

Wie neutral sind medienneutrale Daten?

Alexander Konzelmann

Richard Boorberg Verlag Stuttgart
Scharstr. 2, D-70563 Stuttgart
a.konzelmann@boorberg.de

Schlagnote: Content Management System, Datenhaltung, e-Publishing/elektronisches Publizieren, Medienneutralität, Multimedialität, Speicherformat, Speichermedien, Strukturmodell Fußnote, Strukturmodell Tabelle, XML

Abstract: Der Begriff der Medienneutralität ist nicht festgefügt und eindeutig. Er ist daher in seinen Kontextabhängigkeiten an Beispielen zu erläutern. Das Schlagwort von der medienneutralen Datenhaltung täuscht potenzielle Erwerber von CMS und Bildungswillige über einige Realitäten. Diese realen Mediendependenzen sollen aufgezeigt werden. Konkret bedeutet die Neutralität der Datenhaltung für das electronic publishing im juristischen Bereich die Rückführung aller notwendigen Informationen auf digitalisierte, strukturierte Textdaten im Sinne von XML oder ähnlicher Kunstsprachen.

1. Begriffliches

1.1 Daten

Daten sind – laut DIN 44300 – durch Zeichen aufgrund bekannter oder unterstellter Abmachungen dargestellte Informationen. Diese Zeichen (Codes) sind digital oder analog gespeichert. Folglich sind Daten stets mediengebunden. Solange man nicht die – nicht physikalisch nachweisbare – Existenz von „Information an sich“ postuliert, sind demnach auch Informationen mediengebunden, aber im Unterschied zu Daten sind sie nicht an ein bestimmtes Medium gebunden, sondern in ihrem Fortbestand und in ihrer Funktion nur darauf angewiesen, in irgendeinem Medium zu existieren. Sobald sie in diesem Medium codiert und damit verstehbar dargestellt sind, können sie als Daten bezeichnet werden. Schallwellen von Wörtern in der Luft, Ladungszustände in einem elektronischen Speichermedium, Schalterstellungen, elektromagnetische Wellen definierter Frequenzen und Amplituden, Zeichnungen, Buchstaben und Zahlen auf Papier sind Infor-

mationen und können Daten sein. Im geläufigen Sprachgebrauch sind Daten digital gespeicherte, in so genannte Dateien oder Datenbankfelder portionierte Informationen auf unterschiedlichsten, meist elektronischen, Datenträgern. Sie müssen codiert sein, weisen also ein so genanntes Datenformat auf, das auf Konventionen beruht. Beispiele wären die Textformate ANSI, ASCII, Unicode, RTF, Bildformate wie GIF oder TIFF, Datenbanktabelleformate wie DBase, Excel, Oracle oder Access, Strukturbeschreibungsformate wie XML, HTML und weitere SGML-Derivate, aber auch Seitenbeschreibungssprachen wie PostScript oder PDF.

1.2 Medien

Beim Begriff der Medien ist Vorsicht geboten. Man versteht Medien allgemein als Informationsverbreitungsmittel wie Druckerzeugnisse, Television, Rundfunk, Internetpräsenzen, E-Mails, akustische und audiovisuelle Datenträger wie DVDs, CDs etc. Ein Drucker versteht unter einem Medium eine bestimmte Papiersorte, einen Belichtungsfilm, eine Druckvorlagenplatte. Ein EDV-Spezialist versteht darunter unterschiedliche Speichermedien von der Lochkarte bis zum e-Papier. Andere Berufsgruppen verwenden den Begriff nochmals anders (Physiker, Parapsychologen, ...). Generell ist zwischen Speichermedium und Wiedergabemedium zu differenzieren.

