

Anbietungs- und Archivierungsformate für Datentabellen
Erläuterungen zur Richtlinie des Landesarchivs Baden-Württemberg
AZ V-7511/9-6/Kei, Lg, Nm
Stand: 05.12.2006

Zu Punkt 2: CSV-Format

Das Format CSV ist ein platzsparendes Mittel zur Ablage textbasierter Daten in Tabellenform. Hinsichtlich der Feldtrenner (field separators, field delimiters) und der Textbegrenzer (text delimiters) existieren viele Varianten. Es kann zu Problemen kommen, wenn die zur Tabellenauszeichnung vorgesehenen Zeichen auch in den Feldern der Tabelle auftreten. Solche Felder müssen mit Textbegrenzern maskiert werden.

Für CSV bestehen keine anerkannten Normen. Häufig wird der Algorithmus von MS-Excel verwendet. Das Verhalten von Excel ist aber nicht vollständig dokumentiert.

Bei der IANA (Internet Assigned Numbers Authority) liegt seit Oktober 2005 die MIME-Type-Definition RFC 4180 vor. Sie ist von der IANA offiziell registriert, ist öffentlich zugänglich und definiert das Format ausführlich:

URL: <http://www.rfc-editor.org/rfc/rfc4180.txt>

Die Richtlinie des Landesarchivs entspricht daher weitgehend der Definition der RFC 4180. Als Feldtrenner wird jedoch in der RFC 4180 nur das Komma genannt. Da die meisten Anwendungen im deutschsprachigen Raum stattdessen ein Semikolon benutzen, schreibt das Landesarchiv ebenfalls die Benutzung eines Semikolon vor. Außerdem ist die Einfügung von Feldnamen in der ersten Tabellenzeile verpflichtend.

Zu Punkt 3: Beschreibung von Feldern

Ein Nachteil des CSV-Formats ist die mangelhafte Definition des Feldinhalts. Für die Nutzung einer archivierten Datentabelle sind zusätzliche Informationen erforderlich. Diese Angaben sind daher in einer eigenen CSV-Datei an das Landesarchiv abzugeben.

Zu Punkt 4: Alternativen

Bei komplexeren Datenbanken kann es sinnvoll sein, die Daten in XML oder einem anderen geeigneten Format abzugeben. In diesen Fällen bedarf es weiterer Vereinbarungen zwischen der abgebenden Behörde und dem Landesarchiv.