributed Intelligence" (Kuhlmann et al. 1999) wurden vorgeschlagen, um die jeweiligen Stärken der retrospektiv sowie der prospektiv ausgerichteten Bewertungs- und Unterstützungsinstrumente zu kombinieren. Gerade bei Forschungs- und Technologiefeldern, die potenziell große sozioökonomische Auswirkungen in der Zukunft haben können, sollten auch die Ergebnisse eines möglichst breiten Konsultations- und Diskussionsprozesses für die Entscheidungsfindung über Programme und Maßnahmen herangezogen werden.

Auf europäischer Ebene gibt es unter dem Schlagwort "Science for Policy" derzeit eine aktuelle Debatte welche wissenschaftliche Expertise im Politikprozess benötigt wird, und wie diese im Sinne der Vorbereitung von sozial möglichst "robusten" Entscheidungen organisiert und genutzt werden kann. In einzelnen europäischen Staaten (z. B. Großbritannien) wurden dazu bereits Richtlinien erlassen, die beispielsweise eine Verknüpfung von Evaluierung mit dem nationalen Foresight-Programm und dem ministeriellen Planungsprozess Forward Look garantieren sollen. Die österreichische Forschungs- und Technologiepolitik hat die Diskussion über die Frage nach der Organisation und Verknüpfung von Evaluierung mit den Ergebnissen aus anderen Quellen noch weitgehend vor sich.

Anhang: Wo stehen wir in Österreich? Auf welchen „Standards“ beruht die laufende Evaluierungsarbeit?

Nach mehreren Jahren bewusster Beschäftigung mit Evaluierungsfragen im Bereich der FTE-Politik sind in Österreich derzeit die in der Folge angeführten Entwicklungsstände zu beobachten: Die Punkte sind hier beabsichtigt vereinfacht gehaltene beispielhafte Aussagen. Es geht nicht darum, alles explizit vorzustellen, sondern darum, einen ersten Überblick zu bieten, wohin wir bereits gekommen sind. Die Unterteilung gliedert sich nach Arten bzw. Ebenen der Evaluierung und soweit nötig auch nach zeitlichen Phasen von Evaluierungen.

- **Projektebene ex ante:** Aufgrund der langen erfolgreichen Geschichte der fondsgetragenen Projektförderung ist vor allem das Instrument der ex ante Projektbewertung weit entwickelt und kann sich mit internationalen Vergleichsbeispielen jedenfalls messen. Im Rahmen dieser guten Praxis gibt es einige latente Weiterentwicklungswünsche (vgl. bessere Indikatoren für soziale und ökologische Aspekte) und einige ebenso latente Grundsatzfragen (etwa, wer in welchem Fall evaluieren soll), insgesamt sind hier aber wenig Wünsche offen. Ein Methodenwissen ist ebenso vorhanden wie Qualitätsbewusstsein. Zunehmend findet Projektförderung auch programmgebunden statt und hier fand in den letzten Jahren weiteres Lernen hinsichtlich Evaluierungskriterien und Methoden statt (wie beispielsweise die Einführung von „Extended Peer Review“ Verfahren bei komplexen technologiepolitischen Programmen).
- **Projektebene Monitoring und ex post:** Auch unabhängig von der Größe der Projekte nimmt Monitoring in formalisierterer Form zu; vor allem im Bereich der wirtschaftsorientierten FTE-Förderung, egal ob in rein antragsgebundener oder programmgeleiteter Form. Auch hier kann von einem hohen und der jeweiligen Projektgröße angepassten Niveau gesprochen werden. Die ex post Evaluierung von Projekten hingegen stellt sich als ein Feld dar, in dem ausgehend von guten Ansätzen Arbeitsbedarf vorliegt: In der Förderung der Grundlagenforschung haben entsprechende Arbeiten erst begonnen, in der rein antragsgebundenen wirtschaftsorientierten FTE-Förderung kommen bis dato Instrumente zum Einsatz, bei denen das Verhältnis zwischen Erkenntnisgrundlagen und aufgestellten Wirklichkeitsbehauptungen Diskussionsgegenstand sind. Hinsichtlich der programmgebundenen Förderungen sind diese entweder noch zu jung für ex post Evaluierungen oder es wurde im Rahmen von Programmevaluierungen ein je nach Programm unterschiedlich tiefer Blick auf die Projektebene geworfen².
- **Programmebene ex ante:** Wie wurden bislang neue Programme eingeführt? Meist aufgrund von Initiativen von (im weitesten Sinn Verwaltungs-)Akteuren im öffentlichen Bereich, denen Bedarfsäußerungen aus dem jeweiligen Feld bzw. Bedarfserhebungen in diesem vorangegangen waren. Ein üblicher freilich nicht standardisierter Ablauf dafür ist: Planungsgespräche – Vergabe einer Studie – Bedarfserhebung und Diskussion – Erarbeitung von Programmdokumenten unter Einbeziehung von Betroffenen. Nicht immer waren klare Ablaufpläne vorhanden, die Qualität der analytischen Basis und das Methodenbewusstsein war von unterschiedlicher Qualität. Nur selten gibt es organisiertes Lernen zwischen Programmen. Eine Reihe von externen ExpertInnen dient dazu, dieses Wissen gleichermaßen zu tradieren; mal eher als Legitimation, mal als Lerninstrument. Gute

