

PRIF REPORT

PEACE RESEARCH INSTITUTE FRANKFURT / LEIBNIZ-INSTITUT HESSISCHE STIFTUNG FRIEDENS- UND KONFLIKTFORSCHUNG

MONA KLÖCKNER // SVETLA KOYNOVA // JOHANNA
LIEBICH // LISA NEEF

**ERFAHRUNGEN AUS DER EVALUATIONSPLANUNG
EINES AUSSTEIGERPROGRAMMS**
VORAUSSETZUNGEN FÜR WIRKSAMKEITSERFASSUNG
IN DER TERTIÄREN EXTREMISMUSPRÄVENTION

REPORT-REIHE

PrEval


PRIF Report 6/2021

**ERFAHRUNGEN AUS DER EVALUATIONS-
PLANUNG EINES AUSSTEIGERPROGRAMMS
VORAUSSETZUNGEN FÜR WIRKSAMKEITS-
ERFASSUNG IN DER TERTIÄREN
EXTREMISMUSPRÄVENTION**

MONA KLÖCKNER // SVETLA KOYNOVA //
JOHANNA LIEBICH // LISA NEEF

LEIBNIZ-INSTITUT HESSISCHE STIFTUNG FRIEDENS- UND KONFLIKTFORSCHUNG (HSFK)
PEACE RESEARCH INSTITUTE FRANKFURT (PRIF)

Coverbild:

Lizenzinformationen: asmuSe: <https://pixabay.com/photos/building-architecture-1989816/> (Bearbeitung: Anja Feix)

Textlizenz:

Creative Commons CC-BY-ND 4.0 (Namensnennung/Keine Bearbeitungen/4.0 International).
Das Coverbild unterliegt eigenen Lizenzbedingungen.



Adresse:

Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK)
Baseler Straße 27–31
60329 Frankfurt am Main
Telefon: +49 69 95 91 04-0
E-Mail: preval@hsfk.de
<https://www.hsfk.de>

DOI: 10.48809/prifrep2106

ISBN: 978-3-946459-67-5

DIE AUTOR:INNEN

Mona Klöckner ist Psychologin und wissenschaftliche Mitarbeiterin an der HSK. Ihre aktuellen Arbeitsschwerpunkte liegen in Evaluation und wissenschaftlicher Methodik sowie in der Erforschung des gesellschaftlichen Umgangs mit Extremismen und Verschwörungsideologien in Deutschland.

Svetla Koynova ist Politikwissenschaftlerin und wissenschaftliche Mitarbeiterin bei *Violence Prevention Network (VPN)*. Ihre Forschungsschwerpunkte umfassen Monitoring-, Evaluations- und Lernprozesse, Gruppenbezogene Menschenfeindlichkeit und Radikalisierung.

Johanna Liebich ist Koordinatorin und wissenschaftliche Mitarbeiterin bei *Violence Prevention Network (VPN)*. Ihre Arbeitsschwerpunkte sind Koordination, Monitoring und Evaluation von Projekten der Extremismusprävention, vornehmlich im Phänomenbereich Rechtsextremismus.

Lisa Neef ist Referentin im Ministerium des Innern des Landes Nordrhein-Westfalen in der Abteilung Verfassungsschutz. Ihre Arbeitsschwerpunkte im Referat Prävention und Aussteigerprogramme liegen in den Bereichen Evaluation von Programmen der Extremismusprävention, Qualifizierungsmaßnahmen und Zusammenarbeit mit der Wissenschaft.

HSFK-Reportreihe *PrEval*

Über die Evaluation und die wissenschaftliche Begleitung von Maßnahmen der Extremismusprävention und von Angeboten der politischen Bildung wird aktuell in Fachpraxis und Wissenschaft vielfältig diskutiert. Das liegt nicht zuletzt daran, dass über verschiedene Programme auf Ebene des Bundes, der Länder und der Kommunen seit einigen Jahren immer mehr Fördermittel für solche Maßnahmen bereit stehen, die Fachpraxis sich immer vielfältiger, innovativer und professioneller aufstellt und auch die Forschung zu den Phänomenfeldern, aber auch zu Evaluationsmethoden stetig Fortschritte macht. Folgerichtig fordert beispielsweise auch der Kabinettsausschuss zur Bekämpfung von Rechtsextremismus und Rassismus in seinem Maßnahmenpaket eine langfristige Stärkung der Qualität und Wirksamkeit von Präventionsmaßnahmen und knüpft damit an zahlreiche Initiativen von Bundes- und Landesregierungen, von Fachpraxis und Wissenschaft an.

Das Forschungs- und Transfervorhaben *PrEval* (Evaluationsdesigns für Präventionsmaßnahmen – Multimethodische Ansätze zur Wirkungsermittlung und Qualitätssicherung in der Extremismusprävention sowie den Schnittstellen zur Gewaltprävention und politischen Bildung) setzt hier an. Ziel des Projekts ist es, im Dialog aller an Evaluation in diesem Feld beteiligten Akteursgruppen den aktuellen Wissensstand zu Qualitätssicherung und Evaluation in der Extremismusprävention, der Gewaltprävention und der politischen Bildung zu erheben und gemeinsam Evaluationsdesigns, gerade an den Schnittstellen zwischen verschiedenen Präventionsbereichen, Zugängen, Phänomenfeldern und Akteuren, zu erarbeiten. Dabei befasst sich *PrEval* mit Maßnahmen aus der politischen Bildung, der primären, sekundären und tertiären Prävention sowie Schnittstellen zur Gewalt- und Kriminalprävention. *PrEval* nutzt Analyse-, Monitoring- und Mapping-Formate und entwickelt multimethodische Evaluationsdesigns im Rahmen von ausgewählten Pilotstudien. Ein zentraler Mehrwert des Vorhabens ist die Einbindung verschiedener Präventionsakteure – aus der Fachpraxis, den Sicherheitsbehörden, aus der Verwaltung und der Wissenschaft.

Die *PrEval*-Reportreihe fasst die zentralen Ergebnisse des Projekts zusammen. Dabei wird zum einen ein systematischer Überblick über die Evaluationsbedarfe der deutschen Trägerlandschaft in den Präventionsfeldern Islamistischer Extremismus wie Rechtsextremismus geboten, die bestehenden Kapazitäten für Evaluation und wissenschaftliche Begleitung von Präventionsmaßnahmen erhoben sowie Evaluationsstrukturen international vergleichend analysiert. Zum anderen werden multimethodische Evaluationsdesigns vorgestellt, die im Rahmen der *PrEval*-Pilotstudien für spezifische Präventionsmaßnahmen in enger Zusammenarbeit mit den Trägern (und teilweise auch Mittelgebern) dieser Maßnahmen entwickelt wurden. Möglichkeiten und Grenzen von wissenschaftlicher Begleitung werden diskutiert. Die zehn Beiträge der Reihe eint ein breites Verständnis von Evaluation, um den Komplexitäten des Felds und der darin verorteten Maßnahmen gerecht zu werden. Gleichwohl setzt jeder Report eigene, dem jeweiligen Thema angepasste Akzente in der Begriffsverwendung.

Das *PrEval*-Projekt wird vom Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK) koordiniert und von einem Verbund aus Forschungsinstituten und Trägern aus der Fachpraxis gestützt. Mehr Informationen zum Projekt: <https://preval.hsfk.de/>

Frankfurt/M. Dezember 2021

Übersicht über alle Titel der *PrEval*-Report-Reihe

Erfolgsbedingungen und Herausforderungen für die Zusammenarbeit von Fachpraxis und Wissenschaft bei der Evaluation von Angeboten politischer Bildung von Raphaela Schlicht-Schmälzle // Désirée Theis // Sina Tultschinetski // Melanie Verhovnik-Heinze // Christian Kautz // Maryam Kirchmann. PRIF Report 5/2021.

Erfahrungen aus der Evaluationsplanung eines Aussteigerprogramms. Voraussetzungen für Wirksamkeitserfassung in der tertiären Extremismusprävention von Mona Klöckner // Svetla Koynova // Johanna Liebich // Lisa Neef. PRIF Report 6/2021.

Weitere geplante Titel in dieser Reihe

Monitoring, Evaluation und Lernen: Erfahrungen und Bedarfe der Fachpraxis in der Prävention von Rechtsextremismus und Islamismus von Svetla Koynova // Alina Mönig // Matthias Quent // Vivienne Ohlenforst.

Evaluationskapazitäten im Bereich der Extremismusprävention und der politischen Bildung in Deutschland von Andreas Uhl // Benjamin Zeibig // Manuela Freiheit // Andreas Zick.

Multimethodische Evaluationsdesigns: eine Erhebung der sozialwissenschaftlichen Praxis von Mona Klöckner // Melanie Verhovnik-Heinze // Raphaela Schlicht-Schmälzle // Reiko Nakamura // Julius Strunk.

Webvideos und Livestream zur Aufklärung über Desinformation. Evaluationsansätze zu digitalen Formaten in der politischen Bildung von Désirée Theis // Melanie Verhovnik-Heinze // Stefan Kroll // Keshia Kierek.

Evaluation der politischen Bildung im Jugendstrafvollzug – Ansätze, Chancen, Herausforderungen von Sina Tultschinetski // Alisa Rieth // Désirée Theis.

Developmental Evaluation zur Begleitung von Entwicklungsprozessen in Projekten der Radikalisierungsprävention – Ein Anwendungsbeispiel von Olga Schmidt // Johanna Liebich // Mona Klöckner // Dennis Walkenhorst // Bacem Dziri.

Klientenzentrierte Evaluation in Multi-Agency Settings der Extremismusprävention – Möglichkeiten und Grenzen eines wirkungsorientierten Vorgehens von Susanne Johansson // Julian Junk // Johanna Liebich // Dennis Walkenhorst.

Evaluation und Qualitätssicherung in der Kooperation sicherheitsbehördlicher und zivilgesellschaftlicher Akteure in der sekundären und tertiären Extremismusprävention von Julian Junk // Svetla Koynova // Vivienne Ohlenforst // Maximilian Ruf // Lea Scheu.

Außerdem im Rahmen des *PrEval*-Projekts als GPPi-Study erschienen

Extremismusprävention evaluieren: Institutionelle Strukturen im internationalen Vergleich von Ase-na Baykal // Sarah Bressan // Julia Friedrich // Giulia Pasquali // Philipp Rotmann // Marie Wagner.

Dieser Report legt dar, wie ein wirkungsorientiertes Evaluationsdesign eines Aussteigerprogramms im Phänomenfeld islamistischer Extremismus entwickelt werden kann, welche Herausforderungen es bei Wirkungsmessungen in der tertiären Extremismusprävention gibt und wie wichtig es ist, die Weichen für die Lösung definitorischer, methodologischer und prozessbezogener Herausforderungen bereits bei der Evaluationsvorbereitung zu stellen. Der Report basiert auf einer engen Zusammenarbeit zwischen dem *Aussteigerprogramm Islamismus (API)*, angesiedelt beim Ministerium des Innern des Landes Nordrhein-Westfalen, und dem Forschungsverbund *PrEval*. Gemeinsam wurde die Ausschreibung einer externen wirkungsorientierten Evaluation des Programms vorbereitet. Die Lehren aus dem Prozess sind primär für Institutionen, Träger und Zuwendungsgeber, die Evaluationen in Auftrag geben, aber ebenfalls für Evaluierende sowie Evaluierte von Interesse, gibt es doch bislang keine systematische Reflexion über die Vorbereitung eines wirkungsorientierten Evaluationsdesigns für Aussteigerprogramme.

Das dialogische Vorgehen bei der Vorbereitung einer Evaluation ist zeitintensiv, aber ein zentraler Baustein eines gelungenen Evaluationsprozesses, um die Bedarfe und Erwartungen der beteiligten Akteure bei dem geplanten Evaluationsvorgehen abzubilden. Der Report stellt diesen Prozess in allen Schritten dar und leitet daraus Empfehlungen für alle an der Evaluation beteiligten Akteure ab. Im Zuge der Zusammenarbeit wurde zunächst das Wirkverständnis des *API* herausgearbeitet. In diesem Rahmen wurde eine kausale Beziehung zwischen Programmmaßnahmen und beobachtbaren Ergebnissen dann als hoch wahrscheinlich betrachtet, wenn diese von einer Vielzahl von Datenquellen gestützt wird. Zu drei zentralen Untersuchungsebenen – Wirkannahmen, Wirkindikatoren und Wirkanalyse – wurden im Anschluss mögliche Fragestellungen für die Evaluation gesammelt, die sich an der Chronologie von idealtypischen Beratungsverläufen orientierten.

Die ca. 80 erarbeiteten Fragestellungen wurden anschließend mit geeigneten Datenquellen und Erhebungsmethoden in eine Matrix überführt. Sie ermöglicht den zukünftigen Evaluierenden einen ersten Einblick in zentrale Erkenntnisinteressen der Auftraggebenden und zielt darauf, die Vorbereitung der Datengrundlage und Methodenanwendung zu orientieren, zu vereinfachen und zu beschleunigen. Darüber hinaus bietet sie eine Gesprächsgrundlage für die Evaluierenden und die Mitarbeitenden des *API*, um die eigenen Interessen und Bedarfe nach der intensiven Vorbereitung klarer sortieren und benennen zu können. Ebenso sind Ansprüche des *API* an die Übertragbarkeit und den Praxisnutzen der Evaluationsergebnisse bereits formuliert worden. In der Zusammenarbeit wurden wichtige prozessuale Erkenntnisse gewonnen, die Empfehlungen nach sich ziehen.

Für *Auftraggebende* von Evaluationen empfiehlt der Report:

- das Wirksamkeitsverständnis und entsprechende Erwartungen an Wirkung(en) vor der Gestaltung der Ausschreibung festzulegen, weil sich daraus methodologische Herangehensweisen ergeben, die bei der finanziellen und zeitlichen Ausstattung der Evaluation zu berücksichtigen sind;
- Schwerpunkte (Fragestellungen, thematische Programmschwerpunkte) der evaluatorischen Betrachtung im Vorfeld zu definieren und aufzulisten, um die Vorbereitung der Evaluation zu erleichtern;

- die Güteanalyse vorhandener Wirkindikatoren und ihre formative Weiterentwicklung als Gegenstand der Evaluation mitaufzunehmen, um das interne Monitoring und die Dokumentation der Programme und Projekte zu unterstützen;
- die nutzbare Datengrundlage vor der Ausschreibung zu diskutieren, um die Nutzung von Daten frühestmöglich anzubahnen. Dabei sind vorhandene Daten in Kategorien einzuteilen (etwa: offen zugänglich, blanko/anonymisiert, externe Quellen, deren Verfügbarkeit es zu klären gilt), um den hohen datenschutzrechtlichen Voraussetzungen gerecht zu werden;
- Bei Daten, die es durch die Evaluierenden zu erheben gilt, ist die Wahl passender Erhebungsmethoden für die Zielgruppen inklusiv mit Programmmitarbeitenden zu diskutieren, um sie an die Spezifität der Tertiärprävention und die Vielfalt der relevanten Akteure anzupassen, z.B. bezüglich ihrer Verfügbarkeit. Dabei stellt die Befragung von Klient:innen in diesem Feld eine besondere Herausforderung dar. Eine solche Befragung empfiehlt der Report ausdrücklich, da die Adressat:innen des Programms eine zentrale Informationsquelle für die (subjektive) Bewertung des Maßnahmen Erfolgs sind. Empfohlen wird hier die inklusive Vorbereitung mit und durch Ausstiegsbegleiter:innen, die Einbindung auf freiwilliger Basis möglichst vieler Klient:innen und, falls möglich, ihrer Angehörigen.

Für *Auftraggebende* von Evaluationen und *Evaluierende* empfiehlt der Report:

- den Transfer der Evaluationsergebnisse in die Praxis bereits bei der Konzipierung der Ausschreibung bzw. der Evaluation mitzudenken. Es wird empfohlen, vorab zu klären, welche Rollen externe Evaluator:innen und welche Rollen interne Evaluationsbegleiter:innen dabei spielen sollen. Die Übermittlung von Zwischen- und Endergebnissen und ihre Interpretation (ggf. durch Workshops und/oder Fortbildungen) sollte zeitlich und personell (und dadurch auch finanziell) angemessen ausgestattet sein, damit die Erkenntnisse der Evaluation erfolgreich in die Lern- und Wissensmanagementstruktur der Projekte und Programme eingehen können.

Für *evaluierte Projekte und Programme* empfiehlt der Report:

- eigene Evaluationserfahrungen und -bedarfe sowie praxisrelevante Erkenntnisinteressen im Vorfeld eines Evaluationsvorhabens intern zu diskutieren und auszuformulieren, um sich aktiv in den Prozess der Evaluationsgestaltung einbringen zu können;
- geeignete Qualitätssicherungsstrukturen zu eruieren und zu etablieren, um internes Evaluationswissen und interne Evaluationskapazitäten zu stärken und auf Ergebnisse sinnvoll reagieren zu können.

Für *Zuwendungsgeber* empfiehlt der Report:

- bei der Förderung zu berücksichtigen, dass die stetige Weiterentwicklung und Betreuung der Monitoring- und Dokumentationsstruktur und die Gestaltung, Begleitung und Abnahme von Evaluationen personell und finanziell gedeckt sein müssen;

Dafür sollte es den Trägern ermöglicht werden, flexible Evaluations- und Qualitätssicherungsformate zu implementieren (Evaluationsbeauftragte, AG Evaluation, Pauschale für Monitoring und Evaluation).

