

Arbeitshilfe

zur Evaluierung von Regelungen der Bundesregierung



Januar 2022

Arbeitshilfe

zur Evaluierung von Regelungen der Bundesregierung

Januar 2022

Inhaltsverzeichnis

1	Einleitung.....	4
2	Festlegungen im Regelungsvorhaben	5
2.1	Ist das Vorhaben evaluierungspflichtig?	5
2.2	Darstellung im Regelungsvorhaben.....	7
2.2.1	Ausgangslage und Ziele	7
2.2.2	Bei wesentlichen Vorhaben: Aussagen zur Durchführung oder Nicht-Durchführung der Evaluierung	8
3	Planung der Evaluierung.....	9
3.1	Detaillierungsgrad	9
3.2	Interne oder externe Evaluierung	10
3.3	Ermittlung der Betroffenen	12
3.4	Entwicklung eines Wirkmodells.....	13
3.5	Wahl des Untersuchungsdesigns.....	16
3.6	Wahl der Erhebungsverfahren	18
3.7	Schematische Darstellung einer Evaluierung	21
4	Durchführung einer internen Evaluierung	22
4.1	Darstellung des Gegenstands der Evaluierung.....	22
4.2	Konzeption der Erhebungsinstrumente für Primärdaten	22
4.2.1	Erarbeitung eines Interviewleitfadens	22
4.2.2	Entwicklung eines Fragebogens	24
4.2.3	Pretest	25
4.3	Sammeln von Sekundärdaten	26
4.4	Wirkanalyse	27
4.5	Handlungsempfehlungen	28
4.6	Beteiligung und Qualitätssicherung	28
5	Durchführung einer externen Evaluierung.....	30
5.1	Erstellung einer Leistungsbeschreibung und die Vergabe der Evaluierung.....	30
5.2	Einhaltung der Vorgaben der Leistungsbeschreibung und anerkannter Qualitätsstandards	31
5.3	Stellungnahme zu Handlungsempfehlungen	31
6	Die Veröffentlichung des Evaluierungsberichts	31
7	Der Umgang mit den Ergebnissen.....	32

8	Anhang.....	33
	Anhang I Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 23. Januar 2013	33
	Anhang II Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 26. November 2019	36
9	Impressum	38

1 Einleitung

Wesentliche Regelungsvorhaben müssen evaluiert werden.

Mit der Konzeption zur Evaluierung neuer Regelungsvorhaben, die der Staatssekretärsausschuss Bessere Rechtsetzung und Bürokratieabbau am 23. Januar 2013 beschlossen hat, wurde erstmalig ein Rahmen zur Überprüfung der Wirkung von Gesetzen und Rechtsverordnungen festgelegt. Die Konzeption enthält die Pflicht zur Evaluierung von wesentlichen Gesetzen und Rechtsverordnungen. Der Staatssekretärsausschuss wertete die Erfahrungen aus den ersten Evaluierungsberichten aus und ergänzte die Konzeption durch seinen Beschluss vom 26. November 2019. Dieser Beschluss sieht auch vor, den Bundesministerien eine Arbeitshilfe zur Verfügung zu stellen, in der Arbeitsschritte und Methoden einer Evaluierung dargestellt werden. Neben den genannten Beschlüssen ist auch die **↳ Gemeinsame Geschäftsordnung der Bundesministerien (GGO)** eine für Evaluierungen wesentliche Grundlage.

Die Arbeitshilfe unterstützt Mitarbeiterinnen und Mitarbeiter der Bundesministerien bei Evaluierungen.

Die Arbeitshilfe soll Hilfestellungen für alle Arbeitsschritte der Evaluierung geben und richtet sich an Mitarbeiterinnen und Mitarbeiter in den Bundesministerien sowie deren nachgeordnete Bereiche, die – vielleicht zum ersten Mal – evaluieren. Sie baut auf einschlägigen Empfehlungen der Fachgesellschaften und internationalen Standards auf. Evaluierungen sind so vielfältig wie die Gesetze, die sie behandeln. Es ist nicht Ziel der Arbeitshilfe, alle Evaluierungsformen und Methoden umfassend abzubilden. Sie soll vielmehr den praktischen Einstieg in die Evaluierungspraxis unterstützen. Aus diesem Grund wurden im Vorfeld die Erfahrungen und Erwartungen der Nutzerinnen und Nutzer in mehreren Workshops und Besprechungen abgefragt. Auch zukünftige Erfahrungen, insbesondere mit der Anwendung dieser Arbeitshilfe, werden bei Neuauflagen berücksichtigt werden.

Diese Arbeitshilfe beschreibt die relevanten Arbeitsschritte in chronologischer Abfolge. Diese erschöpfen sich nicht in der Erstellung der Evaluierung. Eine sorgfältige Planung, die schon bei der Formulierung des zu evaluierenden Regelungsvorhabens beginnt, bildet die Grundlage für eine qualitativ hochwertige Evaluierung. Gleichmaßen wichtig ist es, die Ergebnisse bei zukünftigen Regelungsvorhaben zu berücksichtigen und einer politischen Diskussion zuzuführen.

2 Festlegungen im Regelungsvorhaben

2.1 Ist das Vorhaben evaluierungspflichtig?

Das Evaluierungskonzept von 2013 regelt verbindlich, dass wesentliche Regelungsvorhaben evaluiert werden müssen. Die Wesentlichkeit eines Regelungsvorhabens bemisst sich zum einen nach der Höhe des zu erwartenden jährlichen Erfüllungsaufwands. Zum anderen kann sich die Wesentlichkeit aus anderen Gründen ergeben. Bei allen wesentlichen Regelungsvorhaben sind in der Begründung zum Regelungsentwurf gem. § 44 Abs. 7 GGO Aussagen zur Durchführung oder Nicht-Durchführung von Evaluierungen aufzunehmen. Dies gilt unabhängig davon, ob sich die Wesentlichkeit aus dem ex-ante geschätzten Erfüllungsaufwand oder aus anderen Gründen ergibt.

Es handelt sich um ein wesentliches und damit grundsätzlich evaluierungspflichtiges Regelungsvorhaben, wenn ein jährlicher Erfüllungsaufwand von

- 1 Mio. € Sachkosten oder 100 000 Stunden Aufwand für Bürgerinnen und Bürger oder
- 1 Mio. € für die Wirtschaft oder
- 1 Mio. € für die Verwaltung

aufgrund der Ex-ante-Abschätzung zu erwarten ist oder – ist eine solche Abschätzung nicht möglich – nicht ausgeschlossen werden kann. Dabei ist es unerheblich, ob sich der Erfüllungsaufwand in den genannten Größenordnungen erhöht oder reduziert.

Anlass für eine Evaluierung kann auch die Nachmessung des Erfüllungsaufwands sein – wenn erst dann festgestellt wird, dass der tatsächliche Aufwand einen der genannten Schwellenwerte überschreitet. Ergibt die Nachmessung, dass der Schwellenwert unterschritten wird, kann das Ressort seine im Regelungsentwurf getroffenen Evaluierungserwägungen überprüfen.

Sind diese Voraussetzungen nicht erfüllt, muss geprüft werden, ob sich die Wesentlichkeit des Vorhabens nicht aus anderen Gründen ergibt. Andere Gründe für eine Evaluierung können beispielsweise sein:

- ein hoher sonstiger finanzieller Gesamtaufwand, also hohe Haushaltsausgaben oder sonstige Kosten,
- die besondere politische Bedeutung oder
- große Unsicherheiten über Wirkungen oder Verwaltungsvollzug.

Die Wesentlichkeit eines Regelungsvorhabens bemisst sich u. a. nach der Höhe des zu erwartenden jährlichen Erfüllungsaufwands.

Die Wesentlichkeit eines Regelungsvorhabens kann sich auch aus anderen Gründen ergeben, z. B. der besonderen politischen Bedeutung.

Unter bestimmten Voraussetzungen kann auf die Evaluierung wesentlicher Regelungsvorhaben verzichtet werden.

Der Aufwand für die Evaluierung muss in einem angemessenen Verhältnis zu den daraus zu gewinnenden Erkenntnissen stehen. Dies gilt zum Beispiel im Hinblick auf:

- gerichtliche Entscheidungen oder internationale bzw. EU-Vorgaben ohne Umsetzungsspielraum (1:1) sowie
- anderweitige vergleichbare Berichtspflichten – auch gegenüber oder von internationalen bzw. EU-Institutionen – oder eine gesetzlich vorgeschriebene Wirkungsforschung als Daueraufgabe.

Das bedeutet, dass auch wesentliche Vorhaben nicht evaluiert werden müssen, falls die Verhältnismäßigkeitsprüfung negativ ausfällt, etwa, weil schon anderweitige vergleichbare Berichtspflichten ähnliche Fragen beantworten. Wird auf eine Evaluierung verzichtet, so muss die Abwägung in der Begründung dargelegt werden.

Bei der Evaluierung der nationalen Umsetzung von EU-Recht empfiehlt sich der Nachbarcheck.

Evaluierung durch die EU-Kommission und Nachbarcheck

Wird mit einem Regelungsvorhaben in Deutschland eine EU-Richtlinie umgesetzt, sollte geprüft werden, ob auf europäischer Ebene bereits eine Evaluierung stattgefunden hat. Die **EU-Kommission** evaluiert eigene Rechtsvorschriften. Sollte hier bereits eine geeignete Evaluierung vorliegen, ist zu prüfen, inwieweit eine eigene Evaluierung sinnvoll ist. Dies hängt maßgeblich von der konkreten Umsetzung in Deutschland ab und inwieweit bei der Evaluierung auf EU-Ebene eine nationale Umsetzung berücksichtigt worden ist. Sollte eine weitere Evaluierung durch das zuständige Ressort gewünscht sein, ist die Umsetzung der Rechtsvorschrift auf nationaler Ebene zu prüfen, um ggf. Probleme bei der Umsetzung oder bei der entsprechenden Rechtsvorschrift identifizieren zu können. Hier bietet sich auch der sogenannte Nachbarcheck an. Neben den Evaluierungsergebnissen der EU-Kommission sollten demnach in dem eigenen Evaluierungsbericht die Erfahrungen aus anderen Mitgliedsstaaten einbezogen werden. Beispielsweise könnte untersucht werden, ob die Umsetzung in den Niederlanden oder in Frankreich besser funktioniert hat. Best-Practice-Beispiele können so identifiziert und Bessere Rechtssetzung auf europäischer und nationaler Ebene gewährleistet werden.



2.2 Darstellung im Regelungsvorhaben

Die Grundlage für die spätere Evaluierung wird bereits bei der Erstellung des Regelungsvorhabens gelegt. Ob eine Regelung die intendierten Auswirkungen hat, kann bei einer Evaluierung im Regelfall nur dann festgestellt werden, wenn die Ziele des Regelungsvorhabens klar definiert sind und Kriterien benannt werden, die bei der Beurteilung der Zielerreichung zugrunde gelegt werden. Die für die Evaluierung relevanten Inhalte finden sich sowohl im Vorblatt als auch in der Begründung des Regelungsvorhabens.

Die Festlegungen im Regelungsvorhaben bilden die Grundlage für die spätere Evaluierung.

2.2.1 Ausgangslage und Ziele

Nach § 43 Absatz 1 Nummer 1 GGO sind die Zielsetzung und Notwendigkeit des Gesetzentwurfs und seiner Einzelschriften in der Begründung darzustellen. Gleiches gilt nach Anlage 3 zu § 42 Absatz 1 GGO für das Vorblatt. Nummer 1 des St-Beschlusses vom 26. November 2019 sieht vor, dass die Ziele, die bei der Evaluierung zugrunde gelegt werden, ebenfalls benannt werden sollen. Ziele eines Regelungsvorhabens sind dessen beabsichtigten tatsächlichen Wirkungen.

Ziele des Regelungsvorhabens müssen benannt werden.

Ebenso wichtig wie die Darstellung der Ziele ist die Darstellung des Problems, also der Ausgangslage. Eine Evaluierung, welche die tatsächlichen Wirkungen eines Regelungsvorhabens in den Blick nimmt, ist nur dann aussagekräftig, wenn der Ist-Zustand bekannt und dokumentiert ist.

Ist- und Soll-Zustand werden im Vorblatt des Regelungsvorhabens unter „A) Problem und Ziel“ festgelegt.

