



00
Statistische Grundlagen und Übersichten

1787-1700

Bundesamt für Statistik Dateninnovationsstrategie

Zweck, strategische Ziele
und Umsetzungsschritte



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Neuchâtel 2017

Herausgeber: Bundesamt für Statistik (BFS)

Auskunft: Bertrand Loison, BFS, Tel. 058 463 67 70,
bertrand.loison@bfs.admin.ch

Inhalt: Bertrand Loison, BFS;
Diego Kuonen, Statoo Consulting

Reihe: Statistik der Schweiz

Themenbereich: 00 Statistische Grundlagen und Übersichten

Originaltext: Englisch

Übersetzung: Sprachdienste BFS

Layout: Sektion DIAM, Prepress/Print

Titelseite: BFS; Konzept: Netthoewel & Gaberthüel, Biel;
Foto: © vinnstock – Fotolia.com

Copyright: BFS, Neuchâtel 2017
Wiedergabe unter Angabe der Quelle
für nichtkommerzielle Nutzung gestattet

Bestellungen Print: Bundesamt für Statistik, CH-2010 Neuchâtel,
Tel. 058 463 60 60, Fax 058 463 60 61,
order@bfs.admin.ch

Preis: gratis

Download: www.statistik.ch (gratis)

BFS-Nummer: 1787-1700

Inhaltsverzeichnis

Einleitende Bemerkungen	5
Einleitung	7
Dateninnovation aus Sicht des BFS	9
Aktuelle rechtliche Grundlage	9
Primär- und Sekundärdaten	10
Der induktiv-deduktive Verfahrenszyklus	11
Erstes strategisches Ziel	11
Zweites strategisches Ziel	13
Übergreifende Grundsätze	13
Umsetzung der strategischen Ziele	14

Einleitende Bemerkungen

Der Schwerpunkt der Strategie liegt darin, die Produktion bestehender öffentlicher Statistiken des Bundesamtes für Statistik (BFS) dort zu erweitern und/oder zu ergänzen, wo Dateninnovation (gemäss nachfolgender Definition) sinnvoll ist. Der Umfang der Strategie ist, wie unten beschrieben, auf bestimmte Elemente der statistischen Innovation begrenzt. Die Strategie zielt also weder darauf ab, alle Aspekte der Innovation beim BFS vollständig abzudecken, noch soll damit eine Prioritätenliste erstellt werden. Die Strategie wird zudem entsprechend der Erfüllung der nachfolgend definierten Ziele und Umsetzungsschritte fortlaufend überarbeitet und aktualisiert.

Einleitung

Die digitale Revolution stützt sich auf Daten und ist in vollem Gange: Big Data, Cloud, Internet der Dinge, «Internet of Everything», vierte industrielle Revolution, Smart City und Datenwirtschaft sind nicht nur Begriffe in aller Munde, sondern Konzepte, die die Gewohnheiten der Konsumentinnen und Konsumenten sowie der Unternehmen verändern. Diese Veränderung begann mit der ersten Welle der Digitalisierung: Diese war rein technisch und bestand darin, dass analoge Inhalte und Dienstleistungen in digitale umgewandelt wurden (ohne fundamentale Änderungen an den zugrunde liegenden Prozessen), was schliesslich zur (Big-)Data-Revolution führte. Big Data ist jedoch eine Datenmanagementinfrastruktur mit entsprechender Hardware, Software und Architektur und sollte somit nicht bloss zu Sammelzwecken dienen. Bereits 1942 hielt der öffentliche Statistiker W. Edwards Deming fest, dass der «Endzweck des Sammelns von Daten darin besteht, eine Handlungsbasis bzw. -empfehlung bereitzustellen». Es braucht also eine zweite Welle der Digitalisierung, um aus Big Data Erkenntnisse zu gewinnen und mit dieser Revolution einen Mehrwert für die gesamte Gesellschaft zu schaffen: Dies wird beim BFS als **Dateninnovation** bezeichnet. Die grösste Herausforderung liegt dabei in der Richtigkeit des «Datenstammbaumes», d. h. im **Vertrauen in die Daten** bezüglich Zuverlässigkeit, Leistungsfähigkeit, Validität und Qualität sowie in die Transparenz des entsprechenden Produktionsprozesses. In einer von postfaktischer Politik und Fake-News geprägten Welt ist die **Sinnhaftigkeit bzw. die Vertrauenswürdigkeit der Daten** («data veracity») wichtiger denn je.

