



**Nützlichkeit und Nutzen der  
Programmevaluationen im Bereich  
der österreichischen FTI-Politik.  
Metaevaluation der Programmevaluationen  
2003-2014.**

Studie im Auftrag des Rats für Forschung  
und Technologieentwicklung (RFTE)

Kurzbericht  
Dezember 2015

Günther Landsteiner



Der vorliegende Bericht wurde im Rahmen eines freien Auftragsverhältnisses erstellt von:

MMag. Günther Landsteiner  
Unabhängiger Forscher-Berater und Evaluator  
Zieglergasse 75 top 6  
A-1070 Wien  
mail: [g.landsteiner@chello.at](mailto:g.landsteiner@chello.at)

Der Bericht stellt ausschließlich die fachlichen Ansichten und Meinungen des Autors dar, die von denen des Auftraggebers abweichen können.



# 1. Auftrag und Vorgehensweise der Metaevaluation

## 1.1. Motivation

In Österreich wurde über Jahre hinweg eine Evaluationskultur im forschungs-, technologie- und innovationspolitischen Bereich aufgebaut, wobei die Gründung der Plattform fteval 1997 ein entscheidendes Datum darstellt. Die Anzahl an Programmevaluationen, die seitdem durchgeführt wurden, ist im internationalen Vergleich für den Politikbereich jedenfalls als überdurchschnittlich zu erachten. Freilich sind mit der zunehmenden Entfaltung des Unternehmens FTI-Evaluation auch immer wieder Stimmen in der Akteursarena laut geworden, die Enttäuschungen gegenüber ursprünglich hohen Erwartungen thematisieren. Zugleich ist in den letzten Jahren eine internationale Entwicklung zu beobachten, die Frage nach der Wirksamkeit und dem Reflexionspotenzial von Evaluationen im FTI-Bereich aufzuwerfen. Parallel dazu haben über den FTI-Bereich hinaus Evaluationsgesellschaften die Frage der Nützlichkeit von Evaluation und des Wissens über die Evaluationspraxis auf die Tagesordnung gesetzt.

Der RFTE hat deswegen die vorliegende Studie zur Nützlichkeit und zum tatsächlich feststellbaren Nutzen der Programmevaluationen im österreichischen FTI-Bereich beauftragt, mit der der unabhängige externe Evaluator im Juli 2014 betraut wurde. Die Studie zielt darauf ab, evidenzbasierte Einschätzungen darüber zu gewinnen, wie sich die Nützlichkeit der Programmevaluationen gestaltet, und inwieweit etwaiges Verbesserungspotenzial besteht. Die Studie findet dazu ihre Basis in einem allgemein anerkannten, internationalen Wissenskorpus über qualitätsvolle Evaluation und spezifische Aspekte der Evaluationsnutzung. Die konkrete Herangehensweise wurde als diejenige einer Metaevaluation definiert, die um Komponenten einer Nutzungsforschung zur Evaluation angereichert ist. Neben der Frage, welche Grundlagen für die Entstehung von Evaluationsnutzen mit den durchgeführten Programmevaluationen gelegt wurden, wird auch die Frage nach nachvollziehbaren Nutzungsweisen der Programmevaluationen verfolgt. Analysiert wird die Evaluationspraxis im FTI-politischen Bereich über den zwölfjährigen Zeitraum 2003–2014, soweit sie Programmevaluationen betrifft.

## 1.2 Vorgehensweise und Methodeneinsatz

Metaevaluationen stellen systematische Analysen von einer oder mehreren Evaluationen dar, die der Bestimmung von Wert und Güte dieser Evaluationen gelten. Der vorliegenden Studie geht es somit nicht um die inhaltlichen Ergebnisse, die Programmevaluationen erbracht haben, sondern um Gestaltungsweisen der Evaluationsprozesse und Evaluationsberichte, die als essentielle Voraussetzungen und Merkmale der Nützlichkeit von Evaluationen zu erachten sind. Hier verkörpert sich ein Nutzungspotenzial, das die Grundlage für faktische Nutzungsweisen bildet und als solches hinsichtlich von Stärken und Schwächen analysiert werden kann. Nutzungsforschung zur Evaluation identifiziert Ausmaß und Art von Evaluationsnutzen und Einflussfaktoren, die das Zustandekommen dieses Nutzens fördern oder behindern.

Mit den „Standards für Evaluation“ der deutsch-österreichischen Evaluationsgesellschaft DeGEval liegt ein hervorragender Bezugspunkt für eine neutrale und breit abgesicherte Thematisierung von Qualitätsaspekten einer Programmevaluation vor, die explizit mit deren Nützlichkeit verbunden sind. Die „Standards für Evaluation“ verstehen sich als Anleitung für zielgerichtete professionelle Evaluation in allen Politikfeldern und als fachliches Dialoginstrument für einen Austausch über die Qualität von professionellen Evaluationen. Sie zählen zu einer Standardfamilie, die auch die US-amerikanischen „Program Evaluation Standards“ und die schweizerischen SEVAL-Standards umfasst. Die Wahl dieses international hoch relevanten, mit der übergreifenden Gesamtentwicklung von Evaluationstheorie und -forschung verbundenen Bezugspunktes erfolgte in bewusster Abhebung von den Standards der Plattform fteval, die sich die Community der österreichischen FTI-EvaluatorInnen und Auftraggeberinstitutionen gegeben hat. Eine vergleichende Auseinandersetzung mit den fteval-Standards oder deren Kommentierung war nicht Bestandteil des Auftrags. Die Anwendung der Standards und die Vorgehensweise der Metaevaluation wurden in einem Workshop mit den beiden Hauptexperten für die DeGEval- und SEVAL-Standards erörtert. Auf der Grundlage von Überlegungen sowohl konzeptueller als auch pragmatischer Art wurde ein Kriterienset bestimmt, das sich zunächst aus 10 DeGEval-Standards zusammensetzte und während der Durchführung der Evaluation auf 11 Standards erweitert wurde, um erkannten Bedarfslagen noch besser gerecht zu werden.

Für einen summativen und explorativen Zugriff auf die langjährige Evaluationspraxis im Kontext begrenzter Ressourcen wurde eine Stichprobe von 20 publizierten Evaluationsberichten auf der Basis eines mit dem Auftraggeber abgestimmten *theoretical sampling* gezogen. Damit wurde knapp die Hälfte der publizierten Evaluationsberichte zu Programmevaluationen in diesem Zeitraum erfasst. Zur Ergänzung der Informationslage über Evaluationsprozesse und zur Gewinnung von Informationen über tatsächliche Nutzungen von Programmevaluationen wurden eine Online-Befragung unter FTI-EvaluatorInnen sowie Interviews mit AuftraggeberInnen im FTI-politischen Governancesystem und Hauptakteuren der Plattform fteval durchgeführt. Für die EvaluatorInnen-Befragung wurden Inhalte der Standards und der internationalen Nutzungsforschung zur Evaluation operationalisiert, für die Interviews wurden diese Inhalte einem Gesprächsleitfaden zugrunde gelegt. 273 EvaluatorInnen und MitarbeiterInnen von relevanten Instituten in Österreich und im deutsch- und englischsprachigen Ausland, die für die Durchführung einer FTI-Programmevaluation im Beobachtungszeitraum in Frage kamen bzw. bekannt waren, wurden in der Online-Befragung kontaktiert. 37 EvaluatorInnen aus dem In- und Ausland haben die Umfrage beantwortet, wobei rund die Hälfte häufig im österreichischen FTI-Bereich tätige EvaluatorInnen sind, sodass die erhaltenen Umfragedaten ein gut belastbares Bild der österreichischen FTI-Evaluationspraxis liefern. Halbstrukturierte Interviews im Umfang von bis zwei Stunden wurden mit 15 Personen im Bereich der Auftraggeber und HauptadressatInnen von Programmevaluationen geführt, mit gleichmäßiger Abdeckung der relevanten Ressorts und Agenturen auf Bundesebene.