1. Einleitung	1
2. Das <i>Aussteigerprogramm Islamismus</i> des Landes NRW und die Zusammenarbeit mit <i>PrEval</i>	2
2.1 Das Aussteigerprogramm	2
2.2 Die Zusammenarbeit zwischen <i>API</i> und <i>PrEval</i>	4
3. Herausforderungen wirkungsorientierter Evaluationen begegnen	5
3.1 Grundlegende Erwartungen und Vorbehalte	6
3.2 Abbilden von Wirkung	7
3.3 Wahl der Indiktorik	11
3.4 Erkenntnisinteressen und Fragestellungen	12
3.5 Vorhandene Daten: Verfügbarkeit und Bearbeitung	15
3.6 Neu zu erhebende Daten: Analyse- und Erhebungsmethoden	18
3.7 Planung des Ergebnistransfers	24
4. Übertragbarkeit des Vorgehens auf andere Programme	25
5. Fazit und Empfehlungen	27
Literatur	30

1. EINLEITUNG

In den meisten Programmen und Projekten der Extremismusprävention wird die Überprüfung der Wirkung früher oder später Thema. Bisweilen spitzt sich die Wirkungsfrage auf ein einfaches „Wirkt Programm XY?“ zu. Vor dem Hintergrund der vielfältigen Präventionslandschaft und des öffentlichen Drucks im Themenbereich ist dieses Bedürfnis nach Komplexitätsreduktion verständlich. Die Umsetzung von Evaluationen, die zum Ziel haben, die Wirkung(en) von Präventionsprogrammen und -projekten zu erfassen (wirkungsorientierte Evaluationen), ist jedoch voraussetzungsreicher als es auf den ersten Blick scheint. Auf dem Weg vom Erkenntnisinteresse zum letztendlichen Evaluationsdesign muss stets der Individualität der Maßnahmen innerhalb eines Programms Rechnung getragen werden. Operativ müssen dabei zahlreiche Entscheidungen und Definitionen getroffen sowie methodische Herausforderungen gemeistert werden. Dieser Prozess ist geprägt von unterschiedlichen Erkenntnisinteressen aufseiten des Präventionsträgers, erschwerten Datenzugängen für Evaluierende und nicht zuletzt von dem Anspruch, dass die Ergebnisse tatsächlich nützlich für die Praxis sein sollen.

Ob und wie mit diesen Herausforderungen umgegangen werden kann, wird bereits in der Phase der Ausschreibung einer externen Evaluation für ein entsprechendes Programm oder ein Projekt mitbestimmt. Ist ein Auftrag ausgestaltet oder bereits vergeben, kann der Handlungsspielraum schon maßgeblich kleiner und das Zeitfenster verstrichen sein, um die Evaluation bedarfs- und praxisgerecht zu gestalten.

Es ist daher entscheidend, dass alle an einer Evaluation Beteiligten, und insbesondere Auftrag- und Zuwendungsgeber, ein Bewusstsein für die Herausforderungen von wirkungsorientierten Evaluationen haben. Der vorliegende Report zeichnet diese Herausforderungen anhand der Evaluationsvorbereitung des *Aussteigerprogramms Islamismus* des Landes Nordrhein-Westfalen (*API*) nach und illustriert Möglichkeiten, ihnen zu begegnen. Er gibt auf diesem Wege eine mögliche Antwort auf die Frage, wie wirkungsorientierte Evaluationen in der tertiären Extremismusprävention¹ angebahnt werden können und was dabei zu beachten ist. In der HSK-Reportreihe *PrEval*, deren Teil der vorliegende Report ist, deckt er beispielhaft die spezifische Sparte von Aussteigerprogrammen im islamistischen Extremismus ab. Obwohl er einen individuellen Ansatz für das *API* beschreibt, sind viele der Erfahrungen übertragbar auf Evaluationsvorhaben in vergleichbaren Programmen – und in anderen Phänomenfeldern des Extremismus.

Die Situation, in der dieser Report entstand, war in mehrerer Hinsicht besonders. Die Zusammenarbeit zwischen der zu evaluierenden Fachpraxis, vertreten durch das *API*, und der Wissenschaft, ver-

1 Unter Tertiärprävention im Kontext von Radikalisierungsprävention werden hier Maßnahmen verstanden, die sich an bereits radikalisierte Personen richten und das erneute oder fortlaufende Auftreten unerwünschter Entwicklungen in diesem Zusammenhang verhindern sollen (BAG RelEX 2021). Hierzu gehören unter anderem Ausstiegs- bzw. Deradikalisierungsprogramme mit unterschiedlicher methodischer Schwerpunktsetzung und in verschiedenen Phänomenfeldern. Die konkrete Definition tertiärpräventiver Maßnahmen sowie die Abgrenzung zu anderen Präventionsstufen wird durchaus kontrovers diskutiert (Johansson et al. 2022; Greuel 2020).

treten durch ein Team aus dem Forschungsprojekt *PrEval*, fand im Rahmen einer *PrEval*-Pilotstudie statt. Die Pilotstudien haben das explizite Ziel, die Bedarfe der Präventionsfachpraxis in Bezug auf Evaluation abzubilden. Dieses Vorhaben traf zeitlich mit einem besonderen Interesse des *API* zusammen, da dort in diesem Zeitraum eine Evaluation ausgeschrieben werden sollte. Gemeinsam hatten Mitarbeitende von *API* und *PrEval* daher die Möglichkeit, an einem konkreten Beispiel über die Potenziale von Evaluation für die Weiterentwicklung eines tertiären Präventionsprogramms zu reflektieren und die Ergebnisse der Diskussion in eine Ausschreibung einfließen zu lassen. Zum Zeitpunkt des Verfassens ist die Evaluation des *API* offiziell ausgeschrieben, der Evaluationsprozess jedoch noch nicht gestartet.

Der Report zeigt, wie im Rahmen der Zusammenarbeit zwischen *API* und *PrEval* Schritt für Schritt die notwendigen Rahmenbedingungen für eine erfolgreiche Evaluation geschaffen wurden. Während dieses Prozesses beschreibt er die Herausforderungen, die mit jedem Stadium einhergingen und die jeweiligen Lösungen, die im vorliegenden Fall gewählt wurden. Dabei geht er auch auf größere Debatten und Wissensbestände im Feld ein, die die Basis für das Vorgehen im vorliegenden Fall bildeten. Er schließt mit Empfehlungen für alle beteiligten Seiten einer Evaluation: Auftrag- und Zuwendungsgeber, Präventionsakteure (die selbst auch Ersteres sein können) und Evaluierende.

Kapitel 2 beginnt mit einer Vorstellung des *API* und einer kurzen Zusammenfassung der Zusammenarbeit zwischen dem *API* und *PrEval*. Kapitel 3 beschreibt die Herausforderungen von wirkungsorientierten Evaluationen und die im vorliegenden Fall beschlossenen Lösungsstrategien. Kapitel 4 diskutiert, ob und wie die hiesigen Erkenntnisse auf andere Präventionsakteure übertragbar sind, bevor Kapitel 5 abschließende Empfehlungen gibt.

2. DAS AUSSTEIGERPROGRAMM ISLAMISMUS DES LANDES NRW UND DIE ZUSAMMENARBEIT MIT PREVAL

2.1 DAS AUSSTEIGERPROGRAMM

Das seit 2014 bestehende *Aussteigerprogramm Islamismus* ist eines von drei Aussteigerprogrammen der Tertiärprävention des Ministeriums des Inneren des Landes Nordrhein-Westfalen, die beim Verfassungsschutz angesiedelt sind. Es richtet sich an Personen und deren nahe Angehörige, die in der islamistischen Ideologie und Szene fest verankert sind, teilweise Straftaten begangen haben und aussteigen wollen. Häufig wiesen Klient:innen vor der Teilnahme am Programm Szeneaktivitäten in gewaltaffinen Milieus auf, gingen aktiv Propagandatätigkeiten nach, vertraten öffentlich ihre ideologischen Überzeugungen, fielen in polizeilichen Kontexten auf oder waren inhaftiert. Das *API* bietet einen möglichst niedrigschwelligen Zugang, der Gesprächsbereitschaft, Freiwilligkeit und Veränderungsbereitschaft voraussetzt.

Ziel des *API* ist es, das extremistische Personenpotenzial, das nach Definition des Verfassungsschutzes dem Phänomenbereich Islamistischer Extremismus in Nordrhein-Westfalen zugerechnet

wird, zu reduzieren und einschlägige Straftaten zu verhindern. Die Teilnahme am *API* basiert auf einem initialen Ausstiegswillen der Klient:innen und zielt auf die Distanzierung und Deradikalisierung² sowie die Stabilisierung der sozialen Lebensbezüge und die Reintegration in die demokratische Gesellschaft. Die Klient:innen sollen in die Lage versetzt werden, ein selbständiges und selbstbestimmtes Leben ohne Extremismus zu führen. Sie sind häufig mit diversen Problemlagen konfrontiert, sodass im Rahmen der *sozialen Stabilisierung* individuelle Hilfestellungen zur Bewältigung des Alltags angeboten werden. Themen können z.B. Arbeitslosigkeit, finanzielle Schwierigkeiten oder psychische Erkrankungen sein. Ein weiterer wesentlicher Anteil im Rahmen der Tertiärprävention ist die individuelle *ideologische Aufarbeitung*. Dazu werden in persönlichen Gesprächen zwischen Klient:innen und Ausstiegsbegleiter:innen unter anderem Einstiegsprozesse beleuchtet und undemokratische Denkmuster hinterfragt, ohne dass die Religion aufgegeben werden soll.

Die Zahl der Kontaktaufnahmen von ausstiegswilligen Personen im Bereich Islamismus mit dem *API* hat sich seit Start des Programms im Jahr 2014 vervielfacht. Aktuell befinden sich ca. 65 Personen in Begleitung, davon ist rund ein Drittel inhaftiert. Das *API* begleitet sowohl Männer als auch Frauen, wobei mehr Männer am Programm teilnehmen. Die Frauen stellen ungefähr ein Viertel der begleiteten Personen. Bei etwa 45 der aktuellen Begleitungen konnte das *API* bereits eine deutliche Distanzierung von der extremistischen Ideologie und Szene feststellen. Es ist aus Sicht der Begleiter:innen zu erwarten, dass ein Großteil dieser Fälle kurz- bis mittelfristig positiv abgeschlossen werden kann. In über 30 Fällen wurde ein positiver Fallabschluss bereits vermerkt.

Das multiprofessionell aufgestellte Team des *API* bringt professionelle Expertise aus den Bereichen Pädagogik, Psychologie sowie Rechts- und Politikwissenschaften mit und umfasst auch Personen mit Berufserfahrungen aus Polizei, Verfassungsschutz oder Justiz. Dies bietet die Möglichkeit, beispielsweise therapeutische Ansätze, Anti-Aggressions- oder Anti-Gewalttrainings an den Phänomenbereich Islamismus anzupassen, in die Begleitung zu integrieren, auf die individuellen Erfordernisse Rücksicht zu nehmen und bei Bedarf eine zielgerichtete Haftbetreuung zu leisten. Das Referat „Prävention und Aussteigerprogramme“, in dem das *API* angesiedelt ist, steht in vielfachen Kontexten in Verbindung zu wissenschaftlichen Akteuren. Im Referat wurde durch die Schaffung der Stelle einer Referentin, die eigens für das Thema Evaluation zuständig ist (der Evaluationsbeauftragten), außerdem eine besondere Ressource für diesen Bereich geschaffen, die für die vorliegend vorbereitete Evaluation durch eine AG Evaluation ergänzt wurde. In der AG Evaluation sind verschiedene Personengruppen des *API* vertreten, inklusive der Ausstiegsbegleiter:innen.

2 Distanzierung (engl. *disengagement*) beschreibt den Prozess der Loslösung und des Rückzuges eines Individuums aus extremistischen Bestrebungen im Sinne von Verhaltensänderungen. Beispiele sind der Verzicht auf das Engagement in extremistischen Gruppierungen oder auf gewaltbereite Handlungen im Kontext der extremistischen Weltanschauung. Deradikalisierung beschreibt ebenfalls als Prozess verstandene kognitive Veränderungen bezüglich extremistischer Weltbilder und Wertesysteme. Beispiele sind die Ablehnung der extremistischen Ideologie und die Annahme einer mit demokratischen Prinzipien vereinbare Werthaltung. Für weitere Diskussionen der Begrifflichkeiten siehe Baaken et al. 2018.

2.2 DIE ZUSAMMENARBEIT ZWISCHEN API UND PREVAL

Die Zusammenarbeit zwischen Mitarbeitenden von *PrEval* und des *API* erstreckte sich über einen Zeitraum von rund neun Monaten. Die Kontaktabbauung zu einer möglichen Beteiligung an den *PrEval*-Pilotstudien mit dem Ministerium des Innern des Landes Nordrhein-Westfalen und den dort angesiedelten Präventionsprogrammen erfolgte im Mai 2020. Aufgrund anlaufender Vorbereitungen für eine Evaluation des *API* schien das Aussteigerprogramm als Partnerprojekt für eine *PrEval*-Pilotstudie besonders geeignet. Für *PrEval* ergab sich hierdurch wiederum die Möglichkeit, die Perspektive eines institutionalisierten, staatlich getragenen Aussteigerprogramms einzubeziehen.

Im Zuge einer Evaluation des *Aussteigerprogramms Rechtsextremismus* des Landes Nordrhein-Westfalen im Jahr 2015 durch die Hochschule Esslingen konnten bereits positive Erfahrungen mit der Evaluation eines landeseigenen Aussteigerprogramms gesammelt werden. An dieser zurückliegenden wissenschaftlichen Begleitung sollte sich die auszuschreibende Evaluation des *API* inhaltlich und methodisch orientieren.

Im Sommer 2020 wurde die gründliche inhaltliche Vorbereitung der geplanten Evaluationsausschreibung als Ziel des gemeinsamen Arbeitsprozesses festgehalten. Von vornherein war das Interesse des *API* an einer umfassenden wirkungsorientierten Evaluation mit formativen Elementen³ gesetzt. In der Vorbereitung sollte detailliert auf Fragestellungen, Datenquellen, Methoden und Möglichkeiten der Datenanalyse für das Programm eingegangen werden, ohne jedoch die Gestaltungsmöglichkeiten der zukünftigen Evaluierenden einzuschränken. Die Planung der Ausschreibung erfolgte im Rahmen von fünf gemeinsamen Arbeitssitzungen (meist in Form von Online-Workshops) zwischen Juli 2020 und Januar 2021 und wurde zusätzlich durch den bilateralen Austausch und Absprachen mit der programmzuständigen Evaluationsbeauftragten vertieft.

Die Zusammenarbeit wurde dialogisch und partizipativ umgesetzt, um alle für die Arbeit des *API* relevanten Perspektiven einzubeziehen. Aus diesem Grund wurden bereits in der Planungsphase Mitarbeitende des *API* involviert, deren Expertise im Evaluationsprozess gefragt wäre. Konkret wurden jeweils punktuell Ausstiegsbegleiter:innen, Referent:innen des Innenministeriums und Mitarbeitende aus dem Wissenschaft-Praxis-Transfer hinzugezogen. Die Einschätzungen der Mitarbeitenden des Programms waren einerseits sehr wichtig, um sicherzustellen, dass die Evaluation praxisrelevante Ergebnisse anstrebt und andererseits, um in der Folge die Chancen für eine erfolgreiche Integration der Evaluationserkenntnisse in die Praxis zu erhöhen.

Zentrales Produkt des gemeinsamen Arbeitsprozesses von *PrEval* und *API* ist eine Matrix, die den Fragenpool enthält, der gemeinsam mit den involvierten Personen identifiziert und diskutiert wurde.⁴

3 Eine formative Evaluation soll dem Zweck der Verbesserung und Steuerung des Evaluationsgegenstands dienen. Sie richtet sich hierfür primär an Projekt- bzw. Programmverantwortliche, wird in der Regel begleitend zur Maßnahme durchgeführt und ist häufig zyklisch angelegt (DeGEval 2021: 67).

4 Da zum Zeitpunkt der Fertigstellung dieses Reports das Ausschreibungsverfahren der Evaluation noch nicht abgeschlossen ist, muss hier von der Veröffentlichung der Matrix abgesehen werden. In Tabellenform führt sie pro Zeile eine Evaluationsfrage mit I. zur Erhebung geeigneten vorhandenen Datenquellen (etwa: internes Handbuch), II. zu

Insgesamt beinhaltet die Matrix rund 80 für die Evaluation relevante Kernfragen und führt diese mit möglichen Datenquellen und geeigneten Untersuchungsmethoden zusammen. Wertvoll war hierbei besonders die Diskussion mit den Ausstiegsbegleiter:innen des *API* über tätigkeitsrelevante Erkenntnisinteressen, geeignete Informationsquellen und Erhebungsmethoden. Zukünftige Evaluator:innen können sich bei der Finalisierung eines Evaluationsdesigns für das *API* an dieser Grundlage bedienen.

Mitte Dezember 2020 wurde die Evaluationsmatrix und ein Strukturvorschlag zum Aufbau und Inhalt der Ausschreibung durch die Evaluationsbeauftragte der *API*-internen AG Evaluation vorgestellt. Im Januar 2021 konkretisierte die AG mögliche Designoptionen, indem sie weitere geeignete und noch fehlende Datenquellen identifizierte und die Umsetzbarkeit von Erhebungsmethoden diskutierte. Die AG ist auch im weiteren Planungs- und Evaluationsprozess regelmäßig durch die Evaluationsbeauftragte eingebunden. Die durch den gemeinsamen Arbeitsprozess konkretisierte und durch das zuständige Referat finalisierte Evaluationsausschreibung erfolgte im Juni 2021. Für einen Evaluationszeitraum von 12 Monaten steht hierbei ein höherer fünfstelliger Betrag zur Verfügung. Der Evaluationsbericht ist zum Ende des Jahres 2022 vorzulegen.

Parallel zum Abschluss der Designvorbereitung und -planung begann zum Jahresanfang 2021, ebenfalls in Zusammenarbeit von *API* und *PrEval*, die gemeinsame Reflexion über die Erkenntnisse, die aus dem Vorbereitungsprozess im spezifischen Kontext und ähnlich gelagerten Kontexten der Extremismusprävention gezogen werden können. Der vorliegende Report ist ein zentrales Ergebnis dieser Reflexion.

