

Vera Herrmann

Die öffentliche Statistik als Gegenstand internationaler Peer Reviews

Die wirkungsorientierte Verwaltungsführung verlangt nach Evaluationen, und diese verlangen nach öffentlichen Statistiken. Folglich stellt sich die Frage nach deren Qualität. Hierfür gibt es einen Verhaltenskodex, den «European Statistics Code of Practice» (EU CoP), der mittels Peer Reviews überprüft wird. Nachdem sich der Kreis der Länder, die an einer Überprüfung ihrer statistischen Governance interessiert sind, erweitert hat, wurden «Global Assessments» als Evaluierungsinstrument entwickelt und durchgeführt. Diese Evaluationsart wird nachfolgend dargestellt und unter Bezug auf Erkenntnisse aus der Praxis reflektiert.

Categoria di articoli: Contributi scientifici

Citazione: Vera Herrmann, Die öffentliche Statistik als Gegenstand internationaler Peer Reviews, in: LeGes 30 (2019) 1

Inhaltsübersicht

1. Gute Regierungsführung und die Nachfrage nach Daten und Evaluationen
2. Die Rolle der öffentlichen Statistik und deren Code of Conduct
3. Peer Reviews in der öffentlichen Statistik
 - 3.1. Peer-Evaluationen im internationalen Kontext
 - 3.2. Ablauf der Peer Reviews im Bereich der öffentlichen Statistik
4. Erfahrungsbericht und Reflexion über die Global Assessments als spezielle Form einer Peer-Evaluation

1. Gute Regierungsführung und die Nachfrage nach Daten und Evaluationen

[Rz 1] Es gibt heute kaum noch Staaten oder internationale Organisationen, die sich nicht der guten Regierungsführung verpflichtet fühlen. Eine einheitliche Definition dessen, was eine gute Regierungsführung ist, gibt es meines Wissens nicht, obwohl sich zwischenzeitlich ein Forschungszweig zu diesem Thema entwickelt hat und nicht nur internationale Organisationen oder Staaten und deren Ämter und Dienststellen, sondern auch Firmen und Berufsbrachen sich mit *Codes of Good Governance* verpflichten und *Codes of Conduct* entwickelt haben. In einer weltweiten Studie wollen AGUILERA & CUERVO-CAZURRA schon 2008 in 64 Ländern 196 verschiedene solcher Codes gezählt haben; es dürften heute einige mehr sein.

[Rz 2] Mangels klarer Definition versteht man unter guter Regierungsführung eher eine Sammlung von Methoden und Instrumenten. In Bezug auf *Good Governance* eines Staatswesens besteht (noch) Einigkeit darin, dass das Regierungs- und Verwaltungshandeln sich an Fakten orientieren, also evidenzbasiert sein soll. Wenn folglich Regierungs- und Verwaltungshandeln «auf der Basis empirisch zusammengetragener und bewerteter wissenschaftlicher Erkenntnisse erfolgreich»¹ ausgestaltet werden soll, bedarf es Daten für Monitoring und Evaluation.

[Rz 3] Im Hinblick auf Daten, die der Transparenz, Überprüfbarkeit und Vergleichbarkeit wegen möglichst frei zugänglich sein sollten, liefert die öffentliche Statistik (annähernd²) wertfreie Fakten und statistische Analysen von (in der Regel) guter Qualität, die es der Verwaltung als Vollzugs- und Umsetzungsorgan ermöglichen, politische Massnahmen zielgerichtet zu entwickeln, zu überprüfen, zu evaluieren und gegebenenfalls anzupassen. Die öffentliche Statistik hat folglich sehr wohl eine politische Relevanz, denn deren Ergebnisse sollten steuerungsrelevant sein; hingegen muss sie bei deren Ausrichtung und Methodenwahl – im Sinne einer guten Regierungsführung – autonom und frei von Einflüssen sein. In der Schweiz veröffentlichen Bund, Kantone, Gemeinden und weitere Organisationen mit einem staatlichen Auftrag ihre offenen Daten auf dem Datenportal <https://opendata.swiss>.³

¹ Definition gemäss Duden online: <https://www.duden.de/rechtschreibung/evidenzbasiert>.

² Auch die öffentliche Statistik arbeitet mit methodischen, mitunter multidimensionalen Konzepten (wie z. B. bei der Wohlfahrtsmessung) und statistischen Konstrukten (z. B. bei der Bildung von Indikatoren wie die Armutsquote), denen Annahmen oder Werte zugrunde liegen, die gesellschaftspolitisch begründet sind. Ich vertrete daher die Auffassung, dass es selbst in der öffentlichen Statistik weder eine absolute Wertfreiheit noch eine absolute Exaktheit gibt. Entscheidend sind infolgedessen die Einhaltung von Gütekriterien der öffentlichen Statistik und des Code of Conduct, wovon hier die Rede ist.

³ opendata.swiss vereint unterschiedliche Datensätze wie beispielsweise die Gemeindegrenzen der schweizerischen Bevölkerungsstatistiken, aktuelle Wetterdaten, historische Dokumente oder ein Verzeichnis der Schweizer Literatur. Gemeinsam ist den Datensätzen, dass sie alle kostenlos heruntergeladen und weiterverwendet werden können.

[Rz 4] Monitoring meint «die laufende und systematische Sammlung von Informationen über Umfang und Richtung der Veränderungen im interessierenden Handlungs- und Politikfeld» (BALTHASAR 2005: 68). Es handelt sich um eine Art «Dauerbeobachtung», deren Zweck und Ziel es ist, in regelmässigen Zeitabständen zu überprüfen, ob sich der «Beobachtungsgegenstand» in die gewünschte – in der Regel mittels Zielwerten definierte – Richtung bewegt. Gemäss BALTHASAR kann ein Monitoring, da es nur beobachtet, ohne Beurteilungssysteme auskommen. Anders die Evaluationen, die im wörtlichen Sinne⁴ auf eine Bewertung eines beobachtbaren, zu analysierenden Sachverhalts ausgerichtet sind. Die Evaluationswissenschaft, die sich ab den 1960er-Jahren im angelsächsischen und ab den 1980er-Jahren im deutschsprachigen Raum etablierte, unterscheidet zahlreiche Evaluationsarten in Abhängigkeit ihrer Funktion (respektive ihres Zwecks) oder ihrer Form. Zur Einführung oder als Überblick sei stellvertretend auf das Lehrbuch von BANK & LAMES (2010) verwiesen. Im Wesentlichen ist eine Evaluation eine empirische Studie, die die Frage klärt, ob und inwieweit eine Massnahme geeignet erscheint, einen angestrebten Zweck zu erfüllen. Sie dient, je nach gewähltem Untersuchungsansatz, der rückblickenden Wirkungskontrolle oder der vorausschauenden Steuerung. Eine besondere Evaluationsform sind Peer-Evaluationen.