1.3 Neutralität

Meist wird der Begriff „medienneutrale Datenhaltung“ nur hinsichtlich der Wiedergabemedien verwendet. Diese wiederum werden hauptsächlich bestimmt durch die derzeit verfügbaren Peripheriegeräte von Computern. Zuerst einmal wird also von insofern medienneutral gespeicherten Daten verlangt, dass sie schwarzweiß und farbig gedruckt werden können, also für Papier als Wiedergabemedium geeignet sind¹. Weiterhin wird allgemein vorausgesetzt, dass die Daten am Bildschirm oder mittels eines Projektors lesbar beziehungsweise informationsübermittelnd dargestellt werden. Das Erfordernis der Behindertengleichstellung führt dazu, dass Bildern außer-

1 Für farbige Darstellungen ist hierfür eine Beschreibung im CMYK-Farbraum für die sogenannte subtraktive Farbmischung notwendig. Für die additive Mischung von Lichtfarben auf dem selbstleuchtenden Bildschirm hingegen ist eine Beschreibung im RGB-Farbraum nötig. Verständliche Erläuterung zu dieser Unterscheidung zB auf der Seite <http://www.ppt-faq.de/html/monitordrucker.html> von Pia Bork.

dem ein Ersatztext beizufügen ist, dass eine Ausgabe von Texten mittels Braillezeile und die Lesbarkeit für einen sogenannten Screenreader gegeben ist, ein Gerät, das geschriebene Texte durch Vorlesen hörbar macht. Hingegen wird noch nicht verlangt, dass zB für Hörgeschädigte Notenumschreibungen für nicht in Textzeichen darstellbare Klänge aus mp3-Dateien als Alternativen angeboten werden. Dennoch verlangt die Medienneutralität generell, dass auch Musik in einem Format abgespeichert wird, das nicht nur als Klangdatei auf einen Lautsprecher geleitet werden kann, sondern auch in einer Art Notenschrift als strukturierter Zeichencode umrechenbar ist.² Es ist absehbar, dass sämtliche Sinnesorgane aus einer einfachen, textbasierten strukturierten Datenablagersprache heraus bedient werden müssen. Es gibt bereits Kinofilme, bei denen während der Vorführung Düfte zu versprühen sind.³ Neutralität der Datenhaltung bedeutet demnach die Rückführung aller notwendigen Informationen auf digitale, strukturierte Textdaten im Sinne von XML oder ähnlichen Kunstsprachen.

1.4 Multimedialität

Wenn man die Anforderungen an die Medienneutralität festlegen will, geht man am besten vom erwünschten Ergebnis aus. Das heißt, man fragt nach dem Zweck der Forderung. Wer sich die Mühe macht, für Interessenten zu einem bestimmten Thema Inhalte zusammenzutragen, diese EDV-mäßig zu erfassen, regelmäßig zu aktualisieren, zu indizieren und sie zu verbreiten, der möchte zumindest keine Schreib- und Fleißarbeiten mehrmals leisten. Er strebt danach, möglichst viele von seinen Interessenten nachgefragten Wiedergabemedien zu bedienen, ohne den Inhalt einmal in einen Power-Point-Vortrag einzutippen, ihn dann für einen HTML-Auftritt nochmals abzuschreiben, die Grafiken für den Ausdruck neu zu scannen, Tabellen für ein Handout in Word neu zu formatieren und eventuell harte Zeilenumbrüche aus einem Programm in einem anderen wieder löschen zu müssen etc Also ist das Ziel die Multimedialität, dh Bildschirm, Beamer, online- oder offlinebasierte Datenbankrecherche, Drucker, die Belichtungsmaschine in der Setzerei und auch ein E-Mail-Newsletter sollten mit den Inhalten bedient werden können, denn das sind die gängigen Wege zur Ver-

² Vgl hierzu <http://www.musicxml.org/xml.html>.

³ Die Phantasie des Autors hat für Anwendungen über die menschlichen Sinnesorgane hinaus auch ein Schau-Aquarium erdacht, in dem die Elektrosensoren farbiger Fische durch computergesteuerte elektromagnetische Reize so stimuliert werden, dass die Fische einer Choreographie folgen.

breitung von Fachinformationen. Insofern tut eine positive Aufzählung der erwünschten Wiedergabeformen Not, denn mit dem Ansatz „Multimedia-
lität“ kommt man meines Erachtens weiter als mit der schwer greifbaren
„Neutralität“.