Beispiele:

- Weniger Arbeitslose
- Reduzierung des CO₂-Ausstoßes
- Mehr BAföG-Bezieherinnen und -Bezieher

Beispiel für eine Zielbeschreibung:

Ziel der Bundesregierung ist ausweislich des Klimaschutzprogramms 2030, dass in Deutschland bis 2030 sieben bis zehn Millionen Elektrofahrzeuge zugelassen sein sollen. (...) Mit diesem Gesetz sollen die Voraussetzungen dafür geschaffen werden, den Ausbau der Leitungs- und Ladeinfrastruktur für die Elektromobilität im Gebäudereich zu beschleunigen.

Die bloße Umsetzung einer EU-Richtlinie kann für sich genommen nicht evaluierungsfähiges Ziel eines Regelungsvorhabens sein. Dient ein Regelungsvorhaben der Umsetzung einer EU-Richtlinie, so müssen deren Ziele für die Evaluierung zugrunde gelegt werden.

2.2.2 Bei wesentlichen Vorhaben: Aussagen zur Durchführung oder Nicht-Durchführung der Evaluierung

Bei wesentlichen Regelungsvorhaben muss in der Begründung angegeben werden, welche Ziele mit dem Gesetz erreicht werden sollen und welche Indikatoren für die Zielerreichung dabei voraussichtlich herangezogen werden.

Das Evaluierungskonzept der Bundesregierung sieht vor, dass bei wesentlichen Regelungsvorhaben, also solchen, die nach dem Konzept grundsätzlich evaluierungspflichtig sind, Aussagen zur Durchführung der Evaluierung in der Begründung gemacht werden müssen. Dies beinhaltet:

- welche Ziele bei der Evaluierung zugrunde gelegt werden und
- welche Indikatoren für die Zielerreichung dabei voraussichtlich herangezogen werden.

Die Indikatoren sind die tatsächlichen Umstände, anhand deren Veränderung bewertet wird und ob die vom Regelungsvorhaben angestrebten Ziele erreicht werden. Ist Ziel des Regelungsvorhabens beispielsweise, den Ausbau der Leitungs- und Ladeinfrastruktur für die Elektromobilität im Gebäudebereich zu beschleunigen, kann die Wirksamkeit des Gesetzes z. B. anhand der Anzahl der Ladesäulen bemessen werden.

Weiterhin sollen Angaben darüber gemacht werden,

- aufgrund welcher Daten die Zielerreichung ermittelt wird und
- wie die Verfügbarkeit dieser Daten sichergestellt wird sowie über welchen Zeitraum diese Zahlen erhoben werden.

Eventuelle datenschutzrechtliche Fragestellungen sollten frühzeitig bei der Planung, Konzeptionierung und Datensammlung einer Evaluierung berücksichtigt werden.

Regelungsvorhaben werden in der Regel drei bis fünf Jahre nach Inkrafttreten evaluiert.

Schließlich muss festgelegt werden, wann genau (in der Regel drei bis fünf Jahre nach Inkrafttreten) das Vorhaben evaluiert wird.

Beispiel für eine Evaluierungsklausel:

Eine umfassende Evaluierung des neuen Systems der Insolvenzsicherung im Reiserecht soll spätestens nach fünf Jahren erfolgen. Dabei soll evaluiert werden, ob mit der Neugestaltung eine effektive Insolvenzsicherung erreicht wurde. Als Kriterium kann auf die richtlinienkonforme Entschädigung betroffener Reisender abgestellt werden, wobei insoweit Informationen aus den Geschäftsberichten des Reisesicherungsfonds entnommen werden können. Es soll dabei auch untersucht werden, ob zeitgleich mehrere Reisesicherungsfonds die Erlaubnis zur Aufnahme des Geschäftsbetriebs erhalten können sollen oder sogar eine vollständige Marktöffnung möglich ist.

3 Planung der Evaluierung

In der Planungsphase erfolgen die grundsätzlichen Überlegungen zur späteren Durchführung der Evaluierung. Die Planung der Evaluierung erfolgt dabei grundsätzlich in **drei Phasen**:

1. Begleitung des Regelungsvorhabens

Die erste Phase beginnt bei den Überlegungen zur **Ausgestaltung der Evaluierungsklausel**. Darin wird spezifiziert, welche Ziele bei der Evaluierung zugrunde gelegt und welche Kriterien für die Zielerreichung dabei voraussichtlich herangezogen werden.

In dieser Phase sollten, soweit möglich, bereits Überlegungen zum Zeitrahmen der Evaluierung sowie den möglichen Datenquellen vorgenommen werden. Auch die von der Regelung Betroffenen sollten identifiziert werden. Ob genaue Angaben hierzu bereits in die Evaluierungsklausel übernommen werden, entscheidet das Ressort aus fachlicher Sicht. Da sich Regelungen während eines Gesetzgebungsverfahrens noch verändern können, müssen im weiteren Verfahren ggf. auch die Überlegungen zur Evaluierung angepasst werden.

2. Ermittlung und Erfassung der wesentlichen Punkte der Evaluierung

Möglichst zeitnah nach Abschluss des Regelungsvorhabens sollten in einer zweiten Planungsphase dann die Kernpunkte der geplanten Evaluierung strukturell z. B. in einem Grob-Konzept niedergelegt und ggf. bereits Datengrundlagen eruiert bzw. deren Schaffung für die spätere Evaluierung initiiert werden. Auch die Erstellung eines Wirkmodells kann hierbei hilfreich sein.

3. Planung der Durchführung der Evaluierung

Spätestens ein Jahr vor dem Evaluierungsbeginn (zwei bis vier Jahre nach Inkrafttreten des Regelungsvorhabens) muss die letztendliche Durchführung der Evaluierung geplant werden. Dadurch soll die effiziente Durchführung der Evaluierung sichergestellt werden.

3.1 Detaillierungsgrad

Evaluierungen können mit einer unterschiedlichen Detailtiefe erstellt werden. Während in einem Fall eine Aussage über die Gesamtheit der Betroffenen ausreichend sein mag, wird in einem anderen Fall der Einfluss des Regelungsvorhabens auf bestimmte Gruppen von besonderem Interesse sein. Der Detaillierungsgrad und damit der Aufwand für eine Evaluierung soll in einem angemessenen Verhältnis zu den daraus zu gewinnenden Erkenntnissen und zur Bedeutung des Evaluierungsgegenstands stehen. Die Entscheidung hinsichtlich der Tiefe, Methodik und des Umfangs der Evaluierung liegt beim fachlich für das zu evaluierende Regelungsvorhaben zuständigen Ressort.

Über die Detailtiefe einer Evaluierung entscheidet das zuständige Ressort.

Die Wesentlichkeit eines Regelungsvorhabens kann sich aber auch aus anderen Gründen ergeben, z. B. der Intensität von Eingriffen oder der Anzahl der Betroffenen.

Anhaltspunkte für die Bedeutung eines Vorhabens können sich aus der politischen und gesellschaftlichen Diskussion bei der Abstimmung des Vorhabens ergeben, auch die Intensität von Eingriffen und/oder die Relevanz von Leistungen im Einzelfall können unabhängig von der Höhe der Fallzahl auf eine große Bedeutung hinweisen. Die Zahl der Betroffenen sowie die Höhe des geschätzten oder tatsächlichen Erfüllungsaufwands (Nachmessung frühestens zwei Jahre nach Inkrafttreten) können ebenfalls relevant sein. Auch wenn für einen Regelungsbereich eine große Unsicherheit bezüglich Rechtsfolgen besteht, weil in diesem Bereich bislang wenig Evaluierungserkenntnisse vorliegen, kann einer Evaluierung hohe Bedeutung zukommen. Aus Effizienzgründen sollte geprüft werden, ob die Evaluierung und die Nachmessung verzahnt werden können (siehe Kapitel 3.6). Der Beschluss der Bundesregierung, Regelungsvorhaben systematisch zu evaluieren, erfordert jedenfalls ein am Regelungsinhalt orientiertes Vorgehen, unabhängig von seiner politischen Bewertung. Insbesondere wenn Ziel der Regelung ist, eine Wirkung gerade bei einer bestimmten Gruppe von Adressaten auszulösen, sollte die Auswirkung auf diese Gruppe im Rahmen der Evaluierung eine entsprechende Betrachtung erfahren.

Beispiele:

- Eine Regelung soll gerade jungen Unternehmen eine leichtere Teilnahme an öffentlichen Vergaben ermöglichen.
- Eine Regelung soll die berufliche Bildung stärken und insbesondere die berufliche Perspektive von Niedrigqualifizierten verbessern.

3.2 Interne oder externe Evaluierung

In der Literatur wird grundsätzlich zwischen der *Selbstevaluierung* und der *Fremdevaluierung* unterschieden. Eine Selbstevaluierung liegt vor, wenn das für die konkrete Regelung zuständige Fachreferat die Evaluierung selbst durchführt. Wird die Evaluierung von einem anderen Fachreferat des Ressorts oder von nachgeordneten Behörden geplant und durchgeführt, handelt es sich um eine Mischform aus Selbst- und Fremdevaluierung. Eine häufig benutzte Kombination ist auch die, in der Externe spezifische Aufgaben übernehmen oder Arbeitspakete erstellen, die Gesamtbewertung aber vom Fachreferat vorgenommen und der Bericht auch intern erstellt wird. Beide beschriebenen Evaluierungen werden hier unter dem Begriff der internen Evaluierung zusammengefasst. Abgegrenzt dazu handelt es sich bei der vollständigen Fremdevaluierung um eine Vergabe an externe Evaluierende, daher wird diese Form in dieser Arbeitshilfe als externe Evaluierung bezeichnet.

Die Abwägung, ob eine Evaluierung intern oder extern durchgeführt werden soll, liegt beim zuständigen Ressort.

Es obliegt den Ressorts, über die Wahl der Evaluierungsform zu entscheiden. Eine allgemeine Aussage darüber, welche Form der Evaluierung zu präferieren ist, ist nicht möglich. Bei der Entscheidungsfindung können die folgenden Aspekte behilflich sein:

Die interne Evaluierung hat zum Vorteil, dass das Wissen über fach- oder ressortspezifische Hintergründe bei den evaluierenden Stellen liegt und möglicher Änderungsbedarf direkt umgesetzt werden kann. Der mitunter zeitintensive Wissenstransfer bei einer externen Evaluierung wird stark reduziert oder fällt im Falle einer echten Selbstevaluierung sogar weg. Die hohe fachliche Nähe der Evaluierenden ist zum einen der Vorteil der internen Evaluierung, zum anderen sind damit allerdings spezifische Herausforderungen verknüpft. Die Evaluierung soll die Auswirkungen einer gesetzlichen Regelung systematisch, ergebnisoffen und umfassend überprüfen. Der fachliche Schwerpunkt der Evaluierenden kann zu einer verzerrten Evaluierung führen. Diese menschliche Tendenz, die eigenen Annahmen zu bestätigen, wird Bestätigungsfehler (*confirmation bias*) genannt. Dieser Denkfehler kann von Betroffenen selbst dann begangen werden, wenn sie sich dessen bewusst sind. Der Auswirkung eines möglichen Bestätigungsfehlers sollte durch eine methodische Genauigkeit entgegengewirkt werden.

Die interne Evaluierung profitiert vom hohen Fachwissen des Ressorts.

Es empfiehlt sich, das vorhandene Methodenwissen durch eine organisatorisch unabhängige Methodenberatung zu ergänzen¹. Die gewählten Methoden sollten im Evaluierungsbericht transparent aufgezeigt und die erzielten Ergebnisse damit intersubjektiv überprüfbar gemacht werden. Des Weiteren müssen für eine interne Evaluierung ausreichend Ressourcen bereitgestellt werden. Dazu zählen Personen, die nach Möglichkeit bereits mit den Methoden einer Evaluierung vertraut sind und gegebenenfalls Sachmittel für mögliche Datenerhebungen. Darüber hinaus sollte eine interne Evaluierung explizit darauf zielen, bestehende Annahmen z. B. über Wirkmechanismen aufzuzeigen und zu hinterfragen.

Das Statistische Bundesamt unterstützt die Ressorts rund um methodische Fragen.

Die externe Evaluierung wird fremd vergeben, wobei eine enge Begleitung durch das Ressort weiterhin stattfinden muss. Zudem bedarf die Entscheidung über die Erforderlichkeit einer Fremdvergabe der Dokumentation. Für das Vergabeverfahren sollte ausreichend Zeit eingeplant werden. Weitere Details der Vergabe sind ressortintern geregelt. Für das Schreiben einer Leistungsbeschreibung, der Ausschreibung und der konkreten Vergabe sind demnach ungefähr ein halbes Jahr einzuplanen.

Auch bei einer externen Evaluierung muss das zuständige Ressort die Evaluierung fachlich begleiten.