2. Die Rolle der öffentlichen Statistik und deren Code of Conduct

[Rz 5] Ungeachtet ihrer Funktion und Form: Kaum eine Evaluation kommt ohne Daten aus; entweder werden diese im Zuge des Evaluationsprozesses systematisch erhoben und sind integrierter Teil der Bewertung, oder man bedient sich öffentlich zugänglicher Daten.

[Rz 6] Damit kommt der öffentlichen Statistik eine besondere Bedeutung im Kontext von Überprüfung oder Bewertung von Regierungs- und Verwaltungsführung zu: Sie soll statistische Informationen, ergänzt durch Indikatorensysteme (für ein Monitoring) und Analysen zum Zustand sowie Szenarien zur Entwicklung von Bevölkerung, Wirtschaft, Gesellschaft, Raum und Umwelt liefern, die der Meinungsbildung in der Bevölkerung, der politischen Entscheidungsfindung und der Wirkungskontrolle staatlichen Handelns dienen.⁵ Dabei gilt es, die weltweit gültigen Prinzipien statistischer Tätigkeit zu beachten.

[Rz 7] Es besteht Konsens, dass die Erhebung, Aufbereitung und Verbreitung der statistischen Informationen in fachlich unabhängiger Weise erfolgen soll, um nicht eine von bestimmten Interessengruppen beeinflusste und dadurch verzerrte Darstellung der realen Wirklichkeit zu erhalten. Versuche der Einflussnahme gibt es auch hierzulande. Es besteht jedoch nicht immer Einigkeit über den «richtigen» Indikator, der zur Messung eines Phänomens und als Richtwert für die Wirkungskontrolle herangezogen wird.

[Rz 8] Als Schutz vor Einflussnahme hat sich die internationale Statistikgemeinde einen spezifischen Code of Conduct gegeben: Die hierfür geltenden grundlegenden Prinzipien der

⁴ Der Begriff «Evaluation» stammt aus dem angelsächsischen Raum und heisst übersetzt Beurteilung, Bewertung, Auswertung.

⁵ Auszug aus dem Auftrag des Bundesamts für Statistik (BFS), das sich wiederum aus dem Statistischen Mehrjahresprogramm ableitet, das für die Dauer einer Legislaturperiode unter Federführung des BFS für die gesamte Bundesstatistik erstellt wird und als langfristig orientierte strategische Steuerungs- und Planungsgrundlage gedacht ist. Siehe: <https://www.bfs.admin.ch/bfs/de/home/bfs/oeffentliche-statistik/system-oeffentliche-statistik/mehrjahresprogramm.html>.

öffentlichen Statistik wurden 1992 von der Wirtschaftskommission für Europa der Vereinten Nationen (United Nations Economic Council of Europe; UNECE) erlassen. Der Zeitpunkt war nicht zufällig, sollte doch mit diesen Prinzipien, den *UN Fundamental Principles of Official Statistics* nach dem Fall der Berliner Mauer auch ein Leitfaden für den Wiederaufbau der Statistiksysteme der osteuropäischen Staaten vorgegeben werden. Es wurden zehn Grundsätze formuliert, die 1994 von der statistischen Kommission der Vereinten Nationen genehmigt wurden und seitdem einen universellen Standard für den Statistikbereich bilden.⁶

[Rz 9] Die Konferenz der Regionalen Statistischen Ämter der Schweiz (KORSTAT) und das Bundesamt für Statistik (BFS) haben im Mai 2002, in Umsetzung der eben erwähnten UN-Prinzipien, eine Charta für die öffentliche Statistik der Schweiz beschlossen.⁷ Darin halten sie unter anderem fest, dass ihre Tätigkeit einen unerlässlichen «*Service Public*» darstelle, der den Bedürfnissen einer demokratischen Gesellschaft und damit eines modernen Staates entspricht. Relevanz, Qualität und Glaubwürdigkeit der publizierten statistischen Informationen sind Hauptziele der öffentlichen Statistik. Zwecks Förderung und Überwachung dieser Prinzipien wurde ausserdem ein Ethikrat geschaffen. Dieser berufsethische Kodex wurde 2007 unter Berücksichtigung neuerer europäischer Entwicklungen überarbeitet und 2012 ergänzt. Diese Charta wurde von allen regionalen statistischen Ämtern sowie von den meisten Bundesämtern, die eine Statistikstelle unterhalten und zusammen ein statistisches System bilden, unterzeichnet. Damit verpflichten sich diese Statistikstellen zur Einhaltung dieser Qualitätskriterien.

[Rz 10] In der Europäischen Union hat deren statistisches Amt, Eurostat, den Auftrag, die EU mit europäischen Statistiken zu versorgen, die Vergleiche zwischen Ländern und Regionen ermöglichen. Die Schweiz ist mit dem bilateralen Statistikabkommen Schweiz–EU, das zum 1. Januar 2007 in Kraft getreten ist, Mitglied des Europäischen Statistischen Systems (ESS).⁸ Dieses Statistikabkommen legt die Grundlage für eine einheitliche statistische Datenerhebung in der Schweiz und der Europäischen Union (EU) und gewährleistet die Vergleichbarkeit Schweizer Statistiken mit denjenigen aus den EU-Mitgliedstaaten in ausgewählten Bereichen, die in einem Anhang festgelegt wurden und fortlaufend ergänzt werden. Es war erklärtes Ziel dieses Abkommens, dass die Schweiz z. B. als Wirtschaftsstandort dank der Veröffentlichung europakompatibler Statistiken in den EU-Publikationen an internationaler Sichtbarkeit gewinnt.

[Rz 11] Für die Entwicklung, Erstellung und Verbreitung von europäischen Statistiken gelten ebenfalls verbindliche Grundsätze, die im Verhaltenskodex für Europäische Statistiken, dem *European Statistics Code of Practice (CoP)* festgelegt sind. Er wurde 2005 verabschiedet und zwischenzeitlich zwei Revisionen unterzogen. Er umfasst 16 ethische Grundsätze, denen sogenannte Indikatoren zugeordnet sind. Damit unterscheidet sich dieser Verhaltenskodex von den oben erwähnten *UN Fundamental Principles* und der schweizerischen Statistikcharta dahingehend, dass Kriterien für dessen Überprüfung definiert und die Grundlage für eine systematische Evaluierung eines nationalen statistischen Systems gelegt wurden.

⁶ Nähere Informationen zu den *UN Fundamental Principles of Official Statistics* finden sich auf der Webseite der UNECE Statistics Division.

⁷ Nähere Informationen hierzu finden sich auf der Webseite des Ethikrats: www.ethikrat-stat.ch/.