Die öffentliche Statistik hat schon immer mit unterschiedlichen Datenquellen gearbeitet. Um zuverlässige und vertrauenswürdige statistische Dienstleistungen erbringen zu können, muss das BFS seit seiner Gründung im Jahr 1860 laufend die Sinnhaftigkeit bzw. die Vertrauenswürdigkeit seines «Datenstammbaumes» überprüfen. Die Wiederverwendung bestehender Datenquellen und/oder die Nutzung neuer Datenquellen könnte für das BFS jedoch eine Innovationsmöglichkeit darstellen, um die Zahl seiner öffentlichen Statistiken zu erweitern und/oder zu ergänzen sowie um effizientere, aktuelle, wirkungsvolle und umfassendere statistische Dienstleistungen zu erbringen.

Es gilt festzuhalten, dass die Datenrevolution nicht nur in der öffentlichen Statistik relevant ist.

Der Zugang zu und die Nutzung von neuen Datenquellen werden im Rahmen von internationalen und regionalen Statistiktagegen diskutiert. Die «Agenda 2030 für nachhaltige Entwicklung» und deren Ziele («Sustainable Development Goals») sind wegweisend: Daten

und die öffentliche Statistik bekommen nun auf politischer Ebene viel mehr Aufmerksamkeit. Dies erfordert, dass der Zugang zu und die Nutzung von neuen Datenquellen diskutiert, getestet und nach Möglichkeit umgesetzt werden müssen. Der globale Aktionsplan «Cape Town Global Action Plan for Sustainable Development Data», der am 48. Treffen der UNO-Statistikkommission im März 2017 verabschiedet wurde, fordert verschiedene Massnahmen im Zusammenhang mit neuen Datenquellen. Beispielsweise müssen Hindernisse für die Nutzung neuer Datenquellen, einschliesslich Registerdaten, Verwaltungsdaten und anderer neuer Quellen, erkannt und beseitigt und die Bemühungen zu ihrer Verwendung bei der öffentlichen Statistikproduktion koordiniert werden. Die Arbeiten des BFS werden diese Massnahmen unterstützen.

In der Schweiz wurden in den letzten Jahren verschiedene Strategien entwickelt, um den Zugang zu und die Nutzung von neuen Datenquellen und Registerdaten zu fördern (z. B. «Digitale Schweiz», «Schweizer Datenpolitik»). Diese Strategien sind allerdings hauptsächlich technologiegetrieben und nicht auf die Nutzung solcher Daten für statistische Zwecke ausgerichtet. Die öffentliche Statistik und insbesondere das BFS werden vom Parlament zunehmend unter Druck gesetzt: Mittels Register- und Verwaltungsdaten sowie anderer potenzieller, aus den Fortschritten der Digitalisierung entstehenden Datenquellen soll die Belastung der Befragten reduziert werden. Die Arbeiten des BFS werden dazu beitragen, dass die öffentliche Statistik bei der Umsetzung dieser strategischen Initiativen einbezogen wird, sowie dabei helfen, effiziente Lösungen zu entwickeln, um auf den politischen Druck und die entsprechenden Erwartungen zu reagieren.

Dateninnovation aus Sicht des BFS

Ausgehend von der Annahme, dass **Analytik** die Wissenschaft des Lernens aus Daten (des Datenverstehens) ist, wird **Dateninnovation** beim BFS definiert als

die Anwendung von *komplementären Analysemethoden* (z. B. Predictive Analytics [vorausschauende Analyse], die die Methoden von erweiterter Statistik, Data Science und/oder Machine Learning nutzen) auf bestehende (traditionelle) und/oder neue (nicht traditionelle) Datenquellen

zur Festigung der Rolle der öffentlichen Statistik im demokratischen Prozess in der Schweiz. Dabei muss gewährleistet sein, dass die vom BFS gelieferten Informationen zuverlässig, transparent und vertrauenswürdig bleiben.

Die Dateninnovation könnte kosteneffektivere Lösungen ermöglichen und neue Erkenntnisse bringen, die mit den traditionellen Vorgehensweisen und Analysemethoden möglicherweise nicht berücksichtigt wurden. Das Ziel besteht darin, die Qualität, den Umfang und die Kosteneffizienz von Statistiken zu erhöhen und den Aufwand für die befragten Haushalte und Unternehmen zu reduzieren. Die Dateninnovation birgt zudem das Potenzial, die Aktualität und den Detaillierungsgrad der öffentlichen Statistiken zu verbessern.

