

Zwischen Schema F und Innovation: Eine politikfeldübergreifende Diskussion zu methodischen Standards

*Thomas Altenburg*¹

Der Arbeitskreis (AK) Methoden in der Evaluation der DeGEval führte im Rahmen der DeGEval-Jahrestagung 2016 in Salzburg eine Session zum methodischen Erfahrungsaustausch durch. Mit dieser Auftaktveranstaltung wurde der Startschuss für eine politikfeldübergreifende Bestandsaufnahme der verschiedenen Evaluationsansätze und -standards innerhalb der DeGEval gegeben. Ziel ist es, Gemeinsamkeiten und Unterschiede in Evaluationsmethoden und methodische Trends in den jeweiligen Arbeitskreisen der DeGEval herauszuarbeiten, deren Ursachen festzustellen und mögliche Synergieeffekte oder Innovationen auszuloten.

Um einen intensiven Erfahrungsaustausch zu ermöglichen, wurde das Konzept der Fishbowl-Diskussion gewählt. Neben drei festen Diskussionsteilnehmerinnen

- *Vera Hennefeld* (Bereichsleitung Bildung und Kultur, CEval GmbH) – aktuell stellvertretende Sprecherin des AK „Kultur und Kulturpolitik“,
- *Sonja Sheikh* (stv. Direktorin der KMU Forschung Austria) – von 2005 bis 2012 Sprecherin des AK „Forschungs-, Technologie- und Innovationspolitik“ und
- *Susanne von Jan* (stv. Leiterin Evaluierung, Hanns-Seidel-Stiftung e.V.) – aktuelle Sprecherin des AK „Entwicklungspolitik“,

welche exemplarisch die Evaluationserfahrungen aus ihren Arbeitskreisen präsentierten, konnten hierdurch auch die Seminarteilnehmerinnen und Seminarteilnehmer (vorübergehend) in die Diskussionsrunde gelangen und ihre individuellen Sichtweisen und Erfahrungen schildern. Die Diskussion wurde von *Alexandra Caspari* und *Tobias Polak* (Sprecherteam des AK Methoden) moderiert, um sicherzustellen, dass möglichst viele Gäste am aktiven Gedankenaustausch teilnehmen konnten.

Bestandsaufnahme

Vera Hennefeld betonte in ihrem Eingangsstatement, dass im Arbeitskreis **Kultur und Kulturpolitik** zwei Handlungsfelder bestehen, welche getrennt voneinander

1 Berlin

betrachtet werden müssen: Evaluation von Kultur im Sinne von Projekten im Bereich Kunst und Kultur sowie Evaluation im Bereich Kulturpolitik im Sinne von (Auswärtiger) Kultur- und Bildungspolitik. Im Bereich Kultur wird im Vergleich zur Kulturpolitik weniger evaluiert. Gerade in der Kultur ist die Akzeptanz von Evaluation nach wie vor eine große Herausforderung: Notwendig ist hier oftmals die Aufklärung über die Ziele von Evaluation, welche ausdrücklich nicht eine Bewertung von Kultur und Kulturschaffenden selbst beabsichtigt, sondern die Analyse und Bewertung von einzelnen Projekten und deren Zielerreichung. Dieses von Gegnern der Evaluation gerne ins Feld geführte Missverständnis führt vermutlich zu den bestehenden Vorbehalten gegenüber Evaluationsprozessen. Im Bereich Kultur – so *Vera Hennefeld* – werden vornehmlich gängige qualitative Methoden (beispielsweise in Form von Fallstudien) angewandt, aber auch innovative quantitative und qualitative Ansätze werden verfolgt. Beispielsweise wurden kognitive Landkarten genannt, die auch auf andere Evaluationsfelder übertragbar sind. Gegenüber quantitativen Methoden existieren eher Vorbehalte. Eine Ausnahme bildet hier die weit verbreitete und gut etablierte Museumsevaluation und Besucherforschung.