2. Praktische Vorgehensweise

Die Anforderung der Medienneutralität führt bei den Informationsverwal-
tern zu einer Aufsplittung der Betrachtungsweise in externe Datenbanken
und interne Datenbank. Die interne Datenbank enthält in optisch eventuell
nicht ansprechender Form die Inhalte, editierbar und nach Gruppen oder
Stichwörtern recherchierbar und exportierbar. Sie ist nicht zur Datenreprä-
sentation, sondern nur zur Datenhaltung gedacht und enthält jede Informa-
tion nur einmal. Die externen Datenbanken sind nach den Wünschen der
Zielgruppen aufgebaut und werden aus Exporten der internen Datenbank
nach einer Konvertierung der Inhalte in eine präsentable Form bedient.

2.1 Ermittlung des Bedarfs der Zielgruppen (externe Datenbanken)

Zuerst ist – wieder vom Ergebnis her – zu ermitteln, was die Interessenten
wollen. Meist wird dies ein Buch mit einem zusätzlichen Recherchezugriff
oder eine online-Adresse mit einem komfortablen Export- und Druckmana-
gement sein. In manchen Konstellationen wird auch ein rein papierbasier-
tes oder rein elektronisches Informationsangebot das Richtige sein. Es ist
besonderer Wert auf die äußere Darstellung, die Übersichtlichkeit und Ein-
fachheit der Recherche und die rasche Ergebnisauflistung bzw – bei Druck-
erzeugnissen – auf ein gutes Schriftbild zu legen.

2.2 Festlegung der geeigneten Datenformate (interne Datenbank)

Danach ist festzulegen, wie man alle notwendigen Roh-Informationen für
die zu bedienenden externen Datenbanken ohne Redundanzen in einer
internen Datenbank unterbringt. Dabei kommt ein recht komplexes Daten-
konstrukt heraus, das ohne Erläuterung einem Außenstehenden nicht ohne
weiteres einleuchten wird. Das ist aber genau der Datenstamm, der „me-
dienneutral“ genannt wird. In Wirklichkeit ist er hoch spezifisch und nur

Experten zur Bearbeitung und Abfrage zugänglich. Insbesondere trägt der Datenbestand der internen Datenbank eine große Zahl an Metadaten mit sich, das sind Informationen über den aktuellen Datensatz und seine Herkunft, über seine Beziehungen zu anderen Inhalten, seinen sachlichen und zeitlichen Geltungsbereich und auch Hinweise zu seiner Darstellungsform in den externen Datenbanken.

2.3 Erstellung oder Definition der notwendigen Konverter

Daher ist zuletzt an Werkzeugen zu arbeiten, die die hinterlegten Informationen nach Bedarf filtern oder unterdrücken, mit Formatierungen versehen, gruppieren und für die gewählten Präsentationsformen durch Transformationen oder durch Weglassen von Interna in ansprechender Weise zielgruppengerecht aufbereiten. Üblicherweise sind dies Konverterprogramme, im Umfeld von XML zB XSL-T Scripte bzw Templates und Stilvorlagen. Erst nach der Aufbereitung sind die Daten für die Anwender praktisch brauchbar. Solange aber die Aufbereitung automatisiert und ohne erneute Inhaltsskorrekturen bewerkstelligt werden kann, kann man davon sprechen, dass die Ausgangsdaten im Lichte ihrer Zweckbestimmungen „medienneutral“ sind.

3. Beispiele aus der Rechtsinformatik

Wenden wir nun die genannten Prinzipien auf e-Publishing-Aufgaben für Juristen an. Die Rechtsinformatik hat es typischerweise mit Datensammlungen einer begrenzten Anzahl von Textsorten zu tun. Häufig treten auf: „Gerichtsentscheidung“, „Anwaltsschreiben“, „Akte“, „Rechtsvorschrift“, „Lehrbuch“, „Fachaufsatz“, „Formular“, „Kommentar“, „Gutachten“. Hinzu kommen die verschiedenen zugehörigen Metadaten-sammlungen in Form von Indizes über Aktenzeichen, Autorennamen, Spruchkörperbezeichnungen, Fundstellen, persönliche Stammdaten, bibliographische Katalogisierungen, Kalenderdaten, bezogene Rechtsvorschriften, zuständige Stellen. Als Zielgruppen hat der Jurist, der e-Publishing betreibt, Angehörige beratender Berufe sowie staatliche Stellen, jeweils mit einem Schwerpunkt im Bereich „lesende Tätigkeit“ zu berücksichtigen. Die Anforderungen an Datenbanken sind in diesen Zielgruppen vor allem auf schnelle Ergebniseingrenzung und primitive Bedienbarkeit ausgerichtet,