3. HERAUSFORDERUNGEN WIRKUNGSORIENTIERTER EVALUATIONEN BEGEGNEN

Mit dem Begriff Evaluation werden häufig Fragen nach der Wirksamkeit von Programmen, Projekten oder Maßnahmen assoziiert, die einen bestimmten Einfluss auf die Gesellschaft nehmen sollen. Gerade in der tertiären Extremismusprävention kommt Wirksamkeitsfragen ein großes Interesse vonseiten der Öffentlichkeit, aber auch von den beteiligten Akteuren (z.B. politische Entscheidungsträger:innen, Koordinations- und Leitungsebene der ausführenden Träger, inhaltliche Umsetzung, Programmadministration) der Programme selbst zu. Wirkung zu erfassen ist jedoch nicht trivial. Im Folgenden werden die Herausforderungen skizziert, die für wirkungsorientierte Evaluationen in der tertiären Extremismusprävention zu erwarten sind und es wird ein möglicher Umgang mit diesen vorgeschlagen. Teils traten die Herausforderungen auch in der vorliegend beschriebenen Zusammenarbeit zu Tage und konnten so konkret diskutiert werden, teils handelt es sich um solche, die voraussichtlich erst im Zuge der Evaluationsdurchführung relevant und hier nur abstrakt erläutert werden. Die Darstellung der entsprechenden Herausforderungen folgt der Chronologie ihres möglichen Auftretens im Laufe

erhebenden bzw. noch nicht erschlossenen Datenquellen (etwa: zu befragende Interviewpartner:innen), III. für die Datenquellen geeignete Auswertungsmethoden (etwa: qualitative Inhaltsanalyse) und IV. damit eventuell einhergehende Genehmigungsverfahren zusammen. Die Fragen wurden zudem jeweils einer Untersuchungsebene (Wirksamkeiten, Wirkindikatoren oder Wirkanalyse) und einem übergreifenden Themenblock (u.a. Fallaufnahme, Multiprofessionalität im Team, methodische Vielfalt) zugeordnet.

eines Evaluations(planungs)prozesses. Das folgende Kapitel soll eine Übersicht und Diskussionspunkte vorstellen, die auch über diese Fallstudie hinaus informativ sein können.

3.1 GRUNDLEGENDE ERWARTUNGEN UND VORBEHALTE

In der vorliegenden Vorbereitung des Evaluationsdesigns waren zahlreiche Entscheidungen notwendig, bezüglich derer seitens der beteiligten Personen unterschiedliche Erwartungen und Vorbehalte zu vermuten waren. Im hier dargestellten Prozess wurde über all diese Entscheidungen hinweg ein dialogisches und multiperspektivisches Prinzip verfolgt, indem Mitarbeitende des *API* aktiv in den Entwicklungsprozess eingebunden wurden. Besonders wichtig erschien der Austausch mit denjenigen, die die Klient:innen am besten kennen und die (Aus-)Wirkungen der Arbeit des *API* täglich beobachten: den Ausstiegsbegleiter:innen.

Gerade aufseiten der Begleiter:innen in Aussteigerprogrammen können, neben der Schutzbedürftigkeit der Beratung, Sorgen darüber bestehen, ob die Ergebnisse einer Wirksamkeitsevaluation ausschließlich eine sogenannte Rechtfertigungsfunktion für Mittelgebende zur (Nicht-)Weiterfinanzierung eines Programms bedienen, dass also die Evaluation nur dazu genutzt wird, diese zu begründen. Beleuchtet eine wirksamkeitsorientierte Evaluation ausschließlich die Frage, *ob* ein Programm wirkt, kann sie augenscheinlich die Basis für Entscheidungen zur Zukunft des Programms bieten. Zur tatsächlichen Weiterentwicklung der Maßnahmen eines Programms ist sie jedoch nur eingeschränkt hilfreich, da Fragen danach, wie einzelne Maßnahmen wirken, welche Aspekte dafür zentral und welche eher kontraproduktiv sind, nicht vertieft behandelt werden und so organisationales Lernen kaum möglich ist. Dennoch hat die Rechtfertigungsfunktion von Evaluationen ihre Berechtigung, weil Programme häufig aus staatlichen (Förder-)Mitteln finanziert sind und Rechenschaft für die Verwendung dieser Mittel abgelegt werden muss. Auch Praktiker:innen müssen ihre Arbeit im Alltag rechtfertigen können, sei es für das Selbstverständnis der eigenen Arbeit oder im Kontakt mit Klient:innen und/oder Kooperationspartner:innen. Wenn jedoch infrage steht, ob das Programm überhaupt weiterfinanziert wird, können Evaluationen von den betroffenen Praktiker:innen als äußerst bedrohlich wahrgenommen werden.

Diese Sorge war im vorliegenden Fall zwar reduziert, da das *API* dauerhaft institutionell verankert ist, dennoch waren Bedenken möglich, ob die Evaluation reichhaltig genug sein würde, um die Praxis voranzubringen. Dabei sei angemerkt, dass weder die Rechtfertigungsfunktion noch ein Fokus auf Wirksamkeit verhindern, dass eine Evaluation für die Weiterentwicklung der Praxis hilfreiche Ergebnisse erzielt. Aufgrund der häufig verkürzten Assoziationskette von (Wirksamkeits-)Evaluationen mit einer reinen Rechtfertigungsfunktion ohne Entwicklungszweck wird jedoch vielfach eine generelle Skepsis ihnen gegenüber auf Seiten der Präventionsträger vermutet (Möller 2019: 2 nach Gruber/Lützing 2017). Dieser Vermutung stehen erste Erkenntnisse aus einer Umfrage von *PrEval* in Kooperation mit den *Kompetenznetzwerken Islamistischer Extremismus (KN:IX)* und *Rechtsextremismus (KompRex)* entgegen, nach denen es eine grundlegende Skepsis oder gar Ablehnung von Evaluationsvorhaben aufseiten der befragten Präventionsakteure nicht gibt. Stattdessen zeigt die Umfrage, dass die Praktiker:innen die Potenziale einer wirksamkeitsorientierten Analyse sehen, wenn diese

sie mit konkreten Hinweisen darüber informiert, auf welche Maßnahmen es „sich lohnt“, Akzente zu setzen, oder wenn sie konkrete Anpassungen empfiehlt und nicht lediglich ein „Daumen hoch“/„Daumen runter“ vergibt. Eventuelle Bedenken rührten nach Aussage der befragten Präventionsakteure eher daher, dass die konkrete Umsetzung von Wirksamkeitserfassung wie eine *Black Box* erschien und die folgenden Fragen unbeantwortet ließ: Wie wird in der Evaluation die eigene Arbeit beurteilt werden? Anhand welcher Kriterien kann überhaupt eine solche Beurteilung stattfinden? Wie gut kann sie die Arbeit beschreiben und die tatsächlichen Veränderungen spiegeln? Die Befragten betonten, dass Evaluationen dann gewinnbringend seien, wenn Evaluierende Fachexpertise zum Feld der Deradikalisierungsarbeit mitbringen und so die (Arbeits-)Prozesse grundsätzlich nachvollziehen können. Dabei sollten aus ihrer Sicht die angewandten Evaluationsdesigns spezifisch auf das Feld der Tertiärprävention zugeschnitten sein (Koynova et al. 2022).

Das vorliegende Beispiel der Evaluationsvorbereitung macht deutlich, dass eine dialogische Zusammenarbeit unter Beteiligung verschiedener Akteure nicht nur Sorgen ausräumen, sondern auch den Praxisbezug und die Identifikation der Evaluierten mit der Evaluation erhöhen kann. Transparenz und dialogische Gestaltung in der Evaluationsvorbereitung hatten im konkreten Fall mehrere Vorteile: Es konnten bereits im Vorfeld Erwartungen der beteiligten Personen abgebildet und geprüft werden (Erwartungsmanagement), um so programmintern eine höhere Akzeptanz des geplanten Vorhabens zu erreichen. Dadurch wurden mögliche Vorbehalte zu einem frühen Zeitpunkt thematisiert und im Vorbereitungsprozess mitgedacht. Tatsächlich meldeten die Ausstiegsbegleiter:innen im Anschluss an den Prozess zurück, dass bereits die gemeinsame Erarbeitung der Fragestellungen und die Transparenz bezogen auf die zu erhebenden Daten etwaige Befürchtungen reduzierten. Die Identifikation aller Beteiligten mit der Evaluation konnte damit ebenfalls gestärkt werden, da die Vorbereitung im Team stattfand, der Fortschritt für alle sichtbar und damit ein Teamprodukt war. Das gemeinsame Nachdenken über Evaluationsmöglichkeiten und -ziele ermöglichte dabei, bereits in dieser Phase über die Möglichkeiten des Transfers der erwarteten Evaluationsergebnisse in die Praxis und die Rolle der Evaluierenden in diesem Prozess zu diskutieren.

3.2 ABBILDEN VON WIRKUNG

Zuallererst muss bei der Vorbereitung eines Evaluationsvorhabens die Komplexität des Gegenstandes der Arbeit, in diesem Fall der (De-)Radikalisierung, anerkannt werden. Zahlreiche Faktoren sind relevant für (De-)Radikalisierungsprozesse und beeinflussen damit den Ausgang der ebenfalls vielschichtigen Interventionsbemühungen von Präventionsakteuren (*Multifaktorialität*) (Baaken et al. 2018: 7). Dazu zählen etwa gesamtgesellschaftliche Rahmenbedingungen, Umstände und Ereignisse im unmittelbaren sozialen Kontext des Programms und der Klient:innen sowie ihre individuellen Eigenschaften und Voraussetzungen.⁵ Zusätzlich beeinflussen diese Faktoren sich gegenseitig (*Multikollinearität*), was die Isolierung von einzelnen Einflüssen erschwert. Aus diesem Grund ist

5 Hierbei sei außerdem angemerkt, dass sich die Grundlagenforschung zu (De-)Radikalisierung und ihren Einflussfaktoren in stetigem Wandel befindet und ein Konsens zu einigen Kernbegriffen weiterhin fehlt. Dies hängt auch mit dem Wandel des Phänomens, wie es in der Welt beobachtbar ist, sowie mit politischen Konjunkturen und ihrem Einfluss auf Forschungsförderung zusammen (Abay Gaspar et al. 2018).

es schwierig zu belegen, dass Ergebnisse eines Präventionsprogramms auf der individuellen Ebene von Verhaltens- oder Einstellungsveränderungen tatsächlich auf (eine) bestimmte Maßnahme(n) des Programms zurückzuführen sind (*Kausalitätsproblem*) (Stegmüller 1969).

Eine weitere Schwierigkeit liegt im Auftrag von Programmen der Extremismusprävention begründet. Denn während sich das allgemeine Verständnis von Wirkung auf die Veränderung bezieht, die eine Maßnahme oder eine Aktivität herbeiführt, ist das Ziel von tertiärer Präventionsarbeit zumindest teilweise das Nicht-Eintreten von bestimmten Entwicklungen. Ein tertiär-präventives Programm soll einerseits den Weg in die Re-Integration befördern und andererseits weitere Radikalisierung und Gewaltausübung verhindern. Ausbleibende Ereignisse sind schwer zu erfassen und noch schwerer ist der Grund für ihr Ausbleiben belegbar. Die Zahl der Ereignisse, die durch die Präventionsarbeit verhindert wurde, ist daher in der Praxis nicht erfassbar, während ein einziger existierender Vorfall, z.B. die Teilnahme an Szeneveranstaltungen oder im schlimmsten Fall die Beteiligung an Gewaltverbrechen, bereits die Wirkannahmen angreift. Ebenso kann es dazu führen, dass Prävention gerade dann für nicht notwendig befunden wird, wenn sie funktioniert, da das, was sie verhindern soll, nicht sichtbar wird. Es liegt außerdem in der Natur der Sache, dass nicht abbildbar ist, was geschehen wäre, wenn keine oder andere Präventionsarbeit stattgefunden hätte. Sehr deutlich wurde diese Problematik auch im Rahmen der Coronavirus-Pandemie in Deutschland, wo Maßnahmen zur Eindämmung der Virusverbreitung von Einigen für unnötig befunden wurden, wenn die Rate von Erkrankungen (auch durch diese Maßnahmen) niedrig blieb. Dieses als *Präventionsparadox* bezeichnete Problem stellt Evaluationsarbeit vor verschiedene Herausforderungen, welche aufgrund der sicherheitspolitischen Relevanz in der Tertiärprävention besonders spürbar werden.

Für eine Evaluation, die Wirkung abbilden will, ist die eingeschränkte Beobachtbarkeit des Programmserfolgs ein Problem. Dennoch macht sie wirkungsorientierte Evaluationen nicht unmöglich. Um Wirkung nachzuzeichnen, können verschiedene methodologische und paradigmatische Perspektiven eingenommen werden. Darin wird vom abstrakteren gewünschten *Impact* des Evaluationsgegenstandes (im Falle extremismuspräventiver Programme etwa die Verhinderung oder Verringerung religiös oder politisch motivierter Gewaltstraftaten in Deutschland) abgerückt und auf unmittelbare, beobachtbare und positiv formulierbare Resultate (*Outcomes* und *Outputs*) fokussiert.

Ergebnisse von Präventionsprogrammen können differenziert werden bezüglich der Ebenen, auf denen sie eintreten. Im (neu-)deutschen Sprachgebrauch für Evaluationen wird Wirkung daher aufgliedert in *Outputs*, *Outcomes* und *Impacts* (Phineo 2016: 35). Während sich *Outputs* auf die unmittelbar beobachtbaren ausgeführten Aktivitäten beziehen, wie z.B. auf die Durchführung von zehn Beratungsterminen mit insgesamt fünf Teilnehmenden, beschreiben *Outcomes* bereits gewünschte Änderungen bei der Zielgruppe, z.B. dass diese ihre Frustrationstoleranz erhöhen oder Ungleichheitsideologien hinterfragen. Als *Impact* werden Veränderungen in den Sozialräumen der Zielgruppe beschrieben, im Kleinen (wie in der Familie, einer Gruppe, einer Schule, einer Nachbarschaft, einer Vollzugsanstalt) oder im Großen (z.B. in einer Stadt, einer Gesellschaft, einem Staatenbund). Da die Beziehung zwischen Aktivitäten und (nachweisbaren) Wirkungen mit höherer Wirkungsebene immer uneindeutiger wird und zeitliche sowie finanzielle Ressourcen oft eingeschränkt sind, kann sich selbst eine umfassende Evaluation meist maximal auf die Analyse von *Outcome*-Wirkungen konzen-

trieren. Gerade die Offenlegung von unmittelbaren *Outputs* und *Outcomes* kann jedoch sehr wichtig für die konkrete Weiterentwicklung der evaluierten Arbeit sein. In der tertiären Extremismusprävention, in der auf der *Impact*-Ebene ein negativ formuliertes Hauptziel steht (keine weitere Radikalisierung oder ideologisierte Gewalt), können auf *Output*- und *Outcome*-Ebene positiv formulierte Ziele benannt werden, welche einfacher messbar sind.

Eine zentrale und häufig mit Wirkungsüberprüfung assoziierte Methodik sind *Randomized Controlled Trials* (RCTs), welche aus Evaluationen im medizinischen Bereich stammen und dort im Konzept der Evidenzbasierung verankert sind (Armborst 2019). Evidenzbasiert meint grundsätzlich, dass die Verknüpfung zwischen einer Aktivität und dem mit ihr angestrebten Ergebnis wissenschaftlich belegt sein soll. In der medizinischen Forschung gelten dementsprechend solche Behandlungen als evidenzbasiert, welchen im Rahmen von RCTs eine Wirkung nachgewiesen werden kann. In RCTs erhält, vereinfacht gesprochen, eine Gruppe die Behandlung und eine andere nicht (die sogenannte Kontrollgruppe); der Unterschied zwischen den beiden Gruppen im Ergebnis zeigt die Wirkung der Behandlung. Individuen, die bestimmte Kriterien (Einschlusskriterien) zur Teilnahme erfüllen, werden zufällig – also randomisiert – einer der Gruppen zugeteilt. Durch die Einschlusskriterien wird sichergestellt, dass alle Personen geeignet wären für die Behandlung, also etwa die entsprechende Erkrankung und keine Kontraindikation für die Behandlung aufweisen. Durch die Zufallsverteilung soll die Wirkung anderer Einflussfaktoren auf den Behandlungserfolg ausgeglichen werden, sodass kein anderer systematischer Unterschied mehr zwischen den Gruppen vorhanden ist als die Behandlung selbst. Die Randomisierung wird dabei effektiver, je größer die Fallzahlen sind. Im Sinne der Evidenzbasierung kann Wirkung also als eine kausale Folge einer Ursache konzeptualisiert werden, die unabhängig vom individuellen Kontext bei einem Großteil der potenziell zu behandelnden Personen vorhanden ist oder vorhanden wäre, die mithilfe wissenschaftlicher Methoden abbildbar ist und die in der Folge in der Praxis eingesetzt werden kann.

