

Ergebnisdokumentation der Session D4 „Datenaggregation von der Projekt- auf die Programmebene – Herausforderungen und Lösungsansätze im Monitoring und bei Evaluationen“ auf der DeGEval-Jahrestagung 2023 in Magdeburg

Arbeitskreis
Umwelt



Auf der 26. DeGEval Jahrestagung in Magdeburg zum Thema „Valide Daten - rationale Entscheidungen - akzeptierte Steuerung?“ richtete der AK Umwelt am Freitag, dem 15.09.2023 die Session D4 „Datenaggregation von der Projekt- auf die Programmebene – Herausforderungen und Lösungsansätze im Monitoring und bei Evaluationen“ aus.

Kern der Session waren zwei Fachimpulse, im Anschluss wurde in offener Runde diskutiert.

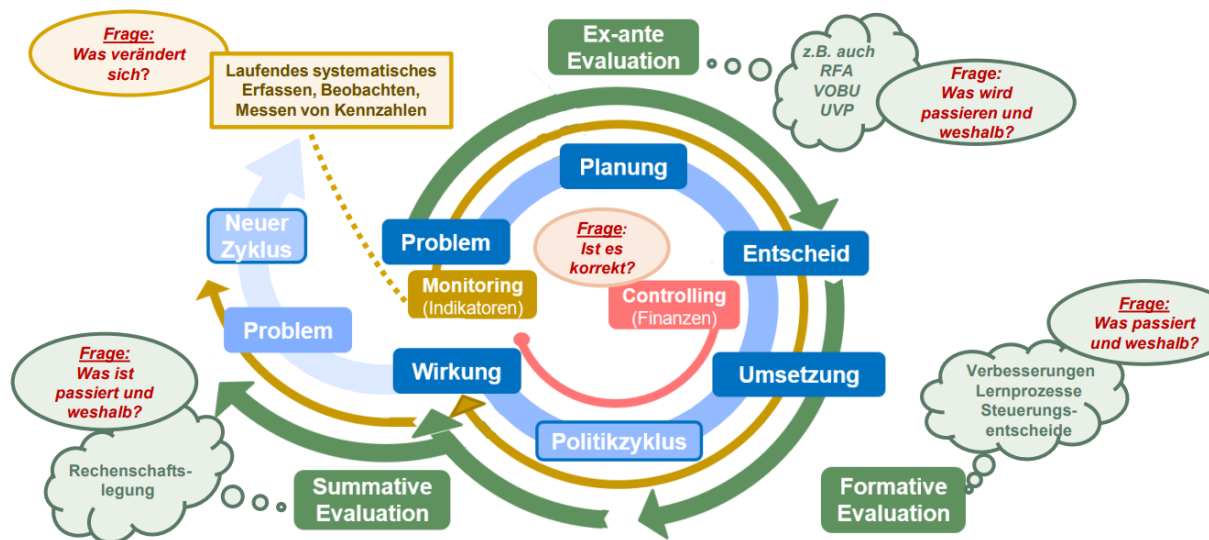
In seinem Impulsvortrag „[Evaluation komplexer Programme mit unterschiedlich gelagerten Teilprojekten](#)“ verweist Reinhard Zweidler auf zwei Beispiele, den Nationalen Maßnahmenplan der Schweiz zur Verringerung der Lärmbelastung und die Verpflichtungen der Schweiz aus dem Kunming-Montreal Protokoll. In beiden Fällen geht es um viele, teils sehr unterschiedlich gelagerte Teilmaßnahmen und dies ist für die Evaluation eine Herausforderung. Monitoringsysteme sollen dazu beitragen, aussagekräftige Daten zu Entwicklungen über längere Zeiträume zu liefern, die auch für die Evaluation verwendet werden können. In der Praxis ist die Güte der Daten allerdings sehr unterschiedlich. Das gilt insbesondere dort, wo es keine einheitliche oder allgemein akzeptierte Indikatorik gibt.

Am Beispiel des Maßnahmenplans Lärm kann das verdeutlicht werden: Die Vielfalt unterschiedlicher Lärmquellen, Wirkungsweisen und Maßnahmentypen im Programm ergibt eine sehr heterogene Datenlage. Im konkreten Fall wurde zunächst eine Machbarkeitsstudie zur Untersuchung der Evaluierbarkeit des Programms vorgeschaltet, ein Ansatz, der auch bei anderen komplexen Programmen erfolgversprechend sein kann. Um die praktische Umsetzung der Evaluation und die Übersichtlichkeit zu verbessern, wurden Maßnahmen zu Bündeln zusammengefasst, für die jeweils gemeinsam eine Wirkungsmodellierung und ein Evaluationskonzept erarbeitet werden konnte.