⁸ Abkommen vom 26. Oktober 2004 zwischen der Schweizerischen Eidgenossenschaft und der Europäischen Gemeinschaft über die Zusammenarbeit im Bereich der Statistik, SR 0.431.026.81.

[Rz 12] Vereinfacht formuliert, ist dieser CoP für die in der öffentlichen Statistik Tätigen das, was die SEVAL-Standards für die Evaluatoreninnen und Evaluatoren sind (SEVAL 2017). Sie weisen Ähnlichkeiten auf, wie Abbildung 1 verdeutlicht.

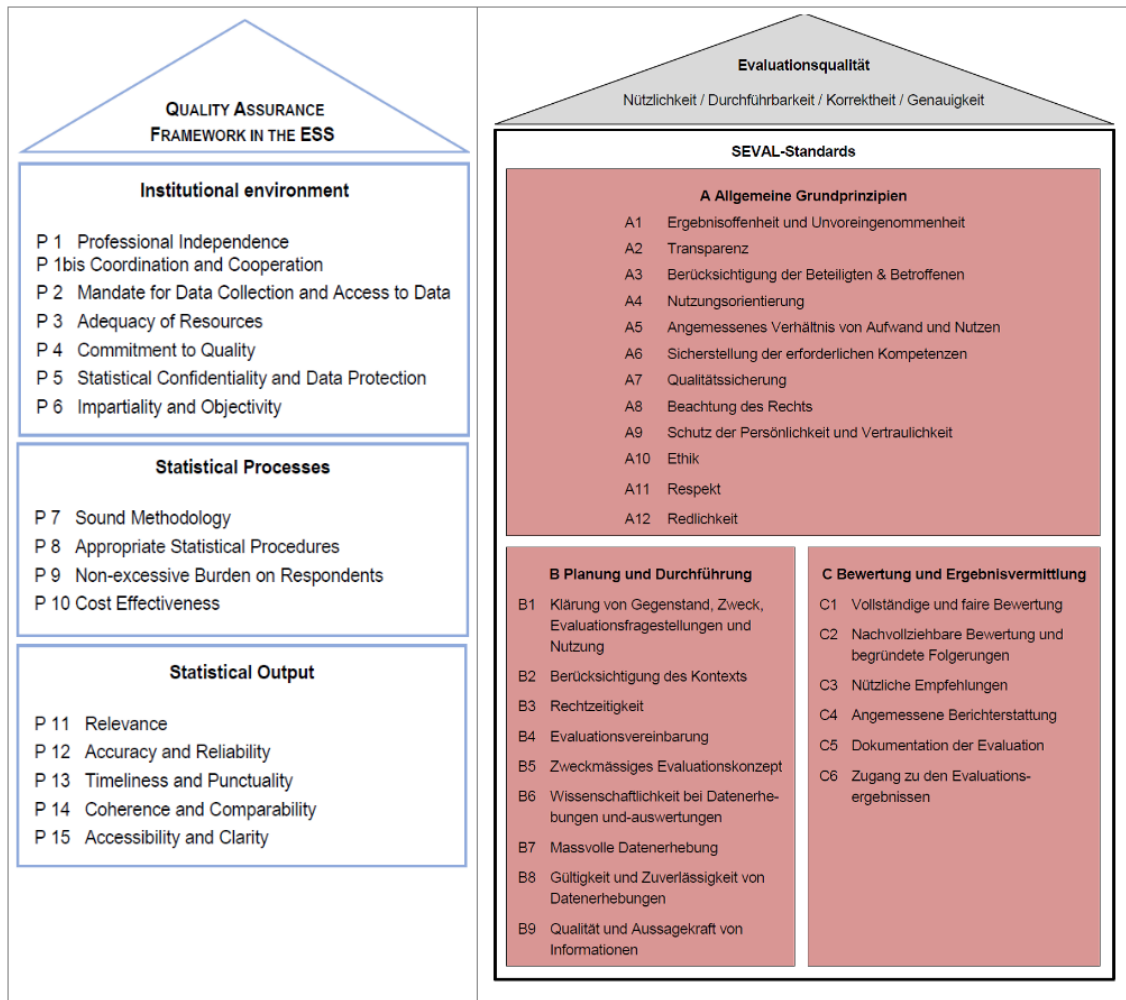


Abb. 1: ESS-Standards für die öffentliche Statistik⁹ und SEVAL-Standards für die Evaluation¹⁰

[Rz 13] Das Bekenntnis zur Qualität über den gesamten Erhebungs- bzw. Evaluationsprozess hinweg ist beiden Regelwerken gemeinsam. Sie verlangen die professionelle Unabhängigkeit bzw. eine Ergebnisoffenheit und Unvoreingenommenheit, die Sicherstellung des Mandats (für die Datenerhebung respektive für die Evaluation) in Form einer Vereinbarung und die Wissenschaftlichkeit (sound methodology) bei gleichzeitiger Wahrung der Vertraulichkeit und des Daten- und Persönlichkeitsschutzes.

⁹ Die aktuelle Ausgabe des European Statistics Code of Practice, die am 16. November 2017 verabschiedet wurde, liegt bisher nur in der englischen Sprachversion vor.

¹⁰ Erläuterungen zu den Evaluationsstandards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards), Ausgabe 2017

[Rz 14] Die Nützlichkeit, Genauigkeit und Glaubwürdigkeit einer Statistik wie auch einer Evaluation hängen von der Gültigkeit und Zuverlässigkeit der ihr zugrundeliegenden Datenerhebungen ab. So verlangt der CoP, dass die Quelldaten, Zwischenergebnisse und der statistische Output regelmässig bewertet und validiert werden, und die SEVAL-Standards verlangen, dass im Rahmen einer Evaluation Vorkehrungen getroffen werden, die die Einhaltung wissenschaftlicher Gütekriterien garantieren.

[Rz 15] Für eine Evaluation sind die Nachvollziehbarkeit der Bewertung und die Begründung der Schlussfolgerungen und Empfehlungen von grosser Wichtigkeit, für die öffentliche Statistik die Kohärenz und Vergleichbarkeit. So müssen die Statistiken auf der Grundlage von einheitlichen Standards in Bezug auf den Geltungsbereich, einheitlichen Definitionen und Klassifikationen erstellt werden und über einen ausreichenden Zeitraum betrachtet vergleichbar sein.

[Rz 16] Beiden Regelwerken ist wiederum die Forderung nach Zugang zu den Daten bzw. den Evaluationsergebnissen sowie das Bemühen um einen angemessenen Aufwand im Verhältnis zum Nutzen unter Berücksichtigung der Belastung der Befragten gemeinsam.