In der Kulturpolitik sind hingegen wesentlich mehr Evaluationen zu beobachten. Ein Schwerpunkt sind klassische Querschnittsevaluationen zum Projektende, wobei großteils qualitative (z.B. Fokusgruppen, leitfadengestützte Interviews) oftmals in Kombination mit quantitativen Methoden, wie z.B. standardisierte Online-Befragungen, angewandt werden. Bislang werden jedoch kaum experimentelle bzw. quasi-experimentelle Designs realisiert. Auch Längsschnittanalysen sind eher eine Ausnahme.

Wünschenswert wäre in beiden Bereichen, wenn Evaluation verstärkt zum Start einer Maßnahme mitgedacht und so mehr Möglichkeiten zum Einsatz von anspruchsvolleren und rigoroseren Designs gegeben wären.

Im Handlungsfeld der **Forschungs-, Technologie- und Innovationspolitik** (FTI) gab *Sonja Sheikh* für den geografischen Bereich Österreich Auskunft – der sich stark von FTI-Evaluationen in Deutschland unterscheidet. Sie erläuterte zu Beginn den gleichwohl großen Evaluationsumfang, zum Teil wird gar für *ein* Programm eine Zwischen- *und* Endevaluation realisiert, welche dann bewusst von verschiedenen Evaluator(inn)en ausgeführt werden müssen. Im Fokus der Evaluation steht meist die Analyse des Gegenstandes, der Umsetzung sowie der Wirkungen, wobei letzteres das ‚Herzstück‘ ist. Der Fokus liegt hier auf einem Methodenmix aus Tiefeninterviews, Logik-Chart-Analysen, Standardbefragungen und/oder Fokusgruppen. Oft sind dabei sehr lange Zeiträume betroffen, wobei das Programmziel eher selten stringent dargestellt wird. Indikatoren für Programmwirkungen sind beispielsweise Publikationen, Veröffentlichungen oder Patente, die mittels einer Befragung erhoben werden. Experimentelle Designs kommen kaum zum Einsatz, obwohl diese unregelmäßig auch in Auftragsbeschreibungen benannt werden. Aus Ressourcengründen, aber auch mangels neutraler Kontrollgruppen, können solche Designs auch kaum realisiert werden. Ein aktueller Trend zeichnet sich derzeit dahingehend ab, dass häufiger dynamische Prozesse (beispielsweise Stakeholder-Workshops als World-Café) im Rahmen von Evaluierungen zum Einsatz kommen.

Im Politikfeld der **Entwicklungszusammenarbeit** (EZ) ist es – vor allem aufgrund der starken Heterogenität von Themen, Ländern und Akteuren – schwierig, von *einem* methodischen Standard zu sprechen, so *Susanne von Jan*. Dennoch können einige Evaluationsmethoden als typisch angesehen werden. Sie nennt hier das Dokumentenstudium (was sie von der Dokumenten*analyse* dezidiert abgrenzt) und qualitative Instrumente (Leitfadeninterviews, semistrukturierte Interviews, Gruppenbefragungen und -diskussionen). Aber auch standardisierte Interviews oder partizipative Methoden (PRA, Mapping) sowie Beobachtungen werden umgesetzt. Standardmäßig werden angestrebte Veränderungen gemessen, allerdings eher als Soll-Ist-Vergleich im Sinne von Zielerreichung, häufig auch über retrospektive Fragen. (Quasi-)Experimentelle Designs oder randomisierte, kontrollierte Studien zur Wirkungsmessung werden zwar auch in diesem Handlungsfeld realisiert, jedoch nicht als Standard. Als ein Zukunftstrend erscheint die vermehrte Erhebung und Nutzung von Baseline-Daten, so dass zukünftig sicherlich verstärkt Vorher-Nachher-Messungen zu erwarten sind. Als innovative Methodologien für die Zukunft gelten aktuell neue Technologien sowie die Nutzung von Big Data. So könnten Umfragen über mobile Endgeräte realisiert oder große Daten aus aktuellen Satellitenbildern (beispielsweise Nachtaufnahmen zur Auswertung der Helligkeit einzelner Regionen als Proxyindikator für Armut) genutzt werden. Im Zusammenhang unter anderem mit den SDG ist auch die stärkere Nutzung von Sekundärdaten in der Diskussion. Abschließend gab sie noch einen Hinweis zum Monitoring: Während das Aktivitäten-Monitoring durchaus stark ausgeprägt ist, stellt das wirkungsorientierte Monitoring viele Projekte noch vor große Herausforderungen.