In der tertiären Extremismusprävention – im Speziellen bezogen auf Aussteigerprogramme – sind RCTs jedoch nur bedingt einsetzbar, insbesondere, um Fragen nach der Wirkung eines Programms im Ganzen zu beantworten. Dies hat mehrere Gründe. Erstens kann es ethisch und auch aus einer Sicherheitslogik heraus problematisch sein, einer Personengruppe bewusst eine Präventionsmaßnahme vorzuenthalten. Um eine geeignete Kontrollgruppe zu erstellen, müsste dazu Personen, die am Programm teilnehmen möchten, diese Teilnahme verwehrt werden. Denn würde man Personen als Kontrollgruppe heranziehen, die kein Interesse am Programm haben, wäre der Ausstiegswille ein Unterschied zwischen den Gruppen, der die Effektschätzung des Programms potenziell stark verfälschen könnte. Das Verwehren von Unterstützungsangeboten könnte allerdings ein unmittelbares Sicherheitsrisiko für die Bevölkerung, aber auch für die Personen selbst darstellen, die in Gefahr geraten können, wenn sie ohne Unterstützung in einer Szene verweilen, von der sie sich eigentlich distanzieren möchten. In anderen sicherheitsrelevanten Kontexten, wie etwa der forensischen Psychiatrie, in der inhaftierten (Gewalt-)Straftäter:innen Therapie- oder Trainingsangebote gemacht werden, wird daher ebenfalls zumeist auf RCTs verzichtet und stattdessen auf Vorher-Nachher-Vergleiche zurückgegriffen (Schmucker 2007: 17). Zweitens ist die Aussagekraft von RCTs vor allem dann gesichert, wenn die Intervention, die Teilnehmende „erhalten“, möglichst standardisiert ist, was in der Aussteigersarbeit aufgrund der individuellen Problem- und

Ressourcenlagen und aufgrund von Beratungsverläufen nur eingeschränkt möglich ist. Drittens müsste die Kontrollgruppe für einen sauberen Vergleich zwischen den Gruppen eine äquivalente Intervention erhalten, um ausschließen zu können, dass etwaige Wirkungen nicht nur darauf zurückzuführen sind, dass interveniert wurde, sondern auch, dass *auf diese Weise* interveniert wurde. Im medizinischen Bereich dienen dazu Placebos; im sozialen Bereich ist das Angebot eines Äquivalents weitaus aufwändiger und die Vergleichbarkeit streitbar. Viertens bestehen Aussteigerprogramme aus zahlreichen Beratungs- und Unterstützungsangeboten, deren Einzelbeitrag in einem Gruppenvergleich von Teilnehmenden vs. Nicht-Teilnehmenden nicht sichtbar würde. RCTs stellen die Frage, ob die Intervention ihre Wirkung im Schnitt entfaltet hat und nicht, welche ihrer Elemente auf welche Art gewirkt haben. Somit könnten nur sehr begrenzte Schlüsse zur Weiterentwicklung des Programms geschlossen werden. Es bestünde zwar, wie in Kapitel 3.6 genauer beschrieben wird, die Möglichkeit, einzelne Elemente des Programms etwa nur mit einer Subgruppe von Aussteiger:innen durchzuführen, dies ist jedoch ebenfalls aufwändig und im System der Ausstiegsarbeit – die nicht streng modularisiert und standardisiert vonstattengeht und gehen kann – oft unrealistisch. Ein weiterer Faktor, der ein solches Vorgehen erschwert, ist, dass für eine tatsächlich randomisierte Zuteilung und die (zumindest teilweise) von vornherein standardisierte Durchführung des Programms idealerweise erst neu eintretende Fälle abgewartet und dann genutzt werden müssten. Bei einer klassischen Evaluationslaufzeit ist dies keineswegs zu gewährleisten.

Andere Paradigmen, wie beispielsweise die theoriebasierte oder realistische Evaluation, verschieben den Fokus von generalisierbaren Kausalbeziehungen hin zu denjenigen, die eine Maßnahme oder ein Programm selbst für sich annehmen. Die Wirkannahmen, auf denen ein Präventionsprogramm aufbaut, werden hierbei als einzig relevanter Referenzpunkt zur Betrachtung seiner Wirkung gesehen. Gefragt wird also nicht schlicht, ob ein Programm funktioniert, sondern ob es so funktioniert, wie angenommen und gewollt. Wie im Folgenden genauer beschrieben, wird sich der kausalen Beziehung zwischen Aktivität und Ergebnis in diesen Ansätzen methodisch angenähert, indem mehrere, durch die Wirkannahmen festgelegte Zielmarker in der Entwicklung von (typischen) Fällen parallel betrachtet werden. Deren Gesamtbild spricht dann für oder gegen eine entsprechende Wirkung. Dieses Vorgehen schafft die Möglichkeit einer weiteren Öffnung der Fragestellung hin zu Wirkungsweisen, also Fragen danach wie, für wen und unter welchen Umständen ein Programm wirkt.

Es gibt zwei Gründe dafür, sich vollständig auf diese interne Projektlogik einzulassen: Erstens erkennt sie an, dass durch die unterschiedlichen Zielgruppen, Radikalisierungsverständnisse und Programmziele im Feld keine allgemeingültigen Annahmen als Maßstab für Wirkungsanalysen dienen können. Zweitens wird die Nützlichkeit der Evaluation für das Programm und die beteiligten Akteure erhöht. Auch in der vorliegenden Erarbeitung der Evaluationsausschreibung für das *API* sollte daher der eigenen Programmtheorie, der darin festgelegten Zielgruppe, dem Problemverständnis und den damit verbundenen Wirkannahmen Rechnung getragen werden. Um sich im hier vorgestellten Evaluationsvorhaben einer „Wirkung“ des *API* anzunähern, standen daher einerseits die Wirkannahmen des Programms im Zentrum und andererseits im Sinne der Wirkannahmen positiv formulierte Indikatoren, welche die realen Wirkungen abbildbar machen sollten. Dazu zählen etwa Veränderungen im Leben einer Person, die beobachtbar sind – neue Bezugspersonen außerhalb der Szene, soziales Engagement, aber auch Auseinandersetzungen mit der Vergangenheit, wie beispielsweise die Über-

nahme von Verantwortung für vergangenes Verhalten. Solche Wirkindikatoren können sich in der Dokumentation von Einzelfällen, in einzelnen Befragungen, im Berichtswesen des Programms und in begleitenden Untersuchungen wie Landes- und Bundesstatistiken finden. Ebenso können Einschätzungen von Personen(gruppen) wertvolle Indikatoren darstellen, die teils jedoch für die Evaluation als Daten noch erhoben werden müssen. Indikatoren können also im Programm standardmäßig vorliegen oder erst noch durch die Evaluierenden erschlossen werden. Denkt man an wissenschaftliche Überprüfungen von Kausalität, kann man also die Wirkannahmen als Hypothesen und die Indikatoren als approximative Hinweise auf deren Erfüllung oder Falsifizierung verstehen.

Damit war im vorliegenden Fall zwar entschieden, dass das Wirkungsverständnis dem des Programms selbst entsprechen sollte, es war aber noch kein Umgang mit dem oben genannten Kausalitätsproblem gefunden. Im Fall des *API* erschien der Ansatz vielversprechend, Daten und Erfahrungen aus möglichst vielen Quellen und Perspektiven zu erfassen (*Triangulation*⁶) (Bernasconi 2009: 99–100). Wenn sich dabei Beobachtungen zur Kausalität häufen, kann man eine kausale Beziehung als hohe Wahrscheinlichkeit formulieren. Damit ist zwar keine Kausalität nachgewiesen, es wird jedoch möglich, sich diesem Ziel anzunähern. Außerdem erhöht sich durch Triangulation die Wahrscheinlichkeit, ein authentisches und unverzerrtes Bild der Präventionsarbeit zu gewinnen, und damit auch die Realitätsnähe der Aussagen der Analyse. Neben den Ausprägungen von Variablen, die in der Falldokumentation zu Beginn der Begleitung, während des Prozesses und beim Fallabschluss beschrieben werden, und der Einschätzung der Ausstiegsbegleiter:innen sind z.B. Erkenntnisse von Bedeutung, die über Rückfallstatistiken und Rückfall-Konstellationen, Interviews mit Kooperationspartner:innen und nicht zuletzt mit den Klient:innen selbst gewonnen werden können.

3.3 WAHL DER INDIKATORIK

Um mit dem gewählten Ansatz eine Wirkanalyse durchzuführen, sollen also mithilfe von Indikatoren beobachtbare Effekte abgebildet und mit den Wirkannahmen abgeglichen werden. In einem solchen Prozedere ist die Wahl geeigneter Indikatoren, die die Veränderungen bei den Klient:innen beschreiben, ausschlaggebend. Die Indikatoren ermöglichen es auch, diese Veränderungen, vormals in der Programmlogik als Ziele definiert, zu explizieren. Dabei müssen für Evaluationen, wie oben angedeutet, nicht unbedingt neue Indikatoren gebildet werden. Aufgabe der Evaluierenden ist jedoch die Überprüfung und gegebenenfalls die Anpassung der bestehenden Indikatorik, die das Programm ohnehin zur Dokumentation und zum Monitoring der eigenen Arbeit aufbereitet.

Die Qualität von Indikatoren kann anhand der Gütekriterien Validität, Objektivität und Reliabilität festgemacht werden. Indikatoren sollten stets einer Validitätsprüfung (Gültigkeitsprüfung) unterzo-

6 Daten-Triangulation kann als Weg verstanden werden, eine Hypothese mit höherer Wahrscheinlichkeit verwerfen oder bestätigen zu können, wenn z.B. Berichte unterschiedlicher Beobachter:innen über die Entwicklung von Klient:innen einbezogen werden. Methoden-Triangulation kann als Ansatz gesehen werden, einen umfassenderen Blick auf die Forschungsfrage zu werfen, wenn z.B. qualitative Befragungen mit interpretativer Auswertung von quantitativen Daten und statistischer Auswertung ergänzt werden. Diskussionen um den Begriff und die Umsetzung der Triangulation haben (auch außerhalb der Evaluationsforschung) eine lange Tradition, zur weiteren Lektüre siehe Flick 2011.

gen werden, um abzuschätzen, welche Aussagekraft sie für welche Annahme entfalten und was sie nicht abschließend beweisen oder erklären können. Objektivität wiederum beschreibt die Unabhängigkeit von einem Untersuchungsergebnis von den Rahmenbedingungen und den Personen, die das Ergebnis festhalten: Werden beispielsweise zwei unterschiedliche Ausstiegsbegleiter:innen den Fortschritt einer Person auf die gleiche Art und Weise beurteilen? Da vollkommene Objektivität in der Realität nicht herstellbar ist, ist bei der Auswertung von Beratungsmaßnahmen neben inhaltlichen Indikatoren auch die wahrgenommene Leistung der einzelnen Begleiter:innen in den Blick zu nehmen, um die Objektivität der inhaltlichen Parameter besser einschätzen zu können. Die Reliabilität beschreibt die Zuverlässigkeit der Ergebnisse auf Basis der Indikatoren bei gleichen Rahmenbedingungen. Im Kontext der Tertiärprävention wäre die Reliabilität dann sichergestellt, wenn beispielsweise vergleichbare Methoden zur Beschreibung der Distanzierung eines:einer Klient:in zum gleichen Zeitpunkt zur gleichen Beurteilung führen.

In der vorliegenden Evaluationsvorbereitung wurde die Ebene der Indikatorik explizit in den Blick genommen. Dies geschah einerseits, indem vorhandene Indikatoren identifiziert wurden, die zur Beantwortung zentraler Evaluationsfragen dienen könnten, und andererseits, indem die Qualität der Indikatoren selbst zur Evaluationsfrage wurde. Verfügbare und bereits erhobene Indikatoren sollen so im Rahmen der anstehenden Evaluation einer Güteanalyse unterzogen und gegebenenfalls in der Zukunft angepasst werden. Während der Vorbereitung der Evaluationsausschreibung wurde das Thema der Qualität von Indikatoren also besprochen, die Güte der möglichen Indikatoren aber lediglich andiskutiert, da erstens die endgültige Wahl der Indikatoren für die Evaluation durch die Evaluierenden noch ausstand und ihre Qualität zweitens ein Fokuspunkt der Evaluation sein sollte. In diesem Zusammenhang wurden auch Aspekte thematisiert, die bei der Datenerhebung zu beachten sein würden, um Verzerrungen der Ergebnisse zu vermeiden, wie in den folgenden Kapiteln ausführlicher beschrieben wird.

3.4 ERKENNTNISINTERESSEN UND FRAGESTELLUNGEN

Um einzuordnen, ob die gewählte Indikatorik geeignet ist, muss geklärt werden, welche Bereiche des Programms in den Blick genommen werden sollen. Aufgrund der inhärenten Komplexität der (De-)Radikalisierung und der Vielfalt der Maßnahmen innerhalb eines Programms können auch bei umfangreichen Evaluationen (wie der hier geplanten) nicht alle Aspekte des Programms gleichermaßen beleuchtet werden. Es gilt daher, die zentralen Erkenntnisinteressen und die daraus folgenden Fragestellungen zu identifizieren.

Bereits von Anfang an war aufseiten des *API* das zentrale Interesse vorhanden, eine umfassende wirkungsorientierte Evaluation von externen Evaluierenden durchführen zu lassen. In Anlehnung an die durch die Hochschule Esslingen durchgeführte Evaluation des *Aussteigerprogramms Rechtsextremismus* des Landes Nordrhein-Westfalen sollten Aussagen über die Wirkung und Wirkungsweise des *API* getroffen werden. Als staatliches Programm, das aus Landesmitteln finanziert wird und in das auch Steuergelder fließen, ist die Nachvollziehbarkeit der Verwendung dieser Fördermittel ein wichtiges Anliegen für politische Entscheidungsträger:innen und die Allgemeinbevölkerung. Diesem

Anliegen kann mithilfe von wirksamkeitsorientierten Evaluationen Rechnung getragen werden. Da es sich um ein institutionalisiertes Programm handelt, sollte die Evaluation jedoch nicht das Ziel haben, über ein Fortbestehen des Programms zu entscheiden. Vielmehr sollten Erfolgsfaktoren auf der einen und Verbesserungspotenziale auf der anderen Seite herausgearbeitet werden.

Für die Zusammenarbeit zwischen *PrEval* und dem *API* wurde beschlossen, die Evaluationsaus-schreibung gemeinsam vorzubereiten. Der Auftrag hierzu umfasste zweierlei, nämlich 1) das Wirkungsverständnis und die Wirkannahmen des *API* zu explizieren und mithilfe von geeigneten Indikatoren abbildbar zu machen sowie 2) Fragestellungen, Datenquellen und Methoden zu sammeln und zu verschriftlichen, die eine Annäherung an die Wirkungsfrage erlauben.

Um den ersten Aspekt zu bedienen, wurden thematische Schwerpunkte definiert, in denen es für das *API* hilfreich sein könnte, Prozesse, Wissen und Annahmen (erneut) explizit zu machen. Im Rahmen der Ausstiegsbegleitung laufen zahlreiche Prozesse parallel und ineinandergreifend, einige sind konzeptionell festgehalten, andere haben sich als Teil der Programmkultur etabliert oder werden von den Begleiter:innen individuell gestaltet, ohne verschriftlicht zu sein. Folgende Schwerpunkte wurden unter anderem zur tieferen Analyse ausgewählt, da sie in der praktischen Arbeit einen zentralen Stellenwert haben:

- Konzept(e) der Ausstiegsarbeit;
- die angewandte Methodik in der praktischen Ausstiegsarbeit;
- der Prozess der Fallaufnahme, der Beratungsverlauf in seinen unterschiedlichen (sensiblen) Phasen, Fallabschlüsse und -abbrüche;
- der Fokus auf die Themenblöcke der *ideologischen Aufarbeitung* und der *sozialen Stabilisierung* als Pfeiler der Ausstiegsarbeit;
- der Einfluss der verschiedenen Professionen und der vielfältigen Methodik in der praktischen Arbeit innerhalb des Teams auf die Ausstiegsbegleitung;
- die Kooperation mit anderen Institutionen

Wie oben beschrieben, sollte auch die Ebene der (Eignung der) Indikatorik in den Blick genommen werden. Daher wurde als Vorbereitung auf die Sammlung und Entwicklung von Fragestellungen zunächst der Begriff der Wirkungsevaluation auf verschiedenen Analyseebenen diskutiert. Diese umfassten erstens die Formulierung von *Wirkannahmen*, zweitens die nutzbaren oder zu überprüfenden *Wirkungsindikatoren* und drittens die auf beiden Ebenen aufbauende *Wirkungsanalyse* in dem jeweiligen Themenbereich. Viertens wurde auch eine Fragenkategorie zum *Transfer wissenschaftlicher Erkenntnisse in die Praxis* gebildet, die im Anschluss jedoch in die Ebene der Wirkannahmen integriert wurde. An diesen Analyse-kategorien orientierte sich der anschließende Prozess der Fragensammlung und -entwicklung.

Die Kategorie „*Wirkannahmen*“ hatte zum Ziel, die expliziten und impliziten Annahmen zur Wirkung (einzelner Aspekte) des Programms zu beschreiben und gegebenenfalls zu definieren („Welche Wirkannahmen bestehen bezogen auf die angebotenen Aktivitäten?“). In der Kategorie „*Wirkindikatoren*“ sollte die zukünftige Evaluation prüfen, welche Indikatoren momentan genutzt werden oder genutzt werden könnten, um Wirkungen zu erfassen („Welche Indikatoren liegen aktuell vor, um diese

Annahmen zu überprüfen?). Darauf aufbauend sollte abgewogen werden, ob diese Indikatoren eine geeignete Annäherung an die zu beschreibenden Wirkungen anbieten. Bei Bedarf können Indikatoren modifiziert oder neue Indikatoren vorgeschlagen werden („Welche Indikatoren sind geeignet, um die Wirkannahmen zu untersuchen und können Wirkungen sichtbar machen?“). Die Kategorie „*Wirkungsanalyse*“ umfasste Fragestellungen, die darauf abzielen, die beobachtbaren Wirkungen anhand der festgelegten Indikatoren auszuwerten und daraus Schlussfolgerungen zu Wirkzusammenhängen abzuleiten („Sind die angenommenen Wirkungen eingetreten?“). Zusätzlich zu Fragen nach den Wirkzusammenhängen selbst konnten auch Fragen danach gestellt werden, inwiefern die geplanten eigenen Arbeitsprozesse im Sinne der Effizienz in der Realität umgesetzt werden. In einem solchen Fall lautete die Frage nach *Annahmen*: „Wie sollen unsere Arbeitsprozesse ablaufen?“, diejenige nach *Indikatoren* lautete: „Anhand welcher Indikatoren können wir sehen, wie die Arbeitsprozesse ablaufen?“ und diejenige auf der *Analyse*-Ebene: „Laufen unsere Arbeitsprozesse wie geplant?“.