² Vgl. die darin völlig unterschiedlichen ITF-Evaluierungen der Schwerpunkte: Energietechnik, Umwelttechnik, Verkehrstechnik, Seed Financing, FlexCIM, Technologietransfer

Praxis in Programmplanung und Management ist bislang oft ein Wunsch und ein eher implizites als explizites Prinzip³ geblieben. Hier ist ausgehend von einigen hoffnungsvollen Ansätzen viel zu tun. Die jüngsten Entwicklungen sind vor allem durch den Rat und seine „Sondermittel“-Empfehlungen geprägt worden. Hier ist eine Absicht zu beobachten, zu stärkerer Strukturierung und Rationalität zu gelangen, eine Absicht, die aber über weite Strecken noch der Realisierung harrt. In den nächsten Jahren wird freilich auch mehr Zeit sein, Programmempfehlungen aufgrund von ex ante Evaluierungen oder evaluierungsähnlichen Mechanismen zu treffen bzw. den Empfehlungsprozess entsprechend abzustützen.

- **Programmebene Monitoring und ex post:** Programmmonitoring findet in Österreich hauptsächlich in Form laufender und in seiner Bedeutung nicht zu unterschätzender Gespräche zwischen Akteuren statt. Manche, vor allem größere Programme werden zwischenevaluert. Die ex post Programmevaluierung wiederum ist ein gut eingeführtes Instrument, das wesentlich zum Aufbau und der Pflege von Evaluierungskultur in Österreich beigetragen hat⁴. Hier wurden verschiedene Methodenansätze erprobt, verwirklicht und auch weiterentwickelt. Anhand derartiger Evaluierungen und damit im Zusammenhang stehender Grundlagenarbeiten wurden große Fortschritte sowohl in den qualitativen als auch in den quantitativen Methoden erzielt. Dabei geht es eher um das Verbindlichmachen bestehender Praxis als um die Formalisierung und Pflege von Standards.
- **Institutionenebene Förderungseinrichtungen:** Förderungseinrichtungen tendieren dazu, im institutionellen Kontext oft nicht als Evaluierungsobjekt wahrgenommen zu werden. Dieser Punkt ist vor allem dort interessant, wo das überwiegende Geschäft solcher Einrichtungen in der Abwicklung von Einzelprojekten liegt. Seit der Einrichtung eines neuen strategischen Beratungsgremiums der Bundesregierung kommen solche Punkte kraft einer „fragenden Instanz“ stärker aufs Tapet. Gerade aus dem Gesagten ergibt sich eine besondere Notwendigkeit sowohl für Auftraggeber, EvaluatorInnen als auch besonders für die evaluierten Einrichtungen selbst, Standards zu haben, die konsensual entwickelt und auch getragen werden. Die Evaluierung der beiden großen Förderungsfonds FWF und FFF ist hier ein Wendepunkt. Wird sie von allen Beteiligten mit Ernst und Anstand bewältigt, so stellt dies einen großen Fortschritt für die heimische Evaluierungskultur dar.
- **Institutionenebene Forschungseinrichtungen:** Die ausführenden Einrichtungen werden über verschiedene Finanzierungsformen erhalten. Dort wo sie Programmgelder in Anspruch nehmen, sind diese Institutionen schnell mit Evaluierungen in Berührung. Dort wo Basisfinanzierung vorliegt, und das ist in der österreichischen FTE-Finanzierung nach wie vor die überwiegende Mehrzahl der Mittel, hängt es davon ab, ob Institutionenevaluierungen vorgesehen sind, vielleicht sogar selbst angestoßen werden oder eben unterbleiben. Das letztere ist eher die Regel, wenn auch bemerkenswerte Initiativen in den vergangenen Jahren zu verzeichnen waren. Insgesamt kann von den Herausforderungen her wohl eine Zweiteilung vorgenommen werden:
 1. *Universitäten* mit den besonderen Merkmalen und Herausforderungen: Verknüpfung Forschung und Lehre, Besonderheiten der Organisation und des Dienstrechts; gegenwärtige Veränderungsprozesse; Versuche, eine indikatorengebundene Finanzierung einzuführen; schließlich Frage der Art der Institutionalisierung der