wie zB die wiederholten Fragebogenaktionen von *Liebwald*⁴ gezeigt haben. Dies ist zwar eine ordentliche Vielfalt, schränkt aber dennoch die Aufgabenstellung der „Neutralisierung“ dieser Informationen bei ihrer EDV-mäßigen Speicherung ein und vereinfacht durch diese Konkretisierung die Entscheidungsfindung und die Festlegung auf eine Reihe von XML-Definitionen zur Beherrschung der juristischen Artenvielfalt. Aus diesem Aufgabenspektrum nun zwei Beispiele.

3.1 Fußnoten

Fußnoten sind meist Zitate, Fundstellenquerverweise und Exkurse. Sie müssen für die Darstellung in Papier gewohnheitsmäßig in eine Fußnotenreferenz und den Fußnoteninhalt zerteilt werden. Für die Bildschirmrepräsentation gibt es spätestens seit HTML die Anforderung, dass sie als klickfähiger Link auszugestalten sind, wobei die Absprungstelle zugänglich bleibt. Um dieses doppelte Ziel zu erreichen, ohne dass eine Fußnote zweimal geschrieben werden muss, ist ihr in der Datenhaltung sowohl das elektronische Linkziel als URL/URI zuzuschreiben als auch in einem Stylesheet (Konverter) die notwendige Information zu hinterlegen, um die Fußnote in Referenz und den Volltext zu zerlegen. Hinzu kommt, dass zB auf Papier erwartet wird, dass wenn eine identische Fußnote zweimal auf derselben Seite referenziert wird, diese nur einmal am Seitenende respektive Beitragsende abgedruckt wird. Praktisch wird die Fußnote mit einem eindeutigen Markup an der Stelle in den Fließtext eingefügt, an dem später auf Papier das Fußnotenzeichen stehen soll. Zusätzlich wird bei echten Querverweisen die Ziel-URL als Attributwert angegeben, dessen Ausgabe beim Print nach Bedarf unterdrückbar wird.

Das bedeutet aber gleichzeitig, dass der abgelegte „neutrale“ Text ohne die Verarbeitung mit den Konvertern für Satz oder Bildschirmrepräsentation keinem der Ansprüche genügt.

3.2 Tabellarische Darstellungen

Wie werden Tabellen medienneutral dargestellt oder abgespeichert?⁵ Tabellen bestehen oberflächlich betrachtet aus sich kreuzenden Spalten und Zeilen. Damit erscheinen sie notwendig zweidimensional und an der Grenz-

4 zuletzt Tagungsband IRIS 2005, 329 ff.

5 lehrreich *Riedel* in: <https://www.tu-chemnitz.de/urz/lehre/doge/rsrc/all-papier.pdf>.

linie von Text zur Grafik. Wenn sie aber von einem Computerprogramm in ein noch nicht bekanntes anderes transportiert werden müssen, dann verbietet sich eine Abspeicherung in einem Format, das nicht ein sequenzielles Textformat ist. Das bedeutet, dass entweder auf die Spalten oder auf die Zeilen im Modell verzichtet werden muss. Die bekanntesten „neutralen“ (besser: plattformübergreifend verwendbaren) Tabellenformate HTML und CALS verzichten auf die Spalten. Das bedeutet, eine Tabelle wird gedanklich und auch tatsächlich definiert als eine (senkrechte) Abfolge von Zeilen, die ihrerseits jeweils eine (waagerechte) Abfolge von Zellen darstellen. Jeder einzelnen Zelle kann nun als Attribut die Nummer der Spalte zugeschrieben werden, zu der die Zelle gehört. Bei Tabellen, die kein regelmäßiges Gitternetz darstellen, ist dies sogar notwendig. Aber insgesamt wird eine Tabelle dadurch universell transformierbar und damit quasi medienneutral, dass man sie wie folgt notiert: Start einer Zeile, Start der ersten Zelle dieser Zeile, Start der nächsten Zelle in derselben Zeile, Start einer neuen Zeile, etc So kann sie zB auch ein Screenreader für Sehbehinderte verständlich vorlesen.