Um den Praxisbezug der Fragestellungen sicherzustellen, wurden sie den Phasen einer idealtypischen Ausstiegsbegleitung (Fallaufnahme, Begleitungsverlauf mit ideologischer Aufarbeitung und sozialer Stabilisierung bis hin zum Fallabschluss) entsprechend chronologisch geordnet und nahmen auf diese Bezug. Das chronologische Vorgehen half gemeinsam mit den oben beschriebenen thematischen Schwerpunkten dabei, wichtige Fragen im Zusammenhang mit dem Beratungsprozess nicht zu übersehen. Zu jeder Phase konnten dabei relevante Fragen auf allen drei der oben genannten Ebenen (Wirkannahmen, -indikatoren und -analyse) gesammelt werden. Zum für die Begleitung wichtigen Element der *sozialen Stabilisierung* könnte sich auf der Ebene der Wirkannahmen etwa die Frage ergeben: „Welche Wirkungen streben wir mithilfe welcher Beratungselemente bezüglich der sozialen Stabilisierung an?“. Deren Fortsetzung auf der Ebene der Wirkindikatoren würde „Welche Kriterien werden zur Beschreibung dieser Ziele im Rahmen des Programm-Monitorings angewendet?“ oder „Welche Indikatoren würden diese Wirkungen bestmöglich abbilden?“ lauten. Auf Ebene der Wirkungsanalyse schliesse sich die Frage „Werden die definierten Kriterien für Wirksamkeit bezüglich der sozialen Stabilisierung erfüllt?“ an.

Fragen zur Multiprofessionalität des Teams und der Methodenvielfalt sowie deren Implikationen für die entfaltenen Wirkungen konnten auf Ebene der *Wirkannahmen* beispielsweise lauten: „Welche Rolle spielt die Multi-Professionalität des Teams bei den Zugängen zu Klient:innen in der Theorie?“. Mit Zugängen ist hierbei der Kontaktaufbau und Beziehungsbeginn mit Klient:innen gemeint, die von der Disziplin, aus der Begleiter:innen stammen, geprägt sein können. Diese Frage wurde auf der *Indikatoren*-Ebene formuliert als: „Anhand welcher Indikatoren können wir den Einfluss von Multiprofessionalität auf die Zugänge beobachten?“ und auf der Ebene der *Wirkungsanalyse* entsprechend: „Welche Rolle spielt die Multi-Professionalität des Teams bei den Zugängen in der Praxis?“.

Dieser Findungsprozess von Schwerpunkten und Fragestellungen war das Herzstück der Zusammenarbeit zwischen *PrEval* und *API*. Mit ihm wurde das gemeinsame Verständnis dessen geklärt, welche Erkenntnisse eine wirkungsorientierte Evaluation bringen sollte (und welche nicht), und wie sich dabei der Wirkung des Programms angenähert werden könnte. Darüber hinaus wurden Fragen und Inhalte diskutierbar, die andernfalls für die Mitarbeitenden des *API* möglicherweise wenig praxisbezogen wirken, wie die adäquate Erhebung und Nutzung von Daten im Rahmen einer Evaluation.

Letztendlich sind neben den Inhalten einer Evaluation auch diese methodischen Aspekte für das evaluierte Programm relevant, weil die Mitarbeitenden unweigerlich bei der Evaluation mitwirken und außerdem diejenigen sind, die von ihren Ergebnissen unmittelbar „betroffen“ sind. Daher war es allen Beteiligten ein Anliegen, die Prinzipien der Multiperspektivität und eines dialogischen Zusammenarbeitens von Anfang an ernst zu nehmen. So wurden zu den Treffen der Evaluationsbeauftragten mit dem *PrEval*-Team regelmäßig weitere Personen aus dem *API*, wie Ausstiegsbegleiter:innen, Referent:innen oder Mitarbeitende aus dem Wissenschaft-Praxis-Transfer, hinzugezogen. Obwohl der Austausch untereinander innerhalb des *API* ohnehin geschehen wäre, konnte der Resonanzraum mit dem *PrEval*-Team über das *API* hinaus erweitert und die dort vorhandenen Perspektiven mit externen konfrontiert werden. Durch die mit dem *PrEval*-Team hinzukommenden Kapazitäten konnte ein intensiver Fokus auf die ausführliche Vorbereitung der Evaluation gelegt werden, der den Ablauf und die (interne) Akzeptanz der wissenschaftlichen Begleitung wesentlich erleichtern kann.

Die Erarbeitung der Fragestellungen ist dabei keinesfalls trivial. Sie erfordert eine Reflexion darüber, welche konkreten Informationen für die Programmdurchführung und -entwicklung gewinnbringend sind – nicht „nur“ aus Sicht der Programmleitung, sondern aller ausführenden und beteiligten Personengruppen. Die detaillierte Ausformulierung der Fragestellungen im Vorfeld der Ausschreibung und die Aufstellung verfügbarer und erwünschter Datenquellen, die im Folgenden beschrieben wird, hatten außerdem den praktischen Nutzen, dass das zweistufige Auswahlverfahren im Ausschreibungsprozess, bei dem evaluierende Einrichtungen zunächst diverse Evaluationsansätze anbieten, entfallen konnte, da die Orientierung der Evaluation bereits in der Ausschreibung konkret benannt werden konnte.

3.5 VORHANDENE DATEN: VERFÜGBARKEIT UND BEARBEITUNG

Nach Sammlung und Sortierung der Fragestellungen sowie einer Schwerpunktsetzung auf diejenigen, die für die Arbeit des *API* am wichtigsten sind, wurden mögliche Datenzugänge und -erhebungsmethoden gesammelt und diskutiert. Die Zugänge zu den notwendigen Daten und die Praktikabilität der Befragungen mit den diversen Beteiligten sind jeweils unterschiedlich gelagert. Daraus ergeben sich Herausforderungen für die Wirkungsanalyse im Kontext der Tertiärprävention.

Für die Bearbeitung der Evaluationsfragestellungen sind viele Datenquellen potenziell nützlich. Einige Daten werden bereits beim *API* gesammelt, z.B. in Monitoringinstrumenten, oder sind anderweitig bereits vorhanden, z.B. in konzeptionellen Dokumenten. Bei solchen Quellen ist zu klären, ob und in welcher Weise sie für die Evaluation nutzbar gemacht werden können. Darüber hinaus besteht die Möglichkeit, weitere Daten zu erschließen, wie etwa öffentlich zugängliche Statistiken, oder im Rahmen der Evaluation neu zu erheben, wie etwa durch Befragungen verschiedener involvierter Personen. Weiterhin dem dialogischen und multiperspektivischen Prinzip folgend, wurde gemeinsam ein Kanon von möglichen Datenquellen gebildet, der einerseits aus Dokumenten, andererseits aus potenziellen Personengruppen für Befragungen bestand. Im Weiteren werden zunächst die Datenquellen präsentiert, die bereits vorliegen und so als Basis einer Evaluation im Bereich der tertiären Extremismusprävention diskutiert werden können.

Eine Dokumenten-Quelle stellte das *API*-Handbuch dar, das die konzeptionelle und theoretische Grundlage des Programms und, genauer, der Begleitungsarbeit darstellt. Dies kann als Grundlage zum Abgleich der Programmtheorie mit der praktischen Umsetzung dienen. Weitere Hinweise auf die konzeptionelle Grundlage lieferte der Handapparat des Programms, der die vorhandene und regelmäßig genutzte Fachliteratur umfasst, auf die in der Beratungsarbeit zugegriffen wird. Ergänzend dazu können die Außenkommunikation bzw. öffentliche Auftritte des *API* zur Analyse hinzugezogen werden, da dort in knapper Form ebenfalls Ziele und Methoden dargestellt werden und darüber hinaus ein Vergleich ermöglicht wird zwischen der internen Programmdefinition und der Außendarstellung. Medien, die dafür vollständig zur Verfügung gestellt werden können, sind Flyer, Social Media-Auftritte und Webseiten. Darüber hinaus können Stellenausschreibungen für Ausstiegsbegleiter:innen sowie Übersichten über ihre Qualifikationen zur Darstellung des beruflichen Anforderungsprofils und der Multiprofessionalität des Teams informativ sein.

Im vorliegenden Fall wurden das *API*-Handbuch, der Handapparat sowie sämtliche ohnehin öffentlich verfügbaren Quellen wie die Außenkommunikation oder etwa polizeiliche Statistiken als soweit wie möglich zugängliche Datenquellen für die spätere Evaluation festgehalten. Dazu zählten auch die Stellenaus- und -beschreibungen für Mitarbeitende, auch wenn die Auskunft über den Gesamtbestand der Mitarbeitenden, die im Programm tätig sind, aus Sicherheitsgründen nur rudimentär bleiben kann.

Neben diesen Dokumenten ist insbesondere die Dokumentation von Fallverläufen für die Evaluation interessant. Erhebungsbögen bilden hier die Basis für die Fallaufnahme und die laufende Beobachtung von Entwicklungen der Klient:innen. Die Fallarbeit wird zusätzlich qualitativ in Falldokumentationen zusammengefasst. Des Weiteren kann die sogenannte „Fallliste aktiver Fälle“ genutzt werden, die die Gesamtheit der Fälle und einzelne Parameter wie Alter und Geschlecht abbildet. Haben Klient:innen das Programm abgeschlossen oder abgebrochen, schwinden mit wachsendem zeitlichen Abstand die Möglichkeiten, Informationen über ihren weiteren Lebenslauf zu gewinnen. Die wenigsten Evaluationen können daher ausreichend langfristig angelegt sein, um Langzeitwirkungen zu beobachten.

Dokumente, die in ausgefüllter Form personenbezogene Daten beinhalten, sollten im Blanko-Format und, in einzelnen Fällen, in anonymisierter Form (falls möglich) zur Verfügung gestellt werden. Entsprechend sind die personalisierten Erhebungsbögen und Falldokumentationen vertraulich zu behandeln und nicht für Dritte zugänglich.

Dieser erste Datenkanon ist extensiv; viele der genannten Daten sind jedoch nicht ohne weiteres zugänglich, insbesondere, wenn es um konkrete Fälle und entsprechende personenbezogene Daten geht. Besonders die Daten von Klient:innen und die konkreten Inhalte der Begleitung sind sensibel und schutzbedürftig. Das Ministerium verpflichtet sich aus diesem Grund den höchsten Datenschutzbestimmungen den Klient:innen gegenüber, auch für die Durchführung der Evaluation. Ihr Einbezug kann außerdem zu Bedenken bezüglich der Begleitung führen, insbesondere, wenn externe Evaluierende hinzugezogen werden oder die Nutzung der Daten intransparent ist: Unterminiert die Evaluation die den Klient:innen zugesicherte Vertraulichkeit oder gefährdet den (positiven) Beratungsverlauf?

Um zu eruieren, welche der möglichen Datenquellen tatsächlich nutzbar gemacht werden können, muss beachtet werden, dass die meisten Quellen in unterschiedlichen Formen verwendbar sind. So können manche fall- und personenbezogenen Daten etwa anonymisiert oder pseudonymisiert werden, ohne dass die für den Evaluationsprozess entscheidenden Informationen verloren gehen. Im Falle einer Anonymisierung werden die Dokumente so bearbeitet, dass keine personenbezogenen Daten mehr enthalten sind, die Rückschlüsse auf Personen oder bestimmte Orte zulassen. Für den vorliegenden Fall könnte das beispielsweise die Veränderung identifizierender Merkmale und Details bedeuten. Eine Anonymisierung geht weiter in der Datenveränderung als eine Pseudonymisierung, die lediglich Namen abwandelt. Aus einer „mittelgroßen Stadt“ kann dann eine „Kleinstadt“ werden, aus einem „Jugendzentrum“, in dem sich eine Szene trifft, entsprechend ein „Sportverein“. Obwohl es für eine Evaluation nicht immer notwendig sein muss, diese Informationen wahrheitsgetreu zu verarbeiten, um die Evaluationsfragen zu beantworten, können spezifische Rahmendaten – Orte, Lokalitäten, Zeiträume – bisweilen durchaus informativ sein. So kann es Programmdurchführende beispielsweise interessieren, wo sie ihre Zielgruppen bereits gut erreichen und wo nicht. Daher muss für die Entscheidung für oder gegen eine Anonymisierung das Erkenntnisinteresse mitbedacht werden. Oberste Priorität hat aber stets der Schutz der Klient:innendaten. Auch sind Anonymisierungen sehr zeitaufwändig und es bedarf zum Einbezug des jeweiligen Erkenntnisinteresses einer gewissen Expertise, sowohl für die Evaluationsfragestellungen als auch für die konkreten Fälle. Eine praktische Lösung ist es daher, dass die Ausstiegsbegleiter:innen die Akten anonymisieren, falls ihre zeitlichen Ressourcen dies zulassen, auch, weil sie am besten einschätzen können, welche Veränderungen notwendig sind, um eine Nicht-Rückführbarkeit sicherzustellen. Anonymisieren die Begleiter:innen die Akten, müssen sie jedoch über die Erkenntnisinteressen der Evaluation mindestens informiert sein, um einschätzen zu können, welche Informationen besonders relevant sind. Evaluator:innen können die mit der Anonymisierung betrauten Personen, etwa im Workshopformat, vor oder während des Evaluationsprozesses hierfür schulen. Im Bericht zu den Erkenntnissen einer Evaluation kann der Schutz der personenbezogenen Daten aufrechterhalten werden, indem auf einem Gruppenlevel aggregierte Ergebnisse berichtet werden. Sollte es entscheidend für den Erkenntnisgewinn des Programms sein, (einzelne) Ergebnisse personenbezogen auszuwerten, kann dies im internen Kreis (etwa zusätzlich zu einer öffentlichen Berichtslegung) geschehen.

Um die letztendliche Nutzbarkeit von vorhandenen Daten zu klären und daraus folgende Entscheidungen für die sinnvollste Quellenkombination zu treffen, bedarf es mehr oder weniger komplizierter Abstimmungs- und Freigabeprozesse. Da diese Prozesse Zeit kosten, ist es sinnvoll, frühzeitig im Vorfeld einer Evaluation programmintern einen Überblick über vorhandene Datenquellen und ihre Nutzbarkeit zu schaffen. Dies war im vorliegenden Fall gegeben. Dem Prinzip der Multiperspektivität folgend, wurden auch hierbei verschiedene Akteure des *API* zu den Gesprächen hinzugezogen und die sich ergebenden Anfragen für Befragungen und Datenaustausch auch mit externen Partner:innen abgeklärt. Die intensive Vorbereitung und Beteiligung der Mitarbeitenden erlaubte es, Bedenken aller Akteure einzubeziehen, ihre Erwartungen anzugleichen und die Praxistauglichkeit der gewählten Quellen bezüglich Nutzbarkeit und Ressourcenintensität zu erhöhen. Im dialogischen Prozess wurden an dieser Stelle die Anforderungen der verschiedenen Akteure mit den vorhandenen Rahmenbedingungen in Einklang gebracht. Darüber hinaus konnten dadurch nach Abschluss der Zusammenarbeit sehr genaue Anforderungen an die potenziellen Evaluierenden gestellt werden. Auch die

Anonymisierung vorhandener Daten konnte so früh geplant und initiiert werden und entsprechende Arbeitsverantwortungen dafür im Rahmen des Programms verteilt werden.

Im vorliegenden Fall wurden nur solche Datenquellen in die Evaluationsausschreibung mit einbezogen, deren Verfügbarkeit final geklärt war. Die letztendliche Nutzung der verfügbaren Quellen wird sich gemeinsam mit den künftigen Evaluierenden klären. Dabei können Datenquellen nicht ohne die Erhebungs- und Analysemethoden gedacht werden, da die Informationen, die aus einer Quelle gezogen werden können, erheblich variieren können. Einen Einfluss auf die Datenqualität können die Evaluierenden und diejenigen, die die Evaluation vorbereiten, am ehesten dort ausüben, wo neue Daten speziell für eine Evaluation erhoben werden. Auch können verschiedene Methoden unterschiedlich aufwändig in der Anwendung sein oder Voraussetzungen in den Daten verlangen, die mitunter nicht gegeben sind. Im Folgenden werden daher relevante Aspekte bei der Auswahl von Methoden beschrieben.

3.6 NEU ZU ERHEBENDE DATEN: ANALYSE- UND ERHEBUNGSMETHODEN

Zentrales Thema in der vorliegenden Evaluationsplanung war es, herauszuarbeiten, wie angemessen einzelne Methoden für den angestrebten Erkenntnisgewinn sein würden und wie Verzerrungen der Ergebnisse vermieden werden könnten. Letzteres kann beispielsweise über die Abfrage mehrerer Perspektiven (z.B. von Klient:innen, Begleiter:innen, Kooperationspartner:innen etc.) bei der Beantwortung der Evaluationsfragen geschehen. Da jede Quelle, seien es vorhandene Dokumente oder Ansichten von Personen(gruppen), nutzbar gemacht werden sollte, bedeutet dies auch immer den Einsatz unterschiedlicher, aber möglichst ertragreicher und angemessener Methoden.

