

Handreichung zu Musterformaten für die Digitalisierung

Die Diskussion, welches Dateiformat im Rahmen von archivischen Digitalisierungsprojekten genutzt werden soll, ist vermutlich so alt wie die Digitalisierung selbst. Insbesondere die Gegensatzpaare unkomprimiert/komprimiert sowie verlustfreie/verlustbehaftete Komprimierung standen dabei im Mittelpunkt. Ursprünglich sehr dogmatisch geführt, hat die internationale Diskussion der letzten Jahre gezeigt, dass die frühere, prinzipiell kritische Sicht auf Komprimierung mehr und mehr hinterfragt wird.¹ Die Frage sollte deshalb inzwischen nicht mehr lauten, ob sich ein Digitalisat vom Original unterscheiden darf, sondern welcher Umfang an Veränderung noch akzeptabel ist.

Bei der Frage nach dem einzusetzenden Format müssen verschiedene Kriterien gegeneinander abgewogen werden. Die Schweizer Koordinationsstelle für die dauerhafte Archivierung elektronischer Unterlagen (KOST) zählt dazu u. a. Lizenzfreiheit, Offenheit, Verbreitung, Speicherdichte und Verifizierbarkeit. Ein Kernkriterium ist außerdem nach wie vor der Grad des Verlustes an Informationen. Aufgrund schlechter Werte insbesondere in dieser Kategorie scheidet eine Vielzahl von Formaten, darunter JPEG, für Archivzwecke aus.

Während also von konkreten Formaten *abgeraten* werden kann, so ist hingegen eine generelle *Empfehlung* für den Einsatz eines bestimmten Dateiformats nach wie vor nicht möglich. Aufgrund der Vielfalt des analogen Ausgangsmaterials und der sich zum Teil deutlich unterscheidenden Einsatzbereiche für Digitalisate im archivischen Alltag sollte die Formatfrage vielmehr im Hinblick auf die Anforderungen des jeweiligen Verwendungszwecks beantwortet werden. Erstes Kriterium ist hierbei die Unterscheidung zwischen der Erstellung von digitalen Mastern und Nutzungsderivaten. Aktuell kommen laut DFG² und KOST³ für digitale Master lediglich drei Formate in Frage: TIFF, JPEG2000 und PDF/A-2. Für Nutzungsderivate eignet sich zudem JPEG.⁴ Die Formate sind, bei einer generellen Auflösung von 300 dpi und einer Farbtiefe von 24 Bit, jeweils für folgende Verwendungszwecke empfehlenswert.

Digitale Masterversionen

Für nicht in größerem Umfang anzufertigende digitale Masterversionen und punktuelle Schutz- und Rettungsdigitalisierung ist trotz seiner Komplexität und der Größe der generierten Dateien nach wie vor TIFF weit verbreitet und empfehlenswert. Das Format wägt seine niedrige Speicherdichte dadurch auf, dass es keine verlustbehaftete Komprimierung nutzt. Ebenso ist seine Offenheit und Verifizierbarkeit von Vorteil. Hierbei ist zu beachten, dass die Empfehlung nicht für alle unter diesem Dateiformat subsumierten Typen, sondern lediglich für definierte Baseline TIFF, die

¹ So zeigen z. B. Auswertungen unmittelbar nacheinander durchgeführter Scans desselben Originals, dass hierbei Abweichungen zwischen zwei unkomprimierten TIFFs auftraten, welche die Abweichungen zwischen komprimierten und unkomprimierten Dateien übertrafen, siehe: https://kost-ceco.ch/cms/dl/dd0ac593881c49cd9939f715f929af89/Kolloquium_Datenkomprimierung_Bericht_v1.pdf, Kolloquium Datenkomprimierung bei Bild, Audio, Video, 11. November 2009, Universität Bern, S. 3, abgerufen am 26.3.2020.

² DFG-Vordruck 12.151 - 12/16 - Praxisregeln „Digitalisierung“, S. 20-21.

³ <https://kost-ceco.ch/cms/empfehlung.html>, abgerufen am 26.3.2020.

⁴ DFG-Vordruck 12.151 - 12/16 - Praxisregeln „Digitalisierung“, S. 14.

