

Empfehlungen für die Abgabe von statistischen Mikrodaten vom Statistischen Verbund an die Landesarchive

Erarbeitet vom Ausschuss „Digitale Archive“ und vom Ausschuss „Records Management“

Status: Endfassung

Stand: 25.2.2018

Redaktion: Dr. Kai Naumann

Seit 2016 liegt ein überarbeitetes Bewertungsmodell der KLA für die Mikrodaten der amtlichen Statistik vor.¹ Vorliegende Empfehlungen sind als Arbeitshilfe gedacht, um die Umsetzung des Bewertungsmodells durch die Landesarchive mit den statistischen Ämtern des Bundes und der Länder zu fördern.

Übergabeverfahren

1. Das Verfahren zur Übergabe der Datenmaterialien ist vorab zwischen Landesarchiv und Statistischem Landesamt abzustimmen. Anzustreben ist ein hinreichender Schutz gegen Ausspähung und Datenverluste. Die Daten dürfen in den statistischen Ämtern erst gelöscht werden, wenn das Landesarchiv die erfolgreiche Übertragung und fehlerfreie Sicherung schriftlich bestätigt hat.

Datenformate (Mikrodaten)

2. Die Datenmaterialien (Mikrodaten, Rohdaten) werden als Textdatei im Zeichenformat ASCII, UTF-8 oder Latin-1 (Codepage 1252) übergeben. Das Zeichenformat wird von der abgebenden Stelle ausdrücklich angegeben. UTF-8 wird bevorzugt.
3. Die Dateinamen werden so vergeben, dass folgende Merkmale daraus hervorgehen:
 - Bezugszeitraum
 - EVAS-Nummer
 - gegebenenfalls Satzart
4. Die Dateiendung ist bei Festbreitenformaten TXT und bei CSV-formatierten Daten CSV.
5. Das CSV-Format folgt der Richtlinie IETF RFC 4180. Die einzelnen Felder enthalten in der ersten Zeile Feldnamen oder Feldkennungen (z.B. „MF14“, „Geschlecht“).
6. Bei Festbreitenformaten, die verschiedene Satzarten kennen, ist jede Satzart in einer separaten Datei zu übergeben.

¹ Vgl. bei der KLA-Arbeitsgruppe „Bewertung von Statistikunterlagen“ unter <http://www.bundesarchiv.de/DE/Navigation/Meta/Ueber-uns/Partner/KLA/kla.html>.

7. Zahlenkompressionsverfahren (auch bekannt als „gepackte“ Zahlen, z.B. zwei Ziffern in einem Byte) sind vor Ablieferung aufzulösen. Das heißt, Zahlenwerte in den Daten werden stets als gültige Zahlzeichen des verwendeten Zeichenformats übergeben (z.B. der Wert „6798“ mit den Integer-Zahlenwerten 6, 7, 9 und 8).
8. In Absprache mit dem zuständigen Archiv können die Mikrodaten auch als XML übergeben werden. Das XML-Schema ist mitzuliefern. Verbreitete Schemata sind gegenüber proprietären Schemata zu bevorzugen.

Metadaten

9. Die Metadaten aus dem Archivierungsmanagementsystem (AMS) der Statistischen Ämter sind in dem in AMS verwendeten XML-Format an die Archive abzugeben. Vorhandene XML-Schemadateien sind mitzuliefern. Die Metadaten enthalten eine Referenz auf die Datenmaterialien (z.B. per Dateinamen). Zusätzlich werden die Metadaten aus dem AMS auch als Textdokumente im Format PDF/A abgegeben.
10. Bei Festbreitenformaten ist zu dokumentieren, an welcher Position welche Werte vorliegen (z.B. Stelle 1-4: laufende Nummer, Stelle 5-25: Nachname, Stelle 26-50 Vorname etc.).
11. Bei Festbreitenformaten, die verschiedene Satzarten kennen, sind die Angaben zu Nr. 10 für alle vorliegenden Satzarten zu wiederholen.
12. Metadaten zur Beschreibung der Mikrodaten müssen enthalten:
 - Beschreibung der Datei
 - Zeichencodierung (vgl. Nr. 2)
 - Gliederungstyp (Festbreite oder CSV)
 - Zeilenanzahl der Datei
 - Trennzeichen zwischen den Zeilen
 - ggf. Trennzeichen zwischen den Feldern (nur bei CSV, vgl. Nr. 5)
 - Mehrere Satzarten in einer Datei?
 - Sonstige Informationen (z.B. bekannte technische Probleme, ungewöhnliche Byte-Reihenfolge etc.)
 - Beschreibung der einzelnen Felder der Datei
 - Feldnamen (z.B. „MF14“)
 - Beschreibung des Felds (z.B. „Anzahl der Pendler pro Tag“)
 - Datentyp jedes Feldes
 - Versatz (Offset) in Bytes (Startposition) für jedes Feld (nur bei Festbreite, vgl. Nr. 10)
 - Breite in Bytes für jedes Feld (nur bei Festbreite, vgl. Nr. 10)
 - Information, dass ein Feld ehemals komprimiert war (vgl. Nr. 7)
 - Informationen darüber, welche Schlüsseldateien genutzt werden müssen (vgl. Nr. 13ff.)

Codierte Werte

13. Schlüsseldateien für codierte Werte müssen mit übergeben werden, oder die codierten Werte müssen in den Daten selbst aufgelöst werden (z.B. 1 für männlich, 2 für weiblich).
14. Schlüsseldateien (Codelisten, Gliederungen, Klassifikationen) sollen in gängigen maschinenlesbaren Formaten abgegeben werden. Dabei sind offene Formate (z.B. CSV, DatML, DDI, TXT, sonstiges XML) gegenüber kommerziellen Formaten (z.B. SAS, SPSS, STATA) zu bevorzugen.
15. Sind die Schlüsseldateien nicht maschinenlesbar (z.B. Word-Dateien, Scans von Schreibmaschinenseiten), so sind sie im Format PDF/A-1 oder PDF/A-2 an die Archive abzugeben.
16. Der Gültigkeitszeitraum von Schlüsseldateien, die sich mit der Zeit verändern, ist für die abgegebenen Datenmaterialien ausdrücklich anzugeben.

Unterstützende Zusammenarbeit

17. Für sogenannte Altdaten, die im AMS nacherfasst werden, gelten bei der Übergabe an das zuständige Archiv die oben genannten Empfehlungen ebenfalls.
18. Falls Archive zur Umsetzung dieser Empfehlungen Rückfragen haben, können sie sich zur Klärung an den KLA-Ausschuss Digitale Archive wenden.