Aktuelle rechtliche Grundlage

Gemäss Bundesstatistikgesetz ermitteln das BFS und andere Institutionen der öffentlichen Statistik in fachlich unabhängiger Weise repräsentative Ergebnisse über den Zustand und die Entwicklung wichtiger Bereiche für datenbasierte politische Entscheidungen, wie Bevölkerung, Wirtschaft, Gesellschaft, Bildung, Forschung, Raum und Umwelt in der Schweiz. Um seinen Auftrag zu erfüllen, verwendet das BFS bereits heute verschiedene Datenquellen. Gemäss dem Gesetz muss das BFS vorgängig prüfen, ob beim Bund, den Kantonen oder Gemeinden oder bei anderen juristischen Personen des öffentlichen Rechts bereits Register- oder Verwaltungsdaten in der nötigen Qualität verfügbar sind. Nur wenn dies nicht der Fall ist und das BFS gemäss dem Mehrjahresprogramm beauftragt ist, statistische Daten zu liefern, sammelt das BFS mittels Volkszählungen und Erhebungen eigene Daten. Dabei werden die Zahl und die Art der Erhebungen auf das Nötigste beschränkt. Die vom BFS erhobenen Daten haben den Vorteil, dass sie ein klares statistisches Ziel verfolgen und dass die Sinnhaftigkeit bzw. die Vertrauenswürdigkeit der Daten und die entsprechende Datenqualität unter der Aufsicht des BFS stehen.

Primär- und Sekundärdaten

Im Zusammenhang mit Dateninnovation teilt das BFS Datenquellen in Primär- und Sekundärdaten ein.

Sogenannte («geschaffene» oder «erhobene») **Primärdaten** werden vom BFS zu statistischen Zwecken erhoben, um die Gültigkeit von spezifischen bestehenden Annahmen anhand der Umsetzung theoretischer Konzepte zu erklären und zu überprüfen. Der Erkenntnisgewinn aus solchen Primärdaten wird als Primäranalyse (oder Top-down-Analyse, d.h. erklärende und bestätigende Analyse) bezeichnet. Das entsprechende Analysemodell ist das «deduktive Vorgehen», das von einer Annahme bzw. einer Theorie ausgeht. Beispiele für Primärdaten sind traditionelle Datenquellen wie Volkszählungen und Erhebungen, die das BFS für statistische Zwecke durchführt.

Sogenannte (beobachtete, «gefundene» oder «organische») **Sekundärdaten** werden aus anderen Gründen erhoben (oft nicht unter der Aufsicht des BFS) und können zum Entwickeln neuer Annahmen bzw. Theorien genutzt werden. Der Erkenntnisgewinn aus solchen Sekundärdaten wird als Sekundäranalyse (oder Bottom-up-Analyse, d.h. explorative und vorausschauende Analyse) bezeichnet. Das entsprechende Analysemodell ist das «induktive Vorgehen», das von den Daten ausgeht («data first»). Beispiele für Sekundärdaten sind nicht traditionelle Datenquellen wie BFS-interne und externe Registerdaten, Verwaltungsdaten und andere digitale Daten von Geräten, Maschinen, Sensoren, Satelliten, Drohnen und den sozialen Medien. Da die Sekundärdaten nicht für die direkte Nutzung im Produktionssystem der öffentlichen Statistik ausgelegt sind, müssen sie so angepasst werden, dass statistische Rückschlüsse, d.h. Schlussfolgerungen zum Zweck der öffentlichen Statistik mittels deduktivem Vorgehen, möglich sind.

Dazu können die Sekundärdaten weiter in identifizierbare Daten und nicht identifizierbare Daten unterteilt werden. **Identifizierbare Daten** können sinnvoll einer einzelnen Einheit an einem bestimmtem Standort und zu einem gegebenen Zeitpunkt zugeordnet werden, etwa einer Einzelperson, einer Institution, einem Produkt oder einem Ort (z. B. Registerdaten, Verwaltungsdaten, Satellitenbilder, raumbezogene Daten und Barcodes von Produkten). **Nicht identifizierbare Daten** lassen sich nicht genau bestimmen (z. B. Google Trends, Twitter Feeds und andere Formen der sozialen Medien). Identifizierbare Sekundärdaten könnten zum Zweck der statistischen Rückschlüsse angepasst werden, wenn ihre Sinnhaftigkeit bzw. ihre Vertrauenswürdigkeit überprüft wurde (wie dies bei der aktuellen Nutzung der BFS-internen Registerdaten der Fall ist), während nicht identifizierbare

Sekundärdaten für statistische Rückschlüsse nur begrenzt genutzt werden können, da ihre Sinnhaftigkeit bzw. ihre Vertrauenswürdigkeit nicht beurteilt werden kann.