4. Medienabhängigkeit der internen Datenbanken

Da Daten immer an ein Speichermedium gebunden sind (so 1), wird die Datenhaltung selbst niemals medienneutral werden. Letztlich sind immer Entscheidungen zu treffen, zum Beispiel anhand von Fragen wie: Müssen die Daten aus Sicherheitsgründen in redundanten aber voneinander im Prinzip unabhängigen Ablageformen mit gegenseitiger Aktualisierung mehrfach gespeichert werden (Beispiel: Komplettausdruck im Leitzordner mit eingehafteter DVD-R oder mit Verweis auf den Fundort einer externen Festplatte)? Sind bestimmte Lesegeräte aufzubewahren, damit zB fünfeinviertel-Zoll-Floppies lesbar bleiben oder müssen mit dem Aufkommen neuer Speichermedien immer alle Altbestände kompatibel zu diesen neu abgespeichert werden? Müssen alte Kataloge und Indizes in die neuen einbezogen werden, damit eine Gesamtrecherche durch einen einheitlichen Bestand hindurch möglich bleibt? Ist für die Umsetzung der Inhalte in Zeichenformate außerhalb unseres Rechtskreises Vorsorge zu treffen? Immerhin hat zB Japan einst das deutsche BGB teilweise rezipiert.⁶

⁶ Quelltext einer kurzen chinesischen XML-Datei, die sicherlich in China als medienneutral gilt, bezeichnenderweise als Grafik, unter der URL <http://www.jeckle.de/images/xml/chinExample.gif>.

Auch wenn alle Daten in einer sequenziellen, sowohl menschen- als auch maschinenlesbaren, definierten und dokumentierten Markup-Sprache wie XML abgelegt sind, ist keine echte Medienneutralität erreicht, sondern lediglich eine medienübergreifende Wiedergabemöglichkeit.⁷ Für das Speichern und für das Konvertieren in die jeweiligen Wiedergabemedien bzw in die verkaufbaren externen Datenbanken hinein bleiben Festlegungen, technische Abhängigkeiten und notwendige Normierungen Pflicht. Dies ist kein Nachteil, sondern ergibt sich aus den praktischen Anforderungen.

5. Nur ein Schlagwort

Der Befund zeigt, dass der Begriff der medienneutralen Datenhaltung kein wissenschaftlich zitierfähiger ist, sondern ein Schlagwort, dessen Bedeutung im Einzelfall von jedem, der für eine Datensammlung verantwortlich ist, genau zu analysieren ist. Es sei im e-Publishing davor gewarnt, sich beispielsweise ein bestimmtes Content Management System⁸ zuzulegen, weil dieses ermöglichen soll, alle notwendigen Daten medienneutral vorzuhalten. Vielmehr hat man zu prüfen, ob die Ausgabeschnittstellen den Anforderungen seiner Zielgruppen entsprechen und die Eingabe- und Verwaltungsschnittstellen der Qualifikation seiner Mitstreiter. Erst danach ist die ebenfalls elementare Frage zu stellen, ob das System intern mit XML arbeitet oder wenigstens einen automatischen XML-Backup der Gesamtinhalte erstellt. Wenn ja, dann ist es so weit medienneutral oder medienübergreifend, wie es für die gestellte Aufgabe sein muss. Denn diese Neutralität ist kein Selbstzweck.

⁷ siehe auch http://de.wikipedia.org/wiki/Medienneutrale_Datenhaltung

⁸ Ein solches „CMS“ integriert die gemeinschaftliche Inhaltssammlung, -aktualisierung und -aufbereitung in der internen Datenbank einerseits und die Inhaltsverwertung durch Schnittstellen zu externen Datenbanken andererseits, wobei stets nur auf eine Datenquelle zurückgegriffen wird und über Benutzerrechte und Workflowverwaltung eine Ablaufsteuerung sichergestellt wird. Vgl http://de.wikipedia.org/wiki/Content_Management_System.