Die externen Anbieter bringen das Methodenwissen zu Evaluierungen mit und übernehmen, je nach Leistungsbeschreibung bzw. Lastenheft, die Datenerhebung und erstellen den Evaluierungsbericht. Die Verantwortung über den Umgang mit den Ergebnissen und den daraus entstandenen Handlungsempfehlungen liegt weiterhin beim zuständigen Ressort. Außerdem muss das vorhandene Fachwissen aus den Ressorts dem externen Unternehmen zugänglich gemacht und ggf. Haushaltsmittel für die Fremdvergabe rechtzeitig eingeplant werden. Dies bindet auch interne personelle Ressourcen.

Die externe Evaluierung profitiert vom hohen Methodenwissen der externen Anbieter.

¹ Zum Beispiel bietet das Statistische Bundesamt eine methodische Beratung an. Erstkontakt: evaluierung@destatis.de

Tabelle 1: Vor- und Nachteile interner und externer Evaluierung

	Interne Evaluierung	Externe Evaluierung
Vorteile	Fachwissen liegt beim Ressort selbst (kein Wissenstransfer notwendig)	Beratungsfirmen, Institute etc. bringen methodisches Fachwissen für Evaluierung mit
	Änderungen sind schnell umzusetzen	Personeller Aufwand des Ressorts geringer als bei interner Evaluierung
Nachteile	Tendenz die eigenen Annahmen zu bestätigen	Um eine gute externe Evaluierung zu gewährleisten, ist eine detaillierte Leistungsbeschreibung nötig
	Mitunter hoher personeller Aufwand	Ebenfalls personeller Aufwand beim Ressort durch Wissenstransfer
	Oftmals fehlt Methodenwissen für eine Evaluierung	Fehlendes Fachwissen

3.3 Ermittlung der Betroffenen

So früh wie möglich sollten von einer Rechtsnorm Betroffene in die Evaluierung einbezogen werden.

Frühzeitig sollten die von einer Rechtsnorm betroffenen Akteure (Normadressaten) ermittelt und in die Evaluierung einbezogen werden. In der Regel können mehrere Adressatengruppen unterschieden werden. Eine Adressatengruppe sind gleichartige Akteure, z. B. die vollziehende Verwaltung, die eine Rechtsnorm umsetzt oder die Wirtschaft, die z. B. neue Vorgaben befolgen muss.

Die Rechtsnorm kann sich sowohl auf die Wirtschaft und die Verwaltung als auch auf die Bürgerinnen und Bürger auswirken. Die Ermittlung dieser Akteure ist eine wichtige Voraussetzung für die Erhebung von Evaluierungsdaten. Die leitende Frage lautet „Wer ist von der Rechtsnorm direkt oder indirekt betroffen?“. In der Regel werden hierzu bereits Erkenntnisse aus dem Gesetzgebungsverfahren vorliegen, z. B. im Rahmen der Beteiligung von Verbänden bei der Erstellung des Gesetzentwurfs oder im Rahmen der öffentlichen Anhörung des federführenden Ausschusses des Bundestages oder aus den Stellungnahmen der Länder bzw. der Stellungnahme des Bundesrates.

Bereits vorliegende Verwaltungsdaten sollten, wenn möglich, als Datengrundlage genutzt werden. Allerdings wird für eine Evaluierung regelmäßig eine zusätzliche Datenerhebung notwendig sein. Ein Wirkmodell (siehe Kapitel 3.4) unterstützt bei der Wahl der notwendigen Daten und Informationen. Die Evaluierenden sollten bereits frühzeitig die Inhaber dieser Daten und Informationen ermitteln. Die Kontaktaufnahme kann auch mit der Hilfe von Interessensverbänden erfolgen.

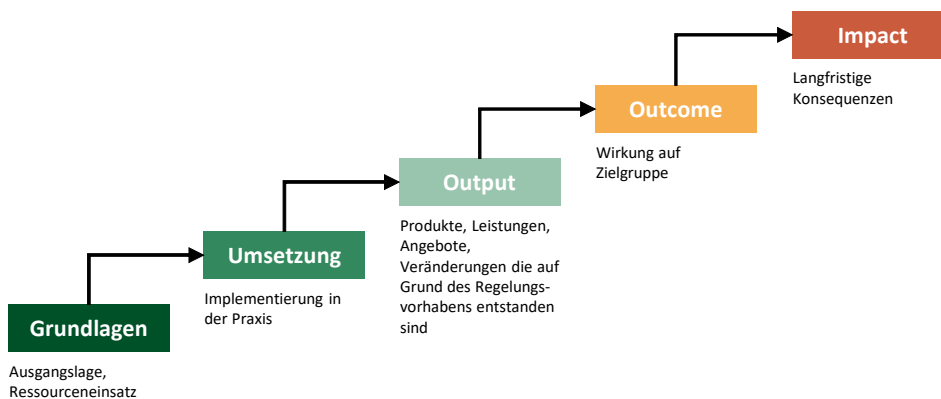
3.4 Entwicklung eines Wirkmodells

Ein Wirkmodell beschreibt die hypothetische logische Ablauffolge zwischen der Rechtsnormänderung und der Zielsetzung. Damit wird explizit dargelegt, über welche Einflussfaktoren der Evaluierungsgegenstand kausal zur Zielerreichung beiträgt. Das Wirkmodell ist ein zentraler Bestandteil der Evaluierung und muss explizit offengelegt werden. Das Wirkmodell sollte mit Expertinnen und Experten und/oder Betroffenen validiert werden. Insbesondere können dadurch ungeplante Auswirkungen der Regelung in das Wirkmodell integriert werden.

Das Wirkmodell ist der Grundpfeiler einer guten Evaluierung und dient als Wegweiser. Es sollte mit der Formulierung des Regelungsvorhabens aufgestellt werden.

Die Mindestanforderung an ein Wirkmodell stellt das logische Modell² dar. Im logischen Modell werden folgende Elemente dargestellt: *Grundlagen, Umsetzung, Output, Outcome* und *Impact*. Die einzelnen Elemente werden anhand eines Beispiels konkret erläutert.

Abbildung 1: Wirkmodell



Die **Grundlagen** beschreiben die Ausgangslage bei der Änderung der Rechtsnorm. Diese sollte im Teil A des Gesetzesvorblatts beschrieben sein. Außerdem werden das Ziel der Rechtsnormänderung und die geplante Umsetzung (Konzept) beschrieben. Explizit sollten an dieser Stelle auch die geplanten Ressourcen erläutert werden.

Beispiel aus dem Bereich der E-Mobilität:

Ziel einer fiktiven Rechtsnormänderung ist die Bereitstellung von mehr Ladesäulen für Elektrofahrzeuge im Straßenverkehr in Deutschland, damit mehr Elektrofahrzeuge gekauft werden. Übergeordnetes Ziel ist es, dass weniger Verbrenner-Fahrzeuge auf deutschen Straßen fahren und somit die CO₂-Emissionen zu reduzieren

² Das logische Modell ist in fünf Phasen unterteilt, die nacheinander stattfinden und den Ablauf eines Regelungsvorhabens darstellen (Abbildung 1). Diese lineare Abfolge der Phasen eines Regelungsvorhabens dient aber nur der Vereinfachung und entspricht selten der Realität. Deutlich komplexere Wirkmodelle (Programmtheorien) sind denkbar, bei denen es Wechselwirkungen und Rückkopplungseffekte zwischen verschiedenen Phasen gibt. Für die Evaluierung setzen solche Modelle aber weitreichende Kenntnisse oder Vermutungen über Wirkzusammenhänge voraus und der Datenbedarf ist deutlich höher.

und den Klimawandel zu bekämpfen. In der Evaluierung müsste nun bei den Grundlagen zuerst die Ausgangslage, also das Problem beschrieben werden. Dies kann in ähnlicher Weise wie in Teil A des Gesetzesvorblatts erfolgen. Im konkreten Beispiel: Zu viel CO₂-Emissionen, weil zu wenige Fahrzeuge mit Elektromotor. Die Annahme wird getroffen, dass Elektroautos unbeliebt sind, weil es nicht genügend Lademöglichkeiten gibt. Um mehr Lademöglichkeiten zu schaffen, sollen diese gefördert werden. Weiter könnte in der Evaluierung dargestellt werden, wie viele Ladesäulen durch die Rechtsnormänderung neu aufgestellt werden sollten und wie viele es vor der Rechtsnormänderung gab (Daten müssen ggf. frühzeitig erhoben werden). Auch der mediale und politische Diskurs vor, während und nach der Rechtsnormänderung kann dargestellt werden, um die damaligen gesellschaftlichen Rahmenbedingungen abzubilden.



Im Anschluss wird unter **Umsetzung** die praktische Implementierung der Rechtsnorm erläutert. Dazu werden alle Tätigkeiten, die zur Zielerreichung beigetragen haben, verstanden. Ebenfalls sollten unterstützende und überprüfende Aktivitäten aufgenommen werden.

Im fiktiven Beispiel E-Mobilität lag der Schwerpunkt der Rechtsnormänderung auf der Schaffung eines Förderprogramms durch die Kreditanstalt für Wiederaufbau (KfW), bei dem jede neue Ladesäule bezuschusst wird. In der Umsetzung könnte nun dargestellt werden, inwieweit konkret die Umsetzung des Förderprogramms bei der KfW funktioniert hat. Ist rechtzeitig ein geeignetes und gut verständliches Antragsverfahren eingerichtet worden? Sind Stellen zur Antragsprüfung und -bewilligung bei der Förderbank geschaffen worden?

Der **Output** beschreibt die unmittelbaren Leistungen, die aus der Rechtsnormänderung entstanden sind. Unter Leistungen sind dabei alle Ausbringungen der Umsetzungsphase zu verstehen. Häufig sind Outputs leicht zu zählen. Neben der reinen Leistungsbereitstellung wird im Output-Teil die Nutzung der Leistung dargestellt.

Die unmittelbaren Leistungen und somit der Output aus der Rechtsnormänderung sind die Ladesäulen. Evaluiert werden müsste also, wie sich die Anzahl der Ladesäulen seit Inkrafttreten der Rechtsnormänderung entwickelt hat. Zum Beispiel könnte die Anzahl der Ladesäulen alle zwei Jahre erhoben und als eine Entwicklung in der Evaluierung abgebildet werden.

Der **Outcome** beschreibt den mittelbaren Effekt einer Rechtsnormänderung, also das eigentliche Ziel einer Rechtsnormänderung. Welche Reaktion/Wirkung/Verhaltensänderung wird angestrebt?

Im obigen Beispiel könnte der Outcome eine höhere Anzahl an Elektroautos sein, die aufgrund der besseren Ladeinfrastruktur gekauft werden. Daher sollte die Zahl der verkauften Elektroautos evaluiert werden.

Der **Impact** beschreibt den langfristigen Einfluss der Rechtsnorm sowohl auf die Zielgruppe als auch auf weitere gesellschaftliche Gruppen. Die Effekte können positiv, negativ oder neutral sein. Hier fallen nicht nur Wirkungen an, die vorher beabsichtigt und im Regelungsvorhaben bei den Zielen dargestellt worden sind, sondern auch nicht-intendierte Nebeneffekte.

Im vorherigen Beispiel ist der Einfluss der Regelung auf die Erfüllung von Klimazielen zu untersuchen (beabsichtigte Wirkung). Andere Gruppen wären zum Beispiel dann betroffen, wenn die neuen Ladesäulen an vorhandenen Parkbuchten gebaut würden und dadurch den danebenlaufenden Radweg einengen (nicht-intendierte Nebeneffekte). Hier wäre zu untersuchen, wie sich die Zufriedenheit der Radfahrenden verändert.

Die Grenzen zwischen Output, Outcome und Impact verlaufen fließend. Während erstere die unmittelbaren und mittelbaren Effekte auf die Zielgruppe umfassen, hilft der Impact bei der Identifizierung von weiteren gesellschaftlichen Effekten.

3.5 Wahl des Untersuchungsdesigns

Das Untersuchungsdesign bestimmt den konkreten Informationsbedarf für die Evaluierung. Die drei möglichen Untersuchungsdesigns können jeweils auf die einzelnen Bestandteile des Wirkmodells angewendet werden.

