

Die amtliche Statistik in digitalen Informationsdiensten

E-Government-Strategie

Auf die Herausforderungen des gesellschaftlichen Wandels zur digitalen Kommunikation haben die staatlichen Stellen reagiert und mit der E-Government-Strategie sowie den Grundsätzen von Open Government und Open Data neue Wege eröffnet.

Die amtliche Statistik verfügt über umfangreiche Daten- und Informationsbestände, die nützliches und unverzichtbares Wissen um Struktur und Entwicklung von Wirtschaft und Gesellschaft sowie für die Planung und Steuerung des Gemeinwesens bereitstellen. Die Veröffentlichung und Verbreitung dieser Datenbestände gehört zum gesetzlichen Auftrag der amtlichen Statistik. Diesem hat sie stets durch eine umfangreiche Publikationstätigkeit entsprochen, die auch die Bereitstellung von Informationsangeboten in elektronischen Medien und im Internet umfasst. Die rasch fortschreitende Digitalisierung der Kommunikationswege und -mittel hat zu einem grundlegenden Wandel in der Art und Weise geführt, wie in der Öffentlichkeit und durch Experten heute Informationen genutzt und verarbeitet werden. Während die Produktion statistischer Informationen schon seit langem in elektronischer Form erfolgt, steht heute die umfassende Einführung der digitalen Veröffentlichung und Verbreitung amtlicher Statistiken an.

Die nationale E-Government-Strategie von Bund, Ländern und Gemeinden vom 24. September 2010 gibt als Leitbild vor, dass E-Government

- am Nutzen für Bürgerinnen und Bürger, Unternehmen und Verwaltung orientiert ist,
- wirtschaftlich und effizient arbeitet,
- Transparenz über Daten und Verwaltungshandeln sowie den Datenschutz sicherstellt,
- die gesellschaftliche Teilhabe der Bürgerinnen und Bürger und Unternehmen unterstützt,
- zukunftsfähige Lösungen fördert und
- leistungsfähige IT-Unterstützung realisiert.

Diesem Leitbild entspricht der Informationsbestand der amtlichen Statistik bislang durch eine umfangreiche Präsentation der Statistiken online und in nationalen und internationalen Datenbanken. Die Entwicklung eines umfassenden und qualitativ hochstehenden Angebots an digitaler Information erfordert eine genaue Analyse des Nutzerbedarfs und muss die IT-technischen Voraussetzungen unterschiedlicher Nutzergruppen einbeziehen.

Spezifische Angebote für Nutzergruppen

Statistische Informationen haben die Eigenheit, dass sie nur mit einem erheblichen Aufwand an Sachwissen und IT-technischer Verarbeitung vollständig nutzbar sind. Die Daten amtlicher Statistik liegen nach Erhebung, Aufbereitung, Plausibilisierung, oft auch noch mit zusätzlichen Berechnungen versehen, zunächst als Dateien vor. Diese Datensätze werden heute übli-

cherweise als statistische Mikrodaten bezeichnet. Hinzutreten muss immer die Beschreibung der Daten – die Metadaten und die Merkmalsdefinitionen –, die überhaupt erst die sachliche Zuordnung der Zahlenwerte zu Aussageinhalten ermöglichen.

Jede Verbreitung statistischer Daten unterliegt weiterhin der Einschränkung, dass statistische Aussagen aus Gründen des Datenschutzes keinen Rückschluss auf den zugrundeliegenden Einzelfall zulassen dürfen. Schließlich stellt jede einzelne Aufbereitung und Interpretation statistischer Ergebnisse lediglich einen Ausschnitt aller potenziell möglichen Merkmalskombinationen und Auswertungen dar. Das setzt einen grundlegenden Widerspruch bei der Verbreitung statistischer Informationen: Je stärker die Geheimhaltung und je verständlicher und interpretativer das Material aufbereitet ist, desto weniger tief ist das Informationspotenzial erschließbar. Eine Lösung findet dieses Problem darin, dass für die jeweiligen Zwecke die digitale Information in je spezifischer Form aufbereitet und verbreitet wird. Hierfür sind pragmatisch Nutzergruppen zu unterscheiden, die hinsichtlich Informationsinteresse, Sachwissen und IT- und Methodenkompetenz jeweils unterschiedlich positioniert sind. Mit Bezug auf die Aufbereitung und Verbreitung statistischer Daten ist es also sinnvoll, mindestens vier Nutzergruppen zu unterscheiden, um die jeweils angemessene Form der Verbreitung für alle zu garantieren:

1. die Bürgerinnen und Bürger als breite Öffentlichkeit,
2. die wissenschaftliche Forschung,
3. Open Data und
4. die Expertennutzer.