3. Peer Reviews in der öffentlichen Statistik

3.1. Peer-Evaluationen im internationalen Kontext

[Rz 17] Peer-Evaluationen (Peer Reviews) sind gemäss BEYWL/NIESTROJ (2009, 201) eine Evaluationsart, die sich dadurch kennzeichnet, dass «Fachleute aus einer auf gleichem Gebiet tätigen, aber externen Organisationseinheit, deren Feld- und Fachkompetenz ähnlich hoch ausgeprägt ist wie die derjenigen, die in der zu evaluierenden Einheit tätig sind (daher «peers» gleich Fachkolleginnen und Fachkollegen), eine Bewertung eines Evaluationsgegenstands vornehmen».

[Rz 18] Peer-Evaluationen haben sich insbesondere im Bildungsbereich gut etabliert: Im Zuge der Europäischen Hochschulreformen werden Peer Reviews zunehmend bei der Evaluierung von Institutionen (bzw. Teileinheiten von Institutionen wie Fakultäten, Fachbereichen, Instituten) eingesetzt. Zahlreiche Schulreformen fordern ebenfalls nach Evaluierung. Selbst- oder Peer-Evaluationen sind inzwischen als gängiges, relativ kostengünstiges Verfahren der (wie auch immer definierten) Qualitätsüberprüfung und Weiterentwicklung oder der Akkreditierung gut etabliert; so ist unter anderem das *European Network for Quality Assurance in Higher Education* (ENQA 2009, 27) unter Leitung der EU-Kommission (Generaldirektion Bildung und Kultur) übereingekommen, «that peer review of agencies should be interpreted as basically the means to achieve the goal of transparency, visibility and comparability of quality of agencies».

[Rz 19] In der internationalen Zusammenarbeit setzte die Organisation für Entwicklung und Zusammenarbeit in Europa (OECD) bzw. deren *Development Co-operation Directorate* mit dem *Development Assistance Committee (DAC)* Massstäbe. Ein Leitfaden für die Evaluierung der Entwicklungszusammenarbeit, die *Quality Standards for Development Evaluation* geben die Leitlinien für die länderspezifischen Peer Reviews und Assessments vor und entwickelten diese als vorherrschendes Evaluationsverfahren, mit dem Ziel «to assess the performance of a given member, not just that of its development co-operation agency, and examine both policy and implementation» (siehe: *DAC - Peer Reviews*). Die auf diese Weise periodisch zu erstellenden Länderberichte dienen jedoch auch der Rechenschaftslegung in der Entwicklungszusammenarbeit. Diese sogenannten DAC-Evaluationskriterien werden derzeit revidiert.

3.2. Ablauf der Peer Reviews im Bereich der öffentlichen Statistik

[Rz 20] Eurostat hingegen hat in Zusammenarbeit mit den nationalen statistischen Ämtern, die im Europäischen Statistischen System (ESS) zusammengeschlossen sind, den Rahmen und die Methodik für die Evaluierung der Compliance anhand des CoP entwickelt. Peer Reviews sind seitdem ein strategisches Element und Teil des Arbeitsprogramms des ESS mit Blick auf die Umsetzung des Verhaltenskodex zur Verbesserung der Integrität, Unabhängigkeit und Rechenschaftspflicht der statistikproduzierenden Institutionen.

[Rz 21] Zwischenzeitlich hat sich der Kreis der Länder, die an einer Überprüfung ihrer statistischen *Governance* interessiert sind, um die Länder, die unter die Europäische Nachbarschaftspolitik (ENP) fallen, erweitert.¹¹ Hauptziel der Zusammenarbeit in der Statistik mit den ENP-Ländern ist die Schaffung eines kontinuierlichen Prozesses der statistischen Harmonisierung. Im Zuge der internationalen Zusammenarbeit der EU (Eurostat) unter Mitwirkung der EFTA und der UNECE (Statistics Division) wurden in der Folge *Global Assessments* (GA) als Evaluierungsinstrument für die öffentliche Statistik entwickelt und durchgeführt. Diese GA sind – aus Sicht der Evaluationsforschung – ebenfalls Peer-Evaluationen. Unterschiede bestehen im Hinblick auf

- *die offizielle Bedeutung*: die rechtlichen Grundlagen des europäischen Statistiksystems und somit auch der Code of Practice stellen für die ENP-Länder kein verbindliches Regelwerk dar;
- *die Anforderungen*: die Teilnahme an solchen Evaluationen ist in den ENP-Ländern freiwillig, innerhalb des ESS hingegen verbindlich;
- *die Dauer und den Umfang der Evaluation*: Peer Reviews innerhalb des ESS sind kürzer bzw. haben geringeren Umfang, da ein gemeinsames Regelwerk bereits besteht;
- *den Fokus der Evaluation*: aus den eben erwähnten Gründen steht bei den ENP-Ländern nicht allein die Überprüfung der Anwendung der Prinzipien des CoP im Mittelpunkt;
- *den Evaluationsgegenstand*: dieser ist bei den ENP-Ländern globaler insofern, als die Produktionsprozesse ausgewählter Statistikbereiche im Hinblick auf europäische Standards analysiert und bewertet werden, mit der Zielsetzung einer schrittweisen Heranführung an europäisch oder international harmonisierte und vergleichbare Statistiken und der Verbesserung internationaler Normen.

[Rz 22] Die nachfolgend erläuterte Methodik ist jedoch ähnlich; es würde den Rahmen dieses Artikels sprengen, auf alle Unterschiede in der Umsetzung einzugehen. Wenn nachfolgend von Peer-Evaluation die Rede ist, sind beide Formen dieser in der internationalen öffentlichen Statistik gebräuchlichen Überprüfung der Compliance mit dem CoP gemeint; wo nötig und sinnvoll, wird auf die innerhalb des ESS durchgeführten Peer Reviews (ESS-PR) oder auf die in der internationalen öffentlichen Statistik üblichen Global Assessments (GA) verwiesen.¹²

[Rz 23] Die ersten Peer Reviews fanden in den Jahren 2006–2008 statt, erste Global Assessments zwischen den Jahren 2008 und 2013. Im Hinblick auf eine zweite Runde, die für beide Arten von Peerevaluationen von vorneherein vorgesehen war, da es schliesslich auch die Fortschritte in der Umsetzung (der Empfehlungen) zu überprüfen galt, wurde die Methodologie – unter Leitung

¹¹ Es sind dies die osteuropäischen Länder Armenien, Aserbaidschan, Belarus, Georgien, Moldau und Ukraine und die im Süden an Europa grenzenden Länder Algerien, Ägypten, Israel, Jordanien, Libanon, Libyen, Marokko, Palästina, Syrien und Tunesien.

¹² Eine weitere Variante sind die Light Peer Reviews (LPR), auf die hier nicht weiter eingegangen wird.

der zuständigen Programmverantwortlichen bei Eurostat in Zusammenarbeit mit einer Expertengruppe bestehend aus Vertreterinnen und Vewrtretern der Mitgliedstaaten und unter Beiziehung eines externen Experten – überarbeitet und angepasst. Das Untersuchungsdesign sowie die Erhebungs- und Auswertungsmethoden sind somit festgelegt und folgen im Wesentlichen dem in Abbildung 2 skizzierten Prozess, der hiermit kurz beschrieben wird.

