
Neuenburg, 31.05.2022

Nutzung beliebiger Datenquellen durch die öffentliche Statistik

1. Ausgangslage

Den folgenden Ausführungen liegt die Überzeugung des Ethikrates zugrunde, dass die öffentliche Statistik ihre derzeitige hohe Akzeptanz und Glaubwürdigkeit nur dann weiter halten können, wenn sie die Herausforderung der digitalen Transformation nutzt. So hat sie z.B. neue, bisweilen unkonventionelle Datenquellen zu erschliessen, methodische Erweiterungen sowie Dienstleistungen im Rahmen der Datenbewirtschaftung oder Datenwissenschaft zu erbringen. Nicht zuletzt unter dem Konkurrenzdruck privater Anbieter von Statistiken, deren Qualitätsstandards nicht immer jenen der öffentlichen Statistik entsprechen oder zumindest nicht deren Anspruch an Transparenz genügen müssen, sind innovative strategische Antworten gefordert, die über die traditionellen statistischen Methoden und Technologien hinausgehen. Das bedingt natürlich die Auseinandersetzung mit den damit einhergehenden datenethischen Fragestellungen.

Beim Bundesamt für Statistik (BFS) wird die Umsetzung der Aufgabenerweiterung in der 2021 erfolgten Anpassung der Organisationsstruktur (neue Abteilungen Interoperabilität und Register IOR, Datenwissenschaft und Statistische Methoden DSSM) klar sichtbar. Die regionalen Statistikstellen gehen die Aufgabenerweiterung insgesamt mit einer gewissen zeitlichen Verzögerung an. Auch sie sind aber zunehmend Kompetenzzentren für Datenbewirtschaftung, einige v.a. grössere Ämter nehmen zudem Aufgaben im Rahmen von OGD wahr. Mit neuen berufs- und datenethischen Fragestellungen und allfälligen Anpassungen ethischer Vorgaben wie der Charta dürften somit alle Stellen der öffentlichen Statistik konfrontiert sein, derzeit aufgrund der ihm vom Bundesrat zugewiesenen neuen Aufgaben allerdings vor allem das BFS.

Es ist daher nicht erstaunlich, dass entsprechende Anregungen und Fragestellungen derzeit eher aus dem BFS kommen. Deren Bearbeitung durch den Ethikrat schliesst jedoch immer die gesamte Trägerschaft und somit auch die regionalen Statistikstellen mit ein. Diese sind ausserdem über diverse Kanäle (Arbeitsgruppen BFS-KORSTAT, Regiostat usw.) in die Überlegungen involviert.

Das BFS und der Ethikrat sehen zur Behandlung datenethischer Fragestellungen ein iteratives Vorgehen vor. Das BFS gelangt, der Entwicklung von Erweiterungsprozess und Projekten folgend, punktuell an den Ethikrat. So wurde er letztes Jahr um Überlegungen zur Verwendung beliebiger Daten durch die öffentliche Statistik angegangen, die im Folgenden dargelegt werden.

2. Fragestellung: Verhaltenskodizes und neue grosse Datenquellen

Als erstes wird der Frage nachgegangen, ob die Verhaltenskodizes wie Fundamental Principles, Code of Practice (CoP) oder Charta bezüglich der Verwendung neuer und grosser Datenquellen durch die öffentliche Statistik irgendwelche Einschränkungen vorsehen. Hierzu finden sich diverse Publikationen. Wir referenzieren hier auf einen Beitrag von D. Rozkrut et al. zum «Mapping the UN Fundamental Principles of Official Statistics against new and big data sources»¹. Untersucht werden die UN-Grundprinzipien der öffentlichen Statistik im Kontext von acht neuen und grossen Datenquellen: Mobiltelefone, Stromverbrauchszähler, Satellitenbilder, Social-Media, Web-Scraping, Straßenverkehrssensoren und Passagierverfolgung, Scanner- und Kassendaten, CCTV-Daten (Sicherheits- / Überwachungsvideos).

Der Beitrag kommt zum Schluss, dass jede dieser Datenquellen für die Produktion von öffentlicher Statistik in Übereinstimmung mit den Fundamental Principles verwendet werden kann. Ja, dass diese Datenquellen gar genutzt werden sollten, wenn die nationalen statistischen Systeme das erste Grundprinzip der Fundamental Principles einhalten wollen, nämlich den Anspruch der Bürgerinnen und Bürger auf öffentliche Information durch die Erstellung und Bereitstellung von amtlichen Statistiken zu erfüllen.

Ohne hier auf Details dieses Papiers einzugehen, stimmt der Ethikrat dieser Erkenntnis zu und ergänzt um die folgenden allgemeinen Überlegungen und Konkretisierungen:

2.1 Allgemeine Überlegungen

Die bestehenden Verhaltenskodizes ...