In der Berichtsanalyse wurde auch die Frage verfolgt, ob für Nützlichkeit relevante Qualitätsmerkmale einer zeitlichen Entwicklung unterlegen sind. Alle eingesetzten Methoden waren darauf ausgerichtet, eine langjährige und vielfältige Evaluationspraxis summativ auffassen zu können und dabei auch Weiterentwicklungen und kumulative Effekte sichtbar machen zu können. Die Ergebnisse aus den drei Erhebungsverfahren werden im Folgenden in synthetisierender, an der Extraktion von Hauptcharakteristika der Evaluationspraxis ausgerichteter Weise dargestellt. Die folgende Darstellung zentraler Ergebnisse und Schlussfolgerungen bezieht sich ausschließlich auf Programmevaluationen im österreichischen FTI-Bereich und kann nicht unmittelbar verallgemeinert werden, auch wenn sich FTI-Evaluation über die letzten beiden Jahrzehnte europaweit entfaltet hat und internationalisierte Verständnisweisen der evaluativen Vorgehensweise im FTI-politischen Bereich nicht von der Hand zu weisen sind.

## **2. Zentrale Ergebnisse**

### **2.1 Nutzung von Programmevaluationen**

Bisherige Programmevaluationen haben durchaus Nutzen generiert. Aus den Auskünften der AuftraggeberInnen und HauptadressatInnen der Programmevaluationen und denen der EvaluatorInnen ergibt sich hier ein hochgradig konsistentes Bild. Dabei stehen Nutzungsweisen im Vordergrund, die von der Evaluationsforschung als instrumenteller und konzeptueller Nutzen von Evaluation bezeichnet werden. Auf Basis von Datenlagen, Schlussfolgerungen und Empfehlungen wurden Entscheidungen über Programme getroffen und neue Sichtweisen gewonnen, die zur Nachschärfung von Programmen oder zur Bereinigung von Zielkatalogen geführt haben. Entscheidungen über evaluierte Programme betrafen vor allem Adjustierungen von weiterlaufenden Programmen. Derartige Entscheidungen können sowohl zu Umsetzungsaspekten in den betrauten Agenturen fallen, entsprechend ihres jeweiligen *Pouvoirs* zum evaluierten Programm, oder seitens der Programmeigentümer für eine nachfolgende Programmphase in Programmdokumenten niedergelegt werden. Hinzu kommen konzeptuelle Einsichten über Merkmale von FTI-Segmenten ebenso wie über das Management von Programmen, die häufig auch da eintraten, wo nicht für außenstehende BeobachterInnen leicht erkennbare Entscheidungen gefällt wurden.

Die Programmevaluationen haben immer wieder Lerneffekte erzeugt, in deren Gefolge Themen verankert und Annahmen über Funktionsweisen von Teilen des FTI-Systems und der darauf gerichteten Steuerungs- und Anreizsysteme verändert wurden. Auch Evaluationen von Programmen, die in der Folge nicht weitergeführt wurden, haben solche systemisch wertvollen Einsichten und Lerneffekte erbracht. Evaluative Erkenntnisse zu einzelnen Programmen haben sich auf Konzeption und Gestaltung thematisch benachbarter Programme ebenso ausgewirkt wie auf Gestaltungsweisen von anderen Programmen im Portfolio derselben Agentur.

Nicht zuletzt ist es in der Entwicklung der Evaluationskultur zu organisatorischen Anpassungen gekommen, die den Umgang mit Evaluationen und deren Ergebnissen unterstützen. Insbesondere wurde in einer Agentur rezent ein Managementprozess für den gezielten Umgang mit Evaluationsergebnissen eingeführt, und in einer weiteren Agentur während der Laufzeit der vorliegenden Metaevaluation die Planung, Durchführung und Präsentation von Evaluationen im Rahmen eines übergreifenden Monitoring- & Evaluationssystems weiter professionalisiert. Organisatorische Anpassungen, die die Evaluationskapazität erhöhen, liegen auch an anderen Systemstellen vor, in unterschiedlichem Ausmaß und insgesamt nicht auf einem gleichen Niveau.

In allen Gesprächen mit AuftraggeberInnen und in der EvaluatorInnen-Befragung wurde ersichtlich, dass die über die Jahre durchgeführten Programmevaluationen als wesentliche Beiträge zu einer Verbreiterung und Vertiefung der Wissensbasis eingeschätzt werden, auf die sich FTI-politisches Handeln gerade auch als aktualitätsbezogenes und voranschreitendes Handeln in dynamischen Umwelten stützt. Zugleich wird auch deutlich, dass es sich beim Eintreten von Nutzen aus Programmevaluationen um Gemengelagen handelt, sodass nicht nur eine Evaluation zu einer Nutzung führt, sondern multiple Effekte auftreten. Freilich handelt es sich bei Umsetzungen von Erkenntnissen aus Programmevaluationen nicht um Automatismen, sondern um Handlungsweisen in Multiakteurs-Konstellationen im Einzelfall, bei denen auch immer wieder Reibungsverluste auftreten. Evaluationsnutzungen sind deutlich von den Konfigurationen der Principal-Agent-Beziehungen geprägt, die sich für die parallel agierenden Segmente des politisch-administrativen Handelns im FTI-Bereich unterschiedlich darstellen. Ob und wie Evaluationsergebnisse in einem dieser Steuerungsbereiche auch über den Kreis der unmittelbar mit einem Programm befassten Personen breiter bekannt gemacht und aufgegriffen werden, erweist sich als unsystematisch und stark vom Engagement von Einzelpersonen abhängig. Zusätzlich erhöht wird die Komplexität der Nutzenentstehung im Überstieg zwischen der administrativen und der politischen Sphäre, wobei auch hier von einer beträchtlichen Variation von Einzelfall zu Einzelfall auszugehen ist.

Als deutlich weniger ausgeprägt erweist sich eine Nutzung von Evaluationsergebnissen, die über die Grenzen der jeweiligen Steuerungsbereiche mit ihren Principal-Agent-Verhältnissen hinaus reicht. Obwohl auch hier relevante Wissenszuwächse beschrieben werden und in den Steuerungsbereichen jeweils davon ausgegangen wird, dass interessierende Information aus anderweitig durchgeführten Evaluationen zumindest prinzipiell zugänglich ist, gibt es keinen systematischen Vorgang im FTI-politischen Governance-System, der das Aufgreifen von in anderen Zuständigkeitsbereichen erbrachten Evaluationsergebnissen und die Auseinandersetzung damit unterstützt. Ausstrahlungswirkungen von evaluativer Information auf interessierte Fachöffentlichkeiten bzw. auf Akteursgruppen in der FTI-Landschaft treten in noch geringerem Maß auf und müssen als volatil gelten, da sie abgesehen von der grundsätzlichen Verfügbarmachung derjenigen Evaluationen, zu denen die Berichte publiziert werden, in aller Regel nicht Gegenstand gezielter Vorgehensweisen sind.