Ausgehend vom Wirkmodell muss der Informationsbedarf ermittelt werden. Der Bedarf ist abhängig vom Untersuchungsdesign, welches wiederum auf die einzelnen Elemente des Wirkmodells angewendet werden sollte (z. B. Output, Outcome, Impact). Grundsätzlich sind drei Untersuchungsdesigns vorstellbar:³

- Soll-Ist-Vergleich
- Vorher-Nachher-Vergleich
- Mit-Ohne-Vergleich

In der Aufzählung steigen die Untersuchungsdesigns sowohl in ihrem Informationsbedarf als auch in der Belastbarkeit und Validität der Ergebnisse. Unabhängig vom gewählten Design sollten bereits bei der Erstellung des Regelungsentwurfs die erforderlichen Datenquellen identifiziert werden. Bei internen Evaluierungen sind dazu sämtliche zuständigen Referate einzubinden. Nur so können Gegenmaßnahmen zu eventuellen Datenengpässen getroffen werden.

Für den Soll-Ist-Vergleich werden Daten zu einem bestimmten Zeitpunkt benötigt, um zu evaluieren, ob das Ziel des Regelungsvorhabens erreicht wurde.

Ein Soll-Ist-Vergleich stellt den Zielerreichungsgrad zu einem bestimmten Zeitpunkt fest. Diese Vorgehensweise setzt voraus, dass die Ziele hinreichend gut ausformuliert und quantifiziert sind (SMARTe-Ziele⁴). Der Soll-Ist-Vergleich eignet sich nicht für Ziele, die lediglich eine „Verbesserung“ beschreiben, ohne festzulegen, wie diese festgestellt und gemessen werden sollen. Reine „Verbesserungsziele“ sind daher zu vermeiden.

Der Informationsbedarf für einen Soll-Ist-Vergleich betrachtet die Daten zu einem Stichtag. Ein Beispiel liefert das **Erneuerbare-Energien-Gesetz (EEG)**: „Ziel des EEG ist es, die Energieversorgung umzubauen und den Anteil der erneuerbaren Energien an der Stromversorgung bis 2050 auf mindestens 80 Prozent zu steigern“. Ein Soll-Ist-Vergleich würde ausschließlich Daten über die Zusammensetzung der Stromversorgung für 2050 benötigen. Bezogen auf das Beispiel der E-Mobilität könnte beim Output die Anzahl der Antragstellungen zur Bezuschussung einer Ladesäule bei der KfW dargestellt werden und beim Outcome der Anteil der Elektrofahrzeuge auf deutschen Straßen fünf Jahre nach Inkrafttreten.

Mit dem Vorher-Nachher-Vergleich kann anhand von Zeitreihen die Zielerreichung dargestellt werden.

Ein Vorher-Nachher-Vergleich zeigt die Entwicklung der Outputs, Outcomes und Impacts im Zeitablauf. Der Datenbedarf für dieses Untersuchungsdesign ist verglichen zum Soll-Ist-Vergleich aufgrund der benötigten Zeitreihe höher. Beim Beispiel E-Mobilität könnte jährlich erhoben werden, wie sich die Anzahl der Ladesäulen (Output) und die Anzahl der Elektrofahrzeuge (Outcome) verändern. Mögliche Datengrundlagen können Interviews, stichprobenartige quantitative Befragungen oder die Auswertung bestehender bzw. die Einführung einer neuen

³ Mischformen der Untersuchungsdesigns sind möglich.

⁴ Ein Ziel ist SMART, wenn es: spezifisch, messbar, erreichbar, relevant und terminiert ist.

amtlichen Statistik⁵ sein. Die Daten sollten vor Inkrafttreten der jeweiligen Regelung als „Vorher-Daten“ vorhanden und damit auch zukünftig belastbar erhebbar sein.

Die bisher genannten Methoden haben ein grundsätzliches Problem gemein. Mit ihnen ist es in der Regel nicht möglich, kausale Zusammenhänge zu identifizieren. Anders gesagt: Mit ihnen kann zwar ermittelt werden, ob ein Ziel erreicht wurde oder ob eine Verbesserung eingetreten ist, allerdings bleibt unklar, ob die Rechtsnormänderung daran einen Anteil hatte und wie groß ggf. der Anteil war. Die Frage nach der Kausalität ist für Evaluierungen allerdings von Bedeutung, denn erst hierdurch werden Aussagen zur Wirkung des Vorhabens eröffnet; sie ermöglichen ein „Lernen“ für zukünftige Vorhaben.

Kausale Zusammenhänge können durch einen Mit-Ohne-Vergleich ermittelt werden. Dieser Versuch zielt darauf ab, Vergleichsgruppen zu bilden, die sich im Idealfall nur durch die effektiven Auswirkungen einer Rechtsnorm unterscheiden. Die Ermittlung von geeigneten Vergleichsgruppen ist eine große Herausforderung für jeden Einzelfall. Gilt eine Rechtsnorm beispielsweise erst ab einer Unternehmensgröße von 50 Mitarbeitenden, können zwei Vergleichsgruppen gebildet werden: Die Kontrollgruppe A mit 40-49 Beschäftigten und die Gruppe B mit 50-59 Beschäftigten. Bezüglich anderer Merkmale (z. B. Branche) sollten die Gruppen ähnlich sein. Die Zahl der Mitarbeitenden ist ebenfalls sehr ähnlich. Daher wird idealtypisch angenommen, dass sich die Unternehmen nur bezüglich der Norm unterscheiden. Die Vergleichbarkeit der Unternehmen sollte anhand der Ausgangswerte (Daten zum Zeitpunkt vor Inkrafttreten des Regelungsvorhabens für beide Gruppen) dargestellt werden. Eine vergleichende Vorher-Nachher-Untersuchung der beiden Gruppen gibt dann einen guten Hinweis darauf, wie sich die Änderung der Norm ausgewirkt hat⁶. Je nach Fragestellung der Evaluierung kommen unterschiedliche Modelle zur Bildung von Vergleichsgruppen in Frage.

Ist ein Mit-Ohne-Vergleich nicht möglich oder nicht verhältnismäßig, kann mit vergleichenden Interviews oder Umfragen zumindest eine Annäherung an die kausalen Zusammenhänge erfolgen.

Untersuchungen, die kausale Zusammenhänge darlegen können, gelten als der „Goldstandard“ der Evaluierung und sollten daher, wo möglich und verhältnismäßig, gewählt werden. Im Detail gibt es viele verschiedene mögliche Untersuchungsdesigns. Eine Kontaktaufnahme mit dem Statistischen Bundesamt zwecks methodischer Ersteinschätzung/Erstberatung wird empfohlen⁷.

Wirklich kausale Zusammenhänge darüber, dass das Regelungsvorhaben für die Zielerreichung verantwortlich ist, können nur mit dem Mit-Ohne-Vergleich getroffen werden.

⁵ Beispiel: Mit dem Prostituiertenschutzgesetz (→§35 ProStSchG) wurde auch eine entsprechende Statistik eingeführt.

⁶ Es empfiehlt sich für diese Art der Evaluierung (auch) Interviews mit Unternehmen beider Gruppen durchzuführen, da eine Rechtsnorm Effekte außerhalb der Zielgruppe erzeugen kann.

⁷ Das Statistische Bundesamt berät die Ressorts bezüglich der Methodik unter evaluierung@destatis.de

3.6 Wahl der Erhebungsverfahren

Für eine Evaluierung kommen viele verschiedene mögliche Datenquellen in Frage. Grundsätzlich kann zwischen Primärdaten und Sekundärdaten unterschieden werden.

Nachdem das Untersuchungsdesign festgelegt wurde, stellt sich die Frage, wie der damit einhergehende Daten- und Informationsbedarf gedeckt werden kann. Zur Erhebung der Daten kommen unterschiedliche Verfahren in Frage, die bei einer Evaluierung häufig kombiniert werden.

Grundsätzlich wird zwischen Primär- und Sekundärdaten unterschieden. Sekundärdaten wurden bereits aus einem anderen Grund als der Evaluierung erhoben. Dazu gehören (amtliche) Statistiken genauso wie qualitative Ausarbeitungen wie zum Beispiel Interviews oder Gerichtsverhandlungen (z. B. Urteile oder Protokolle). Eine sekundäre Datenerhebung ist somit der Abruf von Daten aus einem Datenbestand, der bereits vorliegt. Dabei kann der gesamte Datenbestand oder auch nur ein Teil des Datenbestands genutzt werden. Voraussetzung für die Nutzung bereits vorhandener Datenbestände ist, dass die bereits vorliegenden Daten auf nachvollziehbaren Wegen und nach gefestigten Methoden erhoben wurden.

Die Erhebung von Primärdaten kann grundsätzlich auf drei Arten erfolgen:

- Eigens konzeptionierte und fortlaufende amtliche Statistik
- Leitfadengestützte Interviews
- Schriftliche Befragungen

Eine amtliche statistische Erhebung muss in der Regel frühzeitig beim Entwurf des Regelungsvorhabens festgelegt werden (siehe Kapitel 3.5). Sie erlaubt dann eine passgenaue Erhebung von quantitativen Aussagen. Amtliche Statistiken müssen nach [§ 5 Bundesstatistikgesetz \(BStatG\)](#) gesetzlich angeordnet werden (keine Statistik ohne Gesetz). Darüber hinaus erlaubt [§ 7 BStatG](#) Erhebungen ohne Auskunftspflicht für besondere Zwecke ohne Anordnung.

Die Nachmessung des Erfüllungsaufwands durch das Statistische Bundesamt kann um Evaluierungsfragen ergänzt werden.

Für die Erhebung von Informationen mittels Interviews oder einer schriftlichen Befragung sollte aus Effizienzgründen geprüft werden, ob die Befragung mit der Nachmessung des Erfüllungsaufwands zusammen erfolgen kann.

Das Statistische Bundesamt misst den vorab geschätzten Erfüllungsaufwand (gesamter messbarer Zeitaufwand und Kosten, die bei einem Normadressaten durch die Befolgung bundesrechtlicher Vorschriften entstehen) ungefähr zwei Jahre nach Inkrafttreten nach. Hierzu werden z. B. telefonische Interviews mit Unternehmen geführt, die von einem Regelungsvorhaben betroffen sind. Die konkrete Gestaltung der dafür notwendigen Fragebögen variiert, häufig wird jedoch abgefragt, wie sich Zeitaufwände im Unternehmen durch die Befolgung einer Vorschrift bei einem Prozess verändert haben. Beispielsweise werden folgende Fragen gestellt: Wird die Änderung im Betrieb umgesetzt? Wie hat sich die Bearbeitungszeit verändert? Das Statistische Bundesamt bietet an, Evaluierungsfragen in den Fragebogen mit aufzunehmen. Dadurch müssen Ressorts keine eigene Befragung durchführen und eine doppelte Befragung von Betroffenen kann vermieden werden. Die Ressorts sollten frühzeitig an das Statistische Bundesamt herantreten, damit die Evaluierungsfragen

mit in die Zeit- und Ressourcenplanung der Nachmessung aufgenommen werden können. Das Statistische Bundesamt informiert die Ressorts über anstehende Nachmessungen und bietet die Verzahnung mit den Evaluierungsfragen an (ungefähr zwei Jahre nach Inkrafttreten einer Regelung).

Interviews jeglicher Art (telefonisch, persönlich, Videokonferenzen) sollten nach Möglichkeit anhand eines Leitfadens geführt werden. Gegenüber einem frei geführten Interview bietet diese Vorgehensweise mehrere Vorteile: Der Leitfaden gibt dem Interview eine inhaltliche Struktur. Zudem können durch die Zusammenstellung von vorab überlegten Fragetechniken (unter anderem W-Fragen, Aufzählung, Einordnungsfragen) gezielt die benötigten Informationen in Erfahrung gebracht werden. Leitfadengestützte Interviews ermöglichen außerdem eine bessere, weil objektivere Vergleichbarkeit zwischen mehreren Interviews und erlauben somit eine systematische Auswertung. Allerdings sind sie bei einer großen Befragung mit einem nicht unerheblichen Zeitaufwand verbunden. Aus diesem Grund wird in der Regel nur eine begrenzte Anzahl an Personen befragt. Die Auswahl der Befragten ist wichtig für die Aussagekraft der Evaluierung. Für jede Adressatengruppe (siehe Kapitel 3.3) sollten nach Möglichkeit mehrere Interviews geführt werden. Bei der Auswahl der Betroffenen können auch Multiplikatoren (z. B. Verbände) eingebunden werden. Der Leitfaden kann dabei unterschiedliche Befragungsschwerpunkte für die verschiedenen Adressatengruppen (z. B. Bürgerinnen und Bürger, Wirtschaft, Verwaltung) ermöglichen.

Interviews sollten mit einem Leitfaden geführt werden, um dem Interview eine Struktur zu geben.