- «Fundamental Principles of Official Statistics» der UNO (erstellt 1992 / revidiert 2013),
- «Code of Practice» des Europäischen Statistischen Systems (CoP, 2005 / 2011, 2017)
- die in Anlehnung an den CoP erstellte «Charta der öffentlichen Statistik der Schweiz» (Charta, 2002 / 2008, 2012).

... sind in erster Linie berufsethische Vorgaben für die Stellen der öffentlichen Statistik

- zum Informationsauftrag
- zur fachlichen Unabhängigkeit
- zum Persönlichkeits- und Datenschutz
- zur Wirtschaftlichkeit
- zur Qualität

Im Vordergrund steht die durch **Transparenz der Aufgaben und Arbeitsweisen** anzustrebende Glaubwürdigkeit der öffentlichen Statistik und der von ihr aufbereiteten Statistiken mit hohem Qualitätsstandard. Es geht primär darum, **wie** die Datenbearbeitung (umfassend verstanden als Datengewinnung, -verarbeitung und Publikation der Ergebnisse) im Rahmen «klassischer» Statistik zu erfolgen hat. Einschränkungen bezüglich der zu verwendenden Datenquellen, zumal unter den Möglichkeiten von Big Data, gibt es explizit keine. Wenn sich die Statistikstellen bei der Datenbearbeitung an die berufsethischen Vorgaben der Verhaltenskodizes halten, können sie somit Daten beliebiger Quellen bearbeiten.

¹ Statistical Journal of the IAOS 37 (2021) 161–169

2.2 Konkretisierungen

2.2.1 Informationsauftrag

Dem Geschilderten folgend ist die öffentliche Statistik darauf angewiesen, neue Datenquellen zu erschliessen, um ihrem grundsätzlichen Auftrag nachzukommen, den Bedarf an statistisch relevanten Informationen von gesellschaftlichem Interesse zu decken. Dieser Auftrag ist in den Fundamental Principles und in der Charta jeweils im Prinzip 1 festgehalten. Im CoP – das Prinzip 1 zielt hier auf die Unabhängigkeit der Statistikstellen – ist die Relevanz der statistischen Information im Grundsatz 11 aufgeführt.

2.2.2 Keine Vorgaben bezüglich Datenquellen in den Verhaltenskodizes

Es bleibt somit die Frage, ob die Verhaltenskodizes irgendwelche Vorgaben wie z.B. Einschränkungen bezüglich der Datenquellen machen.

- In den Fundamental Principles hält Prinzip 5 fest, dass für statistische Zwecke Daten aus allen Arten von Quellen bezogen werden können.
- In der Präambel zur 2017 überarbeiteten Version des CoP ist folgendes festgehalten: «Damit soll den jüngsten Veränderungen und Innovationen bei der Entwicklung, Erstellung und Verbreitung amtlicher Statistiken im Europäischen Statistischen System und darüber hinaus Rechnung getragen werden, wie etwa der Entstehung neuer Datenquellen, der Nutzung neuer Technologien, der Modernisierung des Rechtsrahmens oder den Ergebnissen der Peer Reviews zur Umsetzung des Verhaltenskodex.» Mit der revidierten Fassung fand der Hinweis auf «Daten aus vielfältigen Datenquellen» Eingang in den Grundsatz 2 sowie in mehrere Indikatoren des CoP.
- In der Charta ist keine Einschränkung hinsichtlich der Datenquellen ersichtlich. Es wird allerdings nicht wie in den Fundamental Principles und im CoP auf die Möglichkeit vielfältiger Datenquellen hingewiesen. **Die Charta sollte bei der nächsten Revision entsprechend ergänzt werden.**

2.3 Fazit

Aus Sicht der drei Verhaltenskodizes gibt es kein grundsätzliches Problem bezüglich Verwendung beliebiger Datenquellen. Bei der Bearbeitung von neuen grossen Datenquellen eröffnen sich aber Fragestellungen auf verschiedenen Ebenen, die im Folgenden nur ganz kurz (und ohne Garantie auf Vollständigkeit) aufgeworfen werden:

3. Neue Fragestellungen

3.1 Datenbezug von privaten Anbietern

Da private Anbieter gesetzlich in den meisten Fällen nicht zur Datenabgabe an die öffentliche Statistik verpflichtet sind, basiert deren Datenabgabe auf Freiwilligkeit.

3.1.1 Wahren der fachlichen Unabhängigkeit

Für die öffentliche Statistik ist es zentral, die fachliche Unabhängigkeit zu wahren. Dem kann sie mit Transparenz bezüglich der Nennung und Auswahl der Datengeber nachkommen. Umgekehrt ist durch die Statistikstellen den privaten Datenanbietern die Verwendung der Daten für ausschliesslich statistische Zwecke zuzusichern und einzuhalten.