Diskussion

Im Rahmen der anschließenden Diskussion wurden verschiedene Aspekte vertieft. Zunächst wurde betrachtet, **wer Evaluationen durchführt** und wie die **Kommunikation und Kooperation mit der Wissenschaft** aussieht. Die Referentinnen erläuterten beispielsweise, dass Evaluationen im Kulturbereich, in der Entwicklungszusammenarbeit und FTI-Politik meist durch freie Einzelgutachter aber auch außeruniversitäre Einrichtungen wie Consultings oder private Beratungsunternehmen und außeruniversitäre Forschungseinrichtungen durchgeführt werden. Eine große Ausnahme sind Universitäten, wodurch sich der wissenschaftliche Einfluss auf die Evaluationen schwer einschätzen lässt. Allerdings sind einige Consultingunternehmen an Universitäten angesiedelt oder die Akteure sind Professor(inn)en, die die Evaluationen allerdings nicht in ihrer universitären Rolle durchführen.

Ein zentraler Diskussionspunkt betraf die Frage der **Methodenqualität**, d.h. wie ‚sauber‘ wird gearbeitet und was bedeutet ‚methodisch sauberes‘ Arbeiten. Hieraus ergaben sich auch die Frage der **Darstellung und Transparenz** der bei Evaluation genutzten Methoden sowie die Frage, inwieweit **Plausibilitätsüberlegungen** akzeptabel sind.

Dorothee Mack (Bischöfliches Hilfswerk Misereor e.V.) berichtete, dass eine Metaevaluation aus 2016 gezeigt hat, dass die genutzten Methoden selten dargestellt werden und auch in den meisten Berichten nicht erkennbar ist, dass Daten ‚sauber‘ analysiert wurden – hier ist „sehr viel Luft nach oben“. Die zentrale Frage lautet daher, was muss und was kann ein Bericht enthalten? Aus Auftraggebendensicht zeigt sich hier – so *Dorothee Mack* – ein großes Dilemma, da in einem ständigen Balanceakt die knappen Ressourcen entweder in die Methodendarstellung oder Ergebnisdarstellung investiert werden müssen: Die wünschenswerte Qualität in der Methodenanwendung und Datenauswertung treibt jedoch die Kosten schnell in die Höhe und steht dann in keinem Verhältnis mehr zu den Projektkosten. Daher muss pragmatisch vorgegangen werden. So wird durchaus in Auswertungsgesprächen eine sehr saubere Methodenanwendung deutlich, was dann aber aus Kostengründen in den Berichten nicht ausführlich dokumentiert wird. D.h. ‚sauberes Arbeiten‘ zeigt sich insbesondere in einer klaren Verbindung von beschriebener Datenlage zu Schlussfolgerungen, d.h. wie deutlich können einzelne Schlüsse wirklich aus den Daten gezogen werden. Häufig fehlen jedoch in Berichten Zahlen zu Stichprobengrößen. Dann bleibt offen, ob es sich um eine überzeugende Einzelmeinung handelt oder ob diese Meinung trianguliert wurde. Diese Verbindung ist in Berichten teilweise vorhanden aber nicht immer eindeutig dargestellt, wird dann aber in Gesprächen geklärt, so *Dorothee Mack*.

Erwin Geuder-Jilg (DEval) führte aus der Perspektive des freien Gutachters, welcher auch für Misereor tätig war, an, dass z.B. die Bestimmung von Stichprobengröße und Stichprobenauswahl nicht in zwei bis drei Wochen erarbeitet werden kann, sondern ausreichend Vorlaufzeit benötigt. Als positives Beispiel nennt er eine Evaluation in Brasilien, die zwei Missionen umfasste. So konnte im Rahmen der ersten Mission die Gesamtmenge bestimmt werden und daraus die Stichprobe systematisch und transparent herausgearbeitet werden. Ausreichend Zeit für die Entwicklung einer systematischen Methodik ist allerdings seiner Meinung nach sehr selten gegeben – systematisches Vorgehen ist also eine Zeit- und somit Geldfrage.