Die schriftliche Befragung und die Online-Befragung verzichten auf die Flexibilität des leitfadengestützten Interviews. Alle Befragten aus einer Adressatengruppe erhalten die gleichen Fragen mit den gleichen Antwortmöglichkeiten. In der Regel werden die Fragen in der gleichen Reihenfolge gestellt. Die Befragung kann dabei sowohl per ausfüllbarem PDF, per Brief als auch per Onlineumfrage durchgeführt werden. Da die Befragten keine Rückfragen stellen können, muss der Fragebogen selbsterklärend sein. Somit ist eine schriftliche Befragung eine besondere Herausforderung für die Fragebogenkonzeption.

Schriftliche Befragungen bieten sich an, wenn viele Personen befragt werden sollen.

Neben der reinen Anzahl spricht auch die Verfügbarkeit der entsprechenden Adressen⁸ für eine schriftliche Befragung. Die schriftliche Befragung kann auch als explorative Untersuchung verstanden werden. Beispielsweise können durch die Befragung Schwerpunktthemen für noch ungenutztes Potenzial identifiziert werden, die in einem zweiten Schritt mithilfe von leitfadengestützten Interviews vertieft werden können.

Zusammenfassend kann gesagt werden, dass es ein generell zu bevorzugendes Erhebungsverfahren nicht gibt. Vielmehr sollten die Vor- und Nachteile der verschiedenen Methoden (Tabelle 2) gegeneinander abgewogen werden. Dabei können die verschiedenen Erhebungsinstrumente miteinander verbunden werden. Für die Evaluierung sollten die eigenen Erhebungen auch mit weiteren Möglichkeiten (z. B. Verwendung von Sekundärdaten) kombiniert werden.

Jedes Erhebungsverfahren hat Vor- und Nachteile.

⁸ Sollten beispielsweise Behörden befragt werden, lassen sich die Adressen relativ einfach recherchieren, an welche Fragebögen versendet werden können.

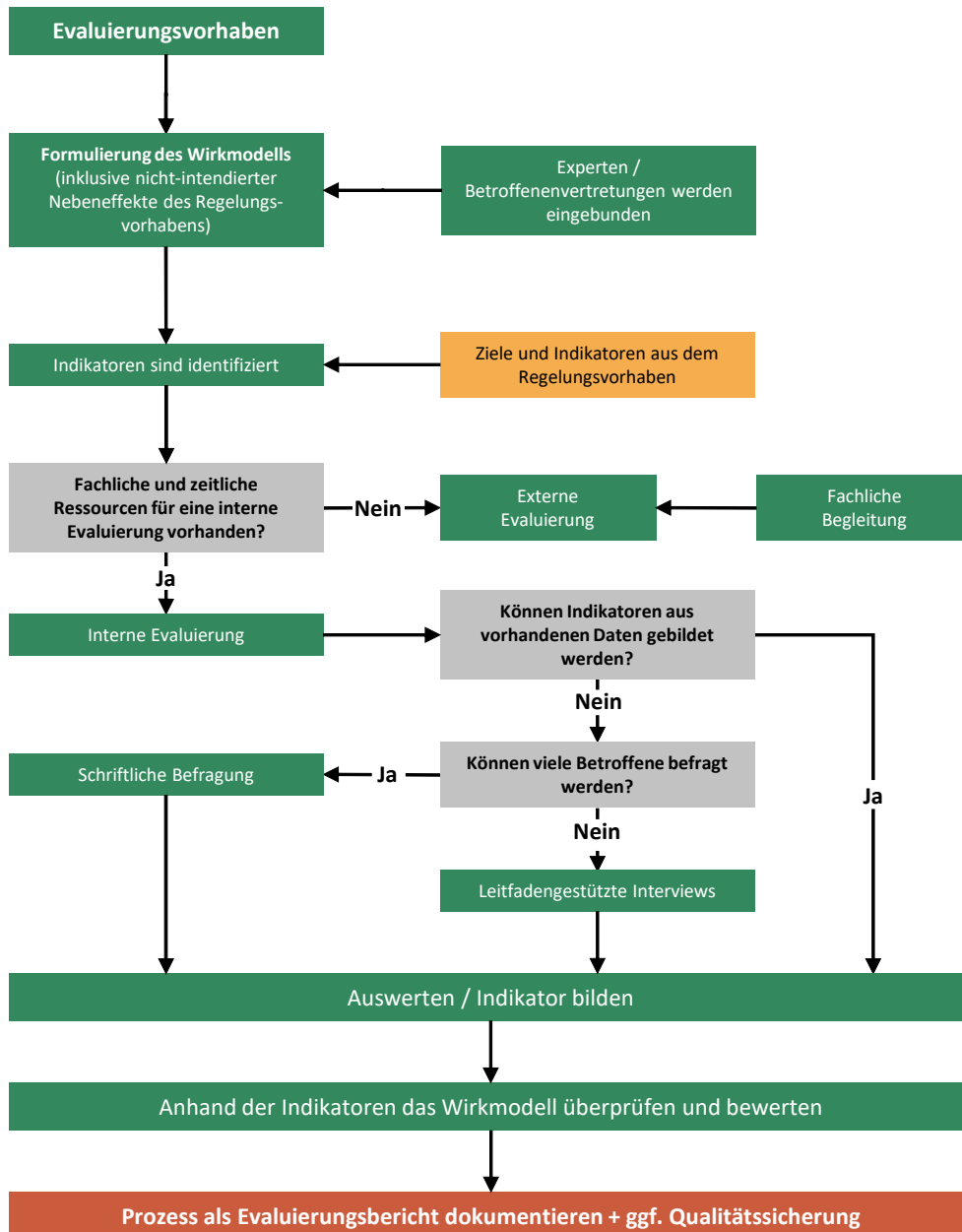
Tabelle 2: Vor- und Nachteile ausgewählter Erhebungsverfahren

Erhebungsverfahren	Vorteile	Nachteile
Neue Statistik einführen (Primärdaten)	Kontrolle über Daten und Datenerhebung Die Statistik kann ggf. auch für weitere Zwecke verwendet werden	Hoher Aufwand, hohe Kosten
Leitfadengestützte Interviews – viele offene Fragen (Primärdaten)	Systematischer Vergleich von Interviews Befragte haben die Möglichkeit, ausführlich die Wirkung eines Regelungsvorhabens zu beschreiben	Hoher Aufwand, wenn viele Interviews geführt werden sollen
Schriftliche Befragung – oftmals viele geschlossene Fragen (Primärdaten)	Viele Befragungen möglich, wodurch bei geeignetem Stichprobendesign repräsentative Aussagen möglich sind Befragungsergebnisse können teilweise in Zahlen übersetzt werden, wodurch statistische Auswertungen möglich sind	Rückfragen der Befragten können kaum beantwortet werden
Auswertung bereits bestehender Statistiken oder Studien (Sekundärdaten)	Kostengünstig	Keine Kontrolle über Datenerhebung

3.7 Schematische Darstellung einer Evaluierung

Die folgende Abbildung zeigt schematisch den möglichen Ablauf einer Evaluierung. Dem Prozess vorgelagert sind die Festlegungen im Rahmen des Regelungsvorhabens.

Abbildung 2: Ablaufplan einer Evaluierung



4 Durchführung einer internen Evaluierung

4.1 Darstellung des Gegenstands der Evaluierung

In dem Evaluierungsbericht sollte kurz das Regelungsvorhaben beschrieben werden.

Zu Beginn der Evaluierung sollte das Regelungsvorhaben dargestellt werden. Die Vorstellung sollte allgemein verständlich gehalten werden. Die Darstellung des Evaluierungsgegenstands ergibt sich im Regelfall aus der Evaluierungsklausel und sollte die folgenden Fragen beantworten:

- Was hat das Regelungsvorhaben geändert?
- In welchem Kontext/Diskurs wurde das Regelungsvorhaben umgesetzt?
- Welches Problem versucht es zu lösen?
- Welche Ziele sollten erreicht werden?
- Welche Akteure sind direkt oder indirekt betroffen?
- Wie fand der Vollzug in der Praxis statt?

Das Wirkmodell der Planungsphase (siehe Kapitel 3.4) kann zur Grundlage für diese Fragestellungen genutzt werden.

4.2 Konzeption der Erhebungsinstrumente für Primärdaten

In der Planungsphase wurde bereits anhand des Wirkmodells der Informationsbedarf an quantitativen und qualitativen Informationen identifiziert. Für die Primärdatenerhebung ohne eigens konzeptionierte Statistik kommen häufig Befragungen in Betracht⁹.

Bei Evaluierungen ist häufig eine Befragung die sinnvollste Methode, um an Daten zu kommen.

Für die Ausgestaltung der Befragung steht eine Vielzahl von Erhebungstechniken zur Verfügung. In diesem Abschnitt sollen die gängigsten vorgestellt und allgemeine Hinweise gegeben werden. Eine Diskussion der grundsätzlichen Vor- und Nachteile der Befragungsmethoden findet sich im Kapitel 3.6.

Unabhängig von der Befragungsmethode sollte das Frageprogramm so kurz wie möglich sein. Informationen, die leicht auf anderem Wege zu beschaffen sind, sollten nur dann erfragt werden, wenn die Validierung dieser Information nötig ist.

4.2.1 Erarbeitung eines Interviewleitfadens

Hauptzweck eines Interviewleitfadens ist es, die Vergleichbarkeit zwischen mehreren Interviews zu sichern und gleichzeitig eine Flexibilität in Form von Nachfragen oder Schwerpunktsetzung zu ermöglichen. In einem Evaluierungsprozess können

⁹ Umfragen können durch andere Erhebungsformate der Verhaltensforschung oder Rechtsgutachten ergänzt werden.

mehrere Interviewleitfäden notwendig sein. So können sich die Befragung von Betroffenen in der Verwaltung von der Befragung von Bürgerinnen und Bürgern oder der Wirtschaft unterscheiden. Diesem Aspekt sollte durch angepasste Interviewleitfäden Rechnung getragen werden.

Anhörungen und Interviews von Expertinnen und Experten

Auch einzelne Personen, zum Beispiel Expertinnen und Experten, können befragt werden, um beispielsweise eine fachliche Einschätzung zu erhalten, welche Daten für die Evaluierung genutzt werden können und wie diese auszuwerten sind. Auch Expertenanhörungen bieten sich an, um verschiedene und mitunter konträre Sichtweisen in die Evaluierung einzubinden. Dabei spielt die Auswahl der einzelnen Expertinnen und Experten eine wichtige Rolle.

Ein Interview sollte zuvor angekündigt und ein entsprechender zeitlicher Freiraum geschaffen werden. Zu Beginn des Interviews oder auch schon bei der Terminvereinbarung sollte der Interviewer oder die Interviewerin den Hintergrund der Befragung erläutern. Dabei sollte auch die Rolle des Interviews für die Evaluierung erläutert werden. Ist geplant, aus einzelnen Interviews direkt zu zitieren, so ist zwingend die Zustimmung des oder der Interviewten einzuholen. Die Erläuterung des Gesprächshintergrunds soll möglichst neutral erfolgen, damit die Einschätzung des Gegenübers nicht beeinflusst wird.

Nach den einleitenden Informationen beginnt das eigentliche Frageprogramm. Es empfiehlt sich, mit einer leicht zur beantwortenden Frage zu einem angenehmen Aspekt zu beginnen. Mit dieser Frage soll ein offenes Gesprächsklima geschaffen werden. Eine kritische Frage zu Beginn birgt die Gefahr, dass der oder die Befragte eine verteidigende oder abwehrende Haltung einnimmt. Die Eingangsfrage sollte offen formuliert und somit nicht einfach mit Ja oder Nein zu beantworten sein.

Es bietet sich an, das Interview mit einer einfachen offenen Frage zu beginnen.

Alle Fragen sollten neutral gestellt werden, um den Befragten nicht bereits in eine Richtung zu beeinflussen (keine Suggestivfragen). Außerdem sollten die Fragen einfach verständlich sein. Es gilt der Grundsatz: „Pro Information eine Frage“. Es ist besser, zum selben Themengebiet mehrere Fragen zu stellen als eine einzige allumfassende Frage. Beispiel: Statt „Wie und in welcher Höhe hat Ihr Unternehmen im letzten Kalenderjahr in die IT-Infrastruktur investiert“ lieber:

Fragen sollten neutral gestellt werden und jeweils nur eine Information abfragen.

1. Hat ihr Unternehmen im letzten Kalenderjahr in die IT-Infrastruktur investiert?
2. Um was für Investitionen hat es sich dabei gehandelt?
3. Wie hoch schätzen Sie die Höhe der Investition in Euro ein?