Digitale Informationen für die Öffentlichkeit

Die statistischen Rohdaten im Zustand unmittelbar nach Erhebung, Verarbeitung und Plausibilisierung galten bislang im Prinzip als nicht öffentlich verbreitbar. Das Produkt des Aufbereitungsprozesses der Statistik ist deshalb stets die Erstellung von Tabellen: die Darstellung statistischer Daten in Zeilen und Spalten, mit Überschriften, die Ort, Zeit und Sachbezug festlegen, mit Erläuterungen und Fußnoten, die Metadaten enthalten. Hierdurch werden die statistischen Rohdaten inhaltlich gegliedert und ausgewählt, jede Tabelle stellt eine bestimmte Sicht auf die Rohdaten dar. Dadurch wird das Informationspotenzial einer Statistik über alle Merkmale und Ausprägungen einer Statistik in der Regel nur sehr ausschnitthaft erfasst. Zugleich macht die tabellarische Aufbereitung

die Rohdaten für die Nutzer überhaupt erst zugänglich und sinnhaft verstehbar. Typische Produkte sind hier die Statistischen Jahrbücher, Statistische Berichte, Querschnittsberichte und Pressemitteilungen. Diese werden noch als Printprodukte, aber zunehmend mehr als CD oder online in geschützten, aber z.T. auch in weiterbearbeitbaren Formaten verbreitet.

Diese Form der Aufbereitung ist für die erste Nutzergruppe geeignet, worunter die Öffentlichkeit, also die Bürgerinnen und Bürger, die Medien und für vielfache Zwecke auch die Wirtschaft, Verbände, Bildungseinrichtungen und Verwaltungen fallen. Im Zuge der Verbesserung der digitalen Aufbereitungsformen wird hier in Zukunft vermehrt die tabellarische Darstellung durch themenspezifische Aufbereitung mit Grafiken, Karten und interpretativen Texten ergänzt. Die Angebote des Amtes für Statistik Berlin-Brandenburg (AfS) werden zunehmend durch Gemeinschaftsveröffentlichungen der Statistischen Ämter des Bundes und der Länder im Statistik-Portal ergänzt.

Forschungsdatenzentren für Wissenschaftler

Für die Erschließung des vollen Informationsgehaltes der amtlichen Statistik ist der Zugriff auf die Mikrodaten, also die statistischen Rohdaten nach Aufbereitung und Plausibilisierung, erforderlich. Seit zehn Jahren stellt die amtliche Statistik in den Forschungsdatenzentren gesicherte Zugangswege für die wissenschaftliche Forschung zur Verfügung. Hier ist der Zugriff auf die Mikrodaten für eine durch Gesetz privilegierte Nutzergruppe in faktisch anonymisierter Form als Scientific Use File, an abgeschotteten Gastwissenschaftlerarbeitsplätzen oder als Fernverarbeitung möglich.

Open Data

In einer offenen, demokratischen Gesellschaft verlangt die in digitaler Kommunikation geschulte und gewohnte Öffentlichkeit auch einen Zugang zu den digitalen Informationen der staatlichen Stellen, also auch zur amtlichen Statistik. Eine dritte Nutzergruppe aus interessierter Öffentlichkeit und Wirtschaft ist in den letzten Jahren entstanden, die den Zugang nach den Grundsätzen des Open Data fordert. Hierbei handelt es sich um versierte IT-Experten, die die Rohdaten der amtlichen Statistik als Input für neu erstellte Anwendungen nutzen wollen. Diese Gruppe stellt, wie die Forschung, über die Metadaten hinaus keine vertieften Anforderungen an Interpretation oder Aufbereitung der Daten durch die Statistischen Ämter.

Die Grundsätze des digitalen Zugangs zu Veröffentlichung und Verbreitung amtlicher Daten sind in den Prinzipien des „Open Government“ bzw. „Open Data“ niedergelegt. Diese können auch auf die Verbreitung der amtlichen Statistik Anwendung finden:

- **Vollständig:** Alle amtlichen Statistiken werden verfügbar gemacht, sofern sie nicht berechtigten Datenschutzerfordernissen unterliegen.
- **Primär:** Die Daten werden medienbruchfrei aus den Quellverfahren heraus zur Verbreitung aufbereitet, dies geschieht unter Beachtung der Statistischen Geheimhaltung im feinst möglichen Detaillierungsgrad.
- **Aktuell:** Die Daten werden ohne Zeitverzug verbreitet.