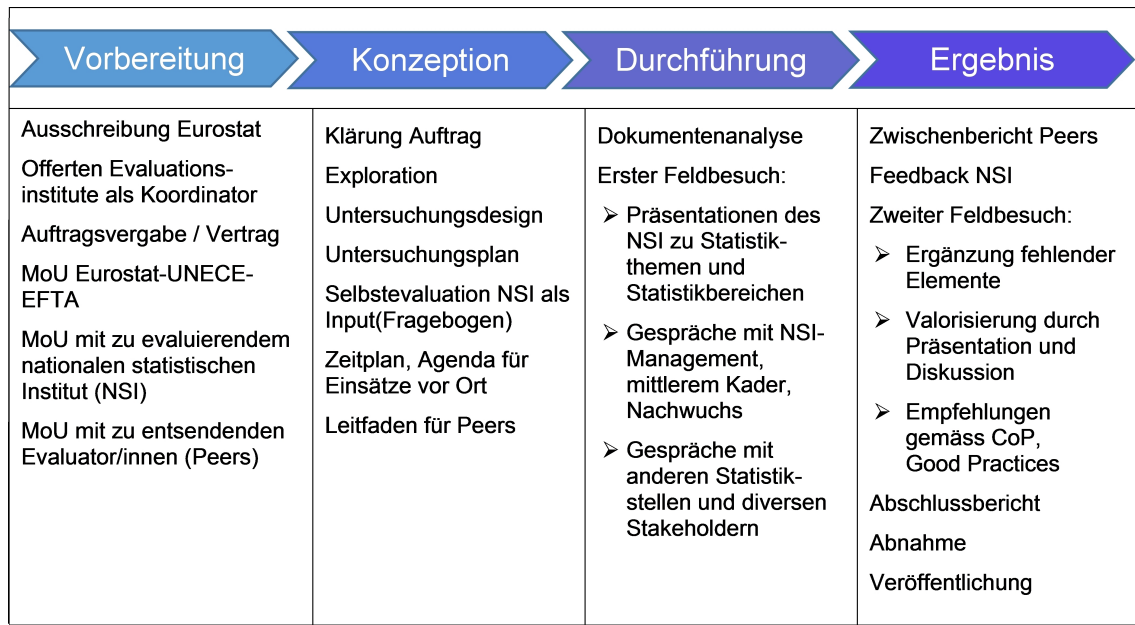


Abb. 2 Untersuchungsdesign für Peer-Evaluationen in der öffentlichen Statistik*

* Eigene Darstellung, vereinfachtes Schaubild, präsentiert am argev-SEVAL-Kongress 2018, Methodenatelier 8

[Rz 24] Die Vorbereitung der Peer-Evaluationen obliegt dem europäischen statistischen Amt, das mit den zu evaluierenden nationalen statistischen Instituten den Evaluatorinnen und Evaluatoren den Evaluationsauftrag in Form eines *Memorandum of Understanding* erteilt. Das Evaluations-team besteht aus mehreren Expertinnen und Experten, die unterschiedliche Fachkompetenzen im Bereich der Statistik mitbringen müssen. Bei den Global Assessments entsenden die auftraggebenden internationalen Organisationen UNECE, EU und EFTA mitunter zusätzliche Experten oder Beobachter, deren Rolle nicht immer klar definiert ist. Zur Vorbereitung gehört des Weiteren die Planung der personellen und finanziellen Ressourcen und die Erstellung eines Zeitplans. Ein Peer Review innerhalb des ESS zieht sich in der Regel über einen Zeitraum von drei bis vier Monaten, ein Global Assessment über sechs bis acht Monate hinweg.

[Rz 25] Konzeption und Methode sind weitgehend harmonisiert, nicht zuletzt um eine Vergleichbarkeit untereinander zu gewährleisten. Ein wichtiger Input ist eine vorgängig durchzuführende Selbstevaluation durch das zu evaluierende nationale statistische Institut (NSI). Dieses «Self Assessment» durch die Verantwortlichen der jeweiligen Teilprogramme des statistischen Arbeitsprogramms ist ein erster Teil der Datenerhebung, verbunden mit einer Zusammenstellung bedeutender Begleitdokumente, die dem Evaluationsteam zur Verfügung gestellt werden und als Vorbereitung für einen späteren Besuch vor Ort dienen. Diese Programmverantwortlichen des NSI (oder anderer nationaler Statistikstellen) bewerten damit ihre eigene Tätigkeit und erhalten bereits in der Selbstreflexion erste Erkenntnisse für potenzielle Verbesserungen. Anzumerken

ist jedoch, dass die NSI relativ frei sind in der Art und Weise, wie sie diese Self Assessments durchführen; entsprechend heterogen ist die Datenlage und damit der Input für einen ersten Länderbesuch, der daraufhin (in der Regel) am Sitz des NSI stattfindet.

[Rz 26] Die Organisation der Länderbesuche («*Country Visits*» in der Sprache der Statistikgemeinde, «Feldbesuche» in der Sprache der Evaluator/innen) obliegt dem jeweiligen nationalen statistischen Amt. Dabei gibt es jedoch eine Reihe von Vorgaben bezüglich des Einbezugs von Stakeholdern, Beteiligten und Betroffenen. So sind Treffen vorgesehen mit

- Vertreterinnen und Vertretern aller statistischen Institutionen, die nationale (öffentliche/amtliche) Statistiken erstellen, darunter auch Vertreterinnen und Vertreter der Nationalbank, des Zolls und registerführender Verwaltungsstellen;
- Vertreterinnen und Vertretern der regionalen statistischen Institutionen, so diese für das jeweilige Statistiksysteem von Bedeutung sind;
- Vertreterinnen und Vertretern der Zivilgesellschaft als Statistikknutzenden, wie z. B. die Sozialpartner, Medien, Verbände, Parlamentarier und Politikerinnen und Politiker;
- und mit Mitarbeitenden aller Hierarchiestufen des NSI.

[Rz 27] In der Vorbereitung und Ausgestaltung dieser Treffen («*Meetings*») sind die nationalen statistischen Ämter ebenfalls frei. So können diese Zusammenkünfte je nachdem den Charakter von qualitativen, leitfadengestützten Interviews, von Gruppendiskussionen oder von relativ formalen Sitzungen haben.

[Rz 28] Kulturelle Unterschiede sind dabei gut beobachtbar. So entschied sich ein Land, das seine heutige Staatsform und Unabhängigkeit 1991 erreichte, für formale, strukturierte, protokollarisch definierte Arbeitssitzungen mittels Übersetzerinnen und Übersetzer, während ein anderes, im Transformationsprozess befindliches Land den kollegialen, teilweise informellen Austausch in einer der beiden UN-Arbeitssprachen vorzog.