Dabei dient die erste Frage auch als Filter für die folgenden zwei Fragen. Wird die erste Frage verneint, so erübrigen sich die weiteren Fragen. Die Anordnung der Fragen ist ebenfalls wichtig, um die Befragten in einen Fluss zu bringen. So sollten thematische Blöcke gebildet werden. Diese Vorgehensweise erlaubt den Teilnehmenden, sich langsam in die einzelnen Aspekte einzudenken. Die einzelnen Blöcke wie-

derum sollten anhand des Prozesses gegliedert sein. Häufig bietet sich eine chronologische Reihenfolge an. Die abschließende Frage des Interviews sollte den Befragten die Möglichkeit geben, weitere Aspekte anzusprechen.

Für die Interviewenden ist es hilfreich, wenn der Interviewleitfaden optisch bereits so gestaltet ist, dass die verschiedenen Themenblöcke schnell voneinander unterschieden werden können. Der Vorteil eines Leitfadens ist es, dass die Reihenfolge der Fragen zwar strukturiert, aber nicht fest ist. Springt der oder die Interviewte eigenständig zu einem eigentlich später angedachten Thema, sollte darauf flexibel reagiert werden können. Ein auch optisch strukturierter Leitfaden hilft dabei.

4.2.2 Entwicklung eines Fragebogens

Die Entwicklung eines Fragebogens bedarf sorgfältiger Überlegungen, da die Befragten in der Regel keine Rückfragen bei unverständlichen Fragen stellen können.

Ein Fragebogen stellt eine standardisierte Zusammenstellung von Fragen dar. Die Formulierung der Fragen bedarf einer besonders intensiven Sorgfalt, da auf mögliche Nachfragen seitens der Befragten in der Regel nicht eingegangen werden kann. Besteht bei einer Frage Interpretationsspielraum, so sollten begleitende Hinweise vorab für Klarheit sorgen. Andernfalls ist eine spätere aussagekräftige Auswertung erschwert.

Grundsätzlich sollte zum Beginn des Fragebogens oder in einem begleitenden Schreiben die Notwendigkeit der schriftlichen Befragung im Rahmen der Evaluierung dargelegt werden. Zusätzlich sollten die Teilnehmenden über den geschätzten Zeitaufwand, der für das Ausfüllen des Fragebogens benötigt wird, aufgeklärt werden.

Das Frageprogramm sollte so strukturiert sein, dass es „vom Allgemeinen zum Speziellen“ führt. Dies ermöglicht den Befragten, sich in das Thema langsam einzudenken. Aus demselben Grund sollten thematisch zusammenhängende Fragen in einem Fragenblock gestellt werden. Die Struktur des Fragebogens sollte durch eine ansprechende optische Gestaltung unterstützt werden. Das gilt insbesondere für Filterfragen. Darunter werden Fragen verstanden, die abhängig von den Antworten zu weiteren/anderen Fragen führen. Die Antwortmöglichkeiten von Filterfragen müssen disjunkt¹⁰ sein. Für alle Befragten muss eine Antwortmöglichkeit zutreffend sein. Komplizierte Filterführungen mit mehr als einer Bedingung sind für selbstauffüllende Befragte zu vermeiden.

Es sollten sowohl offene, geschlossene und halboffene Fragen in einem Fragebogen kombiniert werden.

Der Fragebogen kann (und sollte in der Regel) mehrere Fragetechniken kombinieren. Grundsätzlich wird zwischen offenen und geschlossenen Fragen unterschieden. Eine Mischform stellen die halboffenen Fragen dar.

Die offenen Fragen geben keine Antwortmöglichkeit vor. Die Befragten werden aufgefordert, selbst in Textform zu antworten. Die daraus resultierende Freiheit für die Antwortgebenden ist der Vorteil der offenen Fragen. Sie ermöglichen ganz individuelle Antworten und Begründungen. Die Herausforderung besteht in der zeitintensiven Auswertung der Antworten auf offene Fragen. Offene Fragen bieten sich

¹⁰ Antwortmöglichkeiten müssen einander ausschließen.

insbesondere dann an, wenn das Wirkmodell nicht gesichert ist und validiert/revidiert werden soll. Außerdem bieten sie sich für Verbesserungsvorschläge und Anmerkungen an.

Geschlossene Fragen zeichnen sich durch vorgegebene Antwortmöglichkeiten aus. Die Befragten können dann eine oder mehrere davon ankreuzen. Darunter fallen auch Fragen mit mehrstufigen Antwortskalen wie beispielsweise „(1) stimme gar nicht zu, (2) stimme eher nicht zu, (3) stimme eher zu, (4) stimme voll zu“. Geschlossene Fragen lassen sich schnell ausfüllen und sind auch einfacher auszuwerten. Sie sind dagegen nicht flexibel, und es kann vorkommen, dass Teilnehmende keine der Antworten für (generell) zutreffend halten, sondern lieber eine individuelle Antwort geben möchten.

Die vorgegebenen Antworten haben einen Einfluss auf die späteren Auswertungsmöglichkeiten. Deshalb sollte bereits bei der Gestaltung des Frageprogramms die Auswertung mitgedacht werden. Scheinbar leichte Änderungen der Antwortmöglichkeiten können die Auswertung erschweren.

Beispiel 1: Soll das Alter einer Person abgefragt werden und es werden Altersbereiche vorgegeben (0-19; 20-39; 40-59; =>60), so können im Anschluss nur noch diese Altersgruppen miteinander verglichen werden. Es ist allerdings dann nicht möglich, ein durchschnittliches Alter über alle Befragten festzustellen. Wird dagegen das konkrete Alter in Jahren abgefragt, dann können die oben genannten Gruppen auch im Anschluss noch gebildet werden.

Beispiel 2: Für jede gewünschte Information sollte eine eigene Frage gestellt werden. Doppelfragen wie: „Welches ist der höchste Bildungsabschluss, über den Sie verfügen oder den Sie anstreben?“ sind ungünstig. Im Anschluss kann anhand dieser Frage nicht nach dem höchsten Bildungsabschluss gruppiert werden. Besser wäre eine Aufteilung in zwei Fragen gewesen.

4.2.3 Pretest

Unabhängig davon, ob die Interviews anhand eines Leitfadens oder schriftlich mittels Fragebogen durchgeführt werden, empfiehlt sich ein Pretest. Mithilfe des Tests sollen die Befragungsinstrumente hinsichtlich möglicher Verständnisprobleme oder Ausfüllschwierigkeiten untersucht werden. Der Pretest sollte nach Möglichkeit mit Personen aus der zu untersuchenden Zielgruppe durchgeführt werden, die dann insbesondere eine Einschätzung zur Verständlichkeit abgeben. Die im Pretest entstehenden Antworten sowie die Teilnehmerinnen und Teilnehmer können nicht für die spätere eigentliche Auswertung genutzt werden. Können keine Personen aus der Zielgruppe befragt werden, z. B. weil die Zielgruppe klein ist, so können auch Personen mit Befragungserfahrung für einen Pretest rekrutiert werden.

Ein Pretest sollte eingesetzt werden, um zu überprüfen, wie gut die Befragung in der Praxis funktioniert.

4.3 Sammeln von Sekundärdaten

Die Datensammlung ist das Kernelement der Evaluierung und bedarf einer besonderen Sorgfalt. Insbesondere bei internen Evaluierungen sind hierzu sämtliche im eigenen Haus zuständigen Referate einzubinden.

Die Datenquelle und die Methodik der Erhebung von Daten bestimmen maßgeblich den Personal- und Zeitaufwand, der für die Evaluierung benötigt wird. Deshalb sollte geprüft werden, ob Daten in ähnlicher Qualität bereits vorliegen und für die Evaluierung genutzt werden können. Bei einer eigenen Datenerhebung muss ausreichend Zeit eingeplant werden, um z. B. Befragte im Vorfeld über die Befragung zu informieren oder ein Stichprobendesign auszuarbeiten.

Sekundärdaten haben den Vorteil, dass keine Daten kostenintensiv erhoben werden müssen.

Neben den zuvor beschriebenen Primärerhebungen können auch bereits vorhandene Daten aus seriösen Quellen genutzt und entsprechend aufbereitet werden. Mögliche Anlaufpunkte sind z. B. das [Statistische Bundesamt](#) oder [Das Datenportal für Deutschland](#) sowie ausführende Behörden. Die Datenlage ist dabei keinesfalls auf die Statistik von öffentlichen Stellen beschränkt.

Auch über Verbände können Sekundärdaten gewonnen werden.

Zu bestimmten Fragestellungen kann es z. B. notwendig oder von Vorteil sein, Daten von Verbänden, Kammern oder ähnlichen Einrichtungen abzurufen. Dies gilt insbesondere für Daten, die von der amtlichen Statistik nicht erhoben werden. Daten von solchen externen Institutionen sind insbesondere vor dem Hintergrund der angewendeten Erhebungsmethode und den verwendeten (Ausgangs-)Fragestellungen auf Stichhaltigkeit und Konsistenz zu prüfen. Wurden bei der Datenerhebung z. B. Fragestellungen mit einer eher suggestiven Wirkung verwendet, kann dies zu einer mangelnden Eignung der Daten für eine sekundäre Auswertung im Rahmen einer Evaluierung führen. Auch auf eine hinreichende Repräsentativität ist zu achten. Aus diesem Grund sind nach Möglichkeit nicht nur die Daten selbst, sondern auch die Unterlagen über die Datenerhebung zu beschaffen und zu betrachten.

Wichtig ist, dass die Daten spätestens mit dem Abschluss der Evaluierung öffentlich zugänglich gemacht werden. Liegen dem Ressort Daten vor, die nicht für die Veröffentlichung vorgesehen sind, dann sollen diese keinen Eingang in die Evaluierung finden¹¹.

Bei Sekundärdaten ist immer zu prüfen, inwieweit die Daten und die verwendeten Definitionen zur eigenen Evaluierung passen.

Ein besonderes Augenmerk gilt den möglichen definitorischen Einschränkungen seitens der datenerhebenden Stellen. Beispiel: Für den durchschnittlichen Umsatz von Unternehmen sollen die Daten aus der Umsatzsteuerstatistik genutzt werden. Die Umsatzsteuerstatistik ist eine amtliche Statistik und damit eine seriöse Quelle. Bei der Nutzung der Daten müssen die Besonderheiten der Statistik beachtet werden. Beispielsweise beinhaltet die Statistik keine Kleinunternehmer mit einem Um-

¹¹ Der Evaluierungsbericht soll „intersubjektiv überprüfbar“ sein. Daher sind nicht öffentliche Datengrundlagen zu vermeiden.

satz von weniger als 17 500 €, denn erst oberhalb dieser Schwelle wird die Umsatzsteuer erhoben. Wird für die Evaluierung der durchschnittliche Umsatz über alle Unternehmen benötigt, so ist diese Statistik nicht die geeignete Quelle.

4.4 Wirkanalyse

Die Auswertung der Daten wird als Wirkanalyse bezeichnet. Grundlage ist das zuvor aufgestellte Wirkmodell. Aus dem Modell heraus wurde der Informationsbedarf ermittelt und die erforderlichen Daten gesammelt.

Die Wirkanalyse ist die Auswertung der Daten anhand des Wirkmodells.

Die Wirkanalyse überprüft, ob die gesammelten Daten das theoretische Wirkmodell bestätigen oder ob alternative Hypothesen zutreffend sind. Quantitative Daten müssen entsprechend statistisch aufbereitet und mit geeigneten Methoden analysiert werden¹². Das Spektrum der möglichen Verfahren geht von der Darstellung von Anteilen oder Mittelwerten bis hin zu komplexen statistischen Verfahren wie beispielsweise Regressionsanalysen.

Zur Veranschaulichung geeigneter statistischer Methoden, die hier auf einer Befragung basieren, dient wieder das bereits verwendete Beispiel zur E-Mobilität: Zur Evaluierung sind 1 000 Personen zufällig für eine Befragung ausgewählt worden, die eine Ladesäule mithilfe des Förderprogramms gekauft haben. In der Evaluierung gaben 80 % der Befragten an, dass sie ohne das Förderprogramm keine Ladesäule und folglich auch kein E-Fahrzeug gekauft hätten (Anteile). Zusätzlich sollten die Befragten angeben, wie sie auf einer Skala von 1 bis 10 die Kompliziertheit der Antragstellung bei der Förderbank bewerten (1 = sehr unkompliziert und 10 = sehr kompliziert). Der Durchschnitt über alle Befragten lag bei 7,8. Die Antragstellung war also für viele der Antragstellerinnen und -steller kompliziert, sodass im Rahmen der Evaluierung bereits Verbesserungspotenzial identifiziert werden konnte. Eine Regression nimmt nun mehrere Faktoren gleichzeitig in den Blick. Beispielsweise soll durch eine Regression untersucht werden, welche Faktoren beeinflussen, dass Befragte die Antragstellung als kompliziert einschätzen. Das Ergebnis der Regression in diesem fiktiven Beispiel könnte wie folgt aussehen: Unter sonst gleichen Umständen beurteilen Befragte, die ein kostenlos angebotenes Beratungstelefonat zur Antragstellung wahrgenommen haben, die Antragstellung als nicht so kompliziert, wie Befragte die sich lediglich online informiert haben.