- **Zugänglich:** Die Daten werden unter Beachtung der Statistischen Geheimhaltung dem größtmöglichen Nutzerkreis ohne Einschränkung der Nutzungszwecke zur Verfügung gestellt.
- **Maschinenlesbar:** Die zur Verbreitung gewählten Datenformate erlauben eine automatisierte Weiterverarbeitung.
- **Nicht diskriminierend:** Die Daten sind unter den jeweiligen Bedingungen der statistischen Geheimhaltung für alle Nutzer gleich verfügbar.
- **Nicht proprietär:** Die Daten werden in standardisierten Formaten zur Verfügung gestellt.
- **Creative Commons Lizenz:** Das Basisangebot der statistischen Daten der amtlichen Statistik wird unter offenen Lizenzen bereitgestellt.

Auf Ebene der Verwaltungen der Länder Berlin und Brandenburg sind weite Teile des Datenbestandes noch nicht den Kriterien von Open Data und Open Government entsprechend organisiert. Hierfür sind Online-Dienste und -Informationen zu entwickeln, durch die die öffentlichen Datenbestände für Bürgerinnen und Bürger und Unternehmen sowie Verwaltungsstellen in den Ländern Berlin und Brandenburg in maschinenlesbaren Formaten über offene Schnittstellen unter Wahrung der statistischen Geheimhaltung verfügbar gemacht werden können. Die erforderliche Geheimhaltung der statistischen Daten stellt hier eine Begrenzung des Umfangs der unter Open Data zu vertreibenden statistischen Daten dar. Es können nur Datensätze bereitgestellt werden, bei denen nicht auf Einzelfälle geschlossen werden kann. Die Geheimhaltung muss hier auf den Datensatz insgesamt angewendet werden.

Expertennutzer

Eine tiefer greifende Erschließung des Informationsgehalts amtlicher Statistiken erforderte bisher fast immer die Erstellung spezieller nachfrageorientierter Tabellen, sogenannter Sonderauswertungen. Dies ist für die vierte Nutzergruppe, die Expertennutzer aus Wissenschaft (soweit sie nicht die Forschungsdatenzentren nutzen), Verwaltung und Wirtschaft bislang der einzige Weg, um an Informationen zu gelangen, die nicht bereits in den veröffentlichten Basisdaten enthalten sind. Hierbei handelte es sich um auftragsorientierte, nutzerspezifische Einzelanfertigungen von Tabellen.

Das tabellarische Aufbereitungsprinzip ist bislang in der Regel auch in die digitale Aufbereitung der amtlichen Statistik übertragen worden. In Datenbanken wurden mehrdimensionale Tabellen – Datenquader – gespeichert und über das Internet den Expertennutzern zugänglich gemacht. Dies hat den Umfang der verbreiteten statistischen Daten zwar in neue Dimensionen vergrößert und den Expertennutzern ein erheblich vergrößertes Angebot zur Verfügung gestellt, die prinzipielle Beschränkung der tabellarischen Aufbereitung wurde aber dadurch noch nicht überwunden.

Die in der Regel sehr umfangreichen statistischen Informationsanforderungen der Expertennutzer aus dem Kreise der Bürgerinnen und Bürger, Unternehmen und Verwaltungen können derzeit lediglich über kosten- und zeitaufwändige Sonderauswertungen des statistischen Einzelmaterials abgedeckt

werden. Die Erschließung eines datenbankgestützten Informationssystems für diese Nutzergruppe ist der gegebene Weg, um auch hier zu einer rationellen und kostengünstigen Informationsversorgung zu kommen, die den Kriterien der Qualitätsanforderungen der amtlichen Statistik entspricht.

Hier setzt das Statistische Informationssystem an, das Expertennutzern den Zugang zum gesamten Informationsgehalt der amtlichen Statistik ermöglichen will. Dieses Angebot ist für die Inhalte der amtlichen Statistik ebenfalls öffentlich im Internet zugänglich, es liegt also in der Entscheidung der Nutzer

selbst, ob sie sich als Experten versuchen wollen. Das Statistische Informationssystem wird darüber hinaus auch für Nutzerkreise der Landesverwaltungen mit spezifischen Inhalten implementiert.

*Autoren: Hartmut Bömermann,
Peter Lohauß*