³ Vgl. die standardisierteren Vorgangsweisen etwa in Schweden oder Großbritannien.

⁴ Vgl. sowohl nationale Programme wie die erwähnten ITF-Evaluierungen als auch die „Impact Evaluierungen“ der EU-Rahmenprogrammbeiträgen

Universitäts-evaluierungen (die sich im Entstehen befindliche Qualitätssicherungsagentur)

2. *Andere „außeruniversitäre“ Forschungseinrichtungen* von der Akademie bis zu sehr wirtschaftsnahen Einrichtungen, bei denen unterschiedliche Bereitschaft, unterschiedlich starker finanzieller und politischer Druck, unterschiedliche Voraussetzungen und Notwendigkeiten vorliegen.

Insgesamt ist der Bedarf nach jeweils angepassten („weichen“) Evaluierungsstandards relativ hoch und die Anforderungen für die universitäre Seite doch auch sehr speziell.

- **„Policy Cycle“ und „Governance“:** Hinter diesen Begriffen steht, dass in Evaluierungen stets mehrere Ebenen eingebunden sind, nämlich Besteller, Abwickler/Durchführende und KundInnen/EmpfängerIn. Weiters sind die weiter oben angeführten Ebenen immer in Verbindung miteinander zu sehen: Institutionen beteiligen sich an Programmen und Projekten, die wiederum von Institutionen abgewickelt werden; schließlich bestehen Programme aus Projekten. Zwischen den unterschiedlichen Ebenen gibt es die Notwendigkeiten klarer Arbeitstrennung, aber auch ebenso klarer Kommunikationswege. Nur auf einen soll hier hingewiesen werden: Die Ergebnisse von Evaluierungen müssen „hinauf“, nämlich Teil der Bewertungs-, Lern- und Entscheidungsaktivitäten derer werden, die FTE-Politik machen und das hört eben nicht bei den Programmverantwortlichen auf. Dinge, die bei Nichtschließen dieses „Policy Cycle“ drohen, sind Mikroaktionen, unerwünschte Kontinuitäten ebenso wie unerwünschte Brüche, Handeln wider besseres Wissen (bzw. Wissen-Sollen).
- **„Policy Cycle“ (2):** Nationale Forschungs- und Innovationspläne können in ihrer Argumentationskraft noch ein Stück gewinnen, wenn sie auf vorliegende Evaluierungsstudien Bezug nehmen.
- **Standards und „Standards“:** Was ohne Standards geschieht, kriert auch welche – „Seinesgleichen geschieht“ (nochmals Robert Musil). So gab es natürlich auch in Österreich immer schon gewisse Evaluierungsstandards, freilich unterschiedlich stark vorhandene und in unterschiedlichen Teilsegmenten; regelmäßig ohne größere Bindungskraft. Im Jahr 2003 hat nach umfassenden Vorarbeiten die Plattform Forschungs- und Technologieevaluierung Evaluierungsstandards für diesen Politikbereich beschlossen. Dies bedeutet, dass sich wichtige Akteure, darunter drei Ministerien und die größten Bundesförderer, aber auch Forschungseinrichtungen per Selbstbindung zur Einhaltung bestimmter qualitativer Mindestbedingungen verpflichtet haben. Nun gilt es, diesen bottom up Ansatz mit Leben zu erfüllen.
- **Internationaler Dialog:** Starke Verbindungen zwischen der österreichischen Evaluierungsszene und ihrer weltweiten Kollegenschaft helfen mit, Evaluierungen in Österreich anspruchsvoller und besser werden zu lassen. Ein Höhepunkt war die große Evaluierungskonferenz mit 200 TeilnehmerInnen aus 25 Ländern, die die Plattform mit veranstaltet hat.