[Rz 29] Bereits am Ende eines solchen Länderbesuches werden in einer abschliessenden Sitzung mit dem Kader der nationalen statistischen Institutionen (von denen es insbesondere in föderalen Ländern in der Regel mehrere gibt) erste Befunde vorgestellt und diskutiert. Ein Zwischenbericht soll möglichst zeitnah vorgelegt werden, sodass das NSI dazu Stellung nehmen kann. Diese werden bei der Erstellung des Endberichts berücksichtigt. Ein Peer Review ist (in der Regel) damit abgeschlossen, während bei den Global Assessments ein zweiter Länderbesuch vorgesehen ist, bei dem die Befunde im Zuge weiterer *Meetings* valorisiert werden sollen. Eventuelle Informationslücken oder Divergenzen zwischen Selbst- und Fremdevaluation sollen dabei diskutiert und allenfalls bereinigt werden. Erst dann wird der Endbericht inklusive Empfehlungen vorgelegt und veröffentlicht.

[Rz 30] Die Berichte zu den Peer-Evaluationen sind öffentlich zugänglich. Der Evaluationsbericht zum schweizerischen statistischen System ist auf der Website des BFS abrufbar (*Peer Review Report Switzerland*), die Berichte zu den Peer Reviews innerhalb des Europäischen Systems auf der Website von Eurostat unter *Peer Review in the European Statistical System* und jene zu den Global Assessments unter *Global Assessment Reports*.

[Rz 31] Dank der festgelegten Bewertungskriterien und der Struktur der Berichterstattung sind die Ergebnisse der Peer-Evaluation für den jeweiligen Typ (ESS-PR und GA) summarisch vergleichbar. Der Aspekt der Vergleichbarkeit einer Peer-Evaluation mit einer anderen ist dabei weniger entscheidend als die Reliabilität und Relevanz des einzelnen Peer Reviews oder des Global Assessments. Bei der Beschreibung und Bewertung gilt es, dem nationalen Statistiksysteem

mit den jeweiligen länderspezifischen Elementen Rechnung zu tragen (wie beispielsweise in der Schweiz deren föderativem Aufbau des Staatswesens). Ein weiteres Augenmerk gilt der Zweckmäßigkeit der Evaluation und des Evaluationsberichts, und folglich gibt es nicht nur in Bezug auf Zielsetzung und Evaluationszweck, sondern auch hinsichtlich der Berichterstattung Unterschiede. Während die Evaluation in einem Land Grundlage für das Erstellen eines Aktionsplans zur Weiterentwicklung des nationalen Statistiksystems ist, nutzt ein anderes Land den Evaluationsbericht als Nachweis über die Qualität der Statistikproduktion. Die Kommunikation zur Evaluation richtet sich demnach in einem Fall an interne Kreise, in einem anderen an die interessierte Öffentlichkeit oder eben zielgerichtet an verschiedene Stakeholder.

4. Erfahrungsbericht und Reflexion über die Global Assessments als spezielle Form einer Peer-Evaluation

[Rz 32] Im Hinblick auf den Evaluationszweck gibt es Unterschiede zwischen den Peer Reviews innerhalb des ESS und den Global Assessments. Bei den ESS-PR steht die Überprüfung der Compliance mit dem CoP im Vordergrund, wobei der Fokus auf der Bewertung des institutionellen Rahmens liegt. Die Statistikprozesse, aber auch die Anforderungen an den statistischen Output sind innerhalb des Europäischen Statistiksystems meist gesetzlich in Form von Verordnungen oder Richtlinien oder in Form von verbindlichen Vereinbarungen geregelt und infolge dessen (mehr oder weniger) harmonisiert. Eventuelle Abweichungen sind innerhalb der Prozess- oder Output-Dokumentation begründet und beschrieben, sodass deren Überprüfung im Rahmen der Peer Reviews nicht zwingend erforderlich oder zielführend erscheint.

[Rz 33] Nicht so bei den Global Assessments, die in Ländern durchgeführt werden, deren Statistiksystem nicht an das ESS-Regelwerk gebunden ist. Folglich ist der Evaluationszweck ein anderer, ein erweiterter. Die GA-Methodologie sagt hierzu (EUROSTAT 2016, Anhang): «The main object (...) is to evaluate the level of conformity vis-à-vis European and international standards, incorporating the Code of Practice, the UN Fundamental Principles of Official Statistics and the Eurostat Statistical Requirements Compendium. Through this process the global assessment supports the improvement of the national statistical system and its alignment with European and international recommendations, standards and best practice».

[Rz 34] Im Zentrum der *Global Assessments* steht stets das nationale statistische Amt. Doch in jedem Land gibt es innerhalb der Verwaltung Ministerien, ihnen untergeordnete Ämter oder unabhängige Institutionen, die Statistikstellen unterhalten und amtliche Statistiken erstellen und daher Teil des nationalen Statistiksystems sind.¹³ Sie stehen in einer wechselseitigen Arbeits- und in einer Evaluationsbeziehung. Weitere Beteiligte sind mitunter andere nationale statistische Ämter oder internationale Organisationen, die im Rahmen der technischen Zusammenarbeit als Geldgeber für bestimmte statistische Entwicklungsprojekte fungieren.

¹³ Notabene zur Erinnerung: In der Schweiz sind dies beispielsweise das SECO, ARE, BSV, SBFI, die Zollverwaltung und die Schweizerische Nationalbank sowie ca. 30 weitere Institutionen, die gemeinsam (unter Leitung des BFS) das statistische Mehrjahresprogramm 2016-19 erstellen, das vom Bundesrat verabschiedet wird. Weitere Informationen zum Statistiksystem Schweiz und dessen Arbeitsprogramm finden sich auf der Webseite des BFS.

[Rz 35] Angesichts der vielen Beteiligten (und Betroffenen) zeigen sich bei den Global Assessments immer wieder Schwierigkeiten im Zusammenhang mit

- der Klärung des Evaluationszwecks und des Erreichens eines gemeinschaftlichen Verständnisses betreffend die Evaluation;
- der Rolle der «Peers» und anderer Beteiligter und Betroffener;
- dem systemischen Ansatz der Evaluation.