Der Globale Biodiversitätsrahmen von Kunming-Montreal umfasst vier zentrale Ziele und 23 operative Unterziele, zu deren Erreichung sich die Vertragsstaaten bis 2050 verpflichtet haben. Untersucht wurde, welche Maßnahmen auf nationaler Ebene getroffen werden müssen, um den Staatsvertrag zu erfüllen. Die Bewertung der nationalen Umsetzung ist nicht trivial. Das Kunming-Montreal-Protokoll hat im Herbst 2023 erst für etwa die Hälfte aller Ziele Indikatoren definiert, die restlichen Indikatoren sollten bis zur nächsten COP festgelegt sein. Die nationale Indikatorik, die in bestehenden Monitoringprogrammen eingesetzt wird, kann das Fehlen staatsvertraglicher Indikatoren nur teilweise ersetzen, da die gewonnenen Daten von den nationalen Indikatorsystemen nicht immer einfach auf die internationalen übertragen werden können. Dabei kennt die Schweiz zwei Indikatorsysteme, die MONET-Indikatoren und die BAFU-Indikatoren und beide passen nicht vollständig zur Erfassung der Kunming-Montreal-Ziele. Auch wenn voraussichtlich im Laufe der Zeit für alle Ziele aussagekräftige Indikatoren zur Verfügung stehen, fehlen wenigstens über eine gewisse Zeit zuverlässige Anfangsmessungen, was die Aussagekraft über Fortschritte beeinträchtigt.

Generell kann festgehalten werden: Bei der Formulierung von Programmen sollte ihre Evaluierbarkeit schon mitgedacht und es sollten im Hinblick auf die Evaluation auch passende

Monitoringsysteme entworfen werden. Evaluation und Monitoring stellen dabei jeweils unterschiedliche Fragen: Monitoring fokussiert die beobachtbaren Veränderungen im Laufe der Zeit, Evaluation versucht darüber hinaus immer zu klären, warum Resultate eingetreten sind oder nicht. Dies erklärt, dass die Evaluation nebst den Monitoringdaten auch auf eigene Datenerhebungen angewiesen ist.



Aus den Zielen des Programms und aus den Zwecken der Evaluation leiten sich die Fragestellungen und die Bewertungskriterien ab. Kriterien sollen oft Zustände beschreiben, die schlecht beobachtbar sind, daher müssen sie durch Indikatoren operationalisiert werden. Bei deren Auswahl birgt die Fokussierung auf oft eingeforderte und als aussagekräftig scheinende quantitative Indikatoren das Risiko der Verengung der Perspektive und teilweise gar der „kreativen Buchhaltung“. Manchmal sind auch Daten, die aus bestehenden Indikatorsystemen des Monitorings gewonnen werden können, für die Programmevaluation wenig aussagekräftig. Schwierig wird es, wenn Aussagen in Bereichen gemacht werden sollen, in denen noch keine oder keine akzeptierte Indikatorik besteht.

Zu bedenken ist auch folgendes: Evaluation ist Auftragsforschung und das bedeutet, dass die Auftraggebenden Richtung und Schwerpunkte der Arbeit vorgeben. Gerade in komplexen Situationen kann dies in ungünstigen Fällen dazu führen, dass sich die Evaluation auf Fragen konzentriert, die für eine gesamtheitliche Bewertung des untersuchten Gegenstandes nicht die wirklich entscheidenden sind. Eine Relevanzanalyse kann helfen, die wesentlichen Phänomene im Auge zu behalten.

Im zweiten Impulsvortrag „[Entwicklung eines Monitoring-Systems für ein komplexes Förderprogramm](#)“ berichtete Jonas Rasch vom Projektträger Zukunft – Umwelt – Gesellschaft (ZUG) vom BMUV-Förderprogramm gegen Meeressmüll. Das Förderprogramm ist typisch für Programme im Umweltbereich mit einer hohen Komplexität der Problemlage(n), der Projekte und der Kontexte sowie vielen Akteuren. Es gibt eine hohe Anzahl an Einflussfaktoren, die miteinander vernetzt sind und sich gegenseitig beeinflussen. Beobachtete Systeme sind dynamisch und häufig intransparent. Das bedeutet, dass sowohl Durchführungsorganisationen als auch Evaluierende mit Unkenntnis umgehen müssen.