Im FTI-politischen System dienen Programmevaluationen auch durchaus dazu, andere Akteure in der politischen Sphäre vom evaluierten Programm zu überzeugen oder Entscheidungen über Programme zu rechtfertigen („symbolischer Nutzen“). Hier geht es um eine Überzeugungsarbeit, die in der Multiakteurs-Arena eines von differenzierten Principal-Agent-Verhältnissen, Hierarchien und Kleinteiligkeit geprägten Systems stets zu leisten ist, wenn es um die Zukunft von FTI-Programmen bzw. Steuerungsinstrumentarien geht. Zum Phänomenkomplex der Erzeugung von Evaluationsnutzen zählt freilich auch die Art der Verankerung der Evaluationsfunktion im rechtlich-institutionellen Rahmen der Bundesverwaltung. An der Schnittstelle zwischen Fachressorts und dem Bundesministerium für Finanzen (BMF) kommt den Programmevaluationen eine Funktion der Legitimation der Mittelausgaben zu. Diese Legitimationsfunktion ist im Motivbündel für die Planung und Durchführung von Programmevaluationen stets anwesend. Die Daten zeigen, dass eine symbolische Nutzung von Programmevaluationen andere Nutzenformen keineswegs ausschließt. Allerdings sorgt die gleichzeitige Anwesenheit von Lern- und Legitimationsfunktion für eine innere Spannung in jedem Evaluationsprojekt, die sich letztlich für eine Evaluationspraxis, die sich an der Erzeugung systematischer Wissenszuwächse im Governance-System orientiert, eher abträglich erweist.

Die Ergebnisse der Programmevaluationen können in ihrer Rolle als FTI-politische Informationsmittel als konkurrenzlos gelten, wenn sie auch oftmals nicht die alleinigen Grundlagen FTI-politischer Entscheidungen über Einsätze und Mittelzuweisungen sind. Im Verhältnis zu dieser Rolle von Evaluationsergebnissen ist das grundsätzliche Potenzial, bereits während der Evaluations-

durchführung und unabhängig von den Evaluationsergebnissen aus der Durchführung von Programmevaluationen unmittelbar zu profitieren („Prozessnutzen“), bislang nur wenig ausgeschöpft worden.

## **2.2. Einflussfaktoren auf die Nutzung von Programmevaluationen**

Faktoren, die in der bisherigen Evaluationspraxis Einfluss darauf gehabt haben, ob und wie Evaluationsergebnisse auch genutzt wurden, siedeln sich sowohl im Bereich dessen an, was innerhalb eines Evaluationsprojekts durch dessen Ausgestaltung beeinflusst werden kann, als auch im Bereich des umgebenden Kontexts, der außerhalb dieses Einflussbereichs verbleibt. Anhand von Daten aus der EvaluatordInnen-Befragung können die 20 wesentlichsten Einflussfaktoren bestimmt und gereiht werden. Sie finden in Auskünften von AuftraggeberInnen ihre Entsprechungen, wobei naturgemäß auch Perspektivunterschiede existieren sind und die EvaluatordInnen auch Faktoren bewertet haben, über die AuftraggeberInnen so nicht gesprochen haben. Diese 20 wesentlichsten Einflussfaktoren verteilen sich zu gleichen Teilen auf intrinsische Evaluationsmerkmale und auf Kontextfaktoren.

Unter jenen Faktoren, die sich im Verantwortungsbereich einer einzelnen Programmevaluation ansiedeln, steht die Glaubwürdigkeit der EvaluatordInnen bei den AuftraggeberInnen an erster Stelle. Diese Glaubwürdigkeit wird im österreichischen FTI-politischen Bereich durch die Heranziehung von auf FTI-Evaluation spezialisierten Instituten im In- und Ausland sowie durch die kontinuierliche Kommunikation von österreichischen FTI-EvaluationspezialistInnen mit den Auftraggeber-Institutionen in der Plattform fteval hergestellt. Ähnlich wichtig ist die Klarheit der Berichterstattung (Klarheit der Berichtsaussagen, Vorhandensein eines Executive Summary und dessen Aussagekraft), die von den FTI-EvaluatordInnen ebenso wie von deren AuftraggeberInnen als zentral erachtet wird.

Unter den Kontextfaktoren rangiert die Erwartung der AuftraggeberInnen, dass ihnen die konkrete Programmevaluation von Nutzen sein wird, an erster Stelle. Die bereits genannte Kombination von Lern- und Legitimationszwecken in der institutionell-rechtlichen Verankerung der Programmevaluationen kann als ein wesentlicher Grund dafür angesehen werden, dass deutliche Unterschiede im Umgang mit verschiedenen Programmevaluationen zu bemerken sind und immer wieder Fälle eingetreten sind, in denen Programmevaluationen von vornherein von ihren AuftraggeberInnen als notwendige Übungen betrachtet und kaum mit Nutzenerwartungen verbunden wurden, was sich dann von der Evaluationsplanung weg bis hin zum Umgang mit den Ergebnissen niederschlägt. Ebenfalls bedeutend für Art und Ausmaß der Nutzung von Evaluationsergebnissen ist der Umstand, ob eine Programmevaluation in direktem Konnex mit einem aktuellen FTI-politischen Entscheidungsbedarf oder Problemdruck steht. Ein derartiger Konnex besteht primär durch einen vorgegebenen Evaluations- und Verhandlungsrhythmus für Programmvereinbarungen der Ressorts mit dem Finanzministerium (BMF), während aktuelle FTI-politische Themenstellungen der Ressorts „Windows of Opportunity“ kaum als solche zum Auslöser von direkt auf sie gemünzten Evaluationsaktivitäten werden. Aktuelle Informationsbedürfnisse der Ressorts und Agenturen werden innerhalb dieses Rahmens des Öfteren nur in eingeschränkter Weise befriedigt.

Eine starke Personenabhängigkeit der genaueren Umgangsweise mit einzelnen Programmevaluationen und ihren Ergebnissen tritt in der EvaluatordInnen-Umfrage mit vier Faktoren massiv zutage. Bei diesem „Human Factor“ in der Evaluationsnutzung geht es um die persönlichen Denkstile der jeweils Evaluationzuständigen, um die Konsistenz der Evaluationsergebnisse mit ihren Sichtweisen und Erwartungen, um ihre Erfahrung mit Evaluation, und um ihre Rolle in der jeweiligen Institution. Des Weiteren kommt organisatorischen Anpassungen, Ressourcen und Erfahrungen der auftraggebenden Institutionen ein erheblicher Stellenwert zu. AuftraggeberInnen haben hierauf mindestens ebenso stark hingewiesen wie die EvaluatordInnen, für die diese Kontextfaktoren mit zu den einflussreichsten zählen.

Einige hoch relevante evaluationsmethodische Gesichtspunkte wie die Angemessenheit der Evaluationskriterien, eine ausgewogene Darstellung von Stärken und Schwächen des untersuchten Programms oder die Art des Evaluationsansatzes sind in den 20 wesentlichsten Einflussfaktoren auf eine Nutzenentstehung aus der Sicht der EvaluatordInnen enthalten. Sie fallen jedoch im Gesamtbild hinter einige stärkere Einflussfaktoren merklich zurück, die durch die Vorgehensweise einer Evaluation nicht beeinflusst werden können. Methoden Aspekte im engeren Sinn, wie die Anwendung eines Methodenmix, Triangulation oder die Finesse, mit der bestimmte Methoden eingesetzt werden, kommen unter den 20 wesentlichsten Einflussfaktoren auf Evaluationsnutzung, so wie die EvaluatordInnen sie einschätzen, nicht vor. AuftraggeberInnen sind auf evaluationsmethodische



Aspekte nicht in einer vergleichbaren Detailliertheit eingegangen, haben aber doch gelegentlich auf Mängel hingewiesen, die in der Vergangenheit die Entstehung von Nutzen aus Programmevaluationen beeinträchtigt haben und in den Einzugsbereich der Methodenanwendung fallen. Insgesamt erhärtet sich das Bild, dass die traditionell vor allem in Methodendiskussionen verankerte FTI-Evaluation die tatsächliche Entstehung von Nutzen aus durchgeführten Evaluationen nur in untergeordneter Weise diesem Hauptfokus ihrer Thematisierung der evaluatorischen Vorgehensweisen verdankt.