[Rz 36] Während die Vorstellungen der Auftraggebenden (EU und EFTA, UNECE) und die Aufgabenstellung den Evaluatorinnen und Evaluatoren weitgehend bekannt sind, sind die Erwartungen und Bedürfnisse des nationalen statistischen Amtes, das um das Assessment ersucht und sich damit ebenfalls in der Rolle des Auftraggebenden wähnt, gleichzeitig aber auch Evaluationsgegenstand ist, weniger bekannt. Erfahrungen zeigten, dass die Durchführung von Global Assessments oft mit einer *hidden agenda*, seitens des nationalen statistischen Amtes belegt sind, die sich dem Evaluationsteam nicht sogleich offenbart. Hier einige Beispiele zur Illustration:

- Das NSI ist innerhalb der nationalen Verwaltung von untergeordneter Bedeutung, personell und finanziell schlecht ausgestattet und benötigt eine gute Empfehlung, um nationale oder internationale Geldquellen zu erschliessen.
- Das NSI hat bereits Unterstützung seitens der internationalen Statistikgemeinde oder internationaler Organisationen erhalten (z. B. Finanzmittel oder Knowhow zur Durchführung eines Zensus) und steht nun in der Rechenschaftspflicht, wodurch die Gefahr besteht, dass die Agenda der *Country Visits* und die Zusammensetzung der *Meetings* nicht vollständig und mit einem «bias» belegt sind – schliesslich bedarf es einer guten Bewertung, wenn man auf weitere Finanzmittel hofft.
- Das NSI ist innerhalb der nationalen Verwaltung eng mit der nationalen Regierung verzahnt und nicht vollständig unabhängig. Es erwartet entweder eine Unterstützung zugunsten einer Unabhängigkeit oder eine «Hofberichterstattung».
- Das NSI hat konfliktreiche Arbeitsbeziehungen zu anderen Stakeholdern und erwartet von der Evaluation und den Kolleginnen und Kollegen (resp. Evaluator/innen) eine Schlichtung des Konflikts.
- Innerhalb des NSI werden unterschiedliche, mitunter divergierende, Strategien und Positionen hinsichtlich der künftigen Entwicklung vertreten; die entsprechenden Parteien erwarten eine Mediation.
- Das NSI ortete Unterstützungsbedarf in bestimmten statistischen Bereichen und erwartet von den ausländischen Kolleginnen und Kollegen Beratung und Dienstleistungen (nicht Evaluation).

[Rz 37] Betroffene – in Bezug auf die Arbeits- wie auch auf die Evaluationsbeziehung – sind die nationale Regierung, das Parlament, politische Parteien, die Wissenschaft und Forschung, Medien und die interessierte Öffentlichkeit als Nutzende von Statistiken. Sie werden daher in die Peer-Evaluation (mehr oder weniger) miteinbezogen.

[Rz 38] Die Evaluatorinnen und Evaluatoren erhalten zwar, ganz im Sinne von STUFFEBEAM (2004), Checklisten und Evaluationspläne, finden jedoch im Feld sehr komplexe Arbeitsbeziehungen vor, die mit der bestehenden Methodik nur schwer zu greifen sind. Die kartografische Analyse (Abb. 3) ist eine vereinfachte Darstellung der Arbeitsbeziehungen. Das Schaubild steht für viele solche Lagen, die aufgrund einer Vielzahl von Akteuren komplexer sind als hier dargestellt und die zunächst aufgedeckt werden müssen, ehe mit der Evaluation begonnen werden kann.

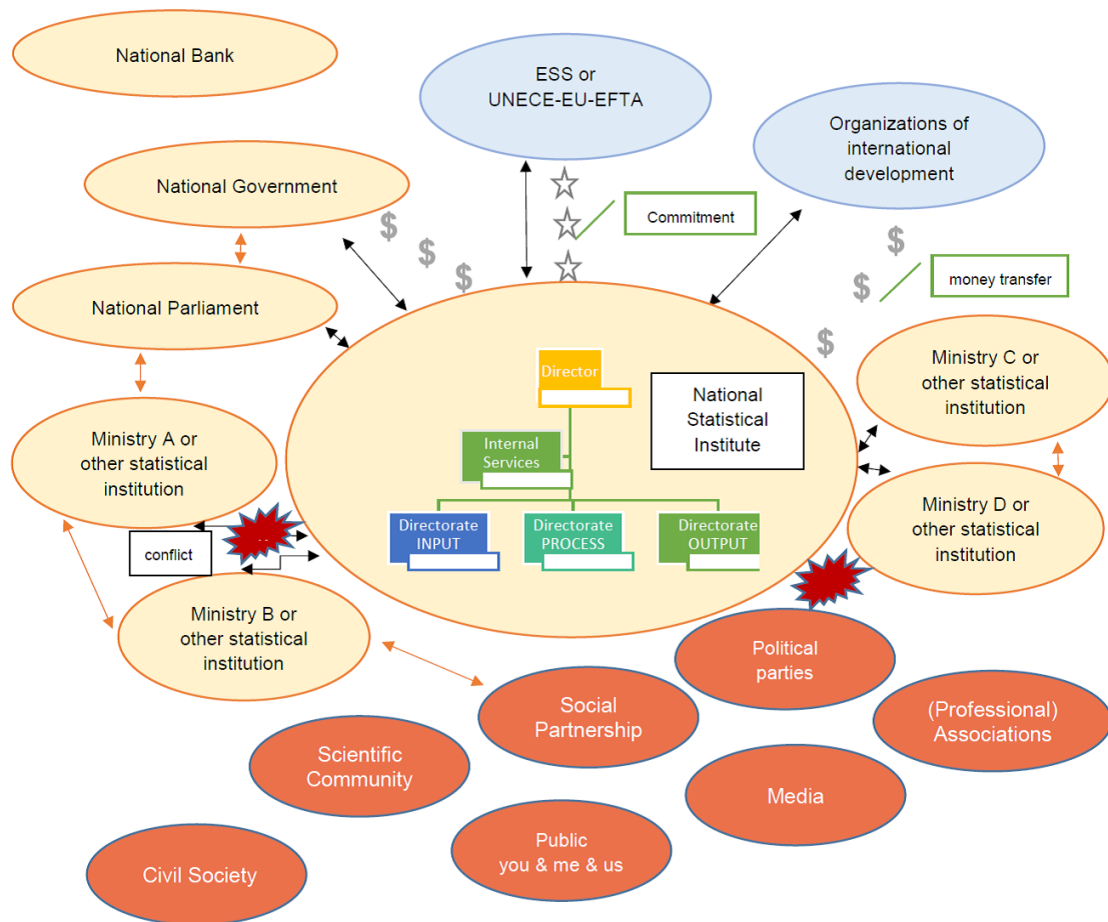


Abb. 3: Kartografische Analyse der Arbeitsbeziehungen*

* Eigene Darstellung, vereinfachtes Schaubild.