3.1.2 Kostenfrage

Zu diskutieren wird wohl die Kostenfrage sein, die sich im Zusammenhang mit dem Bezug von Daten privater Anbieter durch die öffentliche Statistik ergeben könnte. Hier berühren sich datenethische und politische Fragestellungen, insbesondere Fragen des Umfangs des Service public im Bereich der Dateninfrastruktur und deren Nutzung u.a. durch die öffentliche Statistik.

3.2 Ein Detail: Information der Befragten beim Zugriff auf neue Datenquellen?

Der Zugriff auf neue Datenquellen wie z.B. social medias oder auf solche aus der Telekommunikation ist in den Statistikgesetzen zumindest direkt nicht vorgesehen. Die Befragten, d.h. hier wohl die Betreiber / Owner der jeweiligen Datenquelle sind gemäss BStG Art. 4, Abs. 4 und Charta Prinzip 2 (Rechtsgrundlage) über Zweck, Rechtsgrundlage usw. zu informieren. Das ist gegenüber diesen Stellen wohl kein Problem. Reicht es, wenn deren Datengeber über die Weitergabe ihrer persönlichen Daten zu statistischen Zwecken an die öffentliche Statistik via die Allgemeinen Geschäftsbedingungen (AGB) informiert werden? Müssen sie überhaupt informiert werden? Oder wird die Weitergabe auf nicht-persönliche Daten eingeschränkt?

3.3 Neue Herausforderungen bezüglich Methodik und Qualität

Neue grosse Datenquellen sind in der Regel nicht für statistische Zwecke konzipiert. Sie entsprechen weder vom Volumen her, noch strukturell oder inhaltlich herkömmlichen Mustern der öffentlichen Statistik. Definitionen, Standards, Repräsentativität, Abdeckung oder auch Kontinuität sind in neuen Datenquellen nicht zwingend gegeben. Es resultieren daher neue Fragestellungen und Möglichkeiten bezüglich statistischer Methodik und bezüglich Qualitätsanforderungen, die derzeit im Rahmen der Dateninnovationsstrategie des BFS angegangen werden. Welche Kriterien gelten z.B. für die Qualitätskontrolle von Daten, deren Zustandekommen die Statistikstelle nicht kennt und deren Metainformation unter Umständen dürftig ist? Gefordert ist daher die Stärkung der Forschungsabteilungen in der öffentlichen Statistik sowie die Zusammenarbeit mit den Hochschulen und Forschungsinstituten, wie sie in der neu geschaffenen Abteilung DSSM angegangen wird.

3.4 Labeling der Ergebnisse «ausserhalb» der Statistikproduktion

Es können im Rahmen der Dateninnovationsstrategie Ergebnisse resultieren, die mit den Vorgaben der Charta übereinstimmen und in die Statistikproduktion überführt werden können. Andernfalls, wenn z.B. die Qualität (noch) nicht den gängigen Anforderungen entspricht oder ev. in der Entwicklungsphase von Projekten, sind Möglichkeiten zu suchen, die den Bedürfnissen der Stakeholder entsprechen, ohne dabei die Glaubwürdigkeit der öffentlichen Statistik zu gefährden. Der gängige Weg dazu ist im BFS (wie auch im Europäischen Statistischen System, ESS) derzeit die Veröffentlichung solcher Ergebnisse der experimentellen Statistik unter einem Label und mit entsprechender Metainformation.

3.5 Grenzbereiche

Die Vorgaben der Verhaltenskodizes an die öffentliche Statistik sind beim Erstellen von Statistikprodukten unter Anwendung innovativer Methoden zu berücksichtigen. Welche Lösungen finden wir für Fälle wie zum Beispiel ...

3.5.1 ... für experimentelle Statistik

Inwieweit können diese Vorgaben z.B. in experimenteller Statistik überschritten und dann entsprechend vermittelt / gelabelt werden?

- Sind Anpassungen / Ergänzungen der Charta möglich oder bedarf es weiterer ethischer Vorgaben?

3.5.2 ... für Administrativ- und Statistikdaten

Die Grenzen zwischen «Statistik» / «Nicht-Statistik» wie auch z.B. bezüglich Statistik- und Administrativdaten können fließend sein.

- Finden wir mittels datenethischer Vorgaben Abgrenzungen und letztlich Definitionen?
- Wie gehen wir mit der Tatsache um, dass für die Statistik gedachte Information zunehmend auch administrativ benötigt wird und damit wohl die geltende Gesetzgebung und die Verhaltenskodizes tangiert sind.

3.6 Weitere (werden sich mit Prozessfortschritt ergeben)

Im Namen des Ethikrates

Peter Laube