Susanne von Jan führt an, dass Evaluatorinnen und Evaluatoren aus ihrer Erfahrung häufig die in Berichten aufgeführten Schlussfolgerungen auch auf Nachfrage hin nicht mit Daten belegen oder begründen könnten und auf ‚logische‘ Plausibilitätsüberlegungen verweisen. Plausibilitätsüberlegungen sind ihrer Meinung nach allerdings nicht immer angemessen – hier kommt es auf die Fragestellung an. Wenn Wirkungen im Vordergrund stehen und in der Ausschreibung ein quasi-experimentelles Design erwartet wird, dann ist eine Plausibilitätsüberlegung im Sinne von wie ‚realistisch‘ eine Wirkung ist, nicht ausreichend. Auch ihrer Meinung nach sind die finanziellen Mittel das große Problem: Gutachterinnen und Gutachter sollen mit einer kurzen Feldphase alle DAC-Kriterien beantworten – Geld für elaborierte qualitative Designs, wie z.B. eine Kontributionsanalyse oder auch elaborierte quantitative Designs, ist nicht vorhanden. Im Ergebnis greifen Gutachterinnen und Gutachter auf Plausibilitätsüberlegungen, auf ‚logische Überlegungen‘ zurück. Zwar werden bei ‚echten‘ Wirkungsanalysen durchaus mehr Tage, mehr Zeit und mehr Geld bereitgestellt, aber im Normalfall – auch unter Berücksichtigung der zur Verfügung gestellten Arbeitstage – „kann ich mich gar nicht *nicht* damit [Plausibilitätsüberlegungen]

zufrieden geben...“. Auftraggebende sollten daher ihre Anforderungen reduzieren und realistischer gestalten bzgl. dem, was von Auftragnehmerinnen und Auftragnehmern geleistet werden kann. Andererseits erwartet sie von diesen aber Offenheit. In Berichten muss offen dargelegt werden, was tatsächlich ableitbar ist – die Formulierung „das Projekt hat dazu beigetragen, dass...“ ist für *Susanne von Jan* nicht ausreichend, wenn dies ausschließlich auf Plausibilitätsüberlegungen basiert. Dann erwartet sie eine entsprechende Formulierung, im Sinne von „aus den und den Gründen ist es plausibel, dass das Projekt da und da hinführt“. *Dorothee Mack* unterstreicht diesen Hinweis – für sie sind Plausibilitäten im Bereich Wirkungen durchaus akzeptabel, aber „es muss einfach gut nachvollziehbar und belegt sein“. *Susanne von Jan* schlägt als Ausweg aus dem Dilemma Joint Evaluations oder Clusterevaluationen vor: Mehr Ressourcen ermöglichen die methodisch saubere Beantwortung bestimmter Fragestellungen.

Für den Bereich der Kultur teilte *Vera Hennefeld* mit, dass Geld weniger das Problem ist – hier fehlt es eher an Zeit. Verzögerungen in der Planung und Ausschreibung sowie Vergabe von Evaluationen und dann schnell zu treffende Entscheidungen, z.B. über die Weiterführung eines Projektes, gehen oft zu Lasten der Evaluationszeit bzw. ‚sauberer‘ Methodenanwendung. Der Zwang, zeitnah ein Ergebnis vorzulegen, führt dann häufig zu dem Umstand, dass auf Plausibilitätsüberlegungen zurückgegriffen werden muss. Es gehört allerdings zum guten Standard, dies genau in dieser Form offen zu legen – und das hat aus ihrer Sicht nichts mit dem Budget zu tun. D.h. im Bericht ist aufzuführen, dass z.B. Outcomes nicht gemessen werden konnten, aber die Tatsache, dass anhand erbrachter Outputs zumindest die Voraussetzungen gegeben sind, dass Outcomes überhaupt erwartet werden können.