Qualitative Daten aus Interviews und offenen schriftlichen Fragen müssen danach untersucht werden, ob sie das Wirkmodell unterstützen. Die Aufbereitung qualitativer Daten kann sehr zeitaufwändig sein. Bei den qualitativen Verfahren ist es bedeutsam, dass ein umfassendes Bild abgegeben wird. Einzelne Wortmeldungen, die das Wirkmodell bestätigen, sind demnach nicht aussagekräftig.

¹² Die geeignete Methode ist einzelfallabhängig. Das Statistische Bundesamt berät dazu: evaluierung@destatis.de.

Ziel einer Evaluierung ist es, ergebnisoffen Handlungsempfehlungen zu identifizieren, um zukünftig die Rechtsetzung zu optimieren.

4.5 Handlungsempfehlungen

Nachdem die ausgewerteten Daten dem Wirkmodell zugeordnet wurden, können aus den Ergebnissen Rückschlüsse gezogen werden. Das Ziel der Evaluierung ist ausdrücklich **nicht**, das Wirkmodell zu bestätigen. Die Evaluierung ist ergebnisoffen. Beispiel: Wird der erwartete Output in der Praxis nicht erzielt, wohl aber das Outcome, dann wirken vielleicht weitere Wirkkanäle, die bisher nicht bekannt waren. Aus qualitativen Interviews lassen sich mitunter Ansätze für weitere Wirkmechanismen finden.

Aus dem Abgleich der Ergebnisse mit dem Wirkmodell können interessante Erkenntnisse und daraus Handlungsempfehlungen für zukünftige Vorhaben und/oder Anpassungen gewonnen werden. Die Handlungsempfehlungen sind explizit im Evaluierungsbericht zu nennen.

4.6 Beteiligung und Qualitätssicherung

Die Evaluierung und die daraus ggf. entstehenden Handlungsempfehlungen sollten in einem Evaluierungsbericht verständlich und nachvollziehbar dokumentiert werden. Der Bericht ist so transparent zu halten, dass alle Aspekte der Evaluierung für außenstehende Personen verständlich sind. Dazu gehört insbesondere die Beschreibung des Evaluierungsgegenstandes inklusive der gesetzten Ziele und dem beabsichtigten Wirkmodell, die Auswahl, Erhebung und Auswertung von Daten sowie die daraus gezogenen Schlussfolgerungen. Die nachfolgende Aufzählung ist insofern lediglich beispielhaft und nicht abschließend:

- Benennung des Gegenstandes, Anlasses und des Ziels der Evaluierung
- Benennung der Prüfkriterien, Annahmen und des Umfangs der Evaluierung
- Zeitraum der Durchführung der Evaluierung
- Benennung der Datengrundlage und der Datenquellen
- Erläuterung der Datenerhebung / Methodik
- Erläuterung und Bewertung der erhobenen bzw. verwendeten Daten
- Ergebnisse der Evaluierung / ggf. weitergehende Analyse / Bewertung
- Schlussfolgerungen (des Ressorts)

Interne Evaluierungsberichte werden auf Beschluss des Staatssekretärsausschusses für Bessere Rechtsetzung und Bürokratieabbau vom 26. November 2019 grundsätzlich vor ihrer Veröffentlichung von einer fachlich unabhängigen Stelle auf ihre Qualität überprüft. Die Prüfung sollte sich darauf beziehen, ob in der Evaluierung nachvollziehbar und plausibel auf die im Gesetzentwurf getroffenen Annahmen hinsichtlich der Ziele und der Wirkungen eingegangen wurde und welche Daten dafür herangezogen wurden. Der Nationale Normenkontrollrat (NKR) bietet sich für diese Aufgabe an. Die Einbindung der Qualitätssicherung sollte bereits zur Planung des Evaluierungsvorhabens erfolgen.

Interne Evaluierungsberichte müssen auf ihre Qualität durch eine unabhängige Stelle überprüft werden. Der Nationale Normenkontrollrat (NKR) bietet sich für diese Aufgabe an.

Der Evaluierungsbericht wird grundsätzlich auf der zentralen Plattform der Regierung veröffentlicht (siehe Kapitel 6).

5 Durchführung einer externen Evaluierung

Auch eine externe Evaluierung sollte rechtzeitig geplant werden.

Grundvoraussetzung für eine erfolgreiche externe Evaluierung ist die Planungsphase (siehe Kapitel 3). In der Regel wird die externe Evaluierung über eine Ausschreibung vergeben. Für diesen Zweck ist eine Leistungsbeschreibung zu erstellen, die die Evaluierung zu vergeben und zu begleiten. Vor der Vergabe müssen die Haushaltsmittel angemeldet werden. Das liegt zeitlich weit vor der Vergabe. An fehlenden Mitteln können externe Evaluierungen scheitern oder diese stark verzögern. Die konkreten Anforderungen für ein solches Verfahren werden in der Regel ressortintern geregelt sein.

5.1 Erstellung einer Leistungsbeschreibung und die Vergabe der Evaluierung

Eine detaillierte Leistungsbeschreibung, die das Wirkmodell, das Untersuchungsdesign, vergaberrechtliche Vorgaben sowie Hinweise auf mögliche Daten enthält, muss erstellt werden.

Die Leistungsbeschreibung soll einen Überblick über das Evaluierungsvorhaben schaffen und den hierfür nötigen Rahmen vorgeben. Dieser Abschnitt soll die spezifischen Besonderheiten einer Leistungsbeschreibung für die Evaluierungen darstellen. Weitere vergaberrechtliche Vorgaben sind bei der Vergabe zu beachten.

Zu Beginn sollte die Leistungsbeschreibung über den Hintergrund und den Sachstand der Evaluierung bzw. der Rechtsnorm informieren. Zusätzlich sollte sie das (Grob-)Konzept der Evaluierung vorstellen. Dazu gehört insbesondere das Wirkmodell, das die logische Wirkung der Rechtsnormen darstellt. Damit einhergehend sind die Ziele und Betroffenen der Rechtsnorm explizit zu benennen. Außerdem sollten bereits verfügbare Daten genannt und der zeitliche Rahmen der Evaluierung abgesteckt werden.

Das gewählte Untersuchungsdesign (siehe Kapitel 3.5) sollte, wenn bereits festgelegt, in die Leistungsbeschreibung aufgenommen werden. Dazu zählen beispielsweise durchzuführende Befragungen oder Auswertungen von Sekundärdaten. Ein transparenter Meilensteinplan ermöglicht eine bessere Begleitung der späteren Evaluierung. Beispiele für Meilensteine: Fragebogenerstellung, Durchführung der Befragung, Zwischenbericht, Endbericht. Ein Zwischenbericht empfiehlt sich insbesondere für größere Evaluierungsvorhaben.

Nach Möglichkeit sollte der finanzielle Rahmen der Evaluierung grob abgesteckt sein. Allerdings sollten die Kosten nicht als ausschlaggebendes Kriterium genannt werden. Besonders der Evaluierungskompetenz der externen Evaluierenden kommt ebenfalls eine große Bedeutung zu.

Die eingegangenen Angebote sollten in einem ersten Schritt überprüft werden, ob sie die Vorgaben der Ausschreibung erfüllen. Die letztendliche Auswahl eines Anbieters oder eines Konsortiums sollte unter Einbeziehung von Personen mit Methodenkenntnis erfolgen.

5.2 Einhaltung der Vorgaben der Leistungsbeschreibung und anerkannter Qualitätsstandards

Bevor die Evaluierung durchgeführt wird, kann es sinnvoll sein, ein Team zur Projektbegleitung zu bilden. Handelt es sich um eine weniger komplexe Evaluierung, kann die Projektbegleitung auch durch nur eine Person erfolgen. Aufgabe der Projektbegleitung ist die Steuerung der Evaluierung. Die Intensität der Begleitung hängt vom Umfang der Evaluierung ab.

Bei einem ersten Treffen sollten die Evaluierenden alle notwendigen Informationen erhalten. Hierzu gehören beispielsweise auch vorherige Evaluierungsberichte, interne Selbstevaluierungen oder Prozessdaten. Das Auftakttreffen dient auch dazu, die weitere Vorgehensweise abzusprechen. Allerdings dürfen die Auftraggebenden dabei keinen einseitigen Einfluss auf die Evaluierung und/oder den Evaluierungsbericht nehmen. Fortan sollte das Projektteam die fortschreitende Evaluierung anhand der Meilensteine begleiten und den Evaluierenden für Rückfragen zur Verfügung stehen. Es empfiehlt sich auch, die Erhebungsinstrumente seitens des Evaluierungsteams mit zu begleiten.

Die externe Evaluierung sollte durch ein Projektteam oder eine Person aus dem zuständigen Ressort begleitet werden.

5.3 Stellungnahme zu Handlungsempfehlungen

Der externe Evaluierungsbericht sollte die Handlungsempfehlungen streng von der empirischen Analyse trennen. In der Regel werden mögliche Handlungsoptionen in einem eigenen Kapitel dargestellt. Die Ressorts nehmen zu der Frage Stellung, welche Schlussfolgerungen beziehungsweise weiteren Vorgehensweisen sie aus den Evaluierungsergebnissen ziehen¹³. Diese Stellungnahme wird mit dem Evaluierungsbericht veröffentlicht.

6 Die Veröffentlichung des Evaluierungsberichts

Die Evaluierungsberichte sind für verschiedene Zielgruppen von Interesse. In erster Linie werden die vom jeweiligen Regelungsvorhaben Betroffenen und deren Interessenvertretungen wie Verbände wissen wollen, wie dessen Auswirkungen sind. Aber auch für andere im Verlauf des Gesetzgebungsprozesses Beteiligte, den Bundestag, die Länder und die kommunalen Spitzenverbände, sind die Evaluierungsberichte möglicherweise von Interesse.

Nach dem Beschluss von 2019 werden die Evaluierungsberichte und die Stellungnahmen der Bundesregierung hierzu auf einer zentralen Online-Plattform der Bundesregierung veröffentlicht. Darüber hinaus müssen die Evaluierungsberichte an die betroffenen Ressorts, BKAm (Referat 613) und an den NKR übersandt werden.

Evaluierungsberichte und die Stellungnahmen der Bundesregierung hierzu werden auf einer zentralen Online-Plattform der Bundesregierung veröffentlicht.

¹³ ST-Beschluss Bessere Rechtssetzung und Bürokratieabbau vom 26.11.2019.

7 Der Umgang mit den Ergebnissen

Die Verwertung der Evaluierungsergebnisse ist wichtig.

Der Beschluss von 2019 sieht vor, dass die Ressorts zur Frage Stellung nehmen, welche Schlussfolgerungen beziehungsweise weiteren Vorgehensweisen sie aus den Evaluierungsergebnissen ziehen. Die Ergebnisse der Evaluierungsberichte (idealerweise Handlungsempfehlungen) sollten von den Ressorts mindestens auf ihre Umsetzbarkeit hin geprüft werden. Für einen hohen Wirkungsgrad der Handlungsempfehlungen und Transparenz sollte diese Prüfung dokumentiert und Ablehnungen begründet werden („accept or explain“).

Verschiedene Maßnahmen kommen in Betracht, um eine Nutzung zu gewährleisten.

Der Wert der Evaluierungsergebnisse, die unter Umständen mit einem nicht geringen Aufwand gewonnen werden, steht und fällt mit deren Nutzung, z. B. bei künftigen Regelungsvorhaben. Zeitfortschritt, Personalfluktuationen, aber auch fehlende Kenntnis der Hausleitung können dazu führen, dass Evaluierungsergebnisse nicht genutzt werden. Als Maßnahmen, um eine Nutzung der Ergebnisse zu gewährleisten, kommen in Betracht:

- eine regelmäßige Wiederholung der Prüfung auf Umsetzbarkeit,
- Anlegen eines Wissensspeichers im Referat oder
- eine Leitungsvorlage, die über die Ergebnisse der Evaluierung informiert und ggfs. Handlungen empfiehlt.

