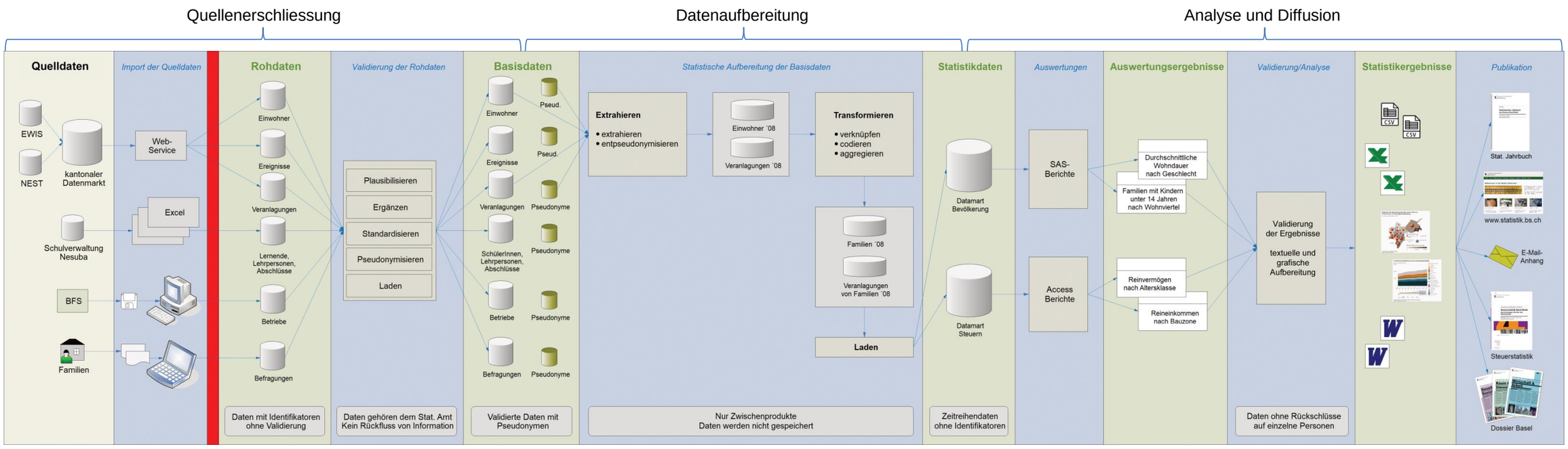
Präsidialdepartement des Kantons Basel-Stadt

Statistisches Amt

Datenfluss-Diagramm – Prozess der Statistikerstellung

Der Prozess der Statistikerstellung lässt sich grob in die drei Hauptprozessschritte «Quellenerschliessung», «Datenaufbereitung» sowie «Analyse und Diffusion» untergliedern. Nachfolgend sind diese drei Schritte im Detail und mit Bezug zu den entsprechenden Datenbeständen und Arbeitsschritten beschrieben. Die Aufbereitung der Daten schliesst die parallele Aufbereitung, Haltung und Nachführung der Metadaten mit ein. Metadaten sind Definitionen und weitere Angaben, die für die korrekte Interpretation der erarbeiteten statistischen Information notwendig sind.



Quellenerschliessung

Quelldaten bezeichnen …

- Daten, die im normalen Verwaltungsvollzug anfallen, wie etwa die Daten über die Bevölkerung im Einwohnerregister,
- Daten von anderen Statistikproduzenten (vornehmlich Bundesamt für Statistik),
- Daten, die im Rahmen von Befragungen erhoben werden.

Der *Import* erfolgt über verschiedene Kanäle wie Web-Services, Excel-Dateien, Papierformulare, etc.

Die importierten Quelldaten liegen nun als Kopie beim Statistischen Amt. Sie befinden sich im selben Zustand, in dem sie importiert worden sind und werden von diesem Moment an als **Rohdaten** bezeichnet.

Bei der *Validierung* werden die Rohdaten einer Qualitätskontrolle unterzogen, pseudonymisiert und als **Basisdaten** in einer Auswertungsdatenbank abgelegt. Der Zuordnungsschlüssel wird in einer separaten Datei, von den Basisdaten getrennt, abgelegt.

Datenaufbereitung

Die für eine Auswertung benötigten Basisdaten werden für die Statistische Datenanalyse vorbereitet. Dieser Arbeitsschritt wird ausschliesslich durch das Datenmanagement-Team durchgeführt.

Falls verschiedene Datenbestände miteinander verknüpft werden müssen (z. B. um das Durchschnittseinkommen verschiedener Haushaltstypen kleinräumig darzustellen), werden die Basisdaten temporär entpseudonymisiert. Die in der statistischen Aufbereitung anfallenden Zwischenprodukte werden nicht gespeichert.

Die verknüpften Daten werden ohne Identifikatoren abgelegt. Ab diesem Zeitpunkt werden sie als **Statistikdaten** bezeichnet.

Die Statistikdaten werden als Zeitreihen abgelegt und gepflegt, d. h. immer bei Vorliegen aktueller Daten (monatlich, quartalsweise, jährlich, usw.) werden die Zeitreihen ergänzt (z. B. Bevölkerungsstand am Monatsende, Leerwohnungsbestand per Stichtag).

Analyse und Diffusion

Die Statistikdaten sind der Ausgangspunkt für sämtliche Auswertungen. Darunter fallen alle Abfragen der Statistikdaten durch die wissenschaftlichen Mitarbeitenden des Statistischen Amtes.

Durch Abfragen der Statistikdaten entstehen die so genannten **Auswertungsergebnisse**. Ein Beispiel: Das durchschnittliche Reineinkommen von Dreipersonenhaushalten im Wohnviertel Gundeldingen beträgt xyz Franken.

- Im Arbeitsschritt *Validierung/Analyse* werden die Auswertungsergebnisse auf Anonymität (kein Rückschluss auf Einzelpersonen) und Aussagekraft geprüft. Zudem werden die Auswertungsergebnisse textlich und grafisch aufgearbeitet.

Nach der Analyse liegen die Informationen nun in der gewünschten resp. benötigten Form vor (Excel-Tabelle, Grafik, Text, Karte). Sie sind validiert und zur Veröffentlichung bereit. In diesem Moment wird von **Statistikergebnissen** gesprochen.

- Im Arbeitsschritt *Publikation* werden die Statistikergebnisse der Öffentlichkeit resp. dem Auftraggeber bekannt gemacht. Beispiele dafür sind das Statistische Jahrbuch, das Dossier Basel, verschiedene Kennzahlenberichte oder die Internetseite des Statistischen Amtes.