Im Rahmen der Diskussion um Methodenqualität, ‚sauberes‘ Arbeiten und Transparenz kam die Frage auf, inwieweit Evaluationsberichte ein **Methodenkapi-
tel** als Anhang enthalten sollten und inwieweit im Text **Quellenverweise** angeführt werden sollten – sowohl im Sinne von Verweisen auf Literatur und Dokumente aber auch im Sinne von Verweisen auf Datenquellen, wie Interviews, standardisierte Befragungen. *Dorothee Mack* verweist darauf, dass dies wiederum Arbeitstage benötigt. Gutachterinnen und Gutachter würden ihr sagen, dass sie Tabellen und Interviewtranskripte vorliegen hätten – bei Rückfragen könnte man darauf zurückgreifen. Aber in den Berichten erwarte sie das nicht – auch wolle das in der Politik ja keiner lesen. *Sonja Sheikh* bestätigt dies aus Auftragnehmerinnensicht: Literatur und Datenauswertungen sind in großem Umfang vorhanden. Natürlich enthalten Evaluationsberichte ein Literaturverzeichnis, aber es werden nicht an jeder Stelle einzelne Verweise eingefügt, das sei zu aufwändig. Stichprobenumfänge etc. würden natürlich im Bericht aufgeführt aber nicht detaillierte Zahlen zu Standardabweichungen oder t-Test-Ergebnissen – „da wird uns schon vertraut“. *Alexandra Caspari* argumentierte, dass durch diese Verweise die Nachvollziehbarkeit erhöht wird: Wenn hinter einer Aussage in Klammern aufgeführt wird, aus welchen Quellen diese Aussage stammt, z.B. Dokumente, Literatur, einzelnen Intensivinterviews sowie standardisierte Befragung, dann ist erkennbar, dass hier Aussagen trianguliert wurden und aus mehreren Quellen stammen – das genau erhöht doch die geforderte Nachvollziehbarkeit und die Plausibilität. *Susanne von Jan* betonte, dass sie als Auftraggebe-

rin sehr wohl Quellenverweise auch zu Literatur erwartet. Zudem begrüßt sie Angaben, die einzelne Aussagen konkreten Gruppen zuordnen, z.B. ZG-Vertretern oder Projektpartnern, so lange die Anonymität der betreffenden Personen gewährleistet bleibt.

Tülin Engin-Stock (uzbonn GmbH und Sprecherin des AK Kultur und Kulturpolitik) warf ein, dass die Diskussion den Eindruck erwecke, seitens der Auftraggebenden gebe es den Konsens, dass in Evaluationsberichten alles plausibel dargestellt werden müsse. Dies deckt sich allerdings nicht mit ihren Erfahrungen als Evaluatorin. Insbesondere wenn sie wissenschaftlich arbeitet und alles nach „bestem Wissen und Gewissen“ niederschreibt, sind die Rückmeldungen eher negativ – von „das kann kein Mensch lesen“ über Kritik am Umfang bis hin zu „das kann ich für meine Öffentlichkeitsarbeit nicht nutzen“. Sie weist darauf hin, dass meist eigentlich zwei Aufbereitungen der Ergebnisse nötig sind – eine interne Aufbereitung, die dem hohen Anspruch nach Legitimität und Transparenz gerecht wird, und eine weitere Aufbereitung der Ergebnisse für den eigentlichen Zweck der Evaluation, die Legitimation, die für die Öffentlichkeit und Verbreitung gedacht ist. Diese doppelte Anforderung macht auch doppelt Arbeit, ist aber in den Evaluationsanforderungen meist nur implizit enthalten. *Susanne von Jan* bestätigt dies teilweise: Es kommt sicherlich auch auf die Person auf Seiten der Auftraggebenden an. Sie selbst wolle diese Informationen gerne dargestellt haben, beispielsweise auch im Anhang. Aber natürlich gibt es verschiedenen Zielgruppen des Berichtes – die Geschäftsführung einer größeren Institution hat sicherlich weniger Interesse an diesen detaillierten Informationen, hier wird möglicherweise eher die Kurzzusammenfassung gelesen.