8 Anhang

Anhang I Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 23. Januar 2013

Vom St-Ausschuss Bürokratieabbau am 23.01.2013 beschlossene Fassung

Konzeption zur Evaluierung neuer Regelungsvorhaben

gem. Arbeitsprogramm bessere Rechtsetzung der BReg vom 28. März 2012, Ziffer II. 3.

Eine Evaluierung stellt einen Zusammenhang her zwischen Ziel und Zweck einer Regelung und den tatsächlich erzielten Wirkungen sowie den damit verbundenen Kosten. Sie ist für alle wesentlichen Regelungsvorhaben nach Maßgabe der folgenden Leitlinien vorzusehen.

I In welchen Fällen wird evaluiert?

Die Wesentlichkeit eines Regelungsvorhabens im Sinne des Arbeitsprogramms bessere Rechtsetzung bemisst sich grundsätzlich nach der Höhe des zu erwartenden jährlichen Erfüllungsaufwands (Schwellenwert). Damit wird der Tatsache Rechnung getragen, dass eine Evaluierung auch einen Beitrag dazu leisten soll, den Erfüllungsaufwand dauerhaft niedrig zu halten.

Als wesentlich gelten danach Regelungsentwürfe, bei denen ein jährlicher Erfüllungsaufwand von mindestens

- 1 Mio. € Sachkosten oder 100.000 Stunden Aufwand für Bürgerinnen und Bürger oder
- 1 Mio. € für die Wirtschaft oder
- 1 Mio. € für die Verwaltung

aufgrund der Ex ante- Abschätzung zu erwarten ist oder – ist eine solche Abschätzung nicht möglich – nicht ausgeschlossen werden kann.

Anlass für eine Evaluierung kann auch die Nachmessung des Erfüllungsaufwands sein – wenn erst dann festgestellt wird, dass der tatsächliche Aufwand einen der genannten Schwellenwerte überschreitet. Ergibt die Nachmessung, dass der Schwellenwert unterschritten wird, kann das Ressort seine im Regelungsentwurf getroffenen Evaluierungserwägungen überprüfen.

Neben der Höhe des jährlichen Erfüllungsaufwands als Anlass für eine Evaluierung können die Ressorts – wie bisher auch – aus anderen Gründen eine Evaluierung vorsehen – wie z.B. bei einem hohen finanziellen Gesamtaufwand, besonderer politischer Bedeutung oder großen Unsicherheiten über Wirkungen oder Verwaltungsvollzug.

Verhältnismäßigkeit

Der Aufwand für die Evaluierung muss in einem angemessenen Verhältnis zu den daraus zu gewinnenden Erkenntnissen stehen. Dies gilt zum Beispiel im Hinblick auf:

- gerichtliche Entscheidungen oder internationale bzw. EU-Vorgaben ohne Umsetzungsspielraum (1:1);
- anderweitige vergleichbare Berichtspflichten – auch gegenüber oder von internationalen bzw. EU-Institutionen – oder eine gesetzlich vorgeschriebene Wirkungsforschung als Daueraufgabe.

Bei allen wesentlichen Regelungsvorhaben sind in der Begründung zum Regelungsentwurf gem. § 44 Abs. 7 GGO Aussagen zur Durchführung oder Nicht-Durchführung von Evaluierungen aufzunehmen.

II. Was wird evaluiert?

Wichtigstes Evaluierungskriterium ist die Zielerreichung, da Regelungen, die das im Regelungsentwurf formulierte Ziel verfehlen, meist unnötigen Erfüllungsaufwand verursachen.

Weitere Prüfkriterien können – je nach Umfang der Evaluierung – sein:

- Nebenfolgen der Regelung (positive oder negative);
- Akzeptanz der Regelung (z. B. Inanspruchnahme staatlicher Angebote);
- Praktikabilität der Regelung, die im Hinblick auf vermeidbaren Erfüllungsaufwand zu prüfen ist (z. B. Bündelung von Aufgaben bei einer Vollzugsbehörde);
- Abwägungen, ob die entstandenen Kosten in einem angemessenen Verhältnis zu den Ergebnissen stehen. Dies muss nicht zwangsläufig eine mo-

...

- 3 -

netäre Betrachtung sein, sondern kann auch als Abwägung von Vor- und Nachteilen einer Regelung erfolgen. Politische Entscheidungen sollen dabei nicht vorweggenommen werden.

III. Wie wird evaluiert?

Eine Evaluierung ist hinsichtlich ihrer Tiefe (z. B. Regelungsvorhaben insgesamt, Teile eines Artikelgesetzes, einzelne Bereiche eines Stammgesetzes, Vollzug), der Methodik (von interner bis hin zu einer wissenschaftlichen Evaluierung) und des Umfangs (von „Zwei-Seiten-Bericht“ bis hin zu einem ausführlichen Bericht – auch abhängig von den zur Verfügung stehenden Ressourcen) nicht vorbestimmt. Diese Entscheidungen obliegen dem jeweils federführenden Ressort. Die Evaluierung kann auf den Ergebnissen der Nachmessung des Erfüllungsaufwands durch das Statistische Bundesamt aufbauen.

Die Dokumentation der Ergebnisse soll mittels Evaluierungsbericht erfolgen, der Aussagen zur Abgrenzung des Untersuchungsgegenstands, zu den zugrundeliegenden Daten und Annahmen sowie zu den relevanten Prüfkriterien enthält. Der Bericht soll in anschaulicher Form die wesentlichen Erkenntnisse der Evaluierung darstellen.

Diese Berichte werden den betroffenen Ressorts sowie dem Koordinator der Bundesregierung für Bürokratieabbau und bessere Rechtsetzung im Bundeskanzleramt und dem Nationalen Normenkontrollrat zur Kenntnis gegeben.

IV. Wann wird evaluiert?

Das federführende Ressort entscheidet nach fachlichem Ermessen über den Zeitpunkt der Evaluierung. Bei der Wahl eines geeigneten Zeitpunkts sind der erwartete Eintritt von Wirkungen sowie ggfs. weitere Änderungen des regulatorischen Umfelds zu berücksichtigen. I.d.R. soll eine Evaluierung 3-5 Jahre nach Inkrafttreten eines Regelungsvorhabens durchgeführt werden.

Dieses Verfahren findet auf Regelungsvorhaben Anwendung, die ab 1. März 2013 in die Ressortabstimmung eingebracht werden.

Anhang II Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 26. November 2019

Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 26. November 2019

Fortentwicklung der Evaluierungskonzeption der Bundesregierung

Der Staatssekretärsausschuss konkretisiert die Konzeption zur Evaluierung neuer Regelungsvorhaben vom 23. Januar 2013 wie folgt:

1. In der Begründung des Regelungsvorhabens soll neben den Angaben nach § 44 Abs. 7 GGO knapp dargestellt werden, welche Ziele bei der Evaluierung zugrunde gelegt werden und welche Kriterien für die Zielerreichung dabei voraussichtlich herangezogen werden. Dies ermöglicht es, Vorsorge dafür zu treffen, dass zum Zeitpunkt der Evaluierung die voraussichtlich erforderlichen Daten zur Verfügung stehen.
2. Soweit Vorschriften, die bereits Gegenstand einer Evaluierung waren, in neuen Regelungsvorhaben geändert werden, stellen die Ressorts dies in der Begründung dar, wenn die Ergebnisse der Evaluierung Eingang in das Vorhaben gefunden haben.
3. Die Bundesregierung erstellt eine Arbeitshilfe, in der Arbeitsschritte und Methoden einer Evaluierung (u.a. Fragestellung der Evaluierung, Datenerhebung und -auswertung, Qualitätssicherung, Aufbereitung der Ergebnisse im Evaluierungsbericht) dargestellt sind.
4. Die Ressorts sollen eine zentrale Arbeitseinheit als Ansprechpartner für Evaluierungen von Regelungsvorhaben intern benennen.
5. Interne Evaluierungsberichte werden grundsätzlich vor ihrer Veröffentlichung durch eine unabhängige Stelle hinsichtlich ihrer Qualität überprüft. Bei Regelungsvorhaben, bei denen der jährliche Erfüllungsaufwand einen Wert von 5 Millionen Euro überschreitet, findet eine Qualitätssicherung stets statt. Die Prüfung sollte sich darauf beziehen, ob in der Evaluierung nachvollziehbar und plausibel auf die im Gesetzentwurf getroffenen Annahmen hinsichtlich der Ziele und der Wirkungen eingegangen wurde und

...

- 2 -

welche Daten dafür herangezogen wurden. Der NKR bietet an, diese Qualitätssicherung durchzuführen.

6. Bei Evaluierungen, die den Erfüllungsaufwand zum Gegenstand haben können, binden die Ressorts das Statistische Bundesamt (StBA) frühzeitig in die Planung von Evaluierungen ein, um eine sinnvolle Verzahnung der Nachmessung des Erfüllungsaufwands und der Evaluierung sicherzustellen. Beim StBA wird eine Kompetenzstelle eingerichtet, die die Ressorts bei Bedarf zur Planung und Durchführung von Evaluierungen berät.
7. Im Rahmen ihrer Weiterbildungsstrategie (Ziffer I.5 des Arbeitsprogrammes Bessere Rechtsetzung und Bürokratieabbau 2018) stellt die Bundesregierung Schulungsangebote für Mitarbeiterinnen und Mitarbeiter der Ressorts zur Verfügung, die Evaluierungen durchführen.
8. Die Ressorts nehmen zur Frage Stellung, welche Schlussfolgerungen beziehungsweise weiteren Vorgehensweisen sie aus den Evaluierungsergebnissen ziehen.
9. Im Rahmen der Evaluierung sollen die Ressorts in geeigneter Weise Länder, kommunale Spitzenverbände, Fachkreise und Verbände soweit betroffen zur Frage der Zielerreichung und ggfs. auch zu den in der Konzeption genannten weiteren Prüfkriterien einbinden. Sie berücksichtigen dabei auch die im Zweiten nationalen Aktionsplan der Bundesregierung im Rahmen der Teilnahme an der Open Government Partnership enthaltene Zielsetzung, die Einbindung von Bürgerinnen und Bürgern und Vertreterinnen und Vertretern der Zivilgesellschaft in das Regierungshandeln zu stärken.
10. Evaluierungen und Stellungnahmen der Bundesregierung nach Ziffer 8 werden grundsätzlich auf einer zentralen Online-Plattform der Bundesregierung veröffentlicht.

9 Impressum

Herausgeber

Statistisches Bundesamt, Wiesbaden
im Auftrag der Bundesregierung

Kontakt

Referat 613 (Bessere Rechtsetzung; Geschäftsstelle Bürokratieabbau)

Bundeskanzleramt
Willy-Brandt-Straße 1
10557 Berlin

↳ buerokratieabbau@bk.bund.de
↳ www.bundesregierung.de/buerokratieabbau

Statistisches Bundesamt

Gruppe I2 – Dienstleistungszentrum der Bundesregierung für Bessere Rechtsetzung
Gustav-Stresemann-Ring 11
65189 Wiesbaden

↳ evaluierung@destatis.de
↳ www.destatis.de/buerokratiekosten

Informationen zur amtlichen Statistik

Internet: www.destatis.de

Zentraler Auskunftsdienst
+49 (0) 611 / 75 24 05

↳ <http://www.destatis.de/kontakt>

Stand

Erste Fassung, erschienen im Januar 2022 auf Grundlage des Beschlusses des Staatssekretärsausschusses vom 26. November 2019

Copyright

© Statistisches Bundesamt (Destatis), 2022
(im Auftrag der Bundesregierung)
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe
gestattet.

Bildrechte

Titelbild: © ipopba / stock.adobe.com / 97169647 / eigene Bearbeitung
Seite 6: © artjazz / Fotolia.com / 93494284 / eigene Bearbeitung
Seite 14: © Michael Flippo / stock.adobe.com / 9753939 / eigene Bearbeitung

Herausgeber

Statistisches Bundesamt, Wiesbaden
im Auftrag der Bundesregierung

Ihr Kontakt zu uns:

Referat 613 (Bessere Rechtsetzung;
Geschäftsstelle Bürokratieabbau)
Bundeskanzleramt
Willy-Brandt-Straße 1
10557 Berlin

buerokratieabbau@bk.bund.de
www.bundesregierung.de/buerokratieabbau

Statistisches Bundesamt
Gruppe I2 – Dienstleistungszentrum der Bundesregierung
für Bessere Rechtsetzung
Gustav-Stresemann-Ring 11
65189 Wiesbaden

erfuellungsaufwand@destatis.de
www.destatis.de