

## Empfohlene Datenformate für die Langzeitarchivierung

Für die Archivierung elektronischer Dokumente müssen geeignete Datenformate gewählt werden. Das Datenformat sollte langfristig eine originalgetreue Reproduktion der Archivdaten sowie ausgewählter Merkmale des ursprünglichen Dokumentmediums (z. B. Papierformat, Farben, Logos, Seitenzahl, Wasserzeichen, Unterschrift) ermöglichen (s. *Empfehlungen des Bundesamtes für Sicherheit in der Informationstechnik*). Grundsätzlich ist die Verwendung von offengelegten, nicht-proprietären und standardisierten Dateiformaten zu empfehlen.

### Strukturformate:

Speziell für die Anforderungen der Langzeitarchivierung wurde PDF/A (Portable Document Format, A für Archivierung) standardisiert. PDF/A enthält alle erforderlichen Informationen in der Datei selber und zwar vollständig, eindeutig, zugänglich und erschließbar.

Ziel beim Einsatz von PDF ist, das Erscheinungsbild eines elektronischen Dokuments unabhängig von der zur Erstellung benutzten Anwendungs-Software, der Hardware-Plattform oder dem Betriebssystem zu bewahren. PDF eignet sich daher in erster Linie für die Archivierung von Dokumenten, bei denen eine Abbildung in Papierform vorgesehen ist bzw. die den Charakter von Briefen wie Briefe und Geschäftsunterlagen haben.

Gegenüber einer rein bildlichen Darstellung (Pixeldarstellung) benötigen Dokumente im PDF-Format meist deutlich weniger Speicherplatz.

### Bildformate:

#### **TIFF:**

Das Format TIFF (Tagged Image File Format) wird zur Speicherung gerasterter Bilder verwendet. Eine TIFF-Datei besteht aus einem Datei-Header und der Bildinformation. Der Header enthält sogenannten Tags, in denen Eigenschaften des aufgenommenen Bildes gespeichert werden, wie z.B. Auflösung oder verwendete Kompressionsverfahren.

Alle gängigen Grafik und Präsentationsprogramme unterstützen das TIFF-Format. Darüber hinaus ist es auch von Archiv- und Workflow-Systeme unterstützt.

TIFF ist in komprimierter Form als Format für die Langzeitarchivierung von Bildern und Bildpräsentationen von Dokumenten geeignet. TIFF ist mit den meisten Kompressionsverfahren kompatibel. Ein verlustfreies Kompressionsverfahren wird empfohlen, um den erforderlichen Speicherbedarf zu minimieren.

## **JPEG:**

JPEG wurde von der Joint Photographic Experts Group entwickelt und ist ideal für Farb- und Graustufenbilder geeignet. JPEG kann anhand einiger Parameter unterschiedlich konfiguriert werden. Je nach Einstellung werden dann unterschiedliche Kompressionsraten erreicht. Auf den höchsten Ebenen der Kompression treten Qualitätsverluste auf.

Alle gängigen Grafik und Präsentationsprogramme unterstützen das JPEG-Format.

## **Audio- und Video-Formate:**

Bei der digitalen Verarbeitung von Audio- und Videodaten entstehen schon bei zeitlich kurzen Aufzeichnungen sehr große Datenmengen. Daher gewinnt eine effektive Kompression bei dieser Anwendung besondere Bedeutung.

Verlustfreie Kompressionsverfahren für Audio- und Videodaten erreichen derzeit jedoch nur Kompressionsraten von etwa 2:1. Gebräuchlicher sind Verfahren, die eine Kompressionsrate bis zu 200:1 erreichen, jedoch nicht verlustfrei arbeiten. Der durch die Kompression entstehende, teilweise erhebliche Datenverlust wird typischerweise in Kauf genommen, solange er mit dem menschlichen Auge bzw. Ohr nicht wahrnehmbar ist bzw. nicht als störend empfunden wird.

Im Folgenden werden einige gängige Formate vorgestellt:

## **MPEG:**

Innerhalb der ISO ist die Motion Pictures Expert Group (MPEG) für die Bearbeitung weltweiter Standards zur Kompression digitalisierter Bewegtbilder verantwortlich.

Derzeit sind drei verschiedene Verfahren bekannt:

- MPEG1: Dieses Format gibt es in drei verschiedenen Layern. Layer 3 ist in der Kurzform MP3 bekannt und als Kompression für Audiodaten verbreitet.
- MPEG2: Dieses Format ist derzeit für die Speicherung von Videodaten auf DVD in Gebrauch und als Standard akzeptiert.
- MPEG4: Dieses Format befindet sich noch in der Entwicklung und ist noch nicht abschließend standardisiert.

## **Datenbank-Formate:**

Ein gebräuchliches und relativ einfaches Dateiformat für Datenbanken ist CSV (Comma-Separated Values, siehe z.B. Wikipedia). Es wird von Konsumenten, Unternehmen und wissenschaftliche Anwendungen unterstützt. Zu den häufigsten Anwendungen gehört der Austausch tabellarische Daten zwischen Programmen, die üblicherweise inkompatible (oft proprietäre und/oder undokumentierte) Formate nutzen. Das funktioniert, weil viele Programme mindestens eine Variation von CSV als alternatives Import / Export-Format unterstützen.

"CSV" ist kein allgemeiner, klar definierter Standard, jedoch wird es im RFC 4180 grundlegend beschrieben. In der Praxis steht der Begriff "CSV" für jede Datei, die

1. Klartext enthält mit einem Zeichensatz wie ASCII, Unicode, EBCDIC oder Shift JIS
2. Aus Datensätzen besteht (in der Regel ein Datensatz pro Zeile)

3. Aufzeichnungen beinhaltet, die in Felder durch Trennzeichen (in der Regel ein einziges reserviertes Zeichen wie Komma, Semikolon, Tab, etc.) getrennt sind
4. Datensätze mit gleicher Sequenz von Feldern enthält.

Innerhalb dieser allgemeinen Grenzen sind viele Variationen im Einsatz. Deshalb sind "CSV"-Dateien oft nicht vollständig übertragbar. Dennoch sind die Unterschiede relativ gering, und viele Implementierungen können Benutzer auf einfache Weise durchführen. Wenn eine CSV-Datei von einem bestimmten Programm nicht gelesen werden, ist es oft möglich die Datei von Hand oder über eine einfache Programmierung zu ändern und so das Problem zu beheben. Somit sind CSV-Dateien in der Praxis recht gut zu importieren.

### **CAD-Formate:**

DXF (Drawing Interchange Format oder Drawing Exchange Format, siehe z.B. Wikipedia) ist ein CAD-Datei-Format von AutoDesk, das den Datenaustausch zwischen AutoCAD und anderen Programmen gewährleisten soll. Das DXF-Format wird von AutoDesk beschrieben und offen dokumentiert. Wegen der gut dokumentierten, einfachen Datenstruktur wird das DXF-Format fast ausschließlich für den programmübergreifenden Datenaustausch auch über verschiedene Betriebssysteme hinaus verwendet. Jedes heutige CAD- und CNC-Programm beherrscht den Import und Export von DXF, womit DXF als Industriestandard den kleinsten gemeinsamen Nenner aller CAD-Systeme bildet. Alle Elemente, die für technische Zeichnungen sinnvoll und implementierbar sind, werden in DXF unterstützt. Versionen von AutoCAD ab Release 10 (Oktober 1988) unterstützen sowohl ASCII- und binäre Formen von DXF. Frühere Versionen unterstützen nur ASCII.

--- Ende des Dokuments