In diesem Kontext dient Monitoring als ein Instrument der begleitenden Erfolgskontrolle. Folgende Überlegungen zum Monitoring sind für die Arbeit im BMUV-Förderprogramm gegen Meeresmüll zentral:

- **Leistungs- oder Wirkungsmonitoring?** Welche Ebenen eines Wirkungsmodells sollten betrachtet werden?
Outputs sind methodisch und organisatorisch in der Regel leichter messbar als Outcomes. Die Erfassung von Outputs ist aber nicht ausreichend, um Outcomes vorherzusagen, folglich müssen auch Outcomes erfasst werden. Umgekehrt ist ein reines Outcome-Monitoring ebenfalls nicht ausreichend, da auch der konkrete Beitrag der Projekte zu den Outcomes erfasst werden muss. In der Konsequenz müssen sowohl Outcomes als auch die jeweiligen Outputs, die zu deren Erreichen beigetragen haben, durch das Monitoring erfasst werden.
- **Attribution und Kontribution:** Rechenschaftslegung fokussiert meist auf Attribution, das heißt, Rechenschaftslegung verlangt Erfassung der Wirkungen, die von den Projekten ausgelöst wurden. Einfach attribuiert werden können im Rahmen des Monitorings aber nur Wirkungen, an deren Entstehen Projekte direkt beteiligt waren (direkter Zusammenhang Output - Outcome) und die noch während der Umsetzung zu beobachten sind (kurzfristige Wirkungen). Zudem kann der Fokus auf Attribution zu einer Zielverschiebung führen: von langfristigen, durch lokale Akteure getragene Veränderungen hin zu kurzfristigen, von Projekten direkt ausgelösten Veränderungen. Folglich sollte das Monitoring sich nicht allein auf Attribution fokussieren, sondern auch Ziele einbeziehen, an deren Erreichen die Projekte nur mittelbar beteiligt sind.
- **Mess-Fixierung:** Verstärkt wird der Gefahr der Zielverschiebung außerdem durch Messfixierung: Indikatoren, die von Ministerium vorgegeben werden, werden von Projektverantwortlichen als Ziele interpretiert. Diese „Messziele“ können die tatsächlichen Ziele verdrängen und so negative Auswirkungen auf den Projekterfolg haben. Deshalb ist es wichtig, nicht nur zentrale Kennzahlen abfragen, sondern auch qualitative Aspekte wie Nachhaltigkeit, lokale Verankerung und Kontribution der Projekte.
- **Quantitative Indikatoren oder qualitative Deskriptoren?** Durch Vorgaben der Finanzkontrolle (BHO / BMF / BRH) konzentriert sich Monitoring vor allem auf quantitative Indikatoren. Allerdings fehlt teilweise Vorwissen, um Messgegenstände und Erhebungsmethoden für quantitative Indikatoren präzise definieren zu können. Zudem sind in der Regel Kontextinformationen für eine angemessene Interpretation einzelner Datenpunkte notwendig. Im Umweltbereich sind Kontexte häufig komplex und können allein über quantitative Indikatoren nicht vollständig erfasst werden, da mit unbekanntem Einflussfaktoren und Zusammenhängen gerechnet werden muss. Ex-ante definierte Indikatoren können folglich diese Komplexität nicht umfänglich abbilden. Unter solchen Umständen setzt ein gutes Programmmonitoring auf eine Kombination aus quantitativen Kennzahlen und offenen qualitativen Fragen zur Erfassung von projektindividuellen Kontextinformationen und unbekanntem Einflussfaktoren. Dabei können die Grenzen zwischen qualitativem Monitoring und Evaluation fließend sein.
- **Überlegungen zur Konzeption des Monitorings:** Bei der Konzeption werden Wirkmodelle auf Projekt- und Programmebene benötigt. Die Identifizierung zentraler Möglichkeiten der Zielerreichung hilft, das Monitoring in einem handhabbaren Rahmen zu halten. Eine hohe Anzahl an Indikatoren und / oder Deskriptoren ist für die Steuerung nützlich, sollte jedoch möglichst so gestaltet sein, dass der Aufwand der Datenerhebung in Grenzen hält (möglichst

wenig zusätzliche Datenerhebung, Abfrage impliziten Wissens). Wichtigste Arbeitsschritte sind also die Identifizierung zentraler strategischer Ansätze und die Entwicklung von messbaren Indikatoren / Deskriptoren-Sets für jeden strategischen Ansatz.