Jörg Faust (Leiter DEval) brachte einen neuen Aspekt in die Runde: Woher erhalten Organisationen **Informationen über existierende Evaluationsmethoden**, d.h. wie funktioniert „Diffusion von Methoden“, was sind die Treiber für **Imitationen** – in Abgrenzung zu ‚Innovationen‘ im Sinne von originärer Methodenforschung. Kommen Imitationen aus anderen Politikfeldern, aus dem internationalen Umfeld und/oder über direkte Peers, also Konkurrenten? Das DEval – so *Jörg Faust* – hat hier eine Sonderrolle inne, da es mit umfangreichen Personal- und Finanzmitteln unterschiedlichste Methodenexpertise zusammentragen kann. Hierdurch kann bei Evaluationen baukastenmäßig auf die unterschiedlichen Expertisen zurückgegriffen werden. In der EZ herrscht darüber hinaus eine große Diversifizierung, die Imitationen kommen somit aus unterschiedlichen Politikfeldern (z.B. Gesundheitsökonomie, Geografie), wobei man sich an internationalen Standards orientiert.

Sonja Sheikh berichtet – auch mit Blick auf ‚sauberes‘ Arbeiten – dass insbesondere Peers eine große Rolle spielen: Im Bereich FTI-Politik besteht zwischen den Evaluatorinnen und Evaluatoren ein großes Wettbewerbsverhältnis. Gerade in einem kleinen geografischen Bereich wie Österreich ist die Anzahl der Auftraggebenden und auch der Auftragnehmenden „überschaubar“, die Wettbewerber kennen sich untereinander. Standards und Qualität müssen einfach eingehalten werden, weil die Community die Qualität kontrolliert. D.h. man *muss* gut arbeiten, denn man hat einen Ruf zu verlieren. Das erzeugt einen viel größeren Druck als von Seiten der Auftraggebenden. Bei Ausschreibungen verzichten Auftraggebende teils gar auf eine detaillierte Methodendarstellung in den Angeboten, da sie wissen, dass methodisch

elaboriert gearbeitet werden kann und wird. Die indirekte Qualitätskontrolle über den Ruf von Evaluierungsanbietern ist im FTI-Bereich möglich, da Evaluationsberichte so gut wie immer publiziert werden. *Jörg Faust* wies darauf hin, dass in der EZ die Transparenz von Evaluationen im Sinne von Veröffentlichungspraxis eher heterogen ist. Hier wird nun allerdings verstärkt über Transparenzinitiativen und Transparenzindizes, die abbilden, inwieweit Organisationen ihre Evaluationsberichte vollumfänglich veröffentlichen, Wettbewerbsdruck zur Standardverbesserung ausgeübt. In der EZ ist der Markt international – somit ist hier die internationale Community die Referenzgröße.

Dorothee Mack berichtet, dass Imitation bei Misereor ebenfalls aus dem internationalen Umfeld kommt – und zwar über die Gutachterinnen und Gutachter, die oft in internationalen Teams arbeiten und somit Methoden, die in anderen Ländern angewendet werden, kennen lernen. Machen sie mit einer neuen Methode gute Erfahrung, bringen sie diese bei anderen Evaluationen zur Anwendung. Ein anderer Einflussbereich sind – so *Dorothee Mack* – Peers und nicht zuletzt die DeGEval und der AK Entwicklungspolitik.