Aussteigerprogramme im Bereich der Extremismusprävention haben gegenstandsbedingt relativ kleine Fallzahlen, was die Aussagekraft mancher Methoden reduzieren und tiefergehende Explikationen notwendig machen kann. Die Nutzung und Analyse von qualitativen Informationen zu einzelnen Beratungsprozessen binden andererseits viele Ressourcen und stellen außerdem hohe datenschutzrechtliche Anforderungen, weil das Anonymisieren von Daten schwer bis unmöglich ist. Daher ist anzunehmen, dass den Evaluator:innen nicht zu allen Fällen die Gesamtheit der verfügbaren qualitativen Informationen zur Verfügung gestellt werden kann. Unter Berücksichtigung der Datenlage muss eine Auswahl exemplarischer Fälle getroffen werden. Die Auswahl möglicher Fallkonstellationen im Sinne der Wirkannahmen sollte vor der Anonymisierung und Pseudonymisierung diskutiert werden, damit eine Repräsentativität der ausgewählten Fälle für die Grundgesamtheit gewahrt wird. Der direkte Vergleich unterschiedlicher Fallverläufe ist im Zeitrahmen einer Evaluation oft nicht möglich: Fälle werden zeitlich asynchron aufgenommen, neben laufenden Begleitungsprozessen gibt es (anlaufende) Clearingverfahren oder ruhende Fälle, die zeitweise keine Beratung wahrnehmen – ganz abgesehen von den höchst unterschiedlichen individuellen Voraussetzungen der einzelnen Klient:innen zu Beginn und im Verlauf der Begleitung.

Die geringe Fallzahl bildet eine Hürde für die Aussagekraft einzelner quantitativer Verfahren. Es wurde bereits angesprochen, dass sich RCTs aus mehreren Gründen nicht gut für die Evaluation von

Aussteigerprogrammen als Ganzes anbieten. Eine Alternative wäre hier, auf natürliche Gruppen von Teilnehmenden vs. Nicht-Teilnehmenden zurückzugreifen und diese zu vergleichen. Hierbei würde sich jedoch ein Selbstselektionseffekt ergeben (Diekmann 2006: 18). Teilnehmende an einem gegenwärtig stattfindenden Aussteigerprogramm wünschen sich die Unterstützungsleistungen zum Ausstieg ausdrücklich, während Nicht-Teilnehmende diese Motivation nicht aufweisen. Somit ist ein systematischer Unterschied neben der Teilnahme am Programm auf jeden Fall vorhanden – nämlich die Motivation zum Ausstieg –, der entscheidenden Einfluss auf die Entwicklung der extremistischen Ansichten und Handlungen von Personen haben müsste. Weitere systematische Unterschiede, etwa in weiteren Faktoren, die mit der Ausstiegsmotivation zusammenhängen, können je nach Kontext und Zusammensetzung der Gruppe ebenfalls aufkommen.

Ein Interesse, das vom *API*-Team formuliert wurde, war, inwiefern Maßnahmen zur ideologischen Aufarbeitung innerhalb des Programms eine zusätzliche Wirkung über die Maßnahmen zur sozialen Stabilisierung hinaus entfalten und als eigene Programmsäule einen Mehrwert darstellen. Bis jetzt werden im Programm beide Ziele in der Klient:innenarbeit parallel verfolgt. Um die Wirkungen einzelner Bestandteile des Programms einzeln untersuchen zu können, könnte forschungsmethodisch erneut an ein experimentelles Vorgehen (wie in einem RCT) gedacht werden. Konkret könnte eine Gruppe von Klient:innen beide Maßnahmen „erhalten“, während eine andere nur Maßnahmen der sozialen Stabilisierung durchläuft (in diesem Fall die Kontrollgruppe). Die Wirkungen, die sich in beiden Gruppen zeigen, könnten dann verglichen werden. Vorhergehend wurde jedoch bereits genannt, dass die Nutzung von experimentellen Methoden, wie RCTs, für Wirkungsermittlungen von ganzen Programmen in der tertiären Extremismusprävention nur eingeschränkt machbar und sinnvoll ist. Dies gilt in ähnlichem Sinne auch für die differentielle Betrachtung einzelner Programmbestandteile. Zwar könnte dadurch klarer herausgearbeitet werden, welche Programmaktivitäten welche Wirkungen auslösen, sodass konkretere Schlüsse für die Weiterentwicklung der Arbeit gezogen werden könnten. Jedoch müsste auch hier die Maßnahme über alle Klient:innen hinweg mehr oder weniger standardisiert und die Klient:innen den Gruppen randomisiert zugeteilt werden, um sonstige systematische Unterschiede auszuschließen. Im Rahmen einer zwölfmonatigen Evaluation müssten dafür allerdings, wie oben angedeutet, genügend neue Fälle gesammelt werden können, was unrealistisch erscheint. Dies ist zwar bei der Gruppe, die beide Maßnahmen erhalten soll, nicht zwingend notwendig, weil auf abgeschlossene Fälle zurückgegriffen werden könnte, zumindest bei der Kontrollgruppe ist es aber nicht vermeidbar. Aufgrund der dadurch nicht-randomisierten Zuteilung zu den Gruppen werden weitere systematische Gruppenunterschiede (in den Eigenschaften der Klient:innen oder aber in der Durchführung der Maßnahmen, die sich z.B. über die Zeit ändern) wahrscheinlicher und ihr Einfluss schwerer einschätzbar. Dadurch schmälert sich die Aussagekraft zur Kausalität – ob unterschiedliche Wirkungen tatsächlich auf die absichtlichen Gruppenunterschiede zurückzuführen sind – deutlich. Bei einer großen Anzahl an Fragestellungen und Interessen ist der Erkenntnisgewinn, der durch ein solches Vorgehen (zur Beantwortung einer einzelnen Frage) entstünde, in den meisten Fällen nicht den Mehraufwand wert. Sollte sich dennoch für ein solches Design entschieden werden, müssten sich alle Akteure einig sein, dass sie selbst die Behandlung der Kontrollgruppe nicht als „unterlassene Leistung“ ansehen und dass sie bereit sind, entsprechenden politischen und medialen Deutungen entgegenzutreten.

Wird davon abgesehen, eine Gruppe bewusst einer Intervention auszusetzen und die andere nicht, können – im Falle genügender Fallzahlen – dennoch manche Informationen für Gruppenvergleiche genutzt werden. Konkret heißt das, dass natürlich vorkommende Unterschiede zwischen Klient:innen, wie etwa das Alter, genutzt werden, um Gruppen zu bilden und statistisch zu eruieren, ob sich bei diesen Gruppen unterschiedliche Erfolgsquoten der Ausstiegsgleitung zeigen. Alternativ könnten auch Korrelationen zwischen dem Einsatz einzelner Programmelemente und Programmwirkungen berechnet werden. Die aus solchen Analysen gezogenen Erkenntnisse können für sich stehen und dabei helfen, das Programm passgenauer auf (bestimmte Subgruppen von) Klient:innen zuzuschneiden. Sie können aber auch als Kontextualisierung für Erkenntnisse zur Wirkung dienen, etwa wenn untersucht wird, ob bestimmte Beratungselemente von jüngeren Personen als hilfreicher empfunden werden als von Lebensälteren. Auch im vorliegenden Fall wurde diskutiert, welche Quellen zur Kontextualisierung und für Vergleiche herangezogen werden könnten. Dies war möglich, weil das *API* vergleichsweise hohe Fallzahlen für ein Aussteigerprogramm aufweist.

Wenn für solche statistischen Vergleiche Erfolg in Zahlen definiert werden soll, werden oft sogenannte Rückfallquoten herangezogen, also die (ggf. erneute) Straffälligkeit von Personen, die das Programm eigentlich abgeschlossen haben. Diese Daten sind schwieriger zusammenzustellen als zuerst angenommen, da sie notwendigerweise nur von Personen vorliegen, die in irgendeiner Form durch staatliche Institutionen erfasst werden. Im vorliegenden Fall wurde in diesem Zuge diskutiert, was als Rückfall zu deuten ist – politisch-motivierte Verbrechen, Straftaten allgemein oder bereits die Wiederaufnahme von Kontakten, die zu der extremistischen Szene gezählt werden. Diese Parameter wird das *API* mit den zukünftigen Evaluierenden nach Klärung der Verfügbarkeit von Quellen weiter definieren.

Mehrere Personengruppen können informative Einblicke in die Annahmen, die etablierten Prozesse und die Wirkung des Programms geben. Dazu zählen die Mitarbeitenden des *API*, angefangen bei Leitungs- und Koordinationsebenen (die angesiedelt sind beim Innenministerium und dem Verfassungsschutz des Landes Nordrhein-Westfalen), über die involvierten ministerialen Referate bis hin zu den Ausstiegsgleiter:innen. Letztere sind diejenigen, die den Klient:innen am nächsten stehen und auf deren Arbeit die Evaluation die direktesten Auswirkungen hat, wodurch ihre Sichtweisen besonders relevant sind. Neben den hauseigenen Akteuren arbeitet das *API* mit mehreren Institutionen zusammen, die teilweise einen Einfluss auf den Prozess des Ausstiegs haben und deren Perspektiven daher ebenfalls wichtig sind. Zu diesen Kooperationspartnern zählen neben Justiz- und Sicherheitsbehörden (Staatsschutz, Landeskriminalamt, Justizvollzugsanstalten, Polizei) auch verschiedene Ämter (Sozialamt, Jugendamt, Ausländerbehörde, Agentur für Arbeit). Auch diese Kooperationsstrukturen spielen natürlich eine Rolle, wenn ein Distanzierungsprozess gelingen soll.⁷ Die Einbindung (externer) Kooperationspartner in die Evaluation ist jedoch nicht ohne weiteres umsetzbar, denn es sind nur bedingt Einflussmöglichkeiten vorhanden, um diese für die Teilnahme an den Erhebungen zu gewinnen, sei es für Datenaustausch oder für Befragungen. Der hier gewählte Anspruch, Erkenntnisse zu triangulieren, der sich aus dem Verständnis des

7 Diese komplexe Einflussdynamik in der sekundären und tertiären Extremismusprävention allgemein wird von Johansson et al. 2022 (im Erscheinen) detaillierter untersucht.

Wirkbegriffs ableitet, räumt diesem Aspekt und den Bemühungen, diese Perspektiven einzufangen, nichtsdestotrotz eine große Wichtigkeit ein.

Über das *API* hinaus ist die zentrale Personengruppe für die adäquate Einschätzung der Arbeit die der Klient:innen, sowohl derjenigen, die sich noch im Programm befinden, als auch derjenigen, die es bereits abgeschlossen oder abgebrochen haben. Soll die Wirksamkeit von Maßnahmen, die Veränderungsprozesse bei Klient:innen in der Tertiärprävention anstreben, beurteilt werden, so kommt man nicht gänzlich ohne Einsichten von ebenjenen Personen aus. Der Einbezug ihrer Perspektive ist außerdem aus ethischer und erkenntnisorientierter Sicht sehr wichtig und war in der vorliegenden Evaluationsplanung von vornherein eingeplant. Wenn sinnvoll und machbar, sollten die Perspektiven der Klient:innen weiter ergänzt werden durch die Perspektiven von Personen aus ihrem persönlichen Umfeld. Im vorliegenden Fall sollte das Bild der Ausstiegserfahrung aus dem Zusammenspiel zwischen den Einblicken von den Klient:innen, ihren Begleiter:innen und ihren Angehörigen entstehen. Dabei boten sich zur Erhebung von Erfahrungswerten, Wissen und Einschätzungen verschiedene Befragungsmethoden an. In der Diskussion um ihre Eignung wurden die folgenden drei Formen als grundsätzlich geeignet festgehalten:

1. *Die mündliche Befragung* anhand teil-standardisierter Leitfadeninterviews bietet die Möglichkeit der intensiven und tiefen Erkundung des Wissens und der Blickpunkte der befragten Personen. Sie ist aber mit einem relativ hohen Zeitaufwand verbunden und erfordert aufwändige und standardisierte Wege der Auswertung (Alemann/Forn dran 2002: 170; Diaz-Bone/Weischer 2015: 243).
2. *Schriftliche Befragungen* anhand standardisierter Fragebögen bieten die Möglichkeit der ökonomischen, möglicherweise anonymisierten Befragung mehrerer Personen gleichzeitig, womit ein vergleichsweise niedriger Zeitaufwand verbunden ist. Diese Befragungsform geht aber mit einer stärkeren Komplexitätsreduktion einher, weil überwiegend standardisierte Wege der Auswertung zur Verfügung stehen (Alemann/Forn dran 2002: 171–172; Diaz-Bone/Weischer 2015: 144).
3. *Fokusgruppengespräche* bieten die Möglichkeit, mehrere Perspektiven zusammenzubringen und Argumente gemeinsam abzuwägen. So entsteht ein Gruppenergebnis aus mehreren *Inputs* von Individuen, bei dem die unterschiedlichen Positionen aber nicht mehr sichtbar sind. Zudem sind Fokusgruppengespräche mit höherem Zeit- und Organisationsaufwand verbunden (Flick 2017: 222–224; Diaz-Bone/Weischer 2015: 140, 167).

Es bleibt den späteren Evaluierenden überlassen, weitere Methoden in das Konzept mit einzubauen oder die besprochenen (situationsbedingt) anzupassen.

Ein nächster Schritt der Planung war die Paarung von geeigneten Erhebungsmethoden mit den Befragungsgruppen: Für welche Gruppen eignen sich welche Formen der Befragung in welchen Fällen? Für die Hinzunahme der Einblicke der Ausstiegsbegleiter:innen wurde die mündliche Befragung durch semi-standardisierte Interviews beschlossen, die eine konkrete Einbindung ihrer Erfahrungen und kontextualisierende Fragen ermöglicht. Für die Befragung der Programmkoordinator:innen und der Programmleitung wurde eine Kombination aus mündlicher und schriftlicher

Befragung anvisiert, die den Hierarchiebeziehungen und den zeitlichen Engpässen einer Landesbehörde gerecht wird. Bei den Kooperationspartner:innen sollten mündliche oder schriftliche Befragungen vorgeschlagen werden, sodass die Partner:innen selbst entscheiden können, was mit ihren Ressourcen und Interessen am besten vereinbar ist. Die Befragung von Klient:innen sollte in mündlicher Form stattfinden. Es sind weiterhin Fokusgruppengespräche denkbar, die gewisse Zielgruppen (z.B. Netzwerkpartner:innen) zusammenbringen.

Als entschieden war, dass Befragungen stattfinden sollten – und insbesondere Befragungen mit Klient:innen – stellte sich die Frage nach der Auswahl geeigneter Fälle für die Erhebung. Hierbei können sich Fallstricke ergeben, wie etwa Selektionseffekte bei der (Vor-)Auswahl von Klient:innen für Befragungen. Wird diese Auswahl nicht zufällig getroffen, können (unbewusst) besonders gut oder besonders schlecht laufende Fälle ausgewählt werden, sodass die Schätzung der Wirkung zu gering oder zu hoch ausfällt. Dennoch gibt es wichtige Gründe für die nicht-zufällige Auswahl von Klient:innen für Befragungen. Dazu zählt insbesondere, dass der Beratungsverlauf und der (psychische) Gesundheitszustand der Klient:innen (etwa durch Retraumatisierungen) nicht durch die Evaluation gefährdet werden dürfen, was bei einigen wahrscheinlicher sein kann als bei anderen. Um eine mögliche Stigmatisierung der Klient:innen im Rahmen der Erhebung zu vermeiden, sollten sich die Evaluierenden vorab mit den Begleiter:innen abstimmen. Beides betrifft sowohl Klient:innen, die aktiv am Programm teilnehmen, als auch solche, die es bereits abgeschlossen haben. An dieser Stelle ist es besonders wichtig, auf die Einschätzung der Ausstiegsbegleiter:innen zu vertrauen und Klient:innen nur dann einzubeziehen, wenn diese eine Interviewsituation bewältigen können und dadurch nicht negativ beeinträchtigt werden.

Ein generelles Ziel der Fallauswahl ist die Repräsentativität der Stichprobe für die Grundgesamtheit aller Klient:innen des *API* mit Blick auf erfolgreiche Entwicklung, Dauer der Begleitung, Grundvoraussetzungen für den Beginn der Begleitung, Alter, Geschlecht, Personenstand, sowie Bildungs- und sozioökonomischem Status. In der Realität sind dieser Repräsentativität Grenzen gesetzt. Diese ergeben sich etwa aus der Bereitschaft der Klient:innen zu ihren Erfahrungen mit dem Programm befragt zu werden, aber auch durch die oben genannte Empfehlung der Begleiter:innen für geeignete Fälle. Dennoch kann und sollte Wert auf die Auswahl diverser Klient:innen-Fallverläufe gelegt werden: Nicht nur „Musterfälle“ sollten berücksichtigt werden und auch länger zurückliegende Begleitungsprozesse sollten in die Auswahl eingehen. Anfragen zum Gespräch sollten sich an mehr als die Mindestanzahl der erwünschten Stichprobe richten, falls mehrere Personen nicht gewillt sind, an der Evaluation teilzunehmen. Auch die Abbrüche des Programms sollten berücksichtigt werden. Über diese kann man verfügbare Informationen anhand der Falldokumentation sammeln, die beteiligten Stellen befragen und bei verfügbaren Kontaktdaten versuchen, auch Abbrecher:innen zu interviewen. Ist volle Repräsentativität nicht zu erreichen – wie in den meisten Evaluationen – sind entsprechend die Aussagen in der Ergebnisdarlegung als nicht-repräsentativ zu kontextualisieren. Trotz Einschränkungen bezüglich der Generalisierbarkeit trägt auch eine nicht-repräsentative Befragung zur Annäherung an die Realität bei, wenn eine ausreichende Triangulation von Daten und Methoden sichergestellt werden kann.