- **Konkrete Umsetzung im Praxisbeispiel Meeresmüll:** Im Programm gibt es 8 strategische Ansätze und je Ansatz 4-10 Indikatoren. Insgesamt sind es 58 Indikatoren / Deskriptoren, davon 15 quantitative Indikatoren und 43 qualitative Deskriptoren. Die Erhebungsmethodik wurde bei quantitativen Indikatoren (noch) nicht vorgegeben, weil Vorwissen fehlt. Aktuell läuft die Sammlung von Erfahrungen aus den Projekten, später folgt ggf. eine Vereinheitlichung. Projekte wählen bei Projektstart relevante Indikatoren-Sets aus und berichten einmal jährlich. Qualitative Indikatoren waren als Fragen formuliert.

Das Fallbeispiel zeigt: Programm-Monitoring ist kein neutrales Untersuchungsinstrument. Allein schon die Vorgabe von Indikatoren durch die Fördermittelgebenden wirkt steuernd auf die geförderten Projekte. Bei jedem Indikator / Deskriptor muss überlegt werden, ob er sich auch als Ziel eignet und evtl. zu Zielverschiebungen führt. Nicht nur Quantität, sondern auch Qualität muss erfasst werden. Bei Programmen mit relativ kurzen Wirkungsketten und gut messbaren Zielen in vorhersehbaren Kontexten ist schlankes, quantitatives Monitoring ausreichend („accounting-based accountability“), komplexe Programme in komplexen Kontexten benötigen einen Ansatz, der Unerwartetes / Unvorhersehbares erfassen kann und Flexibilität zulässt (qualitative Ansätze, „account-based accountability“). Qualitatives Monitoring nimmt implizites und lokales Wissen ernst, setzt Vertrauen voraus, kann dieses aber auch aufbauen.

In der anschließenden Diskussion wurden weiterführende Fragen diskutiert wie etwa:

- *Wie kann man damit umgehen, dass Indikatoren als Ziele verstanden werden?*
Theoretisch sollten alle für den Projekterfolg wichtigen Aspekte durch Indikatoren abgedeckt werden. Dies führt jedoch zu einer sehr umfangreichen Indikatorik. Praktisch sollten neben zentralen Kennzahlen zumindest die auch die wichtigsten qualitativen Aspekte mit Fokus auf die Programmziele abgedeckt werden. Notwendig ist auch eine offene Kommunikation mit den Durchführungsorganisationen, um auf das Risiko hinzuweisen, dass Indikatoren in einzelnen Projekten wie Anreizsysteme genutzt werden könnten. Es obliegt der Programmsteuerung, dafür zu sorgen, dass in den einzelnen Projekten eine angemessene Ausrichtung auf die übergeordneten Programmziele erfolgt.
- *Wie kann das Monitoring für die Projektsteuerung genutzt werden?*
Das im zweiten Impuls vorgestellte Programm-Monitoring dient der Programmsteuerung. Auf Projektebene findet ein eigenes Monitoring statt, das für die Steuerung der einzelnen Projekte genutzt wird. Programm- und Projektmonitoring sind sinnvoll aufeinander abzustimmen.
- *Wie aggregiert man qualitative Daten?*
Eine solche Aggregation ist nicht immer sinnvoll, aber in Deutschland aufgrund BHO-Vorgaben notwendig. Auch qualitative Informationen können quantifiziert werden. Zudem sind individuelle Fließtexte zum Kontext möglich.
- *Aggregation: Wie kommt man von den Einzelindikatoren zu einer Gesamtauswertung?*
Die Auswertung der Erhebung Indikatoren erfordert das gesamte Wissen der im Evaluationsteam vertretenen Fachwissenschaften, wobei mit fortschreitender Datenauswertung der Anteil an Interpretation steigt. Dabei müssen die Indikatoren bzw. die dadurch abgebildeten Kriterien gewichtet und unterschiedliche Sichtweisen auf den untersuchten Gegenstand berücksichtigt werden. In der Regel ist es hilfreich, dabei auf Basis des Wirkmodells des Programms vorzugehen.