Was ‚Imitation‘ betrifft, kocht man – so *Sonja Sheikh* – im Bereich FTI noch ‚im eigenen Süppchen‘. Da alle Evaluationsberichte auf der Plattform fteval² eingestellt sind, kann geschaut werden, was die Konkurrenz macht – aber diese kommen ebenfalls aus dem Bereich FTI. Politikfeldübergreifende Ideen finden sich eher in solchen Diskussionen im Rahmen der DeGEval. Hier kommen Ideen, die meist aufgrund fehlender Zeit trotz gutem Vorsatz nicht nochmals vertiefend betrachtet werden. Aber hier ist ‚mehr Potenzial als wir nutzen‘. *Susanne von Jan* bestätigt dies: Konferenzen wie die DeGEval-Tagung geben Hinweise zu bisher unbekanntem Methoden, die man näher anschauen will, was dann leider nicht immer passiert. Als weitere Quelle für Imitation nennt sie auch wissenschaftliche Artikel.

Im Politikfeld der Kulturevaluation dagegen – so *Vera Hennefeld* – ist Transparenz und somit die Möglichkeit der Kontrolle oder auch des Lernens im Sinne von Imitation annähernd nicht gegeben. So gibt es etwa keine Metaevaluationen und auch Evaluationsberichte werden so gut wie nie publiziert und sind damit nicht zugänglich. Somit ist weder Kontrolle noch Imitation über Peers möglich. Zwar wünscht sich z.B. das Goethe-Institut in seinen Evaluationsausschreibungen explizit ‚innovative Ansätze‘, aber es ist schwierig diese zu finden. Mangels Veröffentlichungen bleiben hier nur der Blick in Methodenbücher oder benachbarte Felder, also auch in Richtung Ökonomie oder auch Betriebswirtschaftslehre und Soziologie. Sinn machen hier theoretische Strömungen, aus denen methodisch etwas abgeleitet werden kann.

Alexandra Dolff (Deutsche Welle) berichtet von dem ‚Leidensdruck‘, dass im Bereich Kulturpolitik häufig Dinge erforscht werden wollen, für die es noch keine Methoden gibt – und somit nur ‚Innovation‘ möglich ist. Hier werden dann gemeinsam mit Peers Methodenprototypen entwickelt, z.B. eine Forschungs-App, um Videos im internationalen Standard vergleichbar bewerten zu können. Die Veröffentlichungspraxis in der Kulturpolitik empfindet sie auch als ‚Katastrophe‘. Allerdings

2 www.fteval.at

verweist sie darauf, dass gemäß Deutsche-Welle-Gesetz veröffentlicht wird: „Wir veröffentlichen alle vier Jahre – dann wenn wir müssen, dürfen wir auch“. Das Wissen ist aufgrund der vielen Studien, die im Forschungsverbund mit anderen Auslandssendern wie der BBC durchgeführt werden, eigentlich sehr umfangreich. Der Forschungsverbund hat aber strenge Veröffentlichungsregeln, daher ist das, was veröffentlicht werden darf, letztendlich so marginal, dass eine Veröffentlichung dann keinen Sinn mehr machen würde.

Ausblick

In der Kürze der Zeit konnten viele Fragen angesprochen, jedoch nicht abschließend diskutiert werden. Kennzeichnend für diese erste Analysesession war jedoch die Feststellung, dass in allen Politikfeldern der Balanceakt zwischen Methodenqualität und Methodentransparenz einerseits und zur Verfügung gestellter Zeit und finanzielle Mittel andererseits gleichermaßen zu bewältigen ist. Auch hat sich gezeigt, dass Evaluationsmethoden vornehmlich aus dem eigenen Politikfeld entnommen werden. Informationen über in anderen Politikfeldern angewandte Methoden werden eher selten eingeholt. Der Arbeitskreis Methoden bietet hier Unterstützung an: Im Schlusswort wies *Alexandra Caspari* darauf hin, dass sich der AK Methoden über Hinweise auf noch eher unbekannt in einem Politikfeld erfolgreich implementierte Methoden, die für andere Politikfelder innovativ sein könnten, sehr freut. Der AK Methoden kann dann evtl. im Rahmen einer Frühjahrstagung oder Session auf der DeGEval-Jahrestagung Arbeitskreise zusammenbringen oder spannende Methoden vertiefen. In einem Jahr wird der AK Methoden – so ist geplant – mit gleichem Format drei weitere Arbeitskreise näher betrachten.