Im vorliegenden Fall bestand Konsens darüber, dass die Wirkungsfrage nicht anhand von Musterfällen, sondern durch die Analyse tatsächlicher Fallverläufe und die Befragung von (ehemaligen) Klient:innen untersucht werden sollte. Eine Teilnahme der Klient:innen sollte in jedem Fall freiwillig erfolgen, nach vorheriger Anfrage der Evaluierenden. Seitens der Ausstiegsbegleiter:innen würden jedoch im Vorfeld jene Fälle ausgeschlossen, bei denen negative Konsequenzen auf den Beratungsverlauf durch eine Interviewsituation zu befürchten wären. Dabei wurde betont, dass die späteren Evaluierenden in Kooperation mit der AG Evaluation und der Evaluationsbeauftragten entsprechend detaillierte Informationen über Ziele und Hintergründe der Erhebungsinstrumente an die Ausstiegsbegleiter:innen weitergeben müssten. Das würde es ihnen ermöglichen, die Klient:innen, die laufend am Programm teilnehmen, entsprechend informiert auszuwählen und auf das Gespräch vorzubereiten. Durch das vertrauensvolle Verhältnis zwischen Klient:in und Ausstiegsbegleiter:in kann die Frage, ob Bereitschaft zu einem Interview besteht, im geschützten Raum gestellt und das Vorgehen genauer erläutert werden. So können vertrauensvolle Zugänge zu möglichen Interviewpartner:innen im Vorfeld einer Befragung aufgebaut werden.

Bezüglich der Befragung von Personengruppen kann zum Schutze der Daten innerhalb einer Evaluation variiert werden, wer die Befragungen durchführt. Obwohl es in den meisten Fällen am sinnvollsten ist, wenn die (methodisch oft erfahreneren) Evaluierenden dies tun, kann diskutiert werden, ob etwa Klient:innen besser von den Begleiter:innen befragt werden sollten. Ob Ausstiegsbegleiter:innen ihre eigenen Klient:innen befragen sollten, ist eine strittige Frage. Hier ist zum einen möglich, dass die Begleiter:innen einen *Bias*, also eine Verzerrung durch ihre Perspektivierung oder Übersetzungsleistung einführen. Aufgrund des bekannten Effekts der sozialen Erwünschtheit (Diekmann 2013: 447–449) kann es außerdem sein, dass Klient:innen weniger ehrlich oder weniger kritisch sind, um aus ihrer Sicht auch die Beziehung zur Begleitperson zu schützen. Andersherum können sie aufgrund ihrer nahen Beziehung besonders kritisch sein. Andererseits könnte diese Konstellation zu einer höheren Bereitschaft der Klient:innen zur Teilnahme führen, weil sie subjektiv eine höhere Vertraulichkeit spüren. Auch die Gefahr der Stigmatisierung ist bei diesen, mit dem Fall sensibilisierten Fachkräften geringer. Die Gefahr für die Beziehung kann unter Umständen vermindert werden, weil die Handlungen der Ausstiegsbegleiter:innen nicht als von außen in Frage gestellt wahrgenommen werden. Dieser Weg sollte aber nur dann gewählt werden, wenn ansonsten die Bereitstellung der Perspektive der Klient:innen gar nicht möglich wäre.

Eine Befragung durch die Evaluator:innen wurde im Falle des *API* der Befragung durch andere, den Klient:innen bis dahin unbekanntes Ausstiegsbegleiter:innen zum Schutz der begleitenden Personen vorgezogen, auch wenn Begleiter:innen ggf. leichteren Zugang zur Zielgruppe finden könnten. Einen solchen Einsatz von „Praxisforschenden“ schlägt u.a. das Zentrum für angewandte Derradikalisierungsforschung (modus | zad; Baaken et al. 2020) als mögliche Lösungsstrategie vor, um Klient:innen möglichst sensibel befragen zu können. Diese Möglichkeit wurde mit den Programmmitarbeitenden des *API* diskutiert.

Zur Umsetzung der Erhebungsmethoden wurde unter anderem besprochen, dass sämtliche Befragungsformen, wie bei Befragungen üblich, in Sprache und Anschauungsbeispielen an die Lebens-

welt der Zielgruppe angepasst sein sollten. Eine Befragung sollte demnach basierend auf den geführten Diskussionen unbedingt so gestaltet werden, dass:

1. die Bereitschaft zur Teilnahme jederzeit gesichert und entsprechend auch widerrufbar ist;
2. eine möglichst breite Streuung der Anfrage in der betreffenden Personengruppe angestrebt wird (bei Klient:innen wenn möglich auch bei abgeschlossenen Fällen), um die Langfristigkeit der Wirkungsbeobachtung zu optimieren. Hier sollten Evaluierende eine Flexibilität bezüglich der Erhebungswahl zeigen: Eine mündliche Befragung wäre grundsätzlich vorzuziehen, eine schriftliche kann aber in Ausnahmefällen, vor allem bei abgeschlossenen Fällen, die einzig umsetzbare sein;
3. keine besonderen Anforderungen an die Fähigkeiten der Befragten zum Verstehen oder Beantworten der Fragen gestellt werden. Hier gilt, wie bei jeder wissenschaftlichen Erhebung, dass Befragungen keine unnötigen Hürden erzeugen sollten, die an der Lebenswelt der Zielgruppe vorbeigehen und die Ergebnisse verfälschen würden.

Unter realen Bedingungen werden auch im vorliegenden Fall eine oder mehrere Verzerrungen wahrscheinlich dennoch auftreten und keine volle Repräsentativität der Fallgesamtheit zustande kommen. Diese Verzerrungen können aber mit den genannten Regelungen auf pragmatischem Wege verringert werden. Eine wirkungsorientierte Evaluation sollte die sich andeutenden oder sicher aufgetretenen Verzerrungen in jedem Fall benennen und bei der Ergebnispräsentation reflektieren.

3.7 PLANUNG DES ERGEBNISTRANSFERS

Evaluation ist kein Selbstzweck und sollte in den meisten Fällen nicht nur der Rechtfertigung, sondern auch der Weiterentwicklung der evaluierten Praxis dienen. Soll es in den Evaluationsergebnissen nicht bei theoretischen Empfehlungen bleiben, die letztlich kaum oder keine Anwendung in der Praxis finden, muss der Erkenntnistransfer im Rahmen der Evaluation bereits mitgedacht werden. Wie können die Ergebnisse der Evaluation konkreten Eingang in die praktische Arbeit des Trägers finden? Wie müssen sie formuliert und transportiert werden, damit sie von der Praxis aufgenommen werden (können)? Der Umgang eines Projekts oder eines Programms mit den Ausführungen und Erkenntnissen der Evaluation bezeichnet man auch als *Uptake*. Dieser kann informell und intern erfolgen oder formalisiert als *Management Response* veröffentlicht werden. Im Zuge dessen können einzelne Ergebnisse hinterfragt oder abgelehnt, vollständig bestätigt und angenommen, die Umsetzung der Empfehlungen geprüft und auf den Weg gebracht und mit konkreten Umsetzungsideen ausgestattet werden.

Die Anknüpfungspunkte der Evaluation an die konzeptionelle Ausgestaltung und den Arbeitsalltag des API sind vielfältig. Dementsprechend wurde es in der vorliegenden Vorbereitung der Evaluation fest eingeplant, dass die Ergebnisse und die Handlungsempfehlungen im Rahmen von Workshops in das Programm rückgeführt werden. Ganz praktisch bedeutet dies, dass die Ergebnisse der Evaluation mit allen Beteiligten des API und im Besonderen mit den Ausstiegsbegleiter:innen analysiert und in konkrete Handlungen übersetzt werden sollen. Demnach werden die Ergebnisse mit allen Mitarbeitenden

besprochen und interpretiert. An dieser Stelle ist die Einbindung verschiedener Expertisen vorteilhaft, da die Handlungsempfehlungen anhand unterschiedlicher Sichtweisen eingeordnet und somit konkret Einzug in die Arbeitspraxis halten können. Ein festgelegter Zeitplan kann an dieser Stelle helfen, die Arbeitsschritte fest zu terminieren, um somit einen Zeithorizont für die vorgeschlagenen Veränderungen zu ermöglichen. Die Verantwortung für die Umsetzung teilen sich dabei die Evaluationsbeauftragte, die AG Evaluation und die für das Programm zuständigen Referent:innen.

4. ÜBERTRAGBARKEIT DES VORGEHENS AUF ANDERE PROGRAMME

Bezogen auf seine finanzielle und personelle Ausstattung ist das *API* als vergleichsweise privilegiert einzustufen. Anders als eine Vielzahl von Projekten im Bereich der Extremismusprävention erhält es keine zeitlich limitierte (Modell-)Projektförderung, sondern ist institutionell beim Verfassungsschutz des Landes Nordrhein-Westfalen verankert; gewisse Evaluationskapazitäten sind regulär im Programm vorhanden. Daraus ergeben sich zahlreiche Faktoren, die förderlich für die Planung und Umsetzung von Evaluationen im Themenfeld sein können. Die mit klaren Kompetenzen, Weiterbildungsmöglichkeiten und entsprechender Verantwortung ausgestattete Person der Evaluationsbeauftragten ist eine davon.

So zeichnet sich das Programm einerseits durch für ein Aussteigerprogramm vergleichsweise hohe Begleitzahlen aus, andererseits ermöglicht es die institutionelle Anbindung, die Klient:innen über mehrere Jahre hinweg zu begleiten. Die Falldokumentation erfolgt hierbei individuell durch die Ausstiegsbegleiter:innen, basiert jedoch auf für das Programm erarbeiteten Monitoringinstrumenten und Dokumentationsvorlagen. Hierdurch verfügt das *API* über eine gut dokumentierte Vielzahl von Begleitungsverläufen, von denen eine wirkungsorientierte Evaluation sehr profitieren kann. Die Falldokumentation informiert bei erfolgreichen Beratungsprozessen bereits gut über den Fallverlauf und kann durch Interviews mit den jeweiligen Klient:innen und Berater:innen ergänzt werden. Bei seitens der Klient:innen abgebrochenen Beratungsprozessen ist sie eine umso wichtigere Informationsquelle. Hier sind die ausgeschiedenen Klient:innen kaum noch erreichbar, die Falldokumentation kann als Zeugnis für mögliche Abbruchgründe herangezogen werden. Die Nutzung eines strukturierten Dokumentationssystems bietet für Akteure der Extremismusprävention viele Vorteile und begegnet den Berichtspflichten gegenüber Mittelgebenden und der (Fach-)Öffentlichkeit sowie internen Qualitätssicherungsansprüchen. Für Evaluationsvorhaben sind systematische Falldokumentationen eine wertvolle Informationsquelle, die bei Bedarf anonymisiert werden kann und gleichzeitig Analysekatgorien mitbringt, die vom Präventionsakteur selbst festgelegt wurden.

Einen weiteren Vorteil für die Evaluationsbestrebungen bieten auch die über Jahre hinweg aufgebauten und etablierten Kooperations- und Netzwerkstrukturen des Programms. Sie ermöglichen es, auch externe Akteure, die in anderen Zusammenhängen mit den *API*-Klient:innen arbeiten (z.B. die Bewährungshilfe, soziale Unterstützungsstellen) in Befragungen einzubeziehen und so eine multiperspektivische Einschätzung der (Aus-)Wirkungen des Programms auf die Klient:innen zu gewinnen. Zusätzlich hat das Programm die Möglichkeit, in den Austausch mit wissenschaftlichen Akteuren zu treten, um einerseits neue wissenschaftliche Erkenntnisse in der eigenen Arbeit zu nutzen und ande-

rerseits das eigene Erfahrungswissen (unter Berücksichtigung der Datenschutzvorgaben) in wissenschaftliche Entwicklungen einfließen zu lassen.

Trotz dieser strukturellen Stärken kostet eine gute Evaluation Zeit und Personalressourcen. Dies gilt nicht nur in der Phase der Datenerhebung, sondern bereits in der Evaluationsvorbereitung und für das anschließende *Uptake*. Dabei sollten die benötigten Kapazitäten möglichst nicht zu Lasten der eigentlichen Programmaktivitäten gehen. Die Möglichkeit, Zeit und Personalressourcen in die intensive Evaluationsvorbereitung investieren zu können, ist abhängig von den strukturellen Rahmenbedingungen, innerhalb derer die Projekte agieren. So vielfältig wie die Landschaft in der Tertiärprävention ist, so vielfältig kann der konstruktive Umgang mit dieser Herausforderung sein, sodass auch kleinere und weniger stetig finanzierte Präventionsakteure Wege finden können, um das Thema Evaluation aktiv für ihre Entwicklung zu nutzen. Gesteigerte Kapazitäten in diesem Bereich werden durch eine zunehmende Professionalisierung des Feldes und wachsende Berichtspflichten ohnehin immer drängender.

Anstatt etwa der Rolle einer:ines Evaluationsbeauftragten, wie beim *API*, kommen auch kleinformige Lösungen in Frage. So kann punktuell externe Evaluationsexpertise „eingekauft“ werden, um die Träger dabei zu unterstützen, das interne Qualitätssicherungssystem aufzubauen, geplante Evaluationen vorzubereiten oder den Transfer von Evaluationsergebnissen in die Praxis einzuleiten. Die gemeinsame Operationalisierung von Zielstellungen und Indikatoren ebenso wie die Unterstützung bei der Erarbeitung von Dokumentationsinstrumenten mit externen Beratenden ebnet den Weg für die systematische Auswertung gesammelter Daten durch die Präventionsakteure selbst. Somit können sie auch langfristig ihre internen Evaluationskapazitäten⁸ stärken. Hier gibt es bisweilen kostenlose Beratungsoptionen, die in den Förderprogrammen mit angelegt sind (wie z.B. Beratungsmöglichkeiten durch die *GesBiT Bildung & Teilhabe GmbH* im Rahmen des Bundesprogramms *Demokratie leben!*). Solche Angebote stärker und breiter in der Förderungslandschaft, also der Praxis des Finanzierens von Prävention und Evaluation, zu verankern, wäre für die Evaluationspraxis sehr förderlich.

Ist eine externe Fremdevaluation vorgesehen, kann die fachliche Beratung, wie im vorliegenden Fall, bei der Vorbereitung der Evaluationsausschreibung unterstützen, damit diese letztlich die vom Träger gewünschte Schwerpunktsetzung enthält, bereits abgestimmte und passgenaue Fragestellungen beinhaltet und geeignete Datenquellen offenlegt. Die Expertise der ausgewählten Evaluierenden oder Beratenden sollte dabei hinreichend ausgeschöpft werden. Die Hinzuziehung von Programmverantwortlichen und Programmtätigen bei der späteren (Zwischen-)Interpretation der Ergebnisse sowie die Diskussion der Formulierung praxisnaher und umsetzbarer Empfehlungen (z.B. in der Form von Workshops) kann, wie im vorliegenden Fall, auch bereits im Rahmen der Ausschreibung gefordert und Gegenstand der Leistungsbeschreibung sein. Damit kann vermieden werden, was häufig

8 Interne Evaluation meint, dass Akteure selbst zu ihrer eigenen Institution (aber nicht unbedingt ihrer höchsteigenen Tätigkeit) evaluativ tätig werden; externe Evaluationen dagegen werden von externen Expert:innen durchgeführt. Verwandt ist der Begriff der Selbst- und Fremdevaluation. Selbstevaluation bedeutet, dass die Evaluierenden ihre eigene Arbeit untersuchen, während Fremdevaluation heißt, dass Personen die konkrete Arbeit anderer evaluieren. Es existieren auch, wie im vorliegenden Fall für die Vorbereitung genutzt, Mischformen, bei denen interne und externe Akteure für die Evaluation zusammenarbeiten (z.B. partizipative Evaluationen).

später passiert: Die Ergebnisse werden schriftlich übergeben, ihre Interpretation und Umsetzung in die Praxis bleibt unklar und unter alleiniger Verantwortung der Evaluierten, die sich parallel jedoch ihrem Tagesgeschäft widmen müssen. Wenn dies der Fall ist, können Evaluationsergebnisse leicht im Sande verlaufen.

Steht eine konkrete Evaluation an, kann wiederum das Hinzuziehen weiterer externer Expertise hilfreich sein, um ein *Management Response*-System für die anstehenden Evaluationsergebnisse und -empfehlungen aufzusetzen. Denn nach der Vorlage des Evaluationsberichtes fängt, wie oben angedeutet, erst die praktische Auseinandersetzung mit den Ergebnissen an: Projekte und Programme können dann intern diskutieren, welche Empfehlungen in ihren Einflussbereich fallen, welche sie grundsätzlich akzeptieren, welchen sie anhand eigener Daten und Erkenntnissen widersprechen möchten und wie sie die angenommenen Empfehlungen umsetzen können und wollen. Dafür ist ein formalisiertes Verfahren sehr förderlich. Wenn dieses einmal etabliert ist, lässt es sich immer wieder anwenden, ohne, dass weitere externe Unterstützung notwendig ist – die Erstellung eines solchen Systems ist somit eine infrastrukturelle Investition.

Darüber hinaus kommt die Gründung einer Evaluations-AG, die verschiedene, von einer Evaluation potenziell „betroffene“ Akteursgruppen an einen Tisch bringt, auch innerhalb kleinerer Träger in Betracht. Hierbei kann einer Überlastung einzelner Mitarbeitenden durch in regelmäßigen Zeitabständen wechselnde Belegung der Arbeitsgemeinschaft begegnet werden. Eine Evaluations-AG hat zudem den Vorteil, dass sie eine kontinuierliche Austauschplattform zu Evaluationsbelangen innerhalb des Trägers sein kann und dabei verschiedene Akteursgruppen gleich miteinbezieht. Erkenntnisinteressen, die zu jeglichem Zeitpunkt im Arbeitsalltag aufkommen, können dadurch direkt festgehalten werden für den nächsten größeren Evaluationszeitpunkt.