[Rz 39] Erfahrungen haben gezeigt, dass Beteiligte und Betroffene im Hinblick auf die Evaluations- und Arbeitsbeziehungen in einer Doppelfunktion auftreten und ihnen die Funktion und die Zielsetzung einer solchen Evaluation sowie ihre Rolle innerhalb des Evaluationsprozesses – als Beteiligte an der Statistikproduktion, aber von der Evaluation Betroffene – nicht immer deutlich ist. Ähnliches gilt für die Evaluatorinnen und Evaluatoren, die Fachleute in einem Statistikbereich und damit nicht zwangsläufig geschulte oder erfahrene Evaluatoren sind, sodass bei den «Peers» mitunter die Trennlinien zwischen kollegialer Beratung, Austausch von «good practices» und Evaluation verschwimmen. Die Unterscheidung zwischen Beratungs- und Evaluationsmandat ist den meisten Beteiligten nicht immer klar; eine Gefahr, die Peer-Evaluations grundsätzlich immanent ist.

[Rz 40] Eine weitere Problematik ergibt sich aus dem Umstand, dass nicht etwa das nationale statistische Amt als Institution, sondern das nationale statistische System zu evaluieren ist. Der Begriff «System» impliziert bereits, dass es sich dabei um die Gesamtheit der statistikproduzierenden Institutionen handelt. Es geht nicht nur um die Bewertung eines einzelnen Statistikamtes, was angesichts der Vielzahl statistischer Prozesse schon komplex genug wäre, sondern um

die Bewertung einer länderspezifischen Entwicklungsstrategie im Bereich der Statistik und der Umsetzung ihres statistischen Arbeitsprogrammes unter Berücksichtigung der im Verhaltenskodex definierten Elemente (als Teil der Public Governance). Folglich kann nur eine summative Evaluation mit bilanzierenden Schlussfolgerungen unter Berücksichtigung der Bewertungskriterien vorgenommen werden. Manchen Programmverantwortlichen in den statistikproduzierenden Institutionen, die sich Unterstützung in ihren Fachgebieten erwarten, ist dies mitunter zu wenig, und so bleibt für sie die Evaluation hinter ihren Erwartungen zurück.

[Rz 41] Gewisse Verbesserungen – ohne weitere Modifikationen in der Methodik – könnten mit folgenden Massnahmen erreicht werden:

- Herbeiführung der Klärung des Evaluationszwecks und -prozesses im Zuge einer eigenen Informationsveranstaltung mit den Beteiligten und Betroffenen vorgängig zur Evaluation;
- Begrenzung des Umfangs und der Schwerpunktsetzung bei der Evaluation, z. B. in Abhängigkeit der Befunde des Self Assessments oder der strategischen Ausrichtung des Arbeitsprogrammes;
- Trennung von Evaluationsergebnissen und Empfehlungen, z. B. durch getrennte Berichte.

[Rz 42] Fazit: Ungeachtet der dargelegten Problematik solcher Peer-Evaluationen überwiegt deren Stärke: «Der Prozess ist das Ziel», denn Erfahrungen zeigen, dass die Selbst- und die Fremdbeurteilung mitunter voneinander abweichen. Die Diskussion der Ergebnisse setzt damit einen Erkenntnisprozess in Gang und lässt sowohl Verbesserungspotenzial als auch gute Praktiken – «good practices» – erkennen. Die Peer-Evaluationen bewegen sich im Spannungsfeld zwischen Qualitätsüberprüfung, systemischer Weiterentwicklung, Rechenschaftslegung oder «Akkreditierung» im dem Sinne, dass eine allgemein anerkannte Instanz einer anderen das Erfüllen gemeinschaftlich gesetzter und anerkannter Standards bescheinigt.

[Rz 43] Die Ansprüche, die seitens Wissenschaft, Politik und Verwaltung oder der interessierten Öffentlichkeit an die öffentliche Statistik gestellt werden, sind die gleichen, die gemäss WIDMER (2001, 19) an die Evaluation gerichtet werden und die dieser mit den Stichworten «Nützlichkeit, Zeitgerechtigkeit, politische Opportunität, Anschlussfähigkeit, Sparsamkeit, Praktikabilität und Korrektheit» umschreibt. Nur wenn diese Bedingungen und die im Code of Practice festgelegten Gütekriterien der Statistik erfüllt sind, können die statistikproduzierenden Institutionen auf Vertrauen und Akzeptanz seitens ihrer Stakeholder hoffen. Auch wenn diese hier skizzierten Peer Reviews der öffentlichen Statistik überwiegend von Evaluationslaien durchgeführt werden, lassen sich die eingangs erläuterten Gütekriterien mittels einer Peerevaluation internationaler Expertinnen und Experten, die durch ein vorhergehendes Self Assessment ergänzt wird, gut überprüfen.

VERA HERRMANN, Dr. rer. pol., Sozialwissenschaftlerin, Bereichsleiterin beim Bundesamt für Statistik und zeitweise als Evaluatorin in Ländern tätig, die unter die Europäische Nachbarschaftspolitik fallen. Dieser Artikel liegt ausschliesslich in der persönlichen Verantwortung der Autorin. E-mail: vera.b.herrmann@gmail.com.

Literaturverzeichnis

AGUILERA, RUTH V./CUERVO-CAZURRA, ALVARO (2009): Codes of Good Governance in Corporate Governance, in: International Review, Volume 17, Nr. 3.

BANK, VOLKER/LAMES, MARTIN (2010): Über Evaluation, 2. Auflage.

BALTHASAR, ANDREAS (2005) Was ist Evaluation und für wen evaluieren wir?, in: LeGes 2005/1, S. 65–80.

BALZER, LARS/BEYWL, WOLFGANG (2015): evaluiert. Planungsbuch für Evaluationen im Bildungsbereich, h.e.p., Bern.

BEYWL, WOLFGANG/NIESTROJ, MELANIE (2009) A-B-C der wirkungsorientierten Evaluation: Glossar – Deutsch / Englisch, Univention Institut für Evaluation (2. Überarb. Aufl. in Print).

European Network for Quality Assurance in Higher Education (2009): Standards and Guidelines for Quality Assurance in the European Higher Education Area.

Eurostat (2016): Second round of Global Assessment – Methodology; ESS-internes unveröffentlichtes Arbeitspapier zur Präzisierung des Mandats (Anhang) einer Peerevaluation.

European Union (2018): European Statistics Code of Practice for the National Statistical Authorities and Eurostat (EU statistical authority); Publications Office of the European Union, 2018. (3. Überarbeitung vom 16. November 2017 auf der Basis der ersten Version, verabschiedet am 24. Februar 2005, und der Revision vom 28. September 2011).

SEVAL (2017): Erläuterungen zu den Evaluationsstandards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards).

STUFFLEBEAM, DANIEL L. (2004): Checkliste Evaluationsplan; übersetzt von Wolfgang Beywl (2006), überarbeitet von Katharina Klockgether (2015).

WIDMER, THOMAS (2001): Qualitätssicherung in der Evaluation – Instrumente und Verfahren, in: LeGes 2001/2, S. 11–39.