Kompressionsverfahren LZW (5), CCITT T.4 (3) und CCITT T.6 (4) sowie Multipage-TIFFs gilt.⁵

Massendigitalisierung und Migrationsformat

Bei der regelmäßigen Erstellung großer Mengen an Digitalisaten muss ein Kompromiss zwischen der Qualität der Dateien und den Kosten für die anfallende große Menge an Speicherplatz gefunden werden. Hier sollte als Alternative zu TIFF auf das als ISO-Standard normierte JPEG2000 zurückgegriffen werden. Sein wichtigster Vorteil gegenüber TIFF ist die höhere Speicherdichte bei verlustfreier Komprimierung. Ggf. kann auch eine minimal verlustbehaftete Kompression angewendet werden⁶, da hier kaum qualitative Nachteile gegenüber unkomprimierten Dateien entstehen. Als Idealwert gilt hier die Komprimierungsstufe von 70, welche eine Speicherplatzersparnis um den Faktor 3,6 bei minimalem messbarem Verlust ergibt.⁷ JPEG2000 ist als Migrationsformat für andere Rasterbildformate geeignet und wird vom DA NRW akzeptiert.

Retrodigitalisierung von seitenbasierten Vorlagen

Bei der Retrodigitalisierung von seitenbasierten Vorlagen wie Druckschriften, Findbüchern oder Karteikarten sollte das als ISO-Standard normierte und lizenzfreie PDF/A-2 eingesetzt werden. Es kann die JPEG2000-Komprimierung nutzen, wodurch eine hohe Speicherdichte möglich wird. Da die wesentlichen signifikanten Eigenschaften von Druckschriften oder Karteikarten der Text und seine Anordnung sind und eine detaillierte fotografische Rekonstruktion nicht notwendig ist, ist für diesen Anwendungszweck eine verlustbehaftete Komprimierung unkritisch ist.⁸ Für PDF/A-2 sprechen zudem seine Verbreitung und seine Verifizierbarkeit.

Digitalisierung von Mikroformen

Die Digitalisierung von Mikrofilmen oder Mikrofiches erfolgt im Regelfall zur massenhaften Erstellung von Derivaten, z. B. für die Nutzung im Lesesaal. Die Mikroformen können hierbei als Master⁹ betrachtet werden. In diesem Fall kann unter Abwägung des Verhältnisses von Speicherplatz, Kosten und Qualitätsanforderungen auf das JPEG-Format zurückgegriffen werden.¹⁰ Eine verlustfreie Komprimierung ist bei diesem Verwendungszweck nicht angezeigt, da bereits die Mikroformen durch ihre ursprüngliche Verfilmung als verlustbehaftet gelten können. Für JPEG sprechen seine sehr weite Verbreitung, seine Lizenzfreiheit, seine hohe Speicherdichte und die Tatsache, dass es mit allen gängigen Bildbearbeitungsprogrammen gelesen werden kann. Hinzu kommt, dass JPEG als ISOStandard normiert ist.¹¹

⁵ https://kost-ceco.ch/cms/dl/358e5b8be2a188ad9be879d4075e368f/TIFF_PreservationPlanning_Studie_v1.0.pdf, TIF Format (Tagged Image File Format). Eine Preservation Planning Studie, S. 6, abgerufen am 26.3.2020.

⁶ Siehe:

https://www.bar.admin.ch/dam/bar/de/dokumente/konzepte_und_weisungen/archivtaugliche_dateiformate.1.pdf.download.pdf/archivtaugliche_dateiformate.pdf, Schweizerisches Bundesarchiv, Standards für die Archivierung digitaler Unterlagen. Archivtaugliche Formate, Version 2018/04, S. 15, abgerufen am 3.4.2020.

⁷ Kolloquium Datenkomprimierung, S. 3.

⁸ <https://kost-ceco.ch/cms/pdf-a-2-bei-bilddaten.html>, abgerufen am 26.3.2020.

⁹ Vgl. hierzu DFG-Vordruck 12.151 - 12/16 - Praxisregeln „Digitalisierung“, S. 26.

¹⁰ Zumal auch so eine vollumfängliche Übernahme der auf dem Mikrofilm enthaltenen Informationen möglich ist. Eine komprimierte Speicherung brächte insoweit genau so wenig Vorteil wie eine höhere Auflösung beim Scannen.

¹¹ <https://kost-ceco.ch/cms/jpeg.html>, abgerufen am 31.3.2020.