Bei größeren Praxisträgern, die mehrere Projekte oder Programme unter ihrem Dach vereinen, können Evaluationsbeauftragte über verschiedene Programme hinweg mit Stundenanteilen querfinanziert werden. Hierbei stehen auch die Mittelgebenden in der Pflicht: Mit den gängigen Finanzierungsmöglichkeiten (etwa: Verwaltungsgemeinkosten) ist den Qualitätssicherungsansprüchen auf der einen und den Informationsinteressen auf der anderen Seite kaum noch zu begegnen. Hier könnte eine deutliche Erhöhung der Gemeinkostenanteile Abhilfe schaffen. Das Ausloben zusätzlicher Mittel – im Speziellen für Monitoring und Evaluation – bietet für Projektträger jedoch noch stärkere Anreize, interne Qualitätssicherungs- und Evaluationsstrukturen zu professionalisieren, ohne dass dies zu Lasten anderer, ebenfalls unverzichtbarer Positionen, geht.

5. FAZIT UND EMPFEHLUNGEN

Wirkungsorientierte Evaluationen im Feld der (insbesondere tertiären) Extremismusprävention stellen Evaluierende wie Präventionsakteure vor Herausforderungen. Bei der Vorbereitung eines Evaluationsvorhabens gilt es zu entscheiden, wie Wirkung verstanden und sich ihr angenähert werden soll, mit welchen wissenschaftlichen Methoden sie erfasst werden kann und welche Datenzugänge genutzt werden sollen. Dabei sind die unterschiedlich gelagerten Interessen, Erfahrungen und Prioritä-

ten aller Beteiligten zu berücksichtigen. *Auftraggebende* sollten also Schwerpunkte (Fragestellungen, thematische Programmschwerpunkte) der evaluatorischen Betrachtung im Vorfeld intern diskutieren, definieren und auflisten, um den Evaluationsprozess zu erleichtern.

Die Erfahrungen mit dem vorliegenden Vorgehen zeigen, wie (zeit-)intensiv und multiperspektivisch die Vorbereitung einer Evaluation durch die und mit den potenziell *Evaluierten* gestaltet werden kann und sollte. Auf diesem Weg kann internes Erwartungsmanagement zur Evaluation betrieben werden, die Identifikation und der Praxisbezug der Evaluation erhöht und die Vorbereitung benötigter Daten frühzeitig eingeleitet werden. Da die Möglichkeiten von Präventionsakteuren, diesbezüglich selbst aktiv zu werden, von ihrer strukturellen Aufhängung abhängig sind, ergeben sich nicht nur Empfehlungen für sie und für die Evaluierenden, sondern auch explizit für die *Institutionen*, in denen sie verankert sind, sowie für die *Mittelgebenden* (die beide jeweils Auftraggebende für Evaluationen sein können).

In der Entscheidungskette, die zu Evaluationsaufträgen führt, sollten spätestens die Auftraggebenden ein Bewusstsein dafür entwickeln, dass die Erwartungen an eine Evaluation vor der Ausschreibung bei den Evaluierten festgehalten werden müssen, weil sich daraus konkrete methodologische Herangehensweisen ergeben. Auch sollten die Schwerpunkte der Betrachtung (leitende Fragestellungen, Programmschwerpunkte) von vornherein entschieden und aufgelistet werden. Evaluierte sollten Evaluationserfahrungen und -bedarfe sowie praxisrelevante Erkenntnisinteressen im Vorfeld eines Evaluationsvorhabens intern diskutieren und ausformulieren, um sich aktiv in den Prozess der Evaluationsgestaltung einbringen zu können. Die Güteanalyse vorhandener Indikatoren und ihre formative Weiterentwicklung sollten als Gegenstand der Evaluation mitaufgenommen werden, da dieser Schritt, wie im vorliegenden Prozess gezeigt, intensive Reflexionen über das Programmmonitoring und die langfristige Informationsaufbereitung eröffnen kann. Nutzbare Datengrundlagen sollten vor der Ausschreibung und mit sämtlichen Präventionsakteuren diskutiert werden, um die operative Vorbereitung der Evaluation frühestmöglich einleiten zu können. Die Einteilung dieser Grundlagen in Kategorien (offen zugängliche Daten; blanko/anonymisierte Daten; externe Quellen, deren Verfügbarkeit es zu klären gilt; Daten, die es zu erheben gilt) hat sich im vorliegenden Fall als vorteilhaft erwiesen. Bei den Daten, die es zu erheben gilt, sollten die Erhebungsmethoden an die Zielgruppen angepasst und die konkrete Umsetzung der Erhebungen (inkl. der Auswahl der Erhebungspartner:innen und der Entscheidung, wer die Erhebung durchführt) diskutiert werden. Auftraggebende, aber auch Evaluierende, sollten zusätzlich beachten, dass die Befragung von Klient:innen in der Tertiärprävention ein Sonderfall ist, dem es mit viel Sensibilität zu begegnen gilt, um Stigmatisierung und Gefährdung der Beziehungsarbeit einerseits zu verhindern und Verzerrung der Ergebnisse andererseits zu minimieren.

Schließlich sollten diese Entscheidungen und Vorbereitungen, wie in diesem Fall geschehen, in eine ausführliche und inklusive Formulierung der Evaluationsausschreibung einfließen, die diese Vorbereitung repräsentiert. Sofern die Zeiträume, die in Ausschreibungen für Evaluationen eingeräumt werden, groß genug sind, können einige dieser Schritte auch in die Arbeit mit den (externen) Evaluierenden verlegt werden. Es sei jedoch stark dazu geraten, nicht gänzlich auf die interne Vorbereitung zu verzichten, da die Präsenz von denjenigen, die ein Programm dem Ersteindruck nach

„überprüfen“ sollen, die offene Diskussion beeinträchtigt. Andersherum kann eine Vorbereitung ohne die Evaluierenden die Kennenlern- und Vertrauensaufbauphase zwischen ihnen und den Evaluierten nicht ersetzen.

Der Transfer der Evaluationsergebnisse in die Praxis sollte bereits bei der Konzipierung der Evaluation und ihrer Ausschreibung von den Auftraggebenden mitbedacht und für die zukünftigen Evaluierenden vorgegeben werden. Folgende Fragen sind hierbei vorab zu klären: Welche Rollen sollen externe Evaluierende, welche die internen Evaluationsbegleitenden spielen? Wie wird der Transfer von Zwischen- und Endergebnissen personell (und dadurch auch finanziell) und zeitlich gestaltet und ausgestattet? Wie können gute Bedingungen für ein gelungenes *Uptake* geschaffen werden? Die Übermittlung von Zwischen- und Endergebnissen und ihre inklusive Interpretation (ggf. durch Workshops und/oder Fortbildungen) muss zeitlich und personell (und dadurch auch finanziell) angemessen ausgestattet sein, damit die Erkenntnisse der Evaluation erfolgreich in die Lern- und Wissensmanagementstruktur der Projekte und Programme der Extremismusprävention eingehen können.

Der Finanzierungsaspekt betrifft dabei auch die Mittel über die Evaluation hinaus. Die stetige Weiterentwicklung und Betreuung der Monitoring- bzw. Dokumentationsstruktur eines Programms und die Gestaltung, Begleitung und Abnahme von Evaluationen inklusive eines *Management Response-Systems* müssen dort personell gedeckt sein und können nicht rein über Zusatzbelastung der regulären Mitarbeitenden aufgefangen werden. Dafür sind flexible Möglichkeiten zu schaffen, wie etwa die Option einer externen Beauftragung, die Position einer Person, die für Evaluation und Wissensmanagement zuständig ist, eine AG Evaluation oder eine finanzielle Monitoring-, Evaluations- oder Lernpauschale. Es ist eine Aufgabe des evaluierten Programms, geeignete Qualitätssicherungsstrukturen zu eruieren und vorzuschlagen, um so im Dialog mit den Zuwendungsgebenden internes Evaluationswissen und -kapazitäten stärken zu können. Das ermöglicht es den Programmen, auf die Evaluationsergebnisse sinnvoll zu reagieren und entsprechende (Prozess-)Anpassungen vorzunehmen.

Mithilfe gestärkter Strukturen und einem Verständnis von Evaluation als Möglichkeit zur Weiterentwicklung der Präventionsarbeit, die nur greifen kann, wenn diejenigen, die die Prävention durchführen, sich mit ihr identifizieren, kann den Herausforderungen von Wirkungsüberprüfungen auf produktivem Wege begegnet werden.

- Abay Gaspar, Hande/Daase, Christopher/Deitelhoff, Nicole/Junk, Julian/Sold, Manjana 2018: Was ist Radikalisierung? Präzisierungen eines umstrittenen Begriffs, PRIF Report 5/2018, Frankfurt am Main.
- Alemann, Ulrich v./Forndran, Erhard 2002: Methodik der Politikwissenschaft. Eine Einführung in Arbeitstechnik und Forschungspraxis (Kohlhammer-Urban-Taschenbücher, Band 204), Stuttgart.
- Armborst, Andreas 2019: Evidenzbasierte Prävention von Extremismus und Radikalisierung. Leerstellen und Handlungsbedarf (Serie: Evaluation, Infodienst Radikalisierungsprävention).
- Baaken, Till/Becker, Reiner/Bjørge, Tore/Kiefer, Michael/Korn, Judy/Mücke, Thomas/Ruf, Maximilian/Walkenhorst, Dennis 2018: Herausforderung Deradikalisierung. Einsichten aus Wissenschaft und Praxis, PRIF Report 9/2018, Frankfurt am Main.
- Baaken, Till/Ruf, Maximilian/Selby, Anne/Walkenhorst, Dennis 2020: „Blinde Flecken“ in der Distanzierungsbeobachtung? Erste Zwischenergebnisse des Projektes DISLEX 3D.
- BAG ReLEX 2021: Wissen – Radikalisierungsprävention, <https://www.bag-relex.de/wissen/radikalisierungspraevention/> (letzter Zugriff: 21.10.2021).
- Bernasconi, Tobias 2009: Triangulation in der empirischen Sozialforschung am Beispiel einer Studie zu Auswirkungen und Voraussetzungen des barrierefreien Internets für Menschen mit geistiger Behinderung, in: Empirische Sonderpädagogik, 1: 1, 96–109.
- DeGEval 2021: Glossar der Standards für Evaluation, <https://www.degeval.org/glossar-der-standards-fuer-evaluation/> (letzter Zugriff: 17.8.2021).
- Diaz-Bone, Rainer/Weischer, Christoph (Hrsg.) 2015: Methoden-Lexikon für die Sozialwissenschaften, Wiesbaden.
- Diekmann, Andreas 2006: Aktuelle Probleme der empirischen Sozialforschung, in: Diekmann, Andreas (Hrsg.): Kölner Zeitschrift für Soziologie und Sozialpsychologie, Opladen, 8–32.
- Diekmann, Andreas 2013: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen, Rororo, Rowohlt's Enzyklopädie, Band 55678, Reinbek bei Hamburg.
- Flick, Uwe 2011: Triangulation. Eine Einführung. Wiesbaden.
- Flick, Uwe 2017: Qualitative Sozialforschung. Eine Einführung, Rororo, Rowohlt's Enzyklopädie, Band 55694, Reinbek bei Hamburg.
- Greuel, Frank 2020: Zum Konzept der Prävention: Ein Plädoyer für engere Grenzen, <https://www.bpb.de/politik/extremismus/radikalisierungspraevention/311923/zum-konzept-der-praevention-einplaedoyer-fuer-engere-grenzen> (letzter Zugriff: 21.10.2021).
- Gruber, Florian/Lützing, Saskia 2017: Extremismusprävention in Deutschland – Erhebung und Darstellung der Präventionslandschaft. Modulabschlussbericht, Wiesbaden.
- Johansson, Susanne/Junk, Julian/Liebich, Johanna/Walkenhorst, Dennis 2022: Klientenzentrierte Evaluation in Multi-Agency Settings der Extremismusprävention – Möglichkeiten und Grenzen von wirkungsorientiertem Vorgehen, PRIF Report im Erscheinen, Frankfurt am Main.
- Koynova, Svetla/Mönig, Alina/Quent, Matthias/Ohlenforst, Vivienne 2022: Monitoring, Evaluation und Lernen: Erfahrungen und Bedarfe der Fachpraxis in der Prävention von Rechtsextremismus und Islamismus, PRIF Report im Erscheinen, Frankfurt am Main.

modus | zad: <https://modus-zad.de/> (letzter Zugriff: 15.11.2021).

Möller, Kurt 2019: Evaluation neu denken. Der Dritte Raum (Serie: Evaluation, Infodienst Radikalisierungsprävention).

Phineo 2016: Das Kursbuch Wirkung, <https://www.phineo.org/kursbuch-wirkung> (letzter Zugriff: 25.10.2021).

Schmucker, Martin 2007: Meta-Analysen zur Sexualstraftäterbehandlung, in: Berner, W./Briken, P./Hill, A. (Hrsg.): Straftäter behandeln mit Psychotherapie und Medikamenten, Köln, 13–30.

Stegmüller, Wolfgang 1969: Kausalitätsprobleme. Ursache und Wirkung Kausalgesetze. Kausale Modalitäten Kausale Erklärungen Das allgemeine Kausalprinzip Determinismus und Indeterminismus, in: Stegmüller, Wolfgang (Hrsg.): Wissenschaftliche Erklärung und Begründung, Berlin, Heidelberg, 428–17.

PRIF REPORT

Die PRIF Reports analysieren Hintergründe politischer Ereignisse und Entwicklungen und präsentieren wissenschaftliche Forschungsergebnisse in Deutsch oder Englisch.

Schlicht-Schmälzle, Raphaela/Theis, Désirée/Tultschinetski, Sina/Verhovnik-Heinze, Melanie/Kautz, Christian/Kirchmann, Maryam (2021): Erfolgsbedingungen und Herausforderungen für die Zusammenarbeit von Fachpraxis und Wissenschaft bei der Evaluation von Angeboten politischer Bildung, PRIF Report 5/2021, Frankfurt/M.

Baldus, Jana/Fehl, Caroline/Hach, Sascha (2021): Beyond the Ban. A Global Agenda for Nuclear Justice, PRIF Report 4/2021, Frankfurt/M.



www.hsfk.de/PRIF-Reports
www.hsfk.de/HSFK-Reports

PRIF SPOTLIGHT

Die PRIF Spotlights diskutieren aktuelle politische und gesellschaftliche Themen.

Schörnig, Niklas (2022): AI for Arms Control. How Artificial Intelligence Can Foster Verification and Support Arms Control, PRIF Spotlight 1/2022, Frankfurt/M.

Bethke, Felix (2021): Civil War in Ethiopia. The Instrumentalization and Politicization of Identity, PRIF Spotlight 16/2021, Frankfurt/M.

PRIF SPOTLIGHT

DIE RICHTIGE FRAU AM RICHTIGEN ORT

Die Bedeutung von Frauen in der Friedens- und Konfliktforschung

Am 16. und 17. März fand das Symposium 'Frauen in der Friedens- und Konfliktforschung' statt. Die Veranstaltung wurde von der PRIF und dem Institut für Friedens- und Konfliktforschung (IFK) organisiert. Die Veranstaltung war ein wichtiger Schritt in der Entwicklung von Frauen in der Friedens- und Konfliktforschung.



Die PRIF ist ein internationaler Forschungsinstitut für Friedens- und Konfliktforschung. Die PRIF ist ein internationaler Forschungsinstitut für Friedens- und Konfliktforschung. Die PRIF ist ein internationaler Forschungsinstitut für Friedens- und Konfliktforschung.

www.hsfk.de/PRIF-Spotlights

PRIF BLOG

Auf dem PRIF Blog erscheinen Beiträge zu aktuellen politischen Fragen und Debatten der Friedens- und Konfliktforschung. Die Blogbeiträge erscheinen in loser Folge in Deutsch oder Englisch.

PRIF BLOG

PRIF RESEARCH INSTITUTE FRIEDENS- UND KONFLIKTFORSCHUNG

<https://blog.prif.org/>

PRIF Reports und PRIF Spotlights sind Open-Access-Publikationen und können kostenlos auf www.hsfk.de heruntergeladen werden. Sie möchten die digitalen Ausgaben abonnieren? Bitte wenden Sie sich an: publikationen@hsfk.de.

www.facebook.com/HSFK.PRIF

www.twitter.com/HSFK_PRIF

<https://blog.prif.org/>

Gefördert durch:



Bundesministerium
des Innern, für Bau
und Heimat

aufgrund eines Beschlusses
des Deutschen Bundestages

PrEval

Evaluationsdesigns für Präventionsmaßnahmen –
Multimethodische Ansätze zur Wirkungsermittlung und
Qualitätssicherung in der Extremismusprävention sowie den
Schnittstellen zur Gewaltprävention und politischen Bildung

MONA KLÖCKNER // SVETLA KOYNOVA //
JOHANNA LIEBICH // LISA NEEF

**ERFAHRUNGEN AUS DER EVALUATIONS-
PLANUNG EINES AUSSTEIGERPROGRAMMS
VORAUSSETZUNGEN FÜR WIRKSAMKEITS-
ERFASSUNG IN DER TERTIÄREN EXTREMISMUS-
PRÄVENTION**

Das *Aussteigerprogramm Islamismus* des Landes Nordrhein-Westfalen (API) und der Forschungsverbund *PrEval* bereiten gemeinsam die Ausschreibung einer wirkungsorientierten Evaluation des API-Programms vor. Im vorliegenden Report beschreiben die Autor:innen das notwendige dialogische Vorgehen bei der Vorbereitung des Evaluationsprozesses und leiten daraus Empfehlungen für die beteiligten Evaluationsakteure ab. Diese systematische Reflexion über die Vorbereitung eines wirkungsorientierten Evaluationsdesigns bietet eine Orientierungshilfe für Auftraggebende von Evaluationen, Evaluierende, evaluierte Projekte und Programme, aber auch Zuwendungsgeber.

ISBN 978-3-946459-67-5