Der induktiv-deduktive Verfahrenszyklus

Es ist wichtig zu wissen, dass sich die beiden Vorgehensweisen der Analytik (d.h. induktives und deduktives Verfahren) *ergänzen und schrittweise hintereinander bzw. sequentiell angewendet werden sollten*, um kontinuierliche Verbesserungen und datenbasierte politische Entscheidungen zu ermöglichen. Das bedeutet, dass die derzeit beim BFS angewendeten Analysemethoden auch künftig gebraucht werden, zusammen mit komplementären Analysemethoden.

So können beispielsweise die im induktiven Verfahren erzeugten Informationen (Erkenntnisse) nicht erklären, ob und weshalb diese Erkenntnisse nützlich sind und inwieweit sie stichhaltig sind. Das deduktive Verfahren wird gebraucht, um die Erkenntnisse zu bestätigen und die Qualität der auf diesen Erkenntnissen basierenden Entscheidungen zu beurteilen.

Erstes strategisches Ziel

Das erste strategische Ziel des statistischen Mehrjahresprogramms des Bundes 2016–2019 besteht darin, die statistische Basisproduktion dauerhaft sicherzustellen und sie laufend den Nutzerbedürfnissen anzupassen. Bei der Festlegung der Umsetzungsschritte wurde unter anderem vorgeschlagen, Leitlinien zur potenziellen Nutzung neuer Datenquellen zu entwickeln.

Das erste strategische Ziel der aktuellen Dateninnovationsstrategie besteht somit darin, aufzuzeigen, dass **Dateninnovation ein wichtiges strategisches Anliegen ist**.

Das erste strategische Ziel wird zudem entsprechend der Umsetzung der nachfolgend definierten Schritte fortlaufend überarbeitet und aktualisiert werden.

Strategisches Ziel 1: Richtlinien für die Dateninnovation entwickeln und die Machbarkeit der Anwendung von komplementären Analysemethoden auf bestehende (traditionelle) und/oder neue (nicht traditionelle) Datenquellen untersuchen, mit dem Ziel, die Produktion bestehender öffentlicher Statistiken dort zu erweitern und/oder zu ergänzen, wo Dateninnovation sinnvoll ist.

In einem **ersten Schritt** sollen mit BFS-internen Pilotprojekten komplementäre Analysemethoden auf bestehende (traditionelle) BFS-interne Primärdatenquellen und bereits verknüpfte identifizierbare Sekundärdatenquellen (falls vorhanden) angewendet werden. Dabei gilt es, die Verwendung bestehender Datenquellen zu erweitern und/oder zu ergänzen, indem sie auf neue Weise genutzt werden, um praktische Erfahrungen zu sammeln und eine Bestandsaufnahme der bestehenden Herausforderungen, Ressourcen, Fähigkeiten und Technologien zur Förderung der Dateninnovation und Überprüfung «schneller Erfolge» («quick-wins») zu machen.

Sofern machbar, könnte dann in einem **zweiten Schritt** die bestehende Statistikproduktion des BFS mit Dateninnovation ergänzt und/oder erweitert werden, indem komplementäre Analysemethoden auf zusätzliche, beim BFS bereits vorhandene Sekundärdaten angewendet werden.

Schliesslich könnten in einem **dritten Schritt** die komplementären Analysemethoden nur auf neue, beim BFS bis dahin noch nie verwendete Sekundärdaten angewendet werden, um in bestimmten Bereichen neue statistische Informationen und Statistiken zu untersuchen und zu produzieren.

Somit soll im Rahmen der Dateninnovationsstrategie des BFS vorzugsweise in der folgenden Reihenfolge auf die Datenquellen zurückgegriffen werden:

1. BFS-interne Primärdatenquellen sowie bereits verknüpfte identifizierbare Sekundärdatenquellen, die in der aktuellen Statistikproduktion des BFS bereits verwendet werden;
2. zusätzliche Sekundärdatenquellen, die beim BFS bereits verwendet werden;
3. neue, beim BFS bis dahin nicht verwendete Sekundärdatenquellen.

Mit dieser Reihenfolge sollen die Effizienz gesteigert, die Kosten gesenkt und der administrative Aufwand für Unternehmen und Einzelpersonen auf ein Minimum reduziert werden. Darüber hinaus soll sie ein besseres Verständnis der Methoden, Technologien und Tools ermöglichen, ohne die Komplexität zu erhöhen. Dies wäre der Fall, wenn direkt mit dem dritten Schritt gestartet würde, da dies rechtliche, technologische (IT) und andere damit verbundene Probleme nach sich ziehen könnte. Darüber hinaus wird es mit zunehmender Erfahrung mit komplementären Analysemethoden leichter werden, neue, auf die spezifischen Ziele der öffentlichen Statistikproduktion ausgerichtete Datenquellen zu finden.

