



2.1 Dateiformate

Daten managen – 2 Datenmanipulation

Stellen wir uns vor, eine Versorgungsforscherin kriegt eine große Menge von Daten aus dem Krankenhaus zugesendet. Darunter befinden sich Excel-Tabellen, PDF-Berichte und einige komische Dateien, mit denen sie nichts anfangen kann. „.dcm“ steht als Dateiendung dabei. Öffnen lässt sich die Datei auch nicht. Sie leitet die Dateien an eine Kollegin weiter. Diese kann mit den komischen Dateien auch nichts anfangen und darüber hinaus auch die Excel-Datei nicht nutzen, da sie keine Microsoft Office Lizenz besitzt. Unsere Forscherin ist frustriert: Warum ist das mit den Formaten so kompliziert und gibt es überhaupt Dateien, mit denen alle arbeiten können?

Um teilbar und langfristig nachnutzbar zu sein, sollten Dateiformate möglichst nicht herstellerspezifisch (offen), standardisiert, langlebig, weit verbreitet, und sowohl von Menschen als auch Maschinen lesbar sein. Zusätzlich sollten die Dateiformate auch möglichst verlustfrei konvertiert werden können (mehr dazu später).

Offene Dateiformate

Ein offenes Dateiformat ist ein standardisiertes Format für digitale Daten, das öffentlich zugänglich ist und von jedem ohne rechtliche oder technische Einschränkungen verwendet, weitergegeben und implementiert werden kann. Im Gegensatz dazu sind herstellerspezifische Dateiformate, auch proprietäre Formate genannt, von einem einzelnen Unternehmen entwickelt und für die Nutzung in der eigenen Software optimierte Formate. Ohne die dazugehörige (meist kostenpflichtige) Software können proprietäre Formate häufig auch nicht geöffnet werden. Deswegen sind für kollaboratives Arbeiten, langfristige Speicherung und das Teilen von Daten offene Formate am sinnvollsten.

Dateityp	Herstellerspezifisches Format (Beispiele)	Alternatives Format
Tabelle Strukturierte Daten	Excel (.xls, .xlsx,), SPSS-Tabelle (.sav)	Comma Separated Values (.csv) Comma Separated Values (.csv), JSON (.json)
Text	Word (.doc, .docx)	Reine Textdatei (.txt), falls Formatierungen relevant sind in PDF/A (.pdf)
Präsentationsdateien Bilder/Fotos	Powerpoint (.ppt, .pptx) Photoshop (.psd) Adobe Illustrator (.ai)	PDF/A (.pdf) TIFF (.tif, .tiff), PNG (.png), SVG (.svg)
Videos	Quicktime (.mov)	MPEG-4 (.mp4)
Audio	Microsoft-Audio (.wma), Apple-Audio (.aac)	.wav, .flac

Grafiken

Digitale Grafiken lassen sich in zwei Kategorien einteilen: Raster- und Vektorgrafiken. Rastergrafiken werden in gleich großen Pixeln – also einzelnen, in einem Raster angeordneten Bildzellen dargestellt. Bei starker Vergrößerung der Bilder werden die einzelnen Pixel sichtbar und der sogenannte „Treppeneffekt“ kann die Kanten von Objekten stufenartig aussehen lassen.



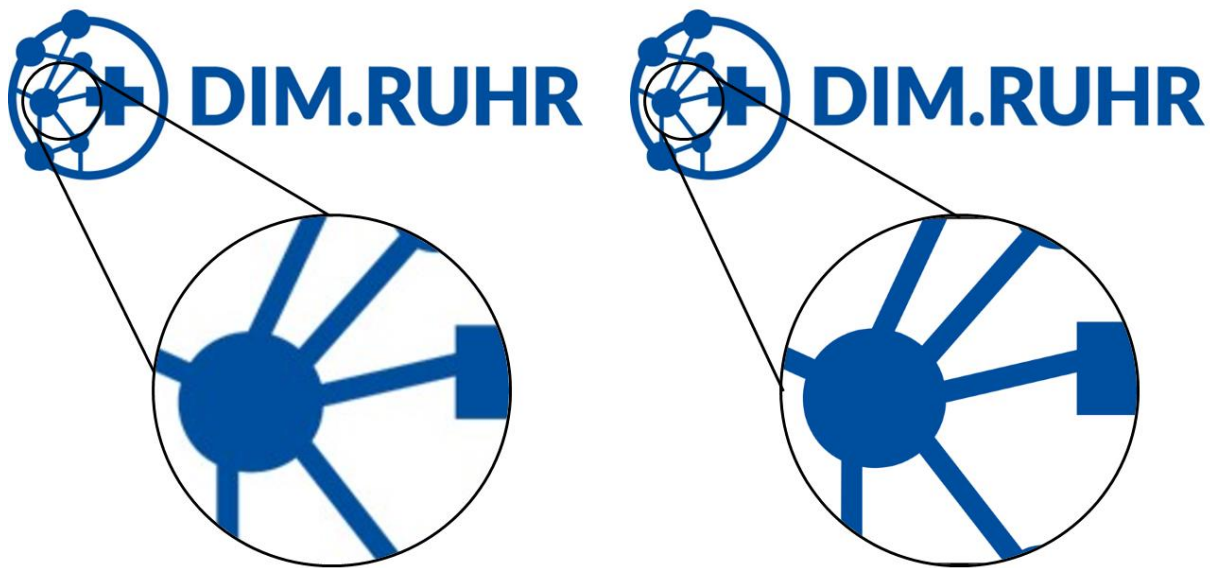


Abb. 1. Vergleich von Raster- und Vektorgrafiken bei starker Auflösung

Eines der beliebtesten Rastergrafikformate ist TIFF (Tagged Image File Format). Eine TIFF-Datei besteht neben dem kodierten Bild aus einem Header, der Titel, Farbmodell, Größe und Auflösung und ggf. weitere Bildparameter als Metadaten binär abspeichert. Es ist ein offenes und nicht-komprimiertes Format.

Ein weiteres, häufiges Format für Rastergrafiken ist JPEG. Sie werden häufig als Vorschaubilder verwendet oder für die Veröffentlichung von Bildern im Internet. Diese Bilder sind verlustbehaftet gespeichert, das heißt durch minimale Änderungen und Zusammenfassungen, die für das menschliche Auge nicht sichtbar sind, werden Informationen reduziert.

Im Kontrast sind Vektorgrafiken aus geometrischen Grundformen (Rechtecke, Kreise etc.) zusammengesetzt, die unterschiedlich groß und überlappend sein können. Die Positionen der Grundformen werden jeweils mit einem Vektor innerhalb der Grafik angegeben – daher der Name. Sie lassen sich anders als Rastergrafiken beliebig vergrößern (skalieren), ohne dass unerwünschte Bildeffekte sichtbar werden.

Ein häufiges Vektorgrafik-Format ist SVG (Scalable Vector Graphics). Bilder werden dort als XML-Dokument gespeichert und können deswegen direkt auch direkt online in HTML-Dokumente (also die Beschreibung der Struktur und des Inhalts von Webseiten) reingeschrieben werden. Sie können also rein theoretisch mit einem einfachen Texteditor erzeugt werden. Ein kostenloses Open-Source-Programm, das Vektorgrafiken erzeugen kann ist [Inkscape](https://inkscape.org/). Das SVG-Format ist vom World Wide Web Consortium (W3C) die empfohlene Auszeichnungssprache für zweidimensionale Vektorgrafiken und deswegen weit im Web verbreitet und langfristig lesbar. Sie werden verlustfrei gespeichert und benötigen vergleichsweise relativ wenig Speicherplatz.

Textformate

Das XML-Format (Extensible Markup Language) ist eine Auszeichnungssprache, welche maschinenlesbar ist. Durch sogenannte Tags können Textpassagen gegliedert und formatiert

werden. Die Tags werden mit < und > umschlossen und ein komplettes Element besteht immer aus einem Start-Tag, dem Inhalt und einem End-Tag.

Das Format bietet viele Vorteile: 1. Es ist erweiterbar, denn es können jederzeit neue Tags zum Vokabular hinzugefügt werden. 2. XML ist sowohl von Maschinen, als auch von Menschen lesbar. 3. Daten können geteilt werden ohne Datenverlust. 4. Das Format ist Software- und Hardwareunabhängig 5. Auch komplexe Daten können gespeichert werden.

Die Auszeichnungssprache HTML, welche hauptsächlich für die Entwicklung von Webseiten und Benutzeroberflächen verwendet wird, erscheint auf den ersten Blick recht ähnlich zu XML aber basiert tatsächlich nicht dieser. Sie wird manchmal auch als „Präsentationssprache“ bezeichnet. XML und HTML werden oft gemeinsam genutzt, um durch XML Daten zu speichern (z.B. auch Metadaten) und HTML die Daten zu formatieren und darzustellen. In HTML werden vordefinierte Tags verwendet, wie beispielsweise <p>, <h1> und <table>, die Textpassagen als Absätze, Überschriften und Tabellen deklarieren. Für XML-Dokumente existieren keine vordefinierten Tags und der Autor muss diese selbst definieren.

Ein kleiner HTML-Abschnitt könnte beispielsweise so aussehen:

```
1 <html>
2   <title>Meine Webseite</title>
3   <body>
4     <h1>Dateiformate- Text und Tabellen</h1>
5     <p>Die vielen verschiedenen Textformate, die es gibt, dienen unterschiedlichen Verwendungszwecken.</p>
6   </body>
7 </html>
```

Abb. 2. Beispiel HTML-Code

Und so würde die dazu passende Webseite aussehen:

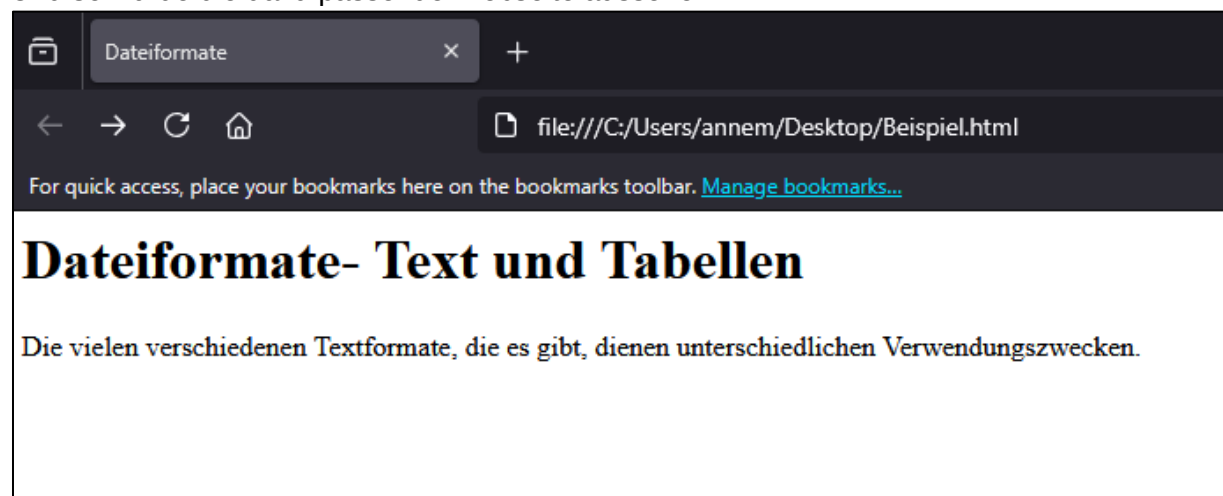


Abb. 3. Beispielwebsite entstehend aus dem HTML-Code

Ein XML-Dokument könnte so aufgebaut sein:

```
1 <?xml version="1.0"? encoding="UTF-8"?>
2 <docs>
3   <author>M. Mustermann</author>
4   <title>Musterpaper</title>
5   <journal>Musterjournal</journal>
6   <date>25-06-2025</date>
7   <description>Dies ist ein wichtiges wissenschaftliches Paper mit wichtigen wissenschaftlichen Erkenntnissen</description>
8   <language>deutsch</language>
9
10 </docs>
```

Abb. 4. Beispiel XML-Code

Auch die wichtigsten Dokumentformate im Büro wie Texte, Tabellen und Präsentationen basieren meistens auf XML. Die bekanntesten sind wahrscheinlich .docx, .xlsx und .pptx basierend auf OOXML, also Office Open XML. Ein herstellerunabhängiges und somit offenes Alternativformat ist ODF (Open Document Format von Sun Microsystems) mit den Dateiformaten .odt, .ods. und .odp. Da nicht garantiert werden kann, dass OOXML- und ODF-basierte Dateien auch in Zukunft gleich dargestellt werden können, sind sie – trotz ihrer weiten Verbreitung – nicht für die Langzeitarchivierung geeignet. Hierfür wird PDF empfohlen. Wer sich intensiver mit XML auseinandersetzen möchte findet [hier](#) Informationen.

Eine weiteres, häufig genutztes Textdateiformat ist JSON (JavaScript Object Notation). Dieses wird genutzt, um Daten strukturiert abzuspeichern. Daten werden meist als Name/Wert-Paare und Listen und gespeichert zu Objekten gekapselt. Das Format ist auch von Menschen lesbar, wird aber meistens zur weiteren Verarbeitung für Maschinen genutzt.

So könnte eine JSON-Datei aussehen:

```

1  {
2    "AutorVorname": "Max",
3    "AutorNachname": "Mustermann",
4    "Titel": "Dr.",
5    "Geburtsjahr": 1961,
6    {
7      "Titel": "Musterartikel 1",
8      "Verlag": "Musterjournal",
9      "CoAutoren": ["Jos Bleau", "Alexandra Ivanova", "Juan Perez"],
10     "veröffentlicht": true,
11   }
12 }

```

Abb. 4. Beispiel JSON-Code

Das PDF (Portable Document Format) ist eine vektorbasierte Seitenbeschreibungssprache. Hier ist die Ausgabe sowohl betriebssystem- als auch geräteunabhängig. Schriftarten und Abbildungen werden hier (meist) direkt eingebunden was zu einer stabilen Darstellung auch über langen Zeitraum führt.

Tabellen

Weitere essenzielle Dateiformate sind Tabellen, die meistens mit Tabellenkalkulationsprogrammen (z.B. Microsoft Excel) erstellt werden. Das „eigene“ Format von Excel (.xlsx) ist nur mit der entsprechenden Software lesbar. Alternativ ist das CSV-Format (Comma Separated Value) ein reines Textformat und deswegen mit jedem Texteditorprogramm lesbar. CSV-Dateien sind wie folgt aufgebaut: Eine Zeile der Datei entspricht einer Tabellenzeile. Die Werte für die Spalten werden durch Trennsymbole (sogenannte Separatoren) getrennt.

In der englischen Version sind diese (wie der Name vermuten lässt) – Kommata. Da im deutschen das Komma aber als Dezimaltrennzeichen genutzt wird, ist hier häufig ein Semikolon als Trennzeichen voreingestellt.

Die folgende Tabelle würde in der deutschen CSV-Variante also wie auf der rechten Seite dargestellt werden:

	A	B	C
1	3,2	4,5	2,7
2	4,8	5,1	4,0
3	1,9	0,8	3,3
4	3,4	5,2	3,9

```

3,2;4,5;2,7
4,8; 5,1; 4,0
1,9; 0,8; 3,3
3,4; 5,2; 3,9

```

Spezifische Dateiformate des Gesundheitswesens

Im Gesundheitswesen gibt es einige spezifische Dateitypen und Datenstandards die verwendet werden. Nachfolgend stellen wir einige der Relevantesten vor:

DICOM

Ist der internationale Standard für die Speicherung, Austausch und Übertragung von medizinischen Bildern und den zugehörigen (Meta-)Informationen. Insbesondere wird er im Bereich der radiologischen Schnittbilddiagnostik, also CT und MRT, eingesetzt. Die Dateiendung für DICOM-Dateien lautet „.dcm“. Um eine solche Datei zu öffnen wird ein spezieller DICOM-Viewer benötigt.

HL7/FHIR

HL7 ist sowohl der Name der Organisation, die Datenstandards im Gesundheitswesen entwickelt und unterstützt, als auch die Benennung zwischenzeitlicher Versionen ihrer Standards, bevor die nächste Generation mit FHIR® (Fast Healthcare Interoperability Resources) begonnen wurde. Lange Zeit war HL7 nur im klinischen Kontext verbreitet und die Praxissoftwareanbieter im niedergelassenen Bereich nutzten eigene Datenaustauschformate, wobei xDT/KVDT (mit den Schnittstellen Abrechnungsdatentransfer (ADT), Gerätedatentransfer (GDT) und Labordatentransfer (LDT)) der wahrscheinlich verbreitetste ist. Aktuell wird daran gearbeitet diese Lücke zu schließen um den Austausch zwischen klinischem und niedergelassenen Sektor zu erreichen. Genau genommen sind diese Standards kein spezifisches Dateiformat, sondern beschreiben wie die Dateien, die meist als JSON oder XML gespeichert werden, auszusehen haben.

ePA und MIO

Die elektronische Patientenakte fungiert als Dokumentenablage: Befundberichte von diagnostischen und therapeutischen Maßnahmen, Befunddaten, Laborbefunde, eArztbriefe sollen zukünftig in der ePA gespeichert werden, soweit die Patient:innen nicht widersprochen haben. Sie soll das Bindeglied zwischen den Patient:innen und zwischen den Akteur:innen im Gesundheitswesen wie Ärzt:innen, Krankenhäusern und Apotheken darstellen. Daten können von den Patient:innen selber oder aber von den Behandlern eingepflegt werden. Beispielsweise können Patient:innen ihre bereits vorliegenden Papierbefunde selbst als Scan oder Foto hochladen. Zum Stand April 2025 können nur Dokumente im Format PDF/A mit max. 25 MB pro Datei eingepflegt werden. Andere Formate wie Bilddateien sind aktuell geplant.

Einen großen Mehrwert der ePA stellt die elektronische Medikationsliste (eML) dar. Sie ist Nachfolgerin des elektronischen Medikationsplans (eMP), der auf der elektronischen Gesundheitskarte gespeichert wird. Dieser ist immernoch für schnelle Abfragen in Notfällen verfügbar. In der eML steht welche Medikamente verschrieben und auch ausgegeben wurden und ermöglicht einen intersektoralen Austausch von aktueller und historischer Medikation zwischen ambulantem und stationären Sektor, Apotheken und Patient:innen. Die Liste enthält den Handelsnamen des Medikaments, den Wirkstoff und –stärke, die Form, Dosierung, das Verordnungsdatum, von welcher Praxis/Arzt sie verordnet wurde, das Abgabedatum des Medikaments und in welcher Apotheke es ausgegeben wurde.

Standardmäßig liegt die eML in PDF-Format vor, einige PVS-Anbieter haben diese aber auch schon in die Benutzeroberfläche integriert, sodass eine Filterung und die Übernahme in die Behandlungsdokumentation möglich sind.

Die sogenannten medizinischen Informationsobjekte (MIO) sollen (textbasierte) Daten in der ePA standardisieren und semantische und syntaktische Interoperabilität ermöglichen. Hierfür wird FHIR® als anerkannter Standard genutzt. Die Daten können als JSON oder XML repräsentiert werden. Festgelegte MIOs sind zum jetzigen Stand beispielsweise der Medikationsplan und die Patientenkurzakte. Unter anderem wird aktuell an MIOs für Bildbefunde, Allergiedokumentation und Krankenhausentlassbriefe gearbeitet. Einsehbar [hier](#)

Quellenverzeichnis

Blümm, M., Fritsch, K., Bock, S., Hackenbuchner, J., Arning, U., & Förstner, K. U. (2024). 08_LE_Daten-Speichern. FDM@Studium.nrw Blended-Learning-Basiskurs „Forschungsdatenmanagement“ (Version 1.0).

https://landesinitiativefdmnrw.github.io/FDMatStudium/thk/texte/08_LE_Daten-Speichern.html

Bundesministerium für Gesundheit (2025). Die elektronische Patientenakte (ePA) für alle. Abgerufen am 30.09.2025, von

<https://www.bundesgesundheitsministerium.de/themen/digitalisierung/elektronische-patientenakte/epa-fuer-alle.html>

HL7 (2025). Welcome to FHIR®. Abgerufen am 30.09.2025, von <https://www.hl7.org/fhir/>

Kaiser, K., Rathmann, T. (2024): Grundwissen: Datenmanagement in Studium & wissenschaftlicher Praxis. FDM@Studium.nrw Selbstlernkurs (Version 1.0) (BUW)

Kassenärztliche Bundesvereinigung (2025). Elektronische Patientenakte. Abgerufen am 30.09.2025, von <https://www.kbv.de/praxis/digitalisierung/anwendungen/elektronische-patientenakte>

Kassenärztliche Bundesvereinigung (2025). Neue Kurzvideos zur elektronischen Medikationsliste. Abgerufen am 30.09.2025, von <https://www.kbv.de/praxis/tools-und-services/praxisnachrichten/2025/07-24/neue-kurzvideos-zur-elektronischen-medikationsliste>

KBr (2017). Praxissoftware: Neue Version der Behandlungs-datentransfer-Schnittstelle. Deutsches Ärzteblatt. Abgerufen am 30.09.2025, von <https://www.aerzteblatt.de/news/praxissoftware-neue-version-der-behandlungsdatentransfer-schnittstelle-9bd9a2ec-8c1f-42d0-aa14-5bd7602c453b>

Medical Imaging Technology Association (MITA) (2025). DICOM: Current Edition. Abgerufen am 30.09.2025, von <https://www.dicomstandard.org/>

Neuroth, H., Putnings, M., & Neumann, J. (2021). Praxishandbuch Forschungsdatenmanagement (p. 587). De Gruyter.

Telekonnect (2025). Elektronischer Medikationsplan/Elektronische Medikationsliste. Abgerufen am 30.09.2025, von <https://www.telekonnect.de/artikel/elektronischer-medikationsplan>

Weiterführende Info

Dezernat Digitalisierung und IT (2025). IT in der Arztpraxis: Datensatzbeschreibung KVDT. Kassenärztliche Bundesvereinigung. Abgerufen am 30.09.2025, von

https://update.kbv.de/ita-update/Abrechnung/KBV_ITA_VGEX_Datensatzbeschreibung_KVDT.pdf

Gevantmakher, M., & Meinel, C. (2020). Medizinische Bildverarbeitung - eine Übersicht. Abgerufen am 30.09.2025, von

https://hpi.de/fileadmin/user_upload/fachgebiete/meinel/papers/Old_Source/TR_Med_Bildverarbeitung.pdf

Kaiser, K., Rathmann, T. (2024): Grundwissen: Datenmanagement in Studium & wissenschaftlicher Praxis. FDM@Studium.nrw Selbstlernkurs (Version 1.0) (BUW)

Kassenärztliche Bundesvereinigung (2024). Medizinische Informationsobjekte. Abgerufen am 30.09.2025, von <https://www.kbv.de/praxis/digitalisierung/anwendungen/elektronische-patientenakte/mio>

Kassenärztliche Bundesvereinigung (2025). Die elektronische Patientenakte. KBV Praxiswissen. Abgerufen am 30.09.2025, von <https://www.kbv.de/documents/infothek/publikationen/praxiswissen/praxiswissen-epa.pdf>

Koordinationsstelle für Archivierung elektronischer Unterlagen (KOST) (2024). Katalog archivischer Dateiformate. Abgerufen am 30.09.2025, von https://kost-ceco.ch/cms/kad_text_de.html

Medical Imaging Technology Association (MITA) (2025). DICOM: Current Edition. Abgerufen am 30.09.2025, von <https://www.dicomstandard.org/>



Das Material ist im Rahmen des Projekts „DIM.RUHR: Datenkompetenzzentrum für die interprofessionelle Nutzung von Gesundheitsdaten in der Metropole Ruhr“ (Förderkennzeichen: 16DKZ2008[A-F]) entstanden. Dementsprechend spiegelt das Material in Version 1.0 den Stand von 2025 wider.

„**Daten managen - 2.1 Datenmanipulation – Dateiformate**“ von Anne Mainz und Sven Meister.

Dieses Werk und dessen Inhalt sind – sofern nicht anders angegeben – lizenziert unter CC BY 4.0. Ausgenommen aus der Lizenz sind die verwendeten Logos.