### **2.3. Nützlichkeit der Evaluationsberichte und Evaluationsprozesse**

Die analysierten Evaluationsberichte entsprechen den herangezogenen DeGEval-Standards auf einem im Großen und Ganzen mittleren Niveau, und mit voranschreitender zeitlicher Entwicklung zunehmend besser. Verbesserungspotenzial ist dennoch vorhanden, wenn es um bestmögliche Programmevaluationen geht, die hohe Nützlichkeit erzielen und das im “Unternehmen Programmevaluation“ angelegte Potenzial bestmöglich ausschöpfen. Eine sehr gute Erfüllung eines der 11 herangezogenen Standards konnte nur in einigen wenigen Fällen attestiert werden. Ebenso selten ist zugleich eine völlige Nichterfüllung eines der Standards, die auch in den letzten Jahren nicht mehr auftritt. Während zu allen herangezogenen Standards grundsätzlich noch Verbesserungspotenzial besteht, erscheinen die folgenden Gesichtspunkte als die relevantesten, um künftig noch nützlichere Programmevaluationen zu erzielen.

Die analysierten Programmevaluationen waren mit Ausnahme einer ex post-Evaluation Zwischenevaluationen oder Teile von Begleitevaluationen. Sie waren in den meisten Fällen sehr breit angelegt, Fragen von der Relevanz der Programme über ihre Effektivität bis hin zu ihrer Wirkung sollten verfolgt werden (sogenannte Multi-Purpose Evaluationen). Es wurden Outputs, Outcomes, und erste Wirkungen der Programme untersucht, sodass Erkenntnisse über die Programme durchaus erzielt wurden. Die Beobachtbarkeit von Programmwirkungen war auf Grund der früh gewählten Evaluationszeitpunkte fast immer deutlich eingeschränkt. Jedoch ist auch hinsichtlich dessen, was zu den Evaluationszeitpunkten bereits grundsätzlich zu den Programmen beobachtbar war, festzustellen, dass in vielen Fällen nicht von einer umfassenden und gründlichen Aufarbeitung der Programme gesprochen werden kann.

Die umfangreichen Evaluationsvorhaben wurden anhand von erhältlichen Monitoringdaten und weiteren, innerhalb der einzelnen Programmevaluationen jeweils selbst erhobenen Daten durchgeführt, die allerdings des Öfteren doch keine analytisch konsequente Ausleuchtung aller Programmkomponenten zuließen. Die Gesamtebenen aller relevanten Programmoutputs und -outcomes, die schrittweise hin zur Erreichung der Programmziele führen sollen, und insbesondere die Verbindungen zwischen diesen Ebenen, wurden nur mit teils deutlichen Einschränkungen greifbar gemacht (Standard N4). Aufgrund dieser Ausschnitthaftigkeit haben die meisten der untersuchten Programmevaluationen letztlich doch den Charakter einer sogenannten „black box“-Evaluation, durch die die genaue Art und Weise, wie ein Programm die intendierten Wirkungen erzielt bzw. an der Erzielung dieser Wirkungen gehindert ist, nicht oder zumindest nicht vollständig erfasst wird. Es zeigen sich des Öfteren Schwierigkeiten mit einer konzisen Gliederung von Programmkomponenten und Umsetzungsschritten zu Zielen unterschiedlicher logisch-hierarchischer Stellung (unmittelbare, intermediäre und übergeordnete Programmziele) und hinsichtlich der Art der Erreichung von direkten und indirekten Zielgruppen. Zugleich haben einige Evaluationen auch Fragestellungen behandelt, die nicht als zentrale Gesichtspunkte für ein tieferes Verständnis des evaluierten Programms zu erachten sind. Wie die EvaluatorInnen angeben, waren Auswahl und Umfang der in den Programmevaluationen herangezogenen Informationen häufig nicht ausreichend, um alle mitgegebenen Evaluationsfragen gut behandeln zu können, und noch weniger, um auch unbeabsichtigte Wirkungen der Programm erfassen zu können. Einige Evaluationsberichte tragen Züge eines „evaluability assessment“, in dem die Bedingungen für eine zielführende Evaluation des Programms erst geklärt werden.

Fast alle analysierten Programmevaluationen haben sich auch mit dem Kontext der evaluierten Programme auseinandergesetzt, in unterschiedlicher Intensität und mit unterschiedlichen Perspektivierungen. Vor allem auf der Basis von qualitativen Untersuchungsstrategien wurden von manchen Evaluationen essentielle Randbedingungen greifbar gemacht, unter denen das jeweilige Programm in seinen Zielgruppen Wirkungen erreichen konnte bzw. daran gehindert war. Etliche Kontextanalysen leiden jedoch darunter, dass zwar einige Faktoren untersucht und für skizzenhafte Bilder fruchtbar gemacht wurden, aber der systematische Stellenwert dieser untersuchten Faktoren

unklar bleibt bzw. keinen expliziten Bezug zu einer strukturierten und gesamthaft verstandenen Wirklogik des jeweiligen Programms aufweist (G2).

Die Evaluationsberichte geben trotz regelmäßig enthaltener Methodenbeschreibungen in der Mehrzahl doch nur unzureichend Auskunft darüber, was warum untersucht wurde, und als wie vollständig und tragfähig die erbrachten Ergebnisse eingeschätzt werden können. Im Verein mit nur sehr breiten und allgemein gehaltenen Angaben über die Untersuchungsschwerpunkte (Standard N2) und einer bemerkenswerten Enthaltensamkeit bei der Angabe von Evaluationsfragestellungen, die den jeweiligen Programmevaluationen zugrunde gelegt waren, ergibt sich so eine nur eingeschränkte Transparenz der Evaluationsergebnisse (Standard G3) und der Schlussfolgerungen, die aus ihnen gezogen wurden (Standard G8). Eine Transparenz der Vorgehensweise erscheint jedoch vor allem von Bedeutung, damit Evaluationsergebnisse auch von Akteuren aufgegriffen werden können, die nicht zum engen Kreis derjenigen Wenigen zählen, die unmittelbar mit der Konzeption und Umsetzung des untersuchten Programms und der dazu durchgeführten Evaluation befasst sind.

Ist anhand der Evaluationsberichte wegen ihrer Gestaltungsweise die Frage oft nicht gut beantwortbar, wie essentiell die erbrachten Ergebnisse im Hinblick auf die Gesamtlogiken der evaluierten Programme jeweils tatsächlich sind (N4, G3), so erscheint ebenso die Frage virulent, wie Programmen insgesamt Wert zugemessen wurde (N5). Hier offenbart sich ein „blinder Fleck“ eines stark datenorientierten und zugleich oftmals eher unsystematischen Zugangs. Während manche Evaluationen nachvollziehbare Bewertungsmaßstäbe in konsequenter Weise in Anschlag gebracht haben, die in einer klaren Verbindung zu den Programmzielen standen, haben andere eher für sich stehende Einzelbewertungen zu einzelnen Beobachtungen vorgenommen, ohne dass in der Kombination von „üblichen“ Betrachtungsweisen ein stringentes Gesamtkonzept greifbar würde. Es wird in der internationalen Evaluationstheorie allerdings davon ausgegangen, dass die Wahl der Bewertungsmaßstäbe ebenso eine tragende Säule jedes Evaluationskonzepts darstellt wie ihre Wissenschaftlichkeit und ihre gezielte Auseinandersetzung mit dem intendierten Nutzen.

Im Zusammenhang mit eingeschränkten Datenlagen waren die EvaluatorInnen immer wieder bestrebt, Lücken durch ihr Hintergrundwissen über das FTI-System und Annahmen über dessen Funktionsweisen oder Eigenschaften von Akteursgruppen wett zu machen (F3, G8). Dies beeinflusste oft merklich den Charakter von Schlussfolgerungen und Empfehlungen, die in unterschiedlicher Weise, aber doch teils recht deutlich, einen Zug von ExpertInnengutachten tragen, in denen das persönliche Wissen der AutorInnen zur Geltung gebracht wird. Dies deckt sich nicht mit dem Grundansatz der Evaluationsstandards, dass alle Aussagen einer Programmevaluation in transparenter Weise in von ihr herangezogenen Fakten und Quellen abgestützt sein sollten.

Festzustellen ist schließlich, dass in den Evaluationsprozessen zahlreiche Schritte, die den Standards zufolge vor allem im Planungsstadium einer Evaluation erfolgen können bzw. sollten, bislang nur ansatzweise wahrgenommen wurden. Hier zeigen sich unter Rückgriff auf Ergebnisse der EvaluatorInnen-Befragung unter anderem deutliche Verbesserungsmöglichkeiten bei der gezielten Auseinandersetzung damit, wie das Evaluationsprojekt auf eine konkret intendierte Nutzung zugeht (N8), und wie es entsprechend LernpartnerInnen einbindet (N1). Innerhalb von kurzen Vorbereitungsphasen der Evaluationen (Beantwortung von Terms of Reference und Hearing) kam es nur eingeschränkt zu einer Mitsprache der EvaluatorInnen, im Rahmen derer sie auf Basis ihrer Kompetenzen die Herangehensweise der Evaluation beeinflussen und schärfen konnten (N4). Über weite Strecken wurden Methoden zum Einsatz gebracht, die von den AuftraggeberInnen erwünscht waren oder von den EvaluatorInnen bzw. ihren Instituten regelmäßig eingesetzt werden. Lediglich in zwei der untersuchten Berichte wurden ungewöhnliche und innovative Methoden eingesetzt, die für die spezifische Aufgabenstellung der betreffenden Programmevaluation als produktiv erachtet wurden. Bewertungsmaßstäbe zur Einschätzung der Programme und Kriterien zur Einordnung von Beobachtungen wurden nur selten zwischen EvaluatorInnen und AuftraggeberInnen vorab gemeinsam geklärt (N5). Die Wahl von Bewertungsmaßstäben wurde oft den EvaluatorInnen überantwortet, und diese zogen entweder Maßstäbe heran, die in ihren Augen denen der AuftraggeberInnen entsprachen, oder verhielten sich unabhängig von solchen Annahmen.

## 2.4 Aktuelle Herausforderungen in der FTI-politischen Arena

Früher gehegte Erwartungen an die Leistungskraft von Programmevaluationen wurden als unrealistisch erkannt. Die verfügbare Ressourcenausstattung von Programmevaluationen wird als wesentlicher Mitgrund dafür erachtet, dass immer wieder Informationsbedürfnisse nur eingeschränkt befriedigt werden konnten. Die Verankerung der Programmevaluationen als Bestandteile der Programmvereinbarungen erzeugt eine Spannung zwischen vorgegebenen Evaluationsfragestellungen und aktuellen Informationsbedürfnissen in einem hochdynamischen System, die wiederholt auf Kosten aktuell relevanter Erkenntnisse gegangen ist. Die Evaluationsfunktion ist im Governancesystem an feststehende Evaluationszeitpunkte und –budgets gebunden, die gemäß den Auskünften der AuftraggeberInnen zwar bisweilen mit einer gewissen Flexibilität gehandhabt werden können, im Großen und Ganzen aber jedenfalls enge Grenzen setzen. Evaluationsprojekte oder Studien evaluativen Charakters, die nicht in Programmdokumenten vorprogrammiert waren, wurden nur in seltenen Ausnahmefällen initiiert.

In allen Ressorts und Agenturen wurden Zuständigkeiten und Kapazitäten geschaffen, um Evaluationen durchführen und Evaluationsergebnisse auf einer strategischen Ebene handhaben zu können. Die Planung und Durchführung der Programmevaluationen, die primär in den Ressorts erfolgt, ist dort an die Fachzuständigkeiten für die evaluierten Programme gekoppelt. Abstimmungsprozesse intern und innerhalb der Principal-Agent-Beziehungen sind erforderlich, die im allgemeinen auf Grund vorhandener Kooperationsbereitschaft erfolgreich verlaufen, aber doch keiner institutionell klar verankerten Systematik folgen. Das Engagement, das für eine einzelne Programmevaluation aufgebracht wird, bemisst sich nicht zuletzt an den zum jeweiligen Zeitpunkt gegebenen Möglichkeiten der fachzuständigen Einzelpersonen im Rahmen auch anderer Agenden. Durchgehend wird dargestellt, dass im Rahmen der gegebenen Kapazitäten keine weiteren Spielräume mehr bestehen.

Die Weitergabe von Evaluationsergebnissen innerhalb der Hierarchien stellt sich als geregelter Vorgang dar. Dabei wird davon ausgegangen, dass Evaluationsergebnisse nur eine Informationsquelle unter vielen sind, auf die sich politische EntscheidungsträgerInnen stützen, und dass auch die politische Aufmerksamkeit für unterschiedliche Programme deutlich variiert. Eine Zirkulation von Evaluationsergebnissen hin zu anderen Fachabteilungen, die zur Stärkung der Wissensbasis in systemischer Hinsicht beiträgt, bemisst sich stark am Engagement von Einzelpersonen. In jüngster Zeit sind verstärkte Bemühungen zu beobachten, durch übergreifende hausinterne Präsentationen Evaluationsergebnisse in Umlauf zu setzen und Diskussionen zu initiieren, in denen auch nicht direkt mit dem evaluierten Programm befasste Abteilungen von den Evaluationsergebnissen profitieren können und strategische Einschätzungen vorgenommen werden können. Eine institutionelle Verankerung derartiger wertvoller Vorgänge ist allerdings nicht gegeben, und eine durchgehende Systematik liegt nicht vor.

Im Rahmen der institutionellen Architektur bestehen einige wenige Berührungspunkte zwischen den Steuerungssegmenten im FTI-politischen Bereich, in denen zumindest potenziell Informationen über geplante und fertiggestellte Evaluationen ausgetauscht werden können. In erster Linie sind es jedoch Personen und Netzwerke, die einen übergreifenden Wissensfluss im Governance-System gewährleisten, sodass sich ein solcher Wissensfluss letztlich als akzidentiell darstellt.

Es besteht allseitiger Bedarf an verstärkt systemisch orientierten Erkenntnissen, durch die die Positionierung einer Maßnahme im breiten Kontext verschiedener Förderungs- und Steuerungsinstrumente ebenso aufgezeigt werden kann wie Optionen, in welcher Weise Bedarfslagen im systemischen Gesamtzusammenhang durch Einsatz und Konfiguration bestimmter Instrumente und Maßnahmen gezielt und in bestmöglicher Weise begegnet werden kann. In einzelnen Systemsegmenten besteht zudem Bedarf an Typen von Politikinformation, die mit den routinisierten Multi-Purpose-Evaluationen nicht gut abgedeckt werden können. Es geht hier (1) um intensivere Auseinandersetzungen mit Zielgruppen und Wirkungsweisen von Maßnahmen auf einer detaillierten Ebene, die in Richtung einer wissenschaftlichen Begleitforschung weisen, (2) um ein möglichst frühzeitiges Erkennen der Realitätshaltigkeit von Annahmen über die Wirkungsweise von Programmen, und (3) um hoch reaktive und schlanke Studien evaluativen Charakters, die das FTI-politische Handeln in dynamischen Umwelten zeitnah unterstützen.

### 3. Empfehlungen

Es wird auf Basis der drei Datenquellen und deren integrierender Analyse ersichtlich, dass es sich bei der Frage der Evaluationsqualität in Bezug auf Nützlichkeit und tatsächlich zustande kommenden Evaluationsnutzen nicht um Einzelursachen handelt, sondern um Syndrome und Faktorenbündel von erheblicher Komplexität. Die bisherige Evaluationspraxis im FTI-Bereich erweist sich als gleichermaßen durch Gestaltungsmerkmale einzelner Programmevaluationen wie durch Kontextfaktoren bedingt. Damit ist auch nicht die *eine* Lösung greifbar, die eine entscheidende Weiterentwicklung über den bisher erreichten Stand hinaus bewirken könnte.

Limitierungen für die Gestaltung von Programmevaluationen und die Entstehung von Evaluationsnutzen ergeben sich aus Merkmalen des institutionellen Arrangements. Evaluationsberichte und die hinter diesen Produkten stehenden Evaluationsprozesse gehen auf das, was die Evaluationsstandards als optimale Schritte hin zu hoher Nützlichkeit bezeichnen, bislang nur bedingt zu. Damit lassen auch Evaluationsprodukte und –prozesse Nutzen entstehen, die die von den Standards empfohlenen bzw. als notwendig erachteten Evaluationseigenschaften nicht optimal verwirklichen. Unter dem Gesichtspunkt einer größtmöglichen Nützlichkeit auf der Basis hervorragender Evaluationsqualität muss es zweifellos angelegen sein, von einer strukturell kompromisshaften Situation zu verbesserten Bedingungen für die Planung, Durchführung, Kommunikation und Nutzung von Programmevaluationen zu gelangen. Die Evaluationsstandards sind als praktische Anleitung zur Bewältigung von Problemen bei der Nutzenentstehung konzipiert, doch können sie Probleme nicht lösen, die außerhalb der Reichweite eines konkreten Evaluationsprojekts liegen, und gute Lösungen entlang der Standards müssen von EvaluationsauftraggeberInnen in der Gestaltung der Evaluationsaufträge auch ermöglicht werden. Strukturell ermöglichte Potenziale für die Planung, Durchführung, Kommunikation und Nutzung von Programmevaluationen bleiben sodann in den jeweiligen Projekten auf der Basis von Kapazitäten und Kompetenzen auszufüllen.

Die Metaevaluation gelangt daher zu Empfehlungen, die sich sowohl auf einer evaluations-theoretischen Ebene als auch auf der Ebene der institutionellen Einbettung der Evaluationsfunktion ansiedeln. Die vorgelegten Empfehlungen sind an der Weiterentwicklung einer in sich dynamischen und systemevolutiven Evaluationspraxis orientiert. Da die bisherige Evaluationspraxis in nachvollziehbarer Weise bereits Nutzen erzeugt hat, setzen die Empfehlungen nicht auf eine radikal-disruptive Veränderung, die aus einer Orientierung an Governancemodellen anderer Länder grundsätzlich abgeleitet werden könnte, aber hinsichtlich tatsächlicher Transferierbarkeit und Eintreten der erhofften Effekte doch auch mit einigen Ungewissheiten einhergeht. Für die evaluationsmethodische Ebene würde eine Benennung aller denkbaren Verbesserungsoptionen freilich darauf hinauslaufen, den gesamten Gehalt der Standards zu referieren. Diesbezügliche Empfehlungen werden nur für diejenigen Gesichtspunkte ausgesprochen, die als die wesentlichsten erscheinen. Letztlich beruht eine hoch entwickelte Evaluationskultur auch auf gesellschaftlich-kulturellen Faktoren wie der Offenheit für sachlich fundierte Kritik und der Bereitschaft zur offenen Diskussion, die sich freilich einer gezielten Beeinflussung entziehen.

Die folgenden 20 Empfehlungen werden ausgesprochen:

1. Programmevaluationen sollten in Zukunft weiterhin durchgeführt werden, da sie in der Vergangenheit wertvolle Beiträge zur zielgerechten Umgestaltung und Neukonzeption FTI-politischer Maßnahmen erbracht haben, die noch über die Ebene der jeweils evaluierten Programme hinaus reichen. Um die Produktivität der Programmevaluationen über das bisherige Maß hinaus weiter steigern zu können, sollten sie mit den folgenden Empfehlungen benannten Schritten einhergehen.
2. Die derzeit gegebene Verankerung der Evaluationsfunktion bei den Institutionen, die für die Konzeption und Umsetzung von FTI-Programmen zuständig sind, sollte beibehalten werden. Entscheidende Kapazitäten für die Planung, Durchführung und Verwertung von Programmevaluationen wurden hier über Jahre hinweg aufgebaut. Die Verankerung bei den Programmverantwortlichen sorgt auch für ein Commitment zu den Programmevaluationen, das für in der Vergangenheit entstandenen Evaluationsnutzen wesentlich war. Eine Weiterentwicklung der Evaluationskultur im FTI-Bereich sollte als pfadabhängige Entwicklung auf dieser wertvollen Grundlage gedacht werden.

3. Programmevaluationen sollten in Zukunft mit denjenigen Ressourcen ausgestattet werden, die eine konzeptgemäße Analyse des evaluierten Programms unter Heranziehung aller für die Evaluationsschwerpunkte und –fragestellungen benötigten Informationsquellen tatsächlich ermöglichen und eine gute Durchführung gemäß dem Qualitätsverständnis der internationalen Standards für Programmevaluation gewährleisten.
4. Programmevaluationen sollten künftig stärker auf eingegrenzte Evaluationsschwerpunkte fokussiert werden. Dadurch können unter Bedingungen begrenzter Ressourcen intensivere und genauere Untersuchungen zu den gewählten Schwerpunkten durchgeführt werden. Jeweils nicht gewählte Evaluationsschwerpunkte können gegebenenfalls durch eine weitere Evaluation verfolgt werden. Dabei können dann auch andere Evaluationsteams zum Einsatz kommen, was zu einer Anreicherung der Sichtweisen auf das untersuchte Programm auf Basis unterschiedlicher Kompetenzen beitragen kann.
5. Programmevaluationen sollten verstärkt in ihrer Prozessqualität begriffen und auf dieser Ebene in Planung und Durchführung gestärkt werden. Die DeGEval-Standards mit ihrem Interpretationshintergrund der Joint Committee-Standards weisen auf Schritte hin, durch die im Planungs- und Durchführungsstadium von Programmevaluationen Qualität in unterschiedlichen Hinsichten gestärkt und sicherstellt werden kann. Die *Plattform fteval* sollte sich mit solchen Möglichkeiten auseinandersetzen, da sie Voraussetzungscharakter für die Erzielung späterer Evaluationsergebnisse und deren Nutzungspotenzial für verschiedene Akteursgruppen haben.
6. Evaluationsberichte sollten in jeder Hinsicht klar und in einer auch für Außenstehende gut verständlichen Weise abgefasst werden. Dies ist insbesondere als Voraussetzung dafür zu verstehen, dass es zu einer verstärkten Nutzung von Programmevaluationen in anderen FTI-politischen Bereichen und nach dem Denkprinzip einer vermehrten systemreferentiellen Selbststeuerung der FTI-Akteure kommen kann.
7. Alle Evaluationsberichte sollten systematisch ein Kapitel beinhalten, in dem die Gesamtvorgehensweise der Evaluation in methodischer wie organisatorischer Hinsicht konzipiert und vollständig dargestellt wird und auch auf Vor- und Nachteile der tatsächlich durchgeführten Analyse hingewiesen wird. Eine derartige kompakte Übersicht über die Gesamtvorgehensweise erscheint insbesondere hinsichtlich einer stärkeren Nutzung von Evaluationsergebnissen in einem gesamt-systemischen Zusammenhang relevant, damit auch Akteure, die mit den unmittelbaren AuftraggeberInnen nicht identisch sind, auf die erbrachten Evaluationsergebnisse gut zugreifen können. In der Darstellung der Vorgehensweisen sollte es auch Mut zum Ausweis von Lücken geben, da keine Programmevaluation alles beleuchten kann, was theoretisch zu einem Programm untersucht werden könnte. Auch ein abholbares Wissen darüber, was noch nicht intensiv untersucht werden konnte, sollte als produktiver Beitrag zum FTI-politischen Wissens- und Informationssystem betrachtet werden, damit dieses im Weiteren produktiv ausgestaltet werden kann.
8. Es sollte eine verstärkte Auseinandersetzung mit der gezielten Anwendung von Bewertungsmaßstäben auf die evaluierten Programme angestrebt werden. Dabei geht es nicht nur darum, wie Zielerreichungen gemessen und eingeschätzt werden, was oft zum Evaluationszeitpunkt in dieser Form noch gar nicht möglich ist, sondern auch und gerade um die wohlbegründete Einordnung der Beobachtungen zu Aspekten der Programmgestaltung. Die von einer Evaluationsstudie angewendeten Bewertungsmaßstäbe sollten als vitale Konzeptfrage begriffen und im Planungsstadium als integraler Bestandteil des übergreifenden Evaluationskonzepts vereinbart und festgelegt werden. Konsistente Bewertungsmaßstäbe verkörpern sich unter anderem in der Verfolgung von Kohärenz und Konsistenz von Programmzielen und Programmkomponenten in ihrer Umsetzung, in einer Festlegung, wie die Sichtweisen verschiedener Akteursgruppen auf das evaluierte Programm zur Gesamteinschätzung führen, in vorab festgelegten Kriterien zur Einordnung späterer Messergebnisse, in oder in der gezielten Bestimmung von Messgrößen (etwa bei einem Programmziel „Kooperation“ die Quantität von Kooperationsbeziehungen versus qualitative Eigenschaften von eingegangenen Kooperationen).
9. Dem Risiko eines Lock-Ins in üblichen Herangehensweisen an Evaluation, die mit Ermüdungserscheinungen der Evaluationspraxis in Zusammenhang stehen, sollte durch eine systematische professionelle Beratung von Evaluationsplanungen und –prozessen gegengesteuert werden. Eine solche Beratung wird vor allem dann ein probates Mittel darstellen, wenn sie nicht nur FTI-spezifische Kompetenzen heranzieht, sondern auch evaluationsmethodische Kompetenzen, die den Konnex zu Entwicklungen und Know-How anderer Bereiche herstellen.

10. Eine Intensivierung der Planungsphasen der Programmevaluationen sollte angestrebt werden, um das Risiko zu minimieren, dass beschränkte Ressourcen in letztlich ergebnisarme Untersuchungsschritte fließen. Dafür bietet sich das international anzutreffende Modell einer sogenannten „*Inception Phase*“ am Beginn einer Programmevaluation an, in der sich die beauftragten EvaluatorsInnen intensiv mit der Datenlage, methodischen Möglichkeiten im Rahmen der gegebenen Ressourcen, und der Beantwortbarkeit der vorgesehenen Evaluationsfragen auseinandersetzen. Diese genaue Abwägung bildet sodann die Grundlage für ein bestmögliches Evaluationsdesign, das der im Anschluss durchgeführten Evaluation zugrunde gelegt wird. Das Modell zielt darauf ab, so realistische Erwartungen wie möglich an eine Evaluation zu entwickeln und die für die Evaluation verfügbaren Ressourcen so gut wie möglich zu nutzen. Vergaberechtliche Voraussetzungen für die Nutzbarkeit dieses Modells bleiben zu prüfen.
11. Die institutionelle Verankerung der Evaluationsfunktion in den Ressorts und Agenturen sollte weiter gestärkt werden. In den auftraggebenden Ressorts und Agenturen existieren HauptansprechpartnerInnen für Evaluationsangelegenheiten und VertreterInnen der Institutionen in der *Plattform fteval*, doch ist bis heute keine dieser Personen ausschließlich mit Evaluationsangelegenheiten betraut, um sich dieser komplexen und anforderungsreichen Materie vollständig widmen zu können. Ressourcen von fachzuständigen MitarbeiterInnen für die Auseinandersetzung mit anderweitig erarbeiteten Programmevaluationen sind kaum vorhanden. Eine spezialisierte, hoch professionelle Evaluationsabteilung oder Stabstelle, die sich mit der Planung der Programmevaluationen, dem Evaluationsmanagement, einer Qualitätskontrolle und der Verwertung und Weitergabe der Evaluationsergebnisse für das ganze Haus befasst, stellt in diesem Zusammenhang das Idealbild dar, das einen entscheidenden Schritt zur Überwindung der Variabilität im Umgang mit einzelnen Evaluationen verkörpern würde.
12. Die Lernfunktion der Programmevaluationen sollte künftig durch eine Flexibilisierung der Auslösung und Intensität der einzelnen Evaluationen weiter gestärkt werden. Frei allozierbare Evaluationsbudgets könnten die Gestaltung von Programmevaluationen im Aktualitätsbezug sowie unter Gewichtung von Informationsbedarfslagen ermöglichen. Nicht alle Programme brauchen in gleicher Weise evaluiert zu werden, um in einem übergreifenden FTI-politischen Informationssystem wesentliche Erkenntnisse zu erzielen. Eine Flexibilisierung würde somit zu zielgerechten Investitionen in anspruchsvollere Evaluationen und Studien und zu einer effektiveren Nutzung der im System vorhandenen Ressourcen beitragen. Programmwirkungen könnten zu passenderen Zeitpunkten analysiert werden, als es bislang der Fall war. Thematische Evaluationen, etwa zu Programmfamilien oder Zielgruppen, und Instrumentenevaluationen könnten verstärkt durchgeführt werden.
13. Programmevaluationen sollten verstärkt als übergreifende und konzise Analysekonzepte verstanden und angelegt werden. Evaluationsmethodische Konzepte und Tools, die für eine möglichst zielführende Evaluation von Programmen über die letzten beiden Jahrzehnte international entwickelt wurden, sollten dabei herangezogen werden. Zu empfehlen ist eine Zuwendung zu Ansätzen, die unter dem Sammelbegriff der Theorie-basierten Evaluation (*theory-based evaluation*) bekannt sind. Diese Ansätze sind gezielt dafür konzipiert, die Einlösbarkeit von Programmannahmen in der realen Programmentfaltung zu beleuchten und geschärfte Umgangsweisen mit der Kausalitätsproblematik zu ermöglichen, wie und inwieweit ein Programm zu intendierten Veränderungen beiträgt. Mit der Zuwendung zu ihnen würde die Evaluationspraxis im FTI-Bereich Analysestrategien zur Anwendung bringen, die in anderen Politikbereichen auf internationaler Ebene und in internationalen Organisationen bereits eingesetzt werden. Die avancierten Ansätze der Realistischen Evaluation (*realistic evaluation*) und der *Contribution Analysis* könnten aufgegriffen werden, um zu einem vertieften Verständnis der Wirkungsweise von Programmen in ihrer Kontextabhängigkeit zu gelangen und komplexe Programme, Programmfamilien, Portfolien und Maßnahmenbündel zielführend und in pragmatischer Weise zu analysieren. Allseitige Ressourcen für die Arbeit mit qualitativen Daten und notwendige Interaktionen zwischen EvaluatorsInnen und AuftraggeberInnen während der Evaluationsdurchführung sind freilich vorausgesetzt.
14. Programmdokumente sollten so eingehend wie möglich darlegen, wie Zielsetzungen systematisch gegliedert sind, welche Outputs die verschiedenen Programmaktivitäten erzeugen sollen, und welche Annahmen darüber gemacht werden, wie diese Outputs zu Outcomes und weiteren Entwicklungen hin zu Zielerreichungen führen. Eine möglichst gute Darstellung der intendierten Wirkungsweise der Programme durch die Programmeigentümer bei der Programmkonzeption bildet den Gegenpol zur evaluatorischen Aufarbeitung einer Programmlogik und deren Ausgestaltung in der Programmwirklichkeit. Die Konzeptualisierung der intendierten

Wirkungsweise der Programme kann im Planungsstadium durch ex ante-Evaluationen unterstützt werden. Freilich können ex ante-Evaluationen spätere Überprüfungen nicht ersetzen, wie sich Programmefekte im realen Operieren des Programms herstellen oder mit Hindernissen konfrontiert sind.

15. Programmevaluationen sollten gemeinsam mit allen verwandten und ergänzenden Bestandteilen eines übergreifenden FTI-politischen Wissens- und Informationssystems durch Publikation verfügbar gemacht werden, um auch Synergien zwischen Studien unterschiedlichen Typs allgemein nutzbar zu machen. Die konkrete Bezeichnung von Programmevaluationen, Reviews, Assessments oder wissenschaftlichen Studien evaluatorischen Charakters sollte nicht zum Anlass werden, wertvolle Informationspotenziale zu beschneiden. Ein Repositorium für alle evaluativen und wissenschaftlichen Studien kann im Bedarfsfall in allgemein zugängliche Bereiche und Bereiche mit Zugangsbeschränkungen gegliedert werden. Nicht-Publikation ist gerechtfertigt und angebracht, wenn in einer systematischen Qualitätskontrolle zum Schluss gekommen wird, dass durch die Publikation unzuverlässige oder irreführende Information zur Nutzung freigegeben würde. Die Nutzbarkeit jedweder evaluativer Information wird von einer adäquaten Dokumentation über den genauen Charakter dieser Information abhängig bleiben. Im Verständnis der Evaluationsstandards ist jeder Nutzung von Evaluationsergebnissen eine umfassende Auseinandersetzung mit der genauer Vorgehensweise und Durchführungsqualität der betreffenden Evaluation vorausgesetzt. Eine bloße Verfügbarkeit von Datenbeständen, die unter nicht genau verstehbaren Ausgangsbedingungen in Bezug auf nicht genau bekannte Informationsbedürfnisse erarbeitet wurden, sollte nicht als ausreichend erachtet werden.
16. Jede Programmevaluation sollte bei ihrer Publikation von einer „Management Response“ begleitet werden, die die Kenntnisnahme der Evaluationsergebnisse auf Ebene des Top Managements bestätigt, eine Positionierung zu diesen Ergebnissen angibt, und damit auch Verbindlichkeit erzeugt. Dabei geht es nicht etwa um eine automatische Übernahme von Evaluationsergebnissen, sondern im Gegenteil um das Produkt einer aktiven Auseinandersetzung mit ihnen. Dieser Weg wird beispielsweise von der Deutschen Forschungsgemeinschaft (DFG) bereits beschritten und wurde neuerdings auch von einer Agentur im österreichischen FTI-Governancesystem eingeschlagen.
17. Der RFTE sollte die ihm zur Verfügung stehenden Mittel nützen, um in der evaluativen Wissensproduktion offen bleibenden Informationsbedarf durch gezielte Vergabe von Studien in aktualitätsbezogener und flexibler Weise zu befriedigen. Dies erscheint im Hinblick auf intensive Analysen zu Themen und Segmenten des FTI-Systems ebenso relevant wie im Hinblick auf übergreifende, systemisch ausgerichtete Analysen. Ein Charakter wissenschaftlicher Begleitforschung, die in den Ressorts und Agenturen keinen Ort hat, könnte dabei zum Tragen kommen. Im Hinblick auf den systemischen Stellenwert solcher Studien erscheint eine Abstimmung mit den relevanten FTI-politischen Akteuren sinnvoll und wichtig.
18. Eine Koordinationsfunktion für FTI-Evaluationen sollte geschaffen werden, die sich mit möglichen Synergiebildungen zwischen an verschiedenen Systemstellen angesiedelten Evaluationsaufgaben und -ressourcen befasst, um durch Abstimmungen und Beratungen die gegenwärtige Zersplitterung der Evaluationsaktivitäten und Kleinteiligkeit im Analytischen zu überwinden. Dadurch kann ein Potenzial ausgeschöpft werden, das aus einer Bündelung von Ressourcen und Erkenntnisinteressen resultiert. Ressort- und Agentur-übergreifende Abstimmungsleistungen könnten erbracht werden, deren Machbarkeit unter den gegebenen Bedingungen eingeschränkt ist. Erträge hinsichtlich stärker systemisch ausgerichteter Fragestellungen zum Stellenwert von einzelnen Maßnahmen und Steuerungen sind zu erwarten.  
  
Dies kann zugleich als sinnvolle Alternative zu ebenso seltenen wie schwer initiiierbaren Großunternehmungen wie der Systemevaluation 2009 erachtet werden, indem systemische Fragestellungen zum Gegenstand eines rollenden Verfahrens werden. Eine solche Koordinationsfunktion ist jedenfalls mit hohen fachlichen Kompetenzen und adäquaten Ressourcen auszustatten. Es bleibt zu prüfen, ob eine Einrichtung möglich ist, ohne bestehende Rechtsbestände anzutasten. Die Konfiguration und Einrichtung sollte durch eine Studie vorbereitet werden, die sich mit internationalen Beispielen auch außerhalb des FTI-politischen Bereichs befasst.
19. Ein Diskussionsforum sollte geschaffen werden, das Evaluationsergebnisse an ein breiteres Fachpublikum heranträgt, das über den engen Kreis der in der Plattform fteval versammelten Akteure hinausreicht und ProgrammangerInnen und Programmverantwortliche an unterschiedlichen Systemstellen genauso anspricht wie Akteursgruppen im FTI-System. Hierdurch

können Wissensflüsse in Gang gesetzt und Diskussionen ausgelöst und angereichert werden, die für ein System systemreferentieller und selbstreflexiver Akteure relevant sind. Der derzeitigen starken Abhängigkeit von Wissensflüssen im FTI-politischen Governancesystem von Personen und Netzwerken würde damit gegengesteuert. Ebenso würde der Umstand, dass auf der Basis einer bloßen Publikation evaluative Information eine Holschuld für etwaige InteressentInnen bleibt, behoben. Ein solches Diskussionsforum kann optional mit der vorgenannten Koordinationsfunktion verbunden werden, aber auch eine getrennt angesiedelte Systemfunktion darstellen.

20. Hinsichtlich einer substantiellen Stärkung der Lernfunktion der Programmevaluationen ist die derzeitige Kombination der unterschiedlichen Evaluationszwecke des Lernens und der Rechenschaftslegung, die für die Programmevaluationen durch deren institutionell-rechtliche Verankerung als Schnittstellenfunktion zwischen Fachressorts und Finanzressort stets gegeben ist, nicht als produktiv zu erachten. Nachdem mit der Wirkungsorientierten Folgenabschätzung (WFA) eine andersartige Evaluationsfunktion im Bezug auf Rechenschaftslegung geschaffen wurde, könnte überlegt werden, inwiefern die Lernfunktion der Programmevaluationen von Zwecken der Rechenschaftslegung künftig getrennt werden kann. Zwecke der Programmdokumentation könnten verstärkt in die Hände der Agenturen gelegt werden, die bereits jetzt wesentliche Teile der Datenbasen erarbeiten, die in Programmevaluationen verwendet werden. Im Gegenzug könnten Evaluationen dann verstärkt Analyseschritte setzen, die nicht der Gefahr eines Lock-Ins in vorab festgelegten Datenstrukturen ausgesetzt sind.