Zweites strategisches Ziel

Die Dateninnovation erfordert einen Paradigmenwechsel, indem deduktive und induktive Vorgehensweisen kombiniert werden, und wird zwangsläufig dazu führen, dass sich die Produktion und Kommunikation von öffentlichen Statistiken dadurch verändert.

Es gilt festzuhalten, dass es in der öffentlichen Statistikproduktion keinen allgemeingültigen methodischen Ansatz für die oben aufgezeigten Vorgehensweisen zur Nutzung der Dateninnovation gibt. Mit jeder Anwendung von Dateninnovation und jeder Art von Daten sind eigene spezifische Herausforderungen und Möglichkeiten verbunden. Um Dateninnovation zu ermöglichen, müssen die aktuellen Prozessabläufe und Prozessmodelle beim BFS allenfalls angepasst und erweitert werden.

Dies bedeutet, dass die **Kommunikation** bei diesen Themen entscheidend sein wird. Dies führt zum zweiten strategischen Ziel.

Strategisches Ziel 2: BFS-interne und externe Kommunikationsmassnahmen entwickeln und umsetzen, um das Bewusstsein für den Mehrwert von Dateninnovation in der öffentlichen Statistik und des entsprechenden Paradigmenwechsels zu stärken.

Übergreifende Grundsätze

Bei der Umsetzung der Dateninnovation misst das BFS den Grundwerten Datenschutz, Datenschutzrichtlinien, Datensicherheit, Data Governance und öffentliches Vertrauen grösste Bedeutung bei.

Umsetzung der strategischen Ziele

Um die strategischen Ziele zu erreichen und Dateninnovation zu ermöglichen, sind folgende spezifischen Massnahmen nötig:

- evaluieren, ob das aktuelle Datenqualitätssystem des BFS angepasst und erweitert werden muss, um die auch Dateninnovation abzudecken (z. B. Berücksichtigung der komplementären Analysemethoden wie vorausschauende Analysen);

- evaluieren, ob beim BFS-Personal Bedarf an neuen Fähigkeiten im Bereich der komplementären Analysemethoden sowie der (IT-)Technologien und Tools besteht, die für die Dateninnovation benötigt werden. Der oben genannte erste Schritt und die zugehörigen Pilotprojekte sind bei der Bestimmung der für die Dateninnovation benötigten Fähigkeiten massgebend. Nach der Umsetzung ist das BFS bei Bedarf befähigt:
 - ein neues BFS-internes Ausbildungsprogramm für BFS-Mitarbeitende im Bereich der komplementären Analysemethoden sowie der dazugehörigen Kompetenzen («soft skills») wie Kreativität, Zusammenarbeit, Neugier, Problemlösungsorientiertheit, Kommunikation und «Storytelling» zu entwickeln («qualifiziertes Personal»), beispielsweise mit Privatunternehmen bzw. Universitäten;
 - die Nutzung eines BFS-internen und/oder externen «Data Labs» («Sandbox») festzulegen und umzusetzen, zur praktischen Umsetzung der neuen, für die Anwendung der komplementären Analysemethoden benötigten Fähigkeiten;
 - den Anpassungsbedarf bei den (IT-)Technologien, der BFS-internen und/oder externen Infrastruktur und der Tools zu bestimmen;

- die rechtlichen Fragen bezüglich Dateninnovation klären, z. B. die rechtlichen Voraussetzungen und Anreize, um das Einverständnis zum Erhalt von Sekundärdaten des Privatsektors einzuholen («Datenhoheit»), ohne dem Geschäftsinteresse der Datenbesitzer zu schaden (für den oben erwähnten dritten Schritt);

- den Informationsaustausch mit allen BFS-Stakeholdern innerhalb des BFS, des statistischen Systems und der Forschungsgemeinschaft sicherstellen.

Erfolgreiche und anhaltende multidisziplinäre Partnerschaften beim Bund sowie in industriellen, akademischen und statistischen Kreisen sind für das BFS zentral, um die für den Erfolg seiner Dateninnovationsstrategie nötigen Kompetenzen und Erfahrungen zusammenzubringen.

Bestellungen

Tel. 058 463 60 60

Fax 058 463 60 61

order@bfs.admin.ch

Preis

Gratis

Download

www.statistik.ch (gratis)

BFS-Nummer

1787-1700

**Statistik
zählt für Sie.**

www